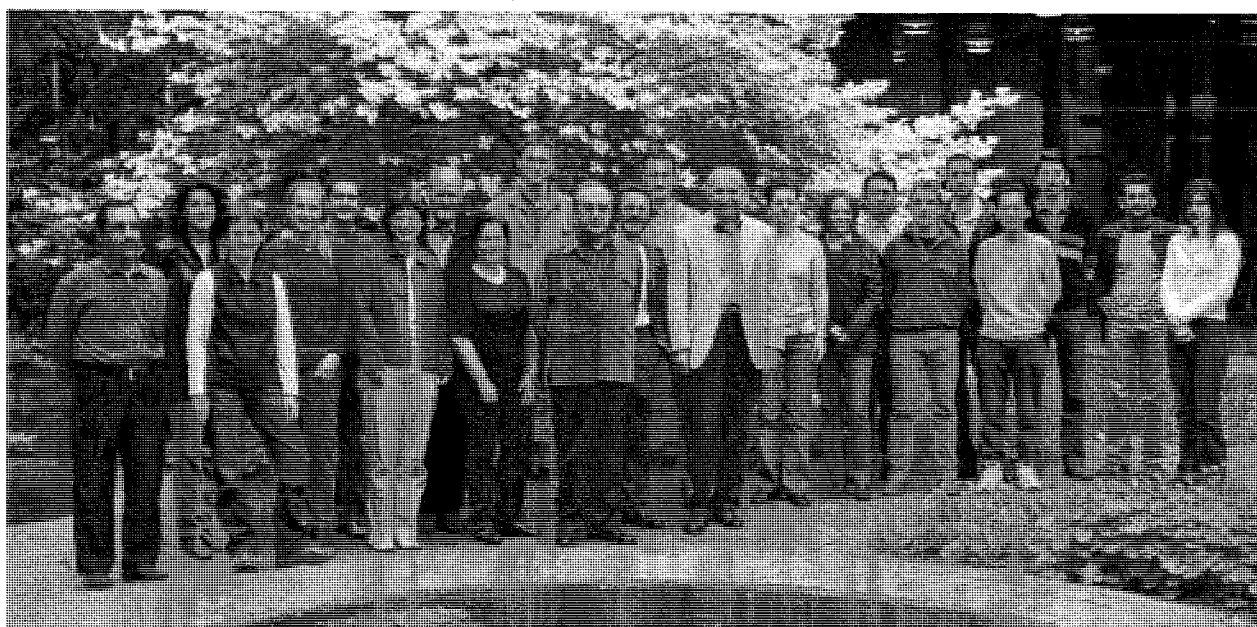


Eidgenössische Technische Hochschule Zürich

Institut für Geodäsie
und Photogrammetrie

Bericht

298



- Professur Photogrammetrie -

**Beiträge zum
XX. Kongress der ISPRS in Istanbul
12.-23. Juli 2004**

A. Grün, Herausgeber

Dezember 2004

Eidgenössische Technische Hochschule Zürich

Institut für Geodäsie
und Photogrammetrie

Bericht

298

- Professur Photogrammetrie –

**Beiträge zum
XX. Kongress der ISPRS in Istanbul
12.-23. Juli 2004**

A. Grün, Herausgeber

Dezember 2004

IGP Bericht Nr. 298

- Professur Photogrammetrie –
Beiträge zum XX Kongress der ISPRS in Istanbul
12.-23. Juli 2004

A. Grün, Herausgeber

© 2004
Institut für Geodäsie und Photogrammetrie
an der Eidg. Technischen Hochschule Zürich
ETH Hönggerberg
CH-8093 Zürich

Alle Rechte vorbehalten

ISBN 3-906467-47-3

Vorwort

Der XX. Kongress der ISPRS in Istanbul (12.-23. Juli 2004) stand unter dem Motto „Geo-Imagery Bridging Continents“. Dieses Leitwort steht stellvertretend für verschiedene Integrationsprozesse, die sich seit einigen Jahren in den Disziplinen, auf die sich die ISPRS abstützt (Photogrammetrie, Fernerkundung, GIS), entwickeln:

- (a) *Vertikale Integration*. Die technologische Basis hat sich, zumindest in Forschung und Entwicklung, ganz klar vereinheitlicht. Heute sind nur noch die digitalen Verfahren relevant, bei Datenakquisition, Administration, Verarbeitung, Analyse und Repräsentation. Damit hat sich das Arbeits- und auch das Methodenumfeld sehr stark homogenisiert. Die Vielfalt der analogen und analytischen Lösungen ist verschwunden. Strenge, hochentwickelte Verfahren benutzen schlussendlich das gleiche Umfeld wie Einfachmethoden. Ein photogrammetrisches Forschungs- und Entwicklungslabor ist heute kaum mehr von einem Informatiklabor zu unterscheiden.
- (b) *Horizontale Integration*. Auch die früheren Barrieren zwischen den Disziplinen bestehen nicht mehr. Die Verfahren und Arbeitswerkzeuge der Photogrammetrie und Fernerkundung nähern sich immer stärker aneinander an. Funktionen für Bilddatenanalyse und Visualisierung sind auch für moderne GIS unverzichtbar. Heute sind Softwareplattformen denkbar, die alle Funktionalitäten für Photogrammetrie, Fernerkundung und GIS gleichzeitig bereitstellen.
- (c) *Interdisziplinarität*. Beiträge zu unseren Wissenschaftsbereichen kommen heute aus einer Vielzahl von Disziplinen. Erfolgreiches Arbeiten erfordert auch Kompetenzen im Bereich inter- und transdisziplinärer Verhaltensweisen. Multidisziplinär zusammengesetzte Forschungsteams sind heute die Regel.
- (d) *Applikations- und Markterweiterung*. Die digitalen Verfahren in Photogrammetrie, Fernerkundung und GIS erlauben es uns heute, vollkommen neue Applikationsbereiche zu bedienen. Es ist zu erwarten, dass sich dieser Trend in Zukunft sogar verstärken wird.

Diese Entwicklungen und Tatbestände werden auch klar in den Arbeiten unserer Gruppe, die wir zum Kongress vorlegten, reflektiert:

- + Grundlegende algorithmische Arbeiten zur Koregistrierung von 3D Oberflächen
- + Modellierung, Kalibrierung und Validierung von neuen Sensoren (Mehrlinienscanner und Panoramakameras)
- + Entwicklung neuer Verfahren zur automatischen DSM-Generierung und Strassenextraktion
- + Untersuchung von Verfahren der automatischen Aerotriangulation und DSM/DTM-Generierung
- + Studien zur Genauigkeit der Auswertungen von hochauflösenden Satellitenbildern
- + Beiträge zur Dokumentation und Analyse von Cultural Heritage – Objekten
- + Entwicklungen in neuen Anwendungsgebieten wie 3D Wolkenmodellierung und Extraktion von Personen aus Videosequenzen

Hier möchte ich deutlich und dankbar auf die Beiträge externer Autoren zu unseren Projekten hinweisen.

Daneben möchte ich auch betonen, dass die Feierlichkeiten zu meinem 60sten Geburtstag, welche meine Mitarbeiterinnen und Mitarbeiter in Istanbul organisierten, mir unglaublich viel Vergnügen bereiteten.

Dafür und für die bemerkenswerten wissenschaftlichen Leistungen möchte ich mich ganz herzlich bei allen bedanken.

Inhaltsverzeichnis

	Seite
Akca, D.: A new algorithm for 3D surface matching	1
Amiri Parian, J., Gruen, A.: An advanced sensor model for panoramic cameras	7
Baltsavias, E., O'Sullivan, L., and L. Zhang, L.: Automated road extraction and updating using the ATOMI system - performance comparison between aerial film, ADS40, Ikonos and Quickbird orthoimagery	13
Eisenbeiss, H., Baltsavias, E., Pateraki, M., Zhang, L.: Potential of Ikonos and Quickbird imagery for accurate 3D point positioning, orthoimage and DSM generation	21
Gruen, A., Remondino, F., Zhang, L.: 3D modeling and visualization of large cultural heritage sites at very high resolution: the Bamiyan valley and its standing Buddhas	29
Niederöst, J.: Image analysis for the history of cartography: Drawing conclusions from the evaluation of Pfyffer's Relief	35
Pateraki, M., Baltsavias, E.: Surface discontinuity modeling by LSM through patch adaptation and use of edges	41
Pateraki, M., Baltsavias, E., Recke, E.: Experiences on automatic image matching for DSM generation with ADS40 pushbroom sensor data	47
Poli, D.: Orientation of satellite and airborne imagery from multi-line pushbroom sensors with a rigorous sensor model	53
Poli, D., Zhang, L., A. Gruen, A.: SPOT-5/HRS stereo images orientation and automated DSM generation	59
Remondino, F.: Character reconstruction and animation from monocular sequence of images	71
Remondino, F., Chen, T.: 35 years of internet, 10 years of ISPRS online	77
Roditakis, A.: Modeling and visualization of clouds from real world measurements ..	89
Sauerbier, M.: Accuracy of automated aerotriangulation and DTM generation for low textured imagery	95
Seiz, G., Poli, D., Gruen, A., Baltsavias, E., Roditakis, A.: Satellite- and ground-based multi-view photogrammetric determination of 3D cloud geometry	101
Talukdar, K.: Extraction and classification of wetland features through fusion of remote sensing images in the Okavango delta, Botswana	109
Tsuno, K., Murai, S., Shibasaki, R., Gruen, A., Zhang, L.: Starimager - a new airborne three-line scanner for large-scale applications	115
Wolff, K.: Approximation of non projective mapping and their efficient application for a geometrically based 3D point determination using multiple views	121
Zhang, Ch., Baltsavias, E.: An operational system for automated road database updating from aerial imagery	127
Zhang, L., Gruen, A.: Automatic DSM generation from linear array imagery data....	133

A NEW ALGORITHM FOR 3D SURFACE MATCHING

D. Akca

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) Zurich
ETH Hoenggerberg, CH-8093 Zurich, Switzerland. E-mail: devrim.akca@geod.baug.ethz.ch

KEY WORDS: Surface matching, Least Squares Matching, Point clouds, Registration, Laser scanning.

ABSTRACT:

A new algorithm for least squares matching of overlapping 3D surfaces, digitized/sampled point by point using a laser scanner device, the photogrammetric method or other techniques, is proposed. In photogrammetry, the problem statement of surface patch matching and its solution method was first addressed by Gruen (1985a) as a straight application of Least Squares Matching. There have been some studies on the absolute orientation of stereo models using DEMs as control information. These works have been known as *DEM matching*. Furthermore, techniques for 2.5D DEM surface matching have been developed, which correspond mathematically with least squares image matching. 2.5D surfaces have limited value, especially in close range applications. Our proposed method estimates the transformation parameters between two or more fully 3D surface patches, minimizing the Euclidean distances instead of Z-differences between the surfaces by least squares. This formulation gives the opportunity of matching arbitrarily oriented surface patches. An observation equation is written for each surface element on the template surface patch, i.e. for each sampled point. The geometric relationship between the conjugate surface patches is defined as a 7-parameter 3D similarity transformation. The Least Squares observations of the adjustment are defined by the observation vector whose elements are Euclidean distances between the template and search surface elements. The unknown transformation parameters are treated as stochastic quantities using proper weights. This extension of the functional model gives control over the estimation parameters. The details of the mathematical modelling of the proposed method, the convergence behavior, and statistical analysis of the theoretical precision of the estimated parameters are explained. Furthermore, some experimental results based on registration of close-range laser scanner and photogrammetric point clouds are presented. This new surface matching technique derives its mathematical strength from the least squares image matching concept and offers high level flexibility for any kind of 3D surface correspondence problem, as well as statistical tools for the analysis of the quality of the final results.

1. INTRODUCTION

Laser scanners can measure directly 3D coordinates of huge amounts of points in a short time period. Since the laser scanner is a line-of-sight instrument, in many cases the object has to be scanned from different viewpoints in order to completely reconstruct it. Because each scan has its own local coordinate system, all the local point clouds must be transformed into a common coordinate system. This procedure is usually referred to as registration. Actually the registration is not a specific problem to the laser scanner domain. Since the problem is more general than the given definition, the emphasis of our work is to investigate the most general solution of the registration problem on a theoretical basis.

In the past, several efforts have been made concerning the registration of 3D point clouds, especially in the Computer Vision area. One of the most popular methods is the Iterative Closest Point (ICP) algorithm developed by Besl and McKay (1992), Chen and Medioni (1992), and Zhang (1994). The ICP is based on the search of pairs of nearest points in the two sets, and estimating the rigid transformation, which aligns them. Then, the rigid transformation is applied to the points of one set, and the procedure is iterated until convergence. The ICP assumes that one point set is a subset of the other. When this assumption is not valid, false matches are created, that negatively influence the convergence of the ICP to the correct solution (Fusiello et al., 2002). Several variations and improvements of the ICP method have been made (Masuda and Yokoya, 1995, Bergevin et al., 1996), but several problems still remain. From a computational expense point of view it is highly time consuming due to the exhaustive search for the nearest point (Sequeira, et al., 1999). Another problem is that it requires every point in one surface to have a corresponding point on the other surface. An alternative approach to this

search problem was proposed by Chen and Medioni (1992). They used the distance between the surfaces in the direction normal to the first surface as a registration evaluation function instead of point-to-nearest point distance. This idea was originally proposed by Potmesil (1983). In (Dorai et al., 1997) the method of Chen and Medioni was extended to an optimal weighted least-squares framework. Zhang (1994) proposed a thresholding technique using robust statistics to limit the maximum distance between points. Masuda and Yokoya (1995) used the ICP with random sampling and least median square error measurement that is robust to a partially overlapping scene. Okatani and Deguchi (2002) propose the best transformation of two range images to align each other by taking into account the measurement error properties, which are mainly dependent on both the viewing direction and the distance to the object surface. The ICP algorithm always converges monotonically to a local minimum with respect to the mean-square distance objective function (Besl and McKay, 1992). Even if good initial approximations for the transformation parameters are provided, in some cases it might converge to a wrong solution due to its closest point (or tangent plane) search scheme. It does not use the local surface gradients in order to direct the solution to a global minimum. Another deficiency of the ICP method is to be not able to handle multi-scale range data. Several reviews and comparison studies about the ICP variant methods are available in the literature (Jokinen and Haggren, 1998, Williams et al., 1999, Campbell and Flynn, 2001).

Since 3D point clouds derived by any method or device represent the object surface, the problem should be defined as a surface matching problem. In Photogrammetry, the problem statement of surface patch matching and its solution method was first addressed by Gruen (1985a) as a straight extension of Least Squares Matching (LSM).

There have been some studies on the absolute orientation of stereo models using DEMs as control information. This work is known as *DEM matching*. The absolute orientation of the models using DTMs as control information was first proposed by Ebner and Mueller (1986), and Ebner and Strunz (1988). Afterwards, the functional model of DEM matching has been formulated by Rosenholm and Torlegard (1988). This method basically estimates the 3D similarity transformation parameters between two DEM patches, minimizing the least square differences along the Z axes. Several applications of DEM matching have been reported (Karras and Petsa, 1993, Pilgrim, 1996, Mitchell and Chadwick, 1999, Xu and Li, 2000). Maas (2000) successfully applied a similar method to register airborne laser scanner strips, among which vertical and horizontal discrepancies generally show up due to GPS/INS accuracy problems. Another similar method has been presented for registering surfaces acquired using different methods, in particular, laser altimetry and photogrammetry (Postolov, Krupnik, and McIntosh, 1999). Furthermore, techniques for 2.5D DEM surface matching have been developed, which correspond mathematically with Least Squares Image Matching. The DEM matching concept can only be applied to 2.5D surfaces, whose analytic function is described in the explicit form, i.e. $z = f(x,y)$. Of course, this formulation has several problems in the matching of solid (3D) surfaces.

Although the registration of 3D point clouds is a very active research area in both Computer Vision and Photogrammetry, there is not such a method that has a complete capability to the following three properties: matching of multi-scale data sets, matching of real 3D surfaces without any limitation, fitting the physical reality of the problem statement as good as possible. The proposed work completely meets these requirements.

The Least Squares Matching concept had been applied to many different types of measurement and feature extraction problems due to its high level of flexibility and its powerful mathematical model: Adaptive Least Squares Image Matching (Gruen, 1984, Gruen, 1985a), Geometrically Constrained Multiphoto Matching (Gruen and Baltsavias, 1988), Image Edge Matching (Gruen and Stallmann, 1991), Multiple Patch Matching with 2D images (Gruen, 1985b), Multiple Cuboid (voxel) Matching with 3D images (Maas, 1994, Maas and Gruen, 1995), Globally Enforced Least Squares Template Matching (Gruen and Agouris, 1994), Least Squares B-spline Snakes (Gruen and Li, 1996). For a detailed survey the author refers to (Gruen, 1996). If 3D point clouds derived by any device or method represent an object surface, the problem should be defined as a surface matching problem instead of the 3D point cloud matching. In particular, we treat it as least squares matching of overlapping 3D surfaces, which are digitized/sampled point by point using a laser scanner device, the photogrammetric method or other surface measurement techniques. This definition allows us to find a more general solution for the problem as well as to establish a mathematical model in the context of LSM.

Our proposed method, Least Squares 3D Surface Matching (LS3D), estimates the 3D transformation parameters between two or more fully 3D surface patches, minimizing the Euclidean distances between the surfaces by least squares. This formulation gives the opportunity of matching arbitrarily oriented 3D surface patches. An observation equation is written for each surface element on the template surface patch, i.e. for each sampled point. The geometric relationship between the conjugate surface patches is defined as a 7-parameter 3D similarity transformation. This parameter space can be extended

or reduced, as the situation demands it. The constant term of the adjustment is given by the observation vector whose elements are Euclidean distances between the template and search surface elements. Since the functional model is non-linear, the solution is iteratively approaching to a global minimum. The unknown transformation parameters are treated as stochastic quantities using proper weights. This extension of the mathematical model gives control over the estimation parameters. The details of the mathematical modeling of the proposed method, the convergence behaviour, and the statistical analysis of the theoretical precision of the estimated parameters are explained in the following section. The experimental results based on registration of close-range laser scanner and photogrammetric point clouds are presented in the third section. The conclusions are given in the last section.

2. LEAST SQUARES 3D SURFACE MATCHING

2.1 The Estimation Model

Assume that two different surfaces of the same object are digitized/sampled point by point, at different times (temporally) or from different viewpoints (spatially). $f(x,y,z)$ and $g(x,y,z)$ are conjugate regions of the object in the *left* and *right* surfaces respectively. The problem statement is finding the correspondent part of the *template* surface patch $f(x,y,z)$ on the *search* surface $g(x,y,z)$.

$$f(x,y,z) = g(x,y,z) \quad (1)$$

According to Equation (1) each surface element on the template surface patch $f(x,y,z)$ has an exact correspondent surface element on the search surface patch $g(x,y,z)$, if both of the surface patches would be continuous surfaces. In order to model the random errors, which come from the sensor, environmental conditions or measurement method, a true error vector $e(x,y,z)$ has to be added.

$$f(x,y,z) - e(x,y,z) = g(x,y,z) \quad (2)$$

The matching is achieved by minimizing a goal function, which measures the Euclidean distances between the template and the search surface elements. Equation (2) is considered observation equations, which functionally relate the observations $f(x,y,z)$ to the parameters of $g(x,y,z)$. The final location is estimated with respect to an initial position of $g(x,y,z)$, the approximation of the conjugate search surface patch $g^0(x,y,z)$.

To express the geometric relationship between the conjugate surface patches, a 7-parameter 3D similarity transformation is used. Depending on the deformation between the template and the search surfaces, the geometric relationship could be defined using any other type of 3D transformation methods, e.g. 12-parameter affine, 24-parameter tri-linear, or 30-parameter quadratic family of transformations.

$$\begin{aligned} x &= t_x + m(r_{11}x_0 + r_{12}y_0 + r_{13}z_0) \\ y &= t_y + m(r_{21}x_0 + r_{22}y_0 + r_{23}z_0) \\ z &= t_z + m(r_{31}x_0 + r_{32}y_0 + r_{33}z_0) \end{aligned} \quad (3)$$

where $r_{ij} = R(\omega, \phi, \kappa)$ are the elements of the orthogonal rotation matrix, $[t_x \ t_y \ t_z]^T$ is the translation vector, and m is the central dilation.

In order to perform least squares estimation, Equation (2) must be linearized by Taylor expansion, ignoring 2nd and higher order terms.

$$f(x, y, z) - e(x, y, z) = g^0(x, y, z) + \frac{\partial g^0(x, y, z)}{\partial x} dx + \frac{\partial g^0(x, y, z)}{\partial y} dy + \frac{\partial g^0(x, y, z)}{\partial z} dz \quad (4)$$

where

$$dx = \frac{\partial x}{\partial p_i} dp_i, \quad dy = \frac{\partial y}{\partial p_i} dp_i, \quad dz = \frac{\partial z}{\partial p_i} dp_i \quad (5)$$

where $p_i \in \{t_x, t_y, t_z, m, \omega, \phi, \kappa\}$ is the i -th transformation parameter in Equation (3). Differentiation of Equation (3) gives:

$$\begin{aligned} dx &= dt_x + a_{10} dm + a_{11} d\omega + a_{12} d\phi + a_{13} d\kappa \\ dy &= dt_y + a_{20} dm + a_{21} d\omega + a_{22} d\phi + a_{23} d\kappa \\ dz &= dt_z + a_{30} dm + a_{31} d\omega + a_{32} d\phi + a_{33} d\kappa \end{aligned} \quad (6)$$

where a_{ij} are the coefficient terms. In the context of adjustment of observation equations, each measurement is related with the function whose variables are unknown parameters. This function constitutes the functional model of the whole mathematical model. In the following definition, the terms $\{g_x, g_y, g_z\}$ are 1st derivatives of this function, which is itself of the search surface patch $g(x, y, z)$. In other words these terms are local surface gradients on the search surface. Using the following notation,

$$g_x = \frac{\partial g^0(x, y, z)}{\partial x}, \quad g_y = \frac{\partial g^0(x, y, z)}{\partial y}, \quad g_z = \frac{\partial g^0(x, y, z)}{\partial z} \quad (7)$$

and substituting Equations (6), Equation (4) gives the following equation:

$$\begin{aligned} -e(x, y, z) &= g_x dt_x + g_y dt_y + g_z dt_z + (g_x a_{10} + g_y a_{20} + g_z a_{30}) dm + \\ &\quad (g_x a_{11} + g_y a_{21} + g_z a_{31}) d\omega + \\ &\quad (g_x a_{12} + g_y a_{22} + g_z a_{32}) d\phi + \\ &\quad (g_x a_{13} + g_y a_{23} + g_z a_{33}) d\kappa - (f(x, y, z) - g^0(x, y, z)) \end{aligned} \quad (8)$$

In matrix notation

$$-e = A x - \ell, \quad P \quad (9)$$

where A is the design matrix, $x^T = [dt_x \ dt_y \ dt_z \ dm \ d\omega \ d\phi \ d\kappa]$ is the parameter vector, and $\ell = f(x, y, z) - g^0(x, y, z)$ is the observation vector that consists of the Euclidean distances between the transformed point using current transformation parameters and its coincident surface element on the other surface. With the statistical expectation operator $E\{\}$ and the assumptions

$$e \sim N(0, \sigma_0^2 Q_{II}), \quad \sigma_0^2 Q_{II} = \sigma_0^2 P_{II}^{-1} = K_{II} = E\{ee^T\} \quad (10)$$

the system (9) and (10) is a Gauß-Markov estimation model.

The unknown 3D similarity transformation parameters are treated as stochastic quantities using proper weights. This extension gives advantages of control over the estimating parameters (Gruen, 1986). In the case of poor initial approximations for unknowns or badly distributed 3D points along the principal component axes of the surface, some of the unknowns, especially the scale factor m , may converge to a wrong solution, even if the scale factors between the surface patches are same.

$$-e_b = I x - \ell_b, \quad P_b \quad (11)$$

The least squares solution of the joint system Equations (9) and (11) gives the unbiased minimum variance estimation for the parameters

$$\hat{x} = (A^T P A + P_b)^{-1} (A^T P \ell + P_b \ell_b) \quad \text{solution vector} \quad (12)$$

$$\hat{\sigma}_0^2 = \frac{v^T P v + v_b^T P_b v_b}{r} \quad \text{variance factor} \quad (13)$$

$$v = A \hat{x} - \ell \quad \text{residual vector for surface observations} \quad (14)$$

$$v_b = I \hat{x} - \ell_b \quad \text{residual vector for additional observations} \quad (15)$$

where $\hat{}$ stands for the Least Squares (LS) Estimator. The function values $g(x, y, z)$ in Equation (2) are actually stochastic quantities. This fact is neglected here to allow the use of the Gauß-Markov model and to avoid unnecessary complications, as typically done in LSM (Gruen, 1985a).

Since the functional model is non-linear, the solution iteratively approaches to a global minimum. In the first iteration the initial approximations for the parameters must be provided:

$$p_i^0 \in \{t_x^0, t_y^0, t_z^0, m^0, \omega^0, \phi^0, \kappa^0\} \quad (16)$$

The iteration stops if each element of the alteration vector \hat{x} in Equation (12) falls below a certain limit:

$$|dp_i| < c_i, \quad i = \{1, 2, \dots, 7\} \quad (17)$$

The theoretical precision of the estimated parameters can be evaluated by means of the covariance matrix

$$K_{xx} = \hat{\sigma}_0^2 Q_{xx} = \hat{\sigma}_0^2 N^{-1} = \hat{\sigma}_0^2 (A^T P A + P_b)^{-1} \quad (18)$$

In a least squares adjustment of indirect observations whose functional model is non-linear, the 1st derivatives (2nd and higher order terms are generally neglected in the Taylor expansion) with respect to unknowns are very important terms, since they direct the estimation towards a global minimum. The terms $\{g_x, g_y, g_z\}$ are numeric derivatives of the unknown surface patch $g(x, y, z)$. Its calculation depends on the analytical representation of the surface elements. As a first method, let us represent the search surface elements as plane surface patches, which are constituted by fitting a plane to 3 neighboring knot points, in the implicit form

$$g^0(x, y, z) = Ax + By + Cz + D = 0 \quad (19)$$

where $A, B, C,$ and D are parameters of the plane. Using the mathematical definition of the derivation, the numeric 1st derivation according to the x -axis is

$$g_x = \frac{\partial g^0(x, y, z)}{\partial x} = \lim_{\Delta x \rightarrow 0} \frac{g^0(x + \Delta x, y, z) - g^0(x, y, z)}{\Delta x} \quad (20)$$

where the numerator term of the equation is simply the distance between the plane and the off-plane point $(x + \Delta x, y, z)$. Then using the point-to-plane distance formula,

$$g_x = \frac{A(x + \Delta x) + By + Cz + D}{\Delta x \sqrt{A^2 + B^2 + C^2}} = \frac{A}{\sqrt{A^2 + B^2 + C^2}} \quad (21)$$

is obtained. Similarly g_y and g_z are calculated numerically.

$$g_y = \frac{B}{\sqrt{A^2 + B^2 + C^2}}, \quad g_z = \frac{C}{\sqrt{A^2 + B^2 + C^2}} \quad (22)$$

Actually these numeric derivative values $\{g_x, g_y, g_z\}$ are x-y-z components of the local surface normal vector at that point.

$$\vec{n} = \frac{\vec{\nabla}g^0}{\|\vec{\nabla}g^0\|} = \frac{\begin{bmatrix} \frac{\partial g^0}{\partial x} & \frac{\partial g^0}{\partial y} & \frac{\partial g^0}{\partial z} \end{bmatrix}^T}{\|\vec{\nabla}g^0\|} = \frac{[A \ B \ C]^T}{\sqrt{A^2 + B^2 + C^2}} \quad (23)$$

In the case of representation of search surface elements as parametric bi-linear surface patches, which are constituted by fitting the bi-linear surface to 4 neighboring knot points $P_{i,j}$:

$$\vec{G}(u, w) = \vec{P}_{0,0}(1-u)(1-w) + \vec{P}_{0,1}(1-u)w + \vec{P}_{1,0}u(1-w) + \vec{P}_{1,1}uw \quad (24)$$

where $u, w \in [0, 1]^2$ and $G, P_{i,j} \in \mathcal{R}^3$. Again the numeric derivative terms $\{g_x, g_y, g_z\}$ are calculated from components of the local surface normal vector on the parametric bi-linear surface patch:

$$\vec{n} = [g_x \ g_y \ g_z]^T = \frac{\vec{\nabla}G}{\|\vec{\nabla}G\|} = \frac{\frac{\partial \vec{G}(u, w)}{\partial u} \times \frac{\partial \vec{G}(u, w)}{\partial w}}{\|\vec{\nabla}G\|} \quad (25)$$

With this approach a better a posteriori sigma value could be obtained due to a smoothing effect. In the case of insufficient initial approximations, the numeric derivatives $\{g_x, g_y, g_z\}$ can be calculated on the template surface patch $f(x, y, z)$ instead of on the search surface $g(x, y, z)$ in order to speed-up the convergence.

2.2 Precision and Reliability Issues

The standard deviations of the estimated transformation parameters and the correlations between themselves may give useful information concerning the stability of the system and quality of the data content (Gruen, 1985a).

$$\hat{\sigma}_p = \hat{\sigma}_0 \sqrt{q_{pp}} \quad , \quad p_i \in \{t_x, t_y, t_z, m, \omega, \varphi, \kappa\} \quad , \quad q_{pp} \in \mathbf{Q}_{xx} \quad (26)$$

As pointed out in (Maas, 2000), the estimated standard deviations of the translation parameters are too optimistic due to stochastic properties of the search surface.

Because of the high level redundancy of a typical data arrangement, a certain amount of occlusions and/or outliers do not have significant effect on the estimated parameters. Baarda's data-snooping method can be favourably used to localize the occluded or gross erroneous measurements.

2.3 Computational Aspects

The computational complexity is of order $O(N^2)$, where N is the number of employed points in the matching process. The actual problem is to search the correspondent element of the template surface on the search surface patch, whereas the adjustment part is a small system, and can quickly be solved using back-substitution followed by Cholesky decomposition. Searching the correspondence is an algorithmic problem, and needs professional software optimization techniques and programming skills, which are not within the scope of this paper.

Since the method needs initial approximations of the unknowns due to the non-linear functional model, one of the methods for pre-alignment in the literature (Habib and Schenk, 1999, Murino et al., 2001, Lucchese et al., 2002, Vanden Wyngaerd and Van Gool, 2002) should be utilized.

Two 1st degree C^0 continuous surface representations are implemented, and explained in detail. In the case of multi-resolution data sets, in which point densities are significantly different on the template and search surface patches, higher degree C^1 continuous composite surface representations, e.g. bi-cubic Hermit surface (Peters, 1974), should give better results, of course increasing the computational expenses.

2.4 Convergence of Solution Vector

In a standard LS adjustment calculus in geodesy and photogrammetry, the function of the unknowns is unique, exactly known, and analytically continuous everywhere, e.g. the collinearity equations in the bundle adjustment. Here the function $g(x, y, z)$ is discretized by using a definite sampling rate, which leads to slow convergence, oscillations, even divergence in some cases with respect to the standard adjustment. The convergence behaviour of the proposed method basically depends on the quality of the initial approximations and quality of the data content, and it usually achieves the solution after 4th or 5th iterations (Figure 1), as typically in LSM.

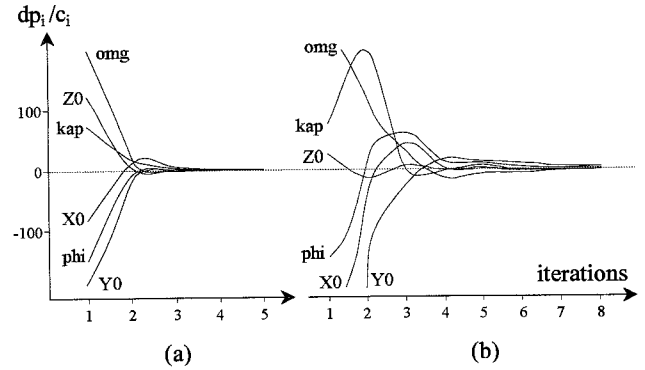


Figure 1: Typical examples for fast convergence (a) and slow convergence (b). Note that scale factor is fixed to unity.

3. THE EXPERIMENTAL RESULTS

Two practical examples are given to show the capabilities of the method. All experiments were carried out using own self-developed C/C++ software that runs on *Microsoft Windows*® OS. Processing times given in Table 1 were counted on such a PC, whose configuration is *Intel*® P4 2.53 GHz CPU, 1 GB RAM. The first example is the registration of three surface patches, which were photogrammetrically measured 3D point clouds of a human face from multi-images (Figure 2). For the mathematical and implementation details of this surface measurement method the author refers to (D'Apuzzo, 2002).

Left and right search surface patches (Figure 2-a and 2-c) were matched to the centre template surface patch (Figure 2-b) by use of LS3D. Since the data set already came in a common coordinate system, the rotation angles $(\omega, \varphi, \kappa)$ of the search surfaces were deteriorated by $\sim 10^5$ in the first iteration. Numerical results of the matching of the left surface and the right surface patches are given at parts I-L and I-R of Table 1 respectively. Relatively high standard deviations for the estimated t_x and φ (note that high physical correlation between x and φ due to a conventional axes configuration) exhibit the narrow overlapping area along the x -axis, nevertheless the matching result is successful. The estimated σ_0 values prove the accuracy potential of the surface measurement method, given as 0.2 mm by D'Apuzzo (2002).

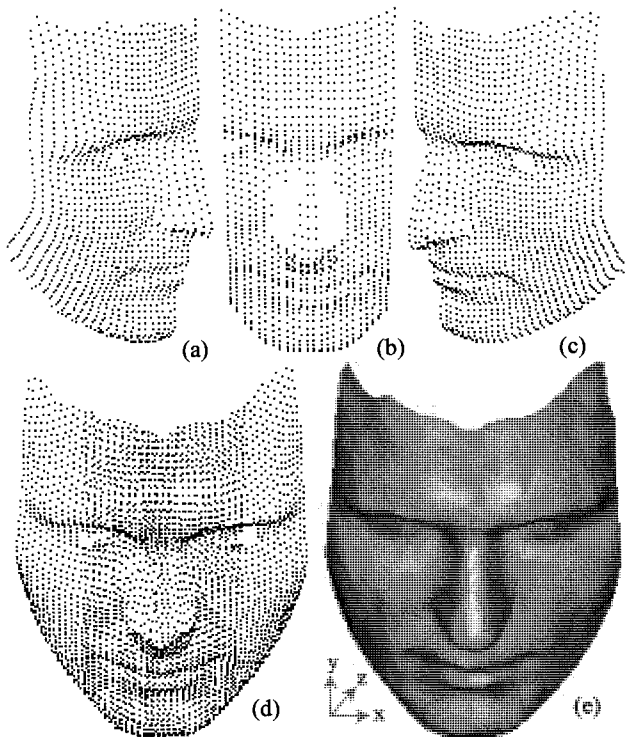


Figure 2: (a) left-search surface, (b) centre-template surface, (c) right-search surface, (d) obtained 3D point cloud after LS3D surface matching, (e) shaded view of the final composite surface.

The second experiment refers to the matching of two overlapping 3D point clouds (Figure 3), which are a part of a chapel in Wangen, Germany, and were scanned using IMAGER 5003 terrestrial laser scanner (Zoller+Fröhlich). Initial approximations of the unknowns were provided by interactively selecting 3 common points on the both surfaces before the matching. Obtained results are given at part II of Table 1. The estimated σ_0 gives valuable information about the sensor noise level and the accuracy limit of the scanner as >1.7 mm.

Table 1: Experimental results

	s	n	i	t	d	$\hat{\sigma}_0$	$\hat{\sigma}_{tx}/\hat{\sigma}_{ty}/\hat{\sigma}_{tz}$	$\hat{\sigma}_w/\hat{\sigma}_\varphi/\hat{\sigma}_\kappa$
					sec	mm	mm	mm
I-L	P	2497	7	0.6	1.5	0.19	0.15/0.07/0.05	0.96/2.44/1.90
	B		7	1.3		0.19	0.15/0.07/0.05	0.96/2.42/1.91
I-R	P	3285	6	0.5	1.5	0.21	0.13/0.03/0.05	0.68/2.25/1.73
	B		6	1.4		0.21	0.13/0.03/0.05	0.69/2.26/1.75
II	P	13461	5	3.8	10	1.74	0.23/0.62/0.01	0.69/0.17/0.46
	B		4	5.6		1.72	0.22/0.61/0.01	0.69/0.17/0.46

I-L: left face surface, I-R: right face surface, II: laser scanner data
s: surface representation, P: plane, B: bi-linear surface, n: number of employed points, i: iterations, t: process time, d: ~ point spacing

The parametric bi-linear surface representation gives a slightly better convergence rate and a better a posteriori sigma value than the triangle plane representation, while increasing the computational expenses. The standard deviation of the z-component of the translation vector shows the excellent data content in the depth direction, but the relative precision is highly optimistic, which is $\sim 1/1000$ of the point spacing.

Since LS3D reveals the sensor noise level and accuracy potential of any kind of surface measurement method or device, it should be used for comparison and validation studies.

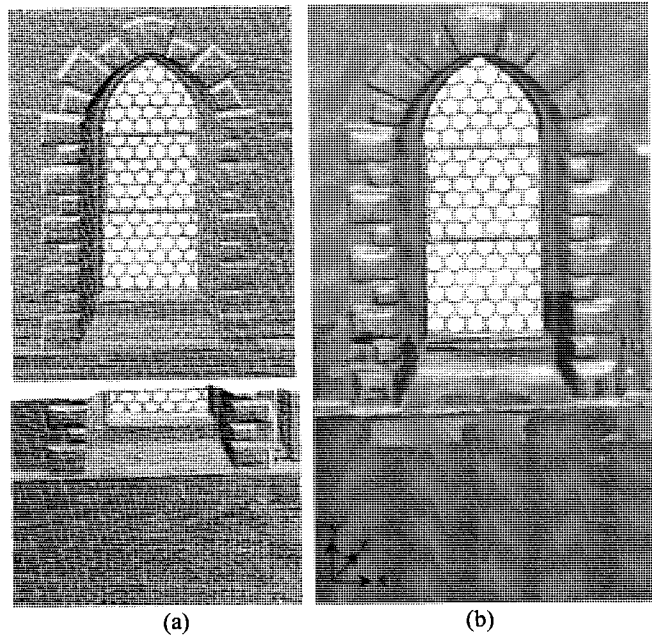


Figure 3: (a) top - template surface patch, (a) bottom - search surface patch, (b) overlay of the shaded surfaces.

4. CONCLUSIONS

LSM is a fundamental measurement algorithm, and has been applied to a great variety of data matching problems due to its strong mathematical model. Two well-known ones are LS image matching in 2D pixel space, and LS multiple cuboid matching in 3D voxel space. The LS3D is bridging the conceptual gap between the LS image matching and the LS cuboid matching.

This new 3D surface matching technique is a generalization of the least squares 2D image matching concept and offers high flexibility for any kind of 3D surface correspondence problem, as well as monitoring capabilities for the analysis of the quality of the final results by means of precision and reliability criteria. Another powerful aspect of this proposed method is its ability to handle multi-resolution, multi-temporal, multi-scale, and multi-sensor data sets. The technique can be applied to a great variety of data co-registration problems. In addition time dependent (temporal) variations of the object surface can be inspected, tracked, and localized using the statistical analysis tools of the method.

ACKNOWLEDGEMENT

The author would like to thank Dr. Nicola D'Apuzzo for providing the face surface data sets, which were measured by use of his own software *Viewtriplet GTK v0.9*©. The laser scanner data set is courtesy of Zoller+Fröhlich GmbH Elektrotechnik (Wangen, Germany).

REFERENCES

- Bergevin, R., Soucy, M., Gagnon, H., Laurendeau, D., 1996. Towards a general multi-view registration technique. *IEEE Pattern Analysis and Machine Intelligence*, 18(5), pp. 540-547.
- Besl, P.J., and McKay, N.D., 1992. A method for registration of 3D shapes. *IEEE Pattern Analysis and Machine Intelligence*, 14(2), pp. 239-256.
- Campbell, R.J., and Flynn, P.J., 2001. A survey of free form object representation and recognition techniques. *Computer Vision and Image Understanding*, 81(2), pp. 166-210.

- Chen, Y., and Medioni, G., 1992. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3), pp. 145-155.
- D'Apuzzo, N., 2002. Measurement and modeling of human faces from multi images. *IAPRS*, 34 (5), pp. 241-246.
- Dorai, C., Weng, J., Jain, A.K., 1997. Optimal registration of object views using range data. *IEEE Pattern Analysis and Machine Intelligence*, 19(10), pp. 1131-1138.
- Ebner, H., and Mueller, F., 1986. Processing of Digital Three Line Imagery using a generalized model for combined point determination. *IAPRS*, 26 (3/1), pp. 212-222.
- Ebner, H., and Strunz, G., 1988. Combined point determination using DTMs as control information. *IAPRS*, 27 (B11/3), pp. 578-587.
- Fusiello, A., Castellani, U., Ronchetti, L., Murino, V., 2002. Model acquisition by registration of multiple acoustic range views. *Computer Vision—ECCV 2002*, Springer, Berlin, pp. 805-819.
- Gruen, A., 1984. Adaptive least squares correlation – concept and first results. *Intermediate Research Project Report to Heleva Associates, Inc.*, Ohio State University, Columbus, Ohio, March.
- Gruen, A., 1985a. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3), pp. 175-187.
- Gruen, A., 1985b. Adaptive kleinste Quadrate Korrelation und geometrische Zusatzinformationen. *Vermessung, Photogrammetrie, Kulturtechnik*, (9), pp. 309-312.
- Gruen, A., 1986. Photogrammetrische Punktbestimmung mit der Buendelmethode. *IGP ETH-Zürich*, Mitt. Nr.40, pp. 1-87.
- Gruen, A., Baltsavias, E.P., 1988. Geometrically Constrained Multiphoto Matching. *PE&RS*, 54(5), pp. 633-641.
- Gruen, A., and Stallmann, D., 1991. High accuracy edge matching with an extension of the MPGC matching algorithm. *Int. Conf. Industrial Vision Metrology, SPIE vol. 1526*, Winnipeg, July 11-12, pp. 42-55.
- Gruen, A., and Agouris, P., 1994. Linear feature extraction by least squares template matching constrained by internal shape forces. *IAPRS*, 30 (3/1), pp. 316-323.
- Gruen, A., and Li, H., 1996. Linear feature extraction with LSB-Snakes from multiple images. *IAPRS*, 31 (3B), pp. 266-272.
- Gruen, A., 1996. Least squares matching: a fundamental measurement algorithm. In: K. Atkinson (ed.), *Close Range Photogrammetry & Machine Vision*, Whittles, pp. 217-255.
- Habib, A., and Schenk, T., 1999. A new approach for matching surfaces from laser scanners and optical scanners. *IAPRS*, 32 (3/W14), pp. 55-61.
- Jokinen, O., and Haggren, H., 1998. Statistical analysis of two 3-D registration and modeling strategies. *ISPRS J.Photog.&RS*, 53(6), pp. 320-341.
- Karras, G.E., and Petsa, E., 1993. DEM matching and detection of deformation in close-range Photogrammetry without control. *PE&RS*, 59(9), pp. 1419-1424.
- Lucchese, L., Doretto, G., Cortelazzo, G.M., 2002. A frequency domain technique for range data registration. *IEEE Pattern Analysis and Machine Intelligence*, 24(11), pp. 1468-1484.
- Maas, H.G., 1994. A high-speed solid state camera system for the acquisition of flow tomography sequences for 3D least squares matching. *IAPRS*, 30 (5), pp. 241-249.
- Maas, H.G., and Gruen, A., 1995. Digital photogrammetric techniques for high resolution three dimensional flow velocity measurements. *Optical Engineering*, 34(7), pp. 1970-1976.
- Maas, H.G., 2000. Least-Squares Matching with airborne laser scanning data in a TIN structure. *IAPRS*, 33 (3A), pp. 548-555.
- Masuda, T., Yokoya, N., 1995. A robust method for registration and segmentation of multiple range images. *Computer Vision and Image Understanding*, 61(3), pp. 295-307.
- Mitchell, H.L., and Chadwick, R.G., 1999. Digital Photogrammetric concepts applied to surface deformation studies. *Geomatica*, 53(4), pp. 405-414.
- Murino, V., Ronchetti, L., Castellani, U., Fusiello, A., 2001. Reconstruction of complex environments by robust pre-aligned ICP. *3DIM 2001*, Quebec, May 28–June 1, pp. 187-194.
- Okatani, I.S., and Deguchi, K., 2002. A method for fine registration of multiple view range images considering the measurement error properties. *Computer Vision and Image Understanding*, 87(1-3), pp. 66-77.
- Peters, G.J., 1974. Interactive computer graphics application of the parametric bi-cubic surface to engineering design problems. In: R. Barnhill and R. Riesenfeld (Eds.), *Computer Aided Geometric Design*, Academic Press, pp. 259-302.
- Pilgrim, L., 1996. Robust estimation applied to surface matching. *ISPRS J.P&RS*, 51(5), pp. 243-257.
- Postolov, Y., Krupnik, A., and McIntosh, K., 1999. Registration of airborne laser data to surfaces generated by Photogrammetric means. *IAPRS*, 32 (3/W14), pp. 95-99.
- Potmesil, M., 1983. Generating models of solid objects by matching 3D surface segments. *Int. Joint Conference on Artificial Intelligence*, Karlsruhe, August 8-12, pp. 1089-1093.
- Rosenholm, D., and Torlegard, K., 1988. Three-dimensional absolute orientation of stereo models using DEMs. *PE&RS*, 54(10), pp. 1385-1389.
- Sequeira, V., Ng, K., Wolfart, E., Goncalves, J.G.M., Hogg, D., 1999. Automated reconstruction of 3D models from real environments. *ISPRS J.Photog.&RS*, 54(1), pp. 1-22.
- Vanden Wyngaerd, J., and Van Gool, L., 2002. Automatic crude patch registration: towards automatic 3D model building. *Computer Vision and Image Understanding*, 87(1-3), pp. 8-26.
- Williams, J.A., Bennamoun, M., Latham, S., 1999. Multiple view 3D registration: A review and a new technique. In *Proc. of IEEE Int. Conf. Sys., Man & Cybernetics*, Tokyo, pp. 497-502.
- Xu, Z., and Li, Z., 2000. Least median of squares matching for automated detection of surface deformations. *IAPRS*, 33 (B3), pp. 1000-1007.
- Zhang, Z., 1994. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2), pp. 119-152.

AN ADVANCED SENSOR MODEL FOR PANORAMIC CAMERAS

Jafar Amiri Parian, Armin Gruen

Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology (ETH) Zurich
(parian, agruen)@geod.baug.ethz.ch

Commission V, WG V/1

KEY WORDS: Close Range, Photogrammetry, Panoramic Camera, Calibration, Sensor, Modeling, Accuracy, Test

ABSTRACT:

Digital terrestrial panoramic cameras constitute an interesting new development, which is currently primarily used for purely imaging purposes such as indoor imaging, landscape recording, tourism advertising and Internet representations. However, the capability of taking high-resolution images continuously over the full horizon generates an efficient means for 3D object reconstruction as well. For that the particular sensor model has to be established and the inherent accuracy potential has to be investigated. We designed a sensor model, which models substantial deviations from the pinhole model using additional parameters. The sensor model maps the object space into the image space. The mapping function is the pinhole model-based perspective transformation in the form of bundle equations. In practice, there are many systematic errors disturbing the ideal model, which can be modeled as additional parameters. Additional parameters relate to the camera itself, the configuration of camera and turntable, and mechanical errors of the camera system during rotation (i.e. tumbling). In this paper we will present the results of calibration with additional parameters for two panoramic cameras, which indicate a subpixel accuracy level for such highly dynamic systems. We also investigate into the problem of temporal stability of the systematic errors. Finally we will demonstrate the systems' accuracy in 3D point positioning, including minimal number of control points adjustment. With these new panoramic imaging devices we do have additional powerful sensors for image recording and efficient 3D object modeling.

1. INTRODUCTION

The first panoramic cameras used in Photogrammetry were film-based aerial cameras. The Manual of Photogrammetry, 1980 lists a number of types, which differ mechanically and optically from each other. A prototype of an aerial panoramic camera can be modeled as a camera with a cylindrical focal surface, in which the image is acquired by sweeping a slit across this surface (Hartley, 1993). Through the integration of CCD technology, new types of airborne and terrestrial digital panoramic cameras were generated, using Linear Array CCDs as imaging devices. The EYESCAN, jointly developed by German Aerospace Center (DLR) and KST Dresden GmbH and the SpheroCam, SpheronVR AG are two different types of line-based panoramic cameras. The EYESCAN camera as used in terrestrial photogrammetric applications was addressed in Scheibe et al., 2001. Schneider and Maas, 2003 and Amiri Parian and Gruen, 2003 have worked on the mathematical modeling of line-based panoramic cameras. Schneider and Maas investigated a geometrical model for a prototype of the EYESCAN panoramic camera and they performed calibration by using a 3D testfield (Schneider and Maas, 2003). They also performed 3D positioning using bundle block adjustment (Schneider and Maas, 2004). We have worked on the mathematical model of general line-based panoramic cameras. We performed calibration and accuracy test using a 3D testfield for EYESCAN and SpheroCam (Amiri Parian and Gruen, 2003). We improved mathematical model by modeling the mechanical error of the rotating turntable, tumbling, and we reported the improvement of the accuracy by a factor of two in the case of using tumbling parameters in the bundle adjustment process (Amiri Parian and Gruen, 2004).

In this paper, by defining image- and block-invariant parameters we put emphasis on 3D positioning using a minimal number of control points. The paper will be organized as follows. Chapter 1 gives a short review of the panoramic cameras SpheroCam and EYESCAN. Chapter 2 addresses our mathematical sensor model. Chapter 4 covers the result of adjustment, included the results of the physical measurement of the tumbling for the SpheroCam, and the calibration results of EYESCAN with/without tumbling parameters. In this chapter we demonstrate the system accuracy for EYESCAN using a testfield with as few as possible control points.

2. PANORAMA TECHNIQUES

Several techniques have been used for panoramic imaging. Mosaicing/stitching of a rotated frame-CCD camera, mirror technology including single mirror and multi mirrors, near 180 degrees with large frame cameras or one shot with fish-eye lens and recently a new technology of creating high resolution panoramic image by rotating a line-CCD camera are some known methods for panoramic imaging. Up to now, these techniques have mainly been used for pure imaging purposes, such as indoor imaging, landscape and cultural heritage recording, tourism advertising and image-based rendering, and recently for efficient Internet representations. Among the mentioned techniques for panoramic imaging, the last one has a possibility to produce a high-resolution panoramic image (more than 300 Mpixels) in one shot. The camera principle consists of a linear array, which is mounted on a high precision turntable parallel to the rotation axis. By rotation of the turntable, the linear array sensor captures the scenery as a continuous set of vertical scan lines.

Table 1. Parameters of EYESCAN and SpheroCam panoramic cameras

Parameters	EYESCAN	SpheroCam
Number of pixel in linear array (vertical format)	3600 or 10200 pixels per line	5300 pixels per line
Horizontal format (depends on the focal lens)	27489 pixels (35 mm lens)	39267 pixels (50 mm lens)
Pixel size	7 or 8 microns	8 microns

In our tests we used two line-based rotating panoramic cameras, a prototype of EYESCAN M3, a joint development between German Aerospace Center (DLR) and KST Dresden GmbH*. The camera is engineered for rugged everyday field use as well as for the measurement laboratory. The other panoramic camera used here is the SpheroCam from the SpheronVR AG** which operates similar to EYESCAN.

2.1. EYESCAN M3

Figure 1 shows the sensor system and Table 1 shows format parameters of the camera. The camera system contains three parts: camera head, optical part, and high precision turntable with a DC-gearsystem motor. The camera head is connected to the PC with a bi-directional fiber link for data transmission and camera control. The optical part of the system uses high performance Rhodenstock lenses. With adjustment rings one can use other lenses. The camera head is mounted on a high precision turntable with a sinus-commutated DC-gearsystem motor (Scheibe et al., 2001), internal motion control and direct controlling by the PC. Rotation speed and scan angle are pre-selectable and correspond to the shutter speed, image size and focal length of the lens. For a more detailed description see Schneider and Maas (2003).

2.2. SpheroCam

The structure of the SpheroCam (Figure 1) includes three parts, the camera head, the optical part which is compatible with NIKON-lenses, and a DC motor to rotate the Linear Array. The SpheroCam is specially designed for use with a fish-eye lens, with a near 180° vertical field of view. As it rotates about its vertical axis, the SpheroCam then captures a complete spherical image. It is designed to capture high quality images. Table 1 contains the format parameters of SpheroCam. For more detail on specifications of the camera see Amiri Parian and Gruen (2003).

3. SENSOR MODEL

The sensor model as a mapping function is based on a projective transformation in the form of bundle equations, which maps the 3D object space information into the 2D image space. The sensor model uses the following coordinate systems:

- Pixel coordinate system
- Linear Array coordinate system
- 3D auxiliary coordinate system
- 3D object coordinate system

Figure 2 shows the pixel coordinate (i, j) system. The original image observations are saved in this system. Figure 3 shows the other coordinate systems: Linear Array (0, y, z), auxiliary (X', Y', Z') and object space (X, Y, Z) coordinate systems. The

* <http://www.kst-dresden.de/>

** <http://www.spheron.com/>

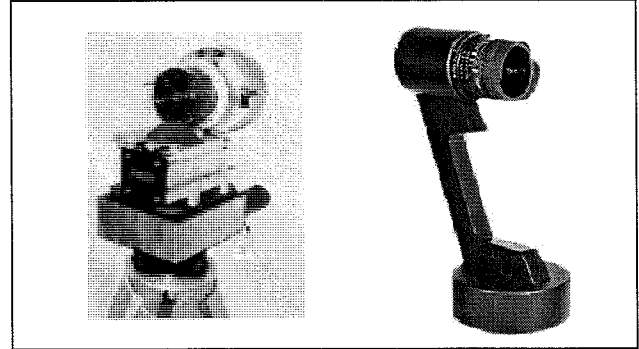


Figure 1. Digital terrestrial panoramic cameras. EYESCAN (left) and SpheroCam (right).

effects of lens distortion and the shift of the principal point are modeled in the Linear Array coordinate system. The rotation of the Linear Array and mechanical errors of the rotating turntable are modeled in the auxiliary coordinate system. The object space coordinate system is used as a reference for determining the exterior orientation parameters of the sensor.

To define the auxiliary coordinate system, an ideal panoramic camera is considered. Here the origin of the auxiliary coordinate system coincides with the projection center O. The rotation axis passes through the projection center and coincides with Z'. X' passes through the start position of the Linear Array before rotation and Y' is defined to get a right-handed coordinate system.

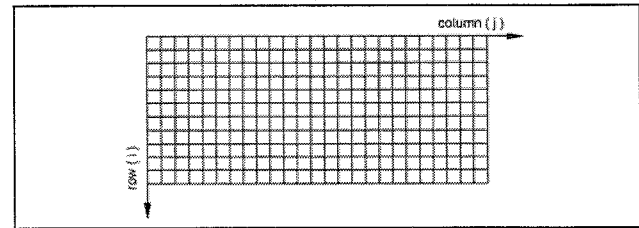


Figure 2. Pixel coordinate system (i, j).

The model, which directly relates pixel observations (i, j) to the object points (X, Y, Z), for an ideal sensor becomes (Amiri Parian and Gruen, 2003):

$$\begin{pmatrix} 0 \\ y \\ -c \end{pmatrix} = \lambda P^{-1} R_z^t(j A_h) M_{w,p,k} \begin{pmatrix} X - X_o \\ Y - Y_o \\ Z - Z_o \end{pmatrix} \quad (1)$$

With

$$P = \begin{pmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad y = (i - \frac{N}{2}) A_v$$

Where, A_h is horizontal pixel size and A_v is vertical pixel size. N is the number of pixel in linear array.

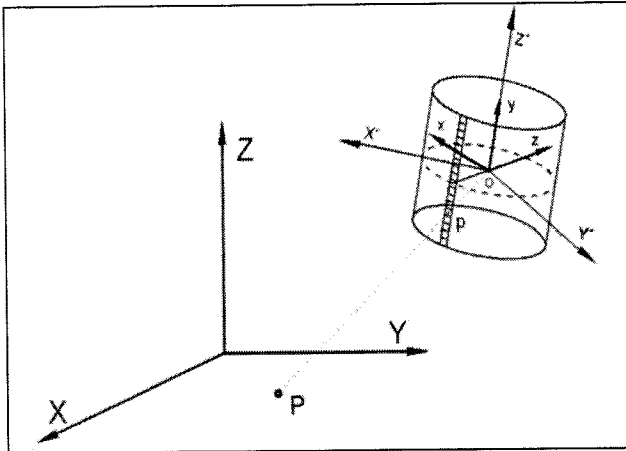


Figure 3. Object coordinate (X, Y, Z), auxiliary coordinate (X', Y', Z') and Linear Array ($0, y, z$) coordinate systems.

There are many systematic errors disturbing the ideal sensor model. The most important ones, with a distinct physical meaning, are:

- 1) Radial lens distortion (2 parameters)
- 2) Shift of principal point (1 parameter)
- 3) Camera constant (1 parameter)
- 4) Tilt and inclination of the Linear Array with respect to the rotation axis (2 parameters)
- 5) Eccentricity of the projection center from the origin of the auxiliary coordinate system (2 parameters)
- 6) Resolution of rotation (1 parameter)
- 7) Tumbling (3 or 6 parameters)

The above errors are modeled as additional parameters for a prototype of a panoramic camera. The results of the modeling for two different cameras were reported in Amiri Parian and Gruen (2003, 2004).

The additional parameters can be divided in four different groups. The first is related to the camera head and optics (parameters of classes 1, 2 and 3). The second group of parameters (Figure 4) is related to the configuration of the camera head and the plane of the turntable (parameters of classes 4 and 5). The third group is related to the turntable itself (parameter of class 6). And finally the fourth group is related to the mechanical errors of the turntable, tumbling, while the camera rotates (parameters of class 7). Tumbling is mainly caused by an incomplete shape of ball bearings and the

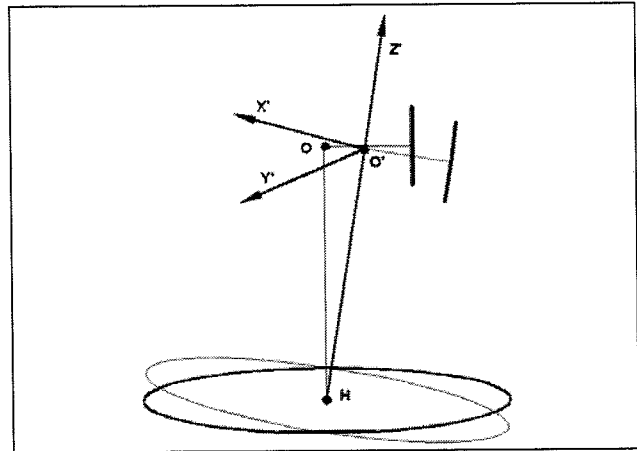


Figure 5. Effect of tumbling: Oscillation of the origin of the auxiliary coordinate system.

contacting surfaces (Matthias, 1961). Tumbling results from the mechanical properties of the instrument. Especially, it is affected by the rotation around the vertical axis and shows its effect as a change of the exterior orientation of the camera head during rotation. From that, one of the main effects of the tumbling is the moving of the origin of the auxiliary coordinate system during rotation (Figure 5). For more detailed information on the mathematical modeling of the tumbling see Amiri Parian and Gruen (2004).

In the next chapter we will report the physical measurement of the tumbling and the result of calibration and accuracy testing with/without the tumbling parameters. We show the results of accuracy tests with minimal number of control points.

4. RESULTS

4.1. Physical Measurement of the Tumbling

The examination of the tumbling error of the SpheroCam was carried out by an inclinometer. In the present case the Zerotron from Wyler Switzerland is used, which provides the inclination in a specific direction. The inclinometer was placed firmly on the top of the turntable near the gravity center of the camera system. Then using the operating software of the camera, the inclinations of at least 3 continuous rotations (1080°) of the turntable at every 15° were recorded. To see

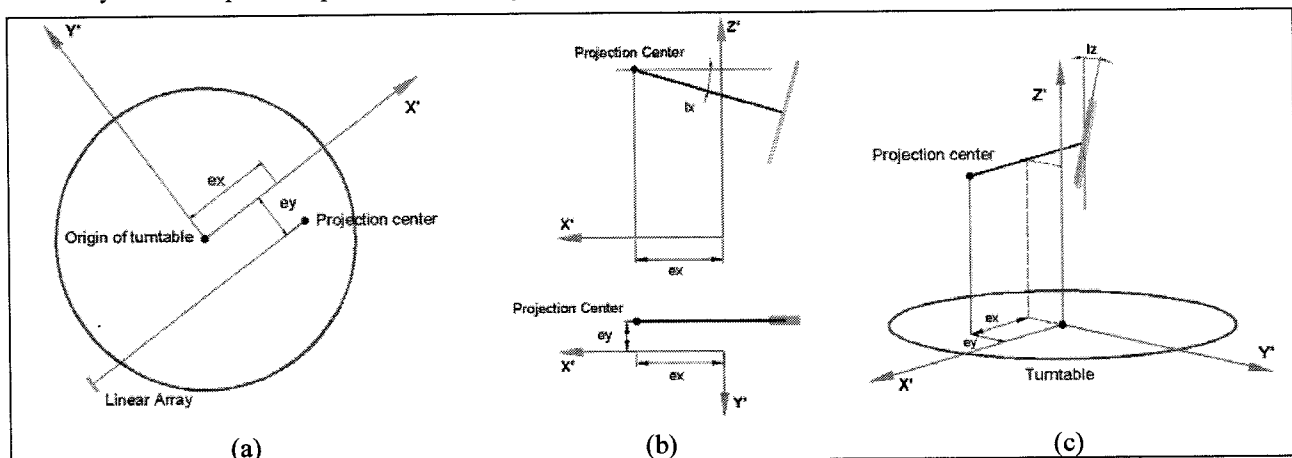


Figure 4. Additional parameters of the configuration of the camera on the turntable. (a) Eccentricity (e_x, e_y), (b) tilt of the linear array (l_x), (c) inclination of the linear array with respect to the rotation axis (l_z).

whether the effect is stationary with respect to time, the measurements were carried out at 4 different epochs. Figure 6 shows the observations for one epoch. A Fourier analysis of the signal was carried out, which shows a high peak at the period near π (Figure 6). The analysis of the other epochs shows that the camera is not stable over time. The instability of the camera causes different amplitudes and periods of the observations. Figure 7 shows the observations and the power spectrum of another epoch. These experiences indicate that the camera has a periodic oscillation.

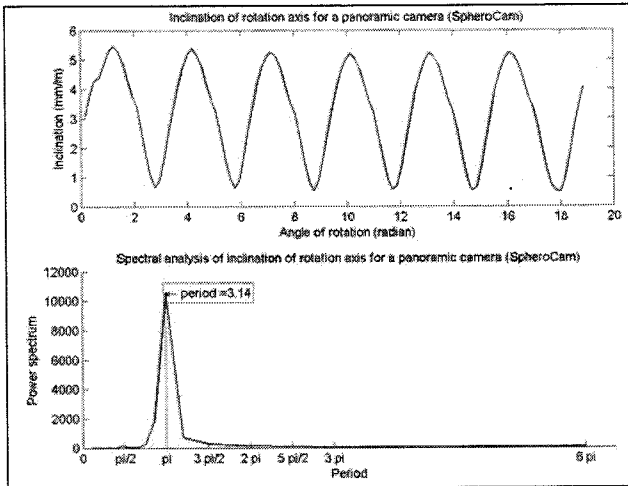


Figure 6. Observations for the inclination of the turntable (top) and the corresponding power spectrum (bottom) for epoch 1.

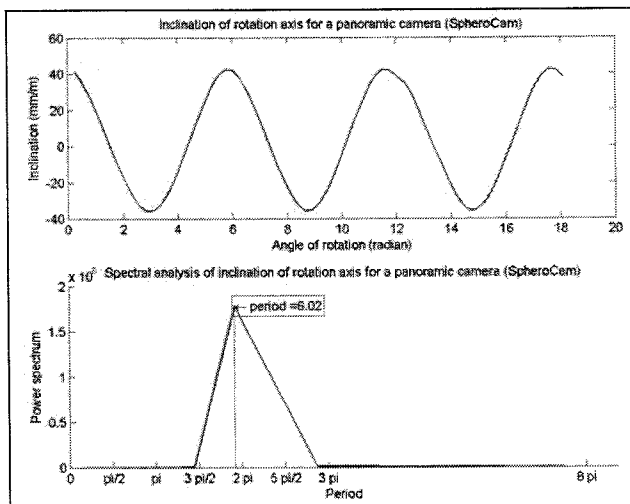


Figure 7. Observations for the inclination of the turntable (top) and the corresponding power spectrum (bottom) for epoch 2.

4.2. Camera Calibration

Camera calibration was performed by the mentioned sensor model using additional parameters. For the analysis of the additional parameters (to find the most influential parameters and those which are stable under the given network condition) we added step by step each parameter to the previous stage of the model and used the correlations for stability checking. Comparing the additional parameters of different images, we found the image and block-invariant parameters. We will report the result of calibration in the following sections for the last stage.

4.2.1. SpheroCam

The camera calibration was performed using a testfield. We established a testfield with 96 circular targets at our institute and used it for the calibration of the SpheroCam. The testfield was measured with a Theodolite with mean precision of 0.3, 0.3, 0.1 mm for the three coordinate axes (X, Y, Z). The camera calibration was performed by the additional parameters mentioned in chapter 3. To model the tumbling error 6 parameters were used. The a posteriori variance of unit weight is 0.59 pixel (4.7 microns) after self-calibration. Figure 8 shows the residuals of the image point observations in the image space for this case. A comparison of the computed tumbling parameters of different images shows that none of the tumbling parameters is block-invariant. To see the effect of tumbling parameters, a camera calibration was performed with the same condition but without tumbling parameters. In this case the a posteriori variance of unit weight is 1.37 pixels (10.9 microns).

4.2.2. EYESCAN

For EYESCAN, we got the image and field observations from Mr. Schneider, TU Dresden. TU's testfield consists of more than 200 control points and the mean precision of control points is 0.2, 0.3, 0.1 mm for the three coordinate axes (X, Y, Z). The camera calibration was performed with the same model and the additional parameters as mentioned in chapter 3. To model the tumbling error of this camera 3 parameters were used. The a posteriori computed variance of unit weight is 0.33 pixel (2.6 microns). Figure 9 shows the residuals of the image point observations in the image space. A comparison of the computed tumbling parameters of different images shows that 2 of 3 tumbling parameters are block-invariant. In the case that tumbling parameters were not used the a posteriori computed variance of unit weight is 1.30 pixels (10.4 microns).

4.3. Block Adjustment with Accuracy Test

An accuracy test was performed for EYESCAN by block triangulation using 5 camera stations and by defining 151 check and 3 control points. Considering the result of camera calibration for different images, totally 8 parameters were used as unknown block-invariant, 1 parameter as priori known parameter (camera constant), and 6 parameters as image-invariant parameters. Table 2 shows the summary of the results of adjustment without the modeling of the tumbling. The RMS errors of check points compared with the standard deviations are too large. The reason is that the mathematical model is not complete and cannot interpret the physical behavior of the dynamic camera system. To complete the mathematical model, tumbling parameters were added and the accuracy test was performed. In this case, 8 parameters were used as unknown block-invariant, 4 parameters as a priori known parameters (camera constant and 3 tumbling parameters), and 6 parameters as image-invariant parameters. The summary of the results of adjustment is in the Table 3. Figures 10 and 11 show the object space residuals for checkpoints in depth axes (X and Y) and lateral axis (Z). The RMS errors of check points, compared with standard deviations, are reasonable and shows the effect the tumbling parameters in the modeling. However, the systematic patterns of the residuals have not been completely removed, but the size of the errors is significantly reduced. The remained systematic errors may come from non-modeled mechanical errors of the camera.

For the accuracy test, similar to the conventional close range photogrammetry (frame CDD cameras), 3 control points were

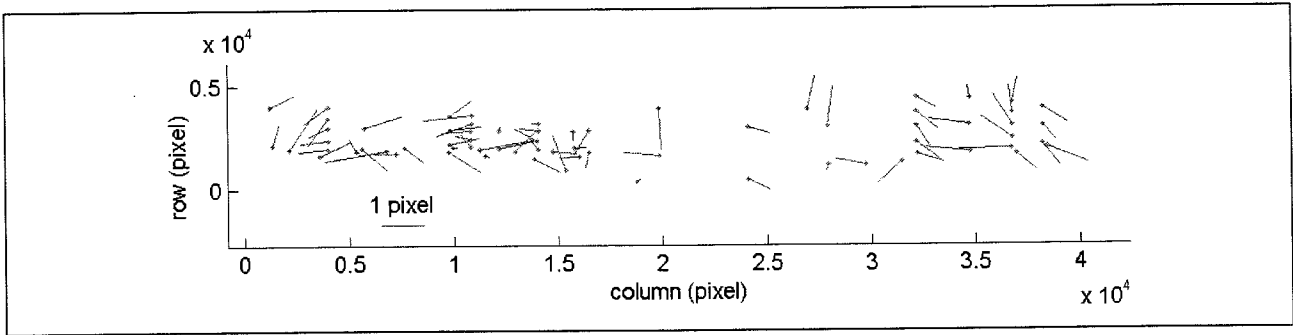


Figure 8. Image space residuals of the observations for the SpheroCam.

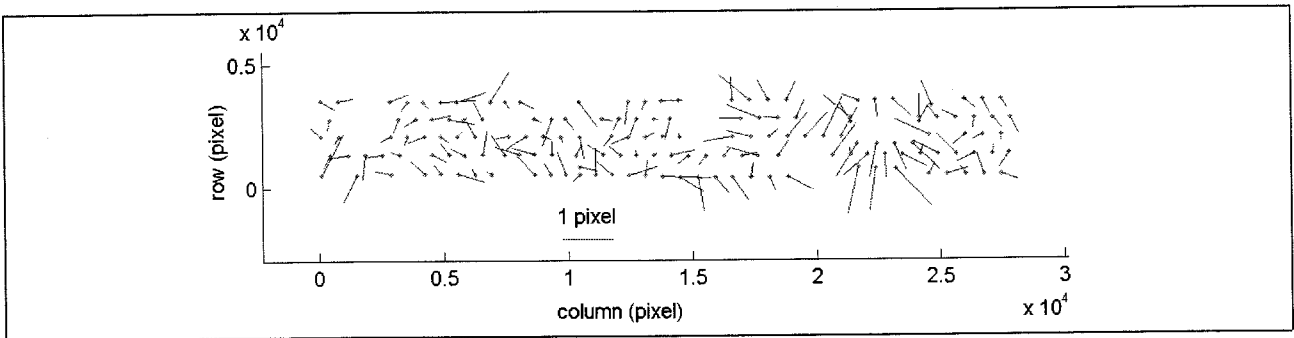


Figure 9. Image space residuals of the observations for the EYESCAN.

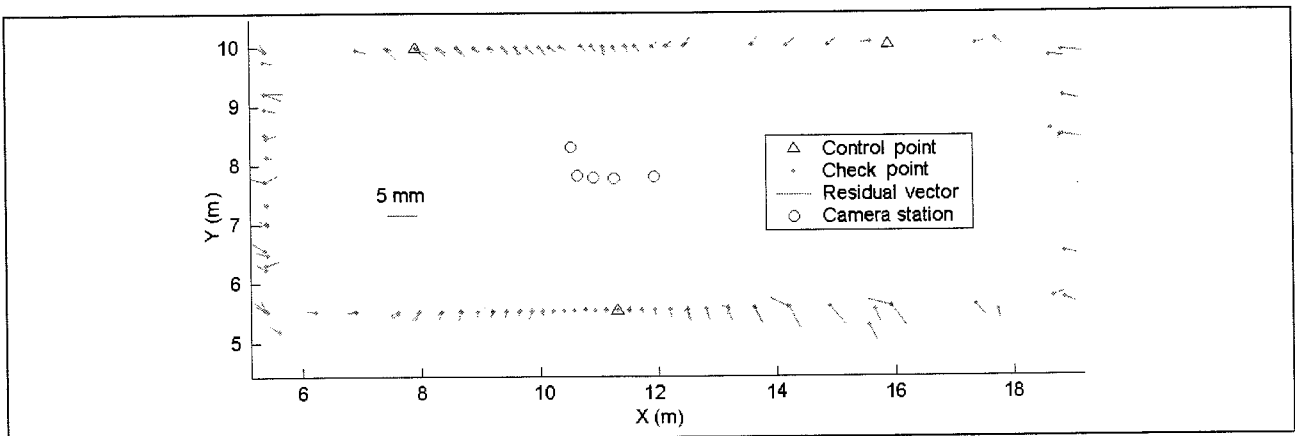


Figure 10. Object space residuals of check points for X and Y axes in XY-plane.

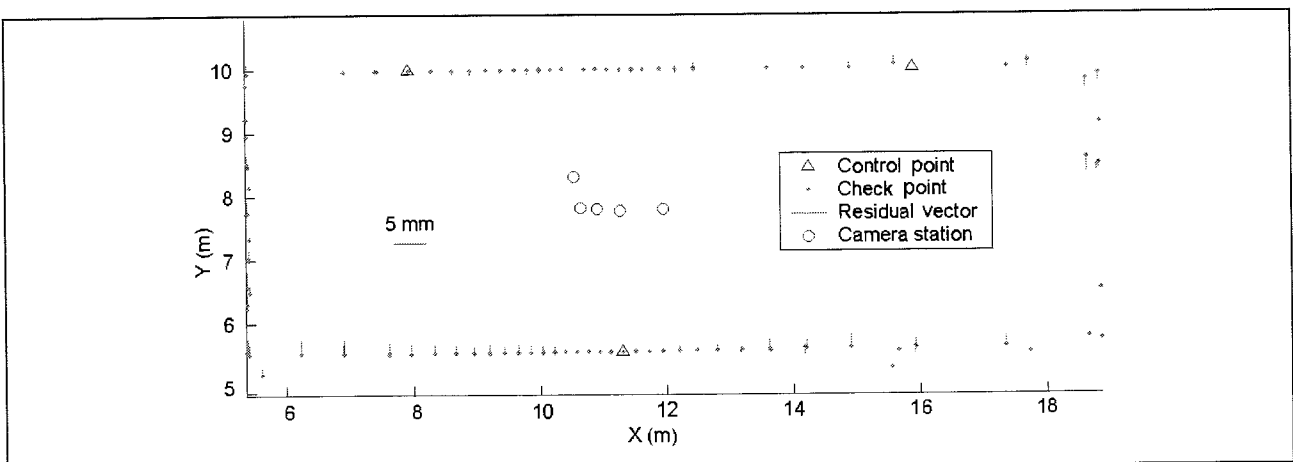


Figure 11. Object space residuals of check points for Z axis in XY-plane.

used to define the datum, since only 6 exterior orientation parameters are datum-dependent. The other parameters were well determined by tie points. However, in the mentioned networks, due to the specific geometry of the camera stations, the camera constant was defined as a priori known parameter to avoid high correlativity of this parameter with check point coordinates. In addition, in the second network tumbling parameters were defined as a priori known parameters since they cannot be determined with a few control points. Although tumbling parameters are not datum-dependent, these parameters model the partial deviations of the orientation parameters of the camera during rotation. Therefore, for a good and reliable estimating, many control points, depending on the number of the tumbling parameters are necessary. We estimated the tumbling parameters of the EYSCAN in the camera calibration process using all control points.

Table 2. Results of accuracy test (without tumbling modeling)

Number of check points	151
Number of control points	3
RMSE of check points (X,Y,Z) (mm)	9.72, 3.72, 3.60
STD of check points (X,Y,Z) (mm)	1.68, 0.64, 0.60
$\hat{\sigma}_0$ (pixel)	0.17 (1.36 microns)

Table 3. Results of accuracy test (with tumbling modeling)

Number of check points	151
Number of control points	3
RMSE of check points (X,Y,Z) (mm)	1.22, 1.04, 0.84
STD of check points (X,Y,Z) (mm)	1.58, 0.60, 0.54
$\hat{\sigma}_0$ (pixel)	0.16 (1.28 microns)

5. CONCLUSIONS

We developed an advanced sensor model for panoramic cameras and showed its accuracy performance. We indicated the improvement of the sensor model by the modeling of the tumbling for two terrestrial panoramic cameras EYSCAN and SpheroCam. We measured the tumbling of the SpheroCam using a physical instrument, an inclinometer. The tumbling of the EYSCAN and also the SpheroCam was estimated after bundle adjustment process, in which the tumbling parameters were defined as additional parameters. We performed self-calibration with/without tumbling parameters for EYSCAN and SpheroCam to show the effect of the tumbling modeling. The estimated standard deviations for the observations in image space are 0.59 pixel for the SpheroCam and 0.33 pixel for the EYSCAN in the case of using all mentioned additional parameters, which shows subpixel accuracy for these dynamic systems.

We also investigated the minimal number of control points for determining additional parameters. For the accuracy test 3 control points and 151 checkpoints were used, in which tumbling parameters were considered as a priori known parameters. The achieved accuracy in object space is 1.22, 1.04, 0.84 mm for the three coordinate axes (X, Y, Z) and is reasonable compared to the computed standard deviations. As mentioned before, tumbling parameters were determined in a camera calibration by means of control points. However, other methods should be investigated for determining the tumbling parameters, such as integration of a real time inclinometer or using additional object space information like straight lines, right angles, etc.

The accuracy test with minimal number of control points confirmed that with these new devices we have additional powerful sensors for image recording and efficient 3D modeling. For the near future we plan to investigate further into aspects of network design based on the characteristics of the panoramic cameras and 3D object reconstruction.

ACKNOWLEDGEMENTS

We appreciate the cooperation of D. Schneider, TU Dresden, who provided us with the image coordinates and control point coordinates for the testing of the EYSCAN camera. We are also grateful to Prof. Dr. L. Hovestadt, ETH Zurich, who rented us his group's SpheroCam for the testfield investigations.

REFERENCES

- Amiri Parian, J. and Gruen, A., 2003. A Sensor Model for Panoramic Cameras. In Gruen/Kahmen (Eds.), 6th Conference on *Optical 3D Measurement Techniques*, Zurich, Switzerland, Vol. 2, pp. 130-141.
- Amiri Parian, J. and Gruen, A., 2004. A Refined Sensor Model for Panoramic Cameras. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, part 5/W16. ISPRS "Panoramic Photogrammetry Workshop", Dresden, Germany, 19-22 February 2004.
http://www.commission5.isprs.org/wg1/workshop_pano/
- Hartley, R., 1993. Photogrammetric Techniques for Panoramic Camera. *SPIE Proceedings*, Vol. 1944, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision. Orlando, USA, pp. 127-139.
- Matthias, H., 1961. Umfassende Behandlung der Theodolitachsenfehler auf vektorieller Grundlage unter spezieller Berücksichtigung der Taumelfehler der Kippachse. Verlag Leemann, Zürich.
- Scheibe, K., Korsitzky, H., Reulke, R., Scheele, M. and Solbrig, M., 2001. EYSCAN - A High Resolution Digital Panoramic Camera. Robot Vision: *International Workshop RobVis 2001*, Auckland, New Zealand, Volume 1998, pp. 77-83.
- Schneider, D. and Maas, H.-G., 2003. Geometric modeling and calibration of a high resolution panoramic camera. In Gruen/Kahmen (Eds.), 6th Conference on *Optical 3D Measurement Techniques*, Zurich, Switzerland, Vol. 2, pp. 122-129.
- Schneider, D. and Maas, H.-G., 2004. Application of a geometrical model for panoramic image data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, part 5/W16. ISPRS "Panoramic Photogrammetry Workshop", Dresden, Germany, 19-22 February 2004.
http://www.commission5.isprs.org/wg1/workshop_pano/
- Slama, C. C., 1980. *Manual of Photogrammetry*, Fourth edition. American Society for Photogrammetry and Remote Sensing, pp. 196-207.

AUTOMATED ROAD EXTRACTION AND UPDATING USING THE ATOMI SYSTEM - PERFORMANCE COMPARISON BETWEEN AERIAL FILM, ADS40, IKONOS AND QUICKBIRD ORTHOIMAGERY

E. Baltsavias^a, L. O'Sullivan^b, C. Zhang^a

^a Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) Zürich, ETH-Hönggerberg, CH-8093 Zürich, Switzerland - (chunsun, manos)@geod.baug.ethz.ch

^b Swiss Federal Office of Topography, Seftigenstr. 264, CH-3084 Wabern, Switzerland - Liam.OSullivan@swisstopo.ch

Commission III, Working Group III/4

KEY WORDS: 3D road reconstruction, automation, updating, mapping, performance evaluation, orthoimage, IKONOS, Quickbird, digital airborne sensor, high-resolution satellites

ABSTRACT:

In the recent years, the automated extraction of roads from digital images has drawn considerable attention due to the need for the efficient acquisition and updating of road data for geodatabases. The development of new digital aerial sensors and high-resolution satellite sensors signifies a revolutionary change in image acquisition and the possibility of fully digital processing from image acquisition to the generation of value-added products for various applications. At ETH Zurich in cooperation with and funded by the Swiss Federal Office of Topography (*swisstopo*), we have developed an operational system for the automated extraction of 3D road networks from imagery that integrates the processing of colour image data and existing digital spatial databases. The system focuses on rural areas, can use stereo or orthoimages and can determine 3D road axes, and possibly other attributes like width if the roads have a minimum width of ca. 3 pixels. Colour is of advantage but not a must, while a DTM or DSM is required. If no road database exists, it can be generated from scratch, using manual measurement of characteristic road seed points. The system has been extensively tested, mainly by *swisstopo*, on areas with diverse terrain relief and landcover types using different resolution stereo and orthoimages with good results. Recently, tests have been performed using ADS40, IKONOS and Quickbird data. This paper reports on the performance comparison of the ATOMI system using different sensor data in two varying test sites. The test results were qualitatively and quantitatively analysed using accurate reference data. Visual analysis and quantitative measures of accuracy, correctness and completeness are presented, with typical completeness and correctness values of over 90% and planimetric accuracy of 0.4 m to 1 m. The advantages and disadvantages using different sensor data for road network updating are also discussed.

1. INTRODUCTION

In modern map production, a shift has taken place from maps stored in analogue form on paper or film to a digital database containing topographic information. A digital topographic database is an essential part of a GIS. Recently, National Mapping Agencies (NMAs), especially in Europe, wish to generate digital landscape/topographic models that conform to reality and do not include map generation effects. In addition, various existing and emerging applications require up-to-date, accurate and sufficiently attributed digital data, especially of roads and buildings, including car navigation, tourism, traffic and fleet management and monitoring, intelligent transportation systems, internet-based map services, location-based services, etc. In 2002, two major European map providers and five car manufacturers started the project NextMAP to identify and evaluate the road database requirements for in-vehicle ITS (Intelligent Transportation Systems) and services applications, as well as the cost consequences involved for data capturing and data production techniques

(<http://www.ertico.com/activiti/projects/nextmap/home.htm>).

Also in 2002, twelve organisations from NMAs, road administrations and private sector key players of road data market submitted the HERDS (Harmonized European Road Data Solution) project proposal for EC funding. Furthermore, in the European Territorial Management Information

Infrastructure project, roads are mentioned together with elevation and hydrography as the only objects, commonly agreed to be important enough to be defined as reference data, needed by most applications (see <http://www.ec-gis.org/etemii/reports/chapter1.pdf>).

To cope with higher product demands, increase the productivity and cut cost and time requirements, automation tools in the production should be employed. As aerial images are a major source of primary data, it is obvious that automated aerial image analysis can lead to significant benefits. In addition, the development of new digital aerial sensors and high-resolution satellite (HRS) sensors signifies a revolutionary change in image acquisition and the possibility of fully digital processing from image acquisition to the generation of value-added products for various applications. At ETH Zurich, in cooperation with the Swiss Federal Office of Topography (*swisstopo*), we have developed a practical system for the automatic extraction of 3D road networks from imagery that integrates processing of colour images and existing digital spatial databases, within the project ATOMI. Some reports on the system performance can be found in Baltsavias and Zhang (2003) and Zhang (2003b). This paper reports on the performance of the ATOMI system using extensive areas with varying relief and landcover and images from different sensors.

2. BRIEF DESCRIPTION OF PROJECT ATOMI

2.1 Aims of ATOMI

The aim of ATOMI is to update roads digitised from 1:25,000 scale maps (part of the national VEC25 dataset) by fitting them to the real landscape, improve their planimetric accuracy to 1m and derive road centerline heights with an accuracy of 1 to 2 m. The topology and the attributes of the existing datasets should be maintained. This update should be achieved by using the image analysis techniques developed at the Institute of Geodesy and Photogrammetry, ETH Zurich (IGP). The whole procedure should be implemented as a standalone software package, should be operational, fast, and most importantly reliable. We do not aim at full automation (ca. 80% completeness is a plausible target), but the "correct" results should be really correct to avoid checking manually the whole dataset. After some initial work, the aims of ATOMI were restricted to improvement of the VEC25 (i.e. no extraction of new roads) with the first target being the open rural areas. More details of ATOMI can be found in Eidenbenz et al. (2000).

The standard input data used includes 1:16,000 scale colour imagery, with 30-cm focal length, and 60%/20% forward/side overlap, scanned with 14 microns at a Zeiss SCAI, a nationwide DTM (DHM25) with 25-m grid spacing and accuracy of 1-3/5-8 m in lowlands/Alps, the vectorised map data (VEC25) of 1:25,000 scale, and the raster map with its 6 different layers. The VEC25 data have a RMS error of ca. 5-7.5 m and a maximum error of ca. 12.5 m, including generalisation effects. They are topologically correct, but due to their partly automated extraction from maps, some errors exist. In some cases, DSM data in the working area was generated using matching (without subsequent editing) on commercial digital photogrammetric workstations with 2-m grid spacing. In the meantime, a much better DTM and DSM (with 2-m spacing and 0.5-m and 1.5-m accuracy in nonforest and forest areas) produced by airborne laser scanning exists for large areas and will be soon finished for all Swiss regions up to 2000 m height, but has not been used up to now.

2.2 The Road Reconstruction System

Our developed system makes full use of available information about the scene and contains a set of image analysis tools. The management of different information and the selection of image analysis tools are controlled by a knowledge-based system. In this section, a brief description of our strategy is given. We refer to Zhang (2003a) for more details. The initial knowledge base is established by the information extracted from the existing spatial data and road design rules. This information is formed in object-oriented multiple object layers, i.e. roads are divided into various subclasses according to road type, landcover and terrain relief. It provides a global description of road network topology, and the local geometry for a road subclass. Therefore, we avoid developing a general road model; instead a specific model can be assigned to each road subclass. This model provides the initial 2D location of a road in the scene, as well as road attributes, such as road class, presence of roadmarks, and possible geometry. A road is processed with an appropriate method corresponding to its model, certain features and cues are extracted from images, and roads are derived by a proper combination of cues. The knowledge base is then automatically updated and refined

using information gained from previous extraction of roads. The processing proceeds from the easiest subclasses to the most difficult ones. Since neither 2D nor 3D procedures alone are sufficient to solve the problem of road extraction, we make the transition from 2D image space to 3D object space as early as possible, and extract the road network with the mutual interaction between features of these spaces.

The system can extract roads with a minimum width of ca. 3 pixels. It focuses on extraction of roads in open rural areas, by excluding roads in forest and urban areas using the existing information about the borders of these landcover classes. The existing road database information is used not only for giving an approximate position but also (a) to bridge and fill-in gaps in the extracted roads, and (b) to copy this information in nonprocessed areas (forest, urban) and connect it to the extracted road network in open rural areas. The aim of these two usages is to provide as final result a complete network (even if partially incorrect) avoiding results which consist of a set of broken and unconnected road segments. The system has been modified to work also with orthoimages, whereby the 3D information is extracted by overlaying the 2D information on the DSM or DTM. Although orthoimages have certain disadvantages compared to 2 or more images, the main being the inaccuracies introduced by the DTM/DSM during their generation, they are much easier to handle, are sensor independent and most importantly lead to reduced input data and much faster processing, a crucial factor for operational production.

Our system includes tools for external evaluation of the extracted results, by comparing the extracted results with precise reference data. The quality measures used in this work aim at assessing completeness and correctness as well as geometric accuracy. Completeness measures the percentage of the reference data that lies within the buffer of the extracted roads, while correctness is the percentage of the extracted roads within the buffer of the reference data (Heipke et al., 1998). The buffer distance is defined using the required accuracy of the project ATOMI, i.e. 1 m. The geometric accuracy is assessed by the mean and RMS of the distances between the extracted roads and the reference data. The detailed description for the computation of the external evaluation measures is presented in Zhang (2003a).

The developed system has been implemented as a stand-alone package initially on SGI platforms for stereo and orthoimages and has been ported to Windows XP only for orthoimage processing, with the same user interface. The system imports imagery, the existing road database and other input data (e.g. DSM/DTM). The extracted road network as well as the computed road attributes including length and width are saved in 3D Arc/Info Shapefile format that is readily imported into existing GIS software. For the technical details of the system, we refer to Zhang (2003a, 2003b). The Windows XP version for orthoimages is termed ATOMIRO (with R standing for roads and O for orthoimages). All current and further improvements of the system and the tests reported here refer to ATOMIRO, while the SGI versions have been frozen.

3. TEST SITES AND DATA DESCRIPTION

Results from two test sites in Switzerland will be presented here, one in Thun and the other one close to the city of Geneva. The selection is mainly based on the consideration that the test

sites should cover as many types of typical landcover in Switzerland as possible. Another consideration is the availability of images from multiple sensors. Both sites are in open rural areas but with different landcover. All road types in Switzerland can be found in the areas. The description of the test sites and the available imagery are listed in Table 1. Fig. 1 shows aerial images of the two test sites. Much larger and different regions have been used for tests by *swisstopo* with a total road length of about 9,000 km.

In Thun, the colour orthoimages were produced by *swisstopo* from aerial images of scale 1:16,000 using the DHM25. The 50-cm orthoimage is part of the nationwide dataset *Swissimage* (produced from 1:30,000 imagery with 15 cm lens) with a planimetric accuracy of about 1 m. The images for 20 cm and 60 cm were taken in spring 2003, and for 50 cm in summer 1998. An orthoimage created from ADS40 summer images using the DHM25 is also available. Due to weaknesses in the control point distribution and the bundle adjustment of the ADS40 images, a discrepancy between the ADS40 orthoimage and the 20cm orthoimage has been observed. A non-exhaustive comparison with manually selected feature points shows that the discrepancy varies between 0 and 80 cm. However, smaller differences also exist between the 1998 and 2003 aerial film orthoimages, caused by errors in the sensor orientation. Thus, the real accuracy of road extraction in image space is higher than the accuracy values derived from comparison between datasets (incl. the reference data), which have varying orientation errors.

	Thun	Geneva
Area (sq. km)	2.66 * 2.68	4.38 * 3.0
Height range (m)	560 ~ 2200	375 ~ 1200
Landscape	Open rural, Villages, Many small settlements	Open rural, Several villages, Forest, Large fields with bare soil
Imagery type (orthoimage pixel size)	aerial film (20cm, 50cm, 60cm) ADS40 (30cm)	aerial film (50cm) IKONOS PSM (1m) Quickbird PSM (70cm)

Table 1. Test site description and image specifications.

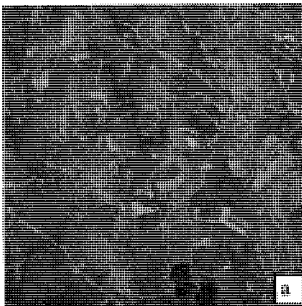


Figure 1. Overview of test sites: (a) Thun, (b) Geneva.

The Geneva test site (Fig. 1b) is near the city of Geneva, containing several larger villages, forest and a river. Another difference to the Thun site is that the scene contains grasslands and large fields of bare soil. In addition, many road-like lines are observed in the fields. The aerial orthoimage came from the *Swissimage* dataset. IKONOS and Quickbird images were also acquired in May 2001 and July 2003 respectively. The pansharpened (PSM) orthoimages of IKONOS and Quickbird were produced by a software system developed at IGP using a 2-m grid laser DTM with 0.5-m accuracy and had a planimetric accuracy of 0.5-0.8 m, estimated using 20-50 check points measured in 25-cm orthoimages of the Canton Geneva, produced using the same laser DTM (however with 1m grid spacing) and with ca. 0.5-m planimetric accuracy).

The reference data for the Thun and Geneva test sites were measured manually, by *swisstopo* in 20-cm pixel size aerial orthoimages and by ETH Zurich in the *Swissimage* orthoimages, respectively. The tests were performed on a DELL PC with Pentium 4, 1.8GHz CPU and 1GB RAM running Windows XP.

4. RESULTS AND DISCUSSION

4.1 Thun Site

Completeness and correctness is sufficient for all images in the Thun site, with slightly inferior results for the 50-cm and 60-cm pixel size orthoimages. Although the pixel size of the ADS40 image is slightly more than the 20 cm of the aerial film orthoimage, the results achieved are almost identical. Typical results of road reconstruction and junction generation are presented in Figs. 2-5, where the VEC25 and the extracted roads are shown as white and black lines. In each figure, (a), (c), (d) are the orthoimages with pixel size 20 cm, 50 cm and 60 cm respectively, while the 30-cm pixel size ADS40 orthoimage is shown in (b).

Fig. 2 is a scene with a four-road junction. Road surface and road sides are clear except at the left side of the figures, where a tree occludes the road. The scenes in Fig. 3 and Fig. 4 are slightly complex compared with that in Fig. 2. More shadows and occlusions are observed. In the settlement areas, some road sides are not defined. In Fig. 5, a first-class road is connected with two third-class roads at two junctions. The roadmarks on the first-class road are visible in all images, but are weak in the lower resolution images. The examples show that roads are generally correctly extracted from all images. Road junctions are also well formed. This observation is confirmed by the external evaluation of the extraction results using the reference data (see Table 2). To account for the discrepancy between the ADS40 and aerial film orthoimages, the buffer distance was set

to 2 m, when assessing the results from the ADS40 orthoimage.

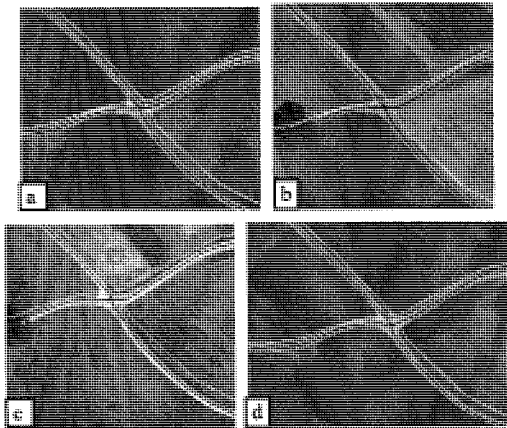


Figure 2. Examples of road extraction and junction generation in scenes with well defined roads.

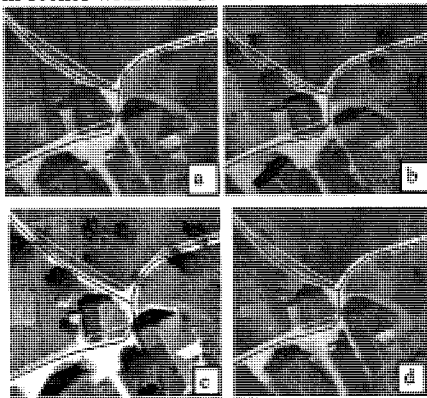


Figure 3. Examples of road extraction and junction generation in scenes with small settlements.

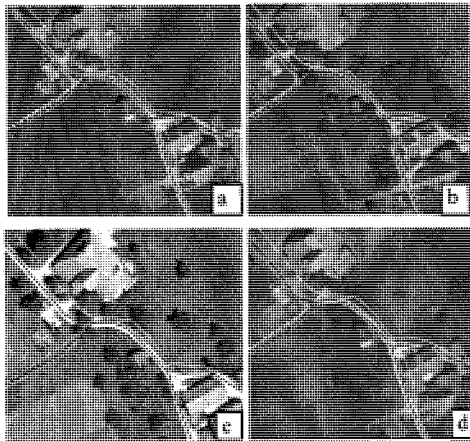


Figure 4. Examples of road extraction and junction generation in scenes with small settlements.

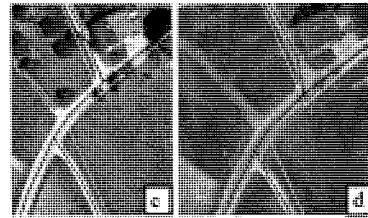
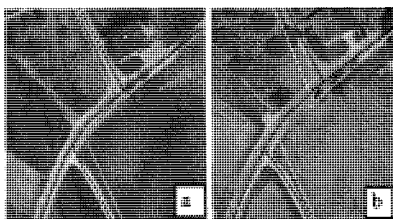


Figure 5. Results in a scene with a first-class road.

Our system delivers the best results with the 20-cm orthoimage. About 95% of the roads are correctly extracted with an accuracy of about 50 cm. The non-extracted or falsely extracted roads are mainly in small villages. Taking into account the discrepancy between the ADS40 and the 20-cm film orthoimage (as indicated by the large mean values), our system performed equally well on ADS40 data. Indeed, visual check over the whole test site shows that the results from ADS40 data are actually at the road centers. Furthermore, images from ADS40 are sharper and have a better radiometric quality compared to scanned film. Image quality is important in general and for object extraction, and depends on such image properties as well, and not only on the ground pixel size (which is often used wrongly as synonymous to image resolution and image quality).

Table 2 also shows that all quality measures are gradually deteriorating with decreasing pixel size. One cause for less completeness is that paths in fields are only partially extracted because the path surface is blurred and road edges are very weak (Fig. 6), while in small villages performance was also worse (Fig. 7). However, the quality deterioration is much less than the pixel size reduction. E.g. for 60-cm vs. 20-cm pixel size, in the first case we have 9 times less data, but completeness, correctness and accuracy deteriorate only by 7.5%, 4.5% and 40%. On the other hand, this slight quality decrease may still mean expensive additional manual editing, so the question of pixel size choice should be carefully considered.

Quality measures	Aerial 20cm	Aerial 50cm	Aerial 60cm	ADS40 30cm	
Completeness (%)	95.44	90.49	88.28	95.27	
Correctness (%)	94.65	92.53	90.68	94.24	
Length of reference (km)	42.08	42.08	42.08	42.08	
Length of extraction (km)	42.43	41.15	40.97	42.54	
RMS error (m)	x	0.45	0.60	0.72	1.00
	y	0.43	0.61	0.75	0.97
Mean error (m)	x	0.04	0.17	0.29	0.57
	y	0.08	0.22	0.19	0.44
Processing time (s)	1872	774	620	1184	

Table 2. Quality evaluation of the results in Thun site.

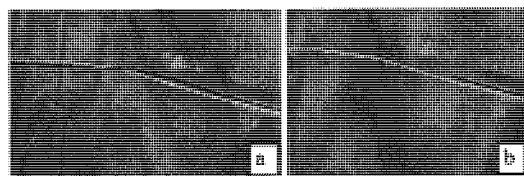


Figure 6. A path extracted in 20-cm pixel size image (a), but only partially extracted in 60-cm pixel size image (b).

The black lines are the extracted results and the white lines are the reference data.

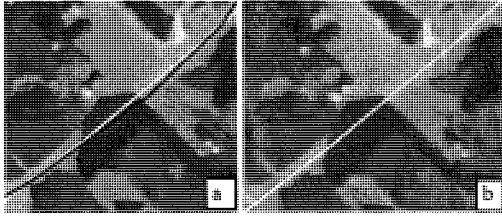


Figure 7. A village road extracted in 20-cm pixel size image (a), but not in 60-cm pixel size image (b). The black lines are the extracted results and the white lines are the reference data.

Table 2 shows that the processing speed of our system is high (e.g. ca. 30 minutes for more than 40-km roads in 20-cm pixel size images), and that processing time decreases almost 1:1 with orthoimage pixel size. The processing time also depends on the road density and to a lesser extent complexity of the scene, increasing with them. Extensive tests at *swisstopo* with 50-cm orthoimages show that roads in an average road density 1:25,000 map sheet covering 210 km² can be extracted in 3-4 hours on a Dell PC with Pentium 4, 2GHz CPU and 2GB RAM running Windows XP. Thus, using this not up-to-date computer configuration, all 1:25,000 map sheets of Switzerland could be processed in 36 days. Note that typical map sheets, excluding large urban centers, large lakes and the Alps, have about 2,500 km of roads, with about 45%-50% of them in rural areas.

4.2 Geneva Site

Our system achieves good results with the 50-cm orthoimage (Swissimage), similar to the ones in Thun (see Table 3). However, the performance (mainly the completeness) with the HRS data is poor, especially the 1-m IKONOS image. In this image, higher-class roads are usually extracted, while most narrow roads such as 4th, 5th and 6th class roads are not, because the system prerequisite of 3 pixel wide roads is not fulfilled. The increased ground resolution in Quickbird makes more roads visible than in IKONOS, and also the road surface and road edges are clearer, resulting in a better performance. However, compared with the 60-cm aerial film orthoimage in Thun, the completeness is still rather low.

Quality measures	Aerial 50cm	IKONOS-PSM 100cm	Quickbird-PSM 70cm
Completeness	90.89%	54.22%	72.68%
Correctness	95.36%	81.22%	89.58%
Length of reference (km)	50.72	50.72	50.72
Length of extraction (km)	48.35	33.87	42.16
RMS error (m)	x	0.62	0.93
	y	0.56	0.82
Mean error (m)	x	0.07	-0.73
	y	-0.05	0.34
Process time (s)	1510	992	924

Table 3. Quality evaluation of the results in Geneva site.

It is apparent that the definition quality of an object does not depend only on the pixel size but other image quality factors too, and that each object type can be favourably extracted

within an object-specific image scale range. Critical factors influencing image quality, like atmospheric and illumination conditions, sensor and sun elevation and image sharpness are much less or not controllable with spaceborne sensors compared to airborne ones, resulting thus in inferior image quality and object definition with the former, even if the ground pixel size is similar. Both HRS images lead to accuracy (RMS) of less than 1m. The mean values are high, due to a systematic bias caused by probable errors in the transformation from the coordinate system of Canton Geneva to the Swiss coordinate system. Thus, in reality the road accuracy from the HRS images is similar or slightly better than that from Swissimage, if the HRS orthoimages are produced with a submeter accuracy DSM/DTM (as in this case) or the sensor elevation is high. Fig. 8 shows several examples of extracted roads and road junctions from the Swissimage, IKONOS and Quickbird orthoimages.

In the Geneva test site, no extraction is applied to the roads inside the villages since the sizes of the villages are large and are classified as urban area. The non-extracted roads are usually those in fields with very weak edges. An example is given in Fig. 9. False extraction in Swissimage occurs when a road in fields is neighbouring with road-like lines (Fig. 10a). Several false extractions are also because the actual road width differs from the width expected for the given road class. This was noticed with several 5th and 6th class roads. An example is shown in Fig. 10b, where a ca. 6.6-m wide 5th class road is incorrectly extracted.

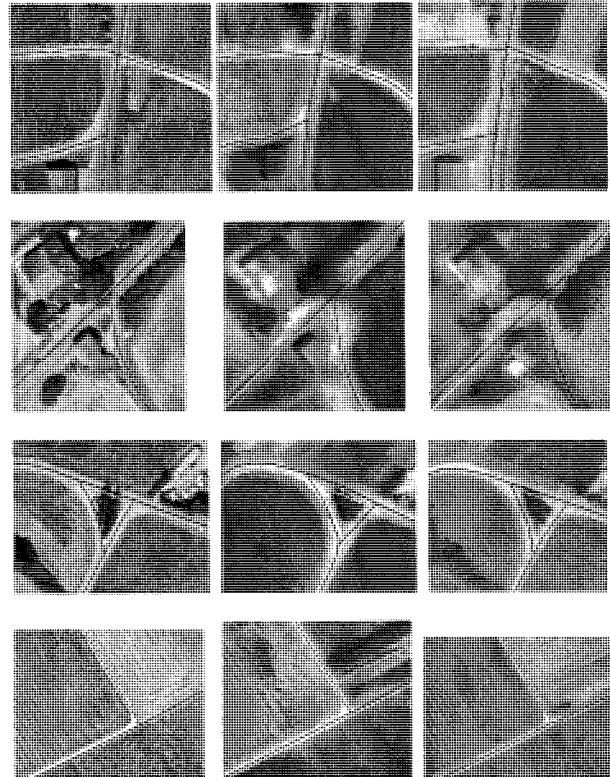


Figure 8. Examples of extracted roads and road junctions in the Geneva site orthoimages. The black lines are the results and the white lines are the VEC25 roads. Left: Swissimage, middle: IKONOS, right: Quickbird.

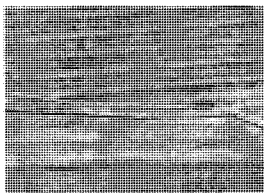


Figure 9. Road in field with weak edges can not be extracted from Swissimage. Black line: reference data. White line: VEC25 roads.

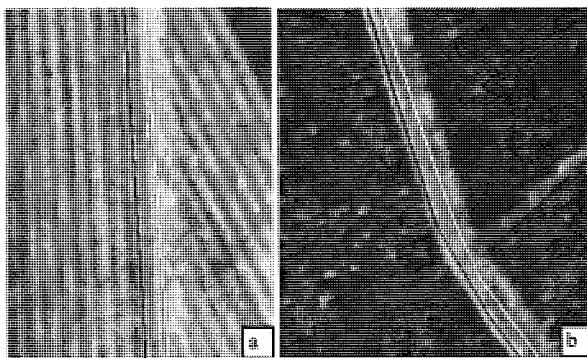


Figure 10. Examples of false extraction from Swissimage. Black line: extraction results. White lines: reference data. (a) a road is incorrectly extracted due to the interference of many road-like features. (b) false extraction caused by assuming wrong road width for the given road class.

Fig. 11 presents two examples (3 images in one row for each) to show the limitation when our system is applied to HRS data. In the figure, the VEC25 roads and extracted results are presented as white and black lines respectively, while the Swissimage, IKONOS and Quickbird orthoimages are shown from left to right. In both examples, the roads are extracted from Swissimage. The road shown in the first scene (first row) is not extracted in the IKONOS image, while the road in the second scene (second row) can not be extracted in the HRS data due to haze. Clouds prohibit road extraction in the example of Fig. 12.

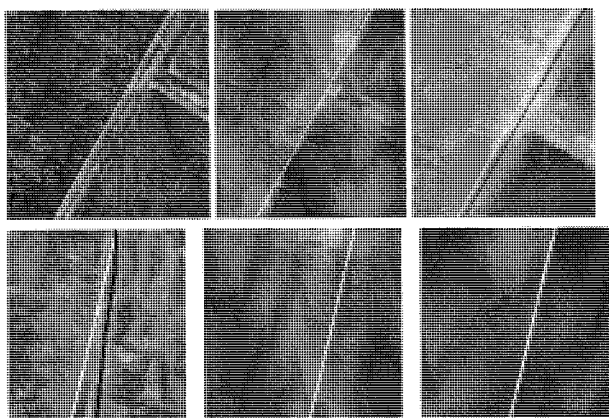


Figure 11. Examples showing limitations of our system applied to HRS data. See text for explanation.

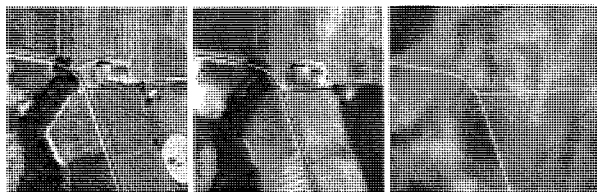


Figure 12. Example showing clouds in the Quickbird image preventing road extraction. Left: Swissimage, middle: IKONOS, right: Quickbird. The VEC25 road and extraction results are presented in white and black lines respectively.

5. CONCLUSIONS

In this paper, we have reported the performance comparison of the ATOMI road reconstruction system between aerial film orthoimages of varying pixel size, ADS40 and HRS orthoimages over two test sites in Switzerland, using accuracy, completeness and correctness quantitative measures and visual control. It is shown that about 95% of roads in rural areas are correctly extracted using aerial film and ADS40 orthoimages with 20-cm and 30-cm pixel size, respectively. With increasing pixel size, the system performance deteriorates but to a much less degree. However, even though the landcover of the two test sites is largely different, our system achieved in both ca. 90% completeness with 50cm aerial orthoimages. Thus, the general conclusion is that the ATOMI system can reconstruct road networks in rural areas using aerial orthoimages with maximum pixel size of ca. 50-60 cm with a completeness and correctness of 90%-95% and an accuracy of 0.4-0.7 m. The speed is sufficient for operational production, while the result includes both extracted and non-extracted (old) data resulting in a complete network with the topology and attributes of the input road database plus new derived attributes like road width. Using manual on-screen digitising of road seed points, the method can be extended to generation of a road database from scratch.

The test shows that the system performance is poor with the HRS data, especially for 1-m IKONOS PSM imagery. Both HRS can deliver submeter accuracy, however the problem lies with the poor object definition and image quality. Only half of the roads in the test sites are reconstructed, mainly higher-class roads with larger width. The surface of the narrow roads (lower-class roads) is usually blurred and the road edges are weak and not clear in the HRS images, thus most of the lower-class roads are not extracted. The test results show that the performance on the 70-cm Quickbird data is considerably better than that on the 1-m IKONOS data, but still of lower completeness than the 60-cm pixel size aerial orthoimage. Other extraction methods, not requiring road widths (ribbons) of 3 pixels or more, may be more appropriate for orthoimages with such pixel size. Generation of Quickbird orthoimages with 60 cm, or deployment of new HRS with 40 cm - 50 cm ground pixel size (license for which the US government has already provided) may pave the way for application of the current approach with good completeness even for such imagery, if the imaging conditions are favourable.

Our system can still be improved, for example, by better use of the existing road vectors to bridge gaps. Post-control on whether the solution conforms in shape and topology to road construction and intersection principles needs to be completed. The self-diagnosis and reliability measures derived for the

extraction results are not robust enough. Use of denser and more accurate laser DSMs/DTMs and of the NIR channel of digital sensors can be used for better quality results. Extension of the method to areas with low buildings and forest borders may be feasible. These and other aspects will be topics of future research.

ACKNOWLEDGEMENTS

We acknowledge the financial support and the data provided for this work and the project ATOMI by the Swiss Federal Office of Topography and the NPOC, Bern. Canton Geneva provided 25-cm orthoimages, a laser DSM and other data in the Geneva test site. Space Imaging USA provided the Rational Polynomial Coefficients for the IKONOS images. We also thank Zhang Li, Henri Eisenbeiss, Oliver Heller and Oliver Gut at ETH Zurich for providing the orthoimages of IKONOS and Quickbird.

REFERENCES

- Baltsavias, E., Zhang, C., 2003. Automated updating of road databases from aerial imagery. Proc. Workshop "Data quality in Earth Observation Techniques", ITC, Enschede, The Netherlands, 21 November.
- Eidenbenz, Ch., Kaeser, Ch., Baltsavias, E.P., 2000. ATOMI – Automated Reconstruction of Topographic Objects from Aerial Images using Vectorized Map Information. International Archives of Photogrammetry, Remote Sensing and SIS, Vol. 33, Part B3/1, pp. 462-471.
- Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1998. External evaluation of automatically extracted road axes. Photogrammetrie Fernerkundung Geoinformation (2), 81-94.
- Zhang, C., 2003a. Updating of cartographic road database by image analysis. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Report No. 79.
- Zhang, C., 2003b. Towards an operational system for automated updating of road databases by integration of imagery and geodata. ISPRS Journal of Photogrammetry and Remote Sensing 58(3-4), 166-186.

POTENTIAL OF IKONOS AND QUICKBIRD IMAGERY FOR ACCURATE 3D POINT POSITIONING, ORTHOIMAGE AND DSM GENERATION

H. Eisenbeiss, E. Baltsavias, M. Pateraki, L. Zhang

Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093, Zurich, Switzerland - (ehenri, manos, maria, zhangli)@geod.baug.ethz.ch

Thematic Session 20 – Applications of High Resolution Data

KEY WORDS: IKONOS, QUICKBIRD, high resolution, radiometric preprocessing, sensor models, point positioning, orthoimage, DSM, accuracy analysis

ABSTRACT:

This paper describes the processing of IKONOS and QUICKBIRD imagery of two different datasets in Switzerland for analyzing the geometric accuracy potential of these images for 3D point positioning, and orthoimage and DSM generation. The first dataset consists of panchromatic and multispectral IKONOS and QUICKBIRD images covering the region of Geneva. In the second area around Thun with a height range of ca. 1650 m, the dataset consisted of a triplet and a stereo pair with an overlap of 50 %. In both areas, laser DTM/DSM existed and in Geneva also aerial orthoimages. GCPs with an accuracy of 0.2-0.4 m have been used in both sites. The investigations for 3D point positioning included 4 different sensor models, different GCP measurement, variable number of control points and area covered by them. The results showed that the Rational Polynomial Coefficient (RPC) model compared to 2D and 3D affine models are more general and can model sufficiently imaging modes that depart from linearity. This is particular so for QUICKBIRD which needs after the use of RPCs an additional affine transformation in order to reach accuracies of 1m or less. With sufficient modeling, the planimetric accuracy was 0.4 – 0.5 m, even for few GCPs and only partly covering the images. Orthoimages were generated from both QUICKBIRD and IKONOS with an accuracy of 0.5-0.8 m, using a laser DTM. A sophisticated matching algorithm was employed in Thun. In spite of various difficult conditions like snow, long shadows, occlusions due to mountains etc., the achieved accuracy without any manual editing, was 1-5 m depending on the landcover type, while in open areas it was about 1 m. Under normal conditions, this accuracy could be pushed down to about 0.5 m. Thus, IKONOS, and to a lesser degree QUICKBIRD, could be an attractive alternative for DSM generation worldwide.

1. INTRODUCTION

1.1 Aims

The topic of this paper is the analysis of the potential of IKONOS and secondary QUICKBIRD (QB) for 3D point positioning, orthoimage and DSM generation. Two test sites, in Geneva and Thun, were used with accurate reference data and partly different aims. In both projects, there was a cooperation with the Swiss Federal Office of Topography (swisstopo) and Space Imaging (SI). In Geneva, the final aim was the investigation whether high-resolution satellite (HRS) imagery can be used for updating the Swiss national maps at foreign border areas, which has as prerequisite the generation of accurate orthoimages. Another aim was the analysis of accuracy of IKONOS and QB for 3D point positioning and orthoimage generation using Rational Polynomial Coefficients (RPCs) and other simpler sensor models. The HRS orthoimages will be compared to alternative information sources regarding feature interpretation and mapping by the swisstopo. In Thun, the main aim was accuracy investigations of IKONOS for point positioning and DSM generation using a block of images (2 strips with 5 images) over a terrain with large height range and very variable landcover. The whole processing was performed exclusively with software based on good quality algorithms and developed at our Institute, most of it part of an operational software package for processing of linear array digital imagery.

1.2 Datasets

In Geneva, we used two slightly overlapping IKONOS images (west and east, each about 10 km x 20 km) and one QB image covering the eastern and 60% of the western IKONOS images. In Thun, one stereo pair (eastern part) and a triplet (western part) of IKONOS images (each image 10 km x 20 km) were used, with each image group acquired on the same day (see Table 1). The two strips in Thun had a ca. 50% overlap, and the triplet images were covered in about 70% of the area by snow, while all images had long shadows. The nadir image in the triplet was very close to one image of the stereopair, which had a suboptimal base/height ratio. All IKONOS images were Geo, 11-bit with DRA off, with 1m panchromatic (PAN) and 4m multispectral (MS) channels (in Thun only PAN was used), while the QB image was Basic 1B, 11-bit, 0.63m PAN and 2.52m MS. IKONOS and QB images had associated RPC files. For the measurement of GCPs in the Geneva site we used in the Canton of Geneva orthoimages with 0.25 m pixel size and ca. 0.5 m accuracy, derived from 1 m laser DTM with 0.5 m accuracy and outside the Canton, Swissimage orthoimages with 0.5 m pixel size and 1 m accuracy, derived from a 25m DTM (DHM25) with ca. 2 m accuracy. The coordinates of the GCPs in Thun were measured with differential GPS. In all cases, GCPs were measured in the images semi-automatically using least squares and intersection of straight, long enough lines or ellipse fit. The control points have an accuracy of 0.2 - 0.4 m in object and image space. In Thun, a 2m laser DSM with an accuracy of 0.5 m - 1 m (1 sigma) for open areas and 1.5 m for vegetation areas was used as reference data for the DSM generation from IKONOS.

Image	Date of acquisition	Scanning mode	Sensor-Azimuth (deg)	Sensor-Elevation (deg)	Numbers of GCPs	GCP accuracy (m)	GCP measurement method
Geneva_Q	2003-07-29	Reverse	286.4	77.6	67	0.3-0.5	Orthoimage / laser DTM
Geneva_I_West	2001-05-28	Forward	253.6	67.2	34	0.3-0.5	Orthoimage / laser DTM
Geneva_I_East	2001-05-28	Reverse	240.2	61.6	44	0.3-0.5	Orthoimage / laser DTM
Thun_I_49_000	2003-12-11	Reverse	140.35	62.78	25	0.2-0.3	GPS
Thun_I_49_100	2003-12-11	Reverse	66.41	63.56	25	0.2-0.3	GPS
Thun_I_51_000	2003-12-25	Reverse	180.39	62.95	24	0.2-0.3	GPS
Thun_I_51_100	2003-12-25	Reverse	72.206	82.13	24	0.2-0.3	GPS
Thun_I_54_000	2003-12-25	Forward	128.17	82.62	24	0.2-0.3	GPS

Table 1. Specifications of used satellite images and respective GCPs (Q stands for QUICKBIRD and I for IKONOS).

2. IMAGE ANALYSIS

2.1 Radiometric Quality

HRS usually employ TDI technology. All IKONOS and QB images have been acquired using 13 stages of the TDI. A higher number of stages would increase the signal but also the danger of saturation, especially for bright objects. TDI results in smoothing and a reduction of the MTF. MTF is always applied by SI and although in the QB metadata nothing is mentioned, it is fairly probable that a similar process is applied. DRA is optional with IKONOS, but with QB although again nothing is mentioned in the metadata, it seems that it is applied by default (this is indicated by the respective histograms which show saturation in the maximum grey value of 2047). The histograms of both IKONOS and QB show that only 8-9 bit are essentially used, while the blue channel has the smallest range of grey values.

The noise characteristics of the images were analysed and quantified using the standard deviation of the gray values in homogeneous (Lake of Geneva, Lake of Thun) and inhomogeneous areas (large image parts without homogeneous areas). The use of homogeneous areas is justified as noise is especially visible in such areas, whereas the use of inhomogeneous areas allows an analysis of the noise variation as a function of intensity and when homogeneous areas are missing. Baltasavias et al. (2001), in their first assessment of IKONOS Geo, give a short description of the method utilised for noise estimation. The method has been modified regarding noise estimation in inhomogeneous areas, in order to adapt computation of the standard deviation according to the number of significant samples in each bin (grey level range). Homogeneous areas existed only in the IKONOS East image of Geneva and the eastern Thun stereo pair. In QB, due to wind, the water surface was not homogeneous and could not be used. The mean standard deviation is computed out of the N% (here 85%) smallest percentage of samples. According to Table 2, the noise in the Thun images is slightly less than in Geneva and the MS exhibit less noise than the PAN ones, possibly due to the 4 times larger pixel size. Considering the fact that the 11-bit data represent actually only 8-9 bit, the noise is quite high for PAN, a fact that could be verified visually by strong image contrast enhancement.

Estimation of noise in inhomogeneous areas uses as input a range of standard deviations in each bin, based on which a percentage is computed. The standard deviation in homogeneous areas is used to compute the input range. For the Geneva IKONOS images, the values of the input range were set to 3.5 for PAN and 1.5 for MS. For the QB PAN, the range

has been empirically set to 1.7. Table 3 shows the results for the PAN channels, whereby the values for IKONOS are average values. Table 3 indicates that noise is intensity dependent for all images, however for QB the noise increases less with intensity. When the number of samples in a bin is less than 50, no value is given. The lower noise of QB may be due to a better preprocessing of the QB images, or due to the imaging conditions (e.g. higher elevation), or due to the fact that QB while scanning the scene, e.g. from North to South continuously rotates from South to North in order to achieve the nominal pixel size for PAN, thus oversampling. But it can also be accidental, or due to uncertainties in noise estimation in inhomogeneous areas. Thus, more tests with QB images involving also homogeneous areas are needed. For the MS channels, in both IKONOS and QB, the noise pattern is similar to PAN, however due to the lower dynamic range (shorter integration time), less bins have a significant number of samples.

Ikonos images	Red	Green	Blue	NIR	PAN
Geneva East	1.89	2.34	2.35	2.02	5.05
Stereo A Thun	1.54	1.98	2.21	1.73	4.20
Stereo B Thun	1.61	1.77	1.93	1.66	4.13

Table 2. Noise estimation for homogeneous areas in IKONOS images.

PAN Scenes	0 - 127	128 - 255	256 - 383	384 - 511	512 - 639	640 - 767	768 - 895
Geneva_I	-	3.46	3.67	4.03	4.20	5.61	6.26
Geneva_Q	1.26	1.35	1.38	1.33	1.47	2.14	2.93
Thun stereo	1.81	1.95	3.26	5.54	-	-	-
Thun triplet	1.82	2.38	2.53	2.99	3.47	4.59	-

Table 3. Noise estimation for inhomogeneous areas and different grey value ranges (bins) in PAN images.

All images were found to exhibit artifacts, which were visible, especially in homogeneous area and/or after strong contrast enhancement. Stripes in flight direction due to imperfect calibration of the sensor elements. Strong reflections in both PAN and MS images, which lead to saturation of the signal and loss of information. Spilling (Fig. 1(a), IKONOS, 1(c) QB) of bright target response in neighbouring lines in the flight direction, visible almost exclusively in the PAN images and blooming (Fig. 1 (b), IKONOS).

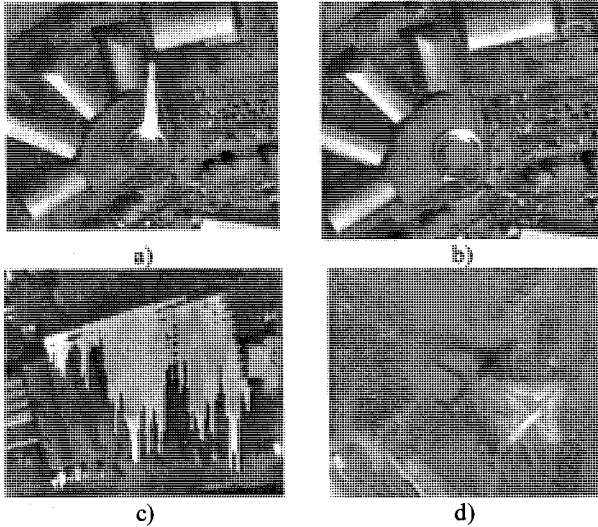


Figure 1. Artefacts.

Spilling is probably the most grave radiometric problem, as it destroys image information and may confuse subsequent feature and object extraction. It increases with smaller pixel size and with smaller angle between the line-of-sight of the sensor and the reflected sun rays. It is pronounced because of the TDI and increases when more TDI stages are used. It is apparent that with bright targets the respective TDI pixels are saturated and the excess signal is not properly discharged, influencing subsequent lines. QB has much more spilling (more often, longer and wider) due to its smaller pixel size but also due to its continuous rotation during imaging. In Geneva, QB had 135 artifacts compared to 10 and 18 for IKONOS East and West. Ghosting of moving objects (Fig. 1 (d), QB) is visible in pansharpened images, due to the time difference in the acquisition of the PAN and MSI images. Another factor influencing image quality are shadows. In both IKONOS and QB images, most shadowed areas (especially in urban areas) did not have significant signal variation, even after strong contrast enhancement. However, in the winter images of Thun, very large open shadowed areas of mountain cliffs covered by snow could be enhanced quite successfully.

2.2 Image Preprocessing

In order to improve the radiometric quality and optimize the images for subsequent processing, a series of filters are applied. The performed preprocessing encompasses noise reduction, contrast and edge enhancement and reduction to 8-bit by non-linear methods. All filters are applied to the 11 bit data.

Noise reduction filters aim at reducing noise, while sharpening edges and preserving corners and one pixel wide lines. The two local filters employed have similar effects although they use different parameters (Baltasvias et al., 2001). In Fig. 2, the Adaptive Edge Preserving Weighted Smoothing is compared to a Gaussian filter. Apart from the visual verification, reduction of noise was quantified by noise estimation in inhomogeneous areas. Comparing Tables 3 and 4, a reduction of noise by a factor of about 2.5 - 3.0 and 1.8 for PAN IKONOS and QB, respectively, is estimated. Following noise reduction, local contrast enhancement is applied using the Wallis filter. Moreover, 11-bit data are reduced to 8-bit by an iterative non-linear method in order to preserve the grey values that are more frequently occurring. Two different approaches are

implemented, one with flat frequency of output grey values and one with Gaussian form frequency, the latter being applied here. The improvement of the image after preprocessing is shown in Fig. 3.

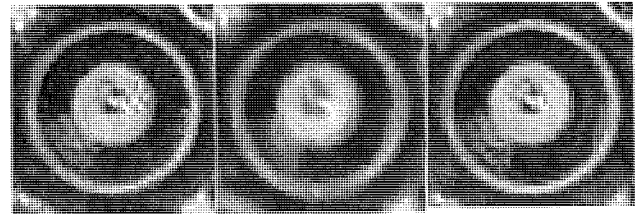


Figure 2. Effect of filtering: (left to right) Original image, Gaussian 5x5 filter, Adaptive Edge Preserving Weighted Smoothing.

PAN Scenes	0 - 127	128 - 255	256 - 383	384 - 511	512 - 639	640 - 767	768 - 895
Geneva_I	-	1.04	1.01	1.17	1.21	1.84	1.90
Geneva_Q	0.80	0.86	0.89	0.88	0.82	0.98	1.24
Thun stereo	0.53	0.54	1.56	2.55	-	-	-
Thun triplet	0.54	0.76	0.81	1.00	1.36	1.94	-

Table 4. Noise level in inhomogeneous areas and different grey value ranges (bins) for IKONOS scenes, after noise reduction.

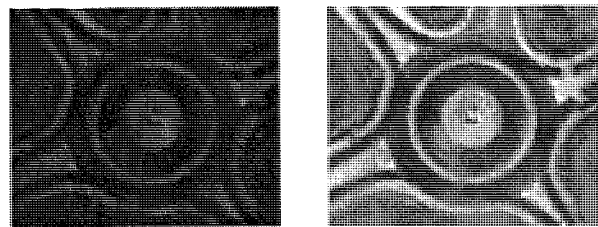


Figure 3. IKONOS image before (left) and after (right) preprocessing.

3. IMAGE ORIENTATION

3.1 Methods and Sensor Models

With the supplied RPCs and the mathematical model proposed by (Groddecki and Dial, 2003), a bundle adjustment is performed. The model used is:

$$x + \Delta x = x + a_0 + a_1 x + a_2 y = RPC_x(\varphi, \lambda, h)$$

$$y + \Delta y = y + b_0 + b_1 x + b_2 y = RPC_y(\varphi, \lambda, h)$$

where a_0, a_1, a_2 and b_0, b_1, b_2 are the affine parameters for each image, and (x, y) and (φ, λ, h) are image and object coordinates of points.

Using this adjustment model, we expect that a_0 and b_0 absorb most errors in the exterior and interior orientation. The parameters a_1, a_2, b_1, b_2 are used to absorb the effects of on-board GPS and IMU drift errors and other residual effects. In our approach, we first use the RPCs to transform from object to image space and then using these values and the known pixel coordinates we compute either two translations (model RPC1) or all 6 affine parameters (model RPC2).

For satellite sensors with a narrow field of view like IKONOS and QB, simpler sensor models can be used. We use the 3D affine model (3daff) and the relief-corrected 2D affine (2daff) transformation. They are discussed in detail in Fraser et al. (2002) and Fraser (2004). Their validity and performance is

expected to deteriorate with increasing area size and rotation of the satellite during imaging (which introduces nonlinearities), while the 3D affine model should perform worse with increasing height range and in such cases is more sensitive than the 2D affine model in the selection of GCPs.

3.2 Measurements of the GCPs

In Geneva, some roundabouts and more straight line intersections (nearly orthogonal with at least 10 pixels length) were measured semi-automatically in the satellite images and the aerial orthoimages (see Fig. 4). Measurement of GCPs by least squares template matching (Baltsavias et al., 2001) was not convenient or possible due to highly varying image content and scale. The height was interpolated from the DTM used in the orthoimage generation. An unexpected complication was the fact that the Canton of Geneva is using an own coordinate system and not the Swiss one! The transformation from one system to the other is not well defined, and based on different comparisons of transformed Geneva coordinates and respective coordinates in the Swiss system, a systematic bias has been observed, indicating that the results listed below could have been better. In Thun, the same image measurement approach was used, however, roundabouts (which are better targets) were very scarce. As expected, well-defined points were difficult to find in rural and mountainous areas, especially in Thun, where they had to be visible in 5 images simultaneously, while shadows and snow made their selection even more difficult. The object coordinates in Thun were measured with differential GPS. GPS requires work in the field, but the accuracy obtained is higher (espec. in height) and more homogeneous than using measurements in orthoimages, which have varying accuracy with unknown error distribution (due to the DSM/DTM). The number of GCPs and their accuracy are listed in Table 1.

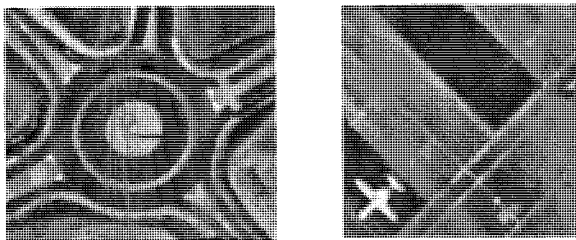


Figure 4. Examples of GCP measurement with ellipse fitting (left) and line intersection (right).

3.3 Comparison of different sensor models

In Geneva, we compared various sensor models, IKONOS vs. QB and analysed the influence of the number of GCPs. Due to lack of space, only the most important results will be shown here.

Tables 5 and 6 show the results for the transformation from object to image space. Three different GCP configurations are used with all, 10 and 4 GCPs. Table 5 shows that with all GCPs, in IKONOS-East, all 4 sensor models have similar performance, with RPC2 being slightly better. In IKONOS-West (with forward scanning) results are similar for RPC1 and RPC2, a bit worse in y with 2D affine and considerably worse for 3D affine. The latter model deteriorates more with reduction of GCPs and is more sensitive to their selection. For the other models, the accuracy reduction from 44 to 4 GCPs is very modest, verifying findings from previous investigations

that the number of GCPs is not so important, as their accuracy and secondary their distribution. The results for the 3D affine were initially by some factors worse than the ones of Table 5, when using geographic coordinates instead of map coordinates (oblique Mercator). The dependency of the results on the coordinate system has been discussed by Fraser (2004), albeit with smaller differences than the ones noted here.

Model	GCP	CP	x-RMS [m]	y-RMS [m]	Max. Δx [m]	max. Δy [m]
rpc1	44	-	0.65	0.56	1.40	1.21
rpc2	44	-	0.54	0.42	1.53	0.98
3daff	44	-	0.55	0.41	1.40	0.81
2daff	44	-	0.55	0.47	1.39	1.18
rpc2	10	34	0.57	0.52	1.52	1.07
rpc2	4	40	0.60	0.50	1.63	1.13
rpc1	4	30	0.63	0.40	1.35	1.40
rpc2	4	30	0.61	0.54	1.63	1.13
3daff	4	30	1.25	4.16	3.83	15.70
2daff	4	30	0.66	0.83	1.39	1.32

Table 5. Comparison of sensor models and number of GCPs with IKONOS-East (Geneva). At the bottom, one example for IKONOS-West. CP are the check points.

QB (see Table 6) is much less linear than IKONOS (expected partly due to its less stable orbit and pointing, and continuous rotation during imaging). Only RPC2 performs with submeter accuracy and only with this model can QB achieve similar accuracy as IKONOS. A residual plot with RPC1 shows a very strong x-shear. The 2D and 3D affine transformations are totally insufficient for modelling. As with IKONOS, a reduction of the GCPs has not any significant influence with RPC2. Thus, using simple RPCs (as in most commercial systems), or even applying 2 shifts in addition, will not lead to very accurate results with QB. It should be noted here that the QB image was Basic, i.e. not rectified. It is expected that a rectified image will show a more linear behaviour, and the respective RPCs will be more stable.

Model	GCP	CP	x-RMS [m]	y-RMS [m]	max. Δx [m]	max. Δy [m]
rpc1	67	-	2.64	0.43	5.57	0.92
rpc2	67	-	0.44	0.43	1.06	0.93
3daff	67	-	12.96	7.47	28.52	22.11
2daff	67	-	8.26	4.83	19.49	15.53
rpc2	10	57	0.46	0.44	1.12	0.97
rpc2	4	63	0.49	0.57	1.34	1.23

Table 6. Comparison of sensor models and number of GCPs with QB. CP are the check points.

For the Thun dataset, the triplet and stereo images were used separately in a bundle adjustment to determine object coordinates (processing of all images together was not possible due to a program limitation). Several semi-automatically measured (with least squares matching) tie points were included. The results for the triplet are shown in Table 7. The previous conclusions were verified, while the 3D affine model was worse compared to Geneva, probably because of the larger height range. A new indication compared to the Geneva data refers to the height accuracy. This is clearly better with RPC2,

and seems to get worse with decreasing number of GCPs, at least for this area with large height differences.

As a next step, we checked the role of the area covered by the GCPs, using always 5 GCPs (Table 8). RPC1 gave more or less similar results in planimetry, verifying previous investigations with the 2D affine model. The height however, is more sensitive to the position of the area covered by the GCPs, deteriorating in accuracy when GCPs were only in flat areas. Surprisingly, RPC2 gives clearly worse results than RPC1, especially when GCPs cover only 1/3 of the image area. This has been also verified with the Geneva images. A possible explanation is that after the RPCs are used, the scales and shears of the affine transformation model very small residual model errors. If in addition the GCP measurements are noisy (see e.g. the particularly high RMS at the mountainous south-west where GCP definition was poor), and the area covered is small, then these parameters may easily take wrong values. Grodecki and Dial (2003) mention the need to use only a linear factor in flight direction if the strip is long (about > 50 km). In future investigations, we will analyse to what extent the 4 scale and shear parameters are significant and determinable. These preliminary results indicate that RPC2 should be used with a GCP distribution covering most of the image area.

4. ORTHOIMAGE AND DSM GENERATION

The focus in the following text will be on the DSM generation in Thun. The results of the orthoimage generation in Geneva are analysed in Heller and Gut (2004). The accuracy of the orthoimages generated with the laser DTM and RPC2 with 10

GCPs gave an exceptional accuracy of 0.5 m - 0.80 m for both IKONOS and QB, with very typical sensor elevation values. These orthoimages are thus more accurate than the national Swissimage orthoimages, however interpretation of objects is more difficult.

4.1 DSM Generation Method

For DSM generation, a hybrid image matching algorithm was used (for details see Zhang and Gruen, 2003, 2004). Our method considers the characteristics of the linear array image data and its imaging geometry. The method can accommodate images from very high-resolution (3-7 cm) airborne Three-Line-Scanner images to HRS images like IKONOS, QB and SPOT-5. It can be used to produce dense, precise and reliable results for DSM/DTM generation. The final DSMs are generated by combining the matching results of feature points, grid points and edges. Matching is performed using cross-correlation and image pyramids. A TIN-based DSM is constructed from the matched features (whereby edges are used as breaklines) at each level of the pyramid, which in turn is used in the subsequent pyramid level for approximations and adaptive computation of the matching parameters. The modified MPGC (Multiphoto Geometrically Constrained Matching) algorithm (Gruen, 1985; Baltasvias, 1991) is employed to achieve sub-pixel accuracy for all points matched (if possible in more than two images) and identify some inaccurate and possibly false matches. Finally, a raster DSM can be interpolated from the original matching results.

Sensor Model	GCP	CP	x-RMS [m]	y-RMS [m]	z-RMS [m]	max. Δx [m]	max. Δy [m]	max. Δz [m]
rpc1	24	-	0.44	0.46	1.06	-1.11	-0.89	2.08
rpc2	24	-	0.39	0.42	0.68	-0.95	-0.84	-1.40
3daff	24	-	2.37	1.07	0.86	-4.87	2.05	1.57
rpc2	20	4	0.40	0.42	0.68	-1.01	-0.93	-1.41
rpc2	12	12	0.41	0.46	0.72	0.90	-0.92	-1.44
rpc2	5	19	0.51	0.43	0.90	-1.37	-0.78	-1.40

Table 7. Comparison of sensor models and number of GCPs in the IKONOS triplet (Thun). CP are the check points.

Sensor Model	GCP	CP	x-RMS [m]	y-RMS [m]	y-RMS [m]	max. Δx [m]	max. Δy [m]	max. Δy [m]
rpc1	5	19	0.45	0.46	1.10	-1.07	-0.99	2.30
rpc2	5	19	0.67	1.70	3.45	1.18	-3.04	6.24
rpc1	5	19	0.50	0.47	1.63	-1.33	0.89	2.93
rpc2	5	19	0.82	0.97	1.75	-1.51	2.02	3.17
rpc1	5	19	0.45	0.46	1.25	-1.05	-0.96	2.74
rpc2	5	19	0.53	0.59	1.50	-1.03	-1.52	3.15
rpc1	5	19	0.49	0.46	1.65	1.06	-1.05	3.35
rpc2	5	19	0.47	0.86	0.92	-0.95	1.95	1.94
rpc1	5	19	0.45	0.46	1.10	-1.06	-1.16	4.11
rpc2	5	19	0.41	0.70	1.05	-1.18	-1.19	-2.33

Table 8. Different distribution of GCPs in the IKONOS triplet Thun. CP are the check points. In the upper table part the GCPs cover 1/3 of the image in south-west, south-east, north-east and north-west, respectively (the most mountainous part is south-west, and then north-east). In the bottom table part, GCPs cover 2/3 of the image.

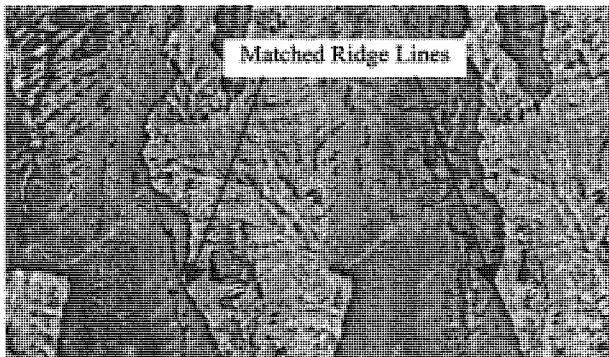


Figure 5. Edge matching results in Alpine area.

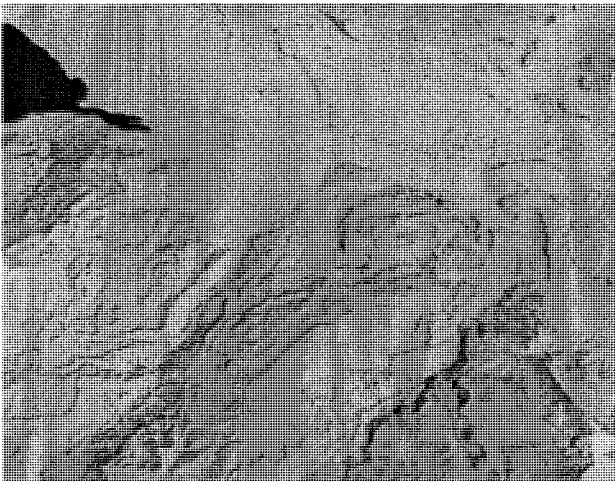


Figure 6. Shaded 5m DSM generated from IKONOS. The city of Thun in the upper left seen from North-West.

The procedure mainly contains the following characteristics:

- 1) It is a combination of feature point, edge and grid point matching. The grid point matching procedure uses relaxation-based relational matching, and can bridge-over areas with no or little texture through local smoothness constraints. The matched edges are introduced to control the smoothness constraints in order to preserve the surface discontinuities.
- 2) The adaptive determination of the matching parameters results in a higher success rate and less blunders. These parameters include the size of the matching window, the search distance and the threshold value for cross-correlation and MPGC. For instance, the procedure uses a smaller matching window, larger search distance and a smaller threshold value in rough terrain area and vice versa. The roughness of the terrain can be computed from the approximate DSM on a higher level of the image pyramid.
- 3) Linear features are important for preserving the surface discontinuities. A robust edge matching algorithm, using the multi-image information and adaptive matching window determination through the analysis of the image content and local smoothness constraints along the edges, is combined into our procedure. One example of edge matching is shown in Fig. 5.
- 4) Edges (in 3D) are introduced as breaklines when a TIN-based DSM is constructed. This DSM provides good approximations for matching in the next pyramid level. The computation of the approximate DSM in the highest pyramid level uses a matching algorithm based on the

“region-growing” strategy (Otto and Chau, 1988), in which the already measured GCPs and tie points can be used as “seed points”.

- 5) If more than two images are available, the MPGC procedure can use them simultaneously and matching results are more robust. Here, the resulting DSM from an image pair can be used as approximation for the MPGC procedure.
- 6) Through the quality control procedure, e.g. using the local smoothness and consistency analysis of the intermediate DSM at each image pyramid, the analysis of the differences between the intermediate DSMs, and the analysis of the MPGC results, blunders can be detected and deleted.

For each matched feature, a reliability indicator is assigned based on the analysis of the matching results from cross-correlation and MPGC. This indicator is used for assigning different weights for each measurement, which are used when a regular grid is interpolated.

4.2 Test Results

For Thun, we used for initial matching the images (and the respective triangulation results) of the triplet and stereopair separately and for the final MPGC all 5 images. The patch size varied from 7^2 to 17^2 for initial matching and was 11^2 for MPGC. Some areas like lakes and rivers were manually defined as “dead areas” via a user-friendly interface. A regular grid DSM with 5m spacing was interpolated from the raw measurements. Fig. 6 shows a visualisation of the generated DSM. In spite of smoothing due to the large area used in each point measurement, the discontinuities are quite well preserved.

Tables 9 and 10 show the DSM accuracy results, without any manual editing. The results are evaluated based on the differences between the heights interpolated in the reference laser DSM at the planimetric position of the DSM from matching and the heights from matching.

The tables show that the DSM accuracy is in the 1-5 m range, depending on the landcover and terrain type. A very high accuracy can be achieved in open areas. In these areas, more than 80% of the differences are less than 2 m. In urban and vegetation areas, the accuracy is worse, which is due to the fact that the reference LIDAR measurements and the parallaxes determined in matching refer to partly different objects. Matching measures higher than LIDAR at trees (in addition, at tress LIDAR sometimes measures below the tree tops) and narrow low-lying objects (like streets). Apart from that, the time difference between LIDAR and IKONOS data acquisition was 3-4 years, and the triplet of IKONOS had snow, up to 2-3 m in the mountains. Other factors that influenced matching were the long shadows (sun elevation was just 19 deg), occlusions, espec. in the W-E mountains, very low textured snow areas (which were improved with our preprocessing) and the patch size used in matching which unavoidably leads to smoothing of abrupt surface discontinuities. The accuracy values deteriorate also due to the high bias (see mean values espec. in Table 9), while height accuracy also gets worse due to the suboptimal base/height ratio (see sensor elevation and azimuth in Table 1). Taking all above factors into account, it becomes clear that IKONOS has a very high geometric accuracy potential and with sophisticated matching algorithms a height accuracy of 0.5 m – 1 m can be achieved in open areas

Area	No. of Compared Points	Mean (m)	RMS (m)	< 2.0 m	2.0-5.0 m	> 5.0 m	Max. (m)
O+C+V+A	29,210,494	-1.21	4.80	60.7%	16.8%	21.3%	424.2
O+C+A	17,610,588	-1.11	2.91	77.0%	13.9%	10.1%	358.9
O+A	14,891,390	-1.24	2.77	79.8%	12.2%	8.0%	358.9
O	11,795,795	-1.00	1.28	90.3%	8.5%	1.2%	37.33

Table 9. Accuracy measures and error classes for the triplet. O-Open areas; C-City areas; V-Tree areas; A-Alpine areas.

Area	No. of Compared Points	Mean (m)	RMS (m)	< 2.0 m	2.0-5.0 m	> 5.0 m	Max. (m)
O+C+V	20,336,024	0.45	4.78	57.7%	21.3%	20.9%	125.2
O+C	13,496,226	-0.33	3.38	68.7%	20.8%	10.3%	47.34
O	3,969,734	-0.97	1.54	83%	15.0%	2.0%	39.4

Table 10. Accuracy measures and error classes for the stereopair. O-Open areas; C-City areas; V-Tree areas.

with cooperative texture. In fact in these areas, the matching accuracy was close to that of LIDAR.

5. CONCLUSIONS

The presented results verify that 3D points can be determined with a submeter accuracy which for the planimetry can be 0.5 m or less, if accurate GCPs are used. This was achieved also with non-GPS GCPs and in mountainous areas with not very well defined GCPs. The number of GCPs can be small, their accuracy being the main point. GCPs can cover only a portion of the image, although caution should be paid in areas with large height differences. QB is not as linear as IKONOS and to achieve equal accuracy, needs an affine transformation after employment of RPCs. The simple models (3D and 2D affine) do not always perform well, thus use of RPCs should be generally preferred. IKONOS and QB orthoimages have been generated with an accuracy of 0.5 – 0.8 m, for typical sensor elevations of 65-75 deg. This requires, however, an accurate DSM/DTM. Sophisticated matching algorithms have derived a 5 m DSM with an accuracy of 1-5 m without editing and under very difficult conditions. In spite of that, accuracy in open textured areas was 1m or below. This potential has been very little exploited up to now, especially with IKONOS which is more stereo capable than QB, and presents an interesting alternative technology for deriving DSMs. Future work will focus on refinement of these investigations and possibly processing of new better quality images in the Thun testfield.

ACKNOWLEDGEMENTS

We thank the Swiss Federal Office of Topography and the NPOC, Bern for providing in Thun the laser DSM and in Geneva the HRS images, Swisssimage and the DHM25, the Canton Geneva for providing the 25-cm orthoimages and the laser DTM and Space Imaging USA for the IKONOS images in Thun and the RPCs of the Geneva IKONOS images. We also thank Oliver Heller and Oliver Gut, students at ETH Zurich, for contributing in this work, especially for the Geneva data.

REFERENCES

Baltsavias, E. P., 1991. Multiphoto Geometrically Constrained Matching. Ph.D. thesis, IGP, ETH Zürich, Mitteilungen No. 49, 221 p.

- Baltsavias, E., Pateraki, M., Zhang, L., 2001. Radiometric and Geometric Evaluation of IKONOS Geo Images and their use for 3D building modelling. Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", Hannover, Germany, 19-21 September (on CD-ROM).
- Fraser, C., Yamakawa, T., 2004. Insights into the affine model for high-resolution satellite sensor orientation. ISPRS Journal of Photogrammetry and Remote Sensing 58(5-6) (in print).
- Fraser, C.S, Baltsavias, E., Grün, A., 2002. Processing of IKONOS images for submetre 3D positioning and building extraction. ISPRS Journal of Photogrammetry & Remote Sensing 56 (3), 177 – 194.
- Grodecki, J., Dial, G., 2003. Block Adjustment of High-Resolution Satellite Images Described by Rational Functions. Photogrammetric Engineering and Remote Sensing 69(1), 59-70.
- Gruen, A., 1985, Adaptive Least Squares Correlation: A powerful Image Matching Technique. South Africa Journal of Photogrammetry, Remote Sensing and Cartography 14 (3), 175-187.
- Heller, O., Gut, O., 2004. Auswertung hochauflösender Satellitenbildern IKONOS/QUICKBIRD. Report for Vertiefungsblock, ETH Zürich, March.
- Otto, G. P., Chau, T. K. W., 1988. A "Region-Growing" Algorithm for Matching of Terrain Images. Proc. 4th Alvey Vision Club, University of Manchester, UK, 31 Aug. – 2 Sept.
- Zhang, L., Gruen, A., 2003. Automatic DSM Generation from StarImager (SI) Data. Proc. 6th Conference on Optical 3-D Measurement Techniques, Gruen, A., Kahmen, H. (Eds.), September 22-25, Zurich, Vol. I, pp. 93-105.
- Zhang, L., Gruen, A., 2004. Automatic DSM Generation from Linear Array Imagery Data. IAPRS, Vol. XXXV, Part B3 (proc. of this Congress).

3D MODELING AND VISUALIZATION OF LARGE CULTURAL HERITAGE SITES AT VERY HIGH RESOLUTION: THE BAMIYAN VALLEY AND ITS STANDING BUDDHAS

Armin Gruen, Fabio Remondino, Li Zhang
Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: <agruen><fabio><zhangl>@geod.baug.ethz.ch

Commission V - WG 6

KEY WORDS: Cultural Heritage, SPOT, IKONOS, Reconstruction, Modeling, Texture, Visualization

ABSTRACT

In this paper we present the modeling and visualization of the cultural heritage area of Bamiyan, Afghanistan. The region is situated ca 200 km north-west of Kabul and it was one of the major Buddhist centres until the ninth century AD. The two standing Buddhas belonged to some of the most famous Buddhist monuments world-wide. In 2001 they were destroyed through an act of vandalism by the Taleban militia. In our previous reports we have already presented the 3D computer reconstruction of the Great Buddha while in this contribution we will describe the terrain modeling of the surrounding area from satellite images, the 3D reconstruction of the entire rock cliff of Bamiyan and the modeling of the two empty niches where the Buddha statues once stood.

1. INTRODUCTION

The region of Bamiyan, ca 200 km North-West of Kabul, Afghanistan, was one of the major Buddhist centres from the second century AD up to the time when Islam entered the area in the ninth century. For centuries, Bamiyan lay in the heart of the famous Silk Road, offering rest to caravans carrying goods across the area between China and Western Empires. Strategically situated in a central location for travellers from North to South and East to West, Bamiyan was a common meeting place for many ancient cultures. In the region, many Buddha statues and a hundreds of caves were carved out of the sedimentary rock. In particular, near the village of Bamiyan, at 2600 meters altitude, there were three big statues of Buddha carved out of a vertical cliff (Figure 1). The famous standing Buddha statues, as well as other small statues in Foladi and Kakrak have been destroyed by the Taleban militia in March 2001 (Figure 2, left).

In 2003, the World Heritage Committee has decided to include the cultural landscape and archaeological remains of the Bamiyan valley in the UNESCO World Heritage List [<http://whc.unesco.org/>]. The whole area is in a fragile state of conservation as it has suffered from abandonment, military actions and explosions. The major dangers are the risk of imminent collapse of the Buddha niches with the remaining fragments of the statues, further deterioration of still existing mural paintings in the caves, looting and illicit excavation.

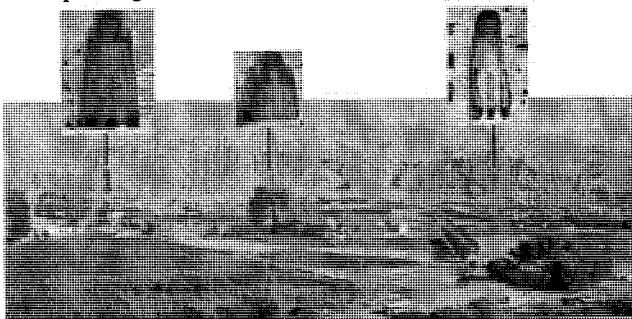


Figure 1: Panorama of the Bamiyan cliff with the three Buddha statues prior to demolition.

In our previous reports we have already presented the 3D computer reconstruction of the Great Buddha of Bamiyan

(Figure 2, right, Gruen et al., 2002, 2003). In this article we describe the 3D modeling procedures for (1) the terrain of the whole Bamiyan area, (2) the entire rock cliff of Bamiyan and (3) the two empty niches where the Buddha statues once stood. For the modeling of the cliff and the niches, images and geodetic measurements acquired during a field campaign in August 2003 were used.

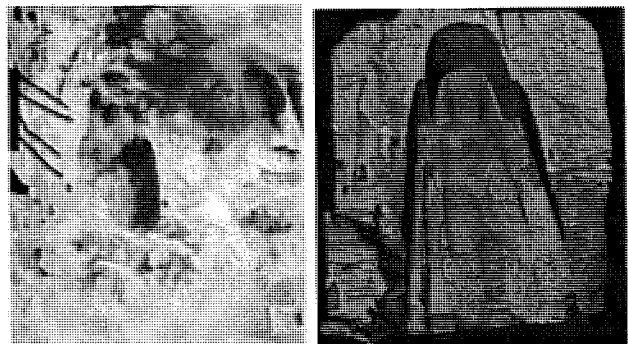


Figure 2: The explosion of March 2001 that destroyed the Great Buddha statue in Bamiyan (left). The 3D computer model of the figure, as reconstructed from existing images (right, Gruen et al., 2002, 2003).

The final goal is the generation of virtual flights over the UNESCO cultural heritage site. This includes a comparison of the previous situation with our high-resolution 3D models and the current situation with the empty niches.

A photo-realistic 3D digital model of the entire Bamiyan area and a detailed view of the cliff with and without the standing Buddhas will be presented.

2. TERRAIN MODELING FROM SATELLITE IMAGERY

For the 3D modeling and visualization of the area of interest, an accurate DTM is required. We had the contours of Russian map 1:50 000 digitised, but when we tried to map an IKONOS image onto the derived DTM we realised that its quality was not sufficient. Aerial images were not available to us and the idea to acquire them was unrealistic, due to the absence of any surveying company operating in that area. So space-based

image acquisition and processing resulted as the only alternative to the aerial photos or any other surveying method. Nowadays space images are competing successfully with traditional aerial photos, for the purpose of DTM generation or terrain study in such problematic countries, as Afghanistan is. Also the resolution and availability of world-wide scenes taken from satellite platforms are constantly increasing. Those scenes are available in different radiometric modes (panchromatic, multispectral) and also in stereo mode.

2.1 The satellite images

The HRG sensor carried on SPOT-5 since May 2002 was suitable for our tasks [Spot Image]. The sensor can acquire stereo images in across-track direction at 2.5m ground resolution in panchromatic mode. The time difference between two successive acquisitions in stereo mode of the same area depends on the incidence angle, with a minimum of one day difference. The satellite flies at a mean height of 832 km, along a quasi-polar and sun-synchronous orbit.

Due to the scientific, social and cultural interest of the project, a B/W stereo pair over Bamiyan was provided at special conditions by the ISIS Program [ISIS]. The two scenes have been acquired on 18th December and 19th December 2003, at 10:40 a.m. and 10:20 a.m. local time respectively. The scenes are 24000 x 24000 pixels large and cover a mountainous area of approximately 60x60 km, centered at (34°50' N, 68° 7' E) and (34°50' N, 67° 48' E). The cloud cover was zero and the ground resolution is 2.5 m.

Furthermore, a B/W Geo level IKONOS image mosaic over the Bamiyan area was provided by Space Imaging [Space Imaging] (Figure 3). The scene was acquired on 15th December 2001 and covers an area of 11.6x18.5 km, centered at (34°46' N, 67°49' E). The image size is 13957 x 21118 pixels and the ground resolution is 1m.

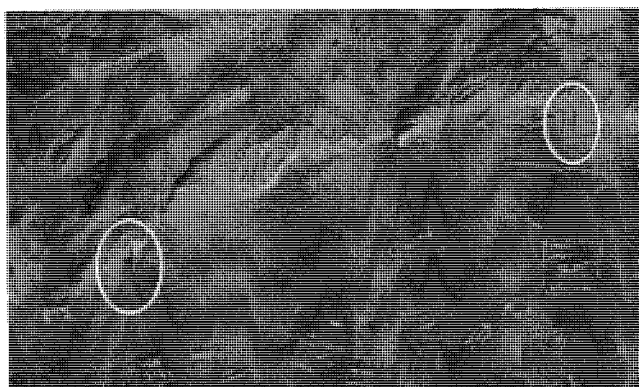


Figure 3: The two empty Buddha niches as observed by IKONOS (courtesy of Space Imaging, Inc.). Left: Great Buddha. Right: Small Buddha.

2.2 GPS measurements

For the georeferencing of the satellite images, seven GPS points were measured (Figure 4) during our field campaign in August 2003. Two Trimble GEO Explorer receivers logging carrier phase data (C/A code) were used; one receiver was set as 'master' (fix position) and the other served as 'rover'. The points were identified on the IKONOS image and then measured in C/A mode.

After the data collection, the observations were post-processed to generate more accurate positions. It was not possible to perform absolute differential corrections as the closest master

station was ca 1500 km away. Table 1 summarises the GPS measurements.

Table 1: GPS measurements and related parameters.

No. of Points	Average satellites	Average PDOP	Horiz. Accuracy	Vert. Accuracy
7	6	2.78	0.11 m	0.3 m



Figure 4: The distribution of the GPS ground control points on the IKONOS image mosaic. In the centre the master station position.

2.3 Scenes orientation and DTM generation

The DTM was generated from level 1A scenes (radiometrically but not geometrically corrected) using SAT-PP (SATellite Precision Positioning) software developed at our Institute. SAT-PP includes different modules for the precise processing of high-resolution satellite image data and allows image pre-processing, orientation, matching, DTM/DSM generation and object extraction. For more details, see [Zhang, Gruen, 2004; Poli et al., 2004]. The IKONOS mosaic orientation was based on a 2D affine transformation. On the other hand, the SPOT scenes orientation was based on a rational function model. Using the camera model, the calibration data and the ephemeris contained in the metadata file, the software estimates the RPC (Rational Polynomial Coefficients) for each image and applies a block adjustment in order to remove systematic errors in the sensor external and internal orientation. The scenes' orientation was performed with the help of the GCPs measured with GPS. Table 2 shows the image orientation results.

Table 2: Image orientation results.

Source	RMSE East (m)	RMSE North (m)	RMSE Height (m)
IKONOS	0.56	0.48	-
SPOT Image Pair	1.22	2.01	1.50

The oriented SPOT stereo pair was then subjected to the automated DTM/DSM generation, using the module of SAT-PP. A 20 m raster DTM for the whole area and 5 m raster DTM for the area covered by the IKONOS image were interpolated from the original matching results, using also some manually measured breaklines near the Buddha cliff (Figure 5). The matching algorithm combines the matching results of feature points, grid points and edges. It is a modified version of MPGC (Multi Photo Geometrically Constrained) matching algorithm [Gruen, 1985; Zhang, Gruen, 2004] and can achieve sub-pixel accuracy for all the matched features.

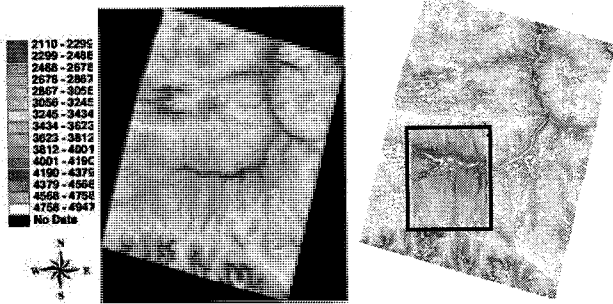


Figure 5: The recovered DTM of the Bamiyan area displayed in colour coding mode (left). The two overlaid DTMs (right).

Finally, for the visualization of the whole Bamiyan area, a 2.5 m resolution ortho-image from SPOT images and a 1 m resolution ortho-image from the IKONOS image were generated. The textured 3D models are presented in Figure 6.

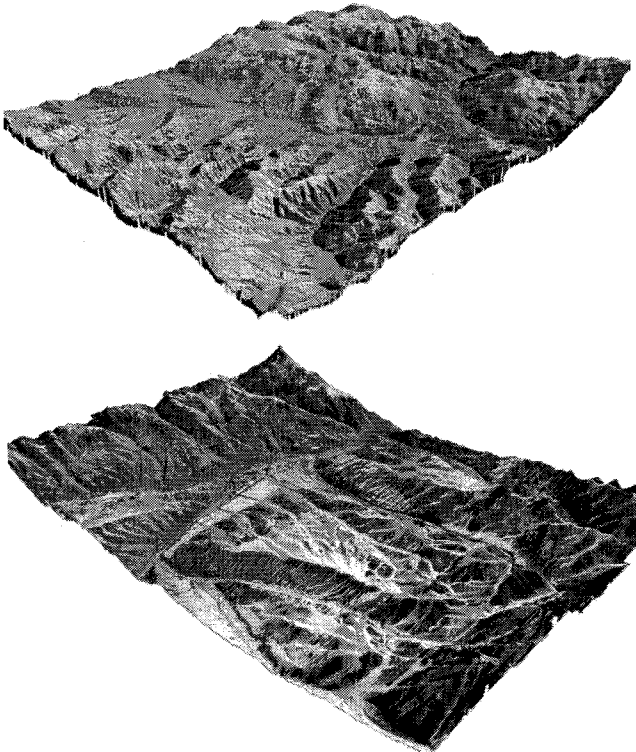


Figure 6: The DTM of the Bamiyan area generated from satellite images and textured with a SPOT5 (above) and an IKONOS ortho-image (below).

In Figure 7 two closer views on the 3D IKONOS textured model of the Bamiyan cliff and the old Bamiyan city (the pyramid-type hill to the left) are presented.

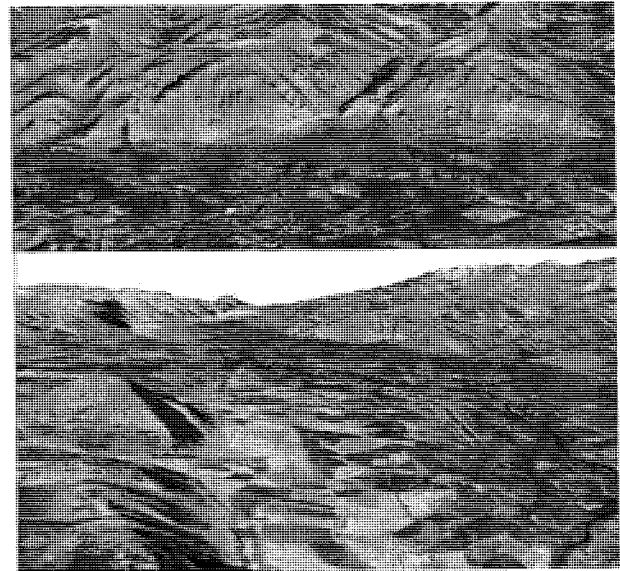


Figure 7: Zoom into the 3D Bamiyan terrain model. The rock cliff with the two empty niches (above). A view of the Bamiyan area: on the left the hill (Shar-i Ghulghulah) where the old city was located while in the centre the new 'bazaar' is visible (below).

3. MODELING OF THE ROCK CLIFF OF BAMBIYAN

For the reconstruction and modeling of the Bamiyan cliff (Figure 8), a series of terrestrial images acquired with an analogue Rollei 6006 camera was used. Furthermore, a geodetic network of 10 stations was used to measure, with a total station, 30 control points distributed all along the rock cliff.

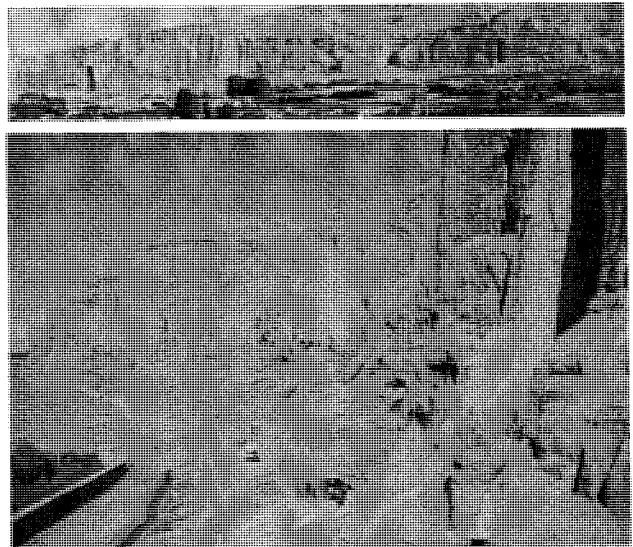


Figure 8: The Bamiyan cliff, approximately 1 km long and 100 m high (above). The cliff as seen from the empty niche of the Small Buddha (below).

The images of the rock facade were acquired along a strip. 39 images were oriented with a self-calibrating bundle adjustment, measuring the tie points in the Analytical Plotter (Figure 9). The average precision of the object points is $\sigma_x = 0.15$ m, $\sigma_y = 0.11$ m, $\sigma_z = 0.23$ m with a standard deviation of unit weight a posteriori at 0.013 mm.

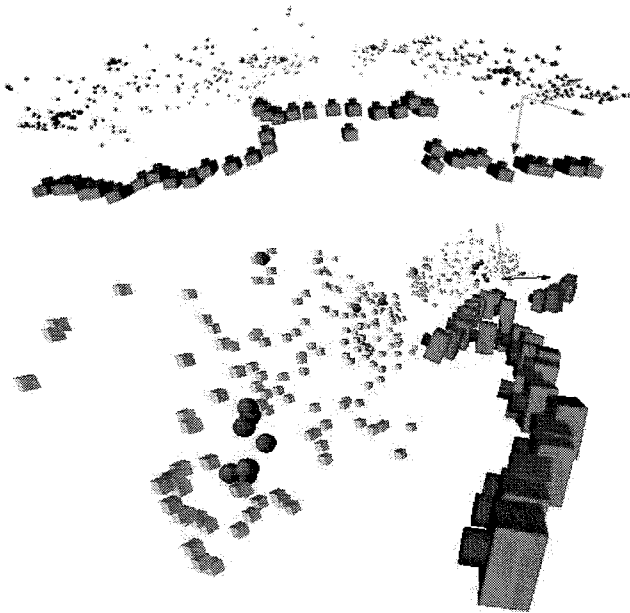


Figure 9: The recovered camera poses of the Rollei images used for the modeling of the rock cliff. The spheres represent the used control points. Top: a view from the top of the strip. Bottom: a view from the side of the empty niche of the Great Buddha.

Because of some software problems with the Analytical Plotter we could not perform stereo measurements with the instrument. Therefore the images were scanned at 20 micron resolution and the modeling of the cliff surface was done by measuring the points with the commercial software Photomodeler. Then a mesh with a 2.5D Delaunay triangulation was generated and a textured model of the cliff was produced (Figure 10).

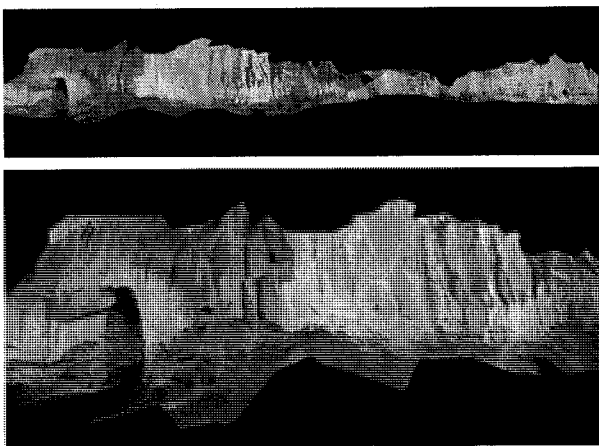


Figure 10: Two different views onto the 3D textured model of the Bamiyan rock cliff.

Because of the network configuration and the complex shape of the rock facade, the recovered geometric model is not complete, in particular in the upper part. In some areas it is not possible to find corresponding features, because of occlusions, different lighting conditions and shadows. This is not such a big problem, because the cliff model is not meant to be used alone, but in a next step it will be integrated into the DTM of the larger environment.

4. MODELING OF THE TWO EMPTY NICHES

The modeling of the two empty Buddha niches was performed using digital images acquired with a Sony Cybershot F707. The image size is 1920x2560 pixels while the pixel size is ca 3.4 μm .

Both niches are now a “national monument with unique importance to humankind” and are safeguarded by UNESCO, the Japanese Government and the Afghan Ministry of Information and Culture (Figure 11).

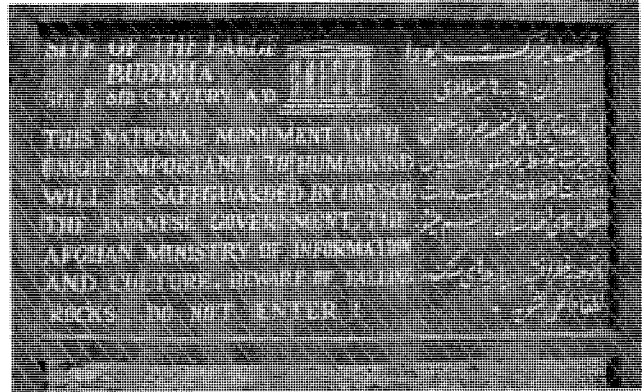


Figure 11: The sign in front of the empty cave of the Great Buddha that declares it a protected national monument.

4.1 The empty niche of the Great Buddha

For the 3D computer reconstruction 5 images were used (Figure 12). The camera parameters were recovered with a self-calibrating bundle adjustment, measuring the tie points semi-automatically by means of Least Squares Matching (LSM) [Gruen, 1985]. The final average standard deviations of the object coordinates are $\sigma_x = 0.014$ m, $\sigma_y = 0.017$ m, $\sigma_z = 0.021$ m.



Figure 12: The empty niche where the Big Buddha once stood, as seen in August 2003.

Afterwards, distortion-free images were generated and imported into the VirtuoZo stereo digitize module [VirtuoZo NT, 1999]. Three stereo-models were set up and points were measured along horizontal profiles, while the main edges were measured as breaklines. Thus a point cloud of ca 12 000 points was generated. The recovered camera poses and the measured points are displayed in Figure 13.

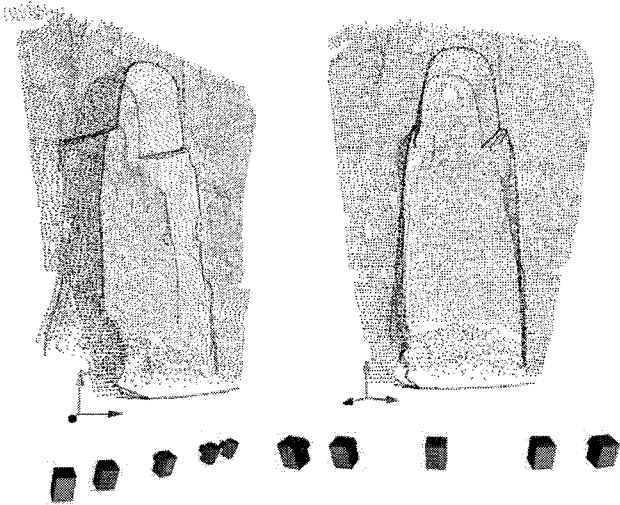


Figure 13: The recovered positions of the cameras and the measured points for the modeling of the empty niche of the Great Buddha.

The surface generation was then performed with a 2.5 Delaunay method, dividing the measured point cloud into separate parts. A mesh was created for each single point cloud and then all the surfaces were merged together with Geomagic Studio [Raindrop]. The final model of the empty niche is displayed in Figure 14.

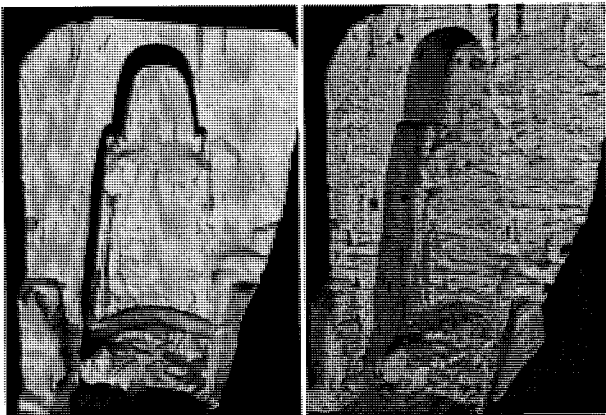


Figure 14: 3D model of the empty niche of the Great Buddha, visualized in shaded (left) and textured mode (right).

The recovered 3D model of the empty niche allows the comparison between the actual situation and the previous one, before the destruction (Figure 15 and 16).

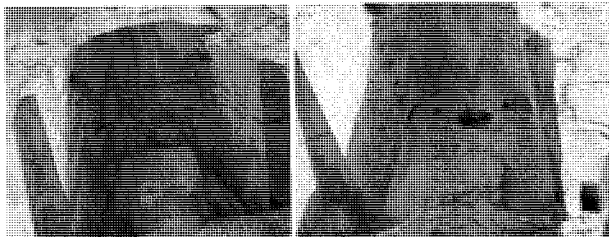


Figure 15: The niche of the Great Buddha: a comparison between an image of the 1970's and the actual situation.

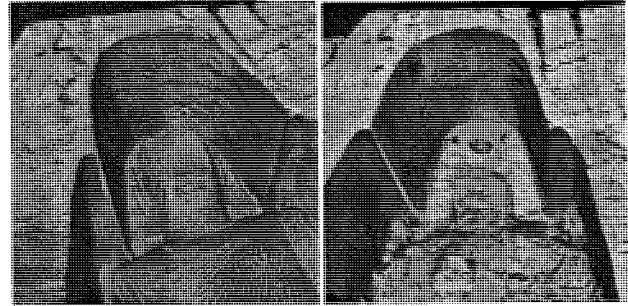


Figure 16: The reconstructed 3D models of the Great Buddha and the actual empty niche. An image of the full 3D model of the Great Buddha is shown in Figure 2.

4.2 The empty niche of the Small Buddha

The modeling of the empty niche of the Small Buddha was performed using 9 images (Figure 17)

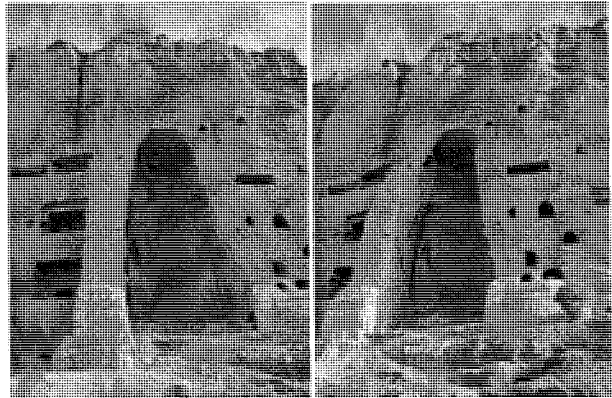


Figure 17: The empty niche of the Small Buddha.

The necessary tie points were measured semi-automatically with LSM and then imported in a self-calibrating bundle adjustment. The final average standard deviations of the object coordinates are $\sigma_x = 0.015$ m, $\sigma_y = 0.019$ m, $\sigma_z = 0.022$ m.

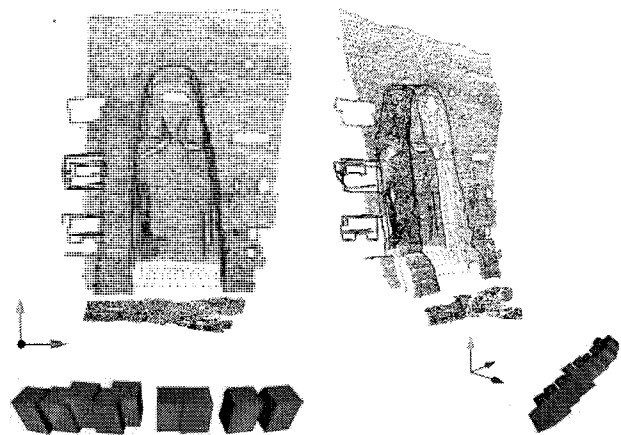


Figure 18: The camera poses and the manually measured points for the modeling of the empty niche of the Small Buddha.

Afterwards manual measurements were performed in Virtuoso on the distortion-free images and a cloud of approximately 17 000 points was recovered (Figure 18).

For the mesh generation, we had to split the point cloud in different parts, to be able to perform 2.5 Delaunay triangulation.

After the surface generation, a textured model was produced, as shown in Figure 19.

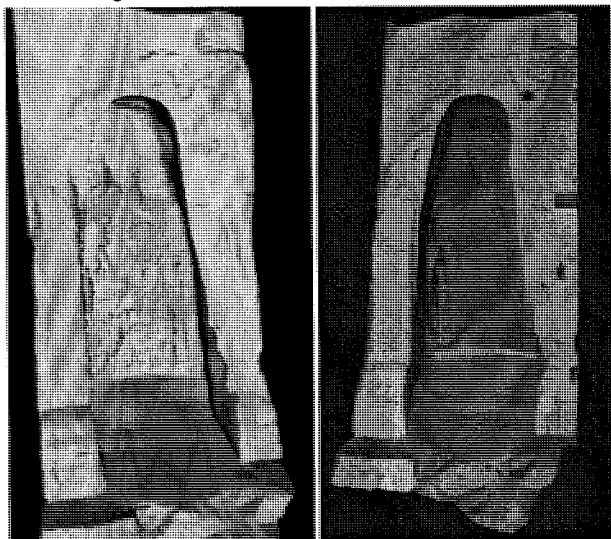


Figure 19: 3D model of the empty niche of the Small Buddha, visualized in shaded (left) and textured mode (centre and right).

5. CONCLUSIONS

In this work we have shown the modeling of the whole cultural heritage site of Bamiyan, Afghanistan. We have used different types of images and produced a detailed terrain model as well as 3D models of the two empty niches and the rock cliff where the two Buddha statues once stood. The 3D reconstruction of the Small Buddha statue is currently underway.

The satellite images were the only possibility for the mapping of the heritage area because of their instant availability and high resolution.

Manual measurements were necessary to reconstruct all those small features that an automatic procedure would miss. The big operational problem was actually the 3D surface modeling of the cliff and the Buddhas. We are still not satisfied with commercial modeling software and we spent more time on modeling than on measurement. In order to have a smooth production process one should adapt the measurement procedure to the capabilities of the 3D modeller.

The photo-realistic 3D digital models of the entire Bamiyan area and the detailed view of the cliff with and without the standing Buddhas will be used for visualization and animation.

We also can now compute the volume of the material that was destroyed during the demolition. This will give clear hints as to whether the leftovers will suffice to be used for physical reconstruction. Results from these considerations will help to decide about the best reconstruction.

In addition, the DTM will now be imported into a GIS software to create a complete information system of the protected area, which may serve for technical, scientific and touristic purposes in the future.

ACKNOWLEDGEMENT

The authors would like to thank Daniela Poli for her help in the satellite image acquisition, Spot Image for providing the HRS images at special conditions and Space Imaging for providing a IKONOS scene for free. We also appreciate the contributions of Natalia Vassilieva, Yuguang Li and Xiaoyun Fu in terms of doing photogrammetric measurements.

REFERENCES

Gruen, A., 1985: Adaptive Least Squares Correlation: A powerful Image Matching Technique. *South Africa Journal of Photogrammetry, Remote Sensing and Cartography*, 14 (3), pp. 175-187.

Gruen, A. Remondino, F. Zhang, L., 2002: Reconstruction of the Great Buddha of Bamiyan, Afghanistan. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIV, part 5, pp. 363-368, Corfu, Greece.

Gruen, A. Remondino, F. Zhang, L., 2003: Image-based Automatic Reconstruction of the Great Buddha of Bamiyan, Afghanistan. *Computer Vision and Pattern Recognition (CVPR) Workshop on 'Application of Computer Vision in Archaeology'*, Madison, USA (on CD-ROM).

ISIS: http://medias.obs-mip.fr/isis/?choix_lang=English, April, 2004.

Poli, D., Zhang, L., Gruen, A., 2004: SPOT-5/HRS Stereo Images Orientation and Automated DSM Generation. *Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXV-B1, in press.

Raindrop: <http://www.geomagic.com>, April, 2004.

Space Imaging, Inc: <http://www.spaceimaging.com>, April, 2004.

Spot Image, <http://www.spotimage.fr>, April 2004.

VirtuoZo NT, 1999: Version 3.1 Manual, Supresoft Inc.

Zhang, L., Gruen, A., 2004: DSM Generation from Linear Array Imagery. *Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXV-B3, in press.

IMAGE ANALYSIS FOR THE HISTORY OF CARTOGRAPHY: DRAWING CONCLUSIONS FROM THE EVALUATION OF PFYFFER'S RELIEF

Jana Niederöst

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology Zurich,
ETH-Hönggerberg, CH-8093 Zurich, Switzerland, jana@geod.baug.ethz.ch

TS-PS WG V/4 Image Analysis and Spatial Information Systems for Application in Cultural Heritage

KEYWORDS: cultural heritage, history of cartography, digital recording, accuracy analysis, surveying reconstruction

ABSTRACT:

Digital photogrammetry offers effective and accurate procedures for the preservation and inventory of cultural heritage. In case of two- and three-dimensional products of early mapmaking the application of these methods goes far beyond documentation and visualization: it primarily aims at the derivation of new information. This paper reports on conclusions which can be drawn from the quantitative evaluation of one of the most spectacular relief models in history, the relief of Franz Ludwig Pfyffer constructed between 1750 and 1786. At first, the techniques and results of image-based reconstruction and accuracy analysis of this masterpiece are briefly summarized. Due to the lack of written documentation concerning the relief and its constructor, the virtual Pfyffer's Relief and the products of image processing represent a great potential for research in the history of cartography. Thus, the main focus of the paper is the reconstruction of Pfyffer's surveying procedures and the interpretation of the reached accuracy in context of European mapmaking of that period. The work leads to a new knowledge about the cartographic, spatiotemporal but also social and political aspects of the relief creation. As opposed to the previous assumptions it can be concluded that Pfyffer performed systematic and repetitive triangulation and height measurements. His procedures, instruments and results correspond with the scientific status quo in the last quarter of the 18th century, whereas his area-wide height survey was a pioneering achievement at that time.

1. INTRODUCTION

Old maps and three-dimensional relief models are an important part of our cultural heritage. They bear witness of the technical development, social, cultural and politic circumstances as well as of the comprehension of space and landscape at the time they had been created. As such they are precious material, pieces of art that help us to understand our history, territorial dynamics and progress of scientific knowledge. The digital recording of two- and three-dimensional historical cartography enables its preservation for the next generations; however, primarily it should serve for the derivation of new information. Due to the lack of primary written sources this "added value" - gained through the recovering of the metrical content of old topographic works by means of statistical and image processing procedures - is very worthwhile for the historical research. This paper shows how the results of accuracy analyses and visualization can be interpreted and hence, which new conclusions can be drawn concerning the early mapmaking at that period.

The object of the research is one of the most spectacular relief models in history, a large bird's eye view of mountainous Central Switzerland constructed by the lieutenant general Franz Ludwig Pfyffer (1716-1802) between 1750 and 1786. The topographic base of this 6.6 x 3.9 m² big masterpiece at a scale of about 1:11'500 are the own measurements of Pfyffer, performed decades before the first Swiss national triangulation was established. The visual verification shows that the relief was a significant improvement of existing maps at that time (Fig. 1) and it also served as a basis for several printed works issued at the end of 18th century. However, because of the very sparse written documentation and the relief complexity, neither its formation nor its geometric characteristics have been investigated up till now. Within the scope of an interdisciplinary project supported by the Swiss National Science Foundation and the local cultural authorities, the methods for the quantitative evaluation of old maps and relief

models are to be developed. The application of these procedures to the Pfyffer's Relief (or more precisely, to its virtual computer model) and related maps of the region should answer the questions of historians concerning the creation, accuracy and historical context of the relief. The indirect analysis of Pfyffer's topographic work as the only possible way of exploring this chapter of Swiss cartography should complete the following three tasks:

- 3D reconstruction of the relief for the documentation of the cultural heritage and for the numerical evaluation of geometric features of the original,
- Accuracy analysis of the relief and related old maps; interpretation of the results in the context of European mapmaking at that period,
- Investigation of Pfyffer's surveying and model constructing procedures; comparison with contemporary and present methods.

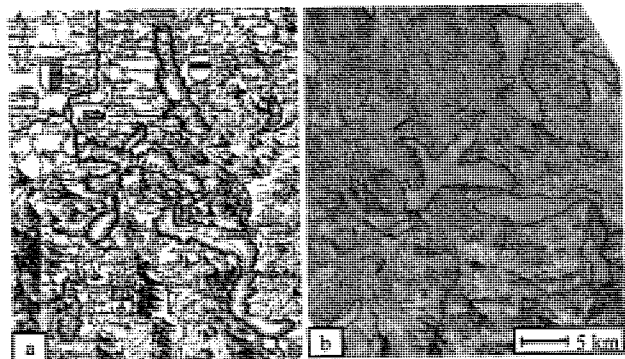


Fig. 1: The depiction of Central Switzerland (a) in the principal map of the region over the whole 18th century (*Nova Helvetiae Tabula Geographica*, J. J. Scheuchzer, 1712), (b) in the georeferenced Pfyffer's Relief overlaid with the current lake contours (blue lines).

In a previous publication (Niederöst, 2003) the methods for the digital recording of Pfyffer's Relief and the quantitative evaluation of the virtual relief and related old maps were presented (they are shortly summarized in Section 2). The analysis yields a number of statistic and visual products, which have a great potential for the research in the history of cartography. Now the most important question is to be answered: Which conclusions can be drawn from the results of the accuracy analysis? This paper aims at the reconstruction of Pfyffer's surveying procedures (Section 3) as well as at the interpretation of the estimated relief accuracy in the context of European cartography in the 18th century (Section 4). The research is based on several sources: a few autographic letters (Pfyffer, 1761) and maps of Pfyffer, archive publications as a testimony of the surveying and cartographic standards of that time and of course, the results of the performed quantitative analyses.

2. DATA PROCESSING

For the generation of a 3D computer model of Pfyffer's Relief, the procedures of image-based object reconstruction were applied. The photogrammetric stereo processing resulted in a digital surface model of 1 cm raster width, an orthoimage and a structured vector data set of significant relief features. The precise digital data set of Pfyffer's Relief including visualization products (Fig. 2) was archived at the Kulturgüterschutz of Lucerne for the documentation of cultural heritage.

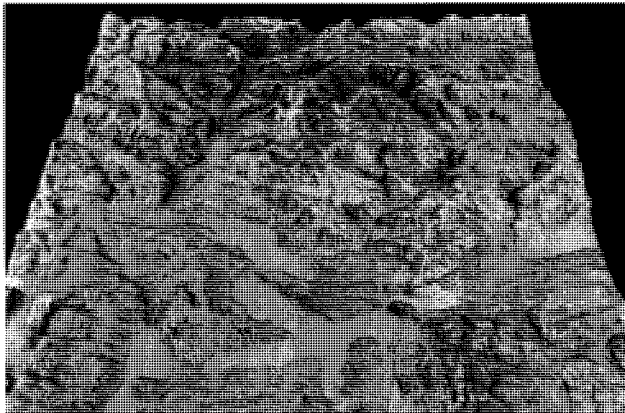


Fig. 2: A view of the photogrammetrically reconstructed texture-mapped Pfyffer Relief

The accuracy analysis of Pfyffer's Relief is based on a number of identical points in the historical and modern data set. Applying a 9-parameter spatial transformation, the accuracy of the relief as well as its metric parameters such as scales and rotation angles were determined. The calculated distortion grid (Fig. 3) allows for a visual presentation of the numeric results of the analysis. Furthermore a new procedure for georeferencing of the historical height, image and vector data was developed. The overlay of the georeferenced data set with the current map information (Fig. 1b) represented a basis for the numerical accuracy analysis of polygonal relief features such as lake forms. Finally, the surface of Pfyffer's model was compared with the national terrain model DHM25 of swisstopo.

In order to understand the spatial and temporal coherences of the relief creation, the following relief-related old maps were also included in the analysis (see also Table 1):

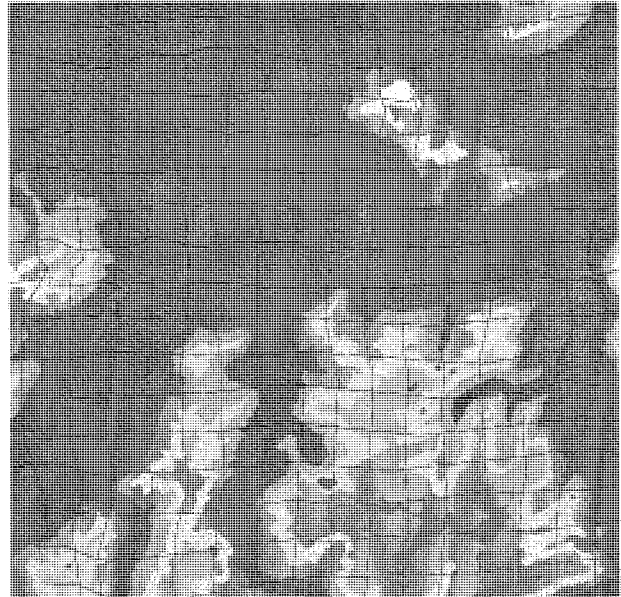


Fig. 3: The distortion grid of Pfyffer's Relief showing the deformation of 2-km coordinate lines on the background of the current national DTM (DHM25, copyright swisstopo).

- The coloured small-scaled drawing of Central Switzerland by Pfyffer's own hand (Fig. 4),
- The fragment of a large-scaled manuscript map of Pfyffer from the region of Rengg valley, which was - re-used as wasted paper - surprisingly found inside the relief during its restoration (Fig. 5),
- The engraved *Carte en Perspective du Nord au Midi*, created by Jakob Joseph Clausner (1744-1795) on the basis of Pfyffer's measurements in 1786 (Fig.6).

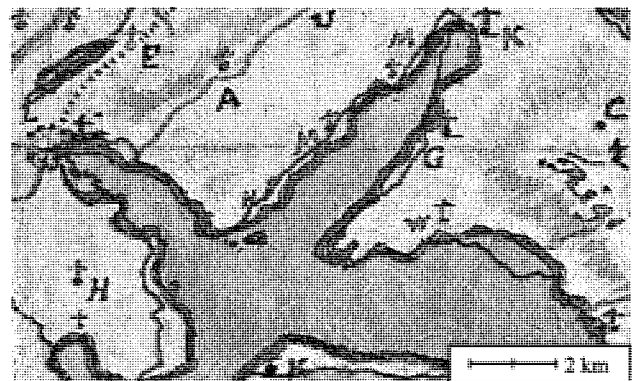


Fig. 4: A part of the georeferenced manuscript map of Central Switzerland of F. L. Pfyffer overlaid with current lake contours.

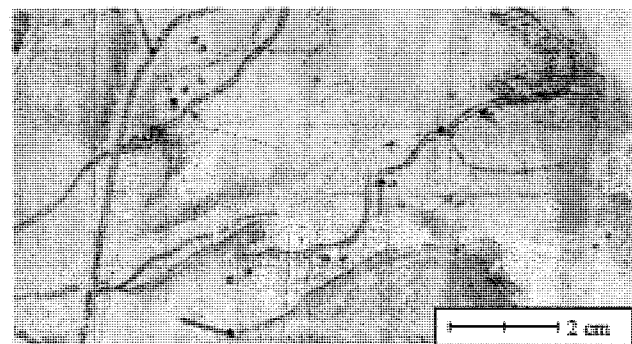


Fig. 5: A part of the large-scaled manuscript map of F. L. Pfyffer overlaid with residuals on identical points (red lines).

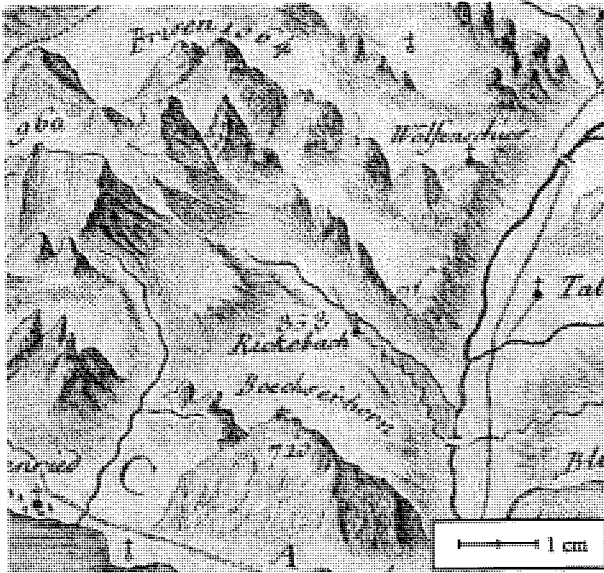


Fig. 6: A part of the map of Central Switzerland by J. J. Clausner created on the basis of Pfyffer's measurements. The map is overlaid with residuals on identical points (red lines).

	(1)	(2)	(3)	(4)
Pfyffer's Relief	208	1:11'700 1:11'400 1:11'200	30.28° south-oriented	470 m 494 m 111 m
Pfyffer's map of Central Sw.	115	1:244'700 1:249'200	-0.65° north-oriented	387 m 528 m
Pfyffer's map of Rengg area	14	1:36'000 1:33'300	15.29° south-oriented	191 m 117 m
Map of Clausner	178	1:125'200 1:123'200	15.62° south-oriented	408 m 483 m

- (1) Number of identical points
- (2) Scale in coordinate directions
- (3) Orientation (or westerly declination, see Fig. 9)
- (4) Sigma a posteriori in coordinate directions

Table 1: Results of coordinate transformations with weighted observations: metric parameters of Pfyffer's Relief and related maps.

3. RECONSTRUCTION OF PFYFFER'S SURVEYING AND RELIEF CONSTRUCTING PROCEDURES

F. L. Pfyffer was a passionate surveyor who spent years performing measurements in the alpine terrain, sometimes with a couple of assistants, but mostly on his own. Although he was a self-educated practitioner, he could build on experiences gained during his four decades' career in the French military service which gave him an opportunity to learn modern surveying methods, to get in touch with reputed scientists and instrument makers and to find access to the scientific literature. Back in Switzerland, he decided to present his mountainous homeland in the most natural and understandable form, in the form of a topographic relief. He was aware of the inaccuracy of existing maps of the region and around 1750 he met the challenge to newly map almost 3'500 km² at a large scale, including a novelty at that time, the height measurements all over the country.

3.1 Triangulation and detailed surveying

On the basis of Pfyffer's letters (Pfyffer, 1761) to the famous Swiss scientist Micheli Du Crest (1690-1766) it is known now

that Pfyffer defined a mesh of large triangles which he surveyed precisely by triangulation. Later on, he densified this primary network by detailed field measurements, using the superordinate triangulation points for forward intersection. The calculated relief accuracy (Table 1) and the more or less homogeneous relief distortions over the whole mapped area (Fig. 3) confirm that the triangulation with its basic principle "from superordinate to subordinate" must have been applied. On the other hand, the accuracy analysis supplies evidence that Pfyffer evaluated his measurements mostly graphically in combination with simple formulas of planar trigonometry instead of adopting more accurate, purely numerical methods under consideration of the curved Earth surface.

Being aware that the requisite for a precise triangulation is a baseline few kilometres long measured with great care, Pfyffer was particularly concerned with the appropriate methods for direct distance measurements. He regularly checked the length of his metal bars and chains (having respectable dimensions from 15 to 120 French feet corresponding to 4.9 and 39 m) along the side of a marmoreal quadrat which he considered to be perfectly shaped. After a long search in mountainous Central Switzerland, Pfyffer finally found a suitable, possibly flat area for the baseline measurements. Then he measured shorter distances in a zigzag manner as well as the corresponding angles and hence he determined the searched baseline length (Fig. 7). With a careful procedure Pfyffer tried to eliminate the impact of systematic errors. Thus, he always laid his chains or bars along a strained cord in order to keep the desired measurement direction and he considered the influence of sloped terrain by a successive horizontal arrangement of the instruments. During the survey for his relief Pfyffer measured several different baselines; only in summer 1761 he mentions about 6 of them. The reason for such excessive measurements was probably the quality control, which was for Pfyffer of great importance.

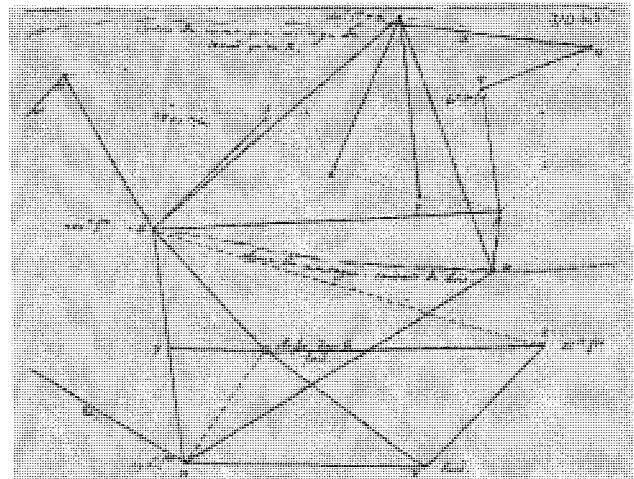


Fig.7: The triangulation sketch of F. L. Pfyffer, found in one of his letters to Micheli Du Crest. The long dotted zigzag line in the upper part was measured using metal chains. Out of these direct measurements Pfyffer determined the length of the baseline MN to 2832 French toises (around 5.5 km).

Before starting angle measurements for the triangulation, Pfyffer had made himself familiar with the territory to be surveyed. During the reconnaissance he checked the reciprocal lines of sight and assigned the future signals and stations. However, as he complains several times in his letters,

he was not allowed to erect permanent signal structures, because those were often removed on purpose. Thus, in order not to attract attention, he marked the places using small posts on the ground level or a piece of iron in the rocks. During the measurements his five or six helpers set a purpose-built vertical signal pole on the assigned location and guarded it until the *operation* - as he called surveying - was finished. For communication with his assistants Pfyffer used a big flag in black and white. From time to time he was also forced to use natural points such as single tall firs as signals; in this case he considered the displacement of the station during the subsequent angle measurements at the concerned position. We can conclude that in the time when surveying was not (yet) a state- or military-supported enterprise but rather a cause for suspicion, Pfyffer solved the problems of missing immovable signal structures by means of good preparation and organization of his work.

Pfyffer ordered his surveying equipment and accessories at various places in Europe. To prove the procedures and the accuracy of his instruments to measure angles, he measured at each station with several different pieces of equipment. As a practitioner he was at most fond of a plane table and a graphometer, both instruments which could deliver the graphical results immediately in the field. Pfyffer constructed some surveying instruments himself, too, such as the alidade with a telescope and two diopters in combination with a plane table (Fig. 8 right). The most precise instrument of Pfyffer was a simple English theodolite which he called *le chassis*, the frame (Fig. 8 left). It was a stable, 2 feet high device with a horizontal and a vertical circle, both divided in degrees and minutes. Instead of a telescope the instrument was provided with two peep sight diopters. The major difference of this "frame" in comparison with his other instruments was that it did not allow the underlying with a piece of paper for graphical interpretation. It belonged to a new generation of instruments coming from England, which supplied the surveyor only with the numerical values whereas the abandonment of drawing accessories made them more compact and more accurate.

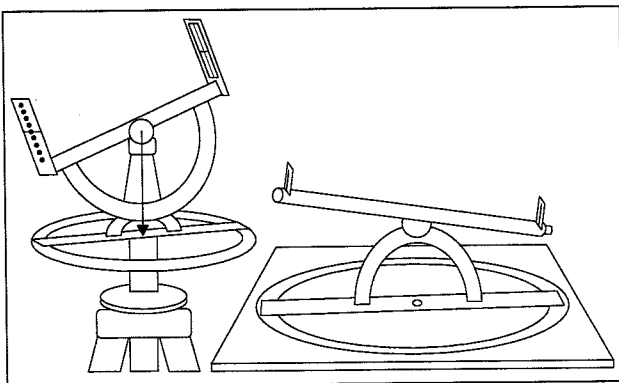


Fig. 8: Two instruments reconstructed according to Pfyffer's description: left an ancestor of a present theodolite, right the alidade with a telescope in combination with a planar table.

The results of numerical analysis using statistical and image processing procedures allowed for coming to a new interesting conclusion concerning Pfyffer's triangulation. As shown in Table 1, the deviation of the relief and the analysed maps from the northern direction amounts to peculiar numbers (around 30 and 15.5 degrees), while the orientation of Pfyffer's manuscript map of Central Switzerland at a scale of about 1:247'000 is almost perfectly northern. What motivated Pfyffer to orient his

major topographic works in such an unusual manner? At the end of the 18th century the northern orientation of small-scaled maps became widely accepted and on the other hand, large scale maps still kept an arbitrary orientation according to the position and form of the territory to be mapped. However, the latter could not be the reason for Pfyffer who himself depicted the same area in a north-oriented manuscript drawing (Fig. 4). The answer to this question can be found when looking at the secular variations of the geomagnetic field of the Earth for over 400 years. On the basis of the old measurements we can recognize a steady progressive change in magnetic declination, or angle between magnetic north and geographic north. The publication of (Fischer, 1994) shows that between 1740 and 1790 the westerly declination in Central Switzerland varied from 15° to 19° with a value of about 15.5° for 1750. Thus, Pfyffer oriented his large-scaled topographic works - including the inner constructional parts of the relief - simply according to his compass (Fig. 9). Since Pfyffer's measurements took place within a long time span and over a large area, changes of magnetic declination certainly had a negative influence on the accuracy of his results. To eliminate these errors, repeated and precise astronomic measurements would have to be undertaken, and this was by Pfyffer probably not the case.

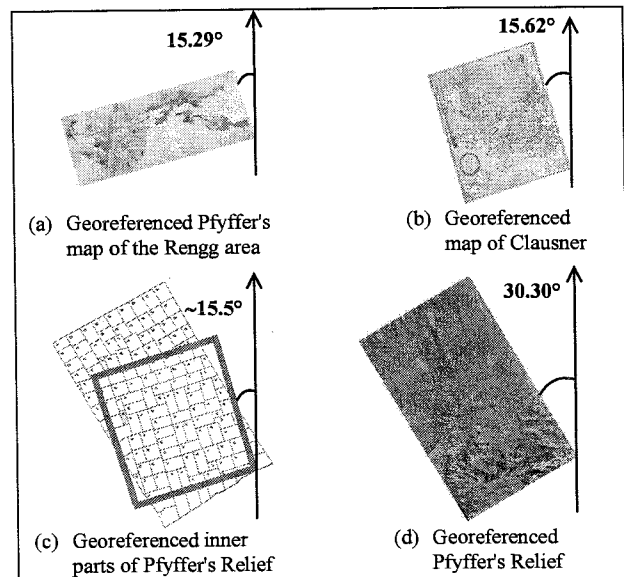


Fig. 9: Orientation of Pfyffer's works, corresponding with the westerly declination of magnetic north (a-c). The original declination of the relief was 15.5° (c). Later on, Pfyffer must have decided - probably for political reasons - to add new constructional parts under a different angle so that the whole relief deviates about 30° from geographic north (c, d).

Pfyffer's surveying was performed before the theory of errors and least squares adjustment had been elaborated. He strived for accuracy in terms of possibly small deviation of the angle sum in a triangle from 180°. His observations were always excessive and thus they enabled the elimination of insecure directions and also a simple averaging of error-afflicted measurements. When content with the triangulation, Pfyffer turned to the detailed measurements. He mapped all the significant terrain objects with his plane table: ways and paths, rivers and small streams, houses, meadows, hedges, ravines and even big stones. He also differentiated oak, beech and fir forests. Under bad weather conditions he transferred the recorded sheets to a large-scaled map and hereafter he began to construct the relief.

3.2 Height measurements

The height measurements represent an area where the pioneering spirit of Pfyffer can be shown in its strongest point. The European maps produced until the first quarter of the 19th century contain no - or only isolated - height information. However, to form the landscape of his relief, Pfyffer needed the area-wide height coverage, which he obtained by performing systematic height measurements as one of the first. The result of his work is not only the relief, but also the map of J. J. Clausner that contains numerous height values (Fig. 6).

Pfyffer's principal instrument to measure heights was the barometer. During his walking-tours he annotated the terrain elevation in regular intervals, obviously based on reading off the barometer. The barometric observation must have been self-evident for Pfyffer, because, as opposed to angle or distance measurements, he never mentions his instrument or the process of his work. The presumption of barometric measurements relies on contemporary travel reports as well as on the results of the accuracy analysis. However, still many questions remain open, particularly Pfyffer's way of derivation of the height values from the quicksilver stand of the barometer, which was a great scientific challenge in the 18th century. The interpretation of Pfyffer's letters and topographic works allows to conclude, that except of barometric observations he applied trigonometric measurements as well. First, he possessed instruments which enabled a straightforward determination of vertical angles (Fig. 8) in his scale-defined trigonometric network. The second clue for trigonometric measurements of Pfyffer are the heights of alpine peaks in the map of Clausner, which were not conquered within the lifetime of Pfyffer (e.g. Jungfrau and Finsteraarhorn, firstly climbed in 1811 and 1812) and thus their heights must have been determined indirectly. As he was aware of the results of the famous arc measurements in Lapland and Peru, he supposedly took the earth curvature into consideration, however, with an unknown amount of flattening, and he probably disregarded the influence of refraction.

Pfyffer had chosen Lake Lucerne as the zero horizon for his height measurements and he estimated its height as 220 toises (428 m) above sea level. A comparison with the lake height in the modern map (434 m) shows an excellent result of Pfyffer's measurements. The small difference of 6 m is evidence of long-term thorough barometric observations as well as of a progressive procedure for the derivation of the heights.

The numerous height values published in Clausner's map constitute the best basis for the evaluation of Pfyffer's measurements. After the conversion from toises above Lake Lucerne into meters above sea level, the heights of 41 well identified localities and summits could be compared with their present values. The average difference (to-be minus is) amounted to -35 m; the negative sign is in accordance with the contemporary phenomena of unreachable mountains considered to be much higher than in reality. In average of absolute values, the 41 tested heights deviate from the present map by 57 m. A clear dependency on terrain elevation can be observed: the average absolute difference in the flat northern area is only 23 m, in variable middle land 61 m, whereas in the mountainous southern part it reaches 82 m. The detailed accuracy analysis of Pfyffer's Relief confirms that due to the same surveying basis the height accuracy of the relief roughly corresponds with the one of Clausner's map. The slight accuracy decrease of the relief in comparison with Clausner's map (the 208 tested identical points of the relief deviate from the current map of 76

m in the height in average of absolute values; Niederöst, 2003) can be explained by the procedure of relief construction. When considering a great number of height measurements undertaken in a large area with the elevation range from 500 m until almost 4'300 m, Pfyffer's achievement is very respectable.

3.3 Relief construction

Pfyffer formed the landscape of his relief on 136 mostly rectangular, transportable and clearly numbered wood plates (Fig. 9c). He marked the position and height of peaks and major localities with vertical wooden sticks and afterwards he modelled the terrain using ordinary materials found in his household: wood, brick shards, carbon rests and cardboard pieces. Then he took a mixture of gypsum and sand for surface formation and to close the gaps between the single constructional parts and he covered it with a layer of beeswax for detailed modelling. Finally, he worked out the surface objects. The forests are made out of green dark- and light-coloured drapery pieces, the water streams of chenille and the ways and small paths of white or red string strained between the nails. The houses appear as red finger-like metal and wax pieces, the churches stick out as small nails plugged upside down into the surface. A surprising result of the accuracy analysis is the unity of relief scales in three coordinate directions: contrary to previous assumptions it can be concluded that Pfyffer did not magnify the heights of the relief on purpose.

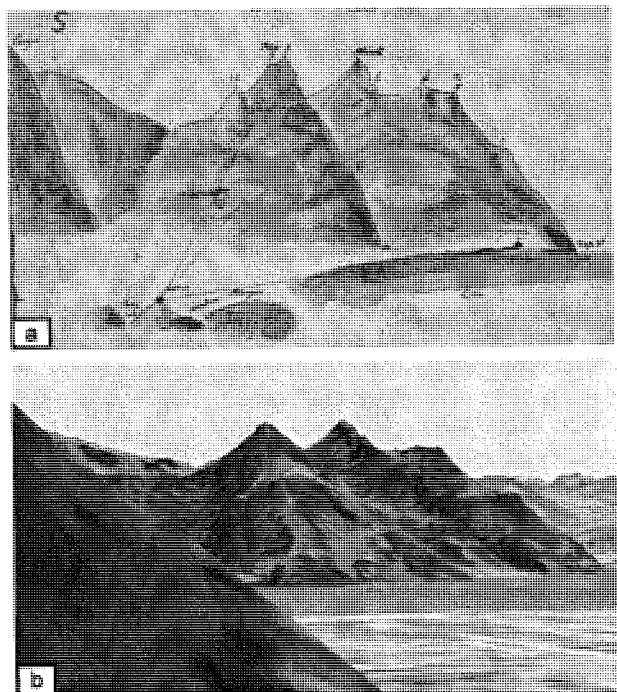


Fig. 10: (a) View of Mt. Rigi, one of 94 existing sketches of Pfyffer, (b) The same view in the present 3D model (DHM25 of swisstopo and a Landsat image of NPOC, www.npoc.ch) showing the distortion of the left-side hill and a perfectly detailed depiction of the background mountain forms.

Except of surveying data, Pfyffer also needed additional information to model mountain forms and the landscape cover of the relief. For this purpose, later relief constructors usually acquired single oblique photographs or terrestrial stereopairs. Pfyffer's way of documenting and indicating the objects to be depicted in the relief was the drawing of coloured landscape

sketches (Fig. 10a). Due to the lines of sight to the surrounded, precisely labelled or numbered summits it was previously assumed that Pfyffer created these drawings in connection with his surveying. However, a recent analysis with the help of landscape models has shown that the sketches were created and also used much later for the formation of the relief surface. Their distorted geometry excludes any systematic, perspective or panorama-like landscape projection, whereas the quality and excessive details of the morphology are amazing (Fig. 10). The computer-generated views confirm Pfyffer's concentration on the forms of selected mountains and his disregarding of geometric aspects in the surrounding area. The crucial point of the analysis was to find the places where the sketches had been drawn, which was not an easy task due to the distortions, unusual fields of view and missing or confusing station descriptions.

4. PFYFFER'S PROCEDURES AND RESULTS IN EUROPEAN COMPARISON

The first product of modern geodesy and systematic land survey was the *Carte géométrique de la France* (1739-1793) at a scale of 1:86'400. The state- and military supported mapping of France under the direction of four generations of the Cassini family was based on one superordinate and two subordinate, computationally evaluated triangulation networks. With the exception of this work, nearly all other 18th century maps were created applying similar procedures and instruments to those of Pfyffer. The following outstanding cartographic products based on graphical triangulation using a plane table and simple instruments to measure angles can be mentioned:

- The huge work of military officers of the Habsburg monarchy at a scale of 1:28'800, created within the first, so called Joseph's survey (*Josephinische Landesaufnahme, 1764-1787*) and covering several countries of today's central and eastern Europe.
- The *Atlas Tyrolensis* at a scale of 1:103'000, created by P. Anich and B. Hueber between 1760 and 1769. The atlas depicts a large area of northern Tyrol in Austria.
- Schmettau's map of Preussen (*Schmettausche Kabinettskarte, 1767-1787*) at a scale of mostly 1:50'000, covering a big part of present Germany.
- Schmitt's map of south Germany (*Schmittsche Karte, 1793-1797*) at a scale of 1:57'600.
- The *Atlas Suisse (1786-1802)* at a scale of about 1:110'000 by J. R. Meyer, J. H. Weiss und J. E. Müller. The atlas displays the entire area of Switzerland.

In order to be able to interpret the geometric quality of Pfyffer's topographic work (Table 1), the results of numerical evaluation of other maps of that period must be taken into consideration. Unfortunately, the research on accuracy analysis of 18th century maps is very rare and in addition, the measures used to express map accuracy are usually very diverse. The numeric results of several investigations are shown in the following:

- According to (Finsterwalder, 1988), the mean position error of 38 identical points of the distortion grid of the mentioned *Schmittsche Karte (1793-1797)* of south Germany is ± 2.5 km. The middle distance error of this map by the average distance of 15 km amounts to ± 1 km.
- (Brunner, 2002) analyzed the accuracy of the first topographic land survey of Sachsen in Germany (so called *Meilenblätter, 1780-1825*). This map at a scale of 1:12'000 is based on theodolite measurements and a numerically evaluated one-step triangulation. Within the accuracy

analysis, the map sheet size was compared with its to-be value of one local mile (6796.56 m). The map is very precise: the 15 selected map sheets created in 1785, 1797 and 1810 deviate from this value in average of 3 m, 68 m and 14 m respectively.

- As a representative of Swiss 18th century maps, the already mentioned *Atlas Suisse (1786-1802)* - considered to be the best map of the country until the first official land survey - was analyzed within the presented project. For the accuracy analysis the sheet Nr. 7 covering the area of Central Switzerland was selected. The evaluation using 115 identical points and applying the same procedures as those for the maps of Pfyffer (Section 2) resulted in the sigma a posteriori of 335 m in the X- and 247 m in the Y-direction respectively.

The way of creation of the above listed European maps provide evidence that at the time of Pfyffer's surveying and relief construction (1750-1786) his procedures and instruments were up to scientific standards of that period. His results hold well with the accuracy of contemporary maps and in particular, his substantial contribution lies in performing pioneering area-wide height measurements and detailed 3D modelling. Considering the conditions under which the relief was constructed - an unfavourable age for landscape exploration and surveying in Switzerland and a large mountainous area modelled at a large scale - Franz Ludwig Pfyffer's achievement is admirable even from today's point of view.

5. CONCLUSIONS

Finished in 1786, Pfyffer's relief attracted numerous visitors from all over Europe to Lucerne. Its unusual bird's eye perspective of an inaccessible mountain barrier was an absolute novelty, a virtual flight in the age of enlightenment. However, with the appearance of new, more precise maps at the turn of the century the opinions on the topographic quality of the relief started to change; the relief was reduced to a kind of handicraft and had almost been forgotten. Due to its complexity and a lack of primary documentation the relief was never investigated until the procedures of photogrammetry and image analysis allowed for drawing conclusions from the geometric evaluation. The results of the presented interdisciplinary research give the painstaking work of Pfyffer its original value back again. The relief must be considered an excellent topographic achievement of that time and a milestone on the way to modern cartography.

REFERENCES

- Brunner, H., 2002: Wie Sachsen vermessen wurde. Die Meilenblätter und die kursächsische Landesvermessung von 1780 bis 1825. Dresden, pp. 91-92.
- Finsterwalder, R., 1988: Massstab und Genauigkeit alter Karten - gezeigt an einigen Kartierungen Bayerns. In: *Cartographia Bavarica: Bayern im Bild der Karte*, München, pp. 193-211.
- Fischer, G., Schnegg, P.-A., 1994: Updating the geomagnetic survey of Switzerland, in *Matériaux pour la Géologie de la Suisse*. Géophysique No 27. Neuchâtel, Fig. 2.
- Niederöst, J., 2003: A bird's eye view on Switzerland in the 18th century: 3D recording and analysis of a historical relief model. *IAPRS, Volume XXXIV-5/C15*, pp. 589-594.
- Pfyffer, F. L., 1761: 10 letters of Pfyffer to Micheli du Crest, from April 28 to September 15. Landecy, Arch. Micheli, Papiers JBMC no 720bis.

SURFACE DISCONTINUITY MODELLING BY LSM THROUGH PATCH ADAPTATION AND USE OF EDGES

M. Pateraki*, E. Baltsavias

Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093, Zurich, Switzerland
(maria, manos)@geod.baug.ethz.ch

Commission III, WG III/6

KEY WORDS: edges, least squares matching, modelling, discontinuities, reliability, DSM, algorithm, evaluation

ABSTRACT:

This paper handles the aspect of least squares matching (LSM) for DSM generation and focuses mainly on the exploitation of edge information in the least squares approach. The mathematical model of LSM employs an affine transformation to model geometrical distortions between the template and the search images. However, the model, when used for single points that lie close or on edges does not suffice to model discontinuities, especially when patch size is large. To alleviate the above-mentioned problem, LSM is firstly extended to edgels using an approximation of the rotation between search image and template, computed from a signal analysis. The signal analysis is performed by computing the signal ellipses from the template and patch grey levels. Furthermore, the template patch can be pre-rotated and its dimensions can be adjusted, so that the patch becomes a thin ribbon along the edge. Secondly, the model is extended to edge features with topological attributes (edges and vertices with link information). The patch size is adapted to include a whole edge and a non-uniform weighting scheme is utilised in the normal equations of LSM, giving a very low weight to all pixels outside the given edge. After matching the edge middle points, the edge is divided in two segments and the procedure is repeated until the edge segments fall below a certain length threshold. The final stage is the matching of the edge end points, where other edges having the same end point are also taken into account. The proposed techniques are compared to each other and to standard LSM. An evaluation of the methods is done with ADS40 imagery. As it is shown, they lead to less discontinuity smoothing and increase of the success rate and accuracy.

1. INTRODUCTION

Least squares matching (LSM) is one of the key techniques in digital photogrammetry applied to solve feature extraction and correspondence problems as well as surface registration issues by means of least squares adjustment. The LSM method has been addressed by Foerstner (1982) and Ackermann (1984), as a generalization of cross-correlation. Several approaches have been further investigated, employing different radiometric and geometric models for surface reconstruction. Gruen (1985) utilized affine transformation for local surface patch matching assuming that the local surface patch is a plane in sufficient approximation. Gruen and Baltsavias (1988) and Baltsavias (1991) present several aspects regarding the extension of LSM to more than two images and the use of geometric constraints. This approach was later extended to linear CCD satellite images with quasi-epipolar geometry and possibly severe multitemporal differences (Baltsavias and Stallmann, 1992). Rosenholm (1987) employed bilinear finite elements to represent the true surface in multi-point matching, and formulated extensions to include breaklines and surface discontinuities (Rosenholm, 1986) but without performing any investigations. Li (1989) used multi-point matching in a multiresolution, multigrid approach and presented methods for the detection and localisation of breaklines, however with breakline assumptions that are unrealistic, especially for natural scenes. The object-space least squares matching method (Helava, 1988; Wrobel, 1991; Ebner and Heipke, 1988) has been introduced to relate information from images directly to an object space model by including height-, optical-, density-, sensor- and illumination model parameters. The model, however, proved to be computationally expensive and complex as more parameters have to be estimated, and is thus also less stable. LSM has been extended also for edge tracking and matching (Gruen and Stallmann, 1992; Tseng and Schenk,

1992; Gruen and Agouris, 1994), applied to warped images for surface reconstruction (Schenk et al., 1990) and tie point extraction (Krupnik and Schenk, 1997).

LSM is considered to be one of the most precise among existing algorithms (sub-pixel accuracy in the range of 0.01-0.05 pixels for ideal targets and without noise), but it is sensitive to large radiometric differences, is computationally expensive compared to cross-correlation, can not model excessive geometric differences, convergence problems (oscillations, false solutions) may arise if the data have insufficient signal content or if the initial approximations in the least squares solution are not close to the real solution, and, most importantly in the context of this paper, the usual LSM implementation uses area patches and thus smooths surface discontinuities. The focus in this paper is the discontinuity modeling as part of DSM extraction through the exploitation of edge information in LSM. Geometrical constraints are included in the model to reduce search space from 2D to 1D. Various modifications of LSM are presented, as well as experimental results using ADS40 images and manually selected typical discontinuity cases with various problems (e.g. outlines of building roofs) and a comparison to standard LSM.

2. DISCONTINUITY MODELING WITH MODIFIED LSM

2.1 General Estimation Model of LSM

Assume that a surface is viewed in $i=0, ..n$ images of different viewpoints. The problem statement is finding the corresponding

* Corresponding author.

part of the template image patch $f(x,y)$ in the search images $g_i(x,y)$, $i=1, \dots, n$, called patches hereafter.

$$f(x,y) - e_i(x,y) = g_i(x,y) \quad (1)$$

Eq. (1) gives the least squares grey level observation equations, which relate $f(x,y)$ and $g_i(x,y)$ image functions or image patches. The true error vector $e_i(x,y)$ is included to model errors that arise from radiometric and geometric differences in the images. For the selection of the geometrical model it is assumed that the object surface is approximated by connected local planar facets and the image patches $f(x,y)$, $g_i(x,y)$ are selected very small with respect to the image size. Therefore, a projective transformation, can be approximated by an affine transformation (Gruen, 1985). Radiometric parameters are not included in the estimation model, instead radiometric corrections are applied prior to (or some of them during) the LSM. These include noise reduction, contrast and edge enhancement and radiometric equalisation among the images.

Geometrical constraints have been proved to strengthen matching in terms of precision and reliability (Baltsavias, 1991) and can be included in the model when information about the sensor orientation model is available. Constraints are employed in the algorithm as additional weighted observation equations. In order to make constraints effective, a very large weight, small standard deviation ($s = 0.001$ pixel) respectively, must be used (Baltsavias, 1991). The least square solution of the joint system is given by Eq. (2):

$$x = (A^T P A + B^T P_i B)^{-1} (A^T P l + B^T P_i t) \quad (2)$$

where x is the vector of unknowns, A and B are the design matrices of grey level and geometric constraints observation equations respectively, P and P_i are the respective weight matrices and l and t the discrepancy vector of the observations. The weights are typically set to unity for all the grey level observation equations.

In the current implementation, the geometric constraint observation equations are introduced by means of epipolar line equations or by the more generalized polynomial equation:

$$t_k = a_n (x + dx)^n + a_{n-1} (x + dx)^{n-1} + \dots + a_1 (x + dx) + a_0 - (y + dy) \quad (3)$$

where $a_n \dots a_0$ the polynomial parameters, (x,y) the approximate pixel coordinates of the point in the search image, (dx, dy) the unknown shifts and t_k the discrepancy vector.

Compared to collinearity equations, epipolar polynomial equations are simpler to implement and can be applied to images with restricted access to their sensor model information. An approximation of height Z and an alteration factor DZ is required to derive the polynomial equation. The point on the template is projected on different height planes and the trajectory of epipolar points on the search image is determined by least squares fit of a curve. The main reason for the generalization from epipolar lines to epipolar curves, modeled by polynomials, is the extension of LSM to images, with nonframe geometry, e.g. acquired from airborne linear sensors, like the ADS40 (Fricker et al., 2000). For ADS40, trajectories of epipolar points on raw imagery follow irregular and random forms due to the unpredictable motions of the aircraft, and they can not be modeled by a

line (Pateraki and Baltsavias, 2003). On the contrary, for spaceborne linear sensors due to the smooth orbit trajectory the epipolar lines can be approximated locally by a line (Baltsavias and Stallmann, 1992). An alternative to the use of higher degree polynomials is to use less strict weights for the epipolar line constraints (e.g. $s = 0.5 - 2$ pixels). The epipolar constraint approach can be also used for ADS40 L1 images, i.e. images rectified on a plane with a user-defined average height. Such images are investigated below.

2.2 Problem Statement

Some characteristics of LSM have been mentioned in Section 1. Here, we will deal with surface discontinuities and edges. If the match points lie along edges, scales are often non-determinable and estimation of the shifts is uncertain, namely in the direction of the edge, however the shift across the edge may be precise. Even if the shaping parameters are weakly determinable, the final position may often lie close to the true position if the match points are selected along the edge, especially if geometrical constraints are enforced and the edges have a sufficiently large angle with the epipolar line. This is however not guaranteed if the approximations are poor and the patch size is small, especially in the presence of different grey values due to occlusions (Fig.1).

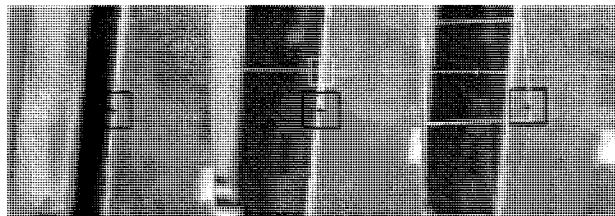


Figure 1. Example of correct (middle) and wrong (right) solution, depending on the approximate values. The template (left) is selected in the Nadir view of ADS40 L1 and the patch in the Backward view. Initial and final position of the patch is indicated with the black and white rectangle respectively. The thin white horizontal line is the epipolar line.

In the case of surface discontinuities, these will either lead to occlusions or large perspective differences. With larger bases between sensor stations higher accuracy can be achieved (better ray intersection), however, occlusions will increase. The solution to this problem is to use more than 2 images, exclude occluded rays and compute the object point from the remaining rays. ADS40 facilitates object point determination due to the quasi parallel projection in flight direction (less perspective distortions) and through the simultaneous use of 3 or more line CCDs of different viewing angles, i.e. the Nadir, Backward, Forward and possibly a spectral line, e.g. Green. In case of large perspective differences, scales and shears will be considerably different than 1 and 0 respectively and LSM will often not converge. Increasing the patch size helps convergence but then discontinuities are increasingly smoothed.

Up to now, to reduce the problems faced at discontinuities our standard multiphoto geometrically constrained matching algorithm has been used along edgels nonparallel to the epipolar lines, with small patch size (e.g. $9^2 - 11^2$ pixels), two shifts and a rotation only, and using gradients instead of grey values. The small patch size reduces discontinuity smoothing and the danger of including in the patch varying grey values due to occlusions. Use of gradients, eliminates the problem of homogeneous surfaces varying due to occlusions (e.g. in one image a black roof

shadow is visible, in another a bright vertical wall). However, the small patch size may lead to instabilities in the determination of the shifts and rotation, while the matching approximations must be very good. The matched edgels must be very dense along the discontinuity to model it nicely, while edgels parallel to the epipolar line can not be matched due to multiple solutions.

Some possible solutions to the above mentioned problems are investigated below through the use of edgels and edges and modifications of LSM. The edgel approach is similar to the one described above, but with a pre-rotation of each patch to be approximately aligned with the template. This improves approximations, stabilises the least squares solution, speeds up convergence and computing time. In addition, the template patch is rotated and made a thin rectangle, so that in case of a straight edge, only this is included in matching and not the flanking regions, which may be partially occluded, thus increasing the success rate and accuracy and leading to less smoothing. The edge approach is quite different than the approach used up to now and is described in Section 2.4. For these initial tests of the developed methods, 20 characteristic roof points covering different problem cases have been manually selected. In all cases, the epipolar constraints were used and the affine transformation of LSM was replaced by a conformal one, due to poor determinability of the affine parameters at edge points.

2.3 Edgel Matching

Matching is applied to edge pixels which are extracted in the template image beforehand by the Canny operator (Canny, 1986) or the gradient thresholding (in-house operator), or by manual selection (e.g. in case of semi-automated measurements for building roof reconstruction). Edges extracted with the Canny operator or the gradient thresholding have width of one pixel or more, respectively, and these edgels of the template image should be matched in the other images.

To derive initial approximations for the rotation, a signal analysis in the template and each patch is made. The signal ellipse is computed in a window centered at the template point and for each patch at the position after the first iteration of LSM with geometric constraints, i.e. on the epipolar line and at an improved approximate position. For edges, the signal ellipse is elongated with the major axis in edge direction. The ellipse measures, size of semi-major (q_a) and semi-minor (q_b) axes, angle (ϕ) between x direction and the semi-major axis and eccentricity (e) of the signal ellipse are computed according to:

$$q_a^2 = \frac{(\sum g_{xx} + \sum g_{yy})}{2} + \sqrt{\frac{(\sum g_{xx} - \sum g_{yy})^2}{4} + \sum g_{xy}^2}$$

$$q_b^2 = \frac{(\sum g_{xx} + \sum g_{yy})}{2} - \sqrt{\frac{(\sum g_{xx} - \sum g_{yy})^2}{4} + \sum g_{xy}^2} \quad (4)$$

$$\phi = 0.5 \cdot \text{atan2}(2 \cdot \sum g_{xy}, \sum g_{xx} - \sum g_{yy})$$

$$e = \sqrt{1 - \frac{q_b^2}{q_a^2}}$$

where $\sum g_{xx}$, $\sum g_{yy}$ and $\sum g_{xy}$ are the sums of the x- and y-gradient products in the window.

If the approximation for one of the patches is poor (e.g. far away from the edge), then the signal ellipse may not be elongated and the procedure below is interrupted. Otherwise, the pre-rotation of each patch is determined by the difference of the angle ϕ between patch and template. Determination of ϕ may be wrong, especially if the patch is large and contains edges in other directions (see Fig. 2, middle). To reduce this problem, smaller patches (e.g. $9^2 - 11^2$) are justified. An example with and without pre-rotation is shown in Fig. 3(a). The seemingly wrong pre-rotation in the right image of Fig. 3(a) is just a visualisation problem. The signal ellipse and ϕ were not determined at the black rectangle shown, but at the result of the first iteration on the epipolar line. At this position, the edge was only partially included in the patch, being otherwise homogeneous, and thus the estimation of ϕ was not reliable. This problem stresses the necessity of good shift approximations before computing the signal ellipse. Furthermore, use of gradients instead of grey values has been tested. This often improves the success rate, see an example in Fig. 3(b).

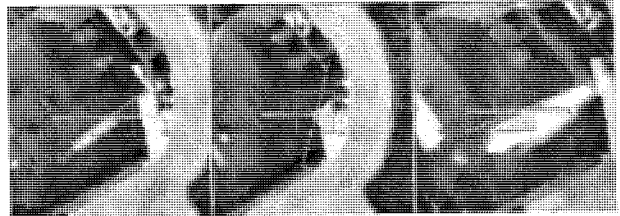


Figure 2. Selection of large patches (17 x 17) on edge discontinuities. Left: template (Nadir), middle: Backward, right: Forward. The dotted line indicates the direction of the ϕ angle with respect to x-axis (horizontal line).

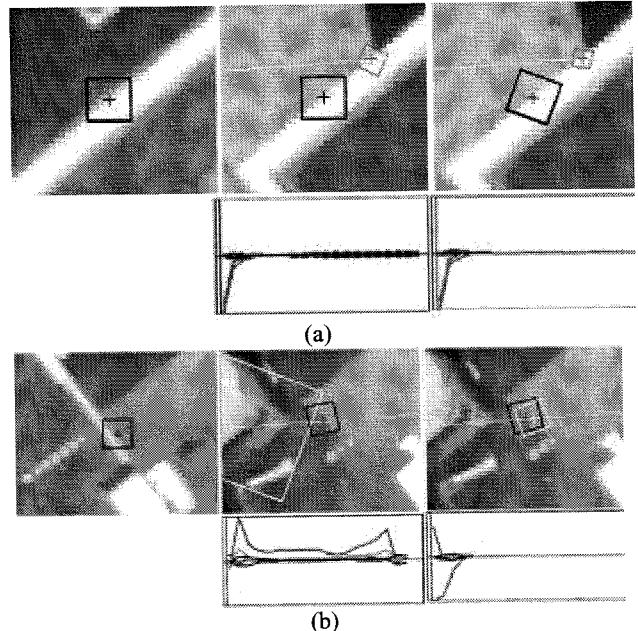


Figure 3. On the left, the selected template (9 x 9) in Nadir view, in the middle and on the right the solution the Backward view. (a) without and with pre-rotation; (b) with pre-rotation but in the middle using grey values and on the right using gradients. Initial (black) and final (grey in (a), white in (b)) position and shaping of the patch are shown. At the lower part of (a) and (b), the alterations of shaping parameters are shown.

For ADS40 L1 images the rotation, scales and shears between the images are small, since the images are rectified. In spite of that, pre-rotation with gradients compared to matching without these two options, and using the 20 test points, led to 3 times less iterations, decrease of oscillations for scales and especially shears (compare bottom of Fig. 3(a)) and increase of the success rate by 35%. Figure 2 (and also Fig. 5) shows well also the quite large perspective differences between the template images (usually selected from the Nadir line CCD) and the images from the Forward CCD with an angle of 28° to the Nadir, while the Backward CCD with 14° angle to the Nadir has less differences to the Nadir images.

To include in matching only a narrow region across the edge, the following procedure is followed. For the template, a thin rectangular patch is used and the long dimension is aligned with the edge direction. The direction of the long patch dimension is computed from the value of ϕ from the signal analysis of the template. The patch dimension across the patch can either be selected small for all edges (e.g. 5 pixels) or can be estimated from the analysis of the grey level ramp for each edge. The patches have the same dimensions as the template patch and are pre-rotated with respect to this new template patch orientation, again using the ϕ values of patch and template. The reshaping and rotation of template (and the patches) occurs before the iterations of LSM start, while the pre-rotation of the patch to fit to the template is estimated after the first iteration, as mentioned above. Fig. 4 (a) shows an example of pre-rotation with wrong matching for the Backward channel, while in Fig. 4 (b) also a template reshaping is included. Tests with the 20 points showed that false matching could be almost eliminated, as long as the derived edge directions in template and patch are reliable.

In all above procedures, the angle ϕ could be also estimated from the orientation of the edgels, e.g. from Canny. This may have the advantage that small gradients (e.g. noise) do not influence the computation of ϕ , as in Eq. (4).

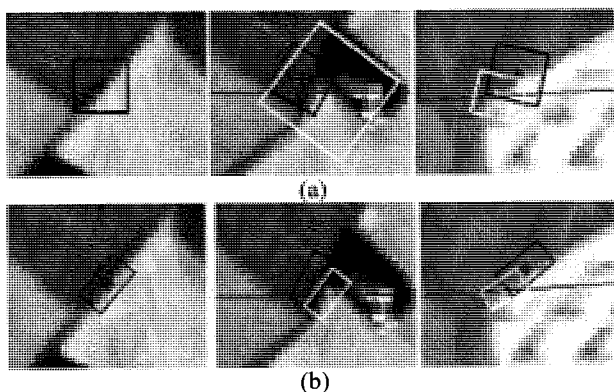


Figure 4. LSM with pre-rotation of patch (a) and with reshaping of template and patch (b). Left: template, middle: Backward, right: Forward. The initial patch and position is shown in black and the final one in white.

2.4 Edge Matching

An alternative approach to edgel matching is to use the full edge length. A feature-based approach of LSM is utilized. Single extracted edgels (one-pixel wide edges, i.e. extracted with Canny) are linked into edges. The edgel aggregation is a sequential process, where the significant edges are aggregated before the weaker ones (Henricsson, 1996) and small gaps are bridged based on criteria of proximity and collinearity. The generated edge graph is post-processed based on defined attributes in or-

der to remove weak edges and fit to the edgels straight edge segments, as long as possible. The attributes can be geometrical (e.g. length and curvature of edge), radiometrical (e.g. strength of edge) or topological (e.g. adjacency, common end points). Edges are extracted only in the template image and for each edge to be matched the patch dimension of the template is expanded in order to contain the edge points and the end points (vertices) of the edge (see left patch in Figure 5 (a), (b), (c), (d), (e)). The patch is centered at the middle point of the edge and initial approximations of the shifts, scales and shears are set to 0, 1 and 0, respectively. In addition, a different weighting scheme is used. For all grey level observation equations that correspond to edge pixels, weights are set to 1.0 and for the remaining equations weights are smaller (0.1). Thus, the contribution of off-the-edge pixels in the least squares solution is reduced. The patch should be large enough (e.g. more than 5×5) to avoid singularities of the system, as the number of observations should be larger the number of unknowns and the smaller weights must not be zero, especially for short edge lengths (and thus small patch sizes).

After matching the edge middle point, other edge points in the template image are selected with a step of 1 pixel. For these pixels, approximations in the patch are derived from the conformal transformation computed between patch and template for the edge middle point. The quality of these approximations depend on (a) the change of height (parallax) along the edge, and (b) the accuracy of the conformal parameters. The further the edge point is from the middle point, the higher the influence of these two factors and the worse in general the approximations. In Fig. 5 in (b) and (c), the dotted edge lines show these approximations. In (c) it is well visible that approximations are poorer for points far away from the middle point. Matching continues for the edge points, except the middle one, as follows. The edge is divided by two, and each edge segment is treated as the whole edge above, i.e. the template patch includes the edge segment end points and matching is performed. However, in this case quite good approximations exist for the shifts, derived from the matching of the whole edge. This continues by dividing each segment in two and so on and so forth, until the edge segments become less than 11 pixels long. At this stage no matching has been performed yet for the two end points of the original edge, but approximations exist from the matching of the middle points of the two outer edge segments. For matching the end points, square patches (9×9 pixels) are used. Before adopting the result as correct, the convergence and the change of the matching solution found from the initial approximation are checked. If the edge end points are end points for more than one edge, then the approximation for matching the end points are derived from the average of the approximation from each edge. Thus, common end points of edges get one common height and the edges are connected in 3D space.

This approach proved relative robust, with fast convergence (3-7 iterations), and due to the often large patch size, requires less good approximations than conventional LSM, reduces the danger of multiple solutions and leads to a safer determination of rotation. Even the case of edges parallel to the epipolar line could be partially treated, when at one of the end points of the edge (that are included in the template patch), other edges with a large enough angle to the epipolar line were included, i.e. corners on roof tops (Fig. 6). However, the problem of edges parallel to the epipolar lines is not generally solved. Furthermore, occlusions may still cause false matching, or even if matching is correct, they may cause wrong scale and rotation estimation in the conformal transformation, and thus wrong approximations for the remaining edge points to be matched.

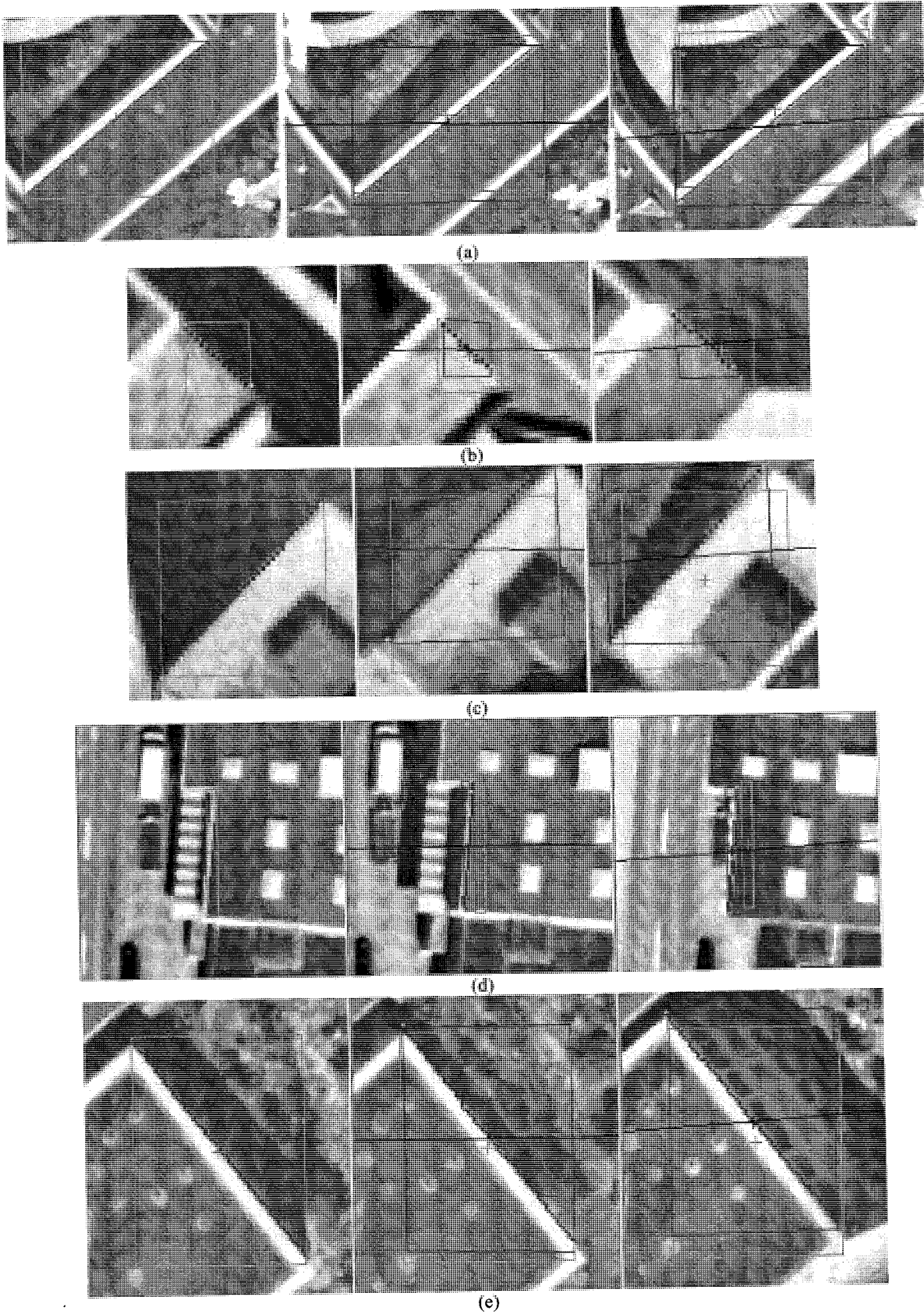
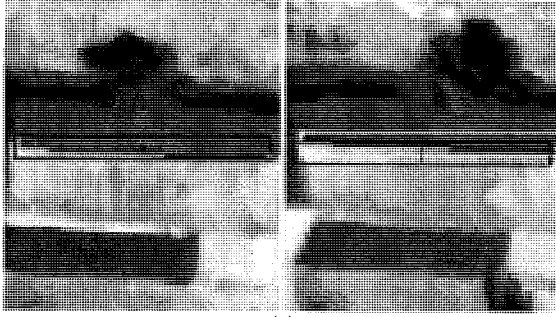


Figure 5. Examples of edge LSM. Edges are extracted in the template image (left) and matched in the remaining images (middle, right). Convergence of the least squares solution is achieved in 3 to 7 iterations. The dots indicate the edge points. Results are illustrated after the LSM solution for the middle of the edge. The light and dark grey rectangles show the initial and final patches respectively.



(a)

Figure 6. LSM for edges parallel to epipolar lines. Initial (black), final (white) position and matched edge are shown.

The edge method could be combined with preshaping of the template as described in Section 2.3, in order to save computation time. However, execution speed is not, at this stage of investigations, the major concern, plus the method of excluding or reducing the influence of pixels in the patch through appropriate weights is general, flexible and elegant.

3. CONCLUSIONS

Various modifications of LSM, for improving its performance at surface discontinuities, have been presented and evaluated. In particular, these methods adapt the template patch orientation and dimensions, improve the approximations for the rotation and shift of the patches, decrease the influence of off-edge pixels and reduce multiple solutions for edges parallel to the epipolar lines. These measures lead to less discontinuity smoothing and increase the success rate and accuracy. The methods should be improved and refined, and also extended to better handle still unsolved problems, e.g. partially occluded edges. After these developments, extensive tests over large areas with manually measured DSMs (incl. building roofs) are planned.

REFERENCES

Ackermann, F., 1984. High Precision Digital image Correlation. In: Photogrammetric Week, No. 9, Institute for Photogrammetry, University Stuttgart.

Baltsavias, E.P., 1991. Multiphoto geometrically constrained matching. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. 49, 221 p.

Baltsavias, E.P., Stallmann D., 1992. Advancement in matching of SPOT images by integration of sensor geometry and treatment of radiometric differences. In: IAPRS, Vol. 29, Part B4, pp. 916-924.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6), 679-698.

Ebner, H., Heipke, C., 1988. Integration of digital image analysis and object surface reconstruction. In: IAPRS, Vol. 27, Part B11, pp. III: 534-545.

Foerstner, W., 1982. On the Geometric Precision of Digital Correlation. In: IAPRS, Vol. 24, Part 3, pp. 176-189.

Fricker, P., Sandau, R., Walker, A.S., 2000. Development of an airborne digital sensor for photogrammetric and remote sensing applications. Proc. ASPRS Annual Conference, Washington DC, USA (on CD-ROM).

Gruen, A., 1985. Adaptive least squares correlation: A powerful image matching technique. South African Journal of

Photogrammetry, Remote Sensing and Cartography 14 (3), 175-187.

Gruen, A., Baltsavias, E.P., 1988. Geometrically Constrained Multiphoto Matching. PERS 54(5), 633-641.

Gruen, A., Stallmann, D., 1992. High accuracy matching of object edges. Proc. SPIE, "Videometrics", Vol. 1820, pp. 70-82.

Gruen, A., Agouris, P., 1994. Linear feature extraction by least squares template matching constrained by internal shape forces. In: IAPRS, Vol. 30, Part3, pp. 316-323.

Helava, U.V., 1988. Object-space least-squares correlation. PERS 54 (6), 711-714.

Henricsson, O., 1996. Analysis of Image Structure using Color Attributes and Similarity Relations. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No.59.

Krupnik, A., Schenk, T., 1997. Experiences with matching in the object for aeriotriangulation. ISPRS Journal of Photogrammetry and Remote Sensing 52 (4), 160-168.

Li, M.X., 1989. Hierarchical multi-point matching with simultaneous detection and location of breaklines. Ph.D. thesis, Royal Institute of Technology, Department of Photogrammetry.

Pateraki, M., Baltsavias, E., 2003. Analysis and Performance of the Adaptive Multi-image matching Algorithm for Airborne Digital Sensor ADS40. Proc. ASPRS Annual Conference, Anchorage, AK, USA, 5-9 May (on CD-ROM).

Rosenholm, D., 1986. Accuracy improvement in digital matching. Photogrammetric Reports No. 52, Transactions of the Royal Institute of Technology, Department of Photogrammetry, Stockholm, Sweden.

Rosenholm, D., 1987. Multi-point matching using least squares technique for evaluation of three-dimensional models. PERS 53 (6), 621-626.

Schenk, T., Li J.-C., Toth, C., 1990. Hierarchical approach to reconstruct surfaces by using iteratively rectified imagery. In: IAPRS, Vol. 28, Part 5/1, pp. 464-470.

Tseng, Y.-H., Schenk, T., 1992. A least squares approach to matching lines with Fourier descriptors. In: IAPRS, Vol. 29, Part B3, pp. 469-475.

Wrobel, B.P., 1991. Least-squares methods for surface reconstruction from images. ISPRS Journal of Photogrammetry and Remote Sensing 46 (2), 67-84.

EXPERIENCES ON AUTOMATIC IMAGE MATCHING FOR DSM GENERATION WITH ADS40 PUSHBROOM SENSOR DATA

M. Pateraki^{1,*}, E. Baltsavias¹, U. Recke²

¹Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093, Zurich, Switzerland
(maria, manos)@geod.baug.ethz.ch

²Leica Geosystems GIS & Mapping GmbH, CH-9035 Heerbrugg, Switzerland
utz.recke@gis.leica-geosystems.com

Commission II, IC WG II/IV

KEY WORDS: pushbroom, sensor, systems, matching, surface, quality, analysis, performance

ABSTRACT:

This paper presents an analysis on automatic image matching for DSM generation with ADS40 airborne push-broom sensor data. ADS40, produced by Leica Geosystems GIS & Mapping (LGGM), offers on a single camera system the possibility to acquire both panchromatic and multispectral images in up to 10 channels (100% overlap), incorporating latest GPS and INS technology for direct sensor orientation and latest developments in sensor technology, optics, electronics, data transfer and storage. The radiometric and geometric characteristics of the sensor can reinforce matching in automated processes through the use of multiple channels with small perspective distortions, direct georeferencing and superior radiometric quality. In this study, the matching performance for DSM generation of the commercial digital photogrammetric system SOCET SET is evaluated. In addition, matching software developed within a joint project of ETH Zurich and LGGM, making use of ADS40 special characteristics, is utilized and compared with the above commercial system. DSM extraction is tested on rectified imagery with adjusted orientation data. The evaluation of the matching performance is done through a qualitative and quantitative analysis in test areas in Switzerland and Japan, of different terrain relief and land cover, as well as of different types of buildings and roofs. Results are analyzed and compared with manually measured reference data.

1. INTRODUCTION

1.1 State-of-the-art and Proposed Evaluation

Single- and multi-line CCDs are employed as research tools in satellite- and airborne- based sensors and are used to acquire panchromatic and multispectral imagery in pushbroom mode for photogrammetric and remote sensing applications. Regarding airborne sensors, several systems have been developed and among them fewer commercial ones, e.g. ADS40 (LGGM), DMC (Z/I Imaging), Starimager (Starlabo), Ultracam-D (Vexcel). New methods, compared to the existing ones for processing of scanned aerial films, are necessary for digital sensors, especially line-CCD-based ones as they have significant differences to the existing film-based cameras, e.g. several (up to 9) CCD-lines with 100% overlap, a non-perspective geometry in flight direction, different radiometric characteristics, simultaneous multispectral imaging capabilities, more complicated imaging geometry and integration of GPS/INS systems for determination of the position and orientation of each line. Investigations regarding airborne linear-CCDs have been already performed regarding camera architecture, direct georeferencing, sensor modeling, ground processing and aerial triangulation for ADS40 (Fricker, 2001; Hinsken et al., 2002; Sandau et al., 2000; Tempelmann et al., 2000; Tempelmann et al., 2003) and for other systems (Fritsch, 1997; Haala et al., 2000; Hoffmann et al., 2000; Leberl et al., 2003; Tianen et al., 2003; Wewel et al., 1998). Notably fewer studies exist on matching methods and DSM generation using airborne linear CCDs (Gwinner et al., 1999; Neukum, 1999; Renouard and Lehmann,

1999; Scholten, 2000). Recently, in the dedicated workshop on digital aerial cameras during the Optical 3D Measurement Techniques conference in 2003, only the papers of Nonin (2003), Pateraki and Baltsavias (2003b) and Zhang and Gruen (2003) have been focused on automatic DSM extraction and only the last two on algorithmic aspects. Regarding digital photogrammetric stations, advancements in matching for tie point extraction and DSM generation have generally not been integrated in the current systems, although images of different sensor orientation models, employing line CCDs, can be imported. To date the algorithms used traditionally for the processing of aerial frame imagery are employed also for airborne linear CCDs.

The aim of this paper is to evaluate the matching performance on ADS40 images, using the commercial SOCET SET digital photogrammetric software package (from this point forward called SS) and the matching software that has been developed at ETH Zurich, adapted to ADS40 characteristics (from this point forward called AIM). However, as each system employs a different matching strategy (see Section 2), the analysis is focused on the quality of the final product, the DSM respectively. The quality of the extracted terrain is evaluated in different areas of terrain relief and land-cover, using manually measured reference data.

1.2 ADS40 and Sensor Plate Design

To date the ADS40 camera system and architecture has been described in several publications, as mentioned already in Section 1.1. In brief, ADS40 consists of seven parallel CCD lines in the focal plane of a single lens system – three panchromatic (forward-, nadir-, backward looking), red, green and blue placed next to each other and near infrared. In addition, ADS40 incor-

* Corresponding author

porates GPS and INS technology (Applanix) for direct georeferencing.

In this study, datasets have been acquired with two different sensor plate (SP) designs. The standard and most known configuration applies to SP1 design with staggered panchromatic lines in backward (PANB), forward (PANF) and nadir (PANN) position, the multispectral red, green and blue lines (RGB) between forward and nadir lines and the near infrared line (NIR) close to nadir. In the SP2 design, the RGB triplet is set at the nadir position (preferred for true orthophoto generation), two panchromatic lines in forward viewing (PANF), one panchromatic line for backward viewing (PANB) and one near infrared line between nadir and forward. In SP2, the staggered mode for the PAN CCDs is not used. In Figure 1, the SP1 and SP2 designs are illustrated and in Table 1, the respective viewing angles of the CCDs are listed.

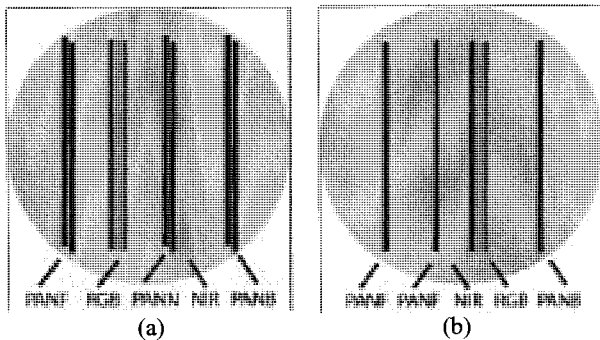


Figure 1. Configuration of the CCD lines on the two different sensor plate designs, SP1 (a) and SP2 (b). The panchromatic lines forward, nadir and backward are indicated as PANF, PANN and PANB respectively, near infrared as NIR and the color triplet as RGB.

CCD lines	SP1	SP2
PAN	+28°, 0°, -14°	+28°, +16°, -14°
RGB	+16°	0°
NIR	-2°	+14°

Table 1. Viewing angles of the CCDs for the two different focal plate designs.

The acquired ADS40 channels are further rectified onto a height plane (Lev1 images) in order to be used for stereo viewing and in automatic matching processes for tie point and DTM/DSM extraction. In Lev1 images, differences between channels due to scale, but also rotation and shear that exist in raw images (Lev0 images) are removed to a large extent. However, the option of utilizing Lev0 images in aerial triangulation (AT) is being investigated (not handled in this paper), in order to accelerate a part of the ground processing chain (rectification and automatic tie point extraction).

2. DATASETS AND SYSTEMS

2.1 Datasets

Two ADS40 datasets were used in these investigations. The first was acquired over the rural area of Waldkirch area in Switzerland and the second over the dense city center of Yokohama in Japan. The Waldkirch block was flown with SP1 camera by LGGM in May 2002, and consisted of 4 parallel and 2 cross-

strips. The coordinate system used was the WGS84. The block of Yokohama consisted of three parallel strips and was flown by Pasco Corp. The coordinate system used was the Japanese grid. Both datasets included panchromatic and multispectral imagery in Lev0 (raw) and Lev1 product with 0.20 m ground sampling distance.

In terms of radiometric quality, the Yokohama compared to the Waldkirch dataset exhibited higher noise. Interpretability of objects was more difficult, as denser and higher buildings existed in combination with strong shadows (in many cases saturated) and poorer radiometric quality (Fig. 2). All images used for DSM extraction have been pre-processed in order to reduce noise, improve feature definition and minimize radiometric differences among channels (Pateraki and Baltasvias, 2003a). This part was essential, notably for the Yokohama dataset, in order to help matching in shadowed areas (Fig. 3).

Regarding geometric quality, each individual camera with SP1 and SP2 has been calibrated over a test field with precisely measured control and check points, and interior orientation and IMU misalignment parameters have been estimated. These have been later used in AT which was carried out for both datasets, using ORIMA software, in order to adjust and refine the orientation parameters acquired from the GPS/INS systems on board. Tie points were automatically measured using Automatic Point Matching (APM) module of SS software. Blunders could be visually controlled and iteratively eliminated. GCPs in the Waldkirch dataset were distributed at the block corners, whereas for Yokohama at the block center. The derived geometric accuracy from bundle adjustment in terms of sigma a-posteriori was 2.5 μ m for the Waldkirch dataset. For the Yokohama dataset the geometric accuracy was lower, due to the poorer radiometric quality (more blunders in automatic tie point measurement), small errors in the recordings of the GPS and the poorer block geometry (no cross strips). Table 2 summarizes acquisition and bundle adjustment parameters of the two datasets.

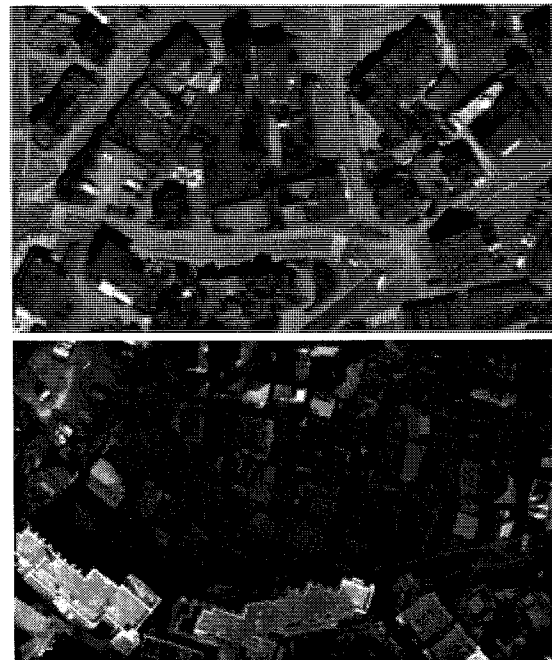


Figure 2. Original image quality of Waldkirch (top) and Yokohama (bottom) dataset.



Figure 3. Preprocessed images of Waldkirch (top) and Yokohama (bottom) dataset.

Parameters	Waldkirch	Yokohama
Coord. system	WGS84	Japan. grid
Acquisition date	May 2002	October 2003
Camera	SP1	SP2
No. strips [paral./cross]	4/2	3/0
Ground pixel size [m]	0.20	0.20
Sensor pixel size [μm]	6.5	6.5
Radiometric quality	Good	Average-poor
No. GCPs	8	5
s0 [μm]	2.5	7.2
Flying height [m]	~2000	~1944

Table 2. Acquisition and bundle adjustment parameters.

2.2 Systems

In SS, the adaptive method or AATE (Adaptive Automatic Terrain Extraction) was used. Adaptive matching can use more than two images, can generate regular grids or triangulated irregular networks (TINs), changes some of the strategy parameters based on an “inference” engine, and computes the mean terrain inclination in small neighbourhoods. Based on this inclination and image exterior orientation the two best ones out of all available images are selected. This selection is preferred (e.g. Bacher, 1998, Baltsavias et al., 2001) and can lead to better results compared to the non-adaptive as problems due to occlusions and large perspective differences can be reduced by an appropriate choice of images. In some cases AATE produces severe errors at image borders, i.e. the terrain is flattened. The matching method, utilized in SS, uses area patches, which lead to smoothing of surface discontinuities. The TIN method is inherently based on the grid matching approach utilised in SS (no interpolation is performed at the last stage, for grid points that have not been successfully matched).

The AIM method is based on a combination of area and feature based matching techniques. Different types of primitives (area patches, single edgels that belong to contours, edges) are combined based on the type of the terrain (rugged, steep, flat). However, since AIM is still an experimental system under fine-

tuning, several parameters are set by the user, according to the area and terrain type. The description of the algorithmic approach exists already in the literature (Pateraki and Baltsavias, 2003b), and below only a brief overview of AIM is given. Two types of matching strategies can be utilized, namely single and multi-template strategy. The first is applied in case of relative flat terrain, whereas the second in more complex areas. Multi-resolution levels are employed in a doublet approach (Pateraki and Baltsavias, 2003a) in order to acquire approximate values. More than two images are matched simultaneously, geometrical constraints are enforced by means of quasi-epipolar curves, and 3D position is computed only from the good rays, following correlation and blunder detection. In the upper levels, a surface approximation is derived by matching of grid points (favorable for faster processing) and which is subsequently refined in the lower levels by inclusion of linked contour points. Initial positions at each level are derived by a multi-patch approach, utilizing cross-correlation and three masks of different size. Least squares matching (LSM) with geometrical constraints is further used for verification and refining the matching solution and is applied for straight edges and single points (edgels and grid points). The main reason for extending LSM to straight edges is to improve modeling of discontinuities and minimize surface smoothing (Pateraki and Baltsavias, 2004).

As automatic matching in each system is based on different strategies, the assessment is focused on the quality of the final product, the DSM respectively. Alternatively, an analysis on a different level, namely forcing the systems parameters to be relatively similar, would not be realistic for SS as it has certain limitations for full control of the matching strategy and blunder detection. For AIM, modifications would be feasible in terms of implementation, to a certain extent (to adapt some of its parameters to the ones of SS, e.g. using area-based grid matching). However, this would be less favorable as the AIM method takes into consideration several characteristics of ADS40 (Pateraki and Baltsavias, 2003a) and uses different primitives for an optimal matching strategy, in contrast to SS. In both cases, the pyramid levels and the initial mask sizes have been set to equal values and the same number of images has been used. The three stereo panchromatic channels and the Green channel taken out of one strip have been used as input in all systems. In SS, the TIN version without additional filtering (elimination of tress/buildings/other objects) has been used. Similarly for AIM, additional smoothing has been excluded and raw matched data have been used in the analysis (irregularly distributed points). Table 3 lists the basic strategy parameters used in each system.

Parameter	SS	AIM
Primitives	Area patches	Area patches/ contour points/ edges
No pyramid levels/ matching passes	6/8	6/6
No of images	4	4
Type of matching	Image matching using two “best” images	Image matching all images simul- taneously

Table 3. Matching parameters.

2.3 Reference Data

In order to check the matching accuracy, reference datasets were derived from ADS40 images. Mass points and breaklines have been manually collected in stereo mode in SS with an estimated accuracy of 20-40 cm and 40-60 cm for Waldkirch and Yokohama dataset respectively. The same orientation used later for

the automatic extraction was used also for the manual measurements. Mass points have been separated in two classes: ground points close to the perimeter of buildings and ground points defining the bare earth (BE) surface excluding points close to buildings. This classification is justified as errors arising from matching should be treated separately from modeling errors of the surface in 2.5D. Breaklines have been extracted along discontinuities on the ground and on buildings or man made objects (MMO), even small features on roofs. Points along the breaklines have been interpolated every 0.20 m, corresponding to approximately 1 pixel in image space. Forest areas and trees have been excluded from the measurements as it was difficult to extract reliable 3D coordinates, due to their imprecise and unclear shape.

Moreover, due to the radiometric quality of the Yokohama dataset, many roofs, especially the smaller ones lying partly in the shadowed areas, could not be accurately extracted and for several manually measured points the estimated accuracy degraded (~ 60 cm).

Roadmarks were not always easily visible due to the shadowing. Less points could be extracted in open surfaces and close to building outlines and therefore the class used to analyze separately modeling errors was not used.

3. RESULTS OF IMAGE MATCHING AND QUALITY EVALUATION

DSM extraction was tested in selected regions, excluding large shadowed areas (especially in the Yokohama dataset). As the Waldkirch region had more variations in land cover, three regions have been selected¹. The relief in area W1052A (Fig. 4 (a)) was undulating and sparse houses, trees and small salt lakes existed. The area W1052B (Fig. 4 (b)) was relatively flat with good texture in agricultural fields. The area W1052C (Fig. 4 (c)) included complex objects, well defined roads, single and groups of trees, in many cases close to the houses. One representative region has been selected in the Yokohama dataset (Fig. 5) and it included high buildings (some over 20 m high) and thus large discontinuities and occlusions. In combination with the low radiometric quality, the degree of difficulty in extracting a reliable DSM of this area increased.

The parameters of each system have been modified according to the characteristics of each area. As mentioned already in Section 2.2, in SS the AATE method was used in all tests in combination with the TIN option. For the Waldkirch test areas W1052A and W1052C, 1 meter grid interval in object space was used and for W1052B, as the terrain was less undulating, the grid interval was set to 1.5 m. For the Yokohama test area a denser matching was employed with 0.5 m grid spacing due to the density of buildings and large parallax differences. In AIM, the strategy of each area varied. Single and multiple template strategies were used for areas W1052A, W1052B and Y0624C, W1052C respectively. An approximate surface in the upper pyramid levels was generated by matching of grid points (5 m interval), and further refined by matching of edges in the lower levels. However, for area W1052B edge matching was used only in the two lower levels as the area was less undulating.

The results of the quantitative analysis, performed for the above areas, are shown in Table 4. The raw DSM points without any postprocessing (e.g. filtering, modeling of breaklines) and the interpolated elevations of the reference-measured points were

¹ The naming scheme and coding of the areas indicates dataset, strip and region of interest. E.g. W1052A is the region A, imaged in strip 1052 of Waldkirch (W) dataset. The number of the strip is also used to indicate that the images used in matching are selected only from this strip in case the area is imaged in multiple strips.

utilized to compute statistics of elevation differences (RMS, mean with sign, absolute maximum and standard deviation).

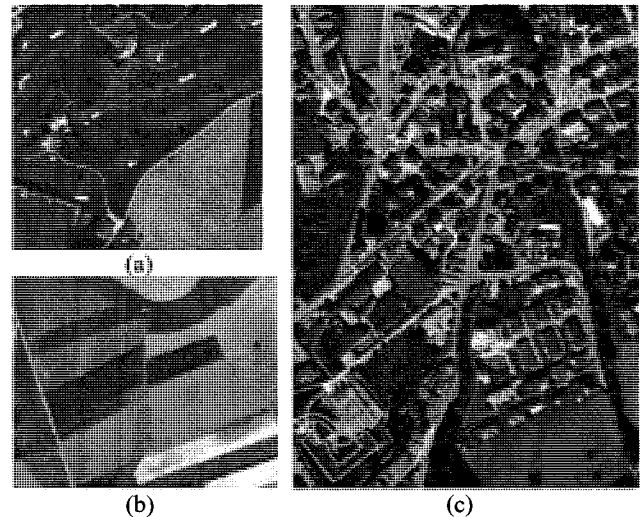


Figure 4. Test areas in the Waldkirch dataset. The size of W1052A (a), W1052B (b) areas was approximately 2000 x 2000 pixels and of W1052C (c) area 2500 x 1800 pixels.



Figure 5. Part of test area Y0624C in the Yokohama dataset. The total size of the area is 1000 x 3000 pixels.

Errors with respect to points on bare earth (BE) and on man made objects (MMO) are analyzed separately (Pts, Brkln). The number of single points on BE was less than the interpolated breakline points and the first were measured in locally flat areas. Modeling errors are separately computed as error statistics from points that lie close to the buildings and on the ground (hPts). In general, AIM delivered more accurate results than SS in all test areas. In relative flat areas (W1052B), less differences in the accuracy could be observed. Blunders in DSMs derived with SS were significantly higher, in locally flat areas and in the case of MMO but also at breaklines. AIM shows a better performance in case of discontinuities, as accuracy increased by approximately two times and blunders were less than SS for all areas in the case of MMO. For both systems accuracy degraded for points close to building outlines, as less points could be automatically derived in these areas. Here, elevation was interpolated from the nearest points, which apparently were either on the ground or on buildings, leading thus often to large errors. In addition, planimetric errors, even if small, can also cause at surface discontinuities large height errors. SS has generated more points than AIM in all areas except W1052B. The reason for this is the inclusion of edges with AIM, leading to extraction of many points in highly-textured areas, e.g. the agricultural fields. In the case of Yokohama, as one could expect, accuracy degraded more, as the geometric and radiometric quality of the data was poorer and the acquisition of the reference data less accurate, compared to Waldkirch. Still, AIM delivered better results, especially along discontinuities on MMO, with an RMS

Test data/ number of matched points {SS/AIM}	Type of points/number of comparison points	System	RMS (m)	Mean with sign (m)	Maximum absolute (m)	Std. dev. (m)
W1052A/ {292157/166009}	BE {Pts}/40	SS	1.29	-0.82	3.26	1.00
		AIM	0.56	-0.22	1.00	0.52
	BE {Pts, Brklin}/2500	SS	0.95	-0.75	2.80	0.60
		AIM	0.70	-0.49	1.69	0.44
	BE (hPts)/43	SS	2.27	-1.90	4.78	1.26
		AIM	2.22	-1.56	3.47	0.90
MMO {Brklin}/1000	SS	1.20	0.09	4.11	1.18	
	AIM	0.66	-0.02	2.32	0.66	
W1052B/ {141512/342692}	BE {Pts}/30	SS	0.86	0.64	1.84	0.68
		AIM	0.58	0.34	1.17	0.50
	BE {Pts, Brklin}/1100	SS	0.37	-0.10	1.84	0.33
		AIM	0.45	-0.36	1.27	0.30
W1052C/ {736749/283956}	BE {Pts}/37	SS	0.72	0.43	2.25	0.59
		AIM	0.55	0.33	1.09	0.45
	BE {Pts, Brklin}/24840	SS	1.38	0.19	15.05	1.36
		AIM	0.89	0.24	7.42	0.86
	BE (hPts)/334	SS	1.79	-0.26	19.24	1.77
		AIM	1.23	-0.46	4.66	1.14
MMO {Brklin}/10000	SS	0.96	0.33	7.89	0.89	
	AIM	0.60	0.17	2.83	0.58	
Y0624C/ {128950/40750}	BE {Pts}/33	SS	4.55	-1.96	14.31	-1.96
		AIM	2.45	-1.71	4.99	1.78
	BE {Pts, Brklin}/6553	SS	4.10	-2.65	19.09	3.10
		AIM	3.28	-2.50	9.80	2.13
	MMO {Brklin}/41738	SS	7.56	1.43	103.96	7.42
		AIM	1.30	0.30	7.87	1.26

Table 4. Error statistics of reference data minus automatically generated DSMs. Single and interpolated points from breaklines are indicated as Pts and Brklin respectively. Points on bare earth (BE) close to building outlines are indicated as hPts.

error of 1.30 m compared to 7.56 m of SS. In addition, SS exhibited gross errors of over 100 m. The good performance of AIM at MMO is also shown by the smaller mean values compared to SS. The mean values for both methods show that matching generally measures higher than the manual measurements, while for MMO matching results are lower. The fact that the accuracy was less for BE than MMO for Y0624C and W1052C is due to the density of the buildings and the relative short distances of the measured points to neighboring MMO. BE breaklines in these areas have been extracted along discontinuities on the ground, which were close to MMO. Therefore, it is possible that modeling errors were introduced.

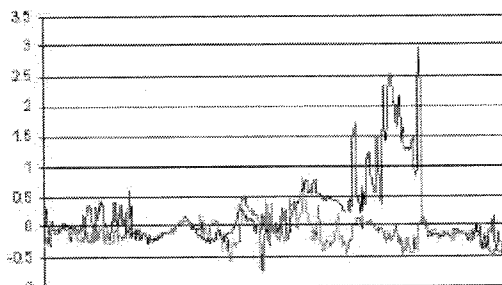


Figure 6. Elevation errors of 900 reference points along a specific breakline on MMO minus automatically derived points with AIM and SS. Larger differences are observed (dark line) for SS compared to AIM.

The example in Fig. 6 shows the elevation differences computed for 900 reference points along a specific breakline defining a roof edge. The differences are smaller for points derived with AIM, compared to SS, because the use of contour points and edges in conjunction with the multi-image matching approach used in AIM, improves the modeling of discontinuities.

4. CONCLUSIONS

In this study, two systems have been evaluated regarding DSM generation using ADS40 imagery acquired over two different test areas. The AIM system showed better performance, compared to SS system, especially along building discontinuities. Accuracy increased, often by factor two or more, and blunders decreased, even in difficult areas, as Yokohama. Various components of the AIM method, especially those that take advantage of the particular characteristics of ADS40, lead to this improved performance. With the exception of breaklines in W1052C, the performance of AIM in the rural areas, even at MMO, was close to the theoretically expected and the accuracy of the manual measurements. However, dense urban areas still pose a problem and need algorithmic improvements but also denser matching and use of more images from neighbouring strips that should have a high overlap. A better configuration of the lines on the focal plane, having e.g. at least two lines with a sufficient base to height ratio in both forward and backward to minimise occlusions, would also be beneficial for urban mapping. Currently, further research in AIM is focused on improving surface modeling, fine-tuning adaptivity of parameters and extending the matching algorithm on Lev0 images with non-adjusted orientation data for tie point extraction in AT.

ACKNOWLEDGEMENTS

This work was performed within the project AIM, a cooperation of ETH Zurich and LGGM. The authors are grateful to Kikuo Tachibana, Tadashi Sasagawa and Hiroyuki Okada, PASCO Corp., for providing the ADS40 dataset of Yokohama. Natasha Vassilieva, IGP-ETH, is especially thanked for measuring the reference data in both Yokohama and Waldkirch datasets. The general support of the ADS40 development team of LGGM and the assistance of Muzzafar Adiguezel and Fernando Shapira, LGGM support team, in data processing are gratefully acknowledged.

REFERENCES

- Baltsavias E., Favey E., Bauder A., Boesch H., Pateraki M., 2001. Digital surface modeling by airborne laser scanning and digital Photogrammetry for glacier monitoring. *Photogrammetric Record*, 17(98): 243-273.
- Bacher U., 1998. Experimental Studies into Automated DTM Generation on the DPW770. In: IAPRS, Vol. 32, Part 4, pp. 35-41.
- Fricker P., 2001. ADS40 – Progress in digital aerial data collection. In: D. Fritsch, R. Spiller (Eds.), *Photogrammetric Week '01*, Wichmann Verlag, Heidelberg, pp. 105 - 116.
- Fritsch D., 1997. Experiences with the Airborne Three-line Photogrammetric Image Acquisition System DPA. In: D. Fritsch, D. Hobbie (Eds.), *Photogrammetric Week '97*, Wichmann Verlag, Heidelberg, pp. 63-74.
- Gwinner K., Hauber E., Hoffmann H., Scholten F., Jaumann R., Neukum G., Puglisi G., Coltelli M., 1999. The HRSC-A Experiment on High Resolution Multispectral Imaging and DEM Generation at the Aeolian Islands. In: 13th Int. Conf. on Applied Geologic Remote Sensing, Vancouver B.C., March 1999, Vol. 1, pp. 560-569.
- Haala N., Fritsch D., Stallmann D., Cramer M., 2000. On the performance of digital airborne pushbroom cameras for photogrammetric data processing - a case study. In: IAPRSSIS, Vol. 33, Part B4/1, pp. 324-331.
- Henricsson, O., 1996. Analysis of Image Structure using Color Attributes and Similarity Relations. Ph.D. Thesis, Report No.59, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland.
- Hinsken L., Miller S., Tempelmann U., Uebbing R., Walker S., 2001. Triangulation of LH Systems' ADS40 imagery using ORIMA GPS/IMU. In: IAPRSSIS, Vol. 34, Part B3/A, pp. 156-162.
- Hoffmann A., van der Vegt J.W., Lehmann F. 2000. Towards automated map updating: Is it feasible with new digital data-acquisition and processing techniques? In: IAPRSSIS, Vol. 33, Part B2, pp. 295-302.
- Leberl W.F., Gruber M., Ponticelli M., Bernoegger S., Perko R., 2003. The UltraCam Large Format Aerial Digital Camera System. In: Proc. ASPRS Annual Conference, (on CD-ROM).
- Neukum G., 1999. The Airborne HRSC-A: Performance Results and Application Potential. In: *Photogrammetric Week '99*, D. Fritsch, D. Hobbie (Eds.), Wichmann, Heidelberg, pp. 83-88.
- Nonin Ph., 2003. Automatic extraction of digital surface models from airborne digital cameras. In: Gruen A., Kahmen H. (Eds.), *Optical 3-D Measurement Techniques VI*, pp. 106-114.
- Pateraki M., Baltsavias E., 2003a. Analysis and performance of the Adaptive Multi-Image matching algorithm for airborne digital sensor ADS40. In: Proc. ASPRS Annual Conference, (on CD-ROM).
- Pateraki M., Baltsavias E., 2003b. Analysis of a DSM generation algorithm for the ADS40 Airborne Pushbroom Sensor. In: Gruen A., Kahmen H. (Eds.), *Optical 3-D Measurement Techniques VI*, pp. 83-91.
- Pateraki M., Baltsavias E., 2004. Surface discontinuity modeling by LSM through patch adaptation and use of edges. In: IAPRS, Vol. 35, Part B3.
- Renouard, L., Lehmann, F., 1999. High Resolution Digital Surface Models and Orthoimages for Telecom Network Planning. In: Fritsch, D., Spiller, R.H. (Eds.), *Photogrammetric Week '99*, Wichmann, Heidelberg, pp. 241-246.
- Sandau R., Braunecker B., Driescher H., Eckardt A., Hilbert S., Hutton J., Kirchhofer W., Lithopoulos E., Reulke R., Wicki S., 2000. Design principles of the LH Systems ADS40 airborne digital sensor. In: IAPRSSIS, Vol. 33, Part B1, pp. 258-265.
- Scholten F., 2000. Digital 3D data acquisition with the High Resolution Stereo Camera – Airborne (HRSC-A). In: IAPRS, Vol. 33, Part B4, pp. 901-908.
- Tempelmann U., Börner A., Chaplin B., Hinsken L., Mykhalevych B., Miller S., Recke U., Reulke R., Uebbing, R., 2000. Photogrammetric Software for the LH Systems ADS40 Airborne Digital Sensor. In: IAPRSSIS, Vol. 33, Part B2, pp. 552-559.
- Tempelmann U., Hinsken L., Recke U., 2003. ADS40 calibration and verification process. In: Gruen A., Kahmen H. (Eds.), *Optical 3-D Measurement Techniques VI*, pp. 48-54.
- Tianen C., Ryosuke S., Murai S., 2003. Development and Calibration of the Airborne Three-Line Scanner (TLS) Imaging System. *PERS*, 69(1): 71-78.
- Wewel F., Scholten F., Neukum G., Albertz J., 1998. Digitale Luftbildaufnahme mit der HRSC – Ein Schritt in die Zukunft der Photogrammetrie. *Photogrammetrie – Fernerkundung Geoinformation*, (6): 337-348.
- Zhang L., Gruen A., 2003. Automatic DSM generation from TLS data. In: Gruen A., Kahmen H. (Eds.), *Optical 3-D Measurement Techniques VI*, pp. 93-105.

ORIENTATION OF SATELLITE AND AIRBORNE IMAGERY FROM MULTI-LINE PUSHBROOM SENSORS WITH A RIGOROUS SENSOR MODEL

Daniela Poli

Institute of Geodesy and Photogrammetry, ETH Zurich, 8093 Zurich, Switzerland- daniela@geod.baug.ethz.ch

Commission I, WG I/5

KEY WORDS: Pushbroom, Sensors, Modelling, Orientation, Self, Calibration

ABSTRACT:

Today CCD linear array scanners are widely used on satellite, airborne and helicopter platforms to provide images with along track stereo viewing. For the orientation of this kind of imagery, models based on the rigorous description of the acquisition geometry, on rational polynomial functions or on affine transformations are used. An overview is presented. Among these approaches, the model developed at the Institute of Geodesy and Photogrammetry (IGP), ETH Zurich, belongs to the class of rigorous models and is applicable to a wide class of pushbroom sensors carried on satellite, airplane and helicopter. The model can be used with single-lens and multi-lens sensors with synchronous and asynchronous stereo acquisition. The sensor position and attitude are modelled with 2nd order piecewise polynomials depending on time. Additional pseudo-observations allow the reduction of the polynomial order from 2 to 1 if the trajectory allows it. In case of sensors carried on aircraft, the observations from GPS and INS instruments are integrated in the piecewise polynomials and are corrected from constant shifts and misalignments between the GPS and INS local systems and the camera one and systematic errors contained in the observations. A self-calibration is also included for the corrections of radial and decentering lens distortions, principal point(s) displacement, focal length(s) variation and CCD line(s) rotation in the focal plane. Using a minimum of 6 Ground Control Points (GCPs) and, additionally, Tie Points (TPs), the external orientation and self-calibration parameters, together with the TPs ground coordinates, are estimated in a least-square adjustment. In order to demonstrate the model flexibility and potentials, different imagery from pushbroom sensors (TLS, EROS-A1, SPOT-5/HRS, ASTER, MOMS-02, MISR) have been oriented. In this paper a summary of the results obtained are presented and discussed.

1. INTRODUCTION

CCD linear array sensors, also called linear scanners, are widely used for the acquisition of images at different ground resolution for photogrammetric mapping and remote sensing applications. They scan the ground surface with an array of CCD elements in pushbroom mode. The image is formed by a side-to-side scanning movement as the platform travels along its path.

CCD linear array sensors for remote sensing and photogrammetric applications are usually mounted on aerial and satellite platforms. Aerial platforms are primarily stable wing aircraft, but also helicopters are used. Acquisition from airborne sensors are often used to collect very detailed images and facilitate the collection of data over virtually any portion of the Earth's surface at any time. In space, the acquisition of images is sometimes conducted from the space shuttle or, more commonly, from satellites for Earth Observation (EO). Because of their orbits, satellites permit repetitive coverage of the Earth's surface on a continuing basis. Cost is often a significant factor in choosing among the various platform options.

The images provided by linear CCD array sensors have very high potentials for photogrammetric mapping at high and low scales.

The triangulation and photogrammetric point determination of pushbroom systems are rather different compared to standard approaches, which are usually applied for full frame imagery, (Haala et al., 2000). Additionally, the use of linear imaging sensors is more difficult with respect to photogrammetric data processing and requires an increased computational effort during the subsequent processing chain including matching, DTM and orthoimage generation.

This work relates to the analysis of the orientation of CCD linear scanners. Different models can be found in literature. The rigorous ones are based on the photogrammetry collinearity equations, which are modified in order to include any external and internal orientation modelling. Other approaches are based on rational polynomial functions, affine models and direct linear transformations. An overview is given in the next section. At the Institute of Geodesy and Photogrammetry (IGP), ETH Zurich, a rigorous sensor model for the georeferencing of imagery acquired by multi-line CCD array sensors, carried on airborne or satellite, has been implemented. The model fulfils the requirement of being as flexible as possible and being adaptable to a wide class of linear array sensors. In fact pushbroom scanners show different geometric characteristics (optical systems, number of CCD lines, scanning mode and stereoscopy) and for each data set specific information are available (ephemeris, GPS/INS observations, calibration, other internal parameters). Therefore the model needs to be dependent on a certain number of parameters that change for each sensor. The results obtained with different satellite sensors will be presented.

2. OVERVIEW OF EXISTING MODELS

For the georeferencing of imagery acquired by pushbroom sensors many different geometric models of varying complexity, rigor and accuracy have been developed, as described in (Fritsch et al., 2000, Hattori et al., 2000 and Dowman et al., 2003). The main approaches include rigorous models, rational

polynomial models, Direct Linear Transformations (DLT) and affine projections.

The rigorous models try to describe the physical properties of the sensors acquisition and are based on collinearity equations, which are extended in order to describe the specific geometry of pushbroom sensors. Some rigorous models are designed for specific sensors, while some others are more general and can be used for different sensors. Few models are designed for both spaceborne and airborne linear scanners.

In case of spaceborne sensors, different approaches have been followed. V. Kratky (Kratky, 1989) developed a software, called SPOTCHECK+, the principle of rigorous photogrammetric bundle formulation is combined with the sensor external orientation modelling. The satellite position is derived from known nominal orbit relations, while the attitude variations are modelled by a simple polynomial model (linear or quadratic). For self calibration two additional parameters are added: the focal length (camera constant) and the principle point correction. The exterior orientation and the additional parameters of the sensor model are determined in a general formulation of the least squares adjustment (Gauss-Helmert model). The use of additional information, e.g. from supplemented data files is not mandatory, but if this information is available it can be used to approximate or pre-set some of the unknown parameters. This model has been used for the orientation of SPOT-2 (Baltasvias et al., 1992), MOMS-02/D2 (Baltasvias et al., 1996), MOMS-02/Priroda (Poli et al., 2000), Landsat TM and JERS-1 (Fritsch et al., 2000) scenes. An advantage of this software is that it can easily integrate new pushbroom instruments, if the corresponding orbit and sensors parameters are known. The model was also investigated and extended in (Fritsch et al., 2000).

The principle of orientation images was used at DLR for the geometric in-flight calibration and orientation of MOMS-2P imagery (Kornus et al., 1999a, Kornus et al., 1999b). This method is based on extended collinearity equations (Ebner et al., 1992). The exterior orientation parameters are determined in the so-called orientation images and between the orientation images the parameters of an arbitrary scan line are interpolated using Lagrange polynomials. For the modelling of the interior orientation for each CCD array five parameters are introduced. All unknown parameters are estimated in a bundle block adjustment using threefold stereo imagery. For the determination of the unknown parameters a large number of tie points which are automatically measured is required.

In the group of Prof. Ebner at TU Munich a mathematical model of photogrammetric point determination for airborne and spaceborne three-line scanners has been developed and tested on MOMS-02/D2 and P2 (Ebner et al., 1992), MEOSS (Ohlhof, 1995), HRSC and WAOSS (Ohlhof et al., 1994) sensors. The model is based on a polynomial approach in case of airborne imagery, whereas orbital constraints are utilised in case of spaceborne imagery. In the airborne case the exterior orientation parameters are estimated only for some so-called orientation points, which are introduced at certain time intervals, e.g. every 100th readout cycle. In between, the external orientation parameters are expressed as polynomial functions (e.g. Lagrange polynomials) of the parameters at the neighboring orientation points. For preprocessed position and attitude data, e.g. acquired by differential GPS and INS, observation equations are formulated. Systematic errors of the position and attitude observations are modelled through additional strip- or block-invariant parameters. By limitation to constant and time-dependent linear terms, which describe the main effects, 12 additional parameters, namely a bias and a drift parameter for each exterior orientation parameter, are introduced. For the

satellite case, the spacecraft's epoch state vector is estimated with the assumption that all scanner positions lie along an orbit trajectory. Due to the lack of a dynamic model describing the camera's attitude behaviour during an imaging sequence, for the spacecraft's attitude the concept of orientation points is maintained.

The University College London (UCL) suggested a dynamic orbital parameter model (Gugan et al., 1988). The satellite movement along the path is described by two orbital parameters (true anomaly and the right ascension of the ascending node), that are modelled with linear angular changes with time and included in the collinearity equations. The attitude variations are modelled by drift rates. This model was successfully applied for SPOT level 1A and 1B (O'Neill et al., 1991), MOMS-02 and IRS-1C (Valadan Zoj et al., 1999) imagery. In (Dowman et al., 2003) this approach was investigated and extended for the development of a general sensor model for along-track pushbroom sensors.

The IPI Institute in Hannover the program system BLUH/BLASPO is used for the adjustment of satellite line scanner images (Jacobsen, 1994). Just the general information about the satellite orbit together with the view directions in-track and across-track are required. Systematic effects caused by low frequency motions are handled by self-calibration with additional parameters. In this model the unknown parameters for each image are 14, that is, 6 exterior orientation parameters for the uniform motion and 8 additional parameters for the difference between the approximate uniform movement and the reality. This program seems very flexible, because it has been successfully used for the orientation of MOMS-02 (B, y, ksalih et al., 2000), SPOT, KFA1000, KVR1000 and IRS-1C (Jacobsen et al., 1998), DPA, IKONOS and Quickbird (Jacobsen et al., 2003) and SPOT-5/HRS (Jacobsen et al., 2003).

In (Westin, 1990) the orbital model used is simpler than in the previous models. A circular orbit instead of an elliptical one is used with sufficient accuracy. Using data from SPOT ephemeris data seven unknown parameters need to be computed for each SPOT image.

Among specific models developed for one sensor, the procedure used at JPL, Pasadena, for the orientation of MISR sensors reproduces the image acquisition using a large number of reference systems and specific MISR parameters measured during laboratory calibration. The external orientation parameters are calculated from precise ephemeris (Jovanovic et al., 1998).

An alternative image orientation approach widely used is the Rational Function Model (RFM), or Rational Polynomial Coefficients (RPC), which provide a means of extracting 2D (3D) information from single (stereo) satellite imagery without explicit reference to either a camera model or satellite ephemeris information. The RFMs describe the relationship between image (line, sample) and object space (typically latitude, longitude and height) coordinates and viceversa through quotients of polynomials, usually of 3rd order (Fraser et al., 2001). In (Grodecki et al., 2003) a block adjustment with Rational Polynomial Coefficients (RPC) is proposed and applied for the orientation of high-resolution satellite images, such as IKONOS. The same model has been implemented at IGP, ETH Zurich, for the orientation of SPOT-5/HRS and SPOT-5/HRG stereo images (Poli et al., 2004).

Other approaches for satellite imagery acquired by CCD linear array scanners are based on affine transformations.

Prof. Okamoto (Okamoto, 1981) proposed the affine transformation to overcome problems due to the very narrow

field of the sensor view. The imaging geometry is converted from the original perspective imagery into an affine projection. Later the method was applied to SPOT stereo scenes of level 1 and 2 (Okamoto et al., 1998). The theories and procedures of affine-based orientation for satellite line-scanner imagery have been integrated and used for the orientation of SPOT, MOMS-2/P (Hattori et al., 2000) and IKONOS (Fraser et al., 2001) scenes. Under this approach an initial transformation of the image from a perspective to an affine projection is first performed, then a linear transformation from image to object space follows, according to the particular affine model adopted. The assumption is that the satellite travels in a straight path at uniform velocity within the model space. The model utilises the Gauss-Krueger projection plane and ellipsoidal heights as a reference system, therefore height errors due to Earth curvature must be compensated. The results demonstrated that 2D and 3D geopositioning to sub-pixel accuracy can be achieved (Fraser et al., 2001).

(Gupta et al., 1997) proposed a simple non-iterative model based on the concept of fundamental matrix for the description of the relative orientation between two stereo scenes. The model was applied on SPOT across-track stereo scenes. The unknown parameters for each pair are: the sensor position and attitude of one scene at time 0, the velocity of the camera, the focal length and the parallax in across-track direction.

The Direct Linear Transformation (DLT) approach has also been investigated. The solution is based only on ground control points and does not require parameters of the interior orientation and ephemeris information. The DLT approach was suggested for the geometric modelling of SPOT imagery (El Manadili et al., 1996) and applied to IRS-1C images (Savopol et al., 1998). In (Wang, 1999) it was improved by adding corrections for self calibration.

In general, the approaches based on 2D and 3D empirical models, as those presented, are advantageous if the rigorous sensor model or the parameters of the acquisition system are not available.

In case of pushbroom sensors carried on airplane or helicopter GPS and INS observations are indispensable, because the airborne trajectories are not predictable. Anyway, the original position and attitude measurements are not enough accurate for high-precision positioning and require a correction.

The IGP at ETH Zurich (Gruen et al., 2002a) investigated three different approaches for the external orientation modelling of the Three-Line Scanner (TLS) developed by Starlabo Corporation: the Direct Georeferencing, in which the translation displacement vector between the GPS and camera systems is estimated for the correction of GPS observations, the Lagrange Polynomials, as used in (Ebner et al., 1992) for spaceborne sensors and the Piecewise Polynomials, where the sensor attitude and position functions are divided in sections and modelled with 1st and 2nd order respectively, with constraints on their continuity. The sensor self-calibration has also been integrated in the processing chain. Further investigations on the models performances are in progress.

In the LH-Systems photogrammetric software for the ADS40 processing, a triangulation is applied for the compensation of systematic effects in the GPS/IMU observations (Tempelmann et al., 2000). These effects include the misalignment between IMU and the camera axes and the datum differences between GPS/IMU and the ground coordinates system. For the orientation of each sensor line the concept of orientation fixes is used. The external orientation values between two orientation fixes are determined by interpolation using the IMU/GPS observations.

From the analysis of the above literature we can see that nowadays both approaches based on rigorous and non rigorous models are widely used. In case of rigorous models the main research interests are the sensor external and internal orientation modelling. The external orientation parameters are often estimated for suitable so-called orientation lines and interpolated for any other lines. A self-calibration process is recommended, as least to model focal length variation and first order lens distortions. In order to avoid over-parameterisation the correlation between the parameters must be investigated and tests on the parameters significance and determinability are required. Moreover it is recommended to take advantage of additional information for the external orientation estimation (orbital elements and ephemeris for spaceborne sensors, GPS and INS measurements for spaceborne and airborne sensors). Few models can be applied for both airborne and spaceborne sensors.

The orientation methods based on rational polynomials functions, affine projections and DLT transformations are mostly used for high-resolution satellite imagery. They can be a possible alternative to rigorous model when the calibration data (calibrated focal length, principal point coordinates, lens distortions) are not released by the images providers or when the sensor position and attitude are not available with sufficient precision (Vozikis et al., 2003).

3. SENSOR MODEL DESCRIPTION

A rigorous sensor model for the georeferencing of a wide class of linear CCD array sensors has been developed at IGP and already applied to different linear scanners carried on satellite and aircraft (Poli, 2003). The photogrammetric collinearity equations describe the perspective geometry in each image line. The sensor position and attitude are modelled with piecewise 2nd order polynomial functions depending on time. The platform trajectory is divided into segments according to the number and distribution of available Ground Control Points (GCPs) and Tie Points (TPs) and for each segment the sensor position and attitude are modelled by 2nd order polynomials. At the points of conjunction between adjacent segments constraints on the zero, first and second order continuity are imposed on the trajectory functions. Additional pseudo-observations can fix some or all parameters to suitable values. For example, if the 2nd order parameters are fixed to zero, the polynomial degree is reduced to 1 (linear functions). This option allows the modelling of the sensor position and attitude in each segment with 2nd or 1st order polynomials, according to the characteristics of the trajectory of the current case study. In case of sensors carried on aircraft, additional GPS and INS observations can be included in the model (Poli, 2002).

The sensor model includes also a self-calibration, which is required for the correction of the systematic errors due to principal point displacement (d_x, d_y), focal length variation (d_c), radial symmetric (k_1, k_2) and decentering lens distortion (p_1, p_2), scale variation in CCD line direction (s_y) and the CCD line rotation in the focal plane (θ). The model can be applied to single lens sensors with synchronous (i.e. TLS, ADS40, WAOSS) and asynchronous (EROS-A1) along-track stereo capability and to multi-lens sensors (i.e. SPOT-5/HRS, ASTER, MOMS-02, MISR).

The functions modelling the external and the internal orientation are integrated into the collinearity equations, resulting in an indirect georeferencing model. Due to their non-linearity, the complete equations are linearized according to the

first-order Taylor decomposition with respect to the unknown parameters. The resulting system is solved with a least square adjustment. As result the coefficients of the polynomials modelling the external orientation, the self-calibration parameters and the coordinates of the tie points are estimated. Statistics on the system s

In case of satellite imagery, the available ephemeris (usually sensor position and velocity at fixed intervals) are used to generate the approximate values for the parameters modelling the sensor external orientation (position and attitude). The required geometric parameters (focal length(s), viewing angles, number and size of CCD elements in each array) are usually available from the imagery provider or from literature. The reference frame used in the adjustment is the fixed Earth-centered Cartesian system, also called ECR.

In case of airborne imagery, the GPS and INS observations are included in the piecewise polynomial equations. The polynomial coefficients model the shift and offset between the GPS and INS local systems and the camera system (centred in the lens perspective centre) and 1st and 2nd order systematic errors contained in the observations.

For the orientation of the pushbroom imagery preliminary tests are made with these objectives:

- determination of best degree for piecewise polynomials and best GCPs configuration, by solving the adjustment without self-calibration, with quadratic functions modelling the external orientation and varying the number of segments and GCPs configuration;
- external orientation modelling with linear and quadratic functions, using the best GCPs configuration and best trajectory segments, without self-calibration;
- self-calibration with best external orientation modelling configuration.

The choice of the unknown self-calibration parameters to include in the modelling is based on the analysis of the cross-correlation between the self-calibration parameters, the external orientation parameters and the ground coordinates of the TPs. Statistics on the adjustment performance and RMS values for the GCPs and Check Points (CPs) are considered for the quality assessment.

4. ORIENTATION OF SATELLITE IMAGES

The model has been applied for the orientation of satellite and airborne images with different acquisition geometry (one-lens and multi-lens optical systems, synchronous and asynchronous acquisition) and ground resolution. As satellite applications concern, in (Poli, 2003), (Giulio Tonolo et al., 2003) and (Poli et al., 2004) the results obtained by the orientation of MOMS-02/P, MISR, EROS-A1 and SPOT-5/HRS are presented, while in (Poli, 2002) the tests carried on the Three Line Sensor (TLS), carried on helicopter, are reported. In the following paragraphs the latest results obtained from SPOT-5/HRS and ASTER are summarised.

4.1 SPOT-5/HRS

Within the HRS-SAP Initiative (Baudoin et al., 2004), a DEM was generated from two stereo images acquired by the High Resolution Stereoscapy (HRS) sensor carried on the newest satellite of SPOT constellation. The sensor model was applied in order to orient the stereopair and estimate the ground coordinates of the CPs. The available ephemeris (sensor position and velocity) were used to generate the approximate

values for the parameters modeling the sensor external orientation (position and attitude) in fixed Earth-centred geocentric Cartesian system. From the available 41 object points, a group of them was used as GCPs and the remaining as CPs. The best results in terms of RMSE in the CPs were obtained by modelling the external orientation with two 2nd order polynomials and with self-calibration. The self-calibration parameters that mostly influenced the model were k_1 , k_2 , p_2 and s_y for both lenses. The other self-calibration parameters could not be estimated due to the high correlation with the TP coordinates and external orientation parameters. By changing the number of GCPs and CPs, the RMSE were always less than 1 pixel. For a more detailed description of the data and processing, see (Poli et al., 2004).

4.2 ASTER

ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) is a high-resolution, multispectral/hyperspectral imaging instrument which is flying on Terra, a satellite launched in December 1999 as part of NASA's Earth Observing System (EOS). ASTER takes data in 14 spectral bands within the Visible and Near Infrared (VNIR), the Shortwave Infrared (SWIR) and the Thermal Infrared (TIR) at ground resolution of 15m, 30m and 90m respectively.

The generation of DEMs is possible with the VNIR instrument, that provided stereo images in along-trak direction. VNIR consists of two independent telescopes operating in band 3 (0.76-0.87 μ m), viewing nadir (channel 3N) and backward (channel 3B, 27.6 $^\circ$ off-nadir) with respect to the spacecraft trajectory. The telescopes scan the ground in pushbroom mode using arrays of CCDs with size 7 μ m x 7 μ m. The number of CCD elements in each array is 4100 for channel 3N and 5000 for channel 3B (4100 are active). The two telescopes allow simultaneous stereo imaging with a 64 sec time delay between the scanning of the same ground target and a B/H of 0.6. Each scene is 4100x4200 pixels large and cover an area of about 60km x 60km. Several types of ASTER data are available at different processing levels. For our purposes, the level 1A is used, because at this level the images are not geometrically processed.

The scene used in this work covers the valley of Shaxi, in the South-East part of China. The images were kindly provided by the Institute for Spatial and Landscape Planning, ETH Zurich, who is involved in a World Monuments Fund project for the economic development of the Shaxi valley. The scenes were acquired on 23rd November 2000 in the morning.

For the orientation of the channels 3N and 3B from the ASTER scene, six GCPs that have been used. The ground coordinates of these points were available from in-situ GPS measurements or measured in local maps. As the sensor external orientation concerns, the ASTER scene metadata file contained the satellite position and velocity in ECR (fixed Earth-centred Cartesian coordinate system) every 400 image lines. These data were used to calculate the satellite attitude at the observations times and calculate the initial approximations for the polynomial coefficients modelling the sensor external orientation. Due to the limited number of object points, only the RMSE for the GCPs have been calculated: 8.1m in X, 8.4m in Y and 10.4m in Z. After the production of five pyramid images, interest points were matched and found progressively in all pyramid levels starting from the low-density features on the images with the lowest resolution. After the process, 320,000 points were successfully matched. The failed matches were mostly in correspondence of areas covered by clouds, due to the cloud movement between the nadir and backward images acquisition.

Using the software Geomatic Studio v.4.1 by Raindrop, the DEM was cleaned by points on clouds and a DSM was generated. In Figure 2 the DSM in the Shaxi valley is shown.

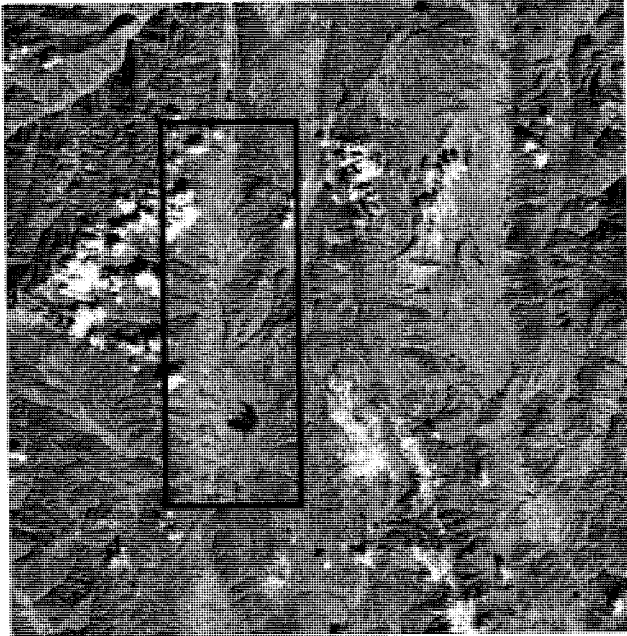


Figure 1. False-colour ASTER-VNIR scene (bands 1, 2, 3N). The Shaxi valley is located in the black box.

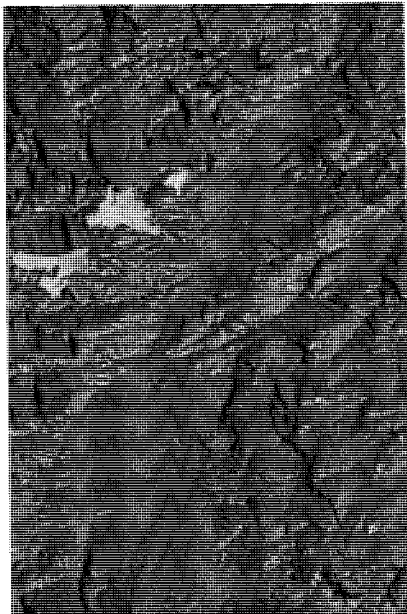


Figure 2. DSM in the Shaxi valley. The holes are due to deleted points on clouds.

5. CONCLUSIONS

In this paper an overview on the alternative models used for the orientation of pushbroom imagery is given. Some of these approaches rigorously describe the physical acquisition geometry and are based on the collinearity equations. Other methods describe the relationship between image and ground coordinates through rational polynomial functions or affine transformations, don't depend on the specific \tilde{O} and are mostly used when the physical sensor model or the required geometric

parameters are not available. The model developed at IGP belongs to the class of rigorous models and has the advantage to be usable with a wide class of pushbroom sensors. The model has been tested on different sensors with different geometric characteristics. In the paper the application on SPOT-5/HRS and ASTER have been briefly presented. In all cases, RMSE in the CPs up to 1 pixel have been reached.

ACKNOWLEDGMENTS

The author would like to thank the Institute for Spatial and Landscape Planning, ETH Zurich for providing the ASTER images and the object points location.

REFERENCES

- Baltsavias E. P., Stallmann D., 1992. Metric Information Extraction from SPOT Images and the Role of Polynomial Mapping Functions. IAPRS, Vol. 29, Part B4, Washington, pp. 358 - 364.
- Baltsavias, E. P., Stallmann, D., 1996. Geometric potential of MOMS-02/D2 data for point positioning, DTM and orthoimage. IAPRS, Vol. 31, Part B4, Vienna, pp. 110-116.
- B,y,ksalih G., Jacobsen K., 2000. Geometric Aspects of MOMS-2P Three-Line Imagery for Mapping Applications. Annual Meeting of the Remote Sensing Society, Lancaster, UK.
- Dowman, I. J., Michalis, P., 2003. Generic rigorous model for along track stereo satellite sensors. ISPRS Workshop iHigh Resolution Mapping from Space 2003i, Hannover, 4-6 October (on CDROM.)
- Ebner, H., Kornus, W., Ohlhof, T., 1992. A simulation study on point determination for the MOMS-02/D2 space project using an extended functional model. IAPRS, Vol. 29, Part B4, Washington D.C., pp. 458-464.
- El-Manadili Y., Novak, K., 1996. Precision rectification of SPOT imagery using the direct linear transformation model. PE&RS, 62(1), pp. 67-72.
- Fraser, C.S., Hanley, H.B., Yamakawa, T., 2001. Sub-metre geopositioning with IKONOS GEO imagery. ISPRS Workshop iHigh-resolution Mapping from Space 2001i, Hanover, September 19-21 (on CDROM).
- Fritsch, D., Stallmann, D., 2000. Rigorous photogrammetric modelling processing of high-resolution satellite imagery. IAPRS, Vol.33, Part B1, Amsterdam, pp.313-321.
- Gruen, A., Zhang L., 2002b. Sensor modeling for aerial mobile mapping with Three-Line-Scanner (TLS) imagery. IAPRS, Vol. 34, Part 2, Xi'an, pp. 139-146.
- Giulio Tonolo, F., Poli, D., 2003. Georeferencing of EROS-A1 high-resolution images with rigorous and rational function model. ISPRS Workshop iHigh resolution mapping from Space 2003i, Hannover, 4-6 October (on CDROM).
- Gruen, A., Zhang, L., 2002. Sensor modeling for aerial mapping with Three-Line Scanner (TLS) Imagery. IAPRS, Vol. 32, Part B2, Xian (on CDROM).
- Grodecki, J., and Dial, G., 2003. Block Adjustment of High-Resolution Satellite Images Described by Rational Functions. PE&RS, Vol. 69, No. 1, pp. 59-70.

- Gugan, D. J., Dowman, I. J., 1988. Accuracy and completeness of topographic mapping from SPOT imagery. *Photogrammetric Record*, Vol. 12, No. 72, pp. 787-796.
- Gupta, R., Hartley, R. I., 1997. Linear Pushbroom Cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No 9, September 1997, pp. 963-975.
- Gupta, R., Hartley, R. I., 1997. Linear Pushbroom Cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No 9, September 1997, pp. 963-975.
- Haala, N., Fritsch, D., Stallmann, D., Cramer, M., 2000. On the performance of digital airborne pushbroom cameras for photogrammetric data processing: a case study. *IAPRS*, Vol. 33, Part B4/1, Amsterdam, pp. 320-331.
- Hattori, S., Ono, T., Fraser, C.S. & Hasegawa, H., 2000. Orientation of high-resolution satellite images based on affine projection. *IAPRS*, Vol. 33, Part 1, Amsterdam, pp. 359-366.
- Jacobsen, K., 1994. Comparison of mapping with MOMS and SPOT images. *IAPRS*, Vol. 30, Part B4, pp. 225-230.
- Jacobsen, K., Konecny, G., Wegman, H., 1998. High Resolution Sensor Test Comparison with SPOT, KFA1000, KVR1000, IRS-C and DPA in Lower Saxony. *IAPRS*, Vol.32, Part 4, Stuttgart, pp. 260-269.
- Jacobsen, K., Konecny, G., Wegmann, H., 1998. Analysis of SPOT HRS stereo data. *ISPRS Workshop iHigh resolution mapping from Space 2003i*, Hannover, 4-6 October (on CDROM).
- Jacobsen, K., Passini, R., 2003. Accuracy of digital orthophotos from high resolution space imagery. *ISPRS Workshop iHigh resolution mapping from Space 2003i*, Hannover, 4-6 October (on CDROM).
- Jovanovic, V. M., Smyth, M. M., Zong, J., Ando, R., Bothwell, G. W., 1998. MISR Photogrammetric Data Reduction for Geophysical Retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 36, No. 4, July 1998, pp. 1290-1300.
- Konecny, G., Lohmann, P., Engel, H., Kruck, E., 1987. Evaluation of SPOT imagery on analytical photogrammetric instrument. *PE&RS*, Vol. 53, No. 9, pp.1223- 1230.
- Kornus, W., Lehner, M., Schoeder, M., 1999a. Photogrammetric block adjustment using MOMS-2P imagery of the three intersecting stereo-strips. *ISPRS Workshop "Integrated Sensor Calibration and Orientation"*, Portland, Maine, USA, June 16-17, 1999.
- Kornus W., Lehner M., Schroeder M., 1999b. Geometric inflight calibration by block adjustment using MOMS 2P imagery of three intersecting stereo strips. *ISPRS Workshop iSensors and Mapping from Spacei*, Hanover, September 27-30 (on CDROM).
- Kratky, V., 1989. Rigorous photogrammetric processing of SPOT images at CCM Canada. *ISPRS Journal of Photogrammetry and Remote Sensing*, No. 44, pp. 53-71.
- Ohlhof, T., Montenbruck, O., Gill, E., 1994. A new approach for combined bundle block adjustment and orbit determination based on Mars-94 three-line scanner imagery and radio tracking data. *IAPRS*, Vol., 30, Part B3/2, Munich, pp. 630-639.
- Ohlhof, T., 1995. Block triangulation using three-line images. *Proceedings of iPhotogrammetric Week 1995i*, Wichmann Verlag ed., Stuttgart, pp. 197-206.
- Okamoto, A., 1981. Orientation and construction of models, Part III: Mathematical basis of the orientation problem of one-dimensional central-perspective photographs. *PE&RS*, Vol. 47, No. 12, December, pp. 1739-1752.
- Okamoto, A., Fraser, C., Hattori, S., Hasegawa, H., Ono, T., 1998. An alternative approach to the triangulation of SPOT imagery. *IAPRS*, Vol. 32, Part 4, pp. 457-462.
- O'Neil, M., Dowman, I., 1991. A new camera model for the orientation of the SPOT data and its application to the OEEPE test of triangulation of SPOT data. *OEEPE Publication*, Issue 26, pp. 153-163.
- Poli, D., Seiz, G., Baltasvias, E. P., 2000. Cloud-top height estimation from satellite stereopairs for weather forecasting and climate change analysis. *IAPRS*, Vol. 33, Part B7/3, Amsterdam, pp.1162-1169.
- Poli, D., 2003. Georeferencing of MOMS-02 and MISR stereoimages with strict sensor model. *ISPRS Workshop iHigh resolution mapping from Space 2003i*, Hannover, 4-6 October (on CDROM).
- Poli, D., Zhang, L., Gruen, A., 2004. SPOT-5/HRS stereo images orientation and automated DSM generation. *IAPRS* Vol. 34, Part B1 (to be published).
- Savopol, F., Armenakis, C., 1998. Modelling of the IRS-1C satellite PAN imagery using the DLT approach. *IAPRS*, Vol. 32, Part 4, Stuttgart, pp. 511-514.
- Tempelmann, U., Boerner, A., Chaplin, B., Hinsken, L., Mykhalevych, B., Miller, S., Recke, U., Reulke, R., Uebbing, R., 2000. Photogrammetric software for the LH systems ADS40. *IAPRS*, Vol. 33, Part B2, Amsterdam, pp. 552-559.
- Valadan Zoj, M. J. V., Foomani, M. J., 1999. Mathematical modelling and geometric accuracy testing of IRS-1C stereo-pairs. *ISPRS Workshop iSensors and Mapping from Space 1999i*, Hannover (on CDROM).
- Vozikis, G., Fraser, C., Jansa, J., 2003. Alternative sensor orientation models for high resolution satellite imagery. Band 12 "Publikationen der Deutschen Gesellschaft f,r Photogrammetrie, Fernerkundung und Geoinformation", Bochum (2003), pp. 179- 186.
- Wang Y., 1999. Automated triangulation of linear scanner imagery, 1999. *Verffentlichungen des Instituts f,r Photogrammetrie und Ingenieurvermessungen, Universitf, Hannover*, Nr. 18. *ISPRS Workshop iSensors and Mapping from Spacei*, Hanover, September 27-30 (on CDROM).
- Westin, T., 1990. Precision rectification of SPOT imagery. *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 2, February 1990, pp. 247-253.

SPOT-5/HRS STEREO IMAGES ORIENTATION AND AUTOMATED DSM GENERATION

Daniela Poli, Zhang Li, Armin Gruen

Institute of Geodesy and Photogrammetry, ETH Zurich, 8093 Zurich, Switzerland

TS HRS(2) - DEM Generation from SPOT-5 HRS Data

KEY WORDS: SPOT, Pushbroom, Sensor model, Orientation, Matching, DEM

ABSTRACT

HRS (High Resolution Sensor), carried on SPOT-5, is the first high-resolution sensor on the SPOT constellation that enables the acquisition of stereo images in pushbroom mode from two different directions along the trajectory. The Institute of Geodesy and Photogrammetry (IGP) participated as Co-Investor in the ISPRS-CNES initiative for the investigation on DEM generation from SPOT-5/HRS stereoscenes. This paper describes the work carried out at IGP on a stereopair acquired on 1st October 2002 over parts of Bavaria and Austria.

For orienting HRS imagery, two alternative approaches have been used: a rigorous sensor model and a rational function model. Both algorithms have been implemented at our Institute. The rigorous sensor model is based on the classical collinearity equations, which are extended by the sensor external orientation modeling with 2nd order piecewise polynomials depending on time and by self-calibration parameters. Using well distributed Ground Control Points (GCPs), the unknown internal and external parameters are estimated with a least squares solution. The alternative method, independent from the camera model, does not describe the physical imaging process, but uses rational polynomials to relate image and ground coordinates to each other. This algorithm consists of two steps: 1) calculation of Rational Polynomial Coefficients (RPC) for each image with a least-squares using the geometric information contained in the metadata files; 2) block adjustment with the computed RPC model using GCPs. Both orientation methods gave RMS errors in Check Points (CPs) in the range of 1 pixel in all coordinate directions.

Using the triangulated orientation elements, the DSM was extracted with algorithms and software packages for CCD linear sensors developed at IGP. After the creation of image pyramids, the matches of three kinds of features (feature points, grid points and edges) on the original images are found progressively in the next levels starting from the low-density features on the images with the lowest resolution. A triangular irregular network based DSM is constructed from the matched points on each level of the pyramid and used in turn in the subsequent pyramid level for the approximations and adaptive computation of the matching parameters. A modified Multi Photo Geometrically Constrained Matching algorithm is employed in order to achieve sub-pixel accuracy for all the matched features. The extracted DSM was compared to the reference DSMs obtained from laser data and map contours at different spacings (5x5 m, 25x25 m and 50x50 m) provided by DLR in Oberpfaffenhofen using both terrain height and orthogonal distances. The results show RMS values between one and two pixels on the average and a systematic error mainly due to the presence of trees. After a manual removal of the main areas covered by trees in the reference DSMs sites those errors have been removed. The final results show a mean error in the range of 1-5 meters.

1. INTRODUCTION

SPOT-5 was launched on 4th May, 2002 by Arianespace from the Kourou Space Centre in French Guyana. After completing two months of in-orbit tests it became fully operational in July 2002.

SPOT-5 belongs to the SPOT (Satellite Pour l'Observation de la Terre) constellation developed by CNES (Centre National D'Etudes Spatiales). The constellation consists of 3 operational satellites (SPOT-2, SPOT-4 and SPOT-5) flying along a near-polar, near-circular and Sun-synchronous orbit at a mean altitude of 832 km, an inclination of 98.7 degrees and a mean revolution period equal to 101.4 minutes. The SPOT satellites orbit the same ground track every 26 days with a nominal cycle of 369 revolutions and cross the equator from North to South at 10:30 a.m. mean local solar time (Gleyzes et al., 2003).

Within the constellation, SPOT-5 is the most innovative satellite. The new HRG (High Resolution Geometry) instruments, derived from the HRVIR instrument on SPOT-4 offer high resolution in across-track direction with up to 2.5m resolution in panchromatic mode. Moreover the new HRS (High Resolution Sensor) allows the acquisition of stereo images in along-track direction, using two telescopes pointing about 20 degrees forward and backward (Gleyzes et al., 2003).

Other payload packages include the same Vegetation instrument as on SPOT-4, and the DORIS instrument, for greater orbital accuracy.

The Institute of Geodesy and Photogrammetry (IGP) of ETH Zurich joined the HRS Scientific Assessment Program (HRS-SAP), organised by CNES and ISPRS. This initiative, announced in Denver in 2002 at the ISPRS Commission I Symposium, has the aim to investigate the potential of SPOT-5/HRS sensor for DEM generation in order to help CNES to improve its future Earth Observation systems and all users to better know and trust the accuracy and quality of the HRS instrument and the derived DEM (Baudoin et al., 2003).

IGP joined the Initiative as Co-Investigator, that is, it processed the data provided by one the Principal Investigators, generated two DEMs with two different orientation methods, compared them to the reference DEMs and produced a quality report.

In this paper the work carried out at IGP within HRS-SAP is reported. After the description of the available data, the processing algorithms applied for images orientation, matching and DEM generation are presented. The results obtained after the comparison between the generated DEMs with the reference ones are reported and analysed. Final comments will conclude the paper.

2. DATA

Within the Initiative, CNES and DLR Oberpfaffenhofen provided the data set number 9 (Chiemsee), consisting of:

- two stereo images from SPOT5-HRS sensor with corresponding metadata files;
- the description of the exact position of 81 object points in Germany, measured with surveying methods;
- reference DEMs produced by Laser data and conventional photogrammetric and geodetic methods.

In the next paragraphs the main data characteristics are reported.

2.1 SPOT-5/HRS scenes

The two stereo images were acquired on 1st October 2002 in the morning from 10:15 to 10:18 (forward) and from 10:18 to 10:21 (backward) over an area of approximately 120x60 km² in Bavaria and Austria.

Each image is 12000 x 12000 pixel large, with a ground resolution of 10m across and 5 m along the flight (parallax) direction. The scenes were acquired in panchromatic mode in stereo viewing along the flight direction with a base over height ratio of 0.8. The two telescopes contained in the HRS instrument scan the ground with off-nadir angles of +20 degrees (forward image) and -20 degrees (backward image). Each telescope has a 580 mm focal length and a focal plane with a CCD line of 12000 pixels, 6.5 μm size. The main sensor characteristics are reported in Table 1.

The scenes cover an area with flat, hilly and mountainous (Alps) terrain, agriculture areas, towns, rivers and lakes. The height ranges between 400 m and 2000 m. Clouds are absent (Figure 2).

The metadata files contain information on the acquisition time and image location, ephemeris (sensor position and velocity from GPS at 30 seconds time interval, attitude and angular speeds from star trackers and gyros at 12.5 seconds interval), sensor geometric (detectors looking angles) and radiometric calibration. For a detailed description of the metadata file see DIMAP site. The on-board determination of the satellite position and the absolute dating are supplied by DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite).

DORIS is a one-way microwave tracking system developed for precise orbit determination (1m RMSE) by GRGS (Groupe de Recherches de Géodésie Spatiale) and IGN (Institut Géographique National). The concept is based on a ground segment (of globally positioned tracking stations) and a space segment (i.e. DORIS as a passenger payload in a satellite consisting of a receiver, an ultra-stable oscillator and an antenna). There is also a control centre as part of the ground segment, located at CNES (Gleyzes et al., 2003). The onboard receiver measures the Doppler shift of uplink beacons in two frequencies (f1 = 2036.25 MHz, f2 = 401.25 MHz), which are transmitted continuously by the DORIS ground network of stations. One measurement is used to determine the radial velocity between spacecraft and beacon, the other to eliminate errors due to ionosphere propagation delays.

Table 1. SPOT5-HRS characteristics (Source CNES).

Mass	90 kg
Power	128W
Dimensions	1 x 1.3 x 0.4 m
Field of view	8°
Focal length	0.580 m
Detectors per line	12,000
Detector pitch	6.5 μm
Integration time per line	0.752 ms
Off-nadir angles:	
-forward	20°
-backward	-20°
Spectral range (PAN)	0.49 μm - 0.69 μm
Ground sample distance:	
-across track	10 m
-along track	5 m
Modulation transfer	> 0.25 function
Signal-to-noise ratio	> 120

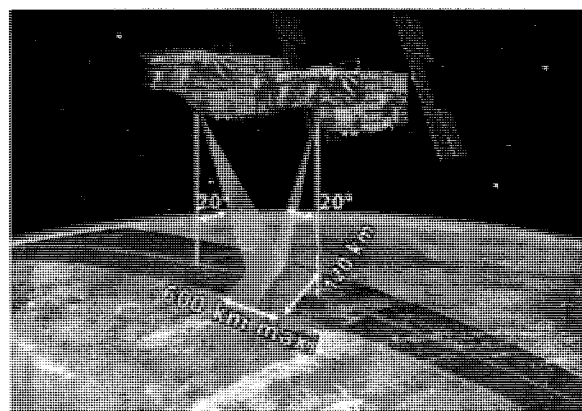


Figure 1. SPOT-5/HRS along-track image acquisition (Source: CNES).

2.2 Reference DEMs

The reference DEMs provided by DLR Oberpfaffenhofen are:

- 4 DEMs in southern Bavaria (Prien, Gars, Peterskirchen, Taching) created from Laser scanner data with a point spacing of 5 meters and an overall size of about 5 km x 5 km. The height accuracy is better than 0.5 m;
- 1 DEM (area of Inzell, total: 10 km x 10 km, 25 m spacing) partly derived from laser scanner data (northern part, height accuracy better than 0.5 m) and partly derived from contour lines 1:10 000 (southern part, height accuracy of about 5 m);
- A large coarse DEM (area of Vilsbiburg, 50 km x 30 km) with 50 m spacing and a height accuracy of about 2 meters, derived by conventional photogrammetric and geodetic methods.

In Table 2 the principal characteristics of the reference DEMs are summarised.

Table 2. Main characteristics of reference DEMs.

DEM	Location	Terrain characteristics	DEM Spacing (m)	Source	DEM Size (kmxkm)	Height Accuracy (m)
1	Prien	Smooth, weakly inclined	5x5	Laser Scanner	5x5	0.5
2	Gars	Smooth, weakly inclined	5x5	Laser Scanner	5x5	0.5
3	Peterskirchen	Smooth, weakly inclined	5x5	Laser Scanner	5x5	0.5
4	Taching	Smooth, weakly inclined	5x5	Laser Scanner	5x5	0.5
5-1	Inzell-North	Rough, strongly inclined	25x25	Laser Scanner	10x1.3	0.5
5-2	Inzell-South	Rolling, strongly inclined	25x25	Contour lines	10x7.7	5.0
6	Vilsbiburg	Rough, weakly inclined	50x50	Photogrammetry	50x30	2.0

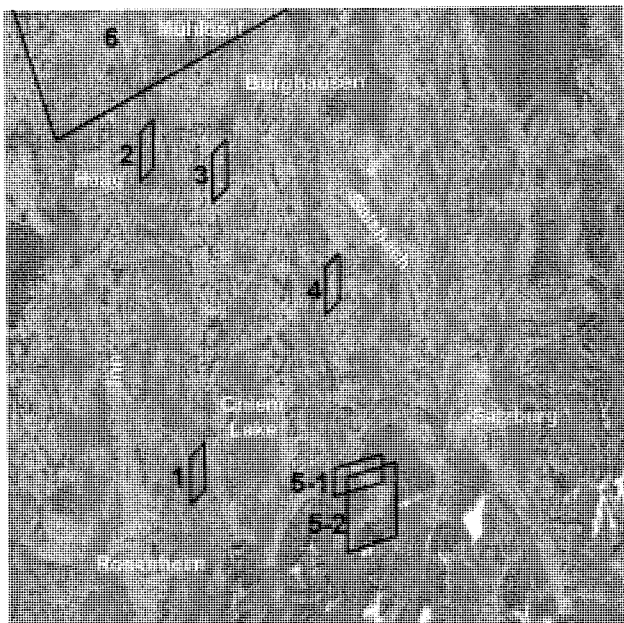


Figure 2. SPOT-5/HRS forward scene with some topographic elements in white (towns, rivers and lakes) and location of reference DEMs (in black) with same notation used in Table 2.

3. POINT MEASUREMENTS

From the available 81 object points, only 41 have been identified in the images. In order to locate them in the images a digital map at 1:50,000 scale (Topo50) was used. The coordinates were given in the Gauss-Krueger system. The exact image coordinates of the points have been measured with unconstrained Least Squares Matching developed at IGP (Baltasvias, 1991), by measuring the points in the master image manually. The final point distribution is shown in Figure 3.

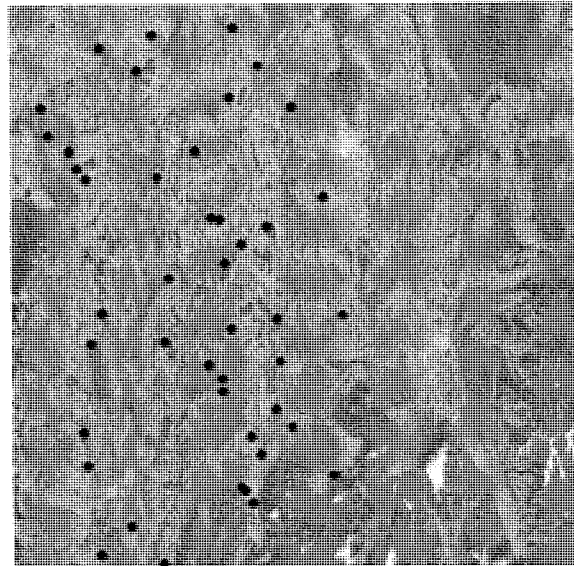


Figure 3. Distribution of 41 object points (in black).

4. IMAGES ORIENTATION

The HRS (High-Resolution Stereoscopic) instrument of SPOT-5 uses linear arrays that scan a single image line at an instant of time in the so-called pushbroom mode. Consequently each line of the HRS image is acquired at a different exposure station with different orientation elements. For the orientation of this kind of imagery two approaches, based on rigorous models and rational function models, are used.

The rigorous model tries to describe the physical properties of the sensor and its image acquisition mode. It is based on collinearity equations, which are extended in order to describe the specific geometry of pushbroom sensors. The adjustment parameters must include the exterior orientation and self-calibration parameters to describe the physical imaging process. Alternatively, rational function models use a general transformation to describe the relationship between image and ground coordinates.

In this work both approaches have been applied. In the next paragraph the algorithms used for the orientation will be described and the results reported.

4.1 Procedure 1: Rigorous model

The aim of rigorous sensor models is to establish a relationship between image and ground reference systems according to the sensor geometry and the available data. For the georeferencing of imagery acquired by pushbroom sensors many different geometric models of varying complexity, rigor and accuracy have been developed, as described in (Fritsch et al., 2000) and (Dowman et al., 2003). A flexible sensor model for the georeferencing of a wide class of linear CCD array sensors has been developed at IGP and already applied to different linear

scanners carried on satellite and aircraft (Poli, 2003). The model is based on the photogrammetric collinearity equations, because each image line is the result of a nearly parallel projection in the flight direction and a perspective projection in the CCD line direction.

The model can be applied to single- and multi-lens sensors. In case of multi-lens sensors, like SPOT-5/HRS, additional parameters describing the relative orientation (displacements and rotations) of each lens with respect to a suitable central point are introduced. During the georeferencing of images from linear CCD array scanners, particular attention must be paid to their external orientation, because each image line is acquired with a different external orientation, that cannot be estimated with a classical bundle adjustment, due to the large number of unknowns (6 for each image line). The sensor position and attitude are modeled with piecewise 2nd order polynomial functions depending on time. The platform trajectory is divided into segments according to the number and distribution of available Ground Control Points (GCPs) and Tie Points (TPs) and for each segment the sensor position and attitude are modeled by 2nd order polynomials. At the points of conjunction between adjacent segments constraints on the zero, first and second order continuity are imposed on the trajectory functions. Additional pseudo-observations can fix some or all parameters to suitable values. For example, if the 2nd order parameters are fixed to zero, the polynomial degree is reduced to 1 (linear functions). This option allows the modeling of the sensor position and attitude in each segment with 2nd or 1st order polynomials, according to the characteristics of the trajectory of the current case study. In case of sensors carried on aircraft, additional GPS and INS observations can be included in the model (Poli, 2002).

The sensor model includes also a self-calibration, which is required for the correction of the systematic errors due to: principal point displacement (d_x, d_y), focal length variation (d_c), radial symmetric (k_1, k_2) and decentering lens distortion (p_1, p_2), scale variation in CCD line direction (s_y) and the CCD line rotation in the focal plane (θ).

Finally the functions modeling the external and the internal orientation are integrated into the collinearity equations, resulting in an indirect georeferencing model. Due to their non-linearity, the complete equations are linearized according to the first-order Taylor decomposition with respect to the unknown parameters. For this operation initial approximations for the unknown parameters are needed. The resulting system is solved with a least squares method. An overview of this sensor model is given in Figure 4.

The sensor model was applied in order to orient the stereopair and estimate the ground coordinates of the Check Points (CPs). The available ephemeris (sensor position and velocity) were used to generate the approximate values for the parameters modeling the sensor external orientation (position and attitude) in fixed Earth-centred geocentric Cartesian system.

The GCPs coordinates were transformed into the same system. From the available 41 object points, a group of them was used as GCPs and the remaining as CPs. Different tests have been carried out in order to choose the best input configuration.

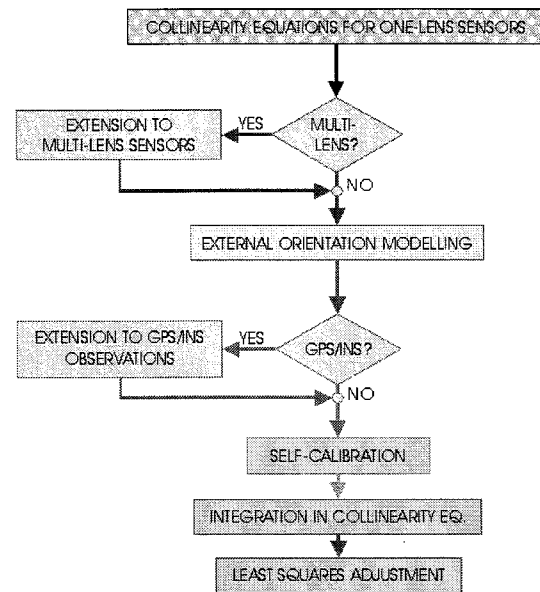


Figure 4. Flowchart with main components of the rigorous sensor model.

The tests were set as follows:

- external orientation modeling with quadratic functions, varying the number of segments and GCPs configurations, no self-calibration;
- external orientation modeling with linear and quadratic functions, best GCPs configuration and best trajectory segments, no self-calibration;
- self-calibration with best external orientation modeling configuration.

The choice of the unknown self-calibration parameters to include in the modeling is based on the analysis of the cross-correlation between the self-calibration parameters, the external orientation parameters and the ground coordinates of the TPs. The best results in the CPs were obtained by modeling the external orientation with 2 segments and 2nd order functions and with self-calibration. The self-calibration parameters that mostly influenced the model were k_1, k_2, p_2 and s_y for both lenses and θ for both lines. The other self-calibration parameters were not used because they correlated highly (>95%) with the external orientation parameters. By changing the number of GCPs and TPs, the RMSE were always less than 1 pixel (Table 3). Figure 5 shows the residuals in planimetry (top) and in height (bottom) using all the object points as GCPs.

Table 3. RMSE for all points using rigorous orientation model.

Number of GCPs + CPs	RMSE in East (m)	RMSE in North (m)	RMSE in Height (m)
8 + 31	3.68	6.52	4.75
16 + 25	3.46	6.22	3.75
41 + 0	3.24	5.52	3.68

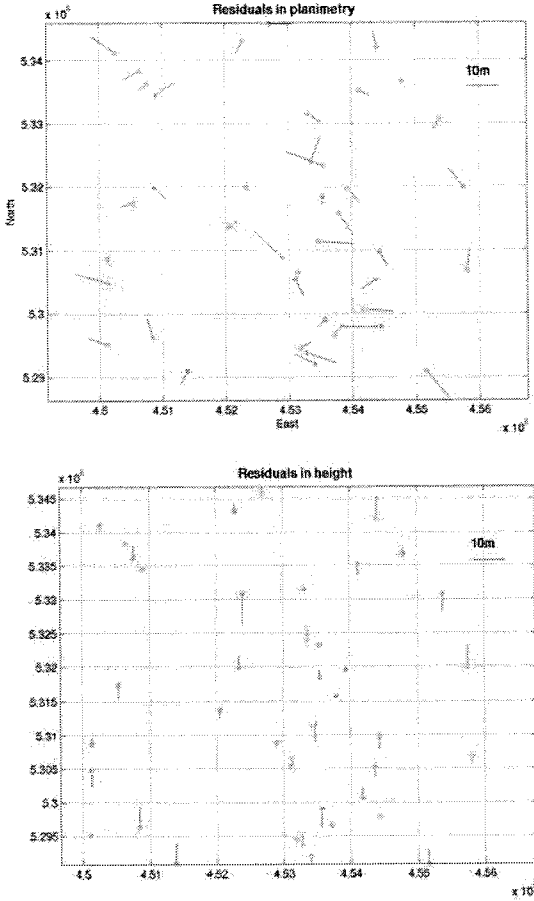


Figure 5. Residuals in planimetry (top) and height (bottom) for all 41 GCPs.

4.2 Procedure 2: Rational functions model

The alternative approach is based on the RPC (Rational Polynomial Coefficients) model. The idea is to describe the camera model contained in the metadata file with suitable rational functions and apply a block adjustment to correct for remaining systematic errors (Zhang et al., 2004). The procedure consists of two main steps:

1. RPC model estimation. After generating a 3D grid of points using the given camera model parameters, the ephemeris and the attitude data attached in the metadata file, the RPC coefficients are determined by a least-squares approach and without GCPs. For details see (Tao et al., 2001). The Equations used for this are rational functions:

$$\begin{aligned} x &= RPC_x(\varphi, \lambda, h) = \frac{NUM_x(\varphi, \lambda, h)}{DEN_x(\varphi, \lambda, h)} \\ y &= RPC_y(\varphi, \lambda, h) = \frac{NUM_y(\varphi, \lambda, h)}{DEN_y(\varphi, \lambda, h)} \end{aligned} \quad (1)$$

Here (φ, λ, h) are normalised object-space geographic coordinates (latitude, longitude and height) and (x, y) are normalised image coordinates, in line and column direction. NUM_x , NUM_y , DEN_x and DEN_y are 3rd order polynomials on (φ, λ, h) , resulting in 67 unknown parameters for each image. The 3D grid of object points is generated from the image-space coordinates, for a set of elevation levels. The RPCs were computed for the whole

test HRS scenes with an internal fitting accuracy of 0.07 pixels (RMSE) and 0.23 pixels maximum difference.

2. Block adjustment with computed RPC model. After the RPC generation in step 1, a block adjustment was performed in order to estimate 6 parameters for each image (affine transformation) to remove remaining systematic errors. As mathematical model of the adjustment, we used the method proposed by (Grodecki et al., 2003). The method is an affine transformation:

$$\begin{aligned} x + a_0 + a_1x + a_2y &= RPC_x(\varphi, \lambda, h) \\ y + b_0 + b_1x + b_2y &= RPC_y(\varphi, \lambda, h) \end{aligned} \quad (2)$$

where a_0, a_1, a_2 and b_0, b_1, b_2 are the adjustment parameters for an image, (x, y) and (λ, φ, h) are image and object coordinates of a GCP or a tie point.

Using this adjustment model, we expect that a_0 and b_0 will absorb any shifts and misalignments in the position and attitude, and the residual interior orientation errors in image line and sample directions. The parameters a_1, a_2, b_1, b_2 are used to absorb the effects of on-board drift errors.

The adjustment results are shown in Table 4 and Figure 6.

Table 4. RMSE for all points with the RPC orientation method.

Number of GCPs + CPs	RMSE in East (m)	RMSE in North (m)	RMSE in Height (m)
4 + 37	5.28	3.87	2.64
8 + 33	5.63	3.96	2.38
41 + 0	4.63	3.66	2.21

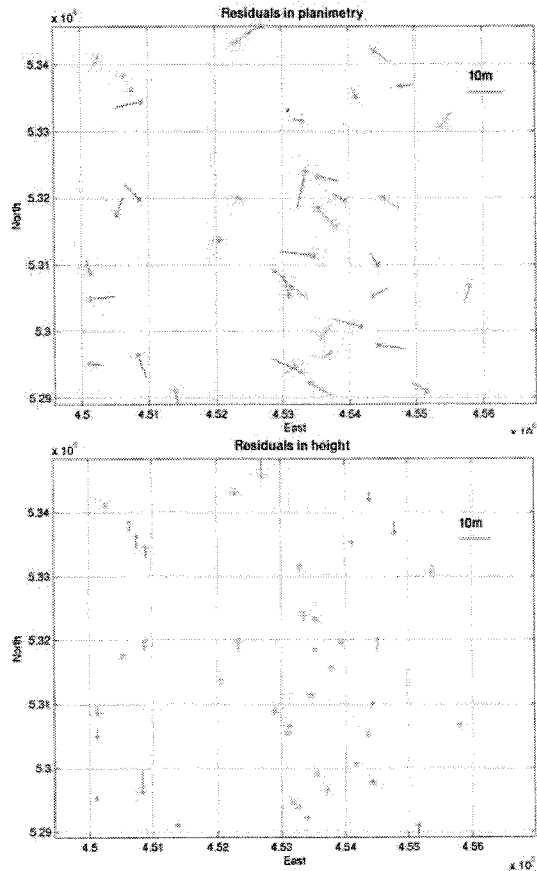


Figure 6. Results in planimetry (top) and height (bottom) from block adjustment with the RPC model and 41 GCPs.

5. MATCHING

In order to automatically extract the DTM / DSMs from the linear array images (airborne or spaceborne), algorithms and software package developed in our group (see Gruen et al., 2002 and Zhang et al., 2003) have been used.

The matching algorithm combines the matching results of the feature points, grid points and edges. It uses a modified version of the MPGC (Multi Photo Geometrically Constrained) matching algorithm (see Gruen, 1985, Gruen et al., 1988 and Baltasvias, 1991) and can achieve sub-pixel accuracy for all the matched features.

Figure 7 shows the workflow of our image matching procedure.

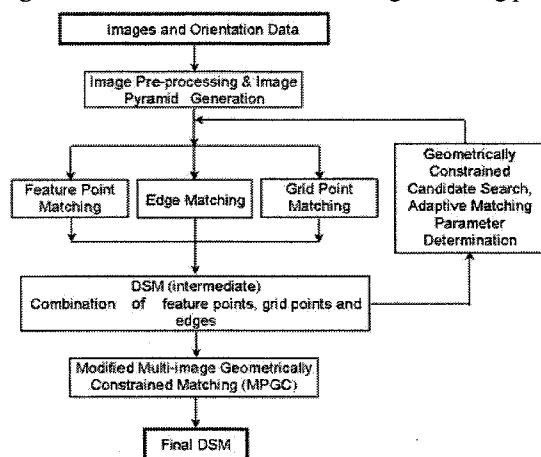


Figure 7. Workflow of the image matching procedure.

For the DSM/DEM generation, the SPOT-5/HRS images and the previously triangulated orientation elements were used. After the pre-processing of the imagery and production of the image pyramid, the matches of three kinds of features, i.e. feature points, grid points and edges, are found progressively in all pyramid levels starting from the low-density features on the images with the lowest resolution. A Triangular Irregular Network (TIN) based DSM is constructed from the matched features on each level of the pyramid and is used in turn in the subsequent pyramid level as approximation for the adaptive computation of the matching parameters. Finally the modified MPGC matching is used to achieve more precise results for all the matched features on the original resolution level (level 0) and to identify some inaccurate and possible false matches. The raster DTM / DSMs are interpolated from the original matching results.

The main features of this matching procedure are:

- It is a combination of feature point, edge and grid point matching. The grid point matching procedure uses relaxation-based relational matching algorithm and can bridge over the non- / little-texture areas through the local smoothness constraints. Edges are introduced to control the smoothness constraints in order to preserve the surface discontinuities.
- The matching parameters include the size of the matching window, the search distance and the threshold value for cross-correlation and MPGC (Least Squares matching). For instance, the procedure uses a smaller matching window, a larger search distance and a smaller threshold value in rougher terrain areas and vice versa. The roughness of the terrain is computed from the approximate DSM on the higher level of image pyramid. The adaptive determination of the matching parameters results in higher success rate and less false matches.

- Line features are important for preserving the surface discontinuities. For this reason a robust edge matching algorithm, which uses the adaptive matching window determination through the analysis of the image contents and local smoothness constraints along the edges, is combined into our procedure.
- Together with point features, edges (in 3D) are introduced as breaklines when a TIN-based DSM is constructed in order to provide good approximations for the matching on the next pyramid level. The computation of the approximate DSM for the highest-level image pyramid uses a matching algorithm based on the *region-growing* strategy (Otto et al., 1988). According to this approach the already measured GCPs and TPs are considered as *seed points*.
- The quality control procedure consists of (1) the local analysis of the smoothness and consistence of the intermediate DSM on each image pyramid level (2) the analysis of the difference between the intermediate DSMs and (3) the analysis of the MPGC results. Blunders can be detected and deleted.
- For each matched feature, a reliability indicator is assigned. Its value is based on the analysis of the matching statistics (cross-correlation and MPGC results). As a consequence, different weights are used during the generation of the grid-based DSM/DEM.

Considering the characteristics of the SPOT-5/HRS image data, some small modifications were introduced in our matching procedure:

- The HRS imagery has 10 meters resolution in cross-track direction and 5 meters in along-track direction (parallax direction). This configuration may result in better accuracy for point determination and DEM generation, but causes some difficulties during the (area-based) matching procedure. In order to avoid the problems, the images have been resampled from 10m x 5m to 10m x 10m and processed with our matching procedure (except the MPGC part). Then the MPGC (Least Squares matching) was run on the original images in order to recover the original matching accuracy. This two-step method results in the reduction of the search distance between corresponding points, which is equivalent to the reduction of the possibility of false matching and the processing time.
- In some difficult areas, like small and steep geomorphologic features (an example is shown in Figure 8), some manually measured points can be introduced as *seed points*. This operation gives better approximations for the matching.

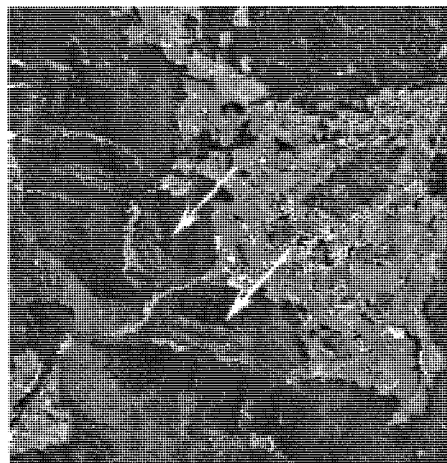


Figure 8. Manually measured seed points in difficult areas (two small hills with steep slopes).

6. DSM GENERATION

The test area includes a mountainous area (rolling and strongly inclined alpine area) on the South part and some hill areas (rough/smooth and weakly inclined areas) on the North part. In order to capture and model the terrain, our image matching software not only generates a large number of mass points (feature points and grid points) but also produces linear features. The TIN based DEM was generated from the matched mass points and the edges (as break-lines). An example of edge matching is shown in Figure 9. As can be seen in this Figure,

even in areas of steep mountains there are many successfully matched linear features. More than 8.4 million points have been matched and 80% of them are marked as 'reliable' points. Some areas, as lakes and rivers, can be manually set as 'dead areas' with a user-friendly interface. In Figure 10 the 3D visualization of the generated DEM is shown. It can be seen that the shape of the DSMs is similar to the reference surfaces, but smoother. This can be expected from the limited resolution of the satellite images.

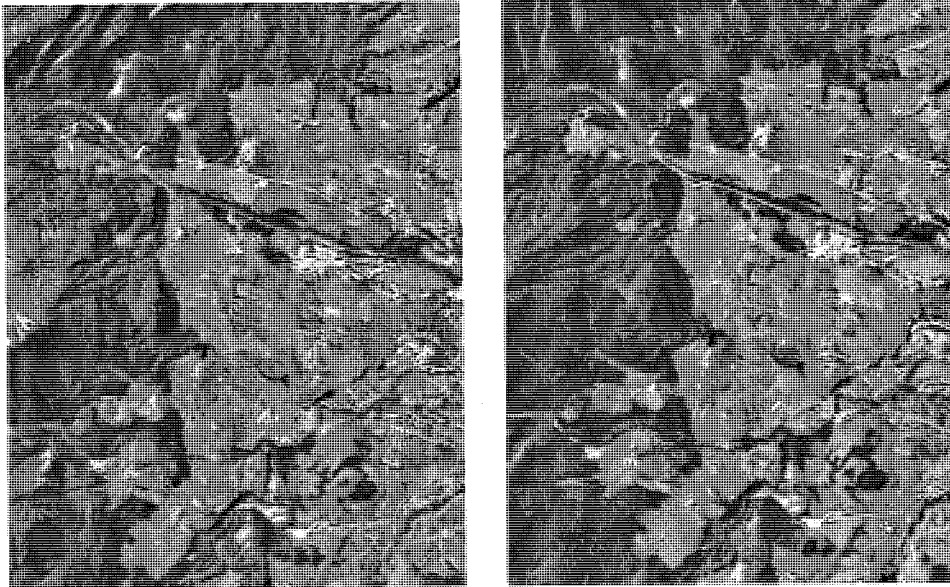


Figure 9. Image window overlapped with matched edges (in violet colour)

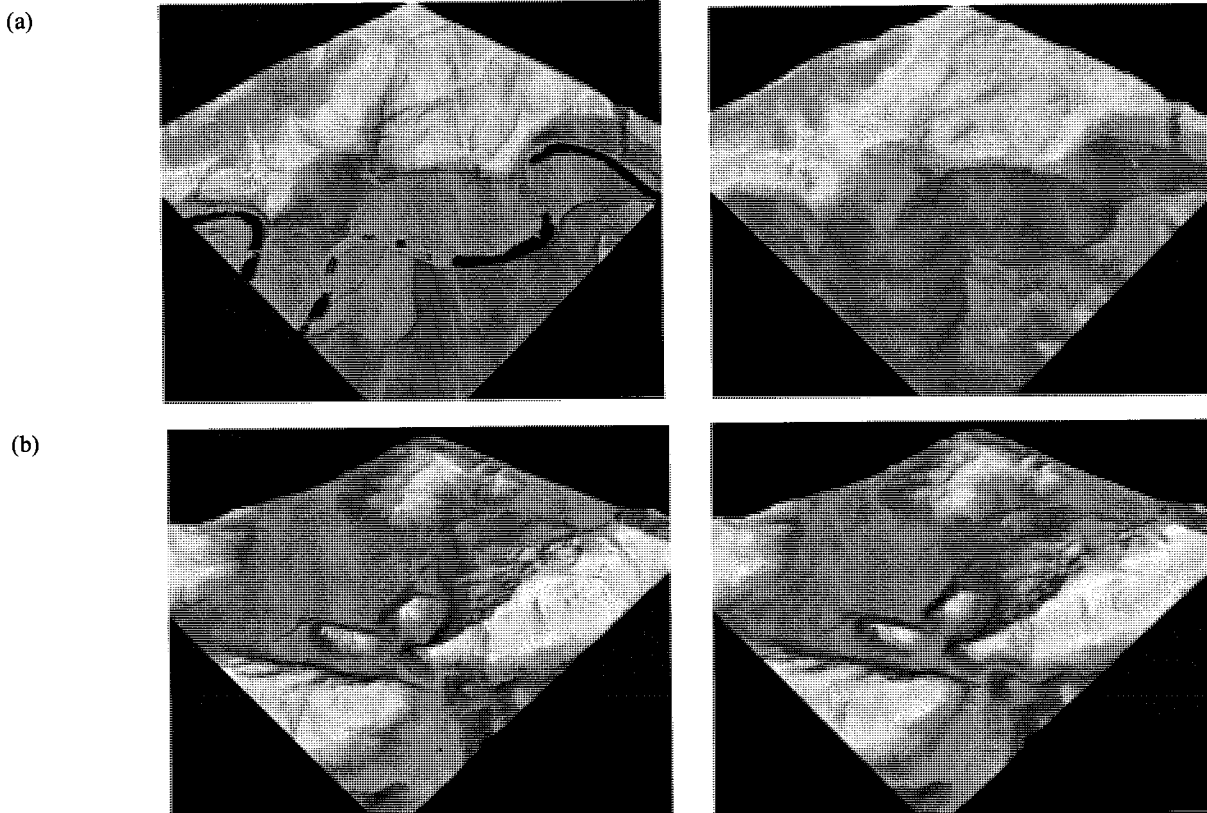


Figure 10. (a): 3D visualization of the reference DEM 2 (5m grid, on the left) and the generated DSM (15m grid, on the right); (b): 3D visualization of the reference DEM 5-2 (25m grid, on the left) and the generated DSM (25m grid, the right).

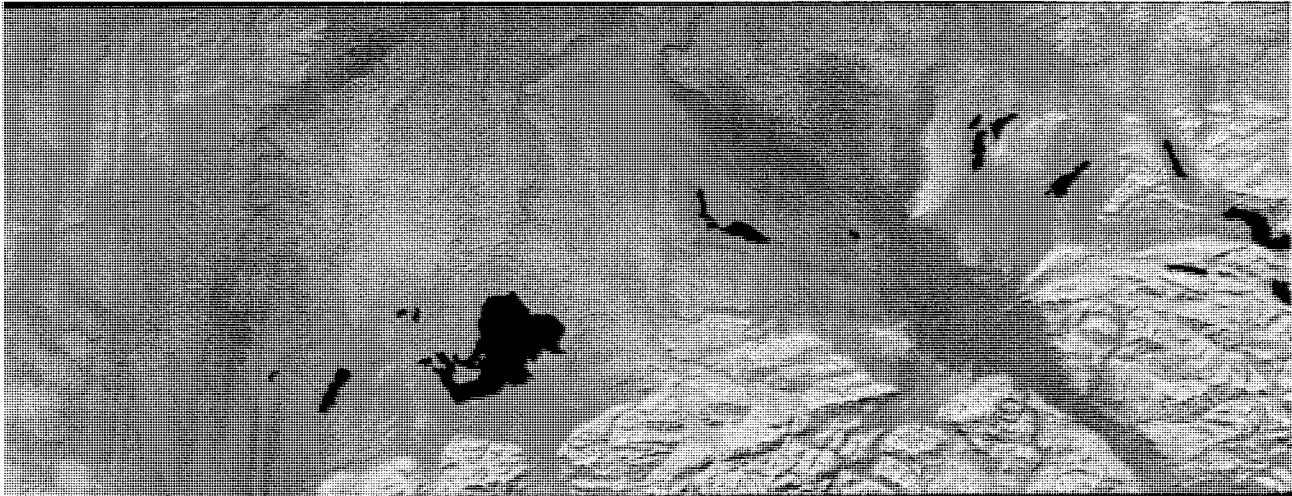


Figure 11. DEM generated for the whole are of 120km x 60km.

7. ACCURACY ANALYSIS

Two DSMs have been generated using the two orientation methods described in Section 4. These DSMs has been compared to the reference DEMs provided by the HRS-SAP. The main characteristics (location, spacing, source, size and height accuracy) of the reference data are shown in Table 2. The coordinate system used in the comparison is the Gauss-Krueger system, Zone 4, with Bessel ellipsoid and Potsdam datum.

Two accuracy tests have been performed in 2.5 D and 3D respectively. In the first test the differences between the heights of the reference DEMs and the corresponding height interpolated from our DSMs have been computed (2.5D). The limit of this approach is that it is highly influenced by the errors of the surface-modelling algorithm. Figure 12 illustrates the concept with a step profile: even if the measurements (green points) are accurate, the modeled profile can not follow the true one. Consequently if the terrain height is compared, in the correspondence of the step, a large difference (Δh) may be measured. For that reason the computation of the 3D orthogonal distance between the surfaces (distance d in Figure 12) is theoretically more correct.

Therefore the second accuracy test is based on the evaluation of the normal distance (3D) between our measurements and the reference DEMs. This test is fundamental in this case study where steep mountains (Alps) are present.

The two tests have been made separately for each DEMs obtained by the procedures described in Section 4. The results are reported in the next paragraphs.

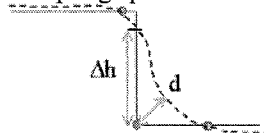


Figure 12. Modelling problems. The true profile is the full black line, the modelled profile is the dashed line.

7.1 Accuracy tests on the terrain height (2.5D)

The results obtained by this test are reported in Table 5. It can be observed that the accuracy of the generated DSM is more or less on 1.0 to 2.0 pixels level, depending on the terrain type. As expected, better accuracy are achieved in smooth and flat areas, while in the mountainous area (DEM 5-1 and 5-2) the RMSE are larger. In all datasets some blunders which failed to be

detected are still present. In the reference datasets called 5-1 and 5-2 some blunders are even above 100 meters, with bias up to 1.0 pixels. Apart from the results of reference DEM 6, all the biases are negative, indicating that our generated DSMs are higher than the reference ones. The results obtained by orientation Procedure 2 appear slightly better than the corresponding results achieved with Procedure 1. The differences between the two orientation approaches are in the order of quarter a pixel.

For further analysis, the frequency distribution of the height differences is shown in the second and third columns of

Table 6. In the frequency distribution of the height differences two peaks occur, one located around value 0.0 and the other one around a negative value (ca. 8m). The relative frequency values are correlated to the percentage of presence of trees. In fact trees causes negative height difference (the green areas), while the open areas have small height difference values. It can be concluded that the bias located at $\approx 8m$ is mainly caused by the trees. This is a main problem for extracting DEM by using the optical imaging systems, as the light cannot penetrate the vegetation. For this reason the areas covered by trees have been manually removed from the images and the accuracy tests have been repeated. The percentage of removed points is 25, 26, 17, 28, 75 and 71 for DEM 1, 2, 3, 4, 5-1, 5-2 and 6 respectively. The results obtained by the new accuracy tests are shown in the last two columns of Table 5. As expected, the negative bias was reduced. This is also graphically confirmed by the new frequency distribution reported in the forth columns of Table 5.

The analysis of the frequency distributions shows that in steep mountain areas (DEM 5-1 and DEM 5-2) there are positive height-difference values. They are probably caused by the presence of blunders or by the local smoothness constraints used in our matching algorithm. These constraints smooth out some steep and small features of the mountain areas under the condition that there are not enough extracted and matched linear features.

7.2 Accuracy tests on the orthogonal distance between two 3D surfaces

This accuracy test has been carried out with the commercial software Geomatic Studio v.4.1 by Raindrop. This software calculates the normal distance between each triangle of a surface (in our case the reference DEMs) and the closest point

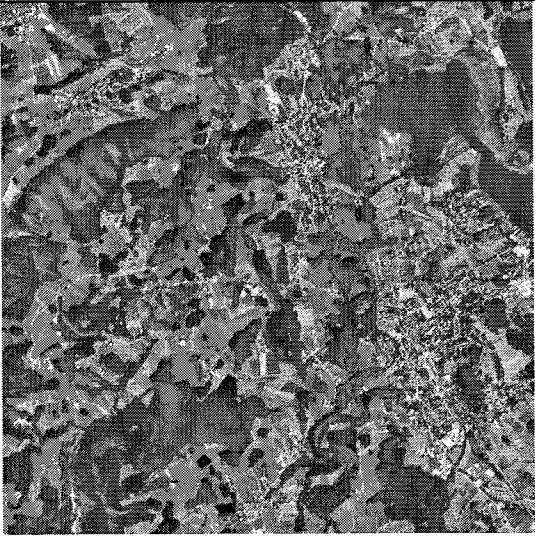
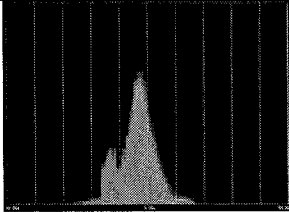
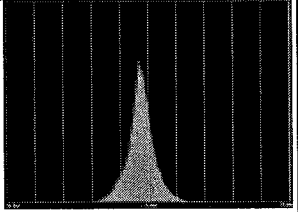
belonging to a point cloud (in our case, the two sets of measurements). The generated DEMs have been compared to the reference ones. The results are reported in Table 7. It can be seen a large reduction of the mean distance and standard deviation in all datasets. This demonstrates that part of the

errors estimated with the 2.5D accuracy tests may be due to modelling errors or to the planimetric errors. Again, the larger errors have been found in mountainous areas (DEM 5-2), while in flat terrains the accuracy of the generated DEMs is very good. The two orientation procedures give very similar results.

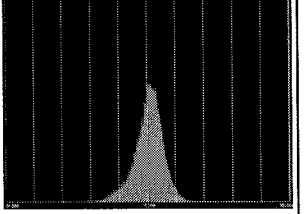
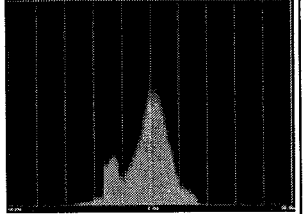
Table 5. Accuracy results using 2.5D comparisons. All results are in meters.

DEM	Number of points		All points Procedure 1				All points Procedure 2				Without trees Procedure 1				Without trees Procedure 2			
	IGP DEM	Reference DEM	Max Diff.	Min Diff.	Mean	RMSE	Max Diff.	Min Diff.	Mean	RMSE	Max Diff.	Min Diff.	Mean	RMSE	Max Diff.	Min Diff.	Mean	RMSE
1	35448	1000000	22.1	-26.1	-4.0	6.2	25.1	-32.9	-2.6	5.7	13.8	-23.6	-3.0	5.4	15.4	-23.7	-1.7	4.6
2	32932	1000000	37.7	-37.1	-3.0	5.5	29.1	-37.1	-1.2	5.0	32.3	-31.2	-1.8	3.9	29.1	-31.7	0.2	3.6
3	33450	1000000	19.7	-17.8	-1.9	4.0	20.7	-17.2	-0.5	3.2	19.6	-14.1	-2.3	3.4	20.7	-13.6	0.1	2.9
4	32067	1000000	11.5	-21.7	-3.8	5.2	13.6	-23.1	-2.5	4.7	9.5	-17.1	-2.8	3.9	10.5	-18.4	-1.2	3.2
5-1	10327	21200	27.1	-36.8	-6.7	9.4	19.2	-33.5	-5.8	8.3	16.1	-23.3	-3.4	5.3	19.1	-13.3	-1.7	4.9
5-2	71795	139200	130.3	-86.1	-5.7	11.2	136.8	-89.3	-4.3	9.5	104.7	-55.5	-4.0	7.4	49.8	-66.8	-1.3	6.7
6	130558	600000	22.7	-19.0	1.1	3.8	26.8	-27.1	1.5	4.0	21.9	-14.6	-0.7	3.9	26.8	-25.9	2.1	4.4

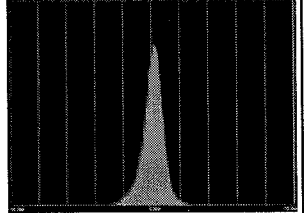
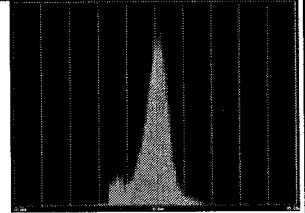
Table 6. Accuracy analysis based on terrain heights. First column: name of reference dataset. Second column: 2D distribution of the height differences. Third column: Frequency distribution of the height differences of all points. Forth column: frequency distribution of the height differences without the tree areas.

DEM	30m -30m	First results	Without trees
1			

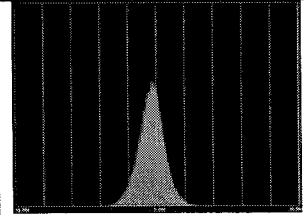
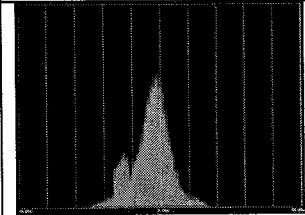
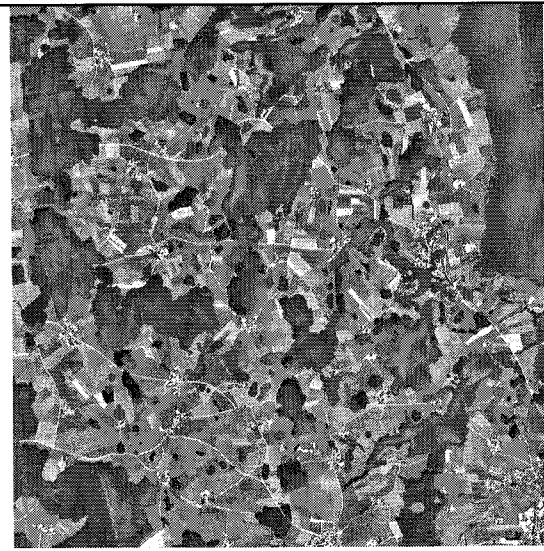
2



3



4



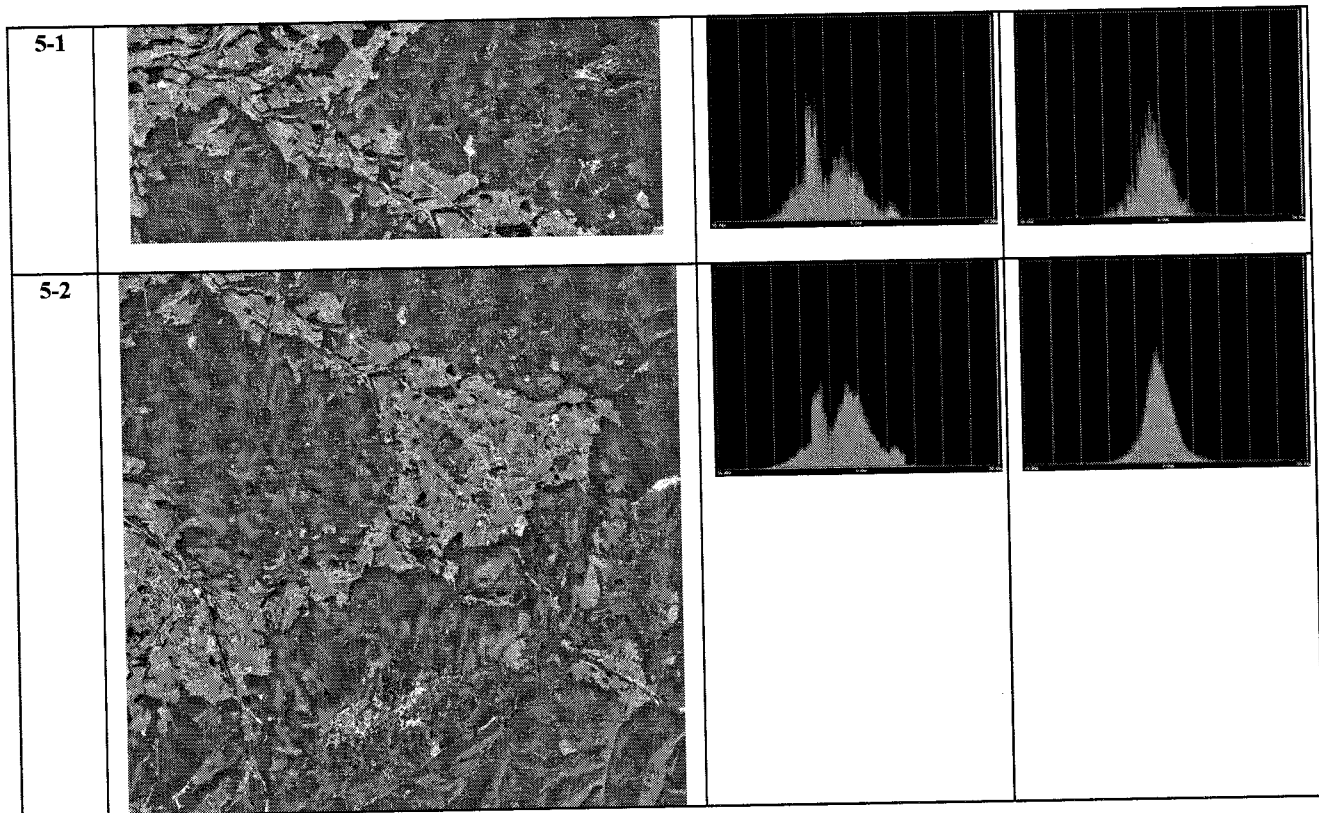


Table 7. Accuracy analysis based on orthogonal distances. In each comparison the absolute maximum distance, the average distance, the standard deviation and the RMSE are shown in meters.

DEM	PROCEDURE 1				PROCEDURE 2			
	Max distance	Average distance	Standard deviation	RMSE	Max distance	Average distance	Standard deviation	RMSE
1	19.60	1.6	1.5	2.2	18.7	2.2	1.7	2.8
2	37.99	1.8	1.8	2.6	37.5	2.7	2.0	3.4
3	22.33	1.4	1.3	1.9	21.4	2.7	1.8	3.2
4	19.74	1.5	1.4	2.1	20.0	2.2	1.6	2.7
5-1	26.25	6.3	4.3	7.6	26.3	6.4	4.4	7.8
5-2	73.60	6.8	5.8	8.9	70.1	6.0	5.0	7.8

8. CONCLUSIONS

In this report the methodology applied and the results obtained during the ISPRS-CNES Initiative about DEM generation from SPOT5-HRS are described.

Our Institute was involved as Co-Investigator in the HRS-SAP Initiative through the processing of the dataset number 9, located in Bavaria (Germany).

All the algorithms used to process the data and generate the DSMs have been developed at our Institute. Using the information contained in the image metafile and a suitable number of GCPs, the images have been oriented according to two different approaches, based on a rigorous sensor model for CCD linear array sensors with along-track stereo capability (Procedure 1) and on Rational Polynomial Functions (Procedure 2). More than eight million image points have been measured in the stereopair with the modified Multi Photo Geometrically Constraint matching algorithm designed for pushbroom imagery. Using the two orientations estimated by Procedures 1 and 2, two distinct DSMs of the full area (120km

x 60km) have been generated and compared to the reference DEMs. For the quality control, a 2.5D (calculation of height differences) and 3D analysis (normal distance between one reference surface and the measured DEM) have been used. Also, the areas covered by trees have been manually removed in order to provide a congruent analysis in order to judge the influence of trees.

From both the 2.5D and 3D quality analysis it resulted an average error between the generated and the reference DSMs of around 1-2 pixels (2.5D analysis) and up to slightly more than 1 pixel (3D analysis), depending on the terrain type. The best results were achieved in smooth and flat areas, while in mountain areas some blunders even exceeded 100 meters. The differences between the DSMs obtained by the two different methods of orientation were less than a fourth of a pixel.

In conclusion, the work carried out at ETH confirmed within the HRS-SAP Initiative the high potential of SPOT5-HRS scenes for DEM generation.

AKNOWLEDGMENTS

We like to thank Mr. Alain Baudoin, CNES, and Mr. Manfred Schroeder, Chairman of ISPRS Commission I, WG 2 "WG I/2 Sensor Calibration and Testing", for the establishment and coordination of the HRS-SAP Initiative and for accepting our Institute as Co-Investigator.

Thanks also go to DLR Oberpfaffenhofen, and in particular to Dr. Peter Reinartz for providing the ground data (reference DEMs, object points) and for supplying all required information.

REFERENCES

Baltsavias, E. P., 1991. Multiphoto Geometrically Constrained Matching. Dissertation, IGP, ETH Z, rich, Mitteilungen No. 49, 221 pages.

Baudoin, A., Schroeder, M., Valorge, C., Bernard, M., Rudowski, V., 2003. The HRS-SAP initiative: A scientific assessment of the High Resolution Stereoscopic instrument on board of SPOT 5 by ISPRS investigators. ISPRS Workshop "High resolution mapping from space 2003", October 2003, Hannover, Germany. Proceedings on CD.

CNES site: <http://www.cnes.fr>

DIMAP site: <http://www.spotimage.fr/dimap/spec/dimap.htm>

Dowman, I. J., Michalis, P., 2003. Generic rigorous model for along track stereo satellite sensors. Proceedings of ISPRS Workshop iHigh Resolution Mapping from Space 2003i, Hannover, 4-6 October. Proceedings on CDROM.

Fritsch, D., Stallmann, D., 2000. Rigorous photogrammetric modelling processing of high resoluition satellite imagery. International Archives of Photogrammetry and Remote Sensing, Vol. 33, Part B1, Amsterdam, pp.313-321.

Gleyzes, J.-P., Meygret, A., Fratter, C., Panem, C., Baillarin, S., Valorge, C., 2003. SPOT5: System overview and image ground segment. IGARSS 2003, Toulouse, France. Proceedings on CD.

Grodecki, J., and Dial, G., 2003, Block Adjustment of High-Resolution Satellite Images Described by Rational Functions. Photogrammetric Engineering and Remote Sensing, Vol. 69, No. 1, pp. 59-70.

Gruen, A., 1985, Adaptive Least Squares Correlation: A powerful Image Matching Technique. South Africa Journal of Photogrammetry, Remote Sensing and Cartography, 14 (3), pp. 175-187.

Gruen, A., Baltsavias, E. P., 1988, Geometrically Constrained Multiphoto Matching. Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 5, pp. 633-641.

Gruen, A., Zhang L., 2002. Automatic DTM Generation from Three-Line-Scanner (TLS) images. IAPRS, Vol. 34, Part 2A, Graz, Austria, pp. 131-137.

Otto, G. P., Chau, T. K. W., 1988. A iRegion-Growingi Algorithm for Matching of Terrain Images. Proc. 4th Alvey Vision Club, University of Manchester, UK, 31 Aug. ñ 2 Sept.

Poli, D., 2003. Georeferencing of MOMS-02 and MISR stereoisimages with strict sensor model. ISPRS Workshop "High resolution mapping from space 2003", October 2003, Hannover, Germany. Proceedings on CD.

Tao, C. V, Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogrammetric Engineering and Remote Sensing, Vol 67, No. 12, pp. 1347-1358.

Zhang L., Gruen, A., 2003. Automatic DSM Generation from StarImager (SI) Data. 6th Conference on Optical 3-D Measurement Techniques September 22-25, 2003 Zurich, Switzerland.

Zhang L., Gruen, A., 2004. Automatic DSM Generation from Linear Array Imagery Data. IAPRS, Vol. 34, Part B3 (to be published).

CHARACTER RECONSTRUCTION AND ANIMATION FROM MONOCULAR SEQUENCES OF IMAGES

Fabio Remondino

Institute for Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: fabio@geod.baug.ethz.ch

Commission V, ICWG V/III

KEY WORDS: Calibration, Orientation, Matching, Reconstruction, Body Modeling, Animation

ABSTRACT

In this paper we present different methods for the calibration and orientation of monocular image sequences and the 3D reconstruction of human characters. Three different situations are considered: a static character imaged with a moving camera, a moving character imaged with a fix camera and a moving character imaged with a moving camera. A self-acquired sequence is used in the first case while in the other cases we used existing sequences available on the Internet or digitized from old videotapes. Most of the image-based techniques use probabilistic approaches to model a character from monocular sequences; on the other hand we use a determinist approach, recovering character's model and movement through a camera model. The recovered human models can be used for visualization purposes, to generate new virtual scenes of the analyzed sequence or for gait analysis.

1. INTRODUCTION

The realistic modeling of human characters from video sequences is a challenging problem that has been investigated a lot in the last decade. Recently the demand of 3D human models is drastically increased for applications like movies, video games, ergonomic, e-commerce, virtual environments and medicine. In this short introduction we consider only the passive image- and triangulation-based reconstruction methods, neglecting those techniques that do not use correspondences (e.g. shape from shading) or computer animation software. A complete human model consists of the 3D shape and the movements of the body (Table 1): most of the available systems consider these two modeling procedures as separate even if they are very closed. A standard approach to capture the *static 3D shape* (and colour) of an entire human body uses laser scanner technology: it is quite expensive but it can generate a whole body model in ca 20 seconds. On the other hand, precise information related to character *movements* is generally acquired with motion capture techniques: they involve a network of cameras and prove an effective and successfully mean to replicate human movements. In between, single- or multi-stations videogrammetry offers an attractive alternative technique, requiring cheap sensors, allowing markerless tracking and providing, at the same time, for 3D shapes and movements information. Model-based approaches are very common, in particular with monocular video streams, while deterministic approaches are almost neglected, often due to the difficulties in recovering the camera parameters. The analysis of existing videos can moreover allow the generation of 3D models of characters who may be long dead or unavailable for common modeling techniques.

3D Shape	Active Sensors	Single-station Videogrammetry <i>Howe [2000]</i> <i>Sidenbladh [2000]</i> <i>Sminchisescu [02]</i> <i>Remondino [02, 03]</i>	Multi-stations Videogrammetry <i>Gavrila [1996]</i> <i>Yamamoto[98]</i> <i>Vedula [1999]</i> <i>DiApuzzo [03]</i>
Movements	Motion Capture		

Table 1: Techniques for human shape and movements modeling.

In this paper we present the analysis of monocular sequences with the aim of (1) generating reliable procedures to calibrate and orient image sequences without typical photogrammetric information and (2) reconstruct 3D models of characters for

visualization and animation purposes. The virtual characters can be used in areas like film production, entertainment, fashion design and augmented reality. Moreover the recovered 3D positions can also serve as basis for the analysis of human movements or medical studies.

2. RECOVERING CAMERA PARAMETERS APPROXIMATIONS FROM EXISTING SEQUENCES

As we want to recover metric information from video sequences (3D characters, scene models or human movement information), we need some metric information about the camera (interior and exterior parameters) and the images (pixel size). The approximations of these parameters are also necessary in the photo-triangulation procedure (bundle adjustment), as we must solve a non-linear problem, based on the collinearity fundamental condition, to obtain a rigorous solution. We assume that we do not know the parameters of the used camera and that we can always define some control points, knowing the dimensions of some objects in the imaged scene.

The *pixel size* is mainly a scale factor for the camera focal length. Its value can be recovered from a set of corresponding object and image coordinates distributed on a plane.

The camera *interior parameters* can be recovered with an approach based on vanishing point and line segments clustering [Caprile et al., 1990; Remondino, 2002] or with orthogonality conditions on line measurements [Krauss, 1996; Van den Heuvel, 1999]. If the image quality does not allow the extraction of lines, the decomposition of the 3x4 matrix of the projective camera model can simultaneously derive the interior parameters given at least 6 control points [Hartley et al., 2000; Remondino, 2003].

Concerning the *exterior parameters*, an approximate solution can be achieved with a closed form space resection [Zeng et al., 1992] or the classical non-linear space resection based on collinearity, given more than 4 points. The DLT method can sequentially recover all the 9 camera parameters given at least 6 control points [Abdel-Aziz et al., 1971]. DLT contains 11 parameters, where two mainly account for film deformation: if no film deformation is present, two constraints can be added to solve the singularity of the redundant parameters [Bopp et al., 1978]. Other approaches are also described in [Slama, 1980; Criminisi, 1999; Foerstner, 2000; Wolf et al., 2000].

3. MODELING A STATIC CHARACTER WITH AN IMAGE SEQUENCE

For the complete reconstruction of a static human model, a full 360 degree azimuth image coverage is required. A single camera, with a fix focal length, is used. The image acquisition can last 30 seconds and this could be considered a limit of the process, as no movement of the person is required. The 3D shape reconstruction from the uncalibrated image sequence is obtained with a process based on (1) calibration and orientation of the images, (2) matching process on the human body surface and (3) 3D point cloud generation and modeling.

3.1 Camera calibration and image orientation

The calibration and orientation of the images have to be performed in order to extract precise 3D information of the character. The process (Figure 1) is based on a photogrammetric bundle adjustment; the required image correspondences (section 3.1.1) are found with an improved version of a process already presented in [Remondino, 2002].

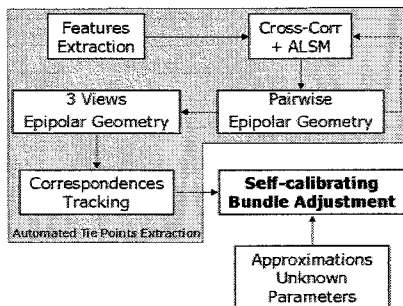


Figure 1: The camera calibration and image orientation pipeline.

3.1.1 Automatic tie points extraction

Most of the presented systems [e.g. Fitzgibbon et al., 1998; Pollefeys et al., 1998; Roth et al., 2000] developed for the orientation of image sequences with automatic extraction of corresponding points require very short baseline between the images (typically called 'shape-from-video'). Few strategy instead can reliable deal with wide-baseline images [Tuytelaars et al., 2000; Remondino, 2002]. Our approach extracts automatically corresponding points with the following 6 steps:

1. *Interest points identification.* A set of interest points or corners in each image of the sequence is extracted using Foerstner operator or Harris corner detector with a threshold on the number of corners extracted based on the image size. A good point distribution is assured by subdividing the images in small patches (9x9 pixel on an image of 1200x1600) and keeping only the points with the highest interest value in those patches.
2. *Correspondences matching.* The extracted features between adjacent images are matched at first with cross-correlation and then the results are refined using adaptive least square matching (ALSM) [Gruen, 1985]. Cross-correlation alone cannot always guarantee the correct match while the ALSM, with template rotation and reshaping, provides for more accurate results. The point with biggest correlation coefficient is used as approximation for the template matching process. The process returns the best match in the second image for each interest point in the first image.
3. *Filtering false correspondences.* Because of the unguided matching process, the found matched pairs often contain outliers. Therefore a filtering of the incorrect matches is performed using the disparity gradient between the found correspondences. The smaller is the disparity gradient, the

more the two correspondences are in agreement. The sum of all disparity gradients of each matched point relative to all other neighbourhood matches is computed. Those matches that have a disparity gradient sum greater than the median of the sums are removed. In case of big baselines or in presence (at the same time) of translation, rotation, shearing and scale between consecutive images, the algorithm can achieve incorrect results if applied on the whole image: therefore the filtering process has to be performed on small image regions.

4. *Epipolar geometry between image pairs.* A pairwise relative orientation and an outlier rejection using those matches that pass the filtering process are afterwards performed. Based on the coplanarity condition, the fundamental matrix is computed with the Least Median of the Squares (LMedS) method; LMedS estimators solve non-linear minimization problems and yield the smallest value for the median of the squared residuals computed for the data set. Therefore they are very robust in case of false matches or outliers due to false localisation. The computed epipolar geometry is then used to refine the matching process (step 3), which is now performed as guided matching along the epipolar line.
5. *Epipolar geometry between image triplets.* Not all the correspondences that support the pairwise relative orientation are necessarily correct. In fact a pair of correspondences can support the epipolar geometry by chance (e.g. a repeated pattern aligned with the epipolar line). These kinds of ambiguities and blunders are reduced considering the epipolar geometry between three consecutive images. A linear representation for the relative orientation of three frames is represented by the trifocal tensor T [Shashua, 1994]. T is represented by a set of three 3x3 matrices and is computed from at least 7 correspondences without knowledge of the motion or calibration of the cameras. In our process, the tensor is computed with a RANSAC algorithm [Fischler et al., 1981] using the correspondences that support two adjacent pair of images and their epipolar geometry. The RANSAC is a robust estimator, which fits a model (tensor T) to a data set (triplet of correspondences) starting from a minimal subset of the data. The found tensor T is used (1) to verify whether the image points are correct corresponding features between three views and (2) to compute the image coordinates of a point in a view, given the corresponding image positions in the other two images. This transfer is very useful when in one view are not found many correspondences. As result of this step, for each triplet of images, a set of corresponding points, supporting the related epipolar geometry is recovered.
6. *Tracking image correspondences through the sequence.* After the computation of a T tensor for every consecutive triplet of images, we consider all the overlapping tensors (e.g. T_{123} , T_{234} , T_{345} , ...) and we look for those correspondences which support consecutive tensors. That is, given two adjacent tensors T_{abc} and T_{bcd} with supporting points $(x_a, y_a, x_b, y_b, x_c, y_c)$ and $(x'_b, y'_b, x'_c, y'_c, x'_d, y'_d)$, if (x_b, y_b, x_c, y_c) in the first tensor T_{abc} is equal to (x'_b, y'_b, x'_c, y'_c) in the successive tensor T_{bcd} , this means that the point in images a, b, c and d is the same and therefore this point must have the same identifier. Each point is tracked as long as possible in the sequence and the obtained correspondences are used as tie points for the successive bundle-adjustment.

3.1.2 Photo-triangulation with bundle-adjustment

A photogrammetric self-calibrating bundle-adjustment is performed using the found image correspondences and the approximations of the unknown camera parameters (section 2). Because of the network geometry and the lack of accurate

control information, usually not all the additional parameters (APs) are recovered.

3.2 Matching process

In order to recover the 3D shape of the static human figure, a dense set of corresponding image points is extracted with an automated matching process [DiApuzzo, 2003]. The matching establishes correspondences between triplet of images starting from some seed points selected manually and distributed on the region of interest. The epipolar geometry, recovered in the orientation process is also used to improve the quality of the results. The central image is used as template and the other two (left and right) are used as search images (slaves). The matcher searches the corresponding points in the two slaves independently and at the end of the process, the data sets are merged to become triplets of matched points. The matching can fail if lacks of natural texture are presents (e.g. uniform colour); therefore the performance of the process is improved with Wallis filter to enhance the low frequencies of the images.

3.3 3D reconstruction and modeling

The 3D coordinates of the 2D matched points are afterwards computed with forward intersection using the results of the orientation process. A spatial filter is also applied to reduce the noise in the 3D data (possible outliers) and to get a more uniform density of the point cloud. If the matching process fails, some holes could be present in the generated point cloud: therefore a semi-automatic closure of the gaps is performed, using the curvature and density of the surrounding points. Moreover, if small movements of the person are occurred during the acquisition, the point cloud of each single triplet could appear misalign respect to the others. Therefore a 3D conformal transformation is applied: one triplet is taken as reference and all the others are transformed according to the reference one.

Concerning the modeling of the recovered unorganized 3D point cloud, we can (1) generate a polygonal surface with reverse-engineer packages or (2) fit a predefined 3D human model to our 3D data [DiApuzzo et al., 1999; Ramsis].

3.4 Results of the modeling of a static character

The presented example shows the modeling of a standing person with a digital still video camera Sony F505 (Figure 2).

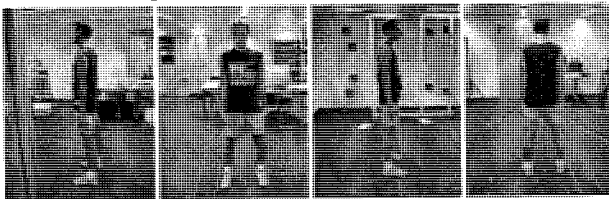


Figure 2: Four (out of 12) images (1200x1600 pixels) used for the 3D static human body reconstruction.

The automatic tie points identification (section 3.1.1) found more than 150 correspondences that were imported in the bundle as well as four control points (measured manually on the body) used for the space resection process and the datum definition. At the end of the adjustment, a camera constant of 8.4 mm was estimated while the position of the principal point was kept fix in the middle of the images (and compensated with the exterior orientation parameters) as no significant camera roll diversity was present. Concerning the distortion parameters, only the first parameter of radial distortion (K1) turned out to

be significant while the others were not estimated, as an over-parameterization could lead to a degradation of the results. The final exterior orientation of the images as well as the 3D coordinates of the tie points are shown in Figure 3.



Figure 3: Recovered camera poses and 3D coordinates of the tie points (left). The influence of the APs on the image grid, 3 times amplified (right).

Afterwards, the matching process between 4 triplets of images produced ca 36 000 2D correspondences that have been converted and filtered in a point cloud of ca 34 000 points (Figure 4). The recovered 3D point cloud of the person is computed with a mean accuracy in x-y of 2.3 mm and in z direction of 3.3 mm. The 3D data can then easily be imported in commercial packages for modeling, visualization and animation purposes or e.g. used for diet management.

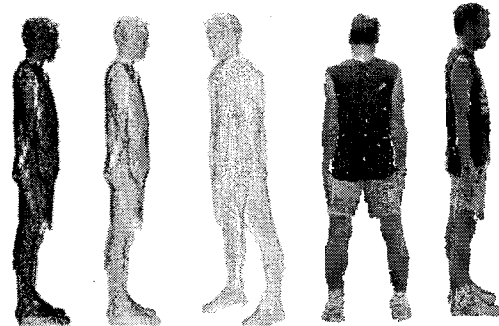


Figure 4: 3D point cloud of the human body imaged in Figure 2 pre and after the filtering (left). Visualization of the recovered point cloud with pixel intensity (right).

4. MODELING A MOVING CHARACTER WITH A FIX CAMERA

Nowadays it is very common to find image streams acquired with a fix camera, like in forensic surveillance, movies and sport events. Due to the complicate shape of the human body, a fix camera that images a moving character cannot correctly model the whole shape, unless we consider small part of the body (e.g. head, arm or torso). In particular, in the movies, we can often see a static camera filming a rotating head. Face modeling and animation has been investigated since 20 years in the graphic community. Due to the symmetric forms and geometric properties of the human head, the modeling requires very precise measurements. A part from laser scanner, most of the single-camera approaches are model-based (requiring fitting and minimization problems) while few methods recover the 3D shape through a camera model. Our solution tries to model the head regarding the camera as moving around it. Therefore it can be considered as a particular problem of the previous case. We have only to assume that the head is not deforming during the movement.

An example is presented in Figure 5. The image sequence, found on the Internet and with a resolution of 256x256 pixels, shows a person who is rotating the head. No camera and scene information is available and, for the processing, we consider the images as acquired by a moving camera around a fix head.



Figure 5: A fix camera imaging a rotating head (16 frames).

Due to the small region of interest (the head) and the very short baseline, the corresponding points for the image orientation are selected manually in the first frame: then they are matched automatically in all the other images using a tracking algorithm based on least squares template matching. For the datum definition and the space resection algorithm, 4 points extracted from a face laser scanner data set are used. Afterwards, with a bundle-adjustment we recovered the camera parameters: no additional parameters (APs) were used and only the focal length was computed. The recovered epipolar geometry is displayed in Figure 6.

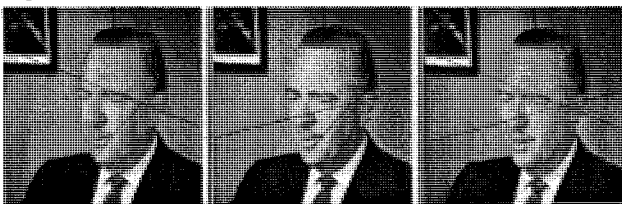


Figure 6: Recovered epipolar geometry between an image triplet.

Finally we applied the matching process described in section 3.2 on 3 triplets of images to get the 3D point cloud of the head. The results, with related pixel intensity, are shown in Figure 7.

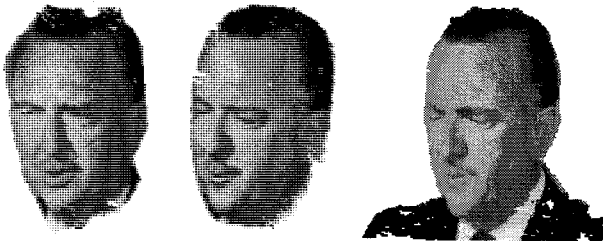


Figure 7: 3D model of the moving head.

5. MODELING A MOVING CHARACTER WITH A MOVING CAMERA

A moving character imaged with a moving camera represents the most difficult case for the character reconstruction problem from image sequences. The camera can be (1) moved with a mechanical arm or on a small railroad or can be (2) stationary but freely rotating on a tripod or on the shoulder of a cameraman. In particular, sport videos are usually acquired with a rotating system and often far away from the scene. In these cases, the baseline between the frames is very short and because of the movements of the character, a standard perspective approach cannot be used, in particular for the 3D modeling of the character. Nevertheless, we recover camera parameters and 3D object information through a camera model, without any model-based adjustment.

5.1 Image acquisition and orientation

A sequence of 60 images has been digitized from an old videotape, using a Matrox DigiSuite grabber, with a resolution of 720x576 pixels (Figure 8). For the orientation and reconstruction only 9 images are used (those images where the moving character has both feet on the ground). From a quick analysis of the images, we can deduce that a rotation is mainly

occurring during the video acquisition while no zooming effects are presents. A right-hand coordinate system with the origin in the left corner of the court is set and some control points are defined knowing the dimensions of the basketball court. Because of the low image quality (interlaced video) the image measurements were performed manually.

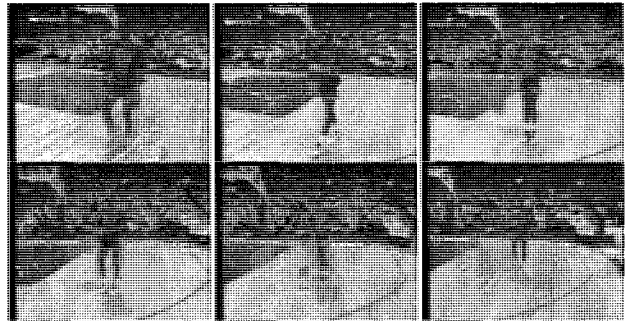


Figure 8: Moving character filmed with a moving camera.

All the measurements are imported as weighted observations and used as tie points in the adjustment. At first, for each single frame, DLT and space resection are used to get an approximation of the camera parameters. Afterwards a bundle adjustment is applied to recover all the parameters, using a block-invariant APs set. We allowed free rotations and very small translation of the camera, weighting the parameters and applying significance tests to analysis the determinability of the APs. The adjustment results ($\sigma_{0\text{-post}}=1.3$ pixel) show a focal length value of 22.7 mm and a pixel aspect ratio of 1.12. The non-unity of the aspect ratio can come from the old video camera or because of the frame grabber used in the digitization process. Concerning the lens distortion, only K1 (radial distortion) turned out to be significant while the others parameters could not be reliable determined. The principal point was kept fix in the middle of the images and compensated with the exterior orientation parameters. Figure 9 shows the global distortion effect on the image grid (3 times amplified) as well as the recovered camera positions.

5.2 3D reconstruction and modeling

For man-made objects (e.g. buildings), geometric constraints on the object (e.g. perpendicularity and orthogonality) can be used to solve the ill-posed problem of the 3D reconstruction from a monocular image. In case of free-form objects (e.g. the human body), we could use a probabilistic approach [Sidenblad, 2000; Sminchisescu, 2002] or other assumptions must be provided [Remondino et al., 2003]:

1. the perspective collinearity model is simplified into a scaled orthographic projection;
2. the human body is represented in a skeleton form, with a series of joints and connected segments of known relative lengths;
3. further constraints on joints depth and segment's perpendicularity are applied to obtained more accurate and reliable 3D models.

This reconstruction algorithm is applied to every frame of the sequence, given the image coordinates of some joints of the human body and the relative lengths of the skeleton segments. For each image, a 3D human model is generated but, because of the orthographic projection, the models are no more in the same reference system. Therefore a 3D conformal transformation is applied using, as common points, the 2 feet (which are always the ground) and the head of the character (known height). The object position of the feet is recovered with a 2D projective

invariance between two planes (basketball court and its image) that undergo a perspective projection [Semple et al., 1952]: the relationship between the two planes is specified if the coordinates of at least 4 corresponding points in each of the two projectively related planes are given. On the other hand, the object position of the head is computed as the middle point between the 2 feet. The invariance property and the conformal transformation are applied to each orthographic model of the sequence and the obtained 3D coordinates are then refined using the camera parameters recovered in the orientation process. The final result is presented in Figure 9, together with the reconstructed scene.

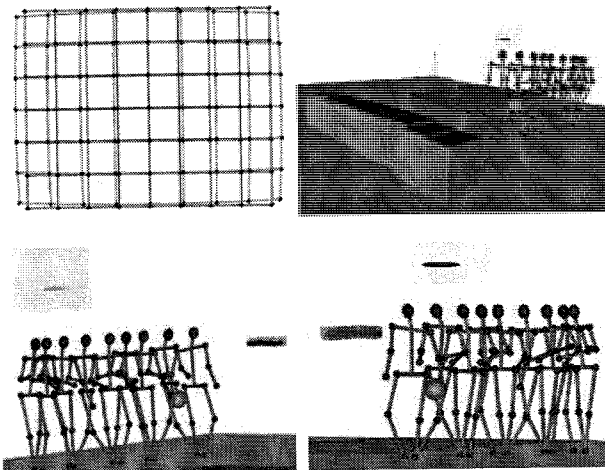


Figure 9: Influence of APs for the analyzed camera (upper left). The camera poses as well as the 3D reconstruction of the basketball court and the moving character (other images).

The recovered poses of the moving human can be used for gait analysis or for the animation of virtual characters in the movie production.

5.3 Other example

Another sequence, presented in Figure 10, is analyzed. The camera is far away from the scene and is rotating (probably on a tripod) and zooming to follow the moving character. The calibration and orientation process, performed with a self-calibrating bundle adjustment with frame-invariant APs sets, recovered a constant increasing of the camera focal length and, again, a non-unity of the pixel aspect ratio ($1.10 \pm 4.5e^{-3}$).



Figure 10: Some frames of a video sequence of a basketball action. The camera is rotating and zooming.

Because of the low precision of the image measurements ($\sigma_{\text{priori}} = 2$ pixel) and the unfair network geometry, the principal point of the camera and the other terms used to model the lens distortion are not computed as very poorly determinable. The final standard deviation resulted 1.7 pixels while the RMS of

image coordinates residuals are 38.45 μm in x direction and 29.08 μm in y direction.

The 3D reconstruction of the moving character is afterwards performed as described in section 5.2. In this case, the orthographic models of each frame could not be transformed into the camera reference system with a conformal transformation. Nevertheless the recovered 3D models are imported in Maya to animate the reconstructed character and generate new virtual scenes of the analyzed sequence (Figure 11).

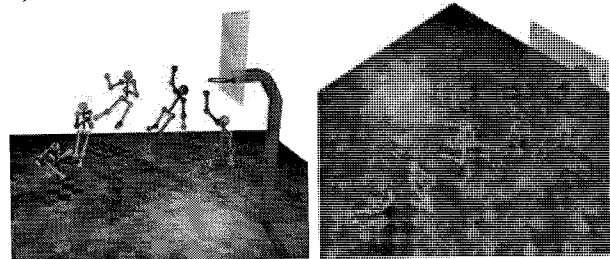


Figure 11: 3D models of the moving character visualized and animated with Maya.

To improve the visual quality and the realism of the reconstructed 3D human skeleton, we fitted a laser scanner human body model [Cyberware] to our data (Figure 12). The modeling and animation features of Maya software allow a semi-automatic fitting of the laser-data polygonal mesh to the skeleton model. The inverse kinematics method and a skinning process are respectively used to animate the model and bind the polygonal mesh with the skeleton [Learning Maya, 2003; Remondino et al., 2003].

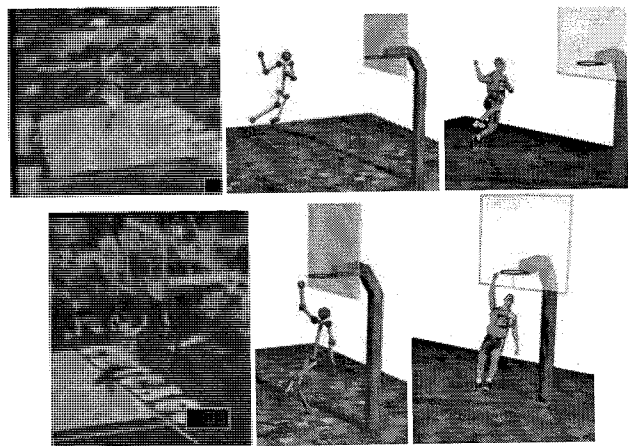


Figure 12: Two examples showing the results of the 3D reconstruction and the modeling process. Original frame of the sequence (left), reconstructed 3D human skeleton (middle) and fitting result, from a slightly different point of view (right).

6. CONCLUSION

The photogrammetric analysis of monocular video sequences and the generation of 3D human models were presented.

The image orientation and calibration was successfully achieved with a perspective bundle adjustment, weighting all the parameters and analysing their determinability with statistical tests. The human modeling, in particular from old videos, showed the capability of videogrammetry to provide for virtual characters useful for augmented reality applications, persons identification and to generate new scenes involving models of characters who are dead or unavailable for common modeling

systems. Finally, the markerless motion measurements and the recovered 3D positions of the human joints can also be used for gait analysis and sport medicine.

REFERENCES

- Abdel-Aziz, Y., Karara, H., 1971: Direct linear transformation from comparator coordinates into object-space coordinates. *Close range Photogrammetry*, pp.1-18, ASPRS, Falls Church, Virginia.
- Bopp, H., Krauss, H., 1978: An orientation and calibration method for non-topographic applications. *PE&RS*, Vol. 44(9)
- Caprile B., Torre, V., 1990: Using vanishing point for camera calibration. *International Journal of Computer Vision*, Vol. 4(2), pp. 127-139
- Criminisi, A., 1999: Accurate Visual Metrology from Single and Multiple Uncalibrated Images. Ph.D. Diss., Oxford University
- Cyberware: <http://www.cyberware.com> [January 2004]
- D'Apuzzo, N., Plankers, R., Fua, P., Gruen, A., Thalmann, D., 1999: Modeling human bodies from video sequences. In El-Hakim/Gruen (Eds.), *Videometrics VI*, Proc. of SPIE, Vol. 3461, pp. 36-47
- DiApuzzo, N., 2003: Surface Measurement and Tracking of Human Body Parts from Multi Station Video Sequences Ph.D. Dissertation, ETH Zurich, Nr. 15271
- Fischler, M., Bolles, R., 1981: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Assoc. Comp. Mach.*, Vol. 24(6), pp. 381-395
- Fitzgibbon, A., Zisserman, A., 1998: Automatic 3D model acquisition and generation of new images from video sequences. *Proceedings of European Signal Processing Conference*, pp. 1261-1269
- Foerstner, W., 2000: New orientation procedure. *Int. Arch. of Photogrammetry and Remote Sensing*, 33(3), pp. 297-304
- Gavrila, D.M., Davis, L., 1996: 3D model-based tracking of humans in action: a multi-view approach. *IEEE CVPR Proc.* pp. 73-80
- Gruen, A., 1985: Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, Vol. 14(3), pp.175-187
- Hartley, R., Zisserman, A., 2000: *Multiview geometry in computer vision*. Cambridge Press
- Howe, N., Leventon, M., Freeman, W., 2000: Bayesian reconstruction of 3D human motion from single-camera video. *Advances in Neural Information Processing System*, Vol. 12, pp. 820-826, MIT Press
- Krauss, K., 1996: *Photogrammetry*, Vol.2. Duemmler Verlag, Bonn
- Learning Maya 5 - Foundation, 2003. Alias Wavefront <http://www.aliaswavefront.com> [January 2004]
- Pollefeys, M., Koch, R., Van Gool, L., 1998: Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters, *IEEE. ICCV Proc.*, pp.90-96
- Ramsis: <http://www.ramsis.de> [January 2004]
- Remondino, F., 2002: Image Sequence Analysis for Human Body Reconstruction. *Int. Arch. of Photogrammetry and Remote Sensing*, Vol. 34(5), pp. 590-595
- Remondino, F., 2003: Recovering Metric Information from Old Monocular Video Sequences. In Gruen/Kahmen (Eds.). *VI Conference on Optical 3D Measurement Techniques*, Vol.2, pp 214-222
- Remondino, F., Roditakis, A., 2003: Human Figures Reconstruction and Modeling from Single images or Monocular Video Sequences. *4th IEEE International 3DIM Conference*
- Roth, G., Whitehead, A., 2000: Using projective vision to find camera positions in an image sequence. *13th Vision Interface Conference*
- Semple, J.G., Kneebone, G.T., 1952: *Algebraic Projective Geometry*, Oxford Press
- Shashua, A., 1994: Trilinearity in visual recognition by alignment. *ECCV, Lectures Notes in Computer Science*, Vol. 800, Springer-Verlag, pp.479-484
- Sidenbladh, H., Black, M., Fleet, D., 2000: Stochastic Tracking of 3D Human Figures Using 2D Image Motion. *European Conference on Computer Vision*, D. Vernon (Ed.), Springer Verlag, LNCS 1843, pp. 702-718
- Slama, C., 1980: *Manual of Photogrammetry*. ASPRS, Falls Church, Virginia
- Sminchisescu, C., 2002: Three Dimensional Human Modeling and Motion Reconstruction in Monocular Video Sequences Ph.D. Dissertation, INRIA Grenoble
- Tuytelaars T., Van Gool, L., 2000: Wide baseline stereo matching based on local, affinity invariant regions, *Proc. British Machine Vision Conference*, Vol. 2, pp. 412-425
- Van den Heuvel, F.A., 1999: Estimation of interior parameters from constraints on line measurements in a single image. *Int. Arch. of Photogrammetry and Remote Sensing*, 32(5), pp.81-88
- Yamamoto, M., Sato, A., Kawada, S., Kondo T., Osaki, Y., 1998: Incremental Tracking of Human Actions from Multiple Views. *IEEE CVPR Proceedings*
- Wolf, P., Dewitt, B., 2000: *Elements of Photogrammetry*. McGraw Hill, New York
- Zheng, Z., Wang, X., 1992: A general solution of a closed-form space resection. *PE&RS*, Vol. 58(3), pp.327-338

35 YEARS OF INTERNET, 10 YEARS OF ISPRS ONLINE

Fabio Remondino^a, Tuan-chih Chen^{b,*}

^a Webmaster, ISPRS; Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland
E-mail: fabio@geod.baug.ethz.ch

^b Events Calendar Editor, ISPRS; Dept. of Civil Engineering, China Institute of Technology, Taipei, Taiwan
E-mail: profchen@ms13.hinet.net

Commission VI

KEY WORDS: Internet/Web, Information, Education, Statistics, On-line

ABSTRACT:

The Internet and the most known part of it, the World Wide Web, are one of the greatest inventions of our time. The formation and the growth of these big Net has changed the way we do business, communicate, entertain, retrieve information, and even educate ourselves. This worldwide medium allows interaction with people without regard for geographic location and it is a great exchange and source of data, information and software. Inside this electronic world, ISPRS is present since 1994 with the goal of providing information, coordination and operations structure of its activities. In this paper, at first a short history of the WWW is presented, with its main facilities and possibilities. In the second part, after a short introduction on the web site of the society, are presented history, presence, and future plans about the use of Internet by ISPRS, and how ISPRS can make use of existing technologies to improve what it now offered, including the educational strategy. The ISPRS Events Calendar is also presented.

INTRODUCTION

The Internet is big and global collection of networks connected together in different ways to form a single entity. The Internet is at once a broadcasting capability in the entire world, a mechanism for information dissemination and a medium for collaboration or interaction between individuals and their computers without regard for geographic location. Sometimes called simply "the Net," it is a public, cooperative, and self-sustaining facility accessible to hundreds of millions of people, a worldwide network of networks in which a user with his computer can get information from any other computer.

Nobody owns the Internet but there are some no-profit organisations that define how we use and interact with the Internet. The most widely used part of it is the World Wide Web, created in 1990 as user-friendly face of the information available on the Internet and, together with e-mails and search engines, provides efficient access to an increasing amount of information. Other services widely used are FTP, e-mail and telnet.

After 35 years of developments and improvements, nowadays there are more than 600 million people online. Everyday more people use an online computer to find information, learn, educate and communicate. But the presence and the access to the Internet are not well distributed in all the countries as there are still part of the world where the use of Internet has yet to grow because it was politically repressed or the infrastructures are still in development.

Inside this big electronic world, ISPRS is present since 1994 with the goal of providing information, coordination and activities of its structure.

The goal of this work is to review the history of the Internet and the WWW, as well as give an overview over the global network, its services and its growth. Then is also described how ISPRS is inserted since 10 years in this network and which

initiatives are taken to present the society to the world-wide public.

INTERNET AND WWW

2.1 The history [ISOC]

The precursor of the Internet is called ARPANet (Advanced Research Projects Agency Network). The US Department of Defence in fact developed it in the late 60's as an experiment in wide-area-networking that would survive to nuclear war. ARPA was the answer to the rising American Cold War fear about military inferiority, fuelled not least by the Russian Sputnik success. In the autumn of 1969 the first ARPANET computer was connected to the ARPANET's IMP node at the University of California at Los Angeles (UCLA) and by the end of the year, the network included four computers (the University of California, Santa Barbara and the University of Utah). All the computers used different operating systems and they were able to talk to each other across the network.

During the 1970's, the ARPANET grew to connect research institutes and laboratories supported by the Department of Defence in various parts of the USA. Many of these were university laboratories studying data processing and computer networks, which developed the TCP/IP network protocol and its applications for the Internet.

During the 1980's, the Internet was still considered to be a temporary system designed for the research world while in the 1984 the TCP/IP data transmission protocol was adopted as the US Department of Defence's official network standard. At the same time that the Internet technology was being experimentally validated and widely used amongst a subset of computer science researchers, other networks and networking technologies were being pursued: USENET, based on a UNIX communication protocols; BITNET (Because It's Time Network), started as a cooperative network at the City

* Corresponding author.

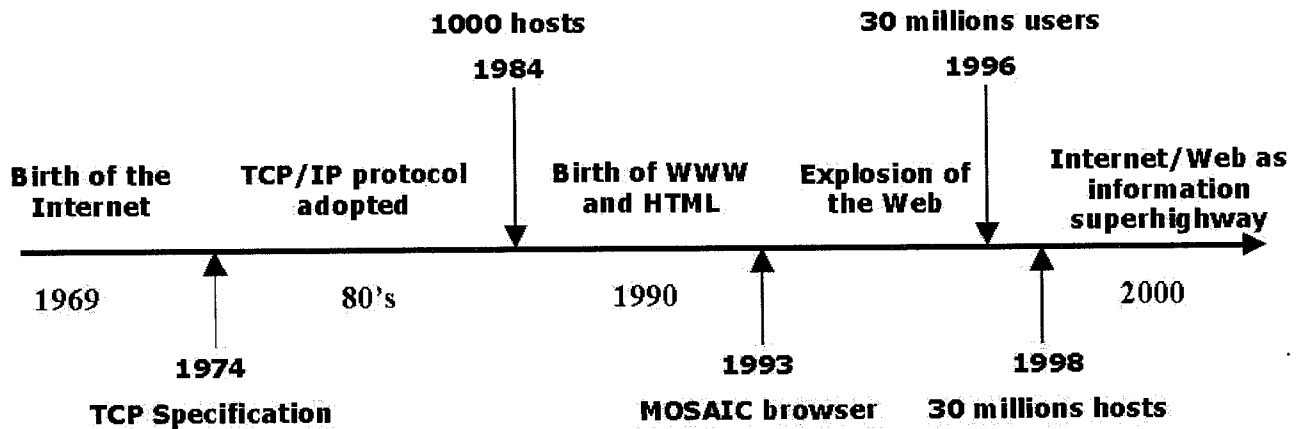


Figure 1: The history of the Internet, summarise in few key points.

University of New York; NSFNET, initiated by the U.S. National Science Foundation (NSF); NSINET, developed by NASA; CSNET (Computer Science NETWORK), developed by a group of computer scientists and some university. By that time Internet was an essential tool for communications, however it also began to create concerns about privacy and security in the digital world and new words, such as "hacker" or "cracker" were created.

In 1990 ARPANET ceased to exist while the country already connected to NSFNET were 28 with ca 300.000 hosts. The NSFNet soon connected to the CSNET, which linked Universities around North America, and then to the EUnet, which connected research facilities in Europe.

Tim Berners-Lee, researcher at CERN of Geneva, noticed that many people were having difficulties in sharing information due to a range of different network information retrieval protocols and a range of workstation types with widely varying display capabilities. For this reason he proposed an Internet-based hypertext system which would have linked together behind a single and easy-to-use interface the various information spread around the Internet. He produced a WWW browser-editor which reads HyperText Markup Language (HTML) documents from Uniform Resource Locator (URL) addresses: i.e. the Web was born [Berners-Lee, T., 1990].

In the next years the Internet Society (ISOC) was founded (1991); the World Bank goes on-line (1992); Mosaic, the first commercial graphical Web browser was released by Eric Bina and Mark Andreessen (1993); "Yahoo" (Yet Another Hierarchical Official Oracle) was invented by two PhD students from Stanford University (1994); NSFNET reverted back to a research project and left the Internet in commercial hands. And in 1995, The Federal Networking Council (FNC), in consultation with the leadership of the Internet and Intellectual Property Rights Communities, unanimously passed a resolution defining the term *Internet* [FNC Resolution, 1995]. Figure 1 shows a brief history of the Internet, from its origin to our time. As the Internet celebrates its 30th anniversary, the military strategies that influenced its birth become historical footnotes. Approximately 500 million people (in almost 250 countries around the world) were already connected to the global net and the traffic on it expanded at a 340% annual growth rate. The number of computer hosts approached 100 million and the Internet passed from a Cold War concept to the Information Superhighway. All the rest is no more history: as new generation will grow up accustomed to communicate using a keyboard, life on the Internet will become an increasing important part of our life.

2.2 How does the Internet work?

Since 1969, the Internet has grown from four host computers in U.S. to millions of computer worldwide. However, even if nobody owns the Internet, it is monitored and maintained in different ways. There are organisations that define how we use and interact with the Internet. The ISOC, Internet Society [ISOC], oversees the formation of the policies and protocols. The World Wide Web Consortium [W3C, 2004], created in 1994, tries to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. The Internet Corporation for Assigned Names and Numbers [ICANN, 2004], a non-profit corporation formed in 1998 to assume responsibility for the IP address space allocation, protocol parameter assignment, domain name system management and root server system management, all functions previously performed under U.S. Government contract by IANA (Internet Assigned Number Authority) [IANA, 2004] and other organisations. They all assume responsibility for technical function, provide reliable information about the Internet and develop interoperable technologies (specifications, guidelines, software and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding.

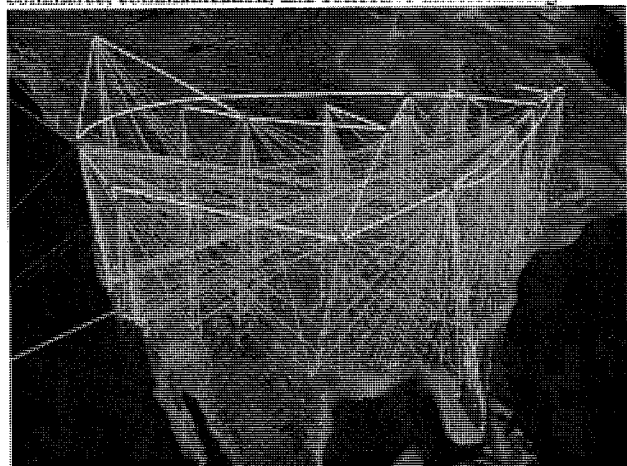


Figure 2: A visualization of the Internet traffic in north America. The colored lines represent virtual connections from the network sites to the backbone [NCSA].

In general, all the machines on the Internet can be divided in two types: servers and clients. Those machines that provide services (like Web servers or FTP servers) to other machines

are servers. And the machines that are used to connect to those services are clients. When a html page is requested to a server, the browser forms a connection to a Web server, requests a page and receives it. In detail (Figure 3), the browser breaks the URL into 3 parts: the protocol ("http"), the server name ("www.isprs.org") and the file name ("isprs.html"). The browser communicated with a domain name server (DNS) to translate the server name "www.isprs.org" into an IP numerical address, which it uses to connect to the server machine. The browser then forms a connection to the server at that IP address on port 80 and following the HTTP protocol, it sends a GET request to the server, asking for the file "http://www.isprs.org/isprs.html". The ISPRS server then sends the HTML text to the browser, which formats the page onto your screen.

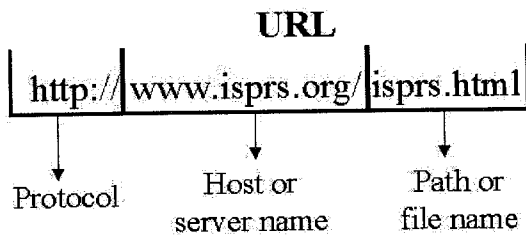


Figure 3: An Internet address (or URL) subdivided in its main parts.

This big network is design in a way that each information can reach its destination using many different paths. When the information is sent through the network, it is split into tiny packets and each packet use a different path to reach its destination. When all the packets reach the destination, they are regrouped to form the original information. If one packet does not reach the destination, the receiving site asks for another copy of this packet. The routes covered by the Internet information are not always the shortest ones, but usually the fastest one (Figure 4).

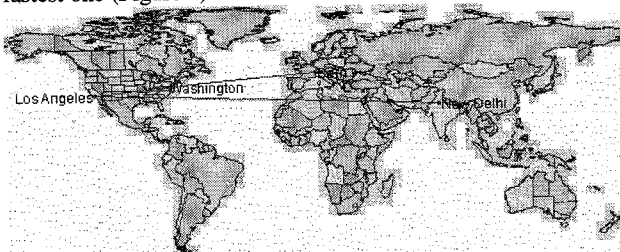


Figure 4: A graphical trace-route of a Internet path covered between a server in Switzerland and the server hosting ISPRS Commission 7 in India. The request passes through US as faster route compared to a Middle East route. Source: [VisualRoute].

2.3 Internet Domain

A domain is a name which identify a web site on the Internet. Domain names have always two or more parts, separated by dots, e.g. 'www.commission5.isprs.org'. The part on the left is the most specific ('commission5'), the part on the right the more general ('isprs'). The extension .org identifies the kind of domain and together with the previous generic part, they are often called *Top-Level Domains* (TLD). There are two types of top-level domains:

- Generic Top-Level Domains (gTLDs), created for use by the Internet public. The most common generic TLD, available since 1984, are: .COM (Commercial), .EDU (Educational), .GOV (US Government), .INT (International Organisations), .MIL (US Dept. of Defence), .NET

(Networks), .ORG (Organisations). On November 2000, ICANN passed a resolution to add seven new generic top-level domains. These new generic Top-Level Domains (gTLDs) are: .BIZ (Business Organisations), .MUSEUM (Museum Organisations), .NAME (Personal), .INFO (Open TLD), .PRO (Professionals as Accountants, lawyers, and physicians), .AERO (Air-transport industry), .COOP (Cooperatives). These were activated in 2001 and 2002. Four of them (.biz, .info, .name, and .pro) are unsponsored while the other three (.aero, .coop, and .museum) are sponsored.

- Country Code Top-Level Domains (ccTLDs), created to be used by each individual country. The Country Code TLD are designed to be assigned to each individual country: it, .fr, .uk, .jp, .ch, .co, .er, They almost follow the ISO 3166 standard that provides codes for the names of countries and dependent areas, even if there are some exceptions. The ISO codes EH (Western Sahara), KP (North Korea), TL (East Timor), CS (Serbia and Montenegro) and AX (Aland Islands), although theoretically available as ccTLDs have never been assigned and do not exist in DNS. On the other hand, eight ccTLDs are currently in use despite not being ISO two-letter codes, namely AC (Ascension Island), GG (Guernsey), IM (Isle of Man), JE (Jersey), SU (Soviet Union), TP (East Timor), UK (United Kingdom) and YU (Serbia and Montenegro). As April 2004, there are 243 country code domains (see [IANA] for an updated list). CcTLD managers, who are also responsible for the operation of the domain, develop the rules regarding who is entitled to domains the ccTLD. But in some cases, anyone in the world can acquire a domain in the ccTLD list, like in the case of Armenia (.AM), Austria (.AT) Cocos Islands (.CC), Germany (.DE), Niue (.NU), Samoa (.WS), Tonga (.TO), Tuvalu (.TV) and Turkmenistan (.TM). This allows domains names like I.am, start.at and go.to.

In addition to gTLDs and ccTLDs, there is a special TLD, .ARPA, which is used for technical infrastructure purposes. ICANN administers this domain name in cooperation with the Internet technical community under the guidance of the Internet Architecture Board.

According to some Internet statistics, the maximum number of characters in one component of a domain name is 63 while the average number of characters in a domain name is 11.

2.4 Wired Internet Connection

The network allows all the computers to communicate with one another and to send and receive data through a network of transmission lines (Figure 5).

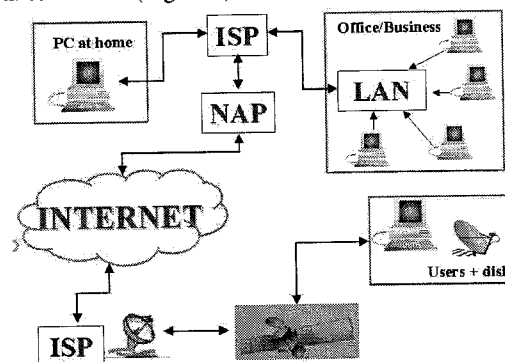


Figure 5: The Internet connection from a home or office computer using wired or satellite mode.

A home computer is usually linked to the Internet using a normal phone line and a modem that talks to an Internet Service Provider (ISP). A computer in a business or university is usually connected to a Local Area Network (LAN) inside the business through a higher-speed connection, like ADSL, ISDN. Usually switches are also used, to provide the different nodes of a company's LAN different connection. The LANs are then attached to an ISP. ISP can be connected to bigger ISP or, using high-bandwidth connections, directly to NAPs (Network Access Point) through communications networks called 'backbones'. The backbones are provided by companies such as AT&T, GTE, IBM, MCI, Netcom, Sprint or UUNET and consist of high-speed links in the T1, T3, OC1 or OC3 ranges (Figure 6). Backbones around the world are connected through world-wide fibre-optic lines, undersea cables or satellite links. In this way, every computer on the Internet is connected to every other computer on the Internet all over the world.

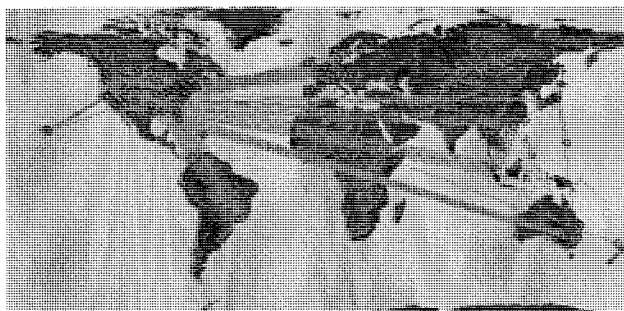


Figure 6: The world-wide backbones provided by UUNET [UUNET].

The Internet connections can be divided in low, medium and high-bandwidth. As reported in Table 1, old modems were very common in the 70's and they were considered high-speed connection (<10 Kbps). Actual modems operate between 14.4 Kbps and 56.6 Kbps, with mild success, at least in one direction. They are very common for Internet connections at home. Recently there has been a lot of effort to develop a higher speed connection for residential users and small office by using an Integrated Services Digital Network (ISDN). ISDN operates at a minimum speed of 64 Kbps on normal copper phone line. For this reason, it is equally available to home and business customers and provides a huge improvement in access speed at only a fractional increase in cost. Data are transmitted in a digital format and the "Integrated" part of ISDN's name refers to the combining of voice and data services over the same wires. A T1 connection is a form of digital, leased private line, which means that a company can lease a point-to-point circuit at a flat rate with a telephone company. A T1 connection allows using the line in the form of 24 channels, running at 64 Kbps each. T1 runs on fibre optic or copper wires. ADSL (Asymmetric Digital Subscriber Line) uses a common phone line and can transfer data using a particular modem with a maximal speed of 6 Mbps. ADSL is getting very common, it does not require new wiring and the phone line for voice call can be used even if an Internet connection is still open. The word *Asymmetric* is used because these modems send data faster in one direction than they do in another. ADSL (and in general DSL technology) exploits the 'extra capacity' of copper wires to carry the information, without disturbing the voice conversation and matching particular frequencies to specific tasks. Following ADSL, the speed jumps to T3, OC1 (Optical Carried level 1) and to the newer version of ISDN, called Broadband ISDN (B-ISDN) which can run at 155 Mbps. Many people who have cable TV can now get a high-speed

connection to the Internet from their cable provider. Cable modems compete with technologies like ADSL and, unlike these, its performance doesn't depend on distance from the central cable office. Another new type of Internet connection can be realised using satellite. Satellite connection can reach a download speed up to 10X faster than dial-up connection and work using normal antenna for Sat-TV.

	Connection	Speed	Time
Low Bandwidth	Old modem	2400 bps	≅ 1 hour
	'Home' modem	< 56.6 Kbps	≅ 4 min.
Medium Bandwidth	ISDN	64 Kbps	≅ 2 min.
High Bandwidth	T1, HDSL2	1.5 Mbps	≅ 5 sec.
	ADSL	6 Mbps	≅ 1.3 sec.
	T3, OC-1	45 Mbps	≅ 0.2 sec.
	B-ISDN	155 Mbps	≅ 0.05 sec.
	OC-XX	> 1 Gbps	< 0.01 sec.

Table 1: Mostly used Internet connections with respective speed and approximately time for downloading the same document.

Despite all these name and speeds, the connection is still a problem for many Internet users: many ISPs that should maintain the "backbones" for an entire nation still work with medium bandwidth connections. And this fact can create many 'stall' problems for local users or for persons who wants to view HTML documents from these countries but has to fight with download speed of 20 Kbps!

2.5 Satellite Connection

As Internet traffic continues to grow world-wide at exponential rates, many Internet Services Providers (ISPs) are facing the problem to keep high-speed connections for their users. The satellite connection between a Network Access Point (NAP) and the Internet backbone can solve this problem. In fact, Internet satellite connections, started at the end of the '90, can assure Internet access whatever the traffic. Moreover, it doesn't matter where the customers are as satellite communication has can deliver bandwidth exactly where and when it is required, without geography and local infrastructure limitations. The satellite connection does not use telephone lines or cable TV systems and, depending on the user's location and requirements, is the best method for skip or extend the terrestrial fibre optic network. In particular, the satellite service is aimed for ISPs or businesses located in areas poorly served by the wired Internet infrastructure. It is the case of the central Asia of Africa, where it could be the (high-speed) transport medium for information and data (Figure 7).



Figure 7: SkyVision World Coverage Map [Sky-Vision] (left). Ses American Internet satellite fleet [SES American] (right).

There are two ways to have an Internet satellite connection: the first possibility is to perform a connection using the same dish that allow you to receive the satellite television; this solution requires you to have an ISP for a dial-up or cable modem for the data you send to the Internet. The second possibility to have a satellite Internet connection is to use a two-way satellite dish

(upload and download). Upload speed is about one-tenth of the 500 kbps download speed. Cable and DSL have higher download speeds, but satellite systems are about 10 times faster than a normal modem. The dish must have a clear view to the south, since the orbiting satellites are over the equator area. Moreover, like satellite TV, trees and heavy rains can affect reception of the digital signal.

Global satellite capacity isn't being used fully. There is enough satellite capacity in the sky to meet the health and education objectives and requirements of every country on earth.

The use of satellite resources for Internet Backbone and other Internet related applications would improve the connections and the capacities of many domains, in particular in developing countries.

2.6 Wireless Internet

Wireless simply means the use of radio-frequency spectrum to transmit and receive voice, data, and video signals for communications.

Regarding wireless Internet, we have to distinguish between laptops and mobile devices (like cell-phone or PDAs). The latter use Wireless Application Protocol (WAP) to access the information on the web. A web site accessible with mobile devices is created with text-only or with low-graphics version and it is generally written in Wireless Markup Language (WML). Only the new generation of PDAs allow the full access to HTML pages [ThunderHawk, 2004]. During the transmission, the data is sent in HTTP form from a web server to a WAP gateway. This system includes the WAP encoder, script compiler and protocol adapters to convert the HTTP information to WML. The gateway then sends the converted data to the WAP client on your wireless mobile device.

Laptops with wireless Internet connection can have Internet access through a Wireless Internet Service Provider (WISP) organization, e.g. an Internet provider that allows subscribers to connect to a server at designated access points (called 'hot spot') using a wireless connection such as Wi-Fi (short for "wireless fidelity"). Wi-Fi is the common term for a high-frequency wireless local area network (WLAN). The Wi-Fi technology is rapidly gaining acceptance in many companies as an alternative to a wired LAN.

The WISPs offer high-speed communication and allow subscriber computers to access the Internet and the web from anywhere within the zone of coverage (several km) provided by the server antenna.

A wireless Internet connection is usually faster than cable or DSL mode, the user is always-on connection and he has the freedom to move around without losing the connection.

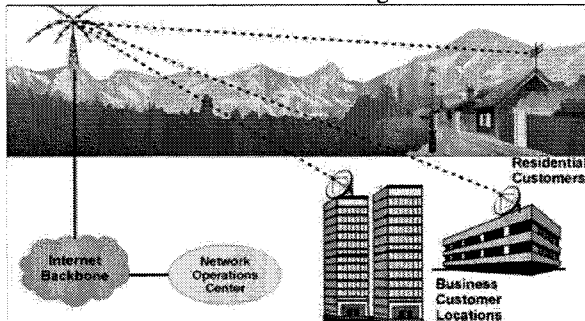


Figure 8: Wireless Internet connection. Source: [Logical Net's].

2.7 Services on the Internet

The World Wide Web is often identified with the Internet, but it is only the most known part of it. Inside the Internet 'live' many other services developed to facilitate the sharing of information through the net. In the following, the main and mainly used services are shortly described.

FTP: the File Transfer Protocol (FTP) is a common format for transmitting and moving data (files) on the Internet.

GHOPE: a system that pre-dates the World Wide Web for organising and displaying files on Internet servers.

TELNET (and SSH): a terminal emulation program used to connect to remote Internet servers.

E-MAIL: short for electronic mail, it is a transmission of electronic messages over networks.

NEWSGROUP (or mailing list): a very big number of people that receive simultaneously an e-mail.

USENET: it is an international network of newsgroups, which are discussion forums that cover many interest groups.

ICQ: it stands for "I-see-you" and it is a revolutionary, user-friendly Internet tool to communicate with other people in real time. It includes the **IRC (Internet Relay Chat)**, system developed in the late 80's as a virtual meeting place where people from all over the world could meet and talk.

WORLD WIDE WEB: it is the most known and used part of the Internet where the users look for information, read articles or see animations.

2.8 Firewall

A firewall is a system or group of systems that enforces an access control policy between two networks. The actual means by which this is accomplished varies widely, but in principle, the firewall can be thought of as a pair of mechanisms: one that exists to block traffic, and the other that exists to permit traffic. Some firewalls place a greater emphasis on blocking traffic, while others emphasize permitting traffic. Probably the most important thing to recognize about a firewall is that it implements an access control policy. If you don't have a good idea of what kind of access you want to allow or to deny, a firewall really won't help you. It's also important to recognize that the firewall's configuration, because it is a mechanism for enforcing policy, imposes its policy on everything behind it. Administrators for firewalls managing the connectivity for a large number of hosts therefore have a heavy responsibility.

2.9 The information on the Web

One of the main problems of Internet users is where to find the right information they are looking for in the less time. Since the 80's, special websites have been created to help the users finding all the information hidden in million of web pages. These sites contain research systems that work with different mechanisms and can be divided in search engines and online directories. Another common resources used to stored information are the Internet "Yellow Pages" Books, which list Internet addresses by categories and can be found in computer and bookstores.

2.9.1 Internet Search Engines

Internet search engines are special web sites designed to help people find information stored on other sites. Before the Web became the most visible part of the Internet, there were already search engines in place to help people find information on the Net. Programs with names like "gopher" and "Archie" kept indexes of files stored on servers connected to the Internet, and

dramatically reduced the amount of time required to find programs and documents. Early search engines held an index of a few hundred thousand pages and documents and received maybe one or two thousand inquiries each day. Today, a top search engine will index hundreds of millions of pages and responds to tens of millions of queries per day. There are differences in the ways various search engines work, but they all perform three basic tasks (Figure 9):

1. they search the Internet or select pieces of the Internet, based on keyword;
2. they keep an index of the words they find and where they are;
3. they allow users to look for words or combinations of words found in that index.

To find information on the hundreds of millions of Web pages that exist, a search engine employs special software robots, called *spiders* or *crawl*, to build lists of the words found on Web sites. When a spider is building its lists, the process is called Web crawling. In order to build and maintain a useful list of words, a search engine's spiders have to look at a lot of pages. When a typical search engine spider looks at an HTML page, it took note of the words within the page and also where the words were found. Words occurring in the title, subtitles and other positions of relative importance were noted for special consideration during a subsequent user search. Each spider takes different approaches but they are always crawling, because of the constantly changing nature of the Web.

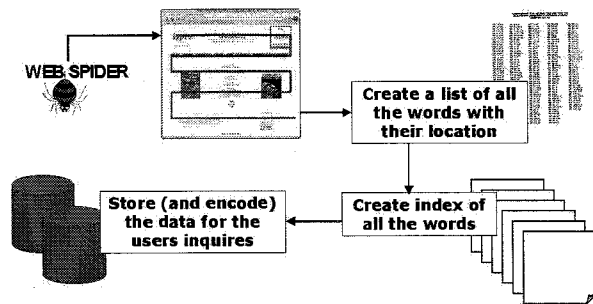


Figure 9: The tasks performed by a good search engine.

These different approaches usually attempt to make the spider operate faster, allow users to search more efficiently. Once the spiders have completed the task of finding information on Web pages, the search engine must store the information in a way that makes it useful. Most search engines store more than just the word and URL. An engine might store the number of times that the word appears on a page and might assign a "weight" to each entry. Each commercial search engine has a different formula for assigning weight to the words in its index. This is one of the reasons that a search for the same word on different search engines will produce different lists, with the pages presented in different orders.

Searching through an index involves a user building a query and submitting it through the search engine. The query can be quite simple, a single word at minimum. Building a more complex query requires the use of Boolean operators that allow you to refine and extend the terms of the search. Each query is analysed and searched in the database of the system. The result of the research is a collection of URL with an associate score (determined by the number of times the search criteria is found in each page), and it is displayed in order from the highest score to the lowest. Some of the most popular search engines are Google (<http://www.google.com>), Altavista (<http://www.altavista.com>), Yahoo (<http://www.yahoo.com>), HotBot (<http://www.hotbot.com>), Lycos (<http://www.lycos.com>), Excite (<http://www.excite.com>), MSM (<http://search.msn.com/>). Some of these search engine entries

present also a main menu organised with directories that can help a user in his research. Elsevier Science has developed a powerful Internet search tool called Scirus (<http://www.scirus.com>). Scirus distinguishes itself from existing search engines by concentrating on scientific content only and by searching both Web and membership sources (articles, presentations, reports). It enables scientists, students and anyone searching for scientific information, locate university sites, and find reports and articles in a clutter-free, user-friendly and efficient manner. Furthermore there are sites, called meta crawler, that use at the same time more search engines to search for a query, as Mamma (<http://www.mamma.com>), Metacrawler (<http://www.metacrawler.com>), Search Engine Guide (<http://www.searchengineguide.com/>).

The qualities of a good search engine web site should be fast answer to the queries, user-friendly, often updated, it should have an advanced research options and a nice display of the results. Few years ago, the search engines were not able to spider all the net and a part of it were hidden to the Internet users. Today this "Invisible Net" is very small as the search engines are more powerful, their databases are often updated and they can recognise also non-text files such as pdf, postscript and other formats.

2.9.2 Online Internet Directories

They are webpages where the information are stored and displayed to the users in thematic channels or categories. Each link is listed with a short description and its URL and these lists can be updated automatically or manually. It is also possible to search inside the web site as a normal search engine. Useful URL are Galaxy (<http://www.galaxy.com/>), Yahoo (<http://www.yahoo.com>), the WWW Virtual Library (<http://www.vlib.org/>), the Educational Virtual Library (<http://www.csu.edu.au/education/library.html>), the Earth Science Portal (<http://webserv.gsfc.nasa.gov/ESD/>). The Earth Science Portal is a searchable links directory together with a web crawler search engine that spans all the web-based information of the NASA Earth Sciences Directorate. AllConferencesNet (<http://www.allconferences.net/>) instead provides interesting links for all kind of events. It is a directory focusing on conferences, conventions, trade shows, exhibits, workshops, events and business meetings. This is a unique search directory that serves users looking for specific information on conference or event information.

2.10 Educational resources on the Web

The possibility to find scientific articles, reports, journals or entire books on the Web is very high. These electronic documents contain nothing different in comparison with the same text and picture of the paper version, except some occasional hyperlink. They are quickly disseminated on the net and everybody can read them. Real e-zines or e-journal have no paper equivalent and are not always free. A big problem of electronic documents is they are not permanent and they can be lost from the permanent record, as subject to changes in positions and unpredictable removal. Instead document on paper or electronic format (CD-ROM) are not transient and can be available and legible for many years, in particular the paper one, which do not require special equipment or knowledge to be read. Therefore to preserve for a longer period also the Internet publication, a timely and active intervention at all stages is required. Educational resources on the web are without limits. A general database of educational material is provided by the "Gateway to Educational Materials" project

(<http://www.thegateway.org>), a consortium effort created to provide easy access to the substantial, but uncataloged, collections of educational materials available on various federal, state, university, non-profit, and commercial Internet sites. Another Internet portal full of resources of relevance to faculties, students, and research staff at the university level is Infomine (<http://infomine.ucr.edu/>). It is a huge database including electronic journals, electronic books, bulletin boards, listservs, online library card catalogues, articles and directories of researchers and many other types of information. More specific resources in Remote Sensing, just to mention few good links, are provided by NASA (<http://rst.gsfc.nasa.gov>), ASPRS (<http://research.umbc.edu/~tbenja1>), CCRS (http://www.ccrs.nrcan.gc.ca/ccrs/learn/learn_e.html) and CEOS CD-ROM (<http://ceos.cnes.fr:8100/>).

Concerning articles, publications and presentations available on the net, they can be easily found using a powerful search engine or through CiteSeer (<http://citeseer.ist.psu.edu/>), a scientific digital library.

2.11 E-Learning

E-learning is the education via Internet, network or standalone computer. E-learning uses electronic applications and processes to teach and learn as the network enables easily transfer of skills and knowledge between users. E-learning applications and processes include Web-based learning, computer-based learning, virtual classrooms and digital collaboration. Content is delivered via the Internet, intranet/extranet, audio or video tape, satellite TV and CD-ROM.

An e-learning web sites collection is provided by the University of Leicester (<http://www.le.ac.uk/cc/rjm1/isp/ele.html>).

2.12 Internet growth and its statistics

It is very difficult to determine how many users and how many domains or hosts are on the net, besides making guesses and estimates. There are many companies that do surveys to estimate the number of users, but we can consider the numbers presented in these surveys to be fairly good estimates of the minimum size of the Internet. And often the results are also in disagreement. Moreover the geographical location of an Internet host is somewhat problematic since a host need not to be located in the country which correspond to its ccTLD; and gTLD has never an explicit geographic designation (e.g. ISPRS, with the server in Zurich, the Headquarter in London and the President in Sidney!). For these reasons is not possible to determine the exact size of the Internet, where host are located or how many users there are.

The growth of the available information can be estimated from the number of registered host (e.g., a computer or machine with a unique IP address on a network). According to the Internet Software Consortium [ISC, 2004], the number of registered hosts states 80,000 in January 1988, 1.3 million in January 1993, 16.1 million in January 1997 and 109.7 million in January 2001 (Figure 10). Internet hosts include network elements such as routers, Web servers, mail servers, workstations in universities and businesses, and ports in modem banks of Internet Service Providers (ISPs). The number of hosts is considered one of the most accurate measures of the size of the Internet, even if their distribution is concentrated mainly in the most developed countries (Figure 11).

In 1997, the OCLC Office of Research started a project aimed at answering fundamental questions about the Web, like how big is it? what does it contain? how is it evolving? The project's objective was to develop and implement a methodology for

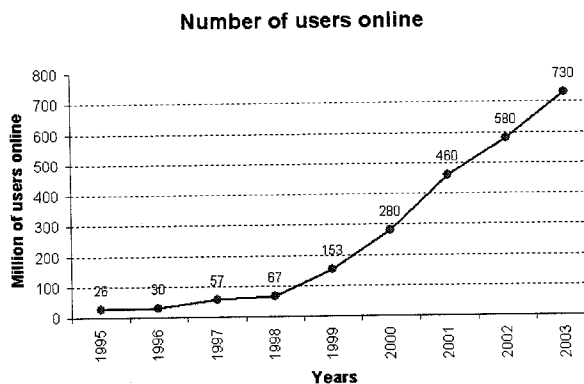


Figure 10: Internet growth represented by the number of Internet hosts. Source: [ISC, NUA].

characterizing the size, structure, and content of the Web, making the results available to both the library community and the public at large. The strategy adopted for characterizing the Web was to harvest a representative sample of web sites, and use this sample as the basis for compute an estimate and make an inference about the Web [O'Neill et al., 1997]. According to OCLC, the web sites can be divided in three categories: (1) public, sites that provide free and unrestricted access to all or at least a significant portion of its content; (2) private, sites whose content is intended for a restricted audience; the restriction can be explicit (e.g., fee payment or authorisation) or implicit (obvious from nature of content); (3) provisional, sites in transitory or unfinished state (e.g., "under construction"), and/or offers content that is, from a general perspective, meaningless or trivial. In the most recent survey (2002) of the OCLC Web Characterization Project [OCLC, 2004], the web contains 9.04 millions of web sites (with 8.7 millions of unique web sites, i.e., the count is adjusted to account for sites duplicated at multiple IP addresses). The growth of unique web site between 1998 and 2002 is ca 231%. OCLC found that 15.5% of the web sites provides information services, 14.2% provides professional, scientific and technical services, 11.8% retail trade and 6.6% provides educational services.

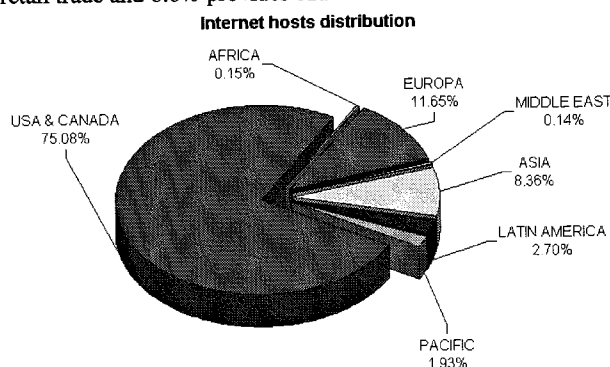


Figure 11: Distribution of Internet hosts. Source [ITU].

Considering the million of users online (Figure 12), the number of people is constantly increasing. In 1995 the Internet population was only 16 million people (app. 0.35% of the world population) while at the middle of 2003 there were more than 730 million people online (11.5% of world population) [NUA]. The number is expected to increase again in the next years and a CIA (Computer Industry Almanac) document reports that in 2007 there will be more than 1400 millions, with 56% of wireless users [CIA].

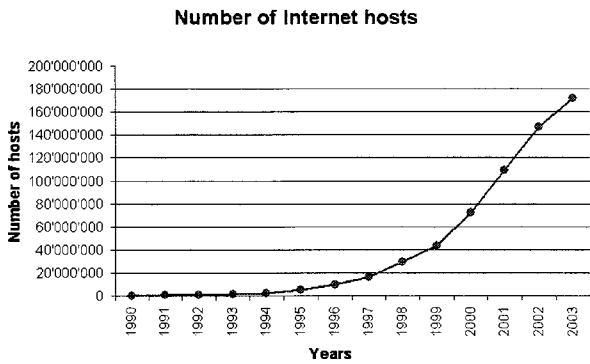


Figure 12: Internet growth represented by the number of users online. Source [ISC, NUA].

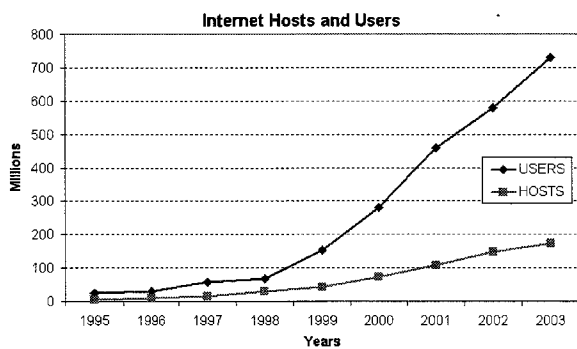


Figure 13: Number of Internet users compared to number of Internet hosts. Source: [ISC, NUA].

Comparing the number of users and host (Figure 13), in 2003 there was an average of 4.2 users per host. With the high quality of service in the United States, there are approximately 1.5 Internet users per host, whereas in some developing there are more than 100 Internet users per host!

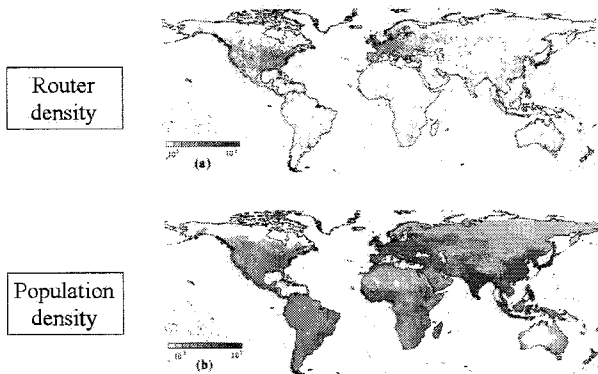


Figure 14: Geographic distribution of Internet routers (device to forward data along networks) against the global distribution of population as in 2001 [Yook, S. et al., 2001].

Despite these number which are just great estimates of the real values, it is widely expected that the Internet population and the number of domains will continue to grow, at least for other few years. But the presence and the access to the Internet is not well distributed in all the country (Figure 14 and Figure 15). As reported by Nua [NUA], the top 10 countries account more than 80% of the world-wide population; moreover, as shown in Table 2, there are country with 0.01% of the country population who has an Internet access while there are European country with more that 60%.

	Max %	Min %
Africa	11.2% (Seychelles)	0.01% (Liberia, Congo)
Asia	59.5% (Hong Kong)	0.11% (Bangladesh)
Pacific	54.8% (Australia)	2.58% (Vanuatu)
Europe	69.8% (Island)	0.34% (Albania)
Middle East	36.8% (UAE)	0.05% (Iraq)
Canada & USA	59.1% (USA)	39.7% (Bermuda)
Latin America	34.1% (Aruba Island)	0.42% (Haiti)

Table 2: Percentage of users online according to the different parts of the world. Source: [NUA].

Online users worldwide

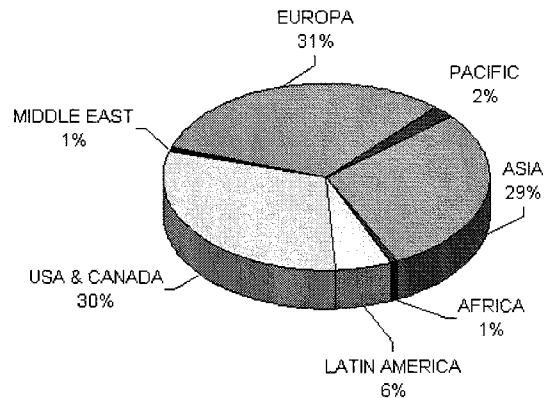


Figure 15: World-wide online population. Source: [NUA].

Considering the demographics of the Internet users, different sources report an average age of the users around 38 years old, with a percentage of female users of 35% and male users of 65%. Furthermore, the Internet usage habits state that 93% of the users use a web browser more than once a day, for personal information (75%), work (65%), education (60%), shopping (50%), while 28% of the users buy online once a month. Moreover, 35.8% of the online population has English as native language, 14.1% Chinese, 9.6 Japanese and 9% Spanish [GR]. Considering the content of the web sites (Figure 16), English language is the most used (72%), followed by German (7%) and Japanese (6%) [OCLC].

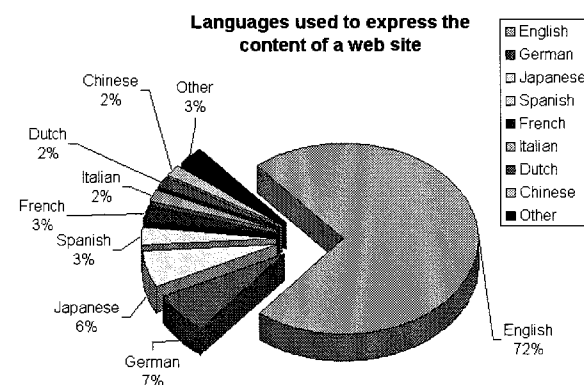


Figure 16: Languages used in the web sites. Source: [OCLC].

2.11 The SPAM messages

"In a single day of May 2003, No. 1 Internet service provider AOL Time Warner (AOL) blocked 2 billion spam messages from hitting its customers' e-mail accounts. Microsoft (MSFT), which operates No. 2 Internet service provider MSN plus e-mail service Hotmail, says it blocks an average of 2.4 billion spams per day" [Buseiness Week Magazine].

SPAM messages or 'bulk e-mails' are becoming one of the most critical problem for anyone who receives e-mails. They are annoying messages containing advertisements and unwanted information that every day fill our mailbox.

The name comes from "SPAM (Shoulder Pork and hAM, SPiced hAM) Luncheon Meat", canned ham produced by Hormel Foods [SPAM, 2004]. For many users, the history of SPAM is very short, but according to some persons, the first SPAM was sent on May 1978, during the Arpanet period, inviting people to a reception. The term got really popular in 1994, when two messages with subject "Global Alert for All: Jesus is Coming Soon" and "Green Card Lottery - Final One?" were posted to every single newsgroup on USENET, the world's largest online conferencing system at that time. There were several thousand such newsgroups, and each one got the advertisement. Soon people called it SPAM and the word started to be used for these kinds of unwanted and multiple posted messages [Templetons, 2004]

Spammers use software to extract names and e-mail addresses automatically from newsgroups or chat rooms; moreover program called *spambots* spider the web looking for the @ sign, that represents an e-mail address. The largest of these companies are able to send billions of spam messages per day. Spam filtering software is available on the market to stop unwanted e-mails. They search keywords like "viagra", "sex", "win", etc., in the text or subject of the e-mails and stop them. Unfortunately these filters can sometimes delete messages that we really want to receive.

The strongest war against spam messages is the elimination of e-mail address: then, if you want to send an e-mail, you have to fill out an online form. Many business companies and the White House of the United States itself, were forced to use this way.

ISPRS ON THE INTERNET

The International Society for Photogrammetry and Remote Sensing is a non-governmental organization devoted to the development of international cooperation for the advancement of photogrammetry and remote sensing and their applications. The Society operates without any discrimination on grounds of race, religion, nationality or political philosophy. It was created in 1910 as International Society for Photogrammetry (ISP), under the leadership of its first President, Eduard Dolezal. After 70 years of functioning under its original name, the Society changed its name in 1980 to the International Society for Photogrammetry and Remote Sensing (ISPRS). Since 1994, ISPRS is online, inserted in the global network to provide information about its activities. In 1996, an ISPRS Working Group (WG VI/4) was established especially for the Internet, and was chaired by Tuan-chih Chen (1996-2000).

3.1 ISPRS Homepage

The ISPRS homepage has turned out to be one of the most important components of ISPRS communications. It provides up-to-date information about the society and links its various activities.

The first html pages regarding ISPRS were inserted on the Internet by Andre' Streilen in December 1994, on a server hosted at ETH Zurich, under the URL <http://www.geod.ethz.ch/isprs>. In September 1999, with to the reservation of the international domain name 'isprs.org', all the information were moved to a meaningful URL and nowadays all different ISPRS activities can have an Internet address ending on the suffix 'isprs.org'. After a short period in T.U.

Delf, in September 2000 the ISPRS server moved back to ETH Zurich where it is actually hosted.

In April 2004, all the information regarding ISPRS are listed in ca 700 HTML pages with approximately 25'000 lines of code; moreover there are ca 1000 PDF files, i.e. a total of ca 2 GB of data available.

3.2 ISPRS Publications

The publications of ISPRS are divided in seven categories: The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, The ISPRS Journal of Photogrammetry and Remote Sensing, the ISPRS Highlights, the Annual Report, the Silver and Blue Book and the ISPRS Brochure.

An updated list of the International Archives is available at <http://www.isprs.org/publications/archives.html>, with links to the online proceedings.

Full articles published in the ISPRS Journal are instead available at Elsevier Science web pages back to 1965 (<http://www.elsevier.nl/locate/isprsjpr>). Online access to full text articles is available to those readers whose library has subscribed to ISPRS Journal via ScienceDirect Digital Collections, or has a current print subscription to ISPRS Journal and has registered for ScienceDirect Web Editions.

3.3 Educational Resources and Job Opportunities through ISPRS web site

The ISPRS educational page tries to collect the wide gamma of educational material and software for Photogrammetry, Remote Sensing and GIS available on the Internet (<http://www.isprs.org/links/tutorial.html>).

It is not a complete list, but some pointers are listed about:

- *Free software*, in particular from CATCON, the Computer Assisted Teaching contest organised by WG VI/2 (<http://www.isprs.org/catcon>). The main objective of the CATCON contest is to promote the development and dissemination of multimedia products, educational information and simulation packages for computer assisted teaching. In general, material submitted by contestants should be non-commercial and provided free of charge for not-for-profit use.
- *Education, training, research and fellowship opportunities in Remote Sensing, GIS and its applications*. (<http://www.ltid.inpe.br/dsr/tania/Rsdir/>). It is an educational Directory that has been developed in the period 1996-2000 as a task of ISPRS TCVI/WG1 on education and as part of the ISPRS Educational Opportunities Program. It is a first attempt to providing a comprehensive directory of education and training services in the remote sensing and spatial information sciences. The Directory was developed from an original document prepared some years ago by the UN Office of Outer Space Affairs in Vienna. In this directory it is possible to get information from all members states that are involved in Space Science. The information contained in this directory for each institution includes its areas of specialisation, the educational and research programmes offered, the facilities available, the prerequisite qualifications, financial information, fellowship opportunities and opportunities for international cooperation. This Directory is necessarily incomplete because of the difficulty in obtaining accurate and timely information about all education institutions around the world in a range of languages. Therefore education

institutions are encouraged to provide their new or updated details of education and training programs in the remote sensing and spatial information sciences.

- *Tutorials* in Photogrammetry, Remote Sensing and GIS
- *News* about satellite missions and launches
- *Glossaries and Acronyms* used in Remote Sensing, GIS, Radar and Cartography
- *Journals* of Photogrammetry, Geodesy and Remote Sensing Working Group VI/1 provides also a big database of education-related links, including training opportunities, online publications and journals, continuing education courses, educational institutions, free software, missions and instruments information (<http://www.commission6.isprs.org/wg1/>).

The ISPRS web site contains also a list of job opportunities. The *Employment Opportunities archives* is a jobs listing intended to provide offers for people who are seeking an employment (http://www.isprs.org/job_opportunities/). Messages are posted via e-mail and last until when the position is taken. Employers and universities are encouraged to submit advertisement in the ISPRS archive.

Educational resources and job opportunities can also be found in the *Academic Sites for Geomatic Engineering*, maintained by Robert Kauper. It is a collection of links to international academic institutions that provide education in the field of Geodesy, Photogrammetry, Surveying, Cartography and GIS (<http://www.lrz-muenchen.de/~t5831aa/WWW/Links.html>).

3.5 ISPRS Events Calendar

The ISPRS Events Calendar (www.isprs.org/calendar.html) is one of the most important parts of ISPRS web pages. The Calendar contains a list of all ISPRS and Sister Societies (FIG, ICA, IAG, IHO, IGU, etc.) sponsored and co-sponsored workshops, symposia, tutorials, conferences, congresses, and other meetings. It also contains details of all international and national events on topics related to the activities of spatial information, photogrammetry, remote sensing, geomatics, surveying, mapping, machine vision, image processing and similar areas. One of the purposes for this Calendar is to allow people of WGs, Commissions and Sister Societies to identify open dates or events, which they may link up with or avoid conflicting with. This Events Calendar was compiled manually by previous editors and contained some incorrect information, or missed some important events. The current Editor (2000-2004) has established an automatic system [Chen, 2002] to search the events information of geo-spatial information, photogrammetry, remote sensing, surveying, geomatics, and GIS on the Internet.

3.6 ISPRS Server statistics

Since the beginning (1995), the ISPRS server statistics are available to analyse the interest of the community. The logfile of the Apache server is examined with Analog [Analog, 2004], a program that analyses servers' files. Figure 17 (upper) shows a monthly report of the number of requests to ISPRS server, in the period January 1995 - April 2004. The increasing interest of the community for the ISPRS homepage is evident. The graph gives a reasonable estimate of the use of the server as only the request for single HTML documents are counted and requests for images, graphics, icons etc. are not taken into account. Moreover these statistics refer only to requests made outside ETH domain, which excludes all the accesses during the maintenance of the documents. The data missing from the figure in the period July-August 2000 is due to the movement of the server from Delft to Zurich.

Figure 17 (lower) shows instead the average request per year. The steady increase of the use of the HTML documents over the years and especially after the registration of ISPRS domain is obvious.

In 1995 the average of monthly requests were 424, in 1998 the average was 5780 while at present ISPRS server has an average of 190 000 requests per month.

The different domain (~country) served at least one by the server were 160 (ca 65% of the registered country code domains); the distinct hosts served (~users) were ca 228 000 with an average of 20Mb of data transferred per day. More than 20 000 different organizations have contacted the ISPRS server with ca 4000 successful requests per day.

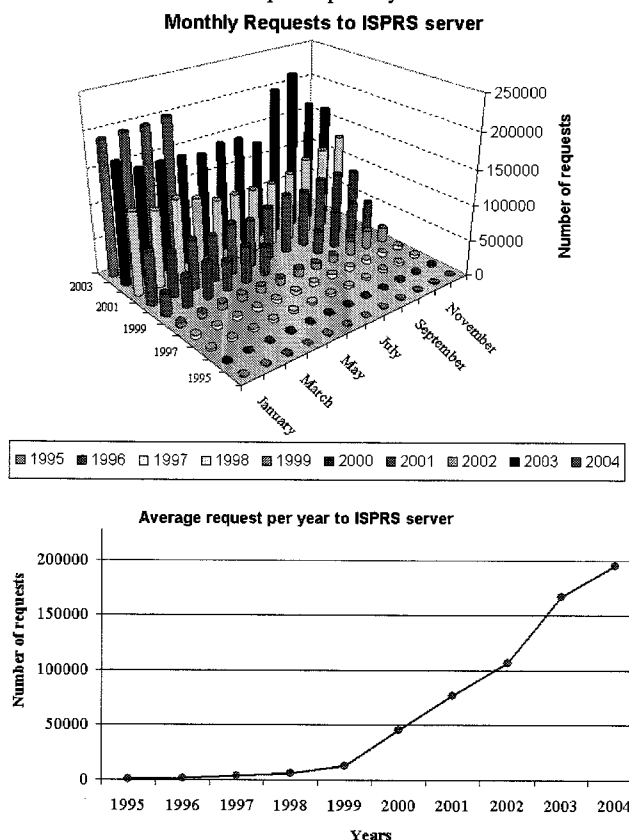


Figure 17: Monthly report (upper) and average requests per year (lower) in the period January 1995-April 2004 on ISPRS server

Another interesting statistic provided by Analog concerns the words and queries used in the search engines to find ISPRS and its related pages: between ca 30 000 search terms, the most used are remote, 'sensing', 'photogrammetry' and 'isprs'. Considering all the queries, the most requested are 'photogrammetry', 'isprs', 'remote sensing', 'International Archive of Photogrammetry and Remote Sensing', 'isprs journal of photogrammetry and remote sensing' and 'orange book'. The browser most used to find information related to ISPRS is Microsoft Internet Explorer followed by Netscape. The majority of the users (67%) has Windows as operating systems; then Unix (Sun + Linux) and Macintosh (Figure 18). The information (directories) more requested contain the publications, the events calendar and the technical commissions while 73% of the consulted files are PDF articles.

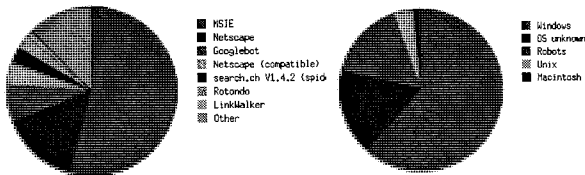


Figure 18: Most used browsers (left) and users operating systems (left) to access the ISPRS web site.

4. INTERNET2 AND NGI

As the Internet grows larger and more crowded, government, scientists, and universities are looking for new ways to send information quickly and powerfully. And for the next generations, many new technologies and developments inside Internet world are ready to be launch. In the future common operation and more interactive applications will be done at home, using an online computer. Two projects have grown out of these needs and are already working: Internet2 [Internet2] and the Next Generation Internet [NGI].

Internet2 is a consortium being led by over 180 universities working in partnership with high-tech companies and government to develop and deploy advanced network applications and technologies, accelerating the creation of tomorrow's Internet. Internet2 is recreating the partnership among academia, industry and government that fostered today's Internet in its infancy. The primary goals of Internet2 are to: (1) create a leading edge network capability for the national research community; (2) enable revolutionary Internet applications; (3) ensure the rapid transfer of new network services and applications to the broader Internet community. Internet2 is not a separate physical network and will not replace the Internet.

The NGI Program is a government project and has been successfully completed and Federal agencies. They are currently coordinating advanced networking research programs under the Large Scale Networking (LSN) Coordinating Group. More than a faster Web or e-mail, both new Internets should develop new and faster technologies to enhance research and communication, and it is expected that both projects will eventually improve the current commercial Internet. Moreover they will enable completely new applications for digital libraries, virtual laboratories, distance-independent learning and tele-immersion.

5. CONCLUSIONS AND OUTLOOK

After 35 years of innovation, data transferring and exchanging, will the Internet with its number of hosts and domain still grow? Which kind of operations or interactive application will we do with an Internet browser? Since 1969, the Internet has grown from a Cold War idea to a superhighway of information. Nowadays we are in the age of the Internet: people communicate, work and plan the holidays over the Internet. Schools use the Internet as a vast electronic library, with untold possibilities. Doctors use the Internet to consult with colleagues half a world away. And even as the Internet offers a single global village and it can create a second class citizenship among those without access. As a new generation grows up as accustomed to communicating through a keyboard as in person, life on the Internet will become an increasingly important part of life on Earth.

There are still part of the world were the use of Internet has yet to grow substantially. These include much of the Asian part of Russia, parts of the Middle East and those part of South East Asia and Africa where it has politically repressed or were the infrastructure are still in development. In this area, satellite connections are really necessary. The absence of Internet from the life of a country is also a lack for the education of the people. Therefore it is expected a growing demand for satellite capacity and links from those parts of the world that may never have full access to international fibre such as most of central Africa. One of the mostly expected explosions is the use of via satellite connection also for residential users, while cable-modem is already common, in particular in US.

Many companies are already using the Internet for Sales, Advertising and Marketing while many web sites are specialized in on-line marketing (e.g. Ebay, Amazon).

It is estimated that the Internet will double in size every year while the WWW will double every 2 months. The World Wide Web and the e-mail service are the greatest success of the Internet. The WWW is an example of a system that started slowly but has grown to huge proportions. For many, it is the user friendly face of the information available on the Internet and has been at least partly responsible for the explosive growth of the Net. The Web, together with the search engines, provides efficient access to an increasing amount of information. But create archives of publications and data is not enough: they must be maintained, updated and managed in a way that users can easily discover and access.

The rapid growth and interest for the WWW introduced many innovations but it has embedded or amplified many problems like security concerns for commercial applications, bandwidth and server saturation, demand for faster access in particular for multimedia data, controlling access to certain types of data, protecting the work of authors (copyright issues), not enough IP addresses to meet the demands. The Internet has opened the access to information to everybody and has changed the way we see the world. But newer and more efficient protocols and ideas are needed to meet the demands for faster access to resources and to solve the problems mentioned before.

A part from any future technological development of the Net, the main source of the Internet remains the people, who use and contribute to make it always bigger. Everyday more people use an online computer to find information, learn, educate and communicate. We have to keep ourselves continuously up-to-date about all new developments and innovations of the Net to really exploit all its capabilities and possibilities.

ISPRS is online since 10 years. Many pages have been created and now there are approximately 2 GB of data available. ISPRS will remain inside the e-world, always improving its appearance. The main board and the web master will continually strive to refine and expand ISPRS online material to provide always more information and services for its users.

REFERENCES

- Analog, <http://www.analog.cx>, April 2004
- Berners-Lee, T., 1990. 'WorldWideWeb: Proposal for a HyperText Project', CERN.
- Business Week Magazine, June 2003.
- Chen, T., 2002: An Automatic Searching and Publishing System on Internet for Geo-Spatial Information. Workshop of ISPRS WG VI/1, Dar es Salaam, Tanzania, CD-ROM.

CIA, Computer Industry Almanac, <http://www.c-i-a.com/>, April 2004.

FNC Resolution, 1995. *'Definition of the Internet'*, http://www.itrd.gov/fnc/Internet_res.html, April 2004.

GR, Global Internet Statistics by Language, <http://glreach.com/globstats/>, April 2004.

Internet2, <http://www.internet2.edu>, April 2004.

IANA, Internet Assigned Numbers Authority, <http://www.iana.org/>, April 2004.

ICANN, The Internet Corporation for Assigned Names and Number, <http://www.icann.org/>, April 2004.

ISC, Internet Software Consortium, <http://www.isc.org/>, April 2004.

ISOC, Internet Society, <http://www.isoc.org/>, April 2004.

ITU, <http://www.itu.int>, April 2004.

Logical Net's, <http://www.logical.net>, April 2004.

Netcraft, Web Server Survey, <http://www.netcraft.com/survey/>, April 2004.

NSCA, Visualization Study of the NSFNET. <http://archive.ncsa.uiuc.edu/SCMS/DigLib/text/technology/Visualization-Study-NSFNET-Cox.html>.

NUA Internet Surveys, http://www.nua.ie/surveys/how_many_online/index.html, April 2004.

OCLC, Online Computer Library Center, Inc., <http://wcp.oclc.org/>, April 2004.

O'Neill, Edward T., Patrick D. McLain, and Brian F. Lavoie. 1997. "A Methodology for Sampling the World Wide Web." Annual Review of OCLC Research.

SES Americom, <http://geamericom.com/sat>, April 2004.

SPAM, <http://www.spam.com>, April 2004.

STARR, STARR Communication Service: <http://www.starrcomm.net>, April 2004.

Sky-Vision, <http://www.sky-vision.com>, April 2004.

Templetons, <http://www.templetons.com>, April 2004.

ThunderHawk, <http://www.bitstream.com/wireless/>, April 2004.

UUNET, <http://global.mci.com/uunet>, April, 2004.

VisualRoute, <http://www.visualware.com>, April, 2004.

W3C, World Wide Web Consortium, <http://www.w3.org>, April 2004.

Yook, S., Jeong H., Barabasi A., 2001: Modeling the Internet's Large-Scale Topology. Available at: <http://arxiv.org/abs/cond-mat/0107417>.

MODELING AND VISUALIZATION OF CLOUDS FROM REAL WORLD DATA

A. Roditakis

ETHZ, Federal Institute of Technology Zurich, CH 8093 Hoenggerberg Zurich, Switzerland
(roditak@geod.baug.ethz.ch)
Commission V, WG V/6

KEY WORDS: Photogrammetry, Meteorology, Modeling, Visualization, Graphics, Hardware, Rendering, Point Cloud.

ABSTRACT:

Visualization of gaseous phenomena, such as smoke, fire or clouds has been a challenging issue in computer graphics due to the complicated geometry of such natural objects. In the early 80's, approaches that faced this problem tried to simplify the representation of such geometry, using simple particles or ellipsoids. During the last decade work was done on rendering fractal volumes and random textures in order to create a realistic turbulent image of random generated volumes, and in parallel, modeling of soft or 'blobby' objects (also known as 'metaballs') was combined with hardware accelerated volume rendering and gave some first convincing images of 3D gas volumes in near realtime. There has been large technological developments in the computer graphics hardware and a lot of work has been done from the computer graphics community on simulation of clouds, still there is not much done in the direction of modeling and rendering of such objects from real world measurements, and work is lacking on the issues that arise from datasets with incomplete spatial or temporal resolution.

In the presented work we use the modeling technique of metaballs based on cloud bottom and top height measurements taken during the CLOUDMAP2 project, in order to construct a 3D volume from point clouds. We present the problems that arise from fitting the volume to the measured points, combination of cloud top and bottom height estimations with different resolutions and the generation of large volume data. Solutions to these problems are presented which include the calculation of cloud volumes, interpolation with cloud top height estimations from simultaneous ground based and satellite observations and finally we present various methods to render such volumes using hardware and software assisted techniques.

1. INTRODUCTION

1.1 Motivation and aim of work

The work presented in this paper, was conducted in the frame of a EU project called CLOUDMAP2. It targeted on the parameterization of macro- and micro-physical properties of clouds through the combination of space-borne and ground based sensors. This fusion of sensors helps to calibrate and validate satellite products that could provide global coverage for the cloud location and distribution, and thus help us understand better the role of clouds in the interaction between Earth's climate and anthropogenic inputs. The final aim of our work was to create a processing chain that would take the cloud top and bottom heights estimations as input, and deliver a three dimensional visualization of cloud fields, that would provide us with an instant image of the cloud shape and distribution over measured areas. Furthermore we want to underline the difficulties that come from using real world data for cloud modelling and rendering, where we witness blunders, errors and areas with missing values. These problems do not appear in case studies for cloud rendering, since we notice that for the demonstration of new modelling and rendering techniques, authors use artificially generated data with uniform resolution and coverage (*perfect data*).

1.2 Overview

Within the EU project Cloudmap2, several measurement campaigns were conducted in the area of Switzerland, where simultaneous measurements from ground digital cameras and satellite sensors were conducted, in order to estimate cloud bottom and top heights and winds. From the bottom side, a configuration of digital cameras provides stereo cloud bottom heights photogrammetrically, while at the same time a satellite passes over the area acquiring images, which provide us with stereo top heights. As the focus of the present work lies on the

modelling and visualization of these measurements into clouds, details on the methods used in the acquisition step can be found in the published work of other participants of the CLOUDMAP2 project (Seiz, 2002), (Mayer et. al, 2003).

Starting from the cloud bottom heights we cleaned the dataset from points that did not belong to clouds, and as a next step we experimented with modelling the cloud bottom height, using polygonal modelling and texturing to simulate the appearance of clouds. Similar work had been done by (Lee, 1996) and we also concluded that this method presents limited options for 3D cloud fields. We continued on the path of previous work done on visualization of fluids and applied some basic volume rendering techniques. In the modeling aspect we tried two different options for a definition of a 3D cloud field, the first being a straightforward 3D regular grid calculation, and the second influenced by the work from Nishita and others (Nishita, 1996) on soft body modelling using 'meta-balls', also known as 'soft-balls'. These two methods interpolated the distinct points over a 3D grid, with the second having the advantage of smooth blending of neighbouring points.

Rendering of this three-dimensional cloud field was also performed by utilizing two different techniques in parallel. We tested the well known method for visualizing three-dimensional iso-surfaces called 'marching cubes', which provided us with rendering of a surface of the cloud field, with a chosen iso-parametric value, and in parallel we implemented a hardware accelerated volume rendering technique, using OpenGL language and textured two-dimensional slices of the volume. Finally we made use of an existing software library, used at the National Center for Atmospheric Research, at Colorado, USA (Lacroute, 1994), (NCAR URL) that although doesn't deliver real-time rendering of the volume field, its high performance in combination with the realistic lighting calculations, gave until now the best results as far as visual quality is concerned. In order to demonstrate the difference of an extrapolated 3D cloud

field, from a cloud field that is present in all three dimensions, we used the same processing chain both in the cloud bottom field, and the local alpine numerical weather prediction model, used from the Swiss Meteorological Office (MeteoSwiss).

1.3 Related work

The problem of modelling and rendering gaseous phenomena has been an active topic since 20 years now either as a problem not solved in computer graphics at that time (Blinn, 1982), or in more practical approaches in the field of cinema special effects (Reeves, 1983). Point based rendering techniques have given a simple, yet flexible and powerful tool for rendering. Levoy and Whitted extended the use of points further than rendering smoke or fire, into traditional geometry visualisation (Levoy, 1985). The last few years were followed by an increase in applications demanding visualisation solutions, together with huge developments in computer hardware. In (Rusinkiewicz, 2000), (Zwicker, 2001) points in combination with texture splatting were used, while Harris (Harris, 2001) used particles and impostors to deliver high frame rates for scenes with many cloud objects. Nishita and others (Nishita, 1996), (Dobashi, 2000) used a particle system to control the metaballs model that composed the cloud objects and in combination with hardware accelerated OpenGL programming, they achieved impressive results of rendering and animation of clouds in near-real-time (Figure 1).



Figure 1: Clouds modelled using metaballs (from Dobash et. al. 2000)

Alternative methods like the planar and curved surfaces with texture used by Gardner (Gardner, 1985) and textures with noise and turbulence functions (Ebert, 1990), taken from the work of Perlin (Perlin, 1985) provided new perspectives, which were difficult though to apply to real world measurements. Volume rendering has undergone many improvements on the speed of algorithms and the work of Lacroute at the University of Stanford (Lacroute, 1994), (Volpack URL) and its development by Schulze (Schulze, 2001) attracted our attention and part of it was included in our implementation.

2. MODELING AND RENDERING OF CLOUDS

2.1 Initial tests

We started our tests on the ground-based (GB) measurements, creating a triangulation of the cloud bottom height (CBH) surface. Afterwards we used the GB images in combination with the point cloud for determining whether the point belonged to a cloud or not (Figure 3). We first performed an adaptive histogram equalization on the image and then classified the points. The results of this procedure were quite good (Figure 3), but is not full-automatic since it demands some input from the user.

The remaining points formed the triangulated model of the CBH surface, which was textured mapped, using the standard interface OpenGL (OpenGL URL) (Woo, 1999), as an attempt

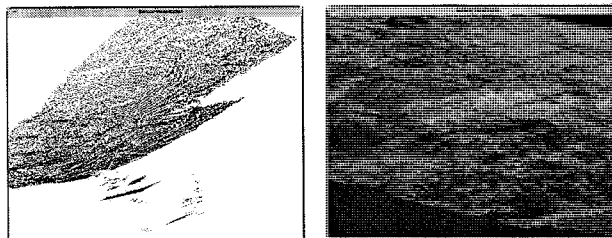


Figure 2: Point cloud from ground measurements (left) and its triangulation (right)

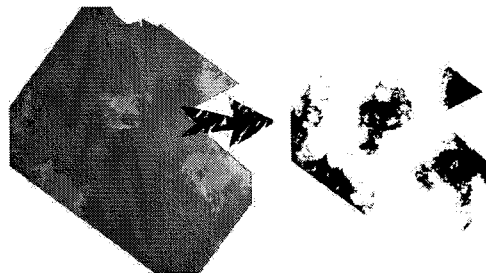


Figure 3: Cloud mask applied in GB measurements

to increase the realism of the visualisation. These first results were fair, but further improvement was not intended since we consider traditional polygonal modelling/rendering as an inefficient method for volumetric phenomena. A part of these first attempts were used, though, in further stages.

2.2 Development of modelling methods

The point cloud which resulted from the first stage, was used as data for developing and testing a point based rendering system, using particles. We constructed a simple system which includes positional and colour information for each particle, based on the point measurements of CBH. We also prepared the source code to import wind direction vectors for the animation of 3D clouds from subsequent measurements, lifetime duration of each particle and normal vector direction. This normal vector direction results from the triangulation of the first stage. Some test animations were performed with GB measurements, adding artificial data for wind direction and speed, with variations in the particle size and anti-aliasing methods with satisfying results as far as performance and memory consumption is concerned.

In the frame of this work, we first looked into different techniques of cloud modelling and concluded that the main issue is the description of clouds as volumes. These volumes consist of cells, where the values of the desired attributes are stored, usually aligned in regular grids. We constructed three-dimensional textures from ground based measurements. The pixel values adjust the material transparency and they are calculated according to the number of point measurements present inside each volume pixel. Volume modelling was tested, using the help of a software used in several weather-related applications, called Vis5d. An open source software, released under the GNU General Public License (Vis5D URL). The positive conclusions are the compression of the volume data, and the ease of including subsequent measurements and creating an animation. The negative conclusions are the absence at that stage of CTH estimations, which lead to an incomplete volume dataset, and a not satisfying final volume rendering. The first impression was that a completed version of the dataset would bring much better results.

The classification method described in section 2.1 was based only upon the radiometric behaviour (colour channels) of the

points measured. A successful segmentation into cloud objects would require to take into consideration the spatial relations between the measured points. An existing method that attracted attention was the modelling of potential 3D fields and the display of iso-potential values, using implicit density functions.

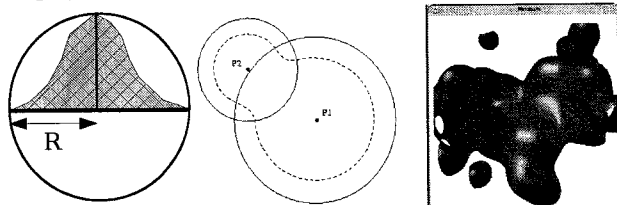


Figure 4: Metaball of radius R (left), blending of metaballs in 2D (middle) and 3D (right)

These set of functions evaluate the value of the field using the distance from specified 'source' points. A sphere around each point defines the area of influence, out of which the density is 0 (Figure 4). This technique was first applied in modelling free form surfaces, which are defined as the iso-potential surface of a value V . The objects modelled in this manner are called 'metaballs', or 'softballs' or 'blobs' (Nishita, 1996). Their

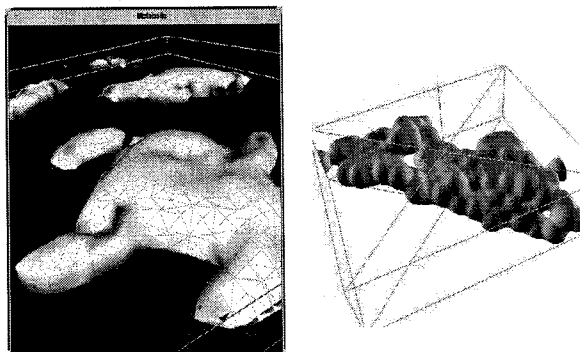


Figure 5: Application of metaballs on GB measurements with different metaball radii

application was extended into modelling clouds by (Dobashi, 2000), in combination with hardware accelerated rendering with realistic results.

Our first effort in this direction was to construct the functions that calculate the field density, based on the distance from the 'source' points. In our case the source points were the CBH point measurements themselves and since the points were organized on a regular grid, we used the resolution of this grid, as the fall-out radius for the metaballs. The visualisation of the isosurface at this stage has been implemented using polygons. The extraction of the iso-surface into a polygonal model was done using the well known 'marching cubes' algorithm. The algorithm can be divided into three sub-routines. First the object space is divided into cubes, at a resolution that we decide and the user supplies the iso-value that desires to extract. Next the algorithm for every cube evaluates the field on every cube vertex and calculates the difference with the desired iso-value. According to the sign of the difference on the cube-vertices, the algorithm finds in a pre-calculated lookup table, in which case falls the intersection of the iso-surface with the cube edges. This speeds up the decision which cube-edges have to be interpolated, in order to create the triangle that represents the iso-surface.

The results depend on the resolution of the cube grid, and in our case this plays a significant role in calculation and rendering times. For example in GB measurements, depending on the average bottom height the coverage varies from 1 to 4 Km in each horizontal direction in a resolution of approximately 5m,

depending again on the cloud average height. The coverage of the scene from the point cloud is irrelevant, since this algorithm covers the whole object space, disregarding holes in the object. The time consumption is largely due to calculation of the intersections and for a scene, with dimensions as described above, 2~3 minutes are required to compute the triangular surface.

We applied the forementioned 'metaballs' technique in order to calculate the pixel values for a cloud scene. (Mayer, 2003), (Nishita, 1996). Around each of the point measurements a sphere of influence is placed, with the amount of influence decreasing as the distance from the centre increases according to a density function (Figure 4). This allows us to estimate a density value on a regular grid, which comes from all 'metaballs' that include the current grid vertex, weighted by their influence. The result of applying this method on a set of ground-based measurements is shown in Figure 5, where the outer shell of the estimated cloud volume is shown in the form of polygonal surface.

Due to the large computational times, for a whole scene taken from EO measurements we started the implementation of hierarchical grid structure, for increasing the storage and speed performance. We noticed that this issue has been addressed in the past by the computer graphics community and we chose a software library used at the National Centre for Atmospheric Research, at Boulder Colorado, for similar purposes as ours (NCAR URL). It is based on the "shear-warp" algorithm for fast volume rendering (Volpack URL), (Lacroute, 1994), and facilitates a complete workflow from reading raw volumetric data, storing colour, transparency and normal vector information in compressed files, to rendering the volume data.

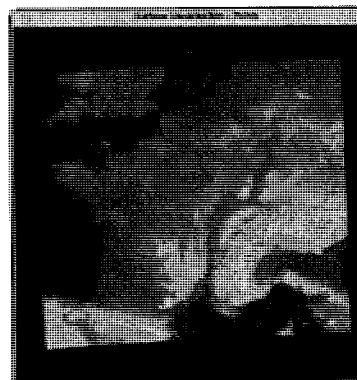


Figure 6: The aLMo coverage

The compression schemes (*rle* and *octree*) allow the storage of the volume data values, together with the transfer functions which determine the colour and opacity of the medium. The *rle* (Run Length Encoded) scheme is optimal for viewing the same volume under different viewing transformations and shading parameters, while the *octree* scheme performs better with volumes with varying transfer function.

The pre-processed volumes have larger file size (up to four times) in comparison with the raw binary data, and this happens due to the material colour/transparency and normal vector information that have to be stored within. Keeping all these information together in one volume file, is though a more important advantage since the processing time is significantly reduced. The software is based on the Tcl scripting language and can be combined with the existing C source easily. The sequence of commands needed to be passed to the Tcl interpreter are saved in script files, and can be called within any C program.

We continued by importing the alpine local model (*aLMo*) (source: Meteo Swiss) into the volume rendering procedure.

Grids from numerical models describe the Cloud Liquid Content in %, in 7x7 Km² cells, covering an area of 1401x1197 Km² (Figure 6). The vertical size is not uniform, and this was an obstacle in importing the grid into a volume suitable for volume rendering. The aLMo NWP data-set is used as test for 3D cloud field modeling and interpolation of irregular grid to regular was performed in IDL programming environment due to the ease of handling 3D arrays of data.

At the current stage we started with importing 3D cloud fields from combined ground-based and satellite observations into the Volsh volume rendering procedure. The data originate from the GB and EO measurements that took place in April 2002, over the Kloten airport near Zurich. The area covered from the EO measurements is 55x55 Km² and approximately 2x2 Km² from the GB observations. The 3D cloud field is described by the cloud top height (CTH), the cloud liquid water (DLR product) estimated on the whole of the grid from the MISR satellite data and the CBH information extrapolated over the area covered from the GB measurements.

2.3 Development of rendering methods

In parallel with the development of 3D cloud fields we used several techniques for visualization of 3D cloud volumes. After the creation of these volumes, based on the metaball technique (see Section 2.2), we implemented a hardware assisted rendering method (Woo, 1999) which is used often in volume rendering. In its simplest form we create planar cross-sections of the volume, oriented towards the viewer, which are in the end blended together. The resulting image is the combination of



Figure 7: Volume Rendering with planar slices

the colour from all planes, multiplied by their transparency. Using the example of the previous section we supply a sample rendering in Figure 7, where we can see that the absence of lighting calculations (shadows) in the volume results in difficulty understanding the depth of a volume. This problem can be solved as in (Dobashi, 2000), where lighting calculations are performed by sorting each of the planes towards the light source, and calculating an attenuation factor for each plane. This factor represents the amount of light that reaches a plane after passing through the other planes that lie in front of the light source, and determine its shade level.

We further improved the interactivity our rendering method by introducing dynamic levels of detail for the 3-dimensional volume. The levels are created by the graphics processor from the original resolution data, and the OpenGL graphics library handles the display and transition between the levels. In order to avoid aliasing effects from undersampling the original

volume and the LoDs' transition we tried trilinear interpolation, which improved the image quality of the texture, but slowed down the frame rate significantly. In parallel we had implemented the import procedure into the Volsh library (see Section 2.1) and tried its rendering capabilities (Figure 8). The procedure is no more real-time, but the rendering times are quite low. For example a volume of 256x128x64 pixels is rendered in approximately 3 seconds.

To the advantages of the last rendering approach we count the option to add shading to the volume, which increases the realism of the resulting image, and the option to adjust the transfer function, which defines the material properties of the medium for every value of the scalar variable, in near real-time. This means that we can interactively change the material properties, and focus on some specific range of the scalar value, leaving the outliers transparent. For the animation, we have the option to record some key-positions and create the intermediate

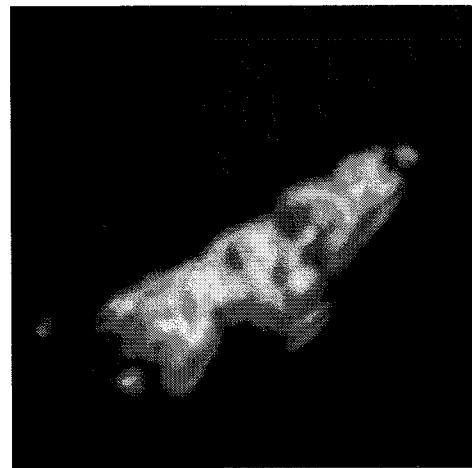


Figure 8: Volume rendering using 'Volsh' library

frames, by interpolating between the neighbour key-positions, thus limiting the number of commands necessary to render the frame sequence. Finally the library is based on the Tcl scripting language, which allows the execution of any external program in combination with the rendering procedure. This allows us, for example, to automate the production of animations, and deliver video from the separate rendered frames.

For the disadvantages we can mention the absence of perspective projection in the library (only parallel projection is available), and the complexity to manipulate the camera flightpath to create key-frames for the animation of volumes.

A rendering of the whole scene (Figure 9) revealed problems with poor rendering quality, which result from the fact that the input data are the surface points measured from the EO observations and do not correspond to a 3D volume, but a surface.

In order to evaluate the performance of the procedure with 3D volumes we imported the CLC - cloud cover- variable from the alpine local model, used at MeteoSwiss (see Section 1), and created an animation of a fly-through around the rendered volume. A snapshot from this animation is given in Figure 10. We continued by visualising the interpolated 3D Cloud Liquid Water, from the combined EO and GB measurements, taken in April 2002. These measurements are a fusion of the MISR Cloud Top Height estimations over Zurich-Kloten and the ground-based CBH estimations, taken at the same time, using the ground camera system developed at ETH Zurich. The measurements cover a different extent of the area and the CBH were extrapolated over the whole extent.

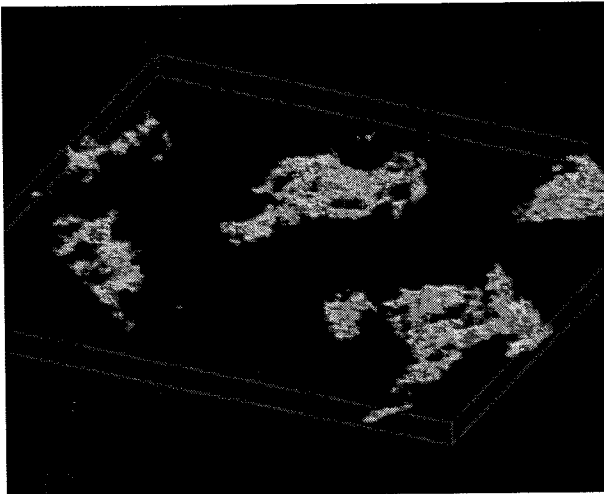


Figure 9: Volsh rendering of a whole GB scene



Figure 10: CLC cloud cover from aLMo NWP model

3. CONCLUSIONS AND FURTHER WORK

From the short literature study that we did, we have seen that the problem of modelling and rendering cloud-type data can be approached from different routes. There is a trade-off between the simulation accuracy of the physical phenomenon and the time required to deliver visualisation. One can focus on keeping a realistic image of cloud formation and movement together with interactive rendering speed, but in scientific visualization this would mean drifting away from the results of the measurements. On the other hand, by using strictly the results of measurements, we move to the other end of visualization difficult to understand without detailed knowledge of the instruments used and possible errors that may happen. In the present work, we demonstrated the several problems and difficulties that arise from using real measurements in cloud modelling and rendering. If visualization is about delivery of desired information in a quick and comprehensive way, then the field of 3D volume rendering has a lot of open issues in fields such as environmental sciences and meteorology.

The present combination of particle systems, soft body modelling, hardware accelerated graphics and software-based volume rendering had the purpose to teach us the benefits that scientific visualization can gain from following and adjusting current advances in computer graphics.

In the current status of the presented work we would like to improve the search algorithm that adds up the metaball weights for every grid cell, since at the moment is a straightforward loop that traverses all the metaballs, without taking into account if they will contribute finally or not. We are interested in continuing the work on temporal interpolation from several time steps in order to deliver animations over time. This will be possible to do on the aLMo NWP model since it provides 3D values every hour. Application of this idea on the combined CBH and CTH measurements seems difficult because of the temporal resolution of the satellite images.

4. ACKNOWLEDGEMENTS

The aLMo NWP data were supplied by Gabriela Seiz and MeteoSwiss, Zurich Weather Office. We would like to thank our CLOUDMAP2 colleague Tobias Zinner at DLR, for supplying us with three-dimensional cloud liquid water from the combined CLOUDMAP2 measurements in Zurich Kloten airport, taken on 12 April 2002. This work was funded by the Bundesamt für Bildung und Wissenschaft (BBW), within the EU project CLOUDMAP2 (BBW Nr. 00.0355-1).

5. REFERENCES

- Blinn, J. Light reflection functions for simulation of clouds and dusty surfaces, *Proceedings of the 9th annual conference on Computer Graphics and interactive techniques*, July 1982, Boston, USA, pp. 21-29.
- Dobashi, Y., Tomoyuki, N., Yamashita, H., Okita, T. Modeling of clouds from satellite images using Metaballs, *The Visual Computer*, 1999, 15(9), pp. 471-482.
- Dobashi, Y., Kaneda, K., Yamashita, H., Okita, T., Nishita, T. A simple, efficient method for realistic animation of clouds, *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, July 2000, pp. 19-28.
- Ebert, S., D., Parent, E., R. Rendering and animation of gaseous phenomena by combining fast volume and scanline A-buffer techniques. *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, August 1990, Volume 24, Issue 4, pp. 357-363.
- Ebert, S., D., Musgrave, F., K., Darwyn, P., Perlin, K., Worley, S., 1998. *Texturing and Modeling – a procedural approach*, second edition, Academic Press.
- Gardner, Y., G., Visual simulation of clouds, *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, July 1985, pp 297-304.
- Harris, J., M., Lastra, A. Real time cloud rendering, *Eurographics Computer Graphics Forum*, 2001, 20(3), Blackwell Publishing, pp. 76-84.
- Lacroute, P., Levoy, M. Fast Volume Rendering Using a Shear-Warp Factorization of the Vieweing Transformation, *SIGGRAPH 94, Proceedings of the 21st annual conference on computer graphics and interactive techniques*, July 1994, Orlando, USA, pp. 451-458.

Lee, C., Kesselman, C., Schwab, S. Near-real-time Satellite Image Processing: Metacomputing in CC++. *IEEE Computer Graphics and Applications*, July 1996, pp. 79-84.

Levoy M., Whitted T. The use of points as a display primitive, *Technical Report 85-022, Computer Science Department*, University of North Carolina at Chapel Hill, January 1985.

Mayer B., Zimmer, T., Seiz, G., Roditakis, A. *CLOUDMAP2 Second Interim Report*. February 2003.++

NCAR URL, *Volsh Volume Rendering Software*, <http://www.vets.ucar.edu/software/volsh/> (accessed 4 Feb. 2004)

Nishta T., Dobashi Y., Nakamae E. Display of Clouds Taking into Account Multiple Anisotropic Scattering and Sky Light, *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, July 1996, pp.379-386.

OpenGL URL, *The industry standard for high performance graphics*, <http://www.opengl.org> (accessed 4 Feb. 2004)

Perlin K. An image synthesizer, *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, July 1985, 19(3), pp 287-296.

Perlin K., Hoffert M. E. Hypertexture, *Proceedings of the 16th annual on Computer graphics and interactive techniques*, July 1989, 23(3), pp. 253-262.

Reeves, W. Particle Systems – A technique for modeling a class of fuzzy objects, *ACM Transactions on graphics*, 2(2), April 1983, pp. 91-108.

Rusinkiewicz, S., Levoy, M. Qsplat – A multiresolution point rendering system for large meshes, *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, July 2000.

Schulze, P., J., Niemeier, R., Lang, U. The Perspective Shear-Warp algorithm in a virtual environment, *IEEE Visualization 2001*, October 21-26 2001, San Diego, California, USA, pp. 207-213.

Seiz, G., Baltasvias, E., Gruen, A. Cloud mapping from the Ground: Use of Photogrammetric Methods. *Photogrammetric Engineering & Remote Sensing*, 68(9), September 2002, pp. 941-951.

Vis5D URL, *The Vis5d home page*, <http://www.ssec.wisc.edu/~billh/vis5d.html>. (accessed 4 Feb. 2004)

Volpack URL, *Computer Graphics at Stanford University*, <http://www.graphics.stanford.edu>. (accessed 4 Feb. 2004)

Woo, M., Neider, J., Davis, T., Shreiner, D. *OpenGL Programming Guide*, October 1999, third edition, Addison Wesley.

Zwicker M., Pfister H., van Baar J., Gross M. Surface Splatting, *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, Los Angeles, California, USA, August 12-17 2000, pp. 371-378.

ACCURACY OF AUTOMATED AEROTRIANGULATION AND DTM GENERATION FOR LOW TEXTURED IMAGERY

M. Sauerbier

ETH Zurich, Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology Zurich
ETH Hoenggerberg, CH-8093 Zurich, Switzerland
www.photogrammetry.ethz.ch
martin.sauerbier@geod.baug.ethz.ch

IC WG II/IV

KEY WORDS: Accuracy, Adjustment, Aerial, Automation, DEM/DTM, Matching, Processing

ABSTRACT:

The Nasca-Project at ETH Zurich aims for a GIS-based analysis of the topography and the geoglyphs carved into the ground by the ancient Nasca (200 B.C. – 600 A.C.) in the desert region of Nasca/Palpa, about 500 km south-east of Lima, Peru. In 1998, three blocks of aerial images have been acquired during a photoflight. Two of these blocks (about 400 images, scale 1:7.500) were processed using an analytical plotter Wild S9 during 4 years of manual measurements, resulting in a high resolution DTM and 3D-vectors of the geoglyphs as well as topographic elements (rivers, streets, houses). Especially aerotriangulation and DTM-measurements have been time-consuming processes. For the third block of Nasca imagery (about 400 images at a scale of approximately 1:9100) we aim for an automated processing (Aerotriangulation and DTM generation) to provide accurate data as a basis for terrain analysis (visibility studies, surface calculation etc.) in relation to the geoglyphs (Grün et al., 2003).

Actual digital photogrammetric stations (DIPS) are examined for this case of low textured imagery due to the deserts characteristic of the landscape. Aerial triangulation and DTMs generated using Z/I's Image Station™ modules ISDM and ISAE, version 4.00, and Supresoft Inc. Virtuozo™ versions 3.1 and 3.3 are compared to reference data measured on an analytical plotter Wild S9. The main focus of this paper is on DTM generation, as for aerial triangulation no strictly comparable results, produced fully automatically, could have been achieved.

1. INTRODUCTION

One of the most important limiting factors concerning automated aerotriangulation and DTM calculation is texture. Actual digital photogrammetric stations allow for good results in most cases if the processed imagery contains texture with high contrast. Problems occur processing images which cover low textured areas, e.g. forests, glaciers, grasslands or deserts. Especially in these areas, matching algorithms fail measuring identical points in two or more images which leads to blunders or even non-measured points. The consequences during aerotriangulation can be instable relative orientations of images if there are not enough points to achieve a regular distribution, and therefore a decreased accuracy of the absolute orientation. DTM generation is affected by blunders or even gaps in areas where matching is impossible (Baltasvias et al., 1996). Both cases require costly manual editing or measurement. The data used for the examination of aerotriangulation consists of a subblock of 26 aerial images (B/W) of the Pampa de Nasca which were triangulated on an analytical plotter Wild S9, using IGP's bundle adjustment software BUN and on Virtuozo 3.1 during a diploma thesis (Keller, 2003). These images were then also processed on Z/I's Image Station. To compare DTM generation, one stereo model of this subblock and 4 stereo models of another subblock were used. For both aerotriangulation and DTM generation the results of the automated processing were compared, no manual measurements or editing were applied. The area covered by the generated DTMs does not contain vegetation or buildings, thus the DSMs calculated by the DIPS can be considered as DTMs.

The used images were scanned at a resolution of 15 microns according to 13.65 centimetres in object space. The overlapping along strip as well as perpendicular to the flight direction is

60%. See figure 1 for an example of the used images and their content.

The landscape covered by the images is mostly flat, here and there interrupted by quebradas, usually dry valleys formed by draining water from the Andes. The surface consists of little stones and sand.

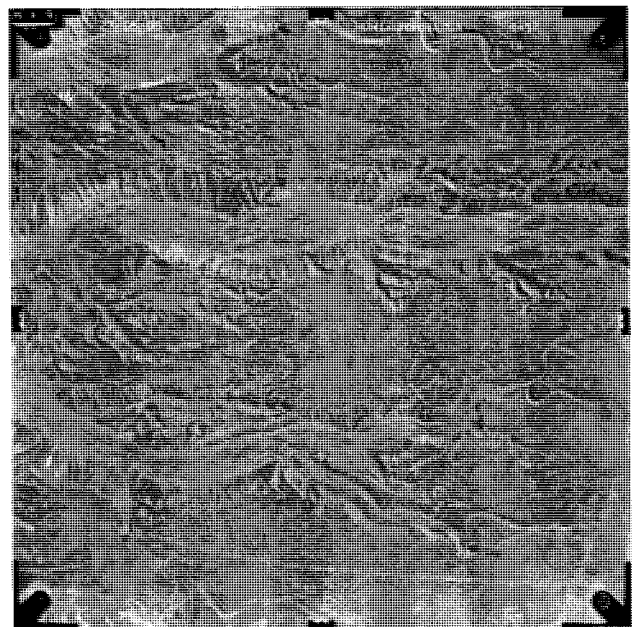


Figure 1: Typical aerial image (117) of the Pampa de Nasca

2. AEROTRIANGULATION

2.1 Test data

To achieve reliable conclusions, high accuracy reference data is required. For one area of investigation, a triangulation was measured on an analytical plotter Wild S9 using 26 aerial images from 8 different strips. The acquired image coordinates of the control and tie points, measured according to the von Gruber distribution, then served as input data for the bundle adjustment package BUN. As a result, a σ_0 of 9.65 microns could be achieved for the processed subblock as global accuracy. Compared to the values we achieved for the two blocks of Palpa imagery (scale 1:7500), $\sigma_0 = 13.3$ microns and $\sigma_0 = 9.5$ microns, the global accuracy of 9.65 microns can be considered as a good result.

The second test area consists of two strips with 26 and 28 images, respectively.

2.2 Z/I Image Station

The workflow of automatic point measurements in Image Station Digital Mensuration (ISDM) is divided into two main parts: the relative orientation, where points are measured in single stereo models, and multiphoto orientation, where points in more than two images can be measured. It is also possible to start with multiphoto orientation directly. The control points have to be measured manually in at least one image using the absolute orientation menu item, and can then also be transferred to other images automatically using multiphoto orientation.

Automated measurement of tie points for relative orientation requires an initial manual measurement of two points in each stereo model before automated processing can be started successfully. The automated transfer of points between images of different strips is possible, and can be used to measure the tie points acquired during the relative orientation in corresponding images of the neighbouring strips. For the automated matching of both, relative and multiphoto orientation, a patch size of 9x9 pixels was used, combined with a number of 5 points maximum to be measured at each of the given 3 positions per image (6 per model). These positions follow the Gruber distribution, but can also be altered by the user to any other distribution. The correlation threshold, the value each measurement has to achieve to be accepted, was chosen to 0.95 and adapted down to 0.75 for single models. After performing the automated measurement with these parameters in relative orientation mode, it could be observed that at each position at least one point could be measured, usually more. Except for one model, the average parallaxes of the models were smaller than 10 microns. Afterwards, using the multiphoto orientation with automated point transfer, bundle adjustment could not be performed successfully due to the large parallaxes for points measured in more than two images, especially in the overlapping areas of strips. The same effect occurred in a second effort with manually measured tie points for strip connection. To clarify the reason for this exactly, some more investigations have to be done. Better results were achieved triangulating the first two strips of 54 images of the Nasca block, where manually measured tie points were used for strip connection. Figure 2 shows the distribution of the automatically measured tie points. In this project, points at almost every given position could be matched although especially strip 2 contains predominantly desertous areas, while some images contain agricultural crop land, which also can be classified as low textured. The position where matching failed significantly is marked in figure 2. As a global accuracy, 21.1 microns could be

achieved after bundle adjustment, which is less than scan pixel accuracy.

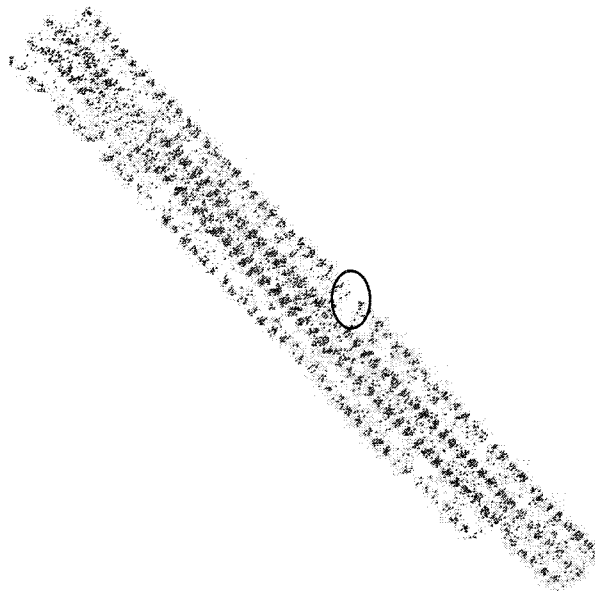


Figure 2: Distribution of automatically measured tie points using strip 1 and 2 (54 images)

2.3 Virtuozo 3.1

The test concerning aerotriangulation was performed using Virtuozo 3.1. At IGP, a license including aerial triangulation is available only for this version. In Virtuozo 3.1, no manual measurements are required for the initialisation of the stereo models, but for tie points between the strips. Using Virtuozo, like in Image Station, the user can influence the distribution of the tie point positions to be measured. Virtuozo supports Gruber and similar distributions and offers the user to choose the number of points to be measured at each position. Further parameters, like the ones mentioned above, can not be changed. The result of the automated tie point extraction in Virtuozo (figure 3) shows, that at several positions no tie points could be extracted. Performing bundle adjustment in PATB, a σ_0 of 6.97 microns was achieved. This value can not be compared directly to the result obtained from the analytical plotter because it is too optimistic due to the fact, that only points of high quality are included into the bundle adjustment, the weak points are completely ignored at the positions where gaps occurred. Thus, the calculated σ_0 does not consider instabilities of relative orientations resulting from missing tie points. In figure 3, red ellipses point up the areas where a significant accumulation of gaps in the tie point distribution appeared, compared to the manually measured points.

For the second subset of Nasca imagery, tie point extraction succeeded only for strip 1 although manual measurements for strip connection had been accomplished in each of the overlapping images. The mode for measurements of tie points for strip connection does not provide a zoom function, therefore the manual measurements were probably not exact enough for successful point matching.

For this reason, no directly comparable results could be achieved for automatic aerotriangulation, the presented results can only be considered as preliminary results of qualitative character.

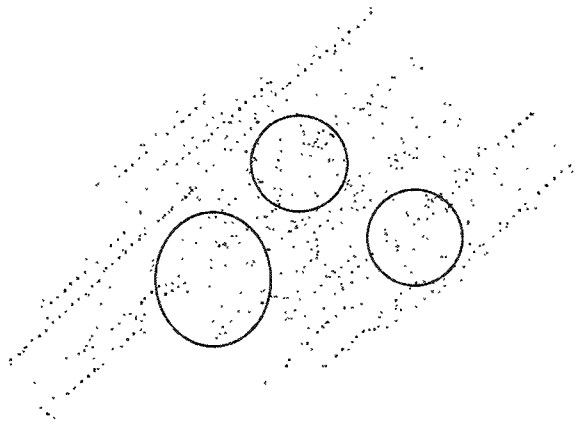


Figure 3: Distribution of tie points, red: automatically generated points, blue: manually measured points

3. DTM GENERATION

3.1 Reference data

In the area of investigation, no control points of superior accuracy exist. Thus, for the comparison of automatically generated DTMs, again accurate reference data measured on an analytical plotter Wild S9 was acquired using the software XMAP by Aviosoft™. For 5 stereo models from two strips, parallel profile measurements were performed with an average point distance of 20m along the profiles and a profile distance of 20m. Breaklines were not measured due to the flat areas except for model 116_117. Using our software DTMZ, grids with 5m mesh size were interpolated from the measured points. The grids then could be imported to ESRI ArcGIS™ 8.3 and directly compared with the grids acquired from the DIPS by calculating the height difference:

$$\Delta Z = Z_{DIPS} - Z_{REFERENCE} \quad (1)$$

From the resulting height difference grid, mean value, RMS error and the minimum and maximum offset were computed. Potential trends in the performance of height errors and their spatial distribution can be extracted. To visualise the distribution of the height errors, histograms of the differential grids and contour lines were produced, which allow for an examination of the geomorphological accuracy.

The topography covered by the processed stereo models is predominantly flat with only few discontinuities except for one model, consisting of the images 16 and 17 in the first strip, which contain parts of 3 quebradas (figure 1).

For DTM generation, the orientations computed using Image Station Digital Mensuration were transferred to the analytical plotter and used for the manual measurements. On Virtuozo, 4 control points per stereo model, transferred from ISDM, were used for absolute orientation of the images. To minimise the marginal differences caused by varying spatial attitudes and positions of the projective centers, Geomagic Studio 4.1 by Raindrop Geomagic Inc. was used to register the automatically derived DTMs to the reference data applying the global registration function.

3.2 Image Station™

The method of DTM generation applied by Image Station Automatic Elevations (ISAE) is described in detail in the manual which is integrated in the graphical user interface (Z/I Imaging Corporation, 2002). For each level of the image pyramid, an initial DTM is derived by matching homologous points, starting with a horizontal plane in the first level. From this initial DTM, a DTM is modeled with bilinear finite elements which then serves as the initial DTM for the next level. For matching, ISAE uses the interest operator and correlation coefficient while the matching area is geometrically defined by a parallax bound and the epipolar line.

ISAE offers a lot of different strategies for DTM generation. Users can choose different terrain types, adaptive or non-adaptive grid, parallax bound and matching modes, correlation thresholds or define terrain types by themselves. Different smoothing filters with user-defined weights and sampling factors can be applied. Some tests with the default terrain types showed, that the best results could be achieved using terrain type "flat" with adaptive mode. The patch size for crosscorrelation was set to 9x9 pixels and the smoothing filter was set to "high", keeping the default values for smoothing weight (2.0) and sampling factor as recommended in the manual. After a first attempt with a correlation threshold of 0.95, a value of 0.75 was chosen because of the low point density and thus a strong terrain filtering achieved with 0.95. The result is a grid with 5m mesh size which the software interpolates from the measured points.

The height differences obtained for the different models show clearly, that not only texture but also terrain characteristics, especially steep slopes like on the border of the valleys in the images 16 and 17, affect the accuracy of the automatically generated DTMs. Table 1 shows the mean height error, its standard deviation and the minimum-maximum range of the height differences between the automatically derived DTM and the manually measured DTM.

Table 1: DTM generated using Image Station compared to the manually measured DTM on an analytical plotter Wild S9

Model	ΔZ	Std. Deviation	Min. - Max.
116_117	0.25m	3.10m	-19.1m - 24.2m
210_211	-	-	-
211_212	0.22m	1.99m	-4.4m - 36.4m
212_213	0.25m	1.53m	-6.3m - 18.8m
213_214	-0.02m	1.33m	-12.9m - 7.0m
223_224	-0.01m	0.77m	-3.3m - 7.6m

Mean height errors, standard deviations and the minimum-maximum-ranges of the different stereo models show a noticeable heterogeneity. In model 210_211, for an unresolved reason no correct DTM could be calculated.

The differential grid of model 116_117 shows another phenomenon: The big blunders show a coincidence with areas of steep slopes and are predominantly positive (figure 4).



Figure 4: Differential grid of model 116_117 (ISAE), positive errors are bright, negative errors dark

Another effect, influencing the geomorphological correctness of the DTM, is the smoothing in the Image Station measurements caused by interpolation. Missing terrain information like marked in figure 5 is the consequence.

From the achieved results for Image Station, no trend for ΔZ can be derived for the single matched points because the generated DTMs are affected by the ascertained smoothing effect which depends on the terrain characteristics.

To reduce the smoothing effect, the DTMs were generated with different filter settings (low, medium and high), but the accuracy did not increase significantly. Compared to earlier investigations (Grün et al., 2000), where an RMS error of 0.7m for Palpa imagery was achieved with ISAE with DTM interpolation of the matched points done with DTMZ, this indicates that the interpolation module of ISAE determines the main part of the height errors. Terrain elevations and sinks of a size up to 100 by 50 metres are modeled as flat terrain and cause blunders in a magnitude of 10 to 20 metres in model 116_117 (figure 5).

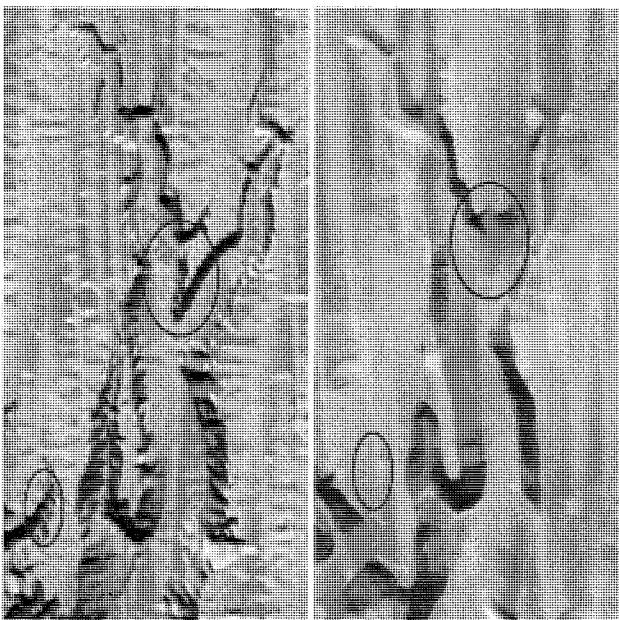


Figure 5: Manually measured DTM (left) and automatically derived DTM using ISAE (right). Marked red: Loss of geomorphological information due to strong smoothing.

3.3 Virtuozo™ 3.1/3.3

Unlike Image Station, Virtuozo uses epipolar images for automatic DTM generation. The algorithm consists of crosscorrelation and a global relaxation matching technique for blunder detection (Baltsavias et al., 1996). The points are calculated based on a regular grid in image space, and therefore the calculated DTM has to be interpolated subsequently. For the comparison of the generated DTMs, digital terrain models were calculated using both versions available at our institute, 3.1 and 3.3, to investigate whether an improvement in matching algorithms affects the obtained results. In Virtuozo, like in Image Station, different strategies for DTM calculation can be chosen by the user. It distinguishes between flat, undulated, hillock, broken and rugged areas. Further parameters that can be influenced are the patch width and height. The value was set to 9 pixels for both width and height for this test. An integration of manually measured breaklines to refine the matched points is also possible. After matching, a visual quality control can be accomplished. The matched points are visualised in the stereo model, divided into 3 classes (red, yellow, green) where the yellow and the green classes contain the more or less reliable points and the red class consists of the unreliable matching results. There is no detailed description of the point classification in the user manual (Supresoft Inc., 1999). To compare them to the reference data set, the points resulting from the matching process still had to be interpolated to a grid, in this case with 5m mesh size using DTMZ.

Table 2 shows the results obtained with Virtuozo 3.1 compared to the manual measurements, while table 3 contains the results for the DTMs generated using Virtuozo 3.3.

Table 2: Results of the DTM-comparison Virtuozo 3.1 vs. the manual measurements

Model	ΔZ	Std. Deviation	Min. – Max.
116_117	0.17m	2.95m	-17.2m – 18.4m
210_211	-0.03m	1.00m	-9.5m – 6.7m
211_212	0.08m	0.74m	-4.5m – 5.5m
212_213	-0.06m	1.35m	-24.9m – 11.9m
213_214	-0.03m	0.98m	-10.8m – 8.3m
223_224	0.02m	0.64m	-5.2m – 4.3m

Table 3: DTM generated with Virtuozo 3.3 compared to the reference DTM

Model	ΔZ	Std. Deviation	Min. – Max.
116_117	0.35m	2.66m	-15.9m – 16.8m
210_211	0.04m	0.81m	-5.6m – 7.7m
211_212	0.13m	0.73m	-6.1m – 4.2m
212_213	0.13m	1.20m	-5.3m – 18.6m
213_214	0.02m	1.03m	-10.7m – 7.4m
223_224	-0.01m	0.65m	-5.2m – 4.3m

Generally, the standard deviations of the measurements produced by Virtuozo 3.3 are marginally smaller than the ones from version 3.1. Also the minimum-maximum ranges become more narrow in the most models. For both automatically generated datasets, the mean height errors are homogeneously close to 0m, having similar standard deviations between 0.65m and 1.35m. Model 116_117 shows a significant difference: the standard deviation of 2.66m and the larger minimum-maximum range for the DTM generated by Virtuozo 3.1 compared to the smaller values of version 3.3 is a sign of larger blunders.

In model 116_117, like in the DTM generated with Image Station, the most blunders occur on the steep slopes, while the height error distribution in the flat areas is very similar for both systems.

Unlike the smoothing effect occurring in the Image Station datasets, both Virtuozo software versions produce very rough surfaces (figure 6). A loss of geomorphological information as appearing in the Image Station data (figure 5) is not obvious. The results of both tested versions of Virtuozo achieved similar results, except for the models 116_117 and 212_213. In both cases, version 3.3 achieved a smaller standard deviation and minimum-maximum range. Generally, the values for ΔZ show that Virtuozo 3.3 tends to measure slightly higher than Virtuozo 3.1 but seems to be less affected by blunders especially for the two problematic models 116_117 and 212_213.

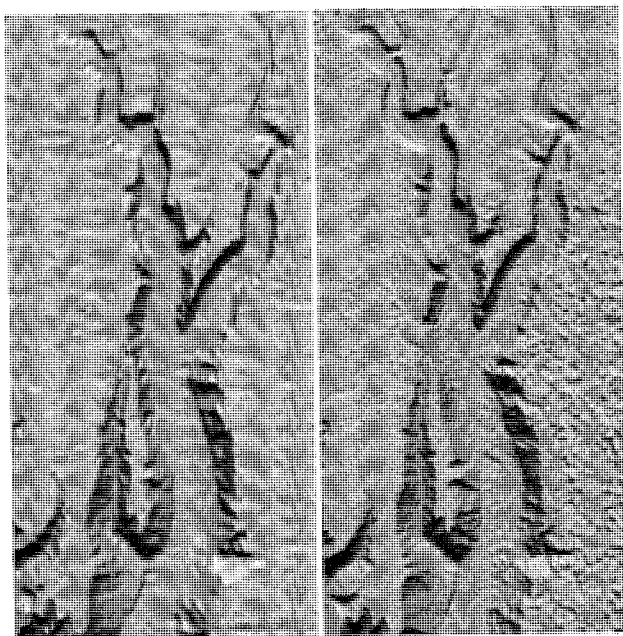


Figure 6: Manually derived DTM (left) and DTM generated by Virtuozo 3.3 (right)

3.4 Comparison

Figures 5 and 6 visualise the general difference between the acquired DTMs on Image Station and Virtuozo: On the one hand, a strong smoothing effect in the Image Station data and on the other hand the rough surface produced by Virtuozo 3.1/3.3.

The accuracies resulting from the comparison of the acquired DTMs (tables 1,2 and 3) show, that in average the Virtuozo system has mean height errors close to 0m with standard deviations mostly between about 0.7m and 1m while the data produced on Image Station differs with a larger variance between about 0.7m and 1.6m in the models covering flat terrain. Both systems show problems in model 116_117, where Virtuozo exceeds standard deviations of 2.5m and Image Station obtains 3.1m. Furthermore, a strong smoothing effect causes a loss of geomorphological information especially in rough terrain (compare figure 5) The differences between manually measured and automatically derived data are caused by the steep slopes in this model, as shown in figure 7.

Concerning the flat areas, both systems achieved throughout good results without big blunders or gaps. Contour lines from a subset of model 116_117, the model for which the worst results

in terms of standard deviations were obtained due to the slopes, are comparatively illustrated (figure 7).

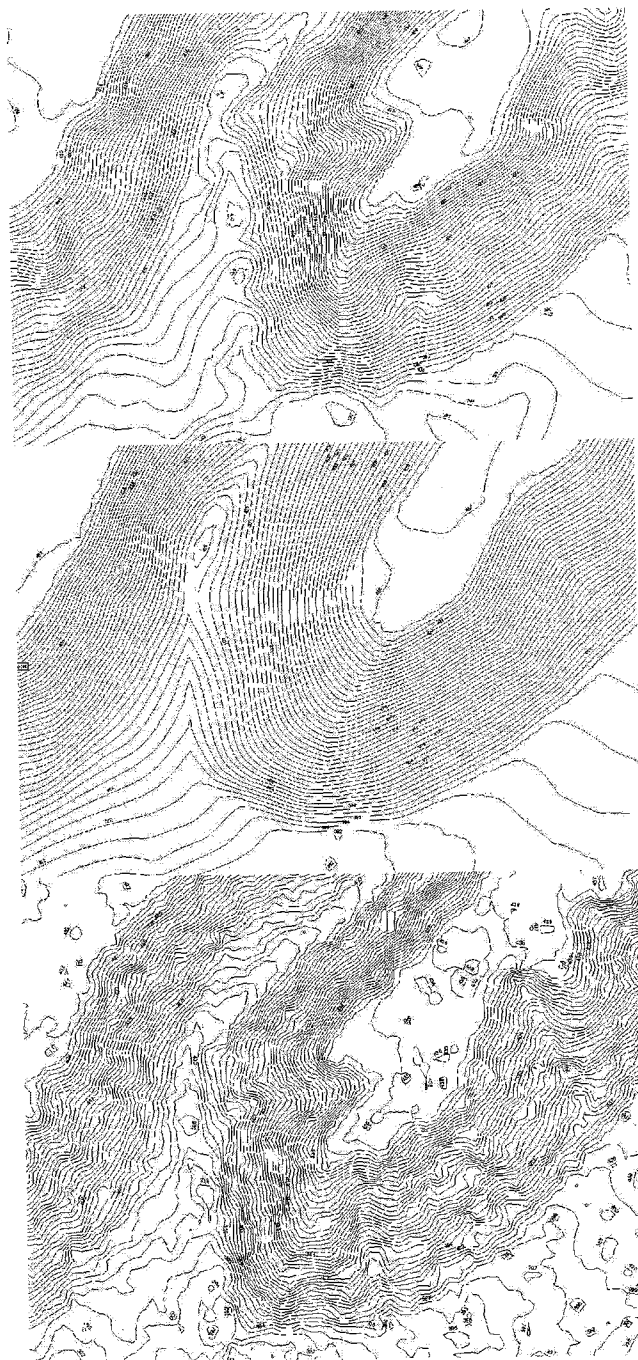


Figure 7: 1m contour lines: Wild S9, Image Station and Virtuozo 3.3 from top to bottom (Model 116_117).

Compared to the contour lines derived from the manually measured DTM, the contour lines based on the Virtuozo measurements are more detailed regarding geomorphological information while the contour lines based on Image Station measurements are too smooth.

A comparison of the contour lines for the model with the best result (223_224), derived from Image Station and analytical plotter, with the difference grid shows that the errors of Image Station data result from the smoothing applied during DTM

interpolation (figure 8). It is clearly visible that morphological details are not considered in a sufficient way.

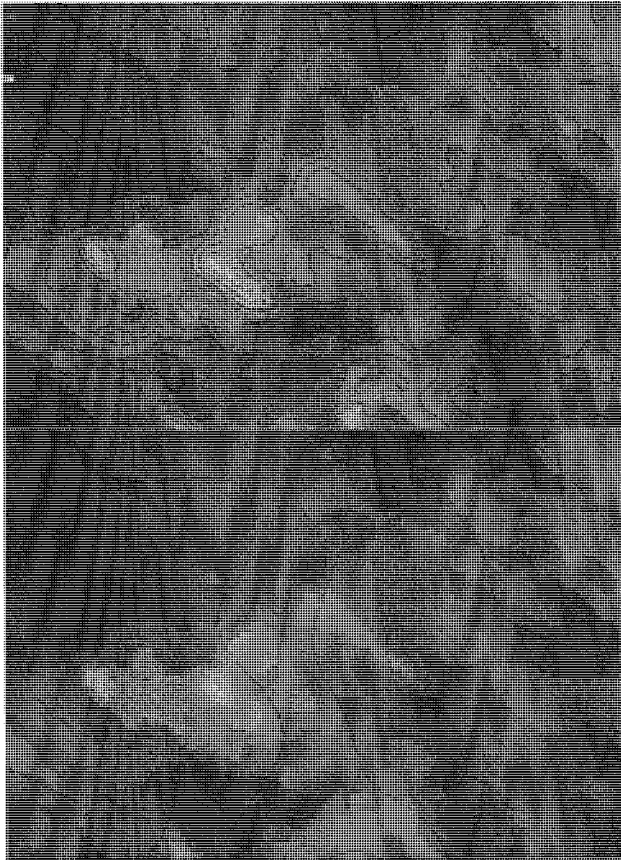


Figure 8: Exemplary comparison of the derived contour lines from analytical plotter (top) and Image Station (bottom) for model 223_224, overlaid on their height difference grid, respectively. Positive height differences are bright, negative differences dark.

4. CONCLUSIONS AND FUTURE WORK

The investigations accomplished so far have shown, that a fully automated image processing for aerotriangulation and DTM generation does not provide results that are as reliable as the reference data measured on an analytical plotter. Concerning aerotriangulation, for the Nasca project a feasible solution would be semi-automated processing using the automated tie point extraction of a digital photogrammetric workstation enhanced by manual measurements in areas of weak point density.

DTM generation can be advanced considerably with digital photogrammetric workstations, compared to manual measurements on analytical plotters. However, the performed tests show that the terrain characteristics affect the obtainable accuracy and reliability significantly. Manual measurements and extensive post editing are required to achieve results which hold the accuracy level of analytically generated DTMs.

The problematic results do not occur in the flat areas with homogeneous texture, where perhaps the automatically derived DTMs are even more accurate than the manually measured DTM due to their high density of matched points. The large blunders, which require manual editing, are mainly located on the steep slopes.

Further investigations on an automated processing of Nasca imagery will be done on an internally developed software package for DTM extraction. First tests, which produced no satisfactory results, have shown, that the software still needs to be improved on aerial imagery. Commercial software packages, including Virtuozo and Image Station, will be further examined regarding aerotriangulation and DTM generation. Especially for Image Station, a direct investigation of the matched points could possibly lead to better results if external DTM interpolation software is used.

Another topic is the integration of manually measured breaklines as pre known data into the matching process for automatic DTM generation, which could improve the accuracy in areas containing discontinuities, but would increase the cost due to the required manual measurements.

ACKNOWLEDGEMENTS

The author would like to thank D. Keller, former ETH Zurich student, for her dedicated work within the scope of Virtuozo and analytical plotter measurements and processing, as well as the students A. Weber and S. Zingg for their aid on image scanning and measurements on the analytical plotter.

REFERENCES

- Baltsavias, E., Käser, C., 1998. DTM and orthoimage generation – a thorough analysis and comparison of four digital photogrammetric systems. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Stuttgart, Germany, Vol. XXXII, Part 4, pp. 42-51.
- Baltsavias, E., Li, H., Gaschen, S., Sinning, M., 1996. DTM generation with the Leica/Helava DSW 770 and Virtuozo digital photogrammetric systems. Photogrammetric Systems. Paper presented at the Symposium Geoinformatics 1996, 16.-19. October, Wuhan, China, pp. 10-17.
- Grün, A., Bär, S., Bühner, T., 2000. DTMs derived automatically – Where do we stand? In: *Geoinformatics*, Vol. 3, No. 5, July/August 2000, pp. 36-39.
- Grün, A., Lambers, K., 2003. The geoglyphs of Nasca: 3D recording and analysis with modern digital technologies. In: *Acts of the XIVth UISPP Congress, University of Liège, Belgium, 2-8 September 2001, section 1: theory and methods - general sessions and posters* (BAR International Series 1145), pp. 95-103.
- Keller, D., 2003. Genauigkeitsuntersuchung zur automatischen Aerotriangulation und DTM-Messung im Projekt Nasca. Diploma Thesis, unpublished.
- Supresoft Inc., 1999. Virtuozo NT User's Manual Version 3.1.
- Supresoft Inc., 2002. Virtuozo User's Manual Version 3.3.
- Z/I Imaging Corporation, 2002. Image Station Automatic Elevations Help (4.00).

SATELLITE- AND GROUND-BASED MULTI-VIEW PHOTOGRAMMETRIC DETERMINATION OF 3D CLOUD GEOMETRY

G. Seiz*, D. Poli, A. Gruen, E. P. Baltsavias, A. Roditakis

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) Zurich
ETH Hoenggerberg, CH-8093 Zurich, Switzerland
{gseiz, daniela, agruen, manos, roditak}@geod.baug.ethz.ch

Commission VII, WG VII/6

KEY WORDS: Atmosphere, Global Change, Sensor Orientation, Matching, Camera, Visualization.

ABSTRACT:

The quantitative 3D description of clouds is important for refined methods in nowcasting and the modeling of weather and climate. The EU project Cloudmap aimed at developing new methodologies for cloud product derivation (heights, type, optical thickness and effective droplet size). The follow-up project Cloudmap2 aimed at producing and exploiting value-added remote sensing data products on macroscopic (e.g. cloud-top height) and microscopic (e.g. cloud droplet radius) properties and water vapour distributions in order to characterize sub-grid scale processes within Numerical Weather Prediction Models (NWP) through validation and data assimilation.

Earth Observation (EO) image data, provided by ESA, EUMETSAT and NASA are used to derive geophysical value-added data products over Europe and the North Atlantic region, whenever possible in near real-time. Ground-based active (cloud radar, ceilometer) and passive (stereo imager system, IR camera) remote-sensing instruments are used to validate the EO-derived products as well as to merge them with the satellite-based results for a full 3D representation of the clouds. The role of our group in Cloudmap2 was to estimate cloud-top height (CTH) and wind (CTW) from stereo images from satellites and cloud-bottom height (CBH) and wind (CBW) from stereo images acquired by our newly developed ground-based stereo imager system. The cloud-top and -bottom results were then combined into a 3D model and visualized. This paper describes the results obtained in CTH and CTW estimation from ATSR2, AATSR, MISR and Meteosat-6/-7, including validation, the CBH and CBW results from the ground-based stereo imager system and a case study where the satellite- and ground-based 3D cloud boundary results are combined.

1. INTRODUCTION

The interest to monitor cloud properties from space and ground-based observations is based on the large influence that clouds have on the Earth and Atmosphere energy balance. The EU-FP4 project Cloudmap aimed at estimating new cloud-top products (heights, type, optical thickness, effective droplet size), especially for cirrus and contrail clouds from existing and new sensors, using three different techniques (brightness temperature with CO₂ slicing method, stereoscopy and Oxygen A-band). These cloud-top products have been validated using airborne sensor underflights, multi-resolution observations from space sensors and ground-based remote sensing instruments. Cloudmap ended in January 2001 and was then continued by the EU-FP5 project Cloudmap2 (www-research.ge.ucl.ac.uk/cloudmap2/) until June 2004.

Cloudmap2 aimed at producing and exploiting value-

added remote sensing data products on macroscopic (e.g. cloud-top height) and microscopic (e.g. cloud droplet radius) properties and water vapour distributions in order to characterize sub-grid scale processes within Numerical Weather Prediction Models (NWP) through validation and data assimilation. Earth Observation (EO) data, provided by ESA, EUMETSAT and NASA are used to derive geophysical value-added data products over Europe and the North Atlantic region, whenever possible in near real-time. Ground-based active (cloud radar, ceilometer) and passive (stereo imager system, IR camera) remote sensing instruments are used to validate the EO-derived products as well as to be merged with the satellite-based results for a whole 3D representation of the clouds. Numerical simulation experiments based on state-of-the-art radiative transfer methods are used to quantify the effect of broken clouds on the Earth's radiation budget and lead to a better representation of clouds within NWP models.

* Corresponding author.

The role of our group in Cloudmap and Cloudmap2 was to estimate cloud-top height (CTH) and wind (CTW) from stereo images from satellites and cloud-base height (CBH) and wind (CBW) from stereo images acquired by our own ground-based stereo imager system, with stereo-photogrammetric techniques. As second step, the cloud-top and -base results were then combined and visualized in 3D.

This paper describes the CTH and CTW retrieval from multi-view satellite sensors (ATSR2/AATSR, MISR) and geostationary satellites (Meteosat-6/-7) and the comparison/validation of these cloud products with other satellite-based products as well as measurements from ground-based instruments (multi-camera system, cloud radar). Finally, a case study is presented where the satellite- and ground-based 3D cloud boundary results were combined.

2. SATELLITE-BASED STEREO ANALYSIS

The satellite-based stereo analysis includes several processing steps which are illustrated in Figure 1. After the data description in Section 2.1, the sensor model for image georeferencing will be presented in Section 2.2. The subsequent processing steps are explained in Section 2.3.

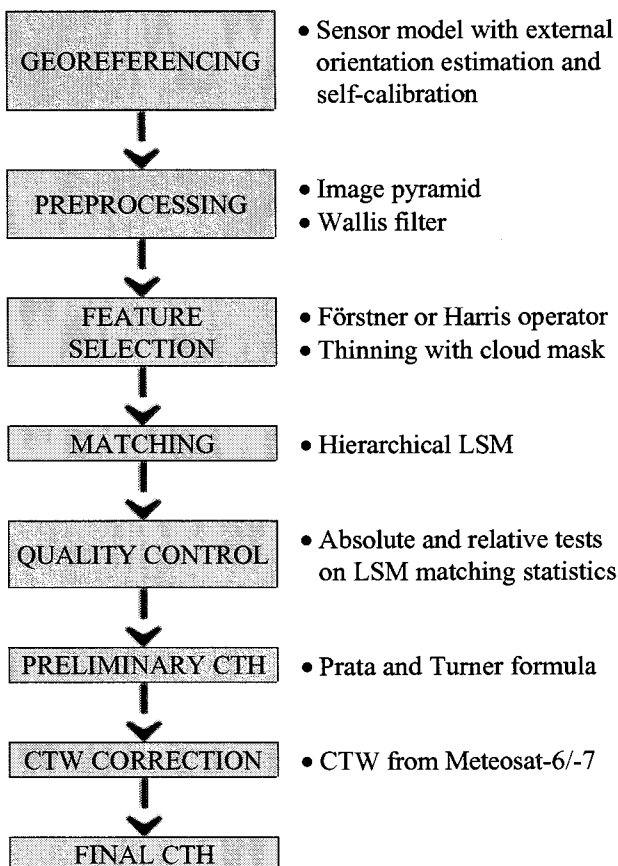


Figure 1. Schematic overview of stereo-

photogrammetric processing of the satellite-based images to derive CTH and CTW.

2.1 Data

2.1.1 ATSR2 / AATSR: The Along Track Scanning Radiometer (ATSR2) instrument is part of the ERS-2 satellite system which was launched in April 1995. The successor sensor, AATSR, is part of Envisat, which was launched in Spring 2002. ERS-2 and Envisat are in a near-circular, sun-synchronous orbit at a mean height of 780 km, an inclination of 98.5° and a sub-satellite velocity of 6.7 km/s. The repeat cycle of ATSR2/AATSR is approximately 3 days.

The ATSR2/AATSR sensor first views the surface along the direction of the orbit track at an incidence angle of 55° as it flies toward the scene. Then, some 120 s later, ATSR2 records a second observation of the scene at an angle close to the nadir. The ATSR2 field of view is comprised of two 500 km-wide curved swaths with 555 pixels across the nadir swath and 371 pixels across the forward swath. The pixel size is 1 km x 1 km at the center of the nadir scan and 1.5 km x 2 km at the center of the forward scan. The sensor records in seven spectral channels, i.e. $0.55 \mu\text{m}$, $0.67 \mu\text{m}$, $0.87 \mu\text{m}$, $1.6 \mu\text{m}$, $3.7 \mu\text{m}$, $10.8 \mu\text{m}$ and $12.0 \mu\text{m}$. All channels have a radiometric resolution of 10-bit. Our CTH retrieval is based on the rectified data products, GBT for ATSR2 and ATS_TOA_1P for AATSR. The geolocation of these rectified products is achieved by mapping the acquired pixels onto a 512 x 512 grid with 1 km pixel size whose axes are the satellite ground-track and great circles orthogonal to the ground-track.

2.1.2 MISR: The Multi-angle Imaging SpectroRadiometer (MISR) is currently the only operational satellite that acquires images from nine different viewing angles. MISR was launched on board the EOS AM-1 Terra spacecraft in December 1999. The orbit is sun-synchronous at a mean height of 705 km with an inclination of 98.5° and an equatorial crossing time of about 10:30 local solar time. The repeat cycle is 16 days. The MISR instrument consists of nine pushbroom cameras at different viewing angles: -70.5° (named DA), -60.0° (CA), -45.6° (BA), -26.1° (AA), 0.0° (AN), 26.1° (AF), 45.6° (BF), 60.0° (CF), and 70.5° (DF). The time delay between adjacent camera views is 45-60 seconds, which results in a total delay between the DA and DF images of about 7 minutes. The four MISR spectral bands are centered at 446 nm (blue), 558 nm (green), 672 nm (red) and 866 nm (NIR). The red-band data from all nine cameras and all spectral bands of the nadir camera are saved in high-resolution with a pixel size of 275 m x 275 m. The data of the blue, green and NIR bands of the remaining eight non-nadir cameras are stored in low-resolution with a pixel size of 1.1 km x 1.1 km. The operational data products from MISR are

described in (Lewicki et al., 1999). The two products used for this study are the L1B1 radiance and the L1B2 ellipsoid-projected radiance data.

The L1B1 product is radiometrically but not geometrically corrected, while the L1B2 ellipsoid-projected radiance product is referenced to the surface of the WGS84 ellipsoid with no terrain elevation included. The MISR georectified product spatial horizontal accuracy requirements are driven by the needs of the geophysical parameter retrieval algorithms. The goal of operational MISR data processing is to achieve an uncertainty better than ± 140 m for both the absolute geolocation of the nadir camera and the co-registration between all nine cameras (Jovanovic et al., 2002). The latest evaluation results of the L1B2 geolocation accuracy as shown in (Jovanovic et al., 2004) are approaching prelaunch requirements, with along- and cross-track errors far below 1 pixel for all cameras (except DA).

The operational L2TC top-of-atmosphere/ cloud product, which contains the operationally derived cloud parameters, like stereo CTH, east-west (EW) and north-south (NS) cloud motion components, as well as many additional parameters from the stereo retrieval (Diner et al., 2001), can be used as comparison data for validation (Seiz, 2003).

2.2 Sensor Modeling

The aim of rigorous sensor models is to establish a relationship between image and ground reference systems according to the sensor geometry of acquisition. In particular, different approaches have been proposed for the georeferencing of pushbroom sensors carried on aircraft (Gruen et al., 2002) and satellite (Poli, 2003). A flexible sensor model that can be applied to a wide class of linear CCD array sensors has been developed in our group and already applied to different linear scanners carried on satellite and aircraft (Poli, 2003). The model is based on the photogrammetric collinearity equations, that are extended in order to include the external orientation modeling with 2nd order piecewise polynomials and a self-calibration for the correction of lens distortions and CCD lines rotations in the focal plane.

The sensor model was applied to georeference the MISR level 1B1 product (Poli, 2003). Two areas of interest, over Germany and South France, were chosen. From the test over Germany a GCP accuracy of 173 m in X, 87 m in Y and 80 m in Z was achieved, corresponding to 0.6, 0.3 and 0.3 pixels (ground pixel size: 275m). For the test over South France, RMS errors of 43 m in X, 45 m in Y and 152 m in Z, corresponding to 0.2, 0.2 and 0.6 pixels, were obtained.

The first results are promising, because the images have been oriented with sub-pixel accuracy. The self-calibration was fundamental because it allowed the estimation of the correct internal and external orientation parameters. In these tests, significant values for the principal point displacement have been estimated. Without self-calibration, the RMS errors in GCPs were larger than one pixel.

All of our MISR CTH and CTW calculations presented in this study (Section 2.3) fully rely on the operational L1B2 georectified radiance data. In the future, however, it is planned to use the georectification from the described in-house sensor model.

2.3 Cloud-Top Height and Motion Estimation

Determination of CTH from ATSR2, AATSR or any two views of MISR proceeds along the same scheme, as illustrated in Figure 1. First, all images were reduced to 8-bit with linear stretching between the minimum and maximum values. As no a-priori values of the cloud heights were given to the matching algorithm, the number of pyramid levels for the hierarchical matching was chosen so that the maximum possible parallax at the highest level was only 1-2 pixels. Three and five pyramid levels were used for ATSR2/AATSR and MISR, respectively. Every pyramid level was enhanced and radiometrically equalized with the Wallis filter. According to the block or filter size, different cloud structures could be enhanced. In general, a block size of about 70 pixels was chosen at the original level, which was then decreased up the pyramid. Points with good texture were selected with the Förstner or Harris interest operator in the first or second pyramid level because it is likely that these same points are readily detectable in the other levels. If a cloud mask was available (e.g. our own cloud mask for ATSR2, L1 RCCM or L2TC cloud masks for MISR), we used it for thinning of the point set to cloud points only, prior to matching.

The unconstrained Multi-Photo Geometrically Constrained (MPGC) least-squares matching (LSM) (Gruen, 1985; Baltsavias, 1991) was applied hierarchically, starting on the highest pyramid level. After each pyramid level, quality control with absolute tests on the LSM matching statistics was performed to exclude the largest blunders from further processing down the pyramid. The patch size was slightly increased from one pyramid level to the next, from 7 x 7 on the highest level to about 15 x 15 on the lowest level.

After applying the MPGC LSM algorithm, the matching solutions were quality-controlled with absolute and relative tests on the matching statistics. Additionally, meteorological criteria can be used in the detection of large blunders, including minimum and maximum cloud heights, minimum and maximum cross-track parallaxes,

which are, after division by the time difference, proportional to the cross-track wind speed, or filtering the cloud heights with the brightness temperature values from the IR channel(s) in the case of ATSR2/AATSR.

The resulting y-parallaxes were converted into cloud-top heights according to Prata and Turner (1997). The zenith angles (e.g. θ_{nadir} and θ_{forward} for ATSR2/AATSR) thereby had to be projected on the along-track plane. The height values of the successfully matched points were finally interpolated to the full resolution grid.

The accuracy of the retrieved cloud-top heights was dependent on the geometric stereo configuration expressed as the base-to-height ratio B/H, the matching accuracy Δy_p , the accuracy of the georectification, including the exact values of the zenith angles, and the along-track motion retrieval accuracy $\Delta v'$. In Table 1 the B/H values and image time differences for ATSR2/AATSR and three different viewing angle combinations of MISR are listed, together with an estimation of the height error Δh given an along-track parallax error Δy_p of 1 pixel from matching or an along-track motion error $\Delta v'$ of 5 m/s.

For all sensors, the height error due to motion errors is very prominent. In contrast to stereo image pairs from scan-synchronized geostationary satellites, stereo image pairs from a single polar-orbiting satellite are never perfectly synchronous. There is a time delay of seconds to minutes between image acquisition at the different viewing angles. The resulting errors in stereo cloud-top height retrievals can be quite large, depending on the along-track cloud motion, as pointed out in Table 1. If more than two non-symmetric views are available, the along-track parallax can be separated into the amount due to cloud height and the amount due to cloud motion. With only two views, or symmetric multiple views, which is the usual case, the along-track cloud motion has to be corrected with data from an independent source. One possible source of independent data is geostationary satellite cloud motion information. In our study, three types of geostationary data from the two European satellites Meteosat-6 and Meteosat-7 (Eumetsat, 2003) were used: the Meteosat-6 5-minute Rapid Scans during MAP, the quasi-operational Meteosat-6 10-minute Rapid Scans and the operational Meteosat-7 30-minute sequences. The launch of the first Meteosat Second Generation (MSG) satellite (called Meteosat-8 since its transition into operational mode in March 2004) in August 2002, with a temporal resolution of 15 minutes, now offers a further data source for accurate CTW retrieval in several spectral bands.

Sensor	B/H ratio	Δt [s]	Δh [m] for $\Delta y_p = 1$ pixel	Δh [m] for $\Delta v' = 5$ m/s
ATSR2/AATSR	0.7-1.2	100-130	830-1430	420-930

MISR AN-AF	0.49	45	560	460
MISR AN-BF	1.02	92	270	450
MISR AN-DF	2.85	204	95	360

Table 1. Height error caused by parallax error and along-track motion for various B/H and time acquisition difference cases.

For cross-track wind retrieval and along-track wind correction, the exact time difference Δt between corresponding pixels in the forward and nadir scans had to be calculated. For ATSR2/AATSR, the time difference varies significantly over the scan and can be calculated from the along-track distance on the ground and the satellite velocity after Lorenz (1985). For MISR, the time difference for different cross-track positions within a block or scene can be assumed as constant. Northerly winds lead to an underestimation of the heights and the along-track wind component has to be added to the y-parallax, while southerly winds result in overestimation of cloud-top heights. The motion vectors retrieved from Meteosat-6 and Meteosat-7 data were resampled to the ATSR2 or MISR grid and the cross- and along-track wind components calculated. Using the time difference between the acquisition of the two views, the along-track wind components were converted into CTH corrections.

3. VALIDATION

As each observation and each retrieval method have its own characteristics, much can be learned from a intercomparison of the results. The objective is to document the relative performance of the different observations of cloud top height, and where possible to understand this. In collaboration with Eumetsat and the Rutherford Appleton Laboratory (RAL), two case studies were analyzed with spaceborne observations from ATSR2, MISR and Meteosat-6/-7 and ground-based observations by the Chilbolton radar and radiosondes. The results are summarized in Table 2; the retrieval methodologies (next to our stereo processing) are described in detail in Tjemkes et al. (2002).

Date	ATSR2	MISR	Radar	Radiosonde
28/06/2000	2.3 ± 0.2	2.4 ± 0.2	2.48 ± 0.03	2.5 ± 0.2
13/06/2001	4.5 ± 0.3	4.4 ± 0.2	4.27 ± 0.08	4.6 ± 0.2

Table 2. CTH results (in km above sea-level) from ATSR2, MISR, radar and radiosondes for two cases (20 June 2000 and 13 June 2001) over Chilbolton, UK.

Based on the validation experience from Cloudmap2, a new comparison study of cloud height assignment methods has been started by Eumetsat in November 2003. Next to the multi-view photogrammetric retrieval from MISR and AATSR, the cloud height will be determined

by optimal estimation from AATSR (by RAL), CO₂ slicing from MODIS and Oxygen A-band from MERIS (by Free University Berlin). The comparison of these different height products with the operational Eumetsat AMV (Atmospheric Motion Vector) product will allow to evaluate the strengths and weaknesses of each individual height assignment method within the AMV production chain. Results from Phase 1 of the study are reported in Fischer et al. (2004).

4. COMBINATION WITH GROUND-BASED STEREO DATA

For the acquisition of ground-based stereo images of clouds, a camera imager system has been developed (Figure 2). The system, data acquisition and data processing is described in detail in Seiz et al. (2002). The imager system was part of the MAP-Special Observation Period (SOP) composite observing system at the Rhine Valley, Switzerland, in autumn 1999 and was later installed at the Zürich-Kloten airport in September 2001 and April 2002. The system was used for the retrieval of cloud-bottom heights, which were then used for the validation of stereo cloud-top heights of vertically thin clouds (e.g. cirrus, contrails) and for the combination of cloud-bottom boundaries with satellite-derived cloud-top boundaries for a 3D representation of the current cloud field. This second application is explained in the next section.

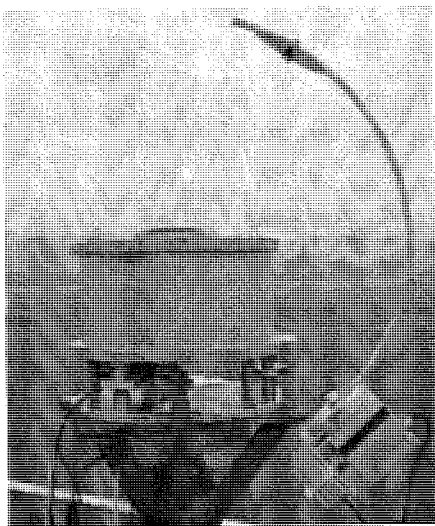


Figure 2. Ground-based imager system (Skycam).

4.1 Combination of Satellite- and Ground-based Datasets

For assimilation experiments with a very high-resolution version of the operational NWP model at MeteoSwiss, several Skycam measurements series were performed at Zurich-Kloten airport in April 2002, in coincidence with satellite (ASTER, MISR) overpasses. In collaboration with the German Aerospace Center (DLR), a 3D cloud

data set was then derived from the combination of MISR stereo cloud-top heights, MODIS microphysical data and Skycam stereo cloud-bottom heights. On a 50 m x 50 m grid, the cloud geometry was defined by the cloud mask and the cloud-bottom heights from Skycam and by the cloud-top heights from MISR. According to the effective radius reached close to the cloud top and the total optical thickness (both from MODIS MOD06 data), a constant liquid water gain and a droplet number density could be derived for each cloudy column. Thus, an internal profile of microphysics of a vertical resolution of 50 m was achieved (Figure 3). More details about the 3D cloud boundary extraction and subsequent application in radiative transfer simulations at DLR can be found in Seiz (2003) and in the Cloudmap2 Final Report, D13/D14 (Cloudmap2, 2004).

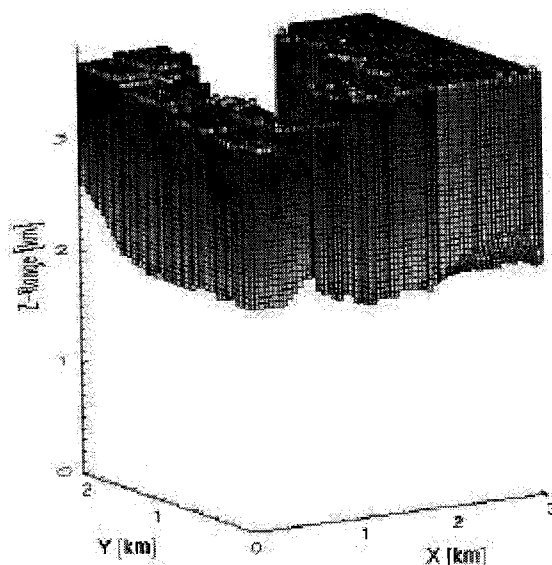


Figure 3. 3D cloud water distribution, derived from 3D cloud boundary data (MISR, ground-based camera system) and MODIS information, 19 April 2002 (image courtesy: Tobias Zinner, DLR).

5. CONCLUSIONS

This paper has shown examples of satellite- and ground-based stereo analysis of clouds. Regarding satellites, there are various sensors currently available which can be used for stereo-photogrammetric cloud retrievals. For stereo CTHs from a single polar-orbiter with only two viewing angles (e.g. ATSR2), it has been proven to be absolutely necessary to correct the preliminary heights with cloud-top wind (CTW) data from another source. Over land and mountainous regions, the cloud motion is most accurately derived from simultaneous images of a geostationary satellite. Over Europe, the Meteosat-6 Rapid Scan trials in 1999 (5min) and in 2000 (10min), and the operational Meteosat-6 10min Rapid Scans (since September 2001) are perfectly suited for this objective. The MISR instrument and its products were

presented as a promising alternative to derive CTH and CTW simultaneously with stereo-photogrammetric methods. The images from our new ground-based imager system showed to be valuable for validation of vertically thin cloud situations and for combination with satellite-based cloud boundaries to derive 3D cloud fields which can then be used for assimilation into numerical weather prediction and climate models.

In co-operation with Eumetsat, Free University of Berlin and RAL, the validation activities for satellite-based cloud height products will be continued. In addition to the geostationary satellites from the Meteosat First Generation (MFG), the data from Meteosat-8 will be used for CTW extraction and eventually stereo cloud-top height estimation in combination with Meteosat-5.

The ground-based photogrammetric validation activities will be continued with a stereo set-up of two operational Whole Sky Imagers at the ARM-SGP site, next to cloud radars, ceilometers and a Raman lidar.

ACKNOWLEDGEMENTS

The Meteosat-6/-7 data were received from the EUMETSAT Archive Facility (MARF), the ATSR2 data via the ESA ATSR2 NRT service, the AATSR data from the Rutherford Appleton Laboratory (RAL) and the EOS-Terra MISR data (level 1B2 and level 2TC) were obtained from the NASA Langley Research Center Atmospheric Sciences Data Center. We thank Veljko Jovanovic, Catherine Moroney, Akos Horvath and Roger Davies for valuable input in the understanding of the MISR products. This work is funded by the Bundesamt für Bildung und Wissenschaft (BBW) within the EU-projects CLOUDMAP (BBW Nr. 97.0370) and CLOUDMAP2 (BBW Nr. 00.0355-1) and by EUMETSAT within ITT-03/527.

REFERENCES

Baltsavias, E., 1991. Multiphoto Geometrically Constrained Matching. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitteilungen Nr. 49.

Cloudmap2, 2004. Cloudmap 2 Final Report. EC 5th Framework Programme, Energy, Environment and Sustainable Development. To become available at <http://www-research.ge.ucl.ac.uk/cloudmap2/publications.html>

Diner, D., Davies, R., Di Girolamo, L., Horvath, A., Moroney, C., Muller, J.-P., Paradise, S., Wenkert, D., Zong, J., 1999. MISR Level 2 Cloud Detection and Classification. JPL Technical Report ATBD-MISR-07, Jet Propulsion Lab., California Inst. of Technol., Pasadena, CA, USA, available at http://eospsp.gsfc.nasa.gov/eos_homepage/for_scientists/atbd/d

<ocs/MISR/atbd-misr-07.pdf> (accessed April 30, 2004).

Eumetsat, 2003. Eumetsat homepage. <http://www.eumetsat.de> (accessed April 30, 2004).

Fischer, J., Preusker, R., Seiz, G., Poli, D., Gruen, A., Poulsen, C., Mutlow, C., Tjemkes, S., Borde, R., De Smet, A., 2004. Validation of cloud top pressure derived from MSG-SEVIRI observations through a comparison with independent observations. EUMETSAT Users' Conference, Prague, 31 May – 4 June.

Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3), 175-187.

Gruen, A., Zhang L., 2002. Sensor modeling for aerial mobile mapping with Three-Line-Scanner (TLS) imagery. *International Archives of Photogrammetry, Remote Sensing and SIS*, Vol. 34, Part 2, pp.139-146.

Jovanovic, V., Bull, M., Smyth, M., Zong, J., 2002. MISR in-flight camera geometric model calibration and georectification performance. *IEEE Transactions on Geoscience and Remote Sensing*, 40(7), 1512-1519.

Jovanovic, V., 2004. MISR Level 1 Quality Statement. http://eosweb.larc.nasa.gov/PRODOCS/misr/Quality_Summaries/L1_Products.html#registration (accessed April 30, 2004).

Lewicki, S., Chafin, B., Crean, K., Gluck, S., Miller, K., Paradise, S., 1999. MISR data products specifications. Technical report, NASA JPL, http://eosweb.larc.nasa.gov/PRODOCS/misr/readme/dps_ne_ic_d.pdf (accessed April 30, 2004).

Lorenz, D., 1985. On the feasibility of cloud stereoscopy and wind determination with the along-track scanning radiometer. *Int. J. Rem. Sens.*, 6(8), 1445-1461.

Poli, D., 2003. Georeferencing of MOMS-02 and MISR stereo images with strict sensor model. ISPRS Workshop "High resolution mapping from space 2003", Hannover, October (on CD-ROM). Available at <http://www.ipi.uni-hannover.de/html/publikationen/2003/workshop/contents.htm> (accessed May 7, 2004).

Prata, A., Turner, P., 1997. Cloud-top height determination using ATSR data. *Rem. Sens. Env.*, 59(1), 1-13.

Seiz, G., 2003. Ground- and satellite-based multi-view photogrammetric determination of 3D cloud geometry. Ph.D. thesis, Institute of Geodesy and Photogrammetry, ETH Zuerich, Switzerland, Mitteilungen Nr. 80.

Seiz, G., Baltsavias, E.P., Gruen, A., 2002. Cloud mapping from the ground: use of photogrammetric methods. *Photogrammetric Engineering & Remote Sensing*, 68 (9), 941-951.

Tjemkes, St., Lutz, H.-J., Duff, C., Watts, Ph., Poulsen, C., Wrench, Ch., Seiz, G., 2002. A multi-sensor analysis of cloud top pressure. EUMETSAT Users' Conference, Dublin, 2-6 September, Eumetsat Conference Proceedings, EUM P36.

EXTRACTION AND CLASSIFICATION OF WETLAND FEATURES THROUGH FUSION OF REMOTE SENSING IMAGES IN THE OKAVANGO DELTA, BOTSWANA

K. Talukdar

Institute of Geodesy and Photogrammetry, ETH Hönggerberg, CH-8093 Zürich, Switzerland -
talukdar@geod.baug.ethz.ch

Commission VII, WG VII/3

KEY WORDS: Remote sensing, land cover, classification, extraction, fusion, multisensor, multiresolution

ABSTRACT:

The Okavango delta in northwestern Botswana is an extremely complex and dynamic wetland ecosystem. The spatial information on diverse wetland features of the delta is needed for hydrological modeling and water resources management. Due to large size and inaccessibility of the delta, satellite images provide the only viable means to reliably map and measure these features. For better identification and delineation of these features in the Okavango delta, efficient image analysis techniques are needed. The synergistic use of images from different sensors with varied spatial and spectral resolutions have the potential for better extraction and classification of features. This paper focuses on extraction and classification of landscape and land cover features through fusion of different resolution images acquired by Landsat 7 ETM+ and SPOT 5 HRG sensors over the Okavango delta. Both multispectral and panchromatic images from these two sensors are used. Different image fusion approaches are examined and used to increase reliability in feature interpretation. The effects of data fusion in recognition and extraction features are examined and illustrated. Thematic information extraction was carried out by means of supervised and unsupervised classification to produce landscape/land cover classes for different spatial resolution data set. The results indicate that as spatial resolution increases, high spatial frequency landscape/land cover features are extracted in increasing detail. However, spatial heterogeneity also increases with increasing spatial resolution.

1. INTRODUCTION

The spatial information on wetland features (i.e. landscape and land cover) of the Okavango delta is needed for hydrological modeling and water resources management. Wetlands are by definition lands with water-tables at or near the surface, either seasonally or permanently (Hughes, 1996). Wetlands and the issue of their management and preservation now engage the increasing attention of public. A goal is to provide better information as wetland ecosystems are influenced or exposed to environmental risks. There are a number of types and classification systems of wetlands to describe them (Anderson *et al.*, 1976). They were categorised into different types and classes based broadly on hydrological, geomorphological, soil and vegetation characteristics of wetlands. There are a variety of information needs in wetland ecosystem management and landscape characterisation.

The Okavango River, which originates in central Angola, after flowing through over a thousand kilometer branching out to form the Okavango delta, one of the largest inland wetland ecosystems in the world, comprising an area between 16,000 to 22,000 km². The whole catchment area of the river plus the delta together known as the Okavango River Basin (Figure 1), comprises an area of nearly 429,400 km². The Okavango sub-basin is part of the larger Makgadikgadi basin, which covers an area of approximately 725,300 km² (Ashton and Neal, 2003).

The delta exhibits great variations in areal extent, spatial complexity and temporal dynamics. Due to large size and inaccessibility of the delta, satellite images provide the only

viable means to reliably map and measure these features. Since the wetland is composed of heterogeneous objects it is difficult to reliably identify and characterize features from satellite images. For better identification and delineation of these features in the Okavango delta, efficient image analysis techniques are needed. The synergistic use of images from

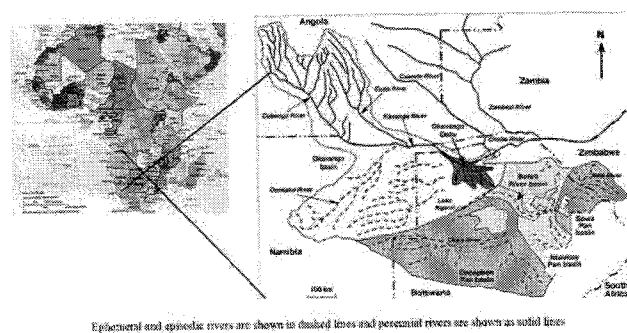


Figure 1. Location of the Okavango River Basin and the Delta, Southern Africa

different sensors with varied spatial and spectral resolutions have the potential for better extraction and classification of features. The principal aim of this study is to examine the effect of fusion of remotely sensed images from different sensors for extracting and classifying wetland features. Wetland features are represented as a collection of pixels in an image. For the

purpose of recognition, we need to describe the properties of pixels or groups of pixels.

2. THE STUDY AREA

The Okavango delta in northwestern Botswana is an extremely complex and dynamic wetland ecosystem. The increasing water demand from basin sharing countries (Angola, Namibia and Botswana) puts the resource under growing pressure and poses complex management problem. Water is a precious and important resource for all three riparian states of the basin. The demand for water is growing due to rapidly rising population and associated activities like irrigation and power generation. The Okavango delta (Figure 2), variedly termed as "the jewel" of the Kalahari (Ross, 1987), Africa's "Wetland Wilderness" (Bailey, 1998) and so on. It is the last remnant of the ancient great Makgadikgadi Lake, which was formed by the Kalahari sands about three million years ago (Ross, 1987). The delta is a large alluvial fan and experience annual flooding. It consists of a series of permanent river channels, semi-permanent drainage channels, lagoons, islands and floodplains that link up and then separate again during the course of an annual flood. The delta has a very low gradient. From Mohembo to Maun with a distance of 240 km, there is roughly 1m height variation for every 4 km distance (Figure 2).

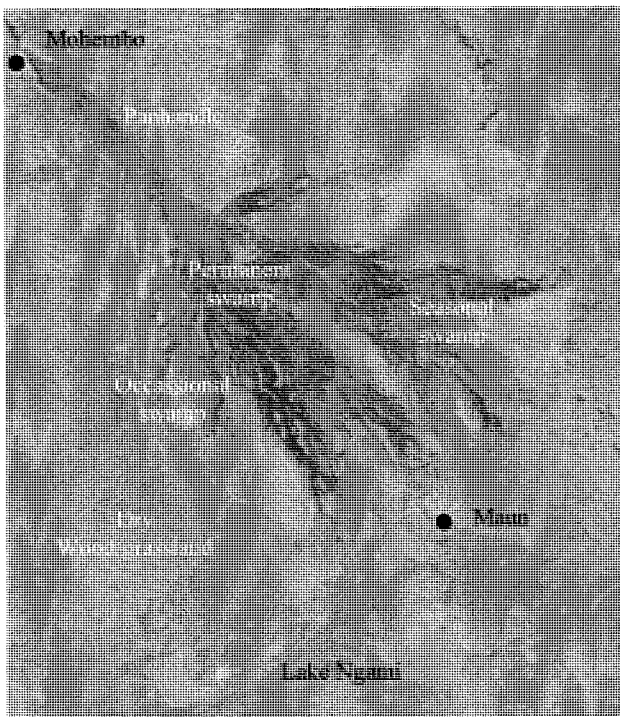


Figure 2. The Okavango delta as seen from Landsat 7 ETM+ (5,4,2 combination) image, April 2000

The Delta is located in a semi-arid region which experiences wide variations in flooding over permanent swamps, seasonal swamps and other intermittently flooded areas. The permanent swamp covers approximately 6,000 km², while the seasonal swamp varies between 4,000 and 10,000 km². Water depth in the permanent swamps averages 1.5m (UNDP, 1977), while in the seasonal swamps it is generally less than 1m (McCarthy *et al.*, 1998).

The river and the delta serve as major source of fresh water to the people of the basin area. The Okavango is one of the few large rivers in the world without notable man-made developments and that has so far remained relatively pristine.

3. DATA AND METHODS

3.1 Data

The primary data used in this study were SPOT 5 High Resolution Geometry (HRG) sensor image acquired on 12 November 2002 and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor image acquired on 28 March 2001. The other data used are 1:50,000 scale topographic maps and GPS measured ground control points. Both SPOT 5 and Landsat 7 images are taken during dry seasons. Figure 3 shows the scene location of SPOT 5 and Landsat 7 over the Okavango delta.

The SPOT 5 satellite was launched on May 2002 with Araine 4 rocket and put on polar sun synchronous orbit at an altitude of 832 km from earth surface, while the Landsat 7 satellite was launched on April 1999 with Delta rocket and put on polar, sun synchronous orbit at an altitude of 705 km from earth surface. Both SPOT 5 and Landsat 7 satellites orbit the earth in sun-synchronous mode at an inclination of 98.2° and 98° degree respectively.

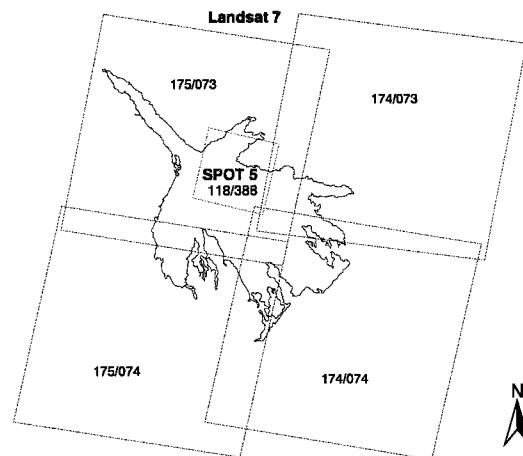


Figure 3. Landsat 7 ETM+ and SPOT 5 HRG scene location over the Okavango delta

SPOT 5 HRG sensor provides multispectral (10m resolution in band 1-3 and 20m resolution in band 4) and one panchromatic band (5m resolution) image. Each SPOT 5 HRG scene covers a ground area of 60 km x 60 km (nadir viewing). The HRG is an optical/near infrared sensor (Table 1), which is based on High Resolution Visible (HRV) and High Resolution Visible and Infrared Sensor (HRVIR). Both HRV and HRVIR are optical/near infrared CCD (*pushbroom*) imaging radiometers carried on SPOT-1 to -3, and SPOT- 4 respectively. The two HRG instruments aboard SPOT 5 are the main imaging instruments. They provide panchromatic (black and white) imagery at a resolution of 2.5 or 5 metres and multispectral (colour) imagery at a resolution of 10 metres.

Table 1. Sensor Characteristics - Landsat 7 ETM+ and SPOT 5 HRG

Wavelength Region	Spectral Range (μm) (with band no.)		Spatial Resolution (m)	
	Landsat 7 ETM+	SPOT 5 HRG	ETM+	HRG
VIS Blue	0.45 - 0.515 (B1)	-	30	-
VIS Green	0.525 - 0.605 (B2)	0.50 - 0.59 (B1)	30	10
VIS Red	0.63 - 0.69 (B3)	0.61 - 0.68 (B2)	30	10
NIR	0.75 - 0.90 (B4)	0.78 - 0.89 (B3)	30	10
MIR/SWIR	1.55 - 1.75 (B5)	1.58 - 1.75 (B4)	30	20
TIR	10.40 - 12.5 (B6)	-	60	-
MIR	2.09 - 2.35 (B7)	-	30	-
PAN	0.52 - 0.90 (B8)	0.48 - 0.71	15	5

Landsat 7 ETM+ sensor provides multispectral (30m resolution in band 1-5 and 7, and 60m resolution in band 6) and one panchromatic band (15m resolution) image. Each Landsat 7 ETM+ scene covers a ground area of 185 km x 170 km. The earth observing instrument on Landsat 7, the ETM+, replicates the capabilities of the highly successful Thematic Mapper (TM) instruments on Landsats 4 and 5. The ETM+ also includes new features that make it a more versatile and efficient instrument for regional to global environmental change studies, land cover monitoring and assessment, and large area mapping. Table 1 shows the detail sensor characteristics.

3.2 Methods

Fusion of remotely sensed images obtained using different sensors need to perform carefully. All data sets to be merged needs accurate registration to one another and resampled to the same pixel size.

3.2.1 Georeferencing

The scanned topographic maps at 1:50,000 scales are geometrically rectified to the UTM projection, Zone 34 S co-ordinates. Clarke 1880 spheroid and Cape datum were used for georeferencing all data sets. On an average half pixel (i.e. 2m) accuracy is achieved in georeferencing of topographic maps. The geocorrected topographic maps and a few Global Positioning System (GPS) measured ground control points (GCPs) are used for georeferencing of SPOT 5 panchromatic and multispectral images.

In case of SPOT 5 panchromatic image geometric correction, 64 well-distributed GCPs and 30 well-distributed check points are selected over the entire scene of 60 x 60 km area. RMS error for GCPs was 4.60 pixels, while for check points RMS error was 4.30 pixels. In case of SPOT 5 multispectral image geometric correction, 64 well-distributed GCPs and 30 check points are selected. RMS error for GCPs was 2.30 pixels, while for check points was 2.42 pixels. A nearest-neighbour interpolation method is used for image transformation.

Landsat 7 ETM+ panchromatic and multispectral images are registered with SPOT 5 panchromatic geocorrected image. In case of Landsat 7 ETM+ panchromatic image rectification, 127 GCPs are used, while in case of multispectral image, 126 GCPs are used. In case of panchromatic image rectification RMS error was 1.79 pixel, while in case of MS image, RMS

error was 1.72 pixel. Again nearest-neighbour interpolation method is used for image transformation.

The relationship between the original pixel co-ordinates (x,y) and the transformed co-ordinates (u,v) in the new projection is specified by a pair of mapping functions:

$$\begin{aligned} u &= f(x,y), \\ v &= g(x,y), \end{aligned} \quad (1)$$

and by an equivalent pair of inverse functions:

$$\begin{aligned} x &= F(u,v), \\ y &= G(u,v). \end{aligned} \quad (2)$$

After geocorrection of all data sets, two test site common to all images of the delta were selected for the present study.

3.2.2 Image fusion

The radiation recorded by a sensor is dependent on, among other things, the spatial resolution of the sensor in relation to the spatial frequency of the terrain (Wulder *et al.*, 2000). Among the most important purposes of fusion of different resolution images is the production of spatially improved images suitable for classification. There are several approaches to fusion of remotely sensed images. Image fusion can be achieved at three levels, the pixel or signal level, the feature level, and the object level (Forstner, 1992; Pohl and van Genderen, 1998; Pohl, 1999). The signal level fusion techniques have been most common in remote sensing applications (Forstner, 1992).

In this study, two well-known methods based on the Brovery transformation and standardised principal components (SPC) transformation are applied to two pair of images acquired by HRG sensor onboard the SPOT 5 satellite and by ETM+ sensor onboard Landsat 7 satellite. Effects of image fusion are examined visually before classification.

3.2.3 Classification

Classification is the process of grouping of pixels or regions of the image into classes representing different ground-cover types. It chooses for each pixel a thematic class from a user-defined set. Two main digital image analysis techniques are available for the segmentation and classification of remotely-sensed, spectral data: unsupervised and supervised. In the unsupervised classification process, a computer algorithm selects a sample of pixels and clusters their radiance values

into similar spectral classes. These signatures are used by a maximum likelihood classifier to generate probability density functions and then each pixel is assigned to the probability density class it most closely resembles. The resulting classified image is assigned with landscape/land cover labels. In the supervised classification process, each pixel in an image is assigned to one of a number of classes (land cover types), by comparing the properties of the pixel with the properties of those pixels (i.e. training data) known to belong to the various classes. In essence, therefore, supervised classification forces the image classification to correspond to user-defined land-cover classes, but does not guarantee that the classes will be separable. In this study both unsupervised and supervised classification methods are used.

Three classes of wetland features are identified based on the objective of the study. They are open water, land-vegetation cover and an intermediate class, i.e. swamp with water-borne vegetation cover. The area is a complex wetland environment of marsh, swamp, channels, and shoreline. Methods to identify land and water in satellite imagery and their interface in wetlands were tested and evaluated to determine the most reliable technique. These methods include spatial image enhancement (convolution filtering), spectral enhancement (Principal component), and spectral classification (ISODATA and maximum likelihood supervised (ML) classification). The ML classifier is one of the most popular pixel-based classification method in remote sensing, in which a pixel with the maximum likelihood is classified into the corresponding landscape/landcover features. ML classifier is popular because of its robustness and simplicity.

Since the purpose of this study is to examine spatial resolution effects in extracting and classifying wetland features, it was necessary to select a proven, dependable process rather than employ a promising but experimental procedure. Accuracy assessment has been carried out through visual inspection, comparison with existing database like topographic maps, land cover maps and field knowledge.

4. RESULTS AND DISCUSSION

On an average half pixel (2m) accuracy was achieved in georeferencing of 1: 50,000 scale topographic maps. This is the best possible accuracy achievable as the quality of maps were poor. In case of SPOT 5 panchromatic image, 4.6 pixel (23m) accuracy was achieved, while in case of SPOT 5 multispectral 2.3 pixel (23m) accuracy was achieved. There were tremendous problems faced during georeferencing of images as topographic maps used are very old, published between 1971 to 1989.

In case of Landsat 7 ETM+ panchromatic image, 1.8 pixel (27m) accuracy was achieved, while in case of Landsat 7 ETM+ multispectral image, 1.7 (25.5m) pixel accuracy was achieved. The results of image fusion and classification are presented in Figures 5, 6, and 7.

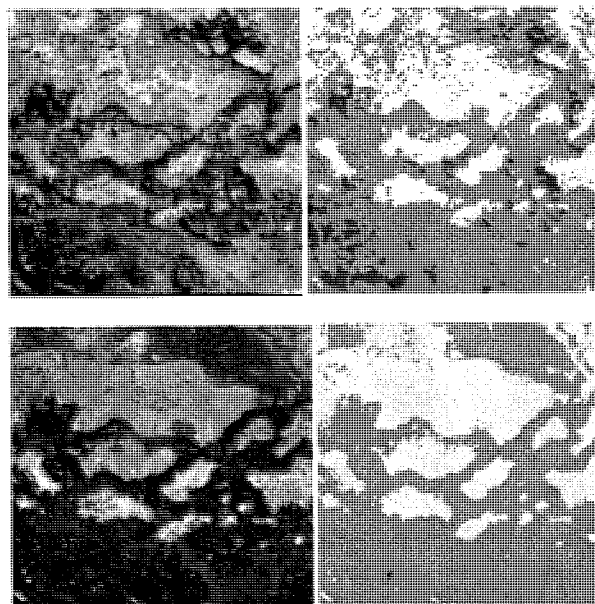


Figure 5. Upper: Left- Landsat 7 MS (30m) 4,3,2 band color composite (156 x 158 pixels), Right: 1-5 & 7 band classification result; Lower: Left- SPOT 5 XS (10m) 1,2,3 band color composite (465 x 472 pixels), Right: 1,3,4 band classification result

Figure 5 shows the results of supervised classification of Landsat 7 MS (30m) and SPOT 5 XS (10m) images. Unsupervised classification did not perform well because of the complexity of wetland features.

Table 2. Classification accuracy of SPOT 5 XS (10m) image

Class	Commission accuracy	Omission accuracy
Open water	99.9	98.8
Swamp	93.4	97.7
Land-vegetation	98.0	94.3

Proportion of unclassified pixels = 4%
 Overall accuracy = 96%
 Kappa statistic = 92.6%

Overall 96% class performance was achieved for three landscape classes from SPOT 5 multispectral image classification. Table 2 shows the accuracy achieved for each of the three landscape/land cover classes along with Kappa statistic. In case of Landsat 7 multispectral image classification, overall 85% accuracy was achieved. SPOT 5 multispectral image classification results in more detail information and more sharper object boundaries than Landsat 7 multispectral image classification.

The result also shows that the best wavelength region for discriminating water from land is the near-infrared and middle-infrared regions at wavelength between 0.75 μm – 2.35 μm . Previous studies also found that land/water interface is well defined in infrared spectral bands of Landsat Thematic Mapper images (Tao *et al.*, 1993; Tittley *et al.*, 1994).

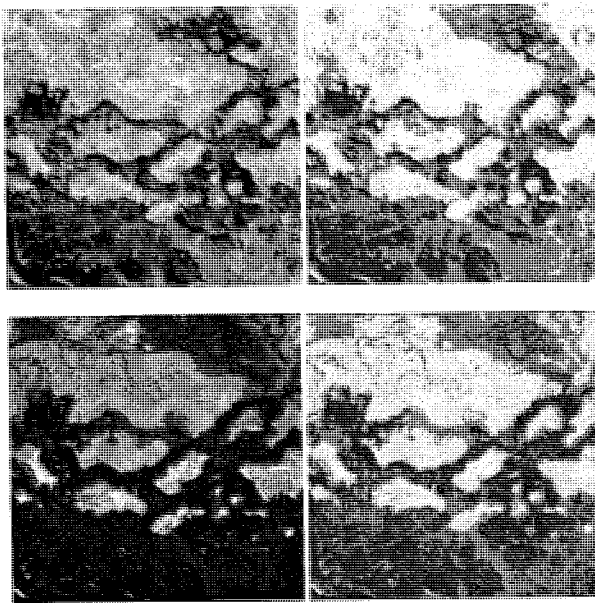


Figure 6. Upper: Left- Landsat 7 Pan (15m) and MS (30m) 4,3,2 band merge image (310 x 315 pixels); Right- classification result. Lower: Left- SPOT 5 Pan (5m) and XS (10m) 1,2,3 band merge image (929 x 943 pixels); Right- classification result.

Figure 6 shows the results of unsupervised classification of Landsat 7 ETM+ and SPOT 5 HRG merged images. Landsat 7 merged images are clustered into 5 clusters with ISODATA algorithms, then classified into three landscape/landcover classes. In case of SPOT 5 merged image, 9 clusters were generated using ISODATA algorithms, then categorise into three landscape/landcover classes. SPOT 5 merged image classification results in more detail information and more sharper object boundaries than Landsat 7 merged image classification. However, SPOT 5 merged image produces more spatial heterogeneity. Visual inspection and existing maps are used to examine the classification results.

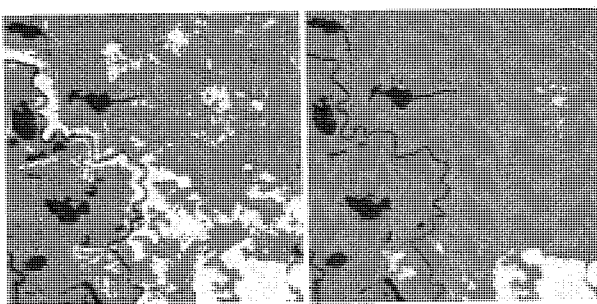


Figure 7. Left: Result of Landsat 7 merge image classification (Pan and band 4,3,2 combination); Right: Result of SPOT 5 merge image classification (Pan and band 1,2,3).

Figure 7 shows the results of supervised classification of Landsat 7 and SPOT 5 merged images for linear feature and water body feature extraction. SPOT 5 merged image provide better accuracy in extracting channels and open water bodies. Very sharp boundaries of these features can be extracted. Due to low spatial resolution, Landsat 7 merged image don't

provide detail spatial information in comparison to SPOT 5 merged image. The accuracy of classification is evaluated using visual inspection, field knowledge as well as existing map information.

5. CONCLUSIONS

Satellite remote sensing images provide a valuable tool for identification and characterisation of wetland features and related land cover types. The synergistic use of images from different sensors with varied spatial and spectral resolutions have the potential for better extraction and classification of features. Extraction and classification of landscape and land cover features from multispectral images as well as different resolution merged images of Landsat 7 ETM+ and SPOT 5 HRG sensors over the Okavango delta shows that classification accuracy and detail information content increases with increasing spatial resolution. Thematic information extraction was carried out using supervised and unsupervised classification to produce wetland landscape/land cover classes for different spatial resolution data set. The results indicate that as spatial resolution increases, high spatial frequency landscape/land cover features are extracted in increasing detail. However, increase in spatial resolution also increases spatial heterogeneity. Further investigation need to be carried out by extracting texture features from panchromatic images and merging them with multispectral images. The segmentation and classification of texture merged image have potentiality to increase classification accuracy.

ACKNOWLEDGEMENTS

The author thanks the Department of Water Affairs of Botswana for highly effective support during field campaigns and ETH Zurich for financial support.

REFERENCES

- Anderson, J.R., Hardy, E.E., Roach, J.T. and Witmer, R.E., 1976. A land use classification system for use with remote sensor data. U.S. Department of the Interior, U.S. Geological Survey Professional Paper 964, Washington D.C.
- Ashton, P. and Neal, M., 2003. An Overview of key strategic issues in the Okavango basin. In: *Transboundary River, Sovereignty and Development: Hydropolitical Drivers in the Okavango River Basin*, Turton, A., Ashton, P. and Cloete, E. (Eds.), AWIRU & Green Cross International, Pretoria & Geneva, pp. 31-63.
- Bailey, A., 1998. *Okavango: Africa's Wetland Wilderness*. Struik Publishers (Pty) Ltd., Cape Town, SA.
- Forstner, W., 1992. Fusing information in remote sensing. Invited paper, *XVII ISPRS Congress*, Washington D.C.
- Hughes, F.M.R., 1996. Wetlands. In: *The Physical Geography of Africa*, Adams, W.M., Goudie, A.S. and Orme, A.R. (Eds.), Oxford University Press, Oxford, UK, pp. 267-286.

McCarthy, T.S., Cooper, G.R.J., Tyson, P.D. and Ellery, W.N., 1998. Seasonal flooding in the Okavango delta, Botswana – Recent history and future prospects. *South African Journal of Sciences* 96, pp. 25-33.

Pohl, C., 1999. Tools and methods for fusion of images of different spatial resolution. In: *International Archives of Photogrammetry and Remote Sensing*, Valladolid, Spain, Vol. 32, Part 7-4-3 W6, pp. 7-11.

Pohl, C. and van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5), pp. 823-854.

Ross, K., 1987. *Okavango: Jewel of the Kalahari*. BBC Books, London.

Tao, Q., Lewis, A.J. and Braud, D.H., 1993. Change detection using multi-temporal feature space with digital TM Data. In: *Proceedings of 1993 ACSM/ASPRS Annual Convention and Exposition (Volume 2)*, New Orleans, Louisiana, February 15-18, American Society for Photogrammetry and Remote Sensing, Bethesda, MD, pp. 364-373.

Tittley, B., Solomon, S.M. and Bjerkelund, C., 1994. The integration of Landsat TM, SPOT, and ERS-1 C-Band SAR for coastal studies in the MacKenzie River Delta, NWT, Canada: A preliminary assessment. *Proceedings Second Thematic Conference on Remote Sensing for Marine and Coastal Environments*, January 31-February 2, New Orleans, LA, pp. I.225-I.236.

UNDP, 1976. Investigation of the Okavango as a Primary Water Resource for Botswana. Technical Report to United Nations Food and Agricultural Organization, UNDP/FAO, BOT/71/506, 3 Vols.

Wulder, M., Niemann, K.O., and Goodenough, D.G., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, 73, pp.103-114.

STARIMAGER – A NEW AIRBORNE THREE-LINE SCANNER FOR LARGE-SCALE APPLICATIONS

Koichi TSUNO*, Shunji MURAI**, Ryosuke SHIBASAKI**, Armin GRUEN and Li ZHANG***

* STARLABO Corporation, 1-20-1, Sanshin Bldg., Shibuya, Shibuya-ku, Tokyo 150-0002, JAPAN
tsuno-koichi@sei.co.jp

** University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN
shmurai@nifty.com, shiba@skl.iis.u-tokyo.ac.jp

*** Institute of Geodesy and Photogrammetry, ETH-Hoenggerberg, CH-8093, Zurich, SWITZERLAND
agruen@geod.baug.ethz.ch, zhangl@geod.baug.ethz.ch

Commission WG I/6

KEY WORDS: Airborne Three-Line-Scanner, Stabilizer, GPS/IMU, Digital Photogrammetry, Aerial Triangulation, DSM Generation, Object Extraction

ABSTRACT:

STARIMAGER is a helicopter-borne high resolution Three-Line Scanner (TLS) imaging system developed by STARLABO Corporation, Tokyo, Japan, jointly with the University of Tokyo and the Institute of Geodesy and Photogrammetry (IGP), ETH Zurich for large scale applications, especially line-shaped ground object recording (such as roads, rivers, railways), and remote sensing applications. As other airborne and space-borne linear array imaging sensors, STARIMAGER is equipped with GPS/IMU to record the exterior orientation parameters of each obtained image line per forward, nadir and backward views during the flight. Since a high-performance stabilizer is used, which absorbs the vibrations of the aircraft and stabilizes the camera's axis within a pixel, the resulting original images have very little waving and blurring, leading to fewer problems for post-processing. A feature of particular interest is that STARIMAGER can currently read out 10 channels simultaneously, so beside providing stereo capabilities it can also be used in multi-spectral modes. The TLS principle brings line perspective images with very little distortion in the flight direction, thus we get less occlusions especially in city and mountainous areas. It can generate seamless and high-quality ortho-images, which have already true ortho-image characteristics in flight direction. This paper focuses on the system hardware configuration and data pre-processing of the STARIMAGER, including the camera system, the stabilizer, GPS/IMU, data controlling and recording units. In addition we will report about the status of the application software development. Given the new sensor geometry a number of new features can be used in the processing algorithms, which are up to now not available on Digital Stations dealing with single frame images. This software includes a modern GUI with image handling and measuring in mono, stereo and multi-image modes. Furthermore we will address the following functions: image orientation and forward intersection, triangulation procedures with three different trajectory models, image rectification using different levels of approximation of object surfaces (plane, DSM), automated DSM generation, ortho-image generation, mono-plotting, 3D city model generation (geometric and textural parts), and support of feature and object extraction in manual and semi-automatic mode. We will also present a variety of projects, featuring the TLS principle in already existing and new application fields.

1. INTRODUCTION

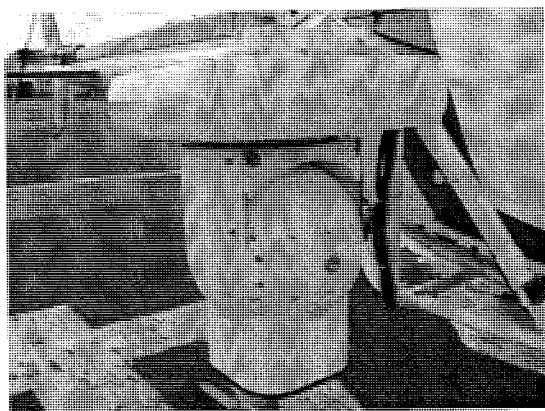


Figure 1. STARIMAGER

An airborne line sensor system is widely used in a satellite for easy multi-spectral image acquisition, while an airborne area

sensor system is coming to the market, positioned as an extended form of a conventional analogue aerial photo system. These two approaches have been intensively studied as airborne fully-digital spatial data acquisition systems. The line sensor system, however, is always ahead of the area sensor system because much more pixels per dimension can be accommodated in a CCD sensing device for the acquisition of a high-resolution image than the area sensor system due to the constraints of semiconductor manufacturing processes (see Murai, S., 1994). This paper explains the system configuration, features and applications of a state-of-the-art helicopter-borne three-line scanner (TLS) system, STARIMAGER, developed by STARLABO Corporation, Tokyo, Japan (see Figure 1).

2. STARIMAGER

2.1 Principle and features

The stabilizer with the TLS camera is equipped on the arm outside a helicopter and absorbs the vibration by the helicopter

movement. As shown in Figure 2, Four CCD line sensor packages are placed in parallel on the focal plane of the camera lens system. The three out of those sensors are for forward, nadir and backward looking. Each sensor package has 3 lines that are R (Red), G (Green) and B (Blue) to generate a color image. In addition, there is another CCD line sensor package for near infra-red (NIR) image between one for the backward looking and one for the nadir looking. Each line sensor can get a high-resolution, two-dimensional image in accordance with the flight of the helicopter, bringing 10 images in total simultaneously which are 100% overlapped with each other. The position and attitude data of the TLS camera for each line image are acquired with GPS (Global Positioning System) and IMU (Inertial Measurement Unit). An antenna for the GPS is equipped on the top of the helicopter, while a set of fiber-optic gyroscopes is incorporated on the TLS camera into the stabilizer. Such a spatial data acquisition system does not need aerial triangulation with ground control points, in principle. STARLABO Corporation acquires the patent right for the system in Japan, the USA, Europe, Australia, and Canada (see Murai, S., 1993).

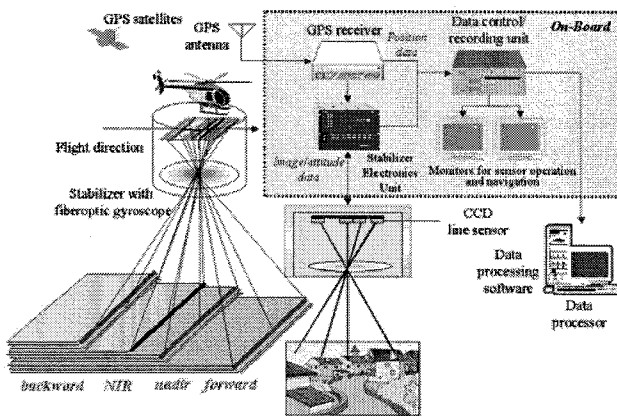


Figure 2. Principle of STARIMAGER

The advantages of STARIMAGER are described as follows (Tsunno, K., 2002a):

- 1) Photograph developing and scanning processes are unnecessary and there is no image deformation such as damage or deterioration on a film. Moreover, since the dynamic range of an output image that is linearly proportional to the luminance of the image plane is wide, it is possible to recover easily even an image part that is in the shade of a building, a cloud, and so on by increasing its image intensity, which leads to the reduction of field survey after data acquisition.
- 2) The ground control points (GCPs) for the orientation can be unnecessary, or fewer due to the existence of a GPS/IMU. It is, therefore, suitable for data acquisition at emergency, on the sea, and so on, where GCPs are difficult to set up.
- 3) Since, due to a high-performance stabilizer, an acquired original image is not waving and has no blurs, it is suitable for the image at emergency and the burden for post-processing can be reduced (see Figure 3). And, the oblique photography that is realized by slanting the stabilizer is effective in the texture acquisition of building walls.

- 4) A helicopter allows low altitude and low speed operation, leading to a high-resolution image. Since an influence by field visibility and weather condition is relatively small due to low altitude, photography can be done more frequently than a conventional aerial photography. And, it can take pictures of the linear-shaped objects efficiently by following it over at a low speed.
- 5) With the TLS principle, an image is very little distorted due to the height of objects in the flight direction (a line-perspective image as compared with a conventional central perspective image), and suitable for ortho-image generation. The system basically does not need mosaic processing in the flight direction without any scrapping parts and it can get the spatial data of linear-shaped objects seamlessly, such as roads, ways, rivers, etc. Moreover, the system can get an image that is redundant and little loss due to the photography in 3 different directions (3 different times), leading to less field survey burden after the data acquisition. Furthermore, a corresponding point searching in stereo matching is easy due to the stereo angles being constant, and measurement precision improves by triplet matching (see Shibasaki, R., 1987 and Morita, K., 2001). It eases a multi-spectral data acquisition and allows getting a color image by integrating RGB three line sensor images and an NIR image as an index which shows vegetation and water in the soil more distinctively than a color image.

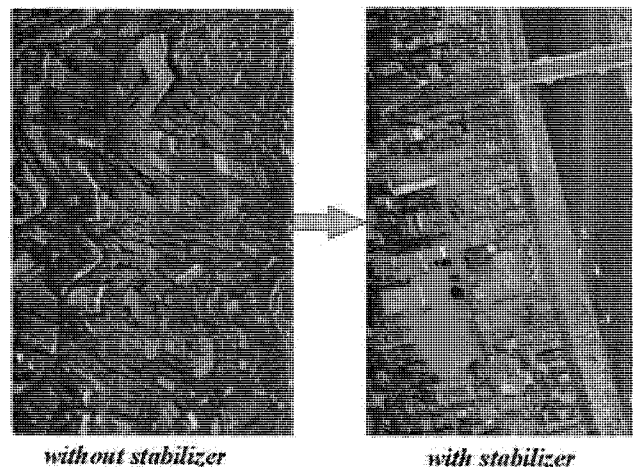


Figure 3. Stabilizer's Effect

2.2 System configuration

The focal distance of the TLS camera lens system is 60 mm, the stereo angles are 17, 23 and 40 degrees, between forward and nadir, nadir and backward, and forward and backward, respectively (see Table 1 for the STARIMAGER SI-250 system specifications). Each line sensor consists of 14,400 CCD pixels with 5 μ m spacing, and acquires 500 line images in 1 second (2 msec acquisition interval, and it is recorded in a controlling and recording device to install in the helicopter cabin. In addition, as shown in figure 4, a GPS antenna acquires the camera position signal at 5 Hz and the IMU on the top of the TLS camera acquires the camera attitude signal at 500 Hz. A Trimble MS750 serves as Rover GPS and collects L1/L2 kinematic data at 5 Hz and another Trimble MS750 serves as Base GPS on the ground.

The stabilizer has a vibration absorbing spring and 5 gimbals, absorbs the shake of the helicopter, and keeps the optical axis

direction of the TLS camera stable within a single pixel of the line sensor. Therefore, an image never waves due to the shake of the helicopter, and never be blurring.

As a result, STARIMAGER produces seamless high-resolution images (5 – 10 cm footprint on the ground) with three viewing directions (forward, nadir and backward). There are two configurations for image acquisition. The first configuration ensures the stereo imaging capability, in which the three CCD arrays working in the red channels are read. The second configuration uses the RGB CCD arrays (see Figure 5 for each spectral channel) to deliver color imagery.

configuration	item		specification
TLS Camera	CCD Elements	pixels/line	14,400
		pixel pitch	5 μ m
	Number of Sensors	10 (3 directions --with RGB - and NIR)	
	Intensity dynamic range	9 bits or more	
	Lens focal length	60 mm	
	Stereo angle	17°, 23°, 40°, etc.	
	Number of capturing lines		125, 250, and 500 lines/sec
Stabilizer	Angle resolution in attitude		0.00125°
	Spatial stability		0.00029°
	Maximum angle velocity		30° /sec
	Data output (acceleration/attitude)		500 Hz
GPS Receiver	2 ^d kinematics (post processes)	planimetric accuracy	2 cm + 2 ppm
		height accuracy	3 cm + 2 ppm
	Data output		5 Hz
Recorder	HDD recording	recording speed	150 MB/sec or more
		recording capacity	320 GB

Table 1. SI-250 System Specifications

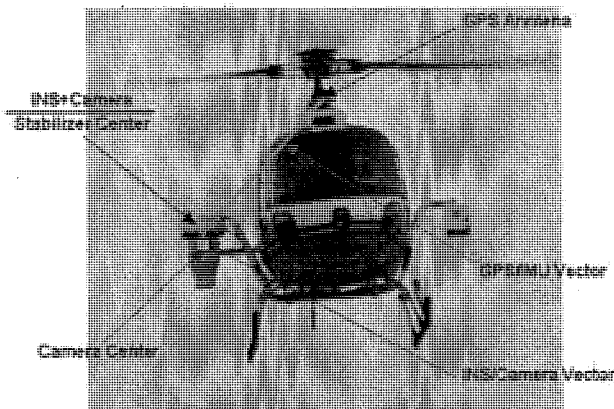


Figure 4. Sensors for STARIMAGER

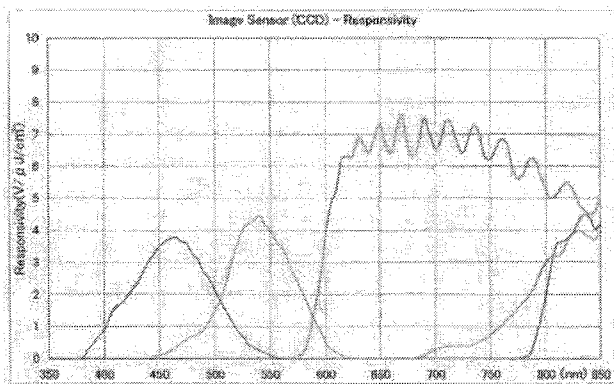


Figure 5. Spectral Channels for SI-250

After the collection of the GPS/INS raw data, the kinematic position and attitude data are calculated. This results in drift values for the IMU observations, which currently have to be

recovered, together with possible other systematic errors, by triangulation.

Unlike with frame-based photography, the three-line geometry is characterized by a nearly parallel projection in the flight direction and perspective projection perpendicular to that. The sensor model for the TLS images is based on modified collinearity equations and uses different forms of trajectory models. These models are used for the improvement of the measured exterior orientation parameters for each scan line by a modified photogrammetric bundle adjustment procedure, called TLS-LAB (Linear Array Bundle adjustment). This is part of a comprehensive package of new methods and the related software for the processing of TLS imagery, which is described in the following chapter.

2.3 Camera Calibration

STARLABO performs both laboratory calibration and field calibration before utilizing the TLS camera for 3D measurement to ensure a certain level of accuracy. The laboratory calibration (see Figure 6) includes geometric tests (CCD pixel position, distortion, MTF, etc.) and radiometric tests (sensitivity, PRNU, linearity, S/N, etc.). The field calibration uses self-calibration to determine the camera distortion data (see Chen, T., 2003).

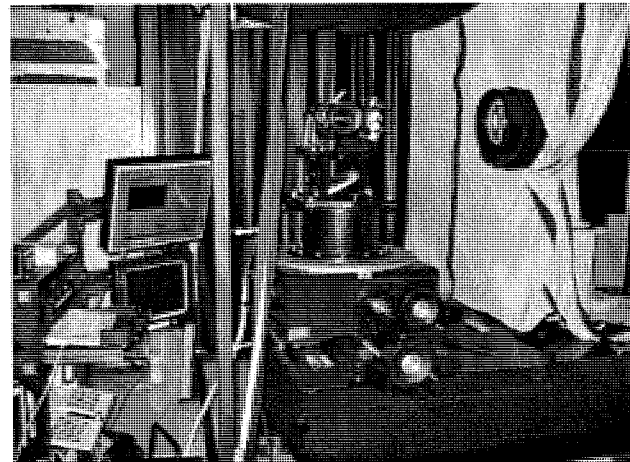


Figure 6. Optical System for Laboratory Calibration

3. DATA PROCESSING

3.1 TLS Digital Photogrammetry System

The outline of the TLS data processing chain can be drawn as shown in Figure 6 (see Gruen, A., 2001). The processing modules include:

- 1) User interface and measurement system: The user interface allows the display, manipulation and measurement of images. It includes the mono and stereo measurement modules in manual and semi-automated mode. It employs large-size image roaming techniques to display the TLS forward, nadir, and backward (plus other channels if available) view direction images simultaneously.
- 2) Triangulation: This module consists of two stages. At the first stage, the directly measured GPS/IMU data is taken as input and the exterior orientation elements for each scan-line are calculated/interpolated for the time of image

capture. The output of this procedure is called “raw orientation data”. The raw orientation parameters are already of fairly good quality and may be used in some applications right away. For high accuracy applications however we recommend a triangulation. The related software is a modified bundle adjustment called TLS-LAB. The software includes a special TLS camera model and offer three different trajectory models: DGR (Direct Georeferencing Model), PPM (Piecewise Polynomial Model) and LIM (Lagrange Interpolation Model). The self-calibration technique for systematic error compensation is currently implemented (see Gruen, A., 2002). The triangulation module also covers new methods for semi- and fully-automated pass and tie point measurement. Tie points in multi-strip/cross-strip configuration, with different image scales and image directions can be measured semi- or fully-automatically and semi-automatically through least squares matching.

- 3) Image Rectification: Here the raw level image data is transformed into quasi-epipolar form in order to reduce the large y-parallaxes caused by high frequency variations of the parameters of exterior orientation. This is absolutely necessary for smooth stereo viewing. Rectification comes in two modes. The coarse version just uses the orientation elements as given (or already derived from triangulation) and projects the raw images onto a pre-defined horizontal object plane. The refined version uses an existing DTM/DSM (of whatever quality) in replacement of the object plane. This latter method reduces the remaining y-parallaxes substantially. It should be realized that the notion of “epipolar” images does not exist for this kind of linear array-based images. Therefore these transformed images can be regarded as “quasi-epipolar”.

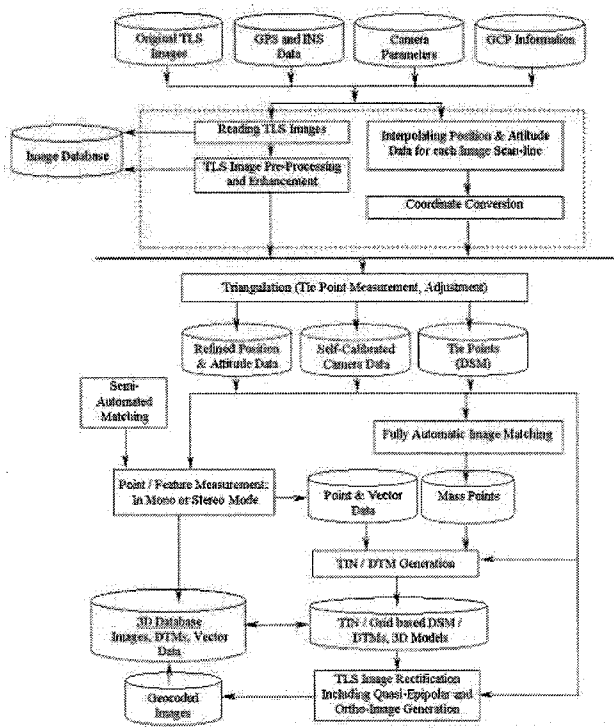


Figure 6. TLS Data Processing Chain

- 4) DSM/DTM generation: A new matching strategy is devised and implemented for the automatic generation of Digital Surface Models, from which Digital Terrain

Models may be derived. This strategy consists of a number of matching components (cross-correlation, least squares matching, multi-image matching, geometrical constraints, edge matching, relational matching, multi-patch matching with continuity constraints, etc.), which are combined in particular ways in order to respond to diverse image contents (e.g. feature points, edges, textureless areas, etc.). The matching module can extract large number of mass points by using multi-images. Even in non-texture image areas reasonable matching results can be achieved by enforcing the local smoothness constraints.

- 5) Ortho-image generation: This is a special solution for fast derivation of ortho-images given geometry and images.
- 6) 3D object extraction and modelling: For 3D object extraction two different semi-automated modes have been realized: (a) Point-based object extraction: This is achieved by interfacing the TLS software with CyberCity Modeler. For details see Gruen et al., 2003. The key points are manually measured in stereomode, producing a 3D pointcloud. The 3D model is generated automatically. (b) Line-based object extraction: The lines of interest are measured manually and monoscopically in the nadir image, while the corresponding lines in the forward and after images are then matched automatically. These procedures and products are represented by the following software components:
 - 7) TLS-SMS: User interface; image measurement in mono and stereo 3-ray forward intersection (point positioning) Image and shadow enhancement
 - 8) TLS-IRS: Quasi-epipolar rectification to plane or via DTM/DSM and Ortho-image generation
 - 9) TLS-LAB: sensor/trajectory modeling, triangulation, Automatic and semi-automatic tiepoint generation
 - 10) TLS-IMS: Image matching for DSM/DTM generation, DSM/DTM modeling and interpolation
 - 11) Adaptation of methods and software for ALOS/PRISM processing
 - 12) Feature/object extraction, e.g. city modeling: CC-TLSAutotext

3.2 Image GIS

It has been an issue to map and plot costly based on aerial photos to construct a GIS (Geographic Information System) system. A digital map may not be necessary to use from the very beginning to fulfill most of field applications. A new concept called, “Image GIS” has been presented to solve the issue. Figure 11 shows an example viewer materializing such a concept. An ortho-image goes as a background image for a 2D map that is obtained from another source and allows one to measure distances and areas on it. In addition, the Image GIS viewer is facilitated with a mono-image measurement system, where, as the center of the ortho-image moves, the target image for the measurement system moves. When one wants to measure the 3D position of a point precisely by pointing to it on the ortho-image, the viewer brings one with its corresponding image set (forward, nadir and backward) for the 3D measurement. In doing so, one can grasp a real 3D situation

surrounding an underlying target object without going to an actual spot and then store the 3D data into a database attached to the GIS viewer. The image GIS viewer can measure the cross-section of a linear-shaped object like a river, a road, etc as shown in Figure 7.

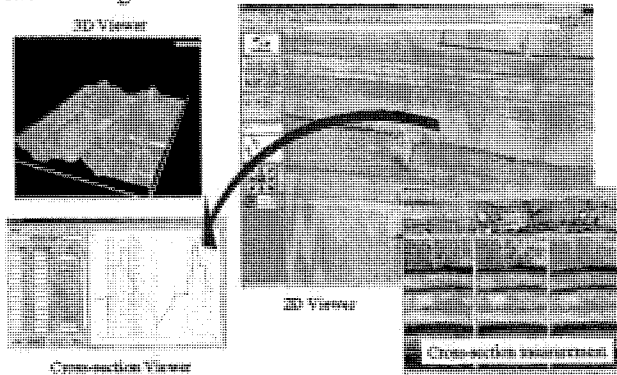


Figure 7. Image GIS Viewer

4. APPLICATIONS

The features of the TLS data are stimulating more demands for 3D seamless information of linear-shaped man-made objects and terrain with high resolution and efficiency. Linear-shaped man-made objects include roads, bridges, railways, power cables, pipelines, etc. for investigation before the construction of those, maintenance and management after the construction, and base data for a variety of GIS's (see Tsuno, K., 2002b). Figure 8 is an example of river, and can be used for the investigation of vegetation, the gravel grain diameter distribution of a dry riverbed, the river floor profile, and so on as described in Fukami, K., 2002.

It also can be applied for the base ground data for flooding simulation with high precision that can judge whether water goes over a Japanese raised floor or below, which can not be judged based on conventional 50 m-meshed terrain data. Three-dimensional city models with high definition, which is expected to use for city planning, landscape simulation, auto-navigation, gaming and so on. When the helicopter flies over along a road, wall texture facing the road can be acquired with a nadir-looking image. The wall textures that are perpendicular to the road can be acquired with either forward-looking or backward-looking image. Textures can be semi-automatically mapped onto 3D polygons. The acquisition of the texture of the building walls facing a trunk road also can be efficiently acquired with oblique-viewing image.

Figure 9 is an example of perfect ortho-images (TrueOrtho™) where the TLS data is taken with highly overlapped between courses and processed with the cooperation of ISTAR in France. A conventional ortho-image has leaning buildings with its sidewalls which occlude roads because the image is generated based on the terrain data only, while a perfect ortho-image is made based on the height of buildings as well. Consequently, it is useful to avoid mistakenly digging in case of construction because the position of gas or water pipes under roads can be well described with high precision and reality.

Furthermore, using the characteristics that can acquire an image of three directions in the constant time lag, the systems can measure the speed of a car, running water, etc. And, it can be

applied to the traffic flow measurement which goes along a road and illegal parking investigation judging from the image deformation of a car, a distance from a road edge, etc.

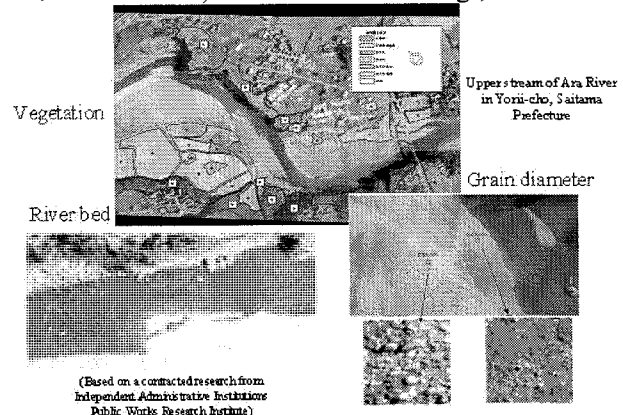


Figure 8. River Environment Investigation

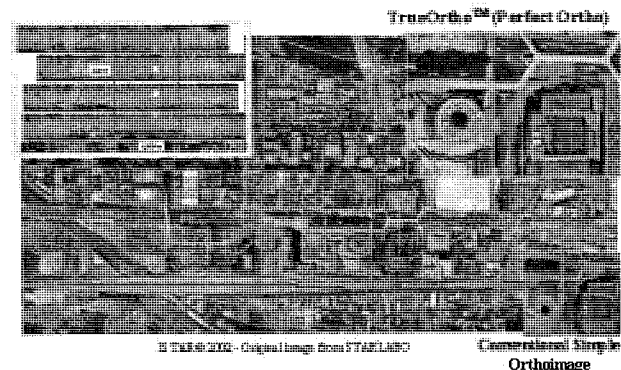


Figure 9. An Example Perfect Ortho-image

5. CONCLUSION

A day when a conventional spatial data acquisition tool based on aerial photographs is substituted with TLS will come soon. The paper presented the principle of TLS and its applications at an early stage. The biggest issue is to improve the measurement accuracy of position and attitude of the TLS camera center by the GPS/IMU for higher accuracy measurement that needs as few GCPs as possible. Furthermore, the extraction of changes between images of two different times, automatic extraction of man-made objects such as buildings and roads (as described in Nakagawa, M., 2001) and speed measurement of cars and rivers should be studied.

Acknowledgements

The authors would like to gratefully acknowledge Prof. A. Gruen, ETH and his lab members, and Prof. R. Shibasaki, the University of Tokyo and his lab members for constructive and useful discussions and advices.

References

- Chen, T., 2003. Development and calibration of the airborne three-line scanner (TLS) imaging system. *Photogrammetric Engineering & Remote Sensing*, Vol. 69, No. 1, January 2003, pp. 71 – 78.
- Fukami, K., 2002. A study on an information collection method for river blocks with TLS. In: *JSPRS Conference*, J-4, 2002.7, pp. 185 – 188.

Gruen, A., 2001. TLS Data Processing. In: *3rd International Image Sensing Seminar ISPRS*, Sep. 2001.

Gruen, A., 2002. Sensor modelling for aerial mobile mapping with three-line scanner (TLS) imagery. In: *Int. Arch. of Photogrammetry and Remote Sensing*, Vol. 34, Part 2, 2002, pp. 139-146.

Gruen, A., 2003. Generation of 3D City Models with Linear Array CCD-Sensors. In: *Proceedings Int. Conference on Optical 3D Measurement Techniques*, Zurich, September 2003, pp.

Morita, K., 2001. Accuracy evaluation of aircraft-borne photographic device. In: *JSPRS Conference Fall 2001*, 2001.11.

Murai, S., 1994. 3D measurement with three-line scanner (TLS). *Photogrammetry and Remote Sensing*, vol. 33, No. 5, 1994, pp. 21 - 25.

Murai, S., 1993. Airborne photographic apparatus (Patent 02807622), 1993.12.

Nakagawa, M., 2001. Study on making 3D urban model with TLS image and laser range data. In: *Asian Conference on Remote Sensing*, November 2001

Shibasaki, R., 1987. A simulation on improvement of the accuracy and the stability of stereo matching using triplet linear array sensor data. *Photogrammetry and Remote Sensing*, vol. 26, No. 2, 1987 pp. 4 - 10.

Tsuno, K., 2002a. Practices in spatial information technology: Three-line scanner and its applications, *JSPRS 2002*.

Tsuno, K., 2002b. Three-line scanner and its applications. *Photogrammetry and Remote Sensing*, Vol. 41, No. 4, 2002 pp. 37 - 40.

APPROXIMATION OF NON PROJECTIVE MAPPING AND THEIR EFFICIENT APPLICATION FOR A GEOMETRICALLY BASED 3D POINT DETERMINATION USING MULTIPLE VIEWS

Kirsten Wolff

Institute of Geodesy and Photogrammetry
ETH Hoenggerberg
CH-8049 Zurich, Switzerland
wolff@geod.baug.ethz.ch

Commission III, WG III/8

KEY WORDS: Orientation, Geometric, Distortion, Matching, Underwater

ABSTRACT:

In this paper an unusual taxonomy for optical mappings is introduced based on their geometric characteristics: 1. type of projection center (single viewpoint, non single viewpoint) and 2. type of transformation (projective, non projective). Under this background we survey the multi media geometry (refraction resulting from different optical media). Strict physical models of non projective mappings can be very complex in dependency on their geometric nature. Therefor different methods for reducing the complexity exist. This paper describes a method of ascertaining a virtual camera to approximate non projective mappings by a projective model and their application for a 3D point determination using multiple views with non projective multi media geometry. As will be seen, the approximation can be used without loosing the quality of the strict model significantly. For the matching process we introduce a new algorithm for multiple views based on geometric constrains alone which uses all images simultaneously.

1 INTRODUCTION

1.1 Motivation

The nature of an optical mapping process between a 3D object space and a 2D image space depends on the geometry of the imaging system, its physical laws and the scene structure.

In this context we use the term *imaging system* instead of camera system, because it should contains all parameters, which have an effect on the nature of the optical mapping. In particular the effect on the way of mapping light-rays, influenced by light refraction or reflection.

Based on the different nature of optical mappings and their characteristic, several kinds of classifications exist (e.g in (Hartley and Zisserman, 2003) whether they have a finite centre or a centre "at infinity", or whether they preserve straight lines or not).

An unusual feature for a classification of mappings is the kind of image distortion, resulting from light refraction or reflexion of the mapping rays. Therefore we introduce a classification of imaging systems, based on the following geometric features, which influence the characteristics of image distortions:

<p>Geometrical characteristics of optical mappings:</p> <ol style="list-style-type: none"> 1. Single Viewpoint (SVP) or Non Single Viewpoint (NSVP) mapping rays intersect in one single point or not 2. Invariance or variance of straight lines straight lines in the scene appear as straight lines in the 2D image space (<i>projective mapping</i>) or appear as curves

From the combination of these geometric features, we get a *classification of imaging systems* with three different classes, summarised in table 1. It is based on the classification of optical

Table 1: Classification of imaging systems

Class	Mapping	Viewpoint	Imaging System	Modeling distortion
1	Projective	Single Viewpoint	Pinhole	-
2	Non Projective	Single Viewpoint	Wide-angle cameras, fish-eye cameras, central catadioptric cameras, Approximation of objective distortion	based on position in image space
3	Non Projective	Non-Single Viewpoints	Wide-angle cameras, fish-eye cameras, catadioptric cameras, camera clusters, moving cameras, multi-media geometry, objective distortion	based on position in object space

mappings in (Wolff and Förstner, 2001) and on the taxonomy of distortions published in (Swaminathan et al., 2003).

1. Class 1 is the *perspective mapping*, also named *pinhole camera*. It is the most specialized and simplest model, where the straight projecting rays intersect in a single viewpoint (the pinhole) and preserves straight lines. This results in no image distortions (not taking distortions into account, which result from the perspective mapping). All cameras modelling central projection are specialisations of the general *projective camera*, therefore we use the term *projective mapping* which could be presented by a projective model. Every deviation from this model causes image distortion.
2. Class 2 is created by mappings with single viewpoints, not preserving straight lines. However, this leads to image distortions, which depends on the image position. Its influence

or error can be calculated in dependency on their image coordinates (*image space based*). No information about the scene structure are needed. An example for such a projection is the general used model for the optical distortion. Strictly, it has not a single viewpoint, but the accuracy of this approximation is well enough.

3. Class 3 is formed by projections with *non single viewpoints which do not preserve straight lines*. The projection rays are no straight lines and they do not intersect in one point. But their envelope forms a locus of viewpoint in three dimensions which is called caustic surface or just caustic (Swaminathan et al., 2001). The resulting image distortions are called caustic distortions. Their exact determination bases on the position of the observed feature in object space (*object space based*). Therefore information about the scene structure are necessary to determine the influence of the distortions. Imagesystems like wide-angle, fish-eye and catadioptric cameras with a spherical and conical reflector based design (Nayar et al., 2000), camera clusters, strict model of objective distortion and multi media geometry (e.g. air and water) belong to this class. In section 2 we will present the caustic of a multi-media system.
4. The combination of a non single viewpoint with an invariance of straight lines is under the valid physical laws not possible.

The influence of image distortion using imaging systems with non single viewpoints is object space based, that means it cannot be determined or corrected without any information of the scene structure. If no information about the scene structure are given, it is necessary to make some assumption about the scene structure (e.g. (Swaminathan et al. 2003)). For the mapping process between the object and the image space, special algorithms are needed. For example the iterative algorithms for the multi media geometry in (Maas, 1995), which could be very complex.

Another method is to replace the non single viewpoint by a single viewpoint, so that the mapping process can be modeled without any information about the object space. Swaminathan et al. presented in (Swaminathan, 2001) a method to determine a single viewpoint by estimation the best location to approximate the caustic by a point for catadioptric cameras. This methods based on the determination of singularities of the caustic.

A method which is used here to define a single viewpoint is first mentioned in (Wolff and Förstner 2000) and was published in more detail in (Wolff and Förstner 2001): the explicit strict physical model with non single viewpoints is replaced and approximated by a less complex projective mapping with a single viewpoint. Therefore no pre-informations about the scene structure are needed. The estimation of the approximation is posed as the minimization of the back projection error in image space. The introduced approximation is applicable for all kinds of optical, non projective mappings. The degree of approximation can be augmented by partitioning the object space into small segments and calculating a local approximation for every part of the object space separately. For this partitioning we need the extension of the observing area approximately. The method was presented in (Wolff and Förstner 2001) used for a matching process based on the trifocal tensor.

1.2 Goal of this paper

In the context of non projective projections, the paper makes the following key contributions:

- Under the background of the taxonomy of imaging systems we survey the non projective multi media geometry (projecting rays passes different media e.g. air, perspex and water). It belongs to class 3 with a caustic as a non single viewpoint.
- We present a new image point matching algorithm for a 3D reconstruction using multiple views, based on geometrically constraints alone. The method uses all images simultaneously. The test of hypotheses is placed on object space.
- The approximation for a non projective mapping by a virtual, projective camera is used for the image point matching process for multiple views with multi media geometry. As we will see, this is implemented without losing the quality of the strict model significantly.
- Different quality tests for the approximation and the point matching algorithm are realized.

1.3 Projective Geometry

We use multiple-view geometry as it has been developed in recent years and is documented in (Hartley and Zisserman 2003).

Assuming straight lines preserving mappings, the projection of object points \mathbf{X} to image points \mathbf{x}' can be modeled with the direct linear transformation (DLT):

$$\mathbf{x}' = \mathbf{P}\mathbf{X} = \begin{pmatrix} \mathbf{1}^T \\ \mathbf{2}^T \\ \mathbf{3}^T \end{pmatrix} \mathbf{X} = \mathbf{K}R(\mathbf{I} - \mathbf{Z})\mathbf{X}$$

for object points \mathbf{X} represented in Plücker coordinates. \mathbf{P} is the projection matrix, \mathbf{K} the calibration matrix, \mathbf{R} the rotation matrix and \mathbf{Z} the projection center of the camera.

2 GEOMETRY OF IMAGING SYSTEMS WITH NON SINGLE VIEWPOINTS

2.1 Caustics as Loci of Viewpoints

For the modeling of point projection we need two relations:

1. A projection relation predicting the image point \mathbf{x}' of a given object point \mathbf{X} .
2. An inverse projection relation, giving the mapping ray \mathbf{L} in the object space. In case of projective mapping a light ray is built by the projection center and the image point. In case of non projective mappings only that part of the broken ray is important, which intersects the object point.

For Class 1 and 2 of our classification the realization of these two relations is geometrically trivial. The mapping ray is built by the object point or rather the image point and the projection center. In the case of image distortion a correction of the image points can be calculated image space based.

For class 3 relation 2 is also trivial. The projecting light rays change their direction because of refraction and reflection (see Fig. 1). These changes can be directly determined using the Snell's refraction law and reflection law. Relation 1 is not as trivial like the others, because the direction of the ray coming from the object point is not directly determinable if the object point and the physical pupil of the lense is given alone. But, as seen in Fig. 1, the envelope of the rays, which do not intersect in one point, forms a locus of viewpoints in three dimensions, the so called caustic. The light rays in object space are the tangent on this surface. Each point on the caustic surface represents the three-dimensional position of a viewpoint and its viewing direction. Thus, the caustics completely describes the geometry of the catadioptric camera (Swaminathan et al., 2001).

Swaminathan et al. uses this caustic in (Swaminathan et al., 2001) for an analysis of an catadioptric camera for its characteristics like field of view, resolution and geometric singularities. They also present a calibration technique to estimate the caustic surface and camera parameters for a conic catadioptric system using known camera motion.

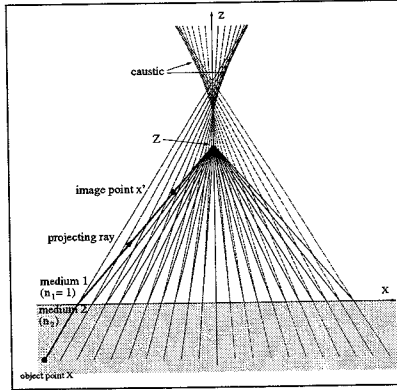


Figure 1: Geometry of a multi media system consisting of air and water, showing the projecting light rays and the caustic of the system.

2.2 Multi Media Geometry

Multi media geometry results from the observation of an object through several transparent physical media with different refraction indices. The light rays are refracted at the interface surfaces between two media. The classical application is to acquire images of objects through the media air, perspex and water, e.g. the 3D tracking of particles for modelling in fluid dynamics (Maas, 1991) or the observation of fluvial transportation processes under water (Wolff and Förstner, 2000).

The parameters of an imaging system with multi media geometry consists of the parameters of the used camera and the parameters of the refraction (exterior orientation of the refracting interfaces and the refraction indices). As seen in Fig. 1, refracted light rays coming from object points do not intersect in one point, but their envelope forms a caustic (Wolff and Förstner 2001). The imaging systems involve the projection center of the camera and the non single viewpoint. Therefore the mapping in general is not projective, it does not preserve straight lines. The resulting image distortions could not be modeled image space based and the complexity of the calculation of the mapping between object and image space increase.

For the strict realisation without approximation of the two relations mentioned in section 2.1, we use the multi media geometric models described in (Maas, 1995). He used a strict multi media geometric model based on Snell's Law for the effect of a ray being twice broken. For relation 1. a radial shift of each object point relative to the nadir point of the camera can be calculated and used as a correction term in the collinearity condition. This projection actually inverts a ray tracing process. As the algebraic expression cannot be directly inverted, the calculation is an iterative procedure. For relation 2 the Snell's Law is used to inverse the projection directly by ray tracing.

3 APPROXIMATION OF NON PROJECTIVE MAPPINGS BY MEANS OF VIRTUAL PROJECTIVE CAMERAS

For the approximation we assume the expected volume V in object space to be approximately known. We assume the non pro-

jective mapping $\mathbf{x}' = f(\mathbf{X})$ to be known, i. e. the orientation and calibration to be available. The task is now to find a projective mapping $\bar{\mathbf{x}}' = \mathbf{P}\mathbf{X}$ such that the systematic errors of the image coordinates $\bar{\mathbf{x}}' - \mathbf{x}'$ are minimum. This leads to the well known problem: determine \mathbf{P} such

$$\Omega = \int_{\mathbf{x} \in V} (\mathbf{x}' - \mathbf{P}\mathbf{X})^T (\mathbf{x}' - \mathbf{P}\mathbf{X}) d\mathbf{X} \rightarrow \min,$$

where the integral is to be taken over the expected volume V of interest. The camera with this projection matrix \mathbf{P} is a *virtual projective camera*, which we use for the approximation. Its quality impair with the enlargement of the volume V . To get the acquired quality of the approximation, we may partition the object space and in corresponding way the image space such that for every part V_i , $\sum_i V_i = V$ of the volume a local DLT with the local projection matrix ${}^a\mathbf{P}$ is solved. To define the partition of the volume V a priori quality analysis of the projective model have to be carried out (Wolff and Förstner, 2001). The determination of the Direct Linear Transformation contains the following steps:

Determination of the Direct Linear Transformation:

1. determination of the parameters of the imaging system orientation using the strict model
2. define of a regular $o \times p \times q$ grid of object points \mathbf{X}_i , $i = 1..(o \cdot p \cdot q)$ lying in the expected object volume.
3. a priori quality analysis of the projective model
4. subdividing of the object volume V into parts V_i accordingly to the a priori quality analysis
5. calculation of the corresponding image points \mathbf{x}'_i , $i = 1..(o \cdot p \cdot q)$ using the strict non projective model
6. estimation of the projection matrices ${}^a\mathbf{P}$ for every V_i by minimizing the back-projection error using the object points \mathbf{X}_i and the image points \mathbf{x}'_i .

If enough well distributed control points and corresponding non projective image points can be measured, the DLT could be also determined by using these real data.

4 MATCHING AND 3D DETERMINATION OF POINTS FOR MULTIPLE VIEWS USING GEOMETRICALLY CONDITIONS

4.1 The Algorithm

The algorithm for finding matching candidates in multiple image views assumes the extracted points of $n > 2$ images to be given with their projection matrix \mathbf{P} . The radiometric information of the images were only used to extract the image points and are not required for the matching algorithm. It has the following characteristics:

1. only geometric conditions: all projection rays of corresponding image points intersect in one object point.
2. all images are used simultaneously.
3. test of hypotheses placed in object space.
4. if necessary, correspondency tests using radiometric information can be easily implemented.

Methods for multiple image matching based on geometrically constraints use often only three or four images simultaneous, like the matching algorithm using the trifocal tensor (Wolff and Förstner, 2000), the quatrifocal tensor or using the intersection of two epipolarlines (Maas 1997). To take all images into account, the algorithms are used for different combinations of three or four images.

The number of ambiguities using geometrically conditions alone, was examined by Maas in (Maas, 1992) for different numbers of images. The complexity of the matching strategy arise with the number of images, but the high amount of ambiguities to be expected requires more images to be reduced. Therefore we use $n > 2$ images and the constrain for an object point, that its image points are seen in at least three images, to eliminate wrong hypotheses.

The presented algorithm uses all images simultaneously, therefor the test of hypotheses is realised in the object space. The geometric condition is that all projection rays of corresponding image points intersect in one object point. Therefor we first find matching hypotheses by using the epipolar constraints defining one image as the starting image. The epipolarlines between the starting image and all other images are calculated and every image point close to the epipolarline is a hypotheses for a corresponding point to the point in the starting image. The epipolarline can be shorten by considering the height extension in object space. To get also the points, which are not seen in the first image, but maybe in at least three other images, this step should be calculated also for other images as starting images. The number of starting images depends on the constellation of the image system. Then we determine the object points belonging to these two point hypotheses. The result is a 3D point cloud, where a group of at least m close points define one object point. The number of points m depends on the number of starting images. To test the hypotheses of correspondences, a clustering of the point cloud is calculated using the k-means algorithm. The resulting clusters containing at least m points belong to one object point. The mean value of the points in one cluster is a first approximation of the 3D point determination. If a higher quality is required, all points can be finally determined with by estimating a bundle adjustment. Therefor we use the image point correspondences resulting from the points belonging to one cluster. We summarize the algorithm into the following steps:

The main steps of the algorithm for 3D prediction of points are:

1. Extraction of points $\mathbf{x}_i^{(j)}$ in all n images, where j is the number of the image and i the number of the point.
2. define one image as the starting image a .
3. for all points \mathbf{x}_i^a in image a determine hypotheses of point correspondences using epipolar lines in all other images.
4. define another image as second starting image b .
5. for all points \mathbf{x}_i^b in image b determine hypotheses of point correspondences using epipolar lines in all other images.
6. if necessary repeat point 4 and 5 for as much different images as it is convenient.
7. clustering of the 3D point cloud P resulting from point 2 to 6 \rightarrow approximated 3D object points $\bar{\mathbf{X}}_i, i = 1..m$.
8. final bundle adjustment of all matched points, using $\bar{\mathbf{X}}_i, i = 1..m$ as approximated values \rightarrow final object points \mathbf{X}_i .

If the imaging system and the resulting images are not projective, then there exist two different possibilities:

1. The specialized strict physical model of the mapping process will be implemented in the algorithm, which is sometimes not possible, or the strict physical model can be very complex and the computational time can rise in dependency on the algorithm.

2. An approximation for the non projective mapping is used for the matching process. For the a priori quality control of the percentage reduction of computation complexity for the replacement of the multi media geometry by a normalized projective model see (Wolff and Förstner, 2001).

4.2 Application for Non Projective Views

The application of the approximation by a virtual projective camera presented in section 3 contains the following steps:

Implementation of the virtual camera for an effective 3D point determination using non projective mappings

1. determine virtual projective mappings aP for the observation space
2. matching the image points using aP
3. final bundeladjustment using the strict model

5 EXAMPLE AND QUALITY CONTROLS

5.1 Data: a surface of a fluvial sediment

Our work on using multi media geometry is motivated by investigations on the generation of fluvial sediments (Wolff and Förstner, 2000). The aim is to derive a physical model of the underlying process of the dynamical sediment transport. The surface of the water is smoothed by a perspex pane. We get the standard case of multi media geometry: air, perspex and water with plane interfaces. The observed sediment surface is shown together with the extracted points of one image in Fig. 2 (for the extraction of interest points see (Förstner, 1994). The surface of the sediment was formed by a jet of water hitting the sediment. We used four Sony XC-77 CE cameras (748 \times 564 pixel) for the acquisition of the images.

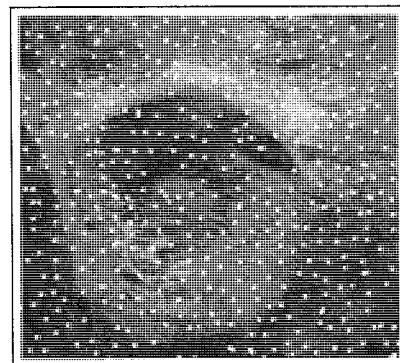


Figure 2: Image of the sediment surface with extracted points.

5.2 Determine the reference data using the strict model

To get reference data for the quality analysis of the matched image points and determined object points we carried out the presented algorithm using the strict multi media model. We use the same software and values for its parameters to calculate reference data and the approximated data.

5.3 Quality analysis

For the quality analysis of the 3D determination of points using the approximation, we want to examine the following points:

Table 2: Quality analysis 1: estimation of the virtual cameras

observing area: left lower corner	(3, 3, -17) [cm]
observing area: right upper corner	(21, 23, -11) [cm]
number of points used for estimation	845
distance between points	2 [cm]
$\hat{\sigma}_{approxVC}(cameraC1)$	0.04 [pel]

Table 3: Coordinates of the projection centre of camera C1 for the strict model and the approximation VC

Projection center	X_1 [cm]	X_2 [cm]	X_3 [cm]
Strict model	3.08	4.53	63.46
Approximation VC	3.07	4.51	85.98

Quality analysis:

1. A priori quality DLT: residuals as backprojection errors in image space
2. Quality DLT: residuals in object space for new points
3. Quality point matching algorithm: comparison of the reconstructed points (before final estimation) using the strict and the approximated model
4. Quality point matching algorithm: comparison of the reconstructed points (after final estimation) using the strict and the approximated model

5.4 Prediction of 3D points using the virtual camera

5.4.1 Estimation of the virtual cameras To define the segmentation of the object space a priori quality test have to be calculated (see (Wolff and Förstner, 2001)). These a priori tests show, that the determination of only one virtual camera (VC) for the whole object space is enough. For the position of the four cameras see Fig.4.

To investigate the quality of the determined virtual cameras (Quality analysis 1), we project the object points which were used for the estimation of P into the image space and get the image points \bar{x}' . The estimated DLT (11 independent parameters) yields residuals $\bar{x}' - x'$ being systematic errors. To get an a priori quality of the projective model we give the r. m. s. error

$$\hat{\sigma}_{approx} = \sqrt{\frac{\sum_i (\bar{x}'_i - x'_i)^2}{2n - 11}}$$

where n is the number of points used.

Tab. 2 gives the entities and results of estimating the virtual cameras of camera C1. The number of points used for the estimation need not to be as high as in this case. Tab. 3 gives the coordinates of the camera projection center for the three different orientations. The multi media geometry influence mostly the height of an object point, which is here the X_3 coordinate of the projection center. Therefore the projection center of the two orientations differ mostly in the height.

5.4.2 Results of the point matching using the approximation

As mentioned above, the algorithm should be calculated for different starting images, to guarantee that also the points, which are not extracted in the starting image, can be found. Here we use four cameras, every camera could see the whole object scene. Together with the constrain, that at least three corresponding image points of an object point are needed, it is enough to have two different starting images. Therefore and because of the constraints,

that the image points of an object point should be seen in at least three image points, we got the constraint for our clustering algorithm: a group of at least three define an object point.

First, we want to examine if the constraint for a object point, that at least three close points in a group define an object point, is sufficient. Fig. 3 shows the hypotheses of two matched image points by there corresponding object points (seen from the side). The distribution of the 3D points shows a very dense part, where the sediment surface is supposed to be. All the other points might be wrong hypotheses and should be deleted by the clustering algorithm. Fig. 4 shows the results after the clustering. All points which differ significantly from the surface are eliminated (Fig. 4 a). Fig. 4 b) shows the distribution of the object points on the surface, which are evenly distributed.

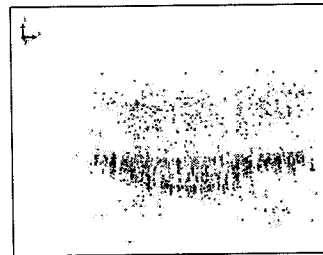


Figure 3: Hypothesis of 3D point matchings before clustering. A group of at least three points define an object point.

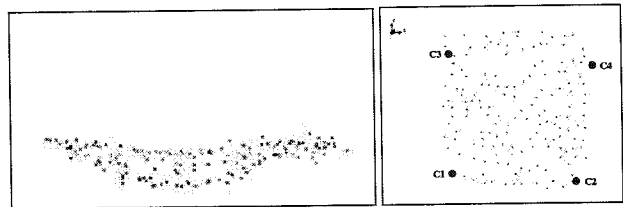


Figure 4: Results after clustering the point hypotheses. The right figure shows the point cloud from the side, the left figure shows is from above together with the positions of the cameras.

Using the strict model gave 156 reconstructed 3D points, the use of the virtual camera VC found 161 points. For the quality analysis 3. we have to compare the two sets of points. Therefore a threshold ϵ is defined, so that a point \bar{X}_k is defined as equal to a referent point X_j if $|\bar{X}_k - X_j| < \epsilon$. The number of points found as equal in dependency of the threshold is shown in Fig. 5.

The main influence of the approximation refers to the height of the object points. The r. m. s. error of the X_3 coordinate of the reconstructed object points $X = (X_1, X_2, X_3)$ is

$$\hat{\sigma}_{X_3} = \sqrt{\frac{\sum_i (\bar{X}_{3i} - X_{3i})^2}{n - 1}}$$

where n is the number of points used. The error of the approximation is given in table 5 in comparison to the referent data before calculating the final bundle adjustment.

5.5 Final 3D determination of the predicted points using the strict model

After the matching process, including an approximated determination of the object point, we calculate a final bundle adjustment for the strict model and for the approximation VC. The clusters resulting from the clustering algorithm contain that points, which were found as corresponding points. To compare this clusters

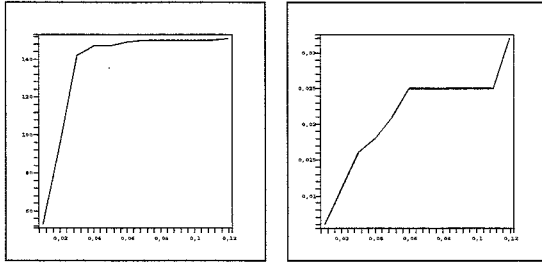


Figure 5: Quality analysis 3: comparison of the point prediction between the strict and the approximated mapping. Left figure: number of points found as equal, depending on the used threshold ϵ . Right figure: histogram of σ_{x_3} , depending on threshold ϵ .

Table 4: Quality analysis 4: Comparison of the final estimation using the strict and the approximated model

	Test
number of points ($\epsilon \leq 0.001$)	149
$x_{max}(\epsilon \leq 0.001)$	$1.0 \cdot 10^{-5}$ [cm]
$y_{max}(\epsilon \leq 0.001)$	$3.0 \cdot 10^{-5}$ [cm]
$z_{max}(\epsilon \leq 0.001)$	$2.0 \cdot 10^{-4}$ [cm]
number of points ($\epsilon \leq 0.1$)	1
number of points ($\epsilon \leq 0.15$)	2
number of points ($\epsilon > 0.15$) or not found	4

(quality analysis 4), the corresponding object points were determined by fixed parameters of the orientation of the imaging system. A cluster is defined as equal, if the difference between the reconstructed points is smaller than 0.001 cm. The results are given in Tab. 4. 149 clusters are identical, 1 object point has a difference which is smaller than 0.1 cm, 2 points smaller than 0.15 cm and 4 points have a bigger difference than 0.15 cm or were not found by using the virtual projective camera as an approximation.

For quality analysis 2 we compare the final estimated 3D points using the strict model of the multi media mapping and the approximation. The error is given in Fig.fig:histogram. The differences are normal distributed.

Fig. 7 shows the digital terrain model of the sediment surface resulting from the estimated object points.

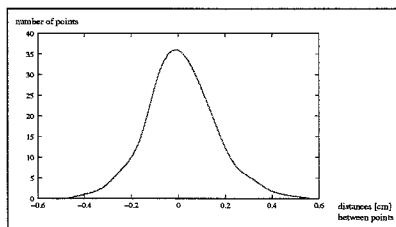


Figure 6: Quality analysis 2: comparison of the final estimated 3D points using the strict model and the approximation.

6 SUMMARY

In this paper we introduced a classification of optical mappings based on the geometry of the imaging system having a single viewpoint or a non single viewpoint. From this classification we got different kinds of image distortions: image space based and object space based. The models for optical mappings belonging to the second kind of mappings need information about

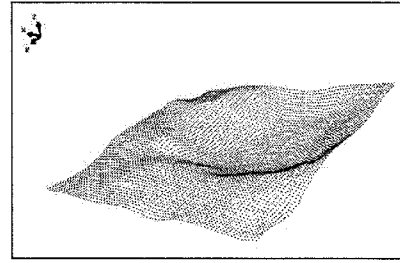


Figure 7: Reconstruction of the sediment surface resulting from the matched points using the virtual camera.

the scene structure and special complex algorithms for the projection between object and image space. Under this background we surveyed the multi media geometry. We presented a method to calculate a virtual projective camera which approximate the strict non projective mapping. The approximation was used for a point matching process using multiple views of a sediment surface with multi media geometry. We introduced a new matching process for multiple views based on geometric constraints alone, which is usable for projective mappings and the approximation of non projective mappings. Different quality tests show, that the approximation is sufficient for the reconstruction of a sediment surface.

ACKNOWLEDGMENT

This work results from a interdisciplinary project *Geometric Reconstruction, Modeling and Simulation of Fluvial Sedimental Transport* in the Special Research Centre (Sonderforschungsbereich) SFB 350 *Continental Mass Exchange and its Modeling*, at the Institute of Photogrammetry, University Bonn, Germany. The author wishes to express her gratitude to the Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland to make it possible to present this work at the ISPRS Congress 2004, Istanbul.

REFERENCES

- R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Second Edition, 2003.
- H.-G. Maas. Complexity analysis for the determination of image correspondences in dense spatial target fields. In *IAPRS*, Vol 29, Part B5, 1992.
- H.-G. Maas. New developments in Multimedia Photogrammetry. In *Proc. Optical 3D Measurement Techniques*. Wichmann-Verlag, 1992.
- H.-G. Maas. *Mehrbildtechniken in der digitalen Photogrammetrie*. ETH Zurich, Institut ff Geodäsie und Photogrammetrie, Mitteilungen Nr.62, Habilitationsschrift, 1997.
- S. K. Nayar and A. D. Karmarkar. 360 x 360 Mosaics. In *Proc. CVPR*, pages I:388-395, 2000.
- R. Swaminathan, M.D. Grossberg, S.K. Nayar. Caustics of Catadioptric Cameras. In *Proc. ICCV*, pages II:2-9, 2001
- R. Swaminathan, M. D. Grossberg and Shree K. Nayar. A Perspective on Distortions. In *Proc. CVPR*, pages 594-601, 2003
- K. Wolff and W. Förstner. Exploiting the Multi View Geometry for Automatic Surface Reconstruction using Feature Based Matching in Multi Media Photogrammetry. In *19th ISPRS Congress*. Amsterdam, 2000.
- K. Wolff and W. Förstner. Efficiency of Feature Matching for Single- and Multi-Media Geometry Using Multiple View Relations. In *Proc. Optical 3D Measurement Techniques*. Wien, 2001.

AN OPERATIONAL SYSTEM FOR AUTOMATED ROAD DATABASE UPDATING FROM AERIAL IMAGERY

C. Zhang, E. Baltsavias

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology Zurich, ETH Hoenggerberg, CH-8093 Zurich,
Switzerland – (chunsun, manos)@baug.geod.ethz.ch

Commission III, WG III/4

KEY WORDS: 3D Road reconstruction, color, context, knowledge base, edge matching, DSM/DTM; roadmarks; multiple cue integration; performance evaluation; semi-automated system

ABSTRACT:

This paper presents a practical system for automated 3D road network reconstruction from aerial images using knowledge-based image analysis. The system integrates processing of color image data and information from digital spatial databases, extracts and fuses multiple object cues, takes into account context information, employs existing knowledge, rules and models, and treats each road subclass accordingly. The key of the system is the use of knowledge as much as possible to increase success rate and reliability of the results, working in 2D images and 3D object space, and use of 2D and 3D interaction when needed. Another advantage of the developed system is that it can correctly and reliably handle problematic areas caused by shadows and occlusions. This work is part of a project to improve and update the 1:25,000 vector maps of Switzerland. The system was originally developed to process stereo images, but it has been modified to work also with orthoimages, thus making it applicable to sensors of unknown geometry. The system has been implemented as a stand-alone software package, and has been tested on a large number of images with different landscape. In this paper, various parts of the developed system are discussed, and the results of our system in the tests conducted independently by our project partner in Switzerland, and the test results with orthoimages in a test site in the Netherlands are presented together with the system performance evaluation.

1. INTRODUCTION

The extraction of roads from digital images has drawn considerable attention in the last few decades. This is increasingly stimulated by various existing and emerging applications requiring in particular up-to-date, accurate and sufficiently attributed road databases. The fact that most vendors of commercial photogrammetric, remote sensing and GI systems do not offer anything useful regarding automation of road extraction (not even practical semi-automatic methods) stresses the importance of this research topic. The existing approaches for road extraction cover a wide variety of strategies, using different resolution aerial or satellite images. A quite extensive overview of such approaches is given in Zhang (2003a). Semi-automatic schemes require human interaction to provide interactively some information to control the extraction. Roads are then extracted by profile matching (Vosselman and Knecht, 1995), cooperative algorithms (McKeown et al, 1988) and dynamic programming or LSB-Snakes (Grün and Li, 1997). Automatic methods usually extract reliable hypotheses for road segments through edge and line detection and then establish connections between road segments to form road networks (Wiedemann et al., 1998). Data from different sources is often useful (Price, 1999). Contextual information is taken into account to guide the extraction of roads (Heipke et al., 2000). Roads can be detected in multi-resolution images (Baumgartner and Hinz, 2000), while Hinz and Baumgartner (2003) use context and scale-dependent models for extraction of urban roads in large-scale images. Several applications use map information (Gerke et al., 2003). The map data is used either as approximation to start tracking or optimization process by Snakes (Bordes et al., 1997; Agouris et al., 2001), or to search for new roads (Vosselman and de Gunst, 1997). The road maps can be updated by map-image matching (Klang, 1998). The

existing approaches show individually that the use of road models and varying strategies for different types of scenes are promising. However, all the methods are based on relatively simplistic road models, and most of them make only insufficient use of a priori information, thus they are very sensitive to disturbances like cars, shadows or occlusions, and do not always provide good quality results. Furthermore, most approaches work in single 2D images, thus neglecting valuable information inherent in 3D processing.

In this paper, we present a knowledge-based system for automatic extraction of 3D roads from stereo aerial images which integrates processing of color image data and existing digital spatial databases. The system has been developed within the project ATOMI (Automated reconstruction of Topographic Objects from aerial images using vectorized Map Information), in cooperation with the Swiss Federal Office of Topography (L+T), with aims to improve road centerlines from digitized 1:25,000 topographic maps by fitting them to the real landscape, improving the planimetric accuracy to 1m and providing height information with 1-2m accuracy (Eidenbenz et al., 2000). The usual input data include 1:16,000 scale color imagery, with 30cm focal length, and 60%/20% forward/side overlap, scanned with 14 microns at a Zeiss SCAI. The other input data include: a nationwide DTM with 25m grid spacing and accuracy of 2-3/5-7m in lowland/Alps, the vectorized map data (VEC25) of 1:25,000 scale, and the raster map with its 6 different layers. The VEC25 data have a RMS error of ca. 5-7.5m and a maximum one of ca. 12.5m, including generalization effects. They are topologically correct, but due to their partly automated extraction from maps, some errors exist. In addition, DSM data in the working area with 2m grid spacing was generated from stereo images using MATCH-T of INPHO.

2. GENERAL STRATEGY

The developed system makes full use of available information about the scene and contains a set of image analysis tools. The management of different information and the selection of image analysis tools are controlled by a knowledge-based system. In this section, a brief description of our strategy is given. We refer to Zhang (2003b) for more details. The initial knowledge base is established by the information extracted from the existing spatial data and road design rules. This information is formed in object-oriented multiple object layers, i.e. roads are divided into various subclasses according to road type, land cover and terrain relief. It provides a global description of road network topology, and the local geometry for a road subclass. Therefore, we avoid developing a general road model; instead a specific model can be assigned to each road. This model provides the initial 2D location of a road in the scene, as well as road attributes, such as road class, presence of roadmarks, and possible geometry. A road is processed with an appropriate method corresponding to its model, certain features and cues are extracted from images, and roads are derived by a proper combination of cues. The knowledge base is then automatically updated and refined using information gained from previous extraction of roads. The processing proceeds from the easiest subclasses to the most difficult ones. Since neither 2D nor 3D procedures alone are sufficient to solve the problem of road extraction, we make the transition from 2D image space to 3D object space as early as possible, and extract the road network with the mutual interaction between features of these spaces.

3. CUE EXTRACTION

When a road from VEC25 is selected, the system focuses on the image regions around the road, defined using the position of the road and the maximal error of VEC25. Then, according to the road attributes a set of image processing tools is activated to extract features and cues. 3D straight edge generation is a crucial component of our procedure because the road sides are among them. The 3D information of straight edges is determined from the correspondences of edge segments between stereo images. An image classification method is implemented to find road regions. With the DSM and DTM data, the above-ground objects and ground objects are separated. We also exploit additional cues such as roadmarks to support road extraction.

3.1 3D straight edge extraction

The edges are extracted by the Canny operator in stereo images. For each edge, the edge attributes are computed, including the geometrical description of the edge and the photometrical information in the flanking regions of the edge. The epipolar constraint is applied to reduce the search space. We then compute a similarity measure for an edge pair by comparing the edge attributes. The locally consistent matching is then determined through structural matching with probability relaxation using the similarity measures as prior information. We refer to Zhang and Baltsavias (2000) for the detailed matching strategy and qualitative performance evaluation. The matched edges are then transformed to object space by finding the corresponding pixels within each matched edge pair (Zhang and Baltsavias, 2002). Finally, 3D straight edge segments are fitted to the 3D edge pixels.

3.2 Image classification for road region detection

We implemented the ISODATA algorithm to classify the color images and separate road regions from other objects. The success of image classification also depends on the input data. The original RGB color image is transformed into different color spaces and is also used to compute several artificial bands/indices to enhance features such as vegetation and shadows so that they are more isolated in feature space. The following 3 bands are selected for image classification: (1) the first component of principal component transformed image, (2) a band calculated with R and G bands in RGB space as $(G-R)/(G+R)$, (3) S band from HSI color space. We then determine 5 classes corresponding to road regions, green objects, shadow areas, dark roofs and red roofs.

3.3 DSM and DTM analysis

The DTM or DSM has been used in our system to reduce search space for straight edge matching. They are also used to verify if a 3D straight edge or a region is on the ground. Because a DSM ideally models the man-made objects as well as the terrain, subtracting the DTM from DSM results in the so-called normalized DSM (nDSM) which enables the separation of above-ground objects (buildings and trees) and ground objects (roads, etc.). Since in ATOMI, the DTM data is not very precise, we extract above-ground objects directly from the DSM using a multiple height bin method presented in Baltsavias et al. (1995). By combining the information of nDSM with image classification data, our system creates redundancy to confirm the existence of roads. Furthermore, it can partly compensate the missing and wrong information in the classification.

3.4 Roadmark and zebra crossing extraction

Roadmarks and zebra crossings are good indications of the existence of main roads and roads in urban areas. Both of them have distinct color (usually white or yellow). Usually roadmarks give the road direction and often the road centerline, while the zebra crossings define the local road width. Thus, they can be used to guide the road extraction process or verify the extraction results. In addition, in many cases the correct road centerlines can be even derived directly from present roadmarks and/or zebra crossings. This is especially useful when the road sides are occluded or not well-defined, such as in cities or city centers.

The roadmarks are detected as white straight lines using an image line model in which the shape of an image line is presented as a second order polynomial (Zhang, 2003a). The extracted straight lines are then transformed into object space by our developed structural matching method. Only those that are on the ground (as defined by nDSM), belonging to the road region (as determined by the classification) and in the buffer defined by VEC25 are kept as detected roadmarks. Zebra crossings are composed of several thin stripes. Using color information, the image is first segmented. Morphological closing is applied to bridge the gaps between zebra stripes. We then obtain several clusters by connected labeling. Only the clusters with a certain size are kept, while the small ones are discarded. Then, the shape of the cluster is analyzed. The rectangle-like clusters are selected as zebra crossings. The center, the short and long axes of the detected zebra crossings are computed using spatial moments.

4. 3D ROAD RECONSTRUCTION

With the extracted features and cues, our next step is to combine them to extract the road. Firstly, irrelevant edges are removed. 3D parallel overlapping edges are then searched for, and are evaluated to find possible road sides. Due to occlusions or shadows, a road and its sides may be totally or partially invisible in images. The road segments with only one side visible in one or two images in the occluded/shadowed areas are inferred and reconstructed. In addition, gaps, where the road sides are totally invisible, are bridged. Finally, the road is reconstructed by finding an optimal path among the road segment candidates that maximizes a merit function. Highways, first class roads and most second class roads are also extracted using the detected roadmarks and the zebra crossings. With the extracted roads, the road junctions are reconstructed and modelled. The main procedures are described below. More details on all above procedures are given in Zhang (2003a).

4.1 Finding 3D parallel road sides

The system checks the extracted edges to find 3D parallel edges. Only edges located in the buffer defined by the VEC25, having a similar orientation to the VEC25 segments and a certain slope are further processed. Edges above the ground are removed by checking with the nDSM. Two edges are considered as parallel if they have similar orientation in 3D space. The edges of a pair must overlap in the direction along the edges, and the distance between them must be within a certain range determined by the road class defined in the VEC25. In addition, the heights of the two edge segments should be similar. The found 3D parallel edges are projected onto the images and evaluated using multiple knowledge. The region between the projected edges must belong to the class road as determined by the image classification. If roadmarks are presented on this road, the extracted roadmarks are used to confirm that the edge pair corresponds to correct road sides. The found 3D parallel edges have high probability of belonging to a road; they are called Possible Road Sides that are Parallel (PRSP). Thus, each PRSP is geometrically described by a pair of 3D straight edges and its corresponding 2D edges, and holds a set of attributes.

4.2 Reconstruction of missing road sides

Not all 3D road segments can be obtained from the procedures described above. The absence of 3D road sides can be caused by shadows, occlusions, or road sides do not actually exist, e.g. in the area where a parking lot is situated next to the road. Depending on the relations between the road segments and the neighbouring objects, sun angle, viewing direction, existence of moving cars on the road etc., there are various types of missing road sides in images. We have made an investigation and classified them into 11 types. Each type is then treated, and the missing road sides are inferred and validated by a specific procedure (Zhang, 2003a). Based on the type of the missing road sides, corresponding types of road segment candidates (RSCs) are obtained. They are then evaluated using the knowledge obtained from the cues. Common reliability measures for all RSCs are evaluated using the results from the image classification and the nDSM. In addition, specific reliability measures for each type of RSC are also computed using the geometric relations between the RSC and its adjacent PRSPs. We refer to Zhang (2003a) for the detailed procedures of RSC evaluation.

4.3 Gap bridging

A gap between neighbouring PRSPs that belongs to a road represents a road part where no road side is visible in the images. In our system, it is bridged either by directly linking the neighbouring PRSPs or by adapting the shape of the VEC25 road corresponding to the gap area. That is, the vertices of the VEC25 road in the gap area are shifted to close the gap, based on the coordinate differences between the end points of the PRSPs and their corresponding points on the VEC25 road. Linking the adjacent PRSPs to close gaps is efficient on straight roads with short gap length, while using the shape of VEC25 road might be more useful for long and curved occlusion areas. We determine the solution for gap bridging in an evaluation process using various knowledge, e.g. the shape of the solution for the gap should approximately comply with that of the VEC25 road; it should be either a road region or a shadow or shadow mixed with road region; or roadmarks are extracted within the hypothesised road area. Based on the evaluation, we compute a measure for the gap hypothesis, s_{gap} . The range of values for s_{gap} is [0,1], with decreasing value for long and inconsistent gaps.

4.4 Road segment linking for 3D road reconstruction

With the extracted PRSPs, the road is reconstructed by linking the PRSPs belonging to a road. The goal of linking road segments is twofold. First, this implies that PRSPs belonging to a road should be selected and connected with the gaps bridged by the linking algorithm. Secondly, this also implies that PRSPs not belonging to a road should be rejected. Therefore, the algorithm must be very selective in which PRSP it adds to a road. The linking function in our system is defined as

$$\sum l_i \cdot s_i + l_{gap} \cdot s_{gap} + l_j \cdot s_j \quad (1)$$

where, i and j are adjacent PRSPs, l_i and l_j are their lengths, s_i and s_j are their reliability measures. l_{gap} is the gap length between i and j , and s_{gap} is the gap evaluation measure. The function takes high values for long curves with a shape similar to the VEC25 road. The linking problem can then be solved by finding a subset among all PRSPs that maximizes the linking function. This can be achieved using dynamic programming (Grün and Li, 1997).

Higher class roads are also extracted using the detected roadmarks and zebra crossings. The roadmarks are linked using a similar method as described in the previous paragraph. This procedure increases the effectiveness and reliability of our system. In complex areas, such as in city centers, the road sides are generally occluded very much, while sometimes they are not defined. However, the road centerlines are successfully extracted by the system using roadmarks. In rural and suburban areas, the extracted road using roadmarks is used by the system to verify the extraction results using 3D parallel edges.

4.5 Road junction generation and modelling

Road junctions are important features of the road network. However, it is even more difficult to model and extract road junctions from images than road segments. This might be one of the reasons that this issue has been rarely touched in past research. In our system, we reconstruct junctions through intersecting the extracted roads guided by the topology of the VEC25 data, and further model the junctions with road class information and the shape of the VEC25. During the process, each road is assigned a weight corresponding to the evaluation

value of the road segment in the junction area. In addition, an angular constraint is applied forcing the angle of the two reconstructed roads at the reconstructed junction point similar to that defined by the corresponding VEC25 roads at the VEC25 junction point. Furthermore, in the angular constraint higher priority is given to higher class roads or continuous roads. In such formulation, the junction point is correctly located, while the shape of the roads at the junction point is well formed: complying with the reality and following the road design rules.

With the extracted roads and road junctions, the road network is obtained. The results inherit other attributes from the VEC25 data with the road lengths updated, and road widths appended. The number of lanes can be inferred from the known road width and also possibly roadmarks. The 3D information permits the derivation of other useful attributes like horizontal and vertical curvatures and road slope.

4.6 Performance evaluation

We introduce two types of measure for self-diagnosis of the extraction results: an overall quality measure for the whole road, and measures for the road segments. If a result does not pass the overall quality test, a further test is conducted to find in which segments the errors occur. The overall quality of the extraction result can be obtained from the following criteria:

- the lengths of the extraction result and the VEC25 road should be similar
- the shape difference between the extraction and the VEC25 road should be small
- total length of PRSPs should cover a large part of the extraction result

We also define an internal quality measure for each road segment using the shape similarity measure between the segment and the corresponding VEC25 road, and the information in the classification result and the nDSM in the segment area (Zhang, 2003b). For higher class roads, the assessment of the reliability of the extraction results is also conducted through comparison of the results by using parallel edges and roadmarks.

External evaluation is done by comparing the extracted results with precise reference data. The quality measures aim at assessing completeness and correctness as well as geometrical accuracy (Heipke et al., 1998; Zhang, 2003b). Additional measures have been defined, e.g. to evaluate the shape quality, but have not been used in the test results presented here.

5. RESULTS

The described system has been implemented as a stand-alone software package with a graphic user interface running both on UNIX SGI and PC Windows XP. The system imports color stereo or ortho-imagery, the existing road database and other input data, and outputs the extracted roads in 3D Shapefile format that is readily imported by existing GIS software. Other data formats can be easily accommodated. The system has been tested using more than 20 models in various landscapes. Some reports of the system performance can be found in Zhang and Baltasvias (2002) and Zhang (2003b). A benchmark test has been conducted independently by our project partner using new flight imagery, in the test site Thun, Switzerland. The terrain height ranges from 550 m to 2200 m. Almost all road types in Switzerland can be found in this area. The images were acquired in October 2001, and the image data have the same

specifications as described in Section 1. During the test, our system is only applied to extract roads in rural areas, while roads in urban and forest areas are not processed. Fig. 1 presents a portion of 3D road extraction and road network generation (only the left image is shown). The landscape of Fig. 1 includes open rural, forest areas and small settlements.

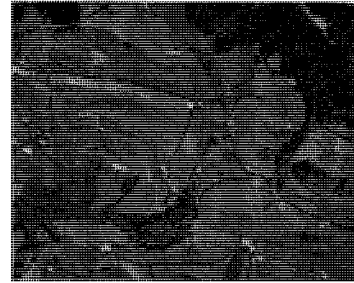


Figure 1. Extracted 3D roads and road network in test site Thun, Switzerland.

The details of automatic 3D road extraction and junction generation in rural areas with varying complexity are presented in Fig. 2, where the VEC25 roads are shown in white lines and the extracted roads in black lines. Note that the road junctions are also well extracted and modeled.

Recently, the system has been under extensive test by our project partner. Around 10,000 km of roads have been processed. Almost all roads in rural areas were correctly extracted using visual check.

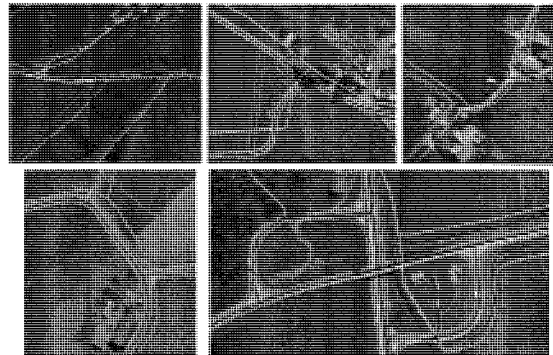


Fig. 2. Details of road extraction and junction generation with varying complexity in test sites in Switzerland. The extracted roads are shown in black lines and the VEC25 roads in white lines.

Table 1 summarises the external evaluation of the extraction results in Fig. 1 using the reference data measured by L+T at an analytical plotter. It can be seen that our system achieves very good results. The completeness and correctness are very high. The accuracy of the extracted road network is about 0.5 m both in planimetry and height, fulfilling the accuracy requirements of the project ATOMI.

Quality Measures		
Completeness		94.2%
Correctness		96.9%
Length of reference (km)		12.4
Length of extraction (km)		11.69
RMS (m)	dx	0.40
	dy	0.33
	dz	0.58

Table 1. Quality measures for the test dataset in Fig. 1.

A higher completeness has been achieved by our partner when spring photography is used. This is because of less tree occlusions. Even some roads along forest borders can be extracted in spring images. In addition, some roads in fields, invisible in summer images, are visible in spring images and are extracted, thus further contributing to a higher completeness of the extracted road network.

We also tested our system on unknown data provided by the National Geographic Institute, Belgium. Although the images are black and white with quite poor radiometric quality, and no DSM is available, the performance of our system is also quite good in flat open rural areas. By comparing the results with the manually measured reference data for ca. 13 km roads, the achieved completeness and correctness are 97.6% and 98.1% respectively (Zhang and Baltasvias, 2002).

The system has been modified to work even with orthoimages, whereby the 3D information is not extracted by image matching, but by overlaying the 2D information on the DSM and DTM. Extensive tests conducted by our Swiss project partner using various resolution orthoimages (0.20 m ~ 0.60 m) have shown that the results are quite similar to that from the stereo imagery, and that an increase of the pixel size leads to a proportionally much smaller deterioration of the geometric accuracy of the extracted roads. With this development, the Dutch Ministry of Transport, Public Works and Water Management (MTPWWM) awarded to us, after an evaluation of various research systems, a project for a feasibility study of semi-automated updating of the Dutch road database, using color orthoimages of 0.5 m pixel size from aerial images of 1:25,000 scale with 15 cm focal length. The study site is situated near the city of Weert in the province Limburg (in the south of the Netherlands), covering an area of 12 * 12 sq. km. The landuse changes gradually from open rural to urban, with the complexity of the scenes increasing correspondingly. The images were taken in June, 2000. The images do not have good quality; they are too green and noisy. In many cases the roads show very poor contrast with surroundings. The image edges are poorly defined; also color shifts between bands are observed. In addition, trees at road sides usually occlude roads very much in these summer images; some roads are even totally occluded. We also observed that the roadmarks on roads are very weakly represented in such images. The old road databases are created by digitizing 1:10,000 topographic maps, with an RMS error of about 10 m. The database allows distinguishing national roads and a small part of the provincial roads in the Netherlands, and provides the number of lanes for them. The other roads are in a single class. There are no clues that can be used to infer the approximate road width. Available height data are from laser scanning (raw and filtered heights). Both datasets have points regularly distributed with a 5m x 5m spacing. During the test, our system is only applied in open rural areas. Fig. 3 shows a portion of the test results. The roads in rural areas are correctly and reliably extracted by our system. In Fig. 4, the details of road extraction and junction generation for this dataset are presented in several examples.

Reference data for the Dutch dataset is not available at moment. The accuracy of the extraction result cannot be accessed. In each test image we computed the ratio of the length of the extracted roads to the length of rural roads in the existing database (the total length of the rural roads in the old database is ca. 500 km). The ratio values range from 80% to 92%, depending on the complexity of the scene. Generally, the performance is worse compared to the performance on the

Swiss dataset. This is mainly caused by: (1) the poor image quality, (2) insufficiently road information in the existing road database, especially the lack of the road classes or road width, (3) the images are taken in summer, many roads are occluded by trees, (4) the worse spatial resolution of 0.5m compared to 0.22m of the Swiss data.

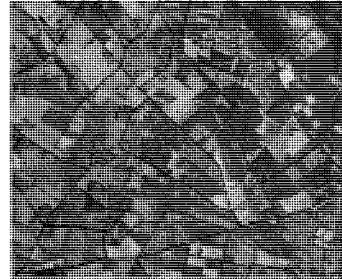


Fig. 3. Extracted 3D roads and road network in the test site in the Netherlands superimposed on image as black lines.

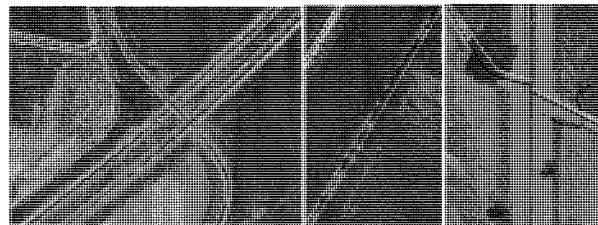


Fig. 4. Details of road extraction and junction generation in the Netherlands dataset. The extracted roads are shown in black lines and the outdated roads in white lines.

6. DISCUSSION AND CONCLUSION

In this paper, we have presented a practical automated system for road extraction from stereo and ortho-images focusing on rural areas. The roads should have a minimum width of about 3 pixels in order that edges on both road sides are extracted. The system has several advantages over other approaches. It uses existing knowledge, image context, rules and models to restrict the search space, treats each road subclass differently, checks the plausibility of multiple possible hypotheses, therefore provides reliable results. The system contains a set of data processing tools to extract various cues about road existence, and fuses multiple cues and existing information sources. This fusion provides not only complementary information, but also redundant one to account for errors and incomplete partial results. Working on stereo images, the system makes an early transition from 2D image space to 3D object space. Road hypotheses are generated directly in 3D object space. This not only enables us to apply more geometric criteria to generate hypotheses, but also largely reduces the search space, and speeds up the process. The hypotheses are evaluated in images using accumulated knowledge information. Whenever 3D features are incomplete or entirely missing, 2D information from stereo images is used to infer the missing features. By incorporating multiple knowledge, the problematic areas caused by shadows, occlusions etc. can be often handled. Based on the extracted roads, the road junctions are generated and modeled, thus the system provides an up-to-date and complete road network for practical uses. We also present in this paper the results of road extraction in benchmark tests conducted independently by our project partner. The quantitative analysis using accurate reference data is also presented. The comparison of the reconstructed roads with such data shows that more than 94% of the roads in rural areas are correctly and reliably

extracted by the developed system, and the achieved accuracy of the road centerlines is about 0.5 m both in planimetry and height, fulfilling the requirements of the project ATOMI. The system can also work with black and white images and no use of DSM without significant performance reduction (as the tests in Belgium have shown). We also showed that a higher completeness can be achieved if spring photography is used. An even better performance can be expected with the increasingly possible use of better quality DTMs and DSMs (e.g. from airborne laser scanning) and a near-infrared channel, which is available in many new digital photogrammetric cameras and high-resolution satellites.

The system can also process orthoimages. The extensive tests at the L+T in rural areas have shown that the performance is quite similar. We have presented the results in the study site of the Netherlands, with the dataset provided by MTPWWM, the Netherlands. Using a visual check, most of the roads in rural areas are extracted. The extraction results are quite good and reliable. There are only very few errors in the extraction results. Better results can be expected by using good quality images, which can be achieved by acquiring better film imagery and by better scanning. Use of new digital cameras, e.g. ADS40 developed by Leica Geosystems, providing near infrared information would be an extra plus. First tests with ADS40 imagery at the L+T have shown that the road extraction results are similar, however with larger pixel size and no use of the infrared channel. Use of digital photogrammetric cameras may prove to be advantageous for road extraction, on top of other advantages like faster data acquisition, avoidance of errors during scanning etc. We are currently also performing tests with Ikonos and Quickbird images and the results will be reported elsewhere.

Regarding 3D reconstruction of new roads not included in the database, this can be achieved by manual on-screen digitisation of characteristic road seed points and continuation with the same procedure described in this paper with minor modifications. This approach can be also followed when no road database exists, for 3D road reconstruction from scratch.

ACKNOWLEDGEMENTS

We acknowledge the financial support for this work and for the project ATOMI by the Swiss Federal Office of Topography, Bern. We also thank the Ministry of Transport, Public Works and Water Management, the Netherlands for letting us use their data for this publication.

REFERENCES

- Agouris, P., Stefanidis, A., Gyftakis, S., 2001. Differential snakes for change detection in road segments. *Photogrammetric Engineering and Remote Sensing* 67(12), 1391-1399.
- Baltsavias, E.P., Mason, S., Stallmann, D., 1995. Use of DTMs/DSMs and orthoimages to support building extraction. In: Grün, A., Kuebler, O., Agouris, P. (Eds.), *Automatic Extraction of Man-made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pp. 189-198.
- Baumgartner, A., Hinz, H., 2000. Multi-Scale road extraction using local and global grouping criteria. *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, Part B3/1, pp. 58-65.
- Bordes, G., Giraudon, G., Jamet, O., 1997. Road modeling based on a cartographic database for aerial image interpretation. In: Förstner, W., Pluemer, L. (Eds.), *Semantic Modelling for the Acquisition of Topographic Information from Images and Maps*, Birkhäuser Verlag, Basel, pp. 23-139.
- Eidenbenz, Ch., Kaeser, Ch., Baltsavias, E.P., 2000. ATOMI – Automated Reconstruction of Topographic Objects from Aerial Images using Vectorized Map Information. *International Archives of Photogrammetry, Remote Sensing and SIS*, Vol. 33, Part B3/1, pp. 462-471.
- Gerke, M., Butenuth, M., Heipke, C., Willrich, F., 2003. Graph supported verification of road databases using aerial imagery. In: Shi, W., Goodchild, M.F., Fisher, P.F. (Eds.), *Proc. 2nd International Symposium on Spatial Data Quality*, pp. 421-430.
- Grün, A., Li, H., 1997. Semi-automatic linear feature extraction by dynamic programming and LSB-Snakes. *Photogrammetric Engineering and Remote Sensing* 63(8), 985-995.
- Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1998. External evaluation of automatically extracted road axes. *Photogrammetrie Fernerkundung Geoinformation* (2), 81-94.
- Heipke, C., Pakzad, K., Straub, B.M., 2000. Image analysis for GIS data acquisition. *Photogrammetric Record* 16 (9), 963-985.
- Hinz, S., Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 58(1-2), 83-98.
- Klang, D., 1998. Automatic detection of changes in road database using satellite imagery. *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 4, pp. 293-298.
- McKeown, D.M., Denlinger, J. L., 1988. Cooperative methods for road tracking in aerial imagery. *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 662-672.
- Price, K.E., 1999. Road grid extraction and verification. *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3-2W5, pp. 101-106.
- Vosselman, G., Knecht, J.D., 1995. Road tracing by profile matching and Kalman filtering. In: Grün A., Kuebler, O., Agouris, P. (Eds.), *Automatic Extraction of Man-made Objects from Aerial and Space Images*, Birkhauser Verlag, Basel, pp. 255-264.
- Vosselman, G., Gunst, de M., 1997. Updating road maps by contextual reasoning. In: Grün, A., Baltsavias, E.P., Henricsson, O. (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Birkhäuser Verlag, Basel, pp. 267-276.
- Wiedemann, C., Heipke, C., Mayer, H., Hinz, S., 1998. Automatic extraction and evaluation of road network from MOMS-2P imagery. *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3/1, pp. 285-291.
- Zhang, C., 2003a. Updating of cartographic road database by image analysis. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Report No. 79.
- Zhang, C. 2003b. Towards an operational system for automated updating of road databases by integration of imagery and geodata. *ISPRS Journal of Photogrammetry and Remote Sensing* 58(3-4), 166-186.
- Zhang, C., Baltsavias, E.P., 2000. Knowledge-based image analysis for 3D edge extraction and road reconstruction. *International Archives of Photogrammetry, Remote Sensing and SIS*, Vol. 33, Part B3/2, pp. 1008-1015.
- Zhang, C., Baltsavias, E.P., 2002. Improving cartographic road databases by image analysis. *International Archives of Photogrammetry, Remote Sensing and SIS*, Vol. 34, Part 3A, pp. 400-405.

AUTOMATIC DSM GENERATION FROM LINEAR ARRAY IMAGERY DATA

Zhang Li, Armin Gruen

Institute of Geodesy and Photogrammetry, Swiss Federal Institute of Technology Zurich
ETH-Hoenggerberg; CH-8093 Zurich, Switzerland
Tel.: +41-1-633 31 57, Fax: +41-1-633 11 01
E-mail: <zhangl><agruen>@geod.baug.ethz.ch

Commission III, WG III/2

KEY WORDS: Linear Array Imagery, Image Matching, DSM

ABSTRACT:

CCD linear array sensors are widely used to acquire panchromatic and multispectral imagery for photogrammetric and remote sensing applications. The processing of this kind of images provides a challenge for algorithmic redesign and this opens the possibility to reconsider and improve many photogrammetric processing components. In addition, the basic capabilities of image matching techniques have so far not been fully utilized yet. This paper presents a matching procedure for automatic DSM generation from linear array imagery data. It can provide dense, precise and reliable results. The method uses a coarse-to-fine hierarchical solution with an effective combination of several image matching algorithms and automatic quality control. The DSMs are generated by combination of matching results of feature points, grid points and edges. Finally, a modified multi-photo geometrically constrained (MPGC) matching algorithm is employed to achieve sub-pixel accuracy for all the matched features with multi-image or multi-strip image data.

The proposed approach in this paper has been applied to different areas with varying textures and terrain types. The accuracy tests are based on the comparison between the high quality DEMs / DSMs derived from airborne Laser Scanner or manual measurements and the automatic extracted DSMs. Results with STARIMAGER, IKONOS and SPOT5 HRS images are reported. We demonstrate with these experiments that our approach leads to good results.

1. INTRODUCTION

In recent years, CCD linear array sensors are widely used to acquire panchromatic and multispectral imagery in pushbroom mode for photogrammetric and remote sensing applications. Linear scanners are carried on aircraft (e.g. ADS40), helicopter (e.g. STARIMAGER) or spacecraft (e.g. IKONOS) and allow for photogrammetric mapping at different scales.

Spaceborne optical sensors like SPOT, IKONOS, and QuickBird provide not only for high-resolution (0.6 – 5.0 m) and multi-spectral data, but also for the capability of stereo mapping. The related sensors are all using linear array CCD technology for image acquisition and are equipped with high quality orbit position and attitude determination devices like GPS and IMU systems.

Progress in the development of airborne linear array imaging system has also been made in the last decade. These systems use the three-line-scanner concept and provide for high resolution (0.5 – 0.03 m) panchromatic and multispectral image data with triplet overlap and along-track base direction. In the year 2000, Starlabo Corporation, Tokyo designed a new airborne digital imaging system, the Three-Line-Scanner (TLS) system (now called STARIMAGER (SI)), jointly with the Institute of Industrial Science, University of Tokyo (Murai, Matsumoto, 2000). The first generation camera STARIMAGER-100 (SI-100) contains three parallel one-dimensional CCD focal plane arrays, with 10200 pixels of 7 μ m each. Starlabo is currently developing a new generation camera system SI-200. This comes with an improved lens system and with 10 CCD arrays on the focal plane (3 \times 3 work in RGB mode, 1 CCD array works in infrared mode). Each CCD array consists of 14 404 pixels at 5 μ m size. The system produces seamless high-resolution images (3 - 10 cm footprint on the ground) with three viewing directions (forward, nadir and backward). For the SI sensor and imaging parameters see Gruen, Zhang, 2002.

The processing of this kind of images provides a challenge for algorithmic redesign and this opens the possibility to reconsider and improve many photogrammetric processing components, like image enhancement, multi-channel color processing, triangulation, orthophoto and DEM generation and object extraction. We have recently developed a full suite of new algorithms and software system for the precision processing of this kind of data.

In this paper, we put particular emphasis on the automatic generation of DSMs. Originally we developed a matching approach and the related software "SI-Matcher" for multi-image

processing of the very high-resolution SI images (Gruen, Zhang, 2003). Now this matching procedure has been extended and has the ability to process other linear array images as well. We will briefly report about the basic considerations for our procedure. Then we will address the key algorithms. We will give experimental results from the processing of SI, IKONOS and SPOT5 HRS images.

2. MATCHING CONSIDERATIONS

Automatic DEM/DSM generation through image matching has gained much attention in the past years. A wide variety of approaches have been developed, and automatic DEM generation packages are in the meanwhile commercially available on several digital photogrammetric workstations. Although the algorithms and the matching strategies used may differ from each other, the accuracy performance and the problems encountered are very similar in the major systems and the performance of commercial image matchers does by far not live up to the standards set by manual measurements (Gruen et al., 2000). The main problems in DEM/DSM generation are encountered with

- (a) Little or no texture
- (b) Distinct object discontinuities
- (c) Local object patch is no planar face
- (d) Repetitive objects
- (e) Occlusions
- (f) Moving objects, incl. shadows
- (g) Multi-layered and transparent objects
- (h) Radiometric artifacts like specular reflections and others
- (i) Reduction from DSM to DEM

The degree to which these problems will influence the matching results is imagescale-dependent. A DSM derived from 5 m pixelsize SPOT5 HRS images or 1 m pixelsize IKONOS images will be relatively better than one derived from 5 cm pixelsize SI images. To extract DSMs from very high-resolution aerial images, we should take into account the occlusions, the surface discontinuities such as man-made objects and trees, large areas with little or even no texture, repetitive patterns, etc.

On the other hand, linear array imagery provides for new characteristics and possibilities for image matching:

- It has the ability to provide 16 bit (effectively 12 bit) images, which should reduce the number of mismatches even in dark shadow areas.

- It has the ability to provide multiple images with multiple channels. Thus it enables the multi-image matching approach, which leads to a reduction of problems caused by occlusions, multiple solutions, surface discontinuities and results in higher measurement accuracy through the intersection of more than two image rays.
- It has the ability to provide for relatively precise orientation elements that can be used to enforce geometric constraints and restrict the search space along quasi-epipolar lines.
- The nearly parallel projection in along-track direction causes less occlusion on the nadir-viewing images.

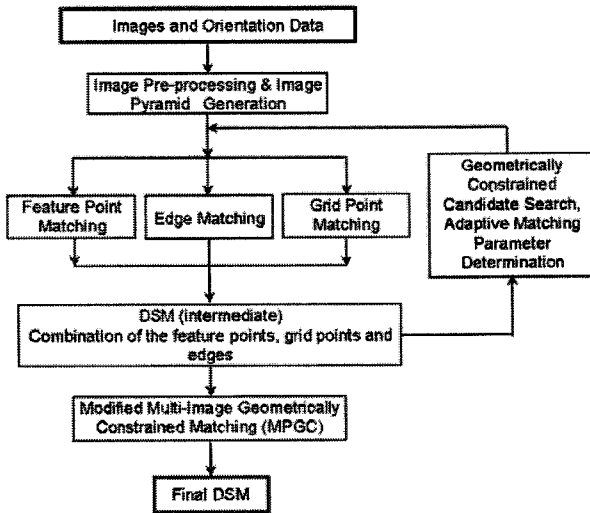


Figure 1: Workflow of our image matching procedure

Among the known matching techniques and algorithms, area-based (ABM) and feature-based matching (FBM) are the two main ones applied to automatic DSM generation in general. ABM and FBM have both advantages and disadvantages with respect to the problems presented above. The key to successful matching is an appropriate matching strategy, making use of all available and explicit knowledge concerning sensor model, network structure and image content. Our matching approach is a hybrid method that combines ABM and FBM. It aims to generate DSMs by attacking the problems (a)-(f) mentioned above. It uses a coarse-to-fine hierarchical solution with a combination of several image matching algorithms and automatic quality control. Figure 1 shows the workflow of our matching procedure. After the image pre-processing and production of the image pyramid, the matches of three kinds of features, i.e. feature points, grid points and edges on the original images are finally found progressively starting from the low-density features on the images with low resolution. A triangular irregular network (TIN) based DSM is constructed from the matched features on each level of the pyramid, which in turn is used in the subsequent pyramid level for the approximations and adaptive computation of the matching parameters. Finally the modified MPGC matching is used to achieve more precise matches for all the matched features and identify some false matches. In the MPGC procedure, multiple strip image data can be introduced and combined. More details of our approach will be provided in the next paragraph.

3. THE MATCHING APPROACH

3.1 Image Preprocessing

In order to reduce the effects of radiometric problems like strong bright and dark regions and to optimise the images for subsequent feature extraction and image matching, a pre-processing method, which combines an edge-preserving smoothing filter and the Wallis filter. First, the edge preserving smoothing filter proposed by Saint-Marc et al., 1991 was applied to reduce the noise, while sharpening edges and preserving even fine detail such as corners and line endpoints. Next, the Wallis filter, which strongly enhances already existing texture patterns, is applied. The examples of Figure 2 indicates that even in

shadow and “homogeneous” areas the image content is enhanced.

The image pyramid starts from the original resolution images. Each pyramid level is generated by multiplying a generation kernel and reduces the resolution by factor 3. The pyramid level number is a pre-defined value that is either a user-input or can be determined according to the height range of the imaging area.

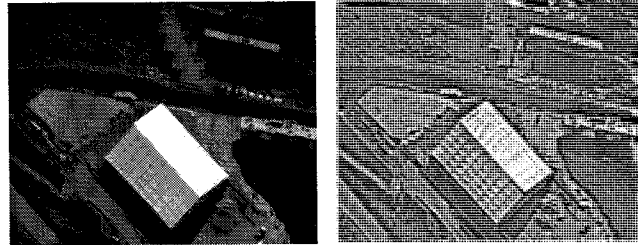


Figure 2: SI image before and after pre-processing

3.2 Feature Point Matching

We use the Foerster interest operator to extract well-defined feature points that are suitable for image matching. Firstly the reference image is divided into small image patches (the nadir viewing SI or satellite image is selected as reference). Only one feature point will be extracted in each image patch. The density of the feature points can be controlled by the size of the patches. The determination of the correspondences of the given points on the search images is carried out using the geometrically constrained cross-correlation method (see Gruen, Zhang, 2003). The matching candidates are computed by cross-correlation with a set of adaptively determined parameters like the image window size w_c , the threshold of the correlation coefficient c_t and the search distance. The approximate DSM that is derived from the higher-level of the image pyramid is used to estimate these parameters.

We incorporate the method proposed by Kanade & Okutomi, 1994 to select an optimal window size by evaluating the image content and the disparity within the matching window. As a result, in flat areas with small image intensity variations, the window size w_c increases and in areas of large terrain elevation variations it decreases. The threshold of the correlation coefficient c_t should also vary according to the roughness of the terrain. We set a larger value in flat areas and smaller value in rough terrain areas. The search window size depends on the terrain elevation variation in a small neighborhood of the given point and on the image geometry. In relatively flat areas the size of the search window decreases and vice versa.

By adaptive selection of these parameters, we can both reduce the processing time and the probability for multiple matching candidates. The number of matching candidates can be further reduced by introducing a third image. For every candidate, its position on the third image can be predicted using the image orientation elements. If the correlation coefficient between the reference and the third image is lower than the threshold, this matching candidate will be dropped. However, we cannot completely avoid the ambiguity problem due to reasons like repetitive patterns. Our procedure takes $n (\leq 5)$ matching candidates with correlation coefficient values above the threshold c_t .

As a result, for each feature point on the reference image several matching candidates can be computed. The correct match is determined by analysing the following quality measures sequentially:

- The correct match should have a clear and sharp maximal correlation coefficient. If there are more than one candidates and the value of the first correlation coefficient peak is more than two times of that of the secondary peak, the candidate that has the largest correlation coefficient value will be considered the correct match.
- Using the same matching parameters, the feature point can be back-matched from the search images to the reference image. If the difference between this two-way matching is less than one pixel, the candidate is assumed to be the correct match.
- Under the condition that the feature point appears on more than two images, the residuals of the photogrammetric forward intersection should be less than 2-3 times the standard deviations of image coordinates of the triangulation adjustment.

The points that pass these tests will be indicated as reliable matches. The points that cannot pass these tests may have multiple solutions. This matching ambiguity will be solved by the following global image matching method through local smoothness constraints (see chapter 3.4).

3.3 Edge Matching

To reconstruct a DSM from very high-resolution images over urban areas we must take into account the problems caused by surface discontinuities, occlusions and the significant perspective projection distortion. Even with satellite images, line features are also important for capturing and modeling terrain features such as ridgelines and breaklines. Matching the edges is a possible solution to these problems. However we should consider the following problems:

- The edge on one image may break into more than one segment due to image noise, occlusions and the deficiencies of the feature extraction algorithms.
- The conjugate edges on different images may have quite different shapes due to the projection distortion.
- There may be many similar features in a search area.

The edge matching procedure presented here is based on the evaluation of the local geometric and photometric attributes of edges for the solution of disambiguities. The quasi-epipolar geometry and the DSM data derived from the higher-level of the image pyramid are used to provide for the matching candidates for each edgel. A figural continuity constraint satisfaction scheme (the disparity along an edge should change smoothly) and a shape matching approach are used to achieve the final results.

The well-known Canny operator is used to locate the intensity discontinuities. Then edgels are linked into free-form edges through a local processing that analyses the characteristics of these pixels in a small neighborhood. This approach is carried out independently on three images. Only edges above a minimum length (30 pixels for SI images and 15 pixels for satellite images) are considered for matching.

The edgels along the given edge are matched with the edges that are defined at the intersection points between the candidate edges and the correspondent epipolar curve within the search window on one of the search images. The search window can be determined by using the same method as in chapter 3.2. There may be several matching candidates within the search window. To solve this ambiguity problem we perform the following three steps sequentially:

- a) Evaluation of the difference of the local edge orientation between the given edgel and its candidates. The local orientation for an edgel is the image intensity gradient computed modulo 2π . Candidates with differences above a threshold will be dropped. Considering the possible relief distortion, this constraint should not be too tight. For example, we use 40 degrees for SI images.
- b) Evaluation of the normalized cross-correlation coefficient of the intensity values on each side of the edgel. Exclusion of those candidates that have a very low correlation coefficient, e.g. less than 0.5.
- c) If more than two images are available, each candidate can be validated on the third image through the indicator used in step b).

After these three steps, the given edgel may still have more than one candidate. The problem will be further solved by using the figural continuity constraint through a relaxation method. This method examines the candidates by computing how much support they receive from their local neighborhood along the edge. We select the candidate that gains the highest support as the correct match for each edgel. For each edge, the edgels that have only one candidate will serve as "anchors" for this relaxation method. By linking the successfully matched edgels in the search images, we obtain the correspondent edge matc. Finally, we do a shape matching between the given edge and its correspondent edge through least squares adjustment. Only the edges with small shape matching errors will be kept.

Figure 3 shows an example of our edge matching.

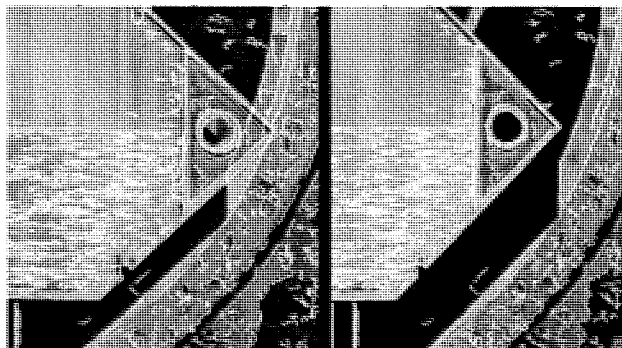


Figure 3: Edge matching with SI images (5 cm resolution)

3.4 Grid Point Matching

We use grid points to create uniformly distributed points over the whole images even in areas with very little or no texture. The correspondences of these grid points could be computed by using the method presented in chapter 3.2. Compared to the feature points, the choice of grid points is blind and thus many grid points are located in areas with weak or no texture. The search for the match of a given grid point has more possibilities to yield multiple candidates, or even no candidate.

To solve this problem, we use a global image matching method with relaxation technique. This method examines the candidates by computation how much support they receive from their local neighborhood and select the candidate that gains the highest support as the correct match. Here we use Prazdny's "coherence principle" model (Prazdny, 1985). We incorporate this idea in our global image matching and get the solution by relaxation technique.

Firstly, the points are selected in form of a regular grid in the reference image. Their matching candidates on the search images are computed. Together with all the matched feature points and edges they construct a TIN. It should be noted here that all the matched points can be categorized into three classes: Points having reliable matches, points having several candidates, and points without matching candidates. In the first case, they are treated as having only one matching candidate and, they serve as "anchors" for the global matching procedure. For the last case, they will be given several "false" candidates (with a very small correlation coefficient value) evenly distributed within the search window. The matched edges serve as break-lines in the TIN structure. They control the weights of the local smoothness constraints.

This method is performed on stereo pairs. The key point of this method, that distinguishes it from the single point matching, is its compatible coefficient function and its smoothness constraint satisfaction scheme. With the smoothness constraint, areas with homogeneous or only little texture can be bridged over, assuming that the terrain surface varies smoothly over the area. In the meantime, the surface discontinuities can be preserved because the smoothness constraints cannot cross the edges. For details of this procedure see (Gruen, Zhang, 2003).

3.5 Matching Through the Image Pyramids

A triangular irregular network (TIN) based DSM is constructed from the matched features on each level of the pyramid, which in turn is used in the subsequent pyramid level for the approximations and adaptive computation of the matching parameters. The matched edges are used as breaklines such that no triangle crosses these edges. The TIN maintains the original matching results without any interpolation. The surface discontinuities of the terrain can be well captured and preserved. The initial DSM for the highest level of image pyramid can be extracted by standard cross-correlation based on a "region growing" matching strategy. This method uses the already measured control and tie points as seed points and matches the points under the assumption that points in a local neighborhood should have similar disparities (Otto, Chau, 1988). This method is justified because the disparity surface can be treated as continuous and smooth on the lowest resolution image pyramid level. In some difficult areas like very rough alpine terrain, some

manually measured points can also be introduced as seed points. It will give better approximations for the matching.

3.6 Refined Matching Based on the Modified MPGC

MPGC (Multi-Photo Geometrically Constrained Matching) was developed by Gruen, 1985, and is described in detail in Baltasvias, 1991. It combines least squares matching and geometric constraints formulated either in image or in object space and permits a simultaneous determination of pixel and object coordinates. Any number of images (more than two) can be used simultaneously. The achieved accuracy is in the sub-pixel range.

Our modified algorithm is an extension of the standard MPGC. It integrates the geometric constraints derived from the Linear Array sensor models. The geometric constraints force the matching to search for a conjugate point only along a small band around the epipolar curve and reduce the possibility of false matches.



Figure 4: MPGC matching with multi-strip SI images.

Top: Images of a strip acquired from west to east

Bottom: Images of the cross-strip

The modified MPGC is used to refine the matched features in order to achieve sub-pixel accuracy. The DSM derived from the approaches (3.2)-(3.4) provides quite good approximations for the MPGC procedure and increases the convergence rate. The initial values of the shaping parameters in MPGC can be predetermined by using the image geometry and the derived DSM data. The corresponding rays of the four corners of the matching window in the reference image are intersected with the derived DSM and their object coordinates can be determined. Through imaging geometry, the corresponding image coordinates in the search images can also be determined. The initial values of the shaping parameters can easily be computed from these four corner points and their correspondences. By this way, features can be matched with multiple images or even with multiple image strips that have different flight directions and image scales. Figure 4 gives an example for the case of cross-strips.

For edge matching the parameters of a spline function of the 3D object edge are directly estimated together with the matching parameters of edges in the image spaces of multiple images. Some points, especially those grid points in non-texture or little texture areas, will fail in MPGC matching. These points are also kept but they are assigned low reliability indicator values.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of our approach for DSM generation it has been applied to different areas with varying textures, terrain types and image scales. In the following we will report about DSM accuracy test results with SI, IKONOS and SPOT5 HRS images.

4.1 SI Image Dataset, GSI area, Japan

In Japan's GSI (Geographical Survey Institute) test area, both SI images and aerial photos are available. The evaluation is based on the comparison between the manually measured DSM from aerial photos and the automatically extracted DSM from the SI images.

The GSI test area is roughly $650 \times 2500 \text{ m}^2$. It is relatively flat with natural and artificial objects. There are a lot of small geomorphological features, but also small discontinuities like cars, isolated trees and large discontinuities and occlusions due to buildings. For the success of a matcher it is very important how it can handle local discontinuities (e.g. buildings or other man-made objects, vertical cliffs, etc.).

Figure 5 shows two image windows from the nadir-viewing SI images. We have 48 control points. They are signalized marks on the ground or on the top of buildings. They appear both in the SI and aerial photos.

Two stereo pairs of color aerial images of 1:8000 image scale, acquired with a film camera of 153 mm focal length, have been used for manual collection of reference data on an Analytical Plotter. The RMSE of the exterior orientation is reported as 5 cm in planimetry and 3 cm in height. Manually measured points, distributed at about 50 cm distance in both X and Y directions, are used as reference data.

Assuming a measurement accuracy of 0.1% of the flying height (best case situation for natural points) we can expect an accuracy of the reference data of 0.12 m as a best case scenario.

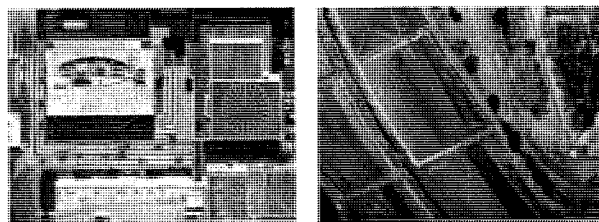


Figure 5: Region (left) of man-made objects. Region (right) of bare terrain area (including some sparse trees). The image patches are from the nadir SI-100 images with the ground resolution of 5.6 cm.

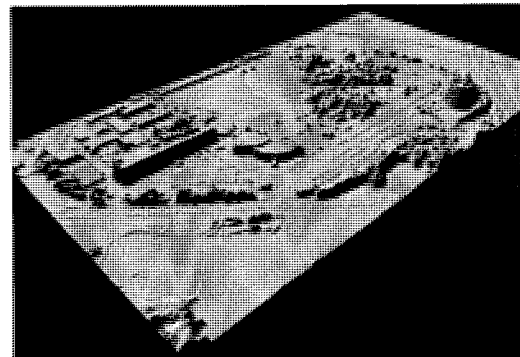


Figure 6: 3D visualization of the shaded DSM of the GSI test area.

Table 1: DSM accuracy evaluation. Manually measurement minus automatically extracted DSM

Area	RMSE (m)	Mean (m)	%				
			0.0-1.0 m	1.0-2.0 m	2.0-3.0 m	3.0-4.0 m	> 4.0 m
(1)	0.28	0.04	98.65	0.91	0.13	0.09	0.03
(2)	1.04	-0.16	74.70	12.15	7.32	4.84	2.36

The processed SI data include three panchromatic images with a footprint of ca. 5.6 cm. As a result of triangulation, 2.8 cm and 5.0 cm absolute accuracy in planimetry and height (as computed from checkpoints) were obtained.

For analysis of the matching accuracy, we divided the reference data into two classes, i.e. (1) the bare terrain area with some sparse trees and small artificial objects (including the large parking areas); (2) the area with man-made objects and trees. Our analysis has been performed for these two classes separately.

A very dense raster DSM with 15 cm interval was interpolated from the automatically matched point cloud and edges. The points with low reliability values were given a small weight in the interpolation procedure. Figure 6 shows the 3D visualization of the extracted DSM.

Comparing the reference points with the raster DSM in two different areas (1) and (2) leads to the results shown in Table 1. The bare terrain area shows much better results, but it still suffers from problems like multi-temporal differences between

aerial photos and SI images, and some sparse trees. In the man-made object area, the accuracy becomes worse. Here 2.4 % of all points have more than 4 meters difference. These points are almost all positioned on the border of buildings and trees. This may be caused by the fact that manually measured points and automatically extracted points are referring to different object points due to small errors in orientation procedure and 3D modeling problems.

4.2 IKONOS Image Dataset, Thun, Switzerland

The test covers the area around the town of Thun, Switzerland. It is about $17 \times 20 \text{ km}^2$. The terrain elevations range from 600 m to 2200 m, with some very steep and high cliffs. One IKONOS stereo pair was acquired on 2003-12-11 10:29 GMT over an area of approximately $11 \times 20 \text{ km}^2$. Another IKONOS triplet was acquired on 2003-12-25, 10:39 GMT over an area of approximately $11 \times 20 \text{ km}^2$. The two areas have nearly 50 % overlap. All images are IKONOS Geo products. The sensor and sun elevation angles (ca. 19 degrees) were less than optimal. The low elevation angle of the sun causes strong shadows, especially in the south part of the images and in general low contrast images. A 2 m resolution reference DSM generated from airborne LIDAR in the year 2000 was obtained from the Swiss Federal Office of Topography, Bern. The reference DSM only covers the south part of the whole area.

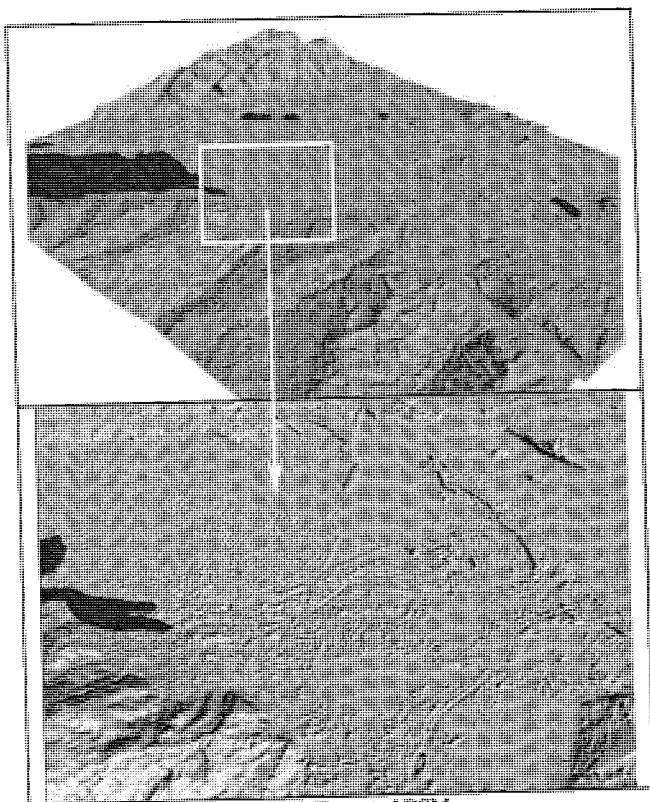


Figure 7: Extracted DSM
Top: The whole area; Bottom: Town of Thun

Firstly, the original 12 bit images were preprocessed and resulted in enhanced images for further image matching. Then the images were orientated with the help of about 40 GCPs measured by GPS. The orientation accuracy is about 0.41 m in planimetry and 0.68 m in height. Finally, a 5 m raster DSM was generated from the matched mass points and the edges. Some areas like lakes and rivers are defined as “dead areas” manually. Figure 7 shows the 3D visualization of the generated DSM (the whole test area and part of the town area). The results show that even small geomorphological features are extracted and surface discontinuities are well preserved.

Table 2: DSM accuracy numbers (west part of test area).
O-Open areas; C-City areas; V-Tree areas; A-Alpine areas.

Area	No. of Compared Points	Mean (m)	RMSE (m)	< 2.0 m	2.0-5.0 m	> 5.0 m
O+C+V+A	29,210,494	1.21	4.80	60.7%	16.6%	21.3%
O+C+A	17,610,688	1.11	2.91	77.0%	13.9%	10.1%
O+A	14,891,390	1.24	2.77	79.8%	12.2%	8.0%
O	11,796,795	1.00	1.28	90.3%	8.5%	1.2%

Table 3: DSM accuracy numbers (east part of test area).
O-Open areas; C-City areas; V-Tree areas; A-Alpine areas.

Area	No. of Compared Points	Mean (m)	RMSE (m)	< 2.0 m	2.0-5.0 m	> 5.0 m
O+C+V	20,336,024	-0.45	4.78	67.7%	21.3%	20.9%
O+C	13,496,226	0.33	3.39	68.7%	20.8%	10.3%
O	3,969,734	0.97	1.54	83.0%	15.0%	2.0%

Table 2 and 3 give the DSM accuracy test results. We compute the differences between the interpolated heights from our DSM minus the heights of the reference DSM. The accuracy of the generated DSM is at 1.3 – 4.8 pixel level. It depends on the terrain type. Higher accuracy can be achieved in open areas. In urban and tree areas the accuracy becomes worse. The analysis shows that points with more than 6 m differences are almost all distributed in the tree and urban areas. In open areas (with some sparse trees and small cluster of houses), more than 70 percent of the points have differences of less than 1 meter. The results show significant biases. This is caused by different point definitions in laserscan and photogrammetric surface points. Also, the different acquisition times of the IKONOS images and the laserscans may play a role. For more details see (Eisenbeiss, et al., 2004).

Table 4: Reference DEMs. Height accuracy in meter

DEM Name	Location	DEM Spacing	Source	DEM Size	Height Accuracy
DLR-DEM-01	Prien	5 x 5	Laser Scanner	5km x 5km	0.5
DLR-DEM-02	Gars	5 x 5	Laser Scanner	5km x 5km	0.5
DLR-DEM-03	Peterskirchen	5 x 5	Laser Scanner	5km x 5km	0.5
DLR-DEM-04	Taching	5 x 5	Laser Scanner	5km x 5km	0.5
DLR-DEM-05-1	Inzell-North	25 x 25	Laser Scanner	10km x 1.3km	0.5
DLR-DEM-05-2	Inzell-South	25 x 25	Contour lines	10km x 7.7km	5.0
DLR-DEM-06	Vilsbiburg	50 x 50		50km x 30km	2.0

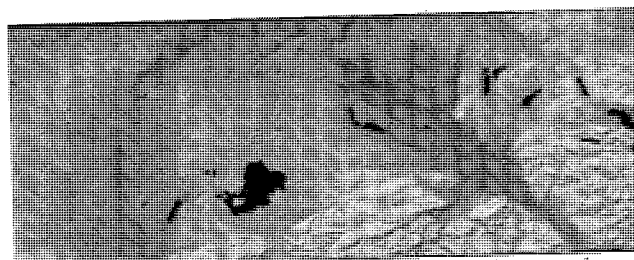


Figure 8a: Color shaded extracted DSM (the whole test area)

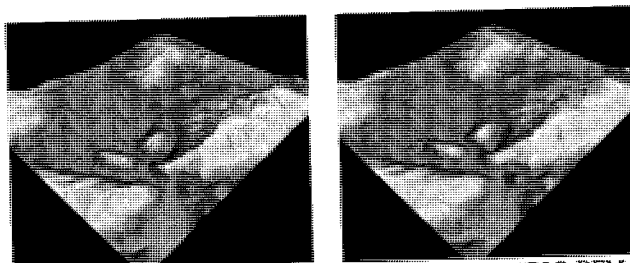


Figure 8b: Reference DEM (left, 25 meter grid of dataset “DLR-DEM-05-2”) and the generated DSM (right, 25 meter grid)

4.3 SPOT5 HRS Image Dataset, Bavaria, Germany

Here we report the work carried out within the ISPRS-CNES Initiative on DEM generation from SPOT5-HRS stereo images. For details see (Poli et al., 2004)

The test area (No. 9) covers a part of South Bavaria and a part of Austria (approximately $120 \times 60 \text{ km}^2$). Table 4 gives information on the reference DEMs. A stereo pair from SPOT5-HRS was acquired on 1 October 2002 in the morning. The

images have the resolution of 5 m in along-track and 10 m in across-track directions.

The test area includes a mountainous area (rolling and strongly inclined alpine area) and hilly areas (rough/smooth and weakly inclined areas). Our image matching software not only generates a large number of mass points, but also produces line features. The TIN based DSM was generated from the matched mass points and the edges (as break-lines).

Figure 8 shows the 3D visualization of the generated DSM. The results show that the shapes of our generated DSMs are similar to the references, but slightly smoothed. This can be expected because of the 5m resolution of the satellite images.

Tables 5 and 6 show the DSM accuracy test results. The orientation accuracy is about 6.3 m in planimetry and 2.6 m in height. We compute the differences between the heights of the reference DEM and the interpolated heights from our DSM. Table 6 shows the DSM accuracy test result by masking out the tree areas manually.

From Tables 5 and 6 it can be seen that:

- The accuracy of the generated DSM is more or less at the 1pixel level or even better. Only the datasets 5 give values at about 2 pixels, but these higher values are mainly caused by trees.
- All datasets still contain some blunders, which our procedures failed to detect.
- The results show systematic errors. In datasets 5-1 and 5-2 the biases are about 1 pixel. Except in case of dataset 6 all biases are significantly negative. This indicates that our generated DSMs are higher than the reference DEMs, an effect which could be expected..

Table 5: DSM accuracy, units are meter

DEM Name	Maximum Difference	Minimum Difference	Average	RMSE
DLR-DEM-01	25.1	-32.9	-2.6	5.7
DLR-DEM-02	29.1	-37.1	-1.2	5.0
DLR-DEM-03	20.7	-17.2	-0.5	3.2
DLR-DEM-04	13.6	-23.1	-2.5	4.7
DLR-DEM-05-1	19.2	-33.5	-5.8	8.3
DLR-DEM-05-2	136.8	-89.3	-4.3	9.5
DLR-DEM-06	26.8	-27.1	1.5	4.0

Table 6. DSM accuracy, units are meter (excluding the tree covered areas)

DEM Name	Maximum Difference	Minimum Difference	Average	RMSE
DLR-DEM-01	15.4	-23.7	-1.7	4.6
DLR-DEM-02	29.1	-31.7	0.2	3.6
DLR-DEM-03	20.7	-13.5	0.1	2.9
DLR-DEM-04	10.5	-18.4	-1.2	3.2
DLR-DEM-05-1	19.1	-13.3	-1.7	4.9
DLR-DEM-05-2	49.8	-66.8	-1.3	6.7
DLR-DEM-06	26.8	-25.9	2.1	4.4

5. CONCLUSIONS

In this paper we have reported about our current matching approaches for fully automated DSM generation from linear array images with different resolutions. We have developed a matching strategy combining feature point matching, grid point matching with neighborhood smoothness constraints, and robust edge matching. The strategy allows us to bridge over areas with little or no texture and at the same time maintain the important contribution of object/image edges. The modified MPGC is used to refine the matching results in order to achieve sub-pixel accuracy. The geometrical constraints are derived from the specific sensor models for the linear array imagery, which can be the rigorous sensor model for aerial and satellite images or the RF (Rational Function) model for satellite images.

As evidenced by a visual inspection of the results we can reproduce even small geomorphological features. The results from the quantitative accuracy test indicate that the presented concept leads to good results. If the bias introduced by trees and buildings is taken out, we can expect a height accuracy of one pixel or even better from satellite imagery (e.g. IKONOS and SPOT) as "best case" scenario. In case of very high resolution aerial images (footprint 8 cm and better) it is obvious that the "one pixel rule" cannot be maintained any more. Alone surface

roughness and modeling errors will lead to large deviations, such that an accuracy of three to five pixels should be considered a good result. This is at the same level as laser scanning results. Of course, the photogrammetric data can also be produced with the same or even better point density. On the other hand, with these accuracies we are still operating at a coarser level than with manual measurements from analogue aerial images, but we do that with the advantage of great gain in processing speed.

A major problem left is the control and automated detection of small blunders, which still infest the results, despite the simultaneous matching of more than two images. This constitutes a relevant topic for further research.

ACKNOWLEDGEMENTS.

We appreciate the support of the Swiss Federal Office of Topography, Bern, which provided the laserscan data. We also thank Henri Eisenbeiss, who helped in setting up the Thun area as a testfield for highresolution satellite image processing.

REFERENCES

- Baltsavias, E. P., 1991, Multiphoto Geometrically Constrained Matching. *Dissertation, IGP, ETH Zürich, Mitteilungen No. 49*, 221 pages.
- Eisenbeiss, H., Baltsavias, E., Pateraki, M., Zhang, L., 2004. Potential of IKONOS and QuickBird Imagery for Point Positioning, Orthoimage and DSM Generation. *IAPRS, Vol. 35, Istanbul, Turkey (to be published)*
- Gruen, A., 1985, Adaptive Least Squares Correlation: A powerful Image Matching Technique. *South Africa Journal of Photogrammetry, Remote Sensing and Cartography*, 14 (3), pp. 175-187.
- Gruen, A., Bär, S., Bühner, Th., 2000: DTMs Derived Automatically From DIPS - Where Do We Stand? *Geoinformatics*, Vol.3, No.5, July/August, pp. 36-39.
- Gruen, A., Zhang, L., 2002. Sensor Modelling for Aerial Mobile Mapping with Three-Line-Scanner (TLS) Imagery. *ISPRS Commission II Symposium on Integrated System for Spatial Data Production*, Xi'an, P. R. China, August 20 - 23.
- Gruen, A., Zhang L., 2003. Automatic DTM Generation from TLS data. *Optical 3-D Measurement Techniques VI*, Vol. I, Zurich, pp. 93-105.
- Kanade, T., Okutomi, M., 1994. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment. *IEEE Transactions on PAMI*, Vol. 16, No. 9, pp. 920-932.
- Murai, S., Matsumoto, Y., 2000. The Development of Airborne Three Line Scanner with High Accuracy INS and GPS for Analysing Car Velocity Distribution. *IAPRS, Vol. 33, Part B2, Amsterdam*, pp. 416-421
- Otto, G. P., Chau, T. K. W., 1988. A "Region-Growing" Algorithm for Matching of Terrain Images. *Proc. 4th Alvey Vision Club*, University of Manchester, UK, 31 Aug. - 2 Sept.
- Poli, D., Zhang, L., Gruen, A., 2004. SPOT-5/HRS Stereo Image Orientation and Automatic DSM Generation. *IAPRS, Vol 35, B 1, Istanbul, Turkey (to be published)*
- Saint-Marc, P., Chen, J-S., Madioni, G., 1991. Adaptive Smoothing: A General Tool for Early Vision. *IEEE Transactions on PAMI*, Vol. 13, No. 6
- Prazdny, K., 1985. Detection of binocular disparities. *Biological Cybernetics*, Vol. 52, pp. 93-99.