Adaptive Multi-Image Matching for DSM Generation from Airborne Linear Array CCD Data

Dr. sc. techn. Maria N. Pateraki

Zurich, 2005

This publication is an edited version of:

DISS. ETH NO. 15915

ADAPTIVE MULTI-IMAGE MATCHING FOR DSM GENERATION FROM AIRBORNE LINEAR ARRAY CCD DATA

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Technical Sciences

presented by
MARIA N. PATERAKI
Dipl. Eng., Aristotle University of Thessaloniki
born 27.01.1977
citizen of Greece

accepted on the recommendation of

Prof. Dr. A. Grün, examiner
Dr. E. Baltsavias, co-examiner
Prof. Dr. C.S. Fraser, co-examiner

2005

Adaptive Multi-Image Matching for DSM Generation
from Airborne Linear Array CCD Data

Maria N. Pateraki

Published by: Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology
CH-8093 Zurich

ISBN 3-906467-53-8

Έφερα τη ζωή μου ως εδώ
Πέτρα ταμένη στο υγρό στοιχείο
Πιο πέρα από τα νησιά
Πιο χαμηλά απ' το κύμα
Γειτονιά στις άγκυρες
Όταν περνούν οι καρίνες σχίζοντας με πάθος
Ένα καινούργιο εμπόδιο και το νικάνε
Και με όλα τα δελφίνια της αυγάζ' η ελπίδα
Κέρδος του ήλιου σε μία ανθρώπινη καρδιά
(Προσανατολισμοί - Οδυσσέας Ελύτης)

To the memory of Rena (1949-1977)

Foreword

In photogrammetry and remote sensing, data acquisition and processing is continuously turning to digital form. The only major exception up to the end of the 90s was airborne photogrammetric film cameras. At that time, the first digital photogrammetric cameras were announced, with the first order being placed for a Leica ADS40 during the 19th ISPRS Congress in Amsterdam, July 2000. Since then, various commercial models have been introduced, with the great majority of orders and sales being for ADS40 (Leica Geosystems), DMC (Z/I Imaging) and UltraCamD (Vexcel). Although the number of film cameras still by far outnumbers the digital ones, it seems that now digital cameras are accepted by the market, and their installation number will soon approach 100. The introduction of digital cameras was a "revolution". It allows a fully digital processing chain and offers various advantages like better radiometric resolution, higher productivity and flexibility, and simultaneous acquisition of panchromatic and four spectral channels. On the other hand, these cameras, independently of whether they use linear or area CCDs, often need, at least partially, different data processing methods. This is particularly true for cameras using linear CCDs, as the previously existing photogrammetric software systems could not process data from such sensors, while linear CCD sensors and ADS40 have special characteristics that allow different and better data processing and performance.

As with older breakthrough developments, like the introduction of the digital photogrammetric stations, at the beginning some "children illnesses" have been observed. Companies often promise too much and deliver less, systems are introduced commercially too early without being mature enough and well tested, and even more importantly the necessary methods and software are missing to a large extent and are underestimated compared to the hardware. Having the older experiences in mind, I thought that such problems should be minimised. Otherwise, frustrated customers and negative experiences could lead to a delay in the acceptance of these advantageous new sensors. Thus, taking into account that ADS40 was the first commercially introduced digital camera and that Leica and ETH Zurich always had a close co-operation, I proposed to Leica the initiation of a Swiss project for development of methods and software for one central aspect of image processing, namely the development of matching methods, tailored to the specific characteristics of ADS40, for the automated generation of Digital Surface Models (DSMs) and possibly automated point transfer in aerial triangulation. My proposal was accepted by Leica and a successful application to KTI (Swiss Innovation Promotion Agency) allowed the initiation of the project AIM (Adaptive multi-Image Matching for automatic terrain model generation from digital airborne sensor data), within which this dissertation was developed, in early 2001.

As we all in universities know, even if a project is well formulated, it fails without the appropriate Ph.D. students to perform good research and "pull the train". Luckily at the end of 1999, a Greek Socrates exchange student from the Aristotle University of Thessaloniki, Maria Pateraki, came to ETHZ and after the end of her Socrates stipendium, stayed over at ETHZ with aim to do her Ph.D. with us. After flirting with quite different topics like medical image analysis or even underwater photogrammetry for archaeological research, and some work within the EU Cloudmap project, she started her Ph.D. on the project AIM, a topic that was a natural continuation of her diploma thesis "Studies on Automatic DTM generation using the Match-T algorithm".

Automated DSM generation using image matching algorithms has been researched for over 20 years. In spite of advances in developed methods and algorithms, the aim of achieving a dense, precise and reliable DSM has not been fully achieved and great weaknesses exist especially in the reliability and success rate of the results. Maria tried to develop in her research new methods, taking into account previous work and the specific characteristics of ADS40, like availability of seven CCD lines with 100% overlap in flight direction, a higher radiometric resolution etc. In parallel, she had to develop software to be integrated in Leica's existing commercial software (at that time SocetSet).

In her thesis, she first gives a very nice and comprehensive overview of airborne digital photogrammetric cameras. First, the CCD and CMOS technologies are compared, listing advantages and disadvantages. Then, a complete and very useful list of airborne digital photogrammetric cameras is given, separated in area and line CCDs, whereby for the frame CCDs a meaningful separation in small, medium and large format cameras is made. Then, an analysis of ADS40 is presented in many details, giving probably the most complete description of ADS40 published to date, which is very useful as general knowledge but also as background information for the rest of the work.

A valuable part of her work deals with radiometric analysis and image preprocessing. These aspects are very important but often underestimated or poorly treated. The analysis and processing methods mentioned are generally applicable to any type of digital images and have been used successfully in many projects at our Institute. The preprocessing methods are of very good quality, reducing noise with parallel enhancing of grey level edges and no removal of thin lines and small structures. The work studies aspects of the reduction of more than 8-bit data to 8-bit, which is getting more important as more and more sensors provide more than 8-bit images, while operational processing is performed with 8-bit.

In the main body of her work, she first studied different methods to implement geometric constraints, for both raw and rectified images. Although some of the implementations have been previously published, new aspects are added on the use of constraints with more complicated form than straight lines. Then, the derivation of approximate values is studied, using image pyramids. Thereby, a new aspect is the use of the so-called doublets, which reduces the amount of necessary interpolations when transferring results from one pyramid level to the other, thus speeding up processing and reducing propagation of errors. Matching using multi-patch-size cross-correlation is proposed in the upper pyramid levels, in order to enable both reliability and precision, and derivation of matching quality measures. A major part of her work refers to modifications of least squares matching, mainly focussing at the very important issue of getting denser and more accurate values along and across surface discontinuities without smoothing them. The thesis presents two interesting and new approaches. The first is precomputing shape, size and rotation of the matching patches, while at the same time improving approximations for the necessary rotation of the search patches. This is proposed to be used for curved or short straight edges. For long straight edges, a second approach using the whole edge and subdividing it sequentially, while using separate weights for the observations is proposed. This approach partly also handles the mainly still unsolved problem of matching at edges parallel to the epipolar lines. Next, two possible matching strategies are presented, one using a single template image, and another one using multiple templates. The use of multiple templates leads mainly to reduction of problems due to occlusions, encountered especially in built-up areas. A last part of the work refers to the most critical matching aspect, namely automated quality evaluation and error detection. The thesis makes use of previously proposed quality measures, while proposing some new ones. Quality measures are proposed for both matching algorithms employed, the multi-patch-size cross-correlation and the modified least squares matching.

The evaluation of the developed methods, in spite of scarce good test data and reference DSMs, and the comparison to one of the most widely used commercial package (SocetSet) verified the improvement brought by this work.

Maria's work expands the current body of knowledge, as well as improves the current state of matching algorithms with various proposed methods and strategies. These are valuable not just for the ADS40 sensor, but can be used with other linear and even frame cameras as well.

Maria's work was hampered to a significant extent by problems regarding poor ADS40 data quality in the project, the lack of reference DSM data, compulsory use of various Leica software libraries in her programs, bugs in and frequent change of these libraries etc. She is quite modest and careful when mentioning these difficulties in the introduction of her thesis. But I know from first hand how many frustrating nights she spent at Hoenggerberg trying to cope with such problems.

Her continuous interest and warm engagement in the project was recognised by Leica, which after the end of the KTI project supported her work financially for over one year, while also offering her a position at Leica, even before finishing her Ph.D. But Maria was not only putting a lot of effort to do a good job regarding her research. Over her whole work within our group, she was investing a lot of time, trying always to do a thorough and good work, and even neglecting her own higher-priority interests, in various tasks of our group, even if these tasks were sometimes not strictly her duty. Lectures of courses, supervision of students, organisation of events, editing of proceedings, various projects (high-resolution 3D modelling of Everest, processing of high-resolution satellite imagery, performance analysis of commercial matching systems, use of photogrammetry in cultural heritage, e-learning) are some examples where she dedicated considerable efforts, with her heart, as Greeks say. Maria was always full of life and energy (the Cretan agile spirit in her!) and socially integrated within our group and in her private life, and a very good company to be with.

A personal note. Supervisors often think about what they want from their Ph.D. students, but I am afraid they rarely think what their students expect from them! Mutual understanding and help and open communication would be ideal but unfortunately do not often exist. In that respect, I have to apologise for insufficient supervision and help during a period of Maria's Ph.D. when I had some serious personal problems. She lived with my ups and downs and showed much understanding and never complained (unfortunately!). Not to mention that her final draft thesis looked after my corrections more red than black, which is not very motivating, I must confess. I could at least have used green colour, as Maria correctly pointed out.

We believe that this work is a valuable scientific contribution to a very important and still unsolved topic and congratulate Maria for her achievements.

Ευχαριστώ Μαρία για την δουλειά και την φιλία σου και καλή τύχη στη ζωή σου!

Manos Baltsavias, Ph.D. thesis supervisor Zurich, July 2005

Table of Contents

Abstract

This thesis investigates the topic of automatic image matching with focus on the generation of Digital Surface Models (DSMs) by using imagery acquired from airborne linear array CCD sensors. The research has been motivated by the recent developments in photogrammetric equipment related to sensor technology, which have introduced a new field of research. Airborne digital cameras that employ linear array CCD sensors, exhibit different radiometric and geometric characteristics compared to conventional film based cameras and new methods are required for processing the data from these sensors and generating products for different applications. In addition, with respect to matching, existing algorithms are geared towards frame imagery, plus the degree of automation is limited. In most cases, current commercial systems show poor success rate and require manual interaction for editing the matching results.

This study is embedded in the framework of the AIM project (Adaptive Image Matching), in which data from the airborne digital camera (ADS40) of LGGM (Leica Geosystems GIS & Mapping) are used as input. Existing matching algorithms are analyzed, further modified and new ones are developed aiming at integrating information from the special characteristics of the sensor in the matching philosophy. First, the camera architecture, the radiometric and geometric properties of the sensor, the calibration of the system, the ground processing workflow and the sensor model are investigated. Then, this research is focused on two main issues, namely the evaluation and enhancement of the image quality and the development of the matching strategy. The radiometric analysis and the preprocessing part include methods for noise estimation, noise reduction, contrast and edge enhancement, radiometric balancing, reduction to 8-bit and processing of multispectral channels. This part is important and if omitted, the matching performance is influenced. The matching strategy consists of different modules that are evaluated individually but also on an integrated basis. In a nutshell, the individual matching aspects that are investigated are: the implementation of geometrical constraints, the derivation of approximate values, the extraction of features, the integration of different matching methods and the quality control and error detection. Geometrical constraints are used to strengthen matching and are employed by means of *quasi*-epipolar lines. Due to the complex geometry of the images (position and attitude information for each line), epipolar lines do not really exist and the epipolar trajectory is modelled over a short length by a second or first degree polynomial equation. Moreover, hierarchical techniques are utilized to gradually refine the matching results. The investigations are performed with respect to generation of image pyramids, for which different filters are utilized and evaluated, and to the selection of the doublets as an optimal strategy towards better time performance and reduction of propagation of matching errors to lower levels. Apart from the above, in the matching strategy feature- and area-based methods, plus methods that aim

at higher reliability and/or at precision, are combined and different primitives (grid points, edgels and edges) are used. The selection of edges, as matching entities, resulted from the evaluation of different feature extraction algorithms (points and edges), based on a set of criteria. Then, the efforts have been mainly focused on edge matching, in order to improve modelling of discontinuities. Different approaches were investigated that led to significant improvements: the use of height and continuity constraints for contour points and extensions of LSM for edge features. In the LSM for edge features both edgels and long, straight edges are handled. Moreover, the ADS40 with the configuration of the channels on the focal plane and their viewing angles permits the use of more than one template and to facilitate the identification of errors occurring in matching, especially occlusions. The role of the different combinations of channels in matching is discussed and the matching block, based either on a single- or multi-template strategy, is described.

Other major aspects of these investigations are the quality control and error detection strategy. Each individual ray is checked based on a set of criteria and pre-defined error types. In the quality control, measures derived from different matching methods (multi-patch matching, LSM, edge matching) are combined, problematic rays are excluded and each 3D point is computed from the good rays only. The performance of the system has been evaluated over different areas of land cover and for different point classes (breaklines and points, on the ground and on anthropogenic objects). A detailed analysis of the results and the statistical measures that have been derived from the tests are presented and discussed. The derived accuracy of the automatic measurements is close to the accuracy of the manual measurements. According to the studies, blunders in the results of AIM are significantly less compared to the results of the commercial system Socet Set 4.4.1 (SS). For AIM the matching accuracy on anthropogenic objects was 0.5-0.66 m, whereas for SS it was > 1 m, especially in dense built areas.

Résumé

Cette thèse étudie le sujet de la mise en correspondance (matching) automatique des images et focalise sur la génération des modèles numériques de surface (DSMs), en employant l' imagerie acquise par senseurs linéaires CCD aéroportées. La recherche a été motivée par les développements récents sur les équipements photogrammétriques liés à la technologie de senseur, qui ont ouverts un nouveau domaine de recherche. Les caméras numériques aéroportées qui utilisent les senseurs linéaires CCD, présent des différentes caractéristiques radiométriques et géométriques, comparées aux caméras conventionnelles à film. De nouvelles méthodes sont nécessaires à traiter les données de ces senseurs et genérer des produits pour différentes applications. En outre, en ce qui concerne le mis en correspondance, les algorithmes existants sont orientes vers l' imagerie conventionnelle, de plus le degré d' automation est limité. Dans la plupart des cas, les systèmes commerciaux courants montrent un faible taux de succès et exigent une interaction manuelle pour éditer les résultats du matching.

Cette étude est inclus dans le cadre du projet AIM. (Adaptive Image Matching), dans lequel les donnιes de caméra numérique aéroportée (ADS40) de LGGM sont employées comme entrιe (input). Les algorithmes de matching existants sont analysés, puis modifiés et de nouveaux algorithmes sont développés avec l' effort d' intégrer l' information des caractéristiques spéciales du senseur dans la philosophie du matching. Premierment, l' architecture de la caméra, les propriétés radiométriques et géométriques du senseur, le calibrage de systéme, le déroulement des opérations de traitement (ground processing) ainsi que le modéle de senseur sont étudiés. Ensuite, la recherche est concentrée sur deux thèmes principaux : l' évaluation et l' optimisation de la qualité d' image, puis le développement de la stratégie du matching. L' analyse radiométrique et la prétraitement incluent des méthodes pour l' estimation de bruit, la réduction de bruit, l' optimisation de contraste et des contours, l' équilibrage radiométrique, la réduction à 8 bits et le traitement des canaux de couleur. Cette partie est importante et si omise, la performance du matching est influencée. La stratégie du matching se compose de différents modules qui sont évalués à la fois individuellement mais aussi sur une base intégrée. En un mot, les différents aspects du matching qui sont étudiés sont : l' implémentation de contraintes géométriques, la dérivation de valeurs approximatives, l' extraction de traits, la integration de différentes méthodes de matching ainsi que le contrôle de qualité et la détection d' erreurs grossières. Les contraintes géométriques sont employées pour renforcer le matching et sont utilisιes au moyen de lignes *quasi* épipolaires. En raison de la géométrie complexe des images (l'information de position et d'attitude pour chaque ligne), les lignes épipolaires n'existent pas vraiment et la trajectoire épipolaire est modelisée sur une courte longueur par une équation de polynôme de deuxième ou de première degré. D' ailleurs, des techniques hiérarchiques sont utilisées pour raffiner graduellement

les résultats du matching. Les investigations sont effectuées en respectant la génération de la pyramide d' image, pour laquelle différents filtres sont utilisés et évalués, et le choix d' une stratégie optimale en terme de temps d' exécution et de réduction d' erreurs de matching à des niveaux plus bas. Indépendemment de ce qui précède, les méthodes basées sur les traits et les secteur, plus les méthodes qui visent soit la fiabilité et soit l' exactitude, sont combinées dans le stratégie de matching et différents primitifs (grid points, edgels, edges) sont employés. La sélection des contours, comme les entités du matching, a résulté de l' évaluation de différents algorithmes d'extraction de traits (points et contours), basée sur un ensemble de critères. Puis, les efforts ont été principalement concentrés sur le mis en correspondane des contours, afin d' améliorer la modelisation des discontinuités. On a recherché différentes approches qui ont mené à des améliorations significatives : l'utilisation de contraintes de hauteur et de continuit: pour les points de contour et les prolongements de LSM pour des traits de contours. D' ailleurs, ADS40 permet avec la configuration des canaux sur le plan focal et leurs angles de vue son employer plus que un template, et facilite l' identification des erreurs se produisant dans le matching, particulièrement les occlusions. Le rôle des différentes combinaisons de canaux dans le matching est discuté et le bloc de matching, basé d' une stratégie mono ou multi-template est décrit.

Un autre aspect principal de cette recherche est le contrôle de qualité et la détection des erreurs grossières. Chaque rayon individuel est vérifié sur la base d' un ensemble de critères et des classes d' erreur prédéfinies. Dans le contrôle de qualité les mesures dérivées de différentes méthodes de matching (multi-patch matching, LSM, edge matching) sont combinées, les rayons problématiques sont exclus et le point 3D est calculé par les bons rayons seulement. La performance du système a été évaluée aux divers secteurs de couverture de terrain et aux diffrentes classes de point (breaklines et points, sur le sol et sur les objets). Un analyse détaillée des resultats et des mesures statistiques qui ont été dérivées des essais sont présentées et discutées. L' exactitude des mesures automatiques est proche de l' exactitude des mesures manuelles. Selon l' analyse, les erreurs grossières dans les résultats du AIM sont significativement moins comparées aux résultats du système Socet Set 4.4.1. Pour AIM l' exactitude de matching sur les objets était de 0.5 -0.65 m et pour SS était > 1 m.

Aknowledgements

When I first came in ETHZ in December 1999 as an exchange student from the Aristotle University of Thessaloniki I didn't plan to do a Ph.D. related to image matching nor to Airborne Linear Digital Cameras as I was attracted, merely to say captivated, by medical imaging at that time. However, the change was inevitable as I joined the Photogrammetry and Remote Sensing Group, recognized for its extensive research in image matching. On the other hand, the general excitement of the photogrammetric community on the Airborne Digital Cameras was considerably larger than my own on medical imaging and this had an obvious effect on my initial decision. It was then more than a challenge for me to work in this field, with this research group and I knew it would be hard. I had no idea. Using data from a sensor under finetuning at the same time as developing methods to deliver good matching points from these data only adds to the stress of a Ph.D. research and makes the thanks I express here even more heartfelt.

Writing a doctorate is an intensely lonely experience, but at the same time it can only be done with the help and co-operation of a number of people. All these people now deserve a special mention.

First mention must go to my direct supervisor, Dr. Emmanuel Baltsavias, whose belief in me from our first meetings and his valuable advices on dealing with different obstacles I encountered while working in this project, critical remarks in image matching but also enlightening discussions about anything i have not be able to understand myself, through to this Ph.D. has given me the lift I needed to get it done. Acknowledgements sections are often full of phrases such as *'without whom'* but in the case of Manos is more than true. Without Manos this work would not have been done. Of course, without being employed as a *Wissentschaftliche Mitarbeiterin* by the Head of the Group, Prof. Dr. Armin Gruen, in the first place, my research output would have been none. Thank you both.

Except the two important supervisors mentioned above I would like to thank Prof. Dr. Clive Fraser from the University of Melbourne for his prompt response and effort in reading the first completed draft of my thesis and his significant remarks, which were appreciated the most, as they were always expressed through means of constructive critic. I am grateful to him not only for investing time for reviewing the scientific part but also for his corrections on the linguistic part.

I could not also omit thanking Prof. Petros Patias, who was not only my supervisor during my undergraduate studies in the Aristotle University of Thessaloniki but also the person who instigated my interest in the field of digital photogrammetry through his enthusiasm as an educator, who's advices all these years were and are still valuable and most important he gave me the possibility of a six month visit in Zurich through Socrates mobility program.

Both the Commission of Technology and Innovation of the Federal Office for Professional Education and Technology and Leica Geosystems GIS & Mapping (LGGM) have provided the required funds in order to realize this research. In this context their contribution is gratefully acknowledged. I would also like to direct my special thanks to a number of persons from the LGGM team, related to the development of the ADS40 sensor, for their assistance during the project AIM. Muzzafar Adiguezel, Peter Fricker, Francois Gervaix, Werner Kirchhoffer, Utz Recke, Tauno Saks, Fernando Shapira from the LGGM offices in Heerbrugg, especially Udo Tempelmann for his constant and vital support in technical and development related difficulties. Robert Uebbing, Belay Beshah and Neil Woodhouse from the LGGM team in San Diego, California for their prompt response and assistance à distance for software integration and sensor model aspects. The contribution of PASCO Corporation from Japan is also acknowledged for providing the Yokohama data set.

Many thanks to my colleagues from the Photogrammetry and Remote Sensing Group for their help, collaboration and good spirit, especially to Dr. Chunsun Zhang, Dr. Gabriela Seiz and Dr. Nicola D'Appuzo for the brainstorming sessions on image matching and coding; Dr. Daniela Poli for her input on pushbroom sensors; Gillian Gruen and Natasha Vasilieva for their endurance in performing streneous stereo measurements of reference data; Beat Rueedin for his IT support and the secretaries Liliane Steinbruckner and Gertrud Rottenberger for their support in administrative issues.

During these years my closest friends, in and out of Zurich, Andreas Chaniotis, Alexander Christides, Christina Doukouzgianni, George Gerostathopoulos, Sofia Karamanoli and Dimitris Meretakis have contributed to a common story whose colourful pages now lead me towards a new chapter. Their optimistic views and moral support enabled me to complete this thesis. My deepest thanks however is owed to Georgia Fotiou who was more than a friend to me, tolerated my personal eccentricities, kept on constantly energizing me in periods of continuous day/night working hours and unexpected changes/disappointments in my life. She was always there and she could turn black into white.

My acknowledgements are extended to Kostas, who's sustained support in the early years was unquestionable and is still unforgettable.

And lastly, there are no words that will suffice to thank my father Nikos and Liana, who believed in me from the very start all those years ago, have always supported me in my endeavors, always given me the strength and encouragement to follow my dreams even when it was hidden behind the spare batteries and lost in thousand lines of code, and have never left me in doubt of their love for me.

Many thanks to all of you and to those I miss ...

Maria N. Pateraki

List of Abbreviations

ABM Area Based Matching

ADS40 Airborne Digital Sensor

AIM Adaptive multi-Image Matching for automatic terrain model generation from digital airborne sensor data

AIMS Airborne Integrated Mapping System

AT Aerial Triangulation

CCD Charge-Coupled Device

CMOS Complementary Metal Oxide Semiconductor

CNES Centre National d'Etudes Spatiales

CVG Coded Vertical Goniometer

DAIS Digital Airborne Imagery System

DDS Digital Sensor System

DLR Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Agency)

DMCS Digital Modular Camera System

DPA Digital Photogrammetric Assembly

DPW Digital Photogrammetric Workstations

DSM Digital Surface Model

DSNU Dark Signal Non Uniformity

DTM Digital Terrain Model

EOS Electro Optical System

FBM Feature Based Matching

FOV Field Of View

FP Focal Plate

FPS Focal Plate coordinate System

GPS Global Positioning System

GSD Ground Sampling Distance

HRSC High Resolution Stereo Camera

HRV High Resolution Visible sensor

HSI Hue Saturation Intensity

IGN Institut Geographique National, France

IMU Inertial Measurement Unit

ISRO Indian Space Research Organisation

LGGM Leica Geosystems GIS & Mapping

LISS Linear Imaging Self Scanning

LSM Least Squares Matching

LUT Look Up Table

MMS Mass Memory System

MPGC Multi Photo Geometrically Constrained

MTF Modulation Transfer Function

NCC Normalized Cross Correlation

ODF Orientation Data File

ORIMA Orientation Management software

PRNU Photo Response Non Uniformity

PSF Point Spread Function

SAD Sum of Absolute Differences

SNR Signal to Noise Ratio

SS Socet Set

SSD Sum of Squared Differences

SUSAN Smallest Univalue Segment Assimilating Nucleus

TLS Three Line Scanner

UHRRC Ultra High Resolution Reconnaissance Camera

WAAC Wide Angle Airborne Camera

YIQ Luminance (Y) In-phase (I) Quadrature-phase (Q)

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The research on this topic has been initiated for three primary reasons. The first relates to the recent development of new digital airborne photogrammetric sensors, which is a milestone in the development of photogrammetric equipment. These sensors offer the possibility of a fully digital production chain from acquisition to various value-added products such as DTMs, DSMs, orthoimages, maps, 3D object modelling, landcover and land use classification and other thematic applications, 3D visualisation, animation and simulation. Airborne line-CCD sensors offer in a single lens camera system the possibility to acquire both panchromatic and multispectral images, thus broadening the possible application areas of these sensors. With line-CCDs the images overlap 100% (with frame imagery typically only 60%), while the perspective distortions in the flight direction are small due to the quasi parallel projection. Moreover, the radiometric quality of digital images compared to scanned film imagery is significantly better. The on-board integration of GPS/INS facilitates the determination of sensor orientation for each image line, and it reduces the need for costly and time-consuming establishment of ground control points, thus significantly increasing the possible use of the sensor in countries where control point establishment is very difficult to impossible.

The second reason is associated with image matching, which is characterized as one of the most fundamental processes in digital photogrammetry and is applied in operations like establishment of the interior and exterior orientation of the images, point transfer and aerial triangulation, feature extraction and DTM and DSM generation. The use of DSMs in the photogrammetric mapping chain is a key to change detection and database updating (Paparoditis et al., 2001), production of true orthoimages (Baltsavias et al., 1995) and generation of 3D CAD models (Brunn et al., 1998; Weidner, 1997). Although several matching methods have been developed over the last 25 or more years and significant efforts continue towards fully automated matching, the correspondence problem remains a difficult and complex task, for which no general solution has been found. With respect to automated DTM and DSM generation, existing methods in various digital photogrammetric systems show poor reliability and in some cases low success rates (Baltsavias and Käser, 1998). Thus, the use of automated DSM and DTM generation has found limited practical use and relevant research has not led to significant improvements in the last decade. In addition, all existing matching algorithms are geared towards frame aerial imagery. For the present airborne linear CCD sensors, matching methods have not significantly advanced, and there has been little explicit exploitation of the specific characteristics of linear CCDs in matching.

Last but not least, with respect to the extensive know-how within image matching in our group at ETH Zurich, and with the development of the world-wide first commercial photogrammetric airborne digital sensor (ADS40) by LGGM, this research appeared to be as challenging as it finally proved to be.

1.2 Research aims of project AIM

The work presented in this thesis has been embedded in the project AIM (Adaptive multi-Image Matching for automatic terrain model generation from digital airborne sensor data), financed by the Commission of Technology and Innovation (KTI) of the Federal Office for Professional Education and Technology and the company LGGM, Heerbrugg. The main aim was to develop methods for finding automatically and accurately corresponding features in two or more images of linear array CCD cameras, with a focus on automatic generation of DSMs from rectified images. Another aim, of much less importance, is the use of the same or slightly modified matching methods to find tie points in raw image data for digital aerial triangulation. Therefore, the methods should not be restricted only to the processing of rectified images, and should be easily extendable and adaptable to raw images as well. The input data are digital data from the photogrammetric airborne digital camera (ADS40) of LGGM. The tasks in this scientific work include:

a. Test and compare various processing and matching algorithms and methods and evaluate their performance on acquired ADS40 imagery.
b. Refine, develop and combine processing and matching algorithms. Various side aspects of the matching procedure are to be investigated individually, but also on an integrated basis.
c. Quality evaluation of the automatically derived DSMs using different datasets.

To a certain extent, there are some promising approaches mentioned in research literature, but they have not been tested extensively and/or refer to simpler applications compared to DSM generation from aerial imagery, e.g. interior orientation, semi automatic measurement of targets. Our group has worked researchwise on many aspects of image matching and possible approaches have been formulated, but these have to be tested with respect to ADS40 images. The promising approaches are kept and are further refined. Matching techniques, especially those which are commercial, have in most cases restricted themselves to one of the two common matching classes, area-based or feature-based. Both have advantages and disadvantages, which also depend on the terrain roughness and cover, and the image characteristics. In our approach, such techniques are combined. The key aspects of the new approach are: (a) the use of multi-image matching, leading to substantial reduction of problems caused by occlusions, multiple solutions, image noise, and surface discontinuities, and to higher measurement accuracy through the intersection of more than two image rays, (b) consideration of the fact that the base between all images is in one direction, while in classical frame aerial imagery the base between overlapping pairs of images varies, and (c) consideration of the almost parallel projection in flight direction, which leads to less perspective differences and occlusions. Moreover, the incorporation of a priori information about the sensor orientation is investigated to reduce the search space. Matching results are compared with manually measured reference data and statistical measures are extracted. In addition, the results are further compared with the matching results derived from the commercial system Socet Set 4.4.1. (SS), during this research[1] the only system able to fully process ADS40 images.

However, one has to accept that difficulties are part of every research. In general, these are related mostly to weaknesses in the scientific hypothesis and the adopted approaches, the development and testing of methods and further the limitations in data availability. In this research work, the majority of encountered problems was related to hardware faults in the ADS40 sensor system and software aspects of the ADS40 workflow, while other problems were due to poor quality test data and the need to use specific libraries in the own software. These problems caused significant time delays in the development, implementation, testing and refinement of methods and also in the time invested for own research, considering that in parallel an amount of time was used to identify and report errors that existed in the input data sets. The below briefly summarized problems can be justified to a large extend, as ADS40 was a newly introduced sensor under fine tuning:

- The ADS40 sensor model, designed and implemented by LGGM, was used in this work and the problems were related to the sensor model implementation (software bugs, high computing time and memory leaks). Observed errors in the sensor model transformations influenced the overall geometric accuracy of the model and consequently its further usage in own developed methods.
- Inheritance of the Socet Set libraries in software development, according to the guidelines that were defined in the project description. The developed code had to be dynamically linked with libraries of Socet Set (mainly for the integration of the LGGM sensor model). The problem was that there was no control over the source code of Socet Set libraries and their integration was a relative complex task due to insufficient documentation. Moreover, frequent updates in their content and structure, thus different incompatible versions were the main reason of higher integration and software development effort.
- Poor initial data sets, due to the start difficulties and fine tuning of the sensor, with respect either to their radiometric properties and the amount of artefacts in the image and/or to the geometric integrity (e.g. problems with the GPS/IMU measurements, geometric sensor model, block geometry). These data sets could not be used for benchmark tests and exploitation of the matching performance.
- Reference data were not available for specific datasets or were difficult to acquire. Collection of reference data over different areas (e.g. by manual measurements in stereo mode) was time consuming (approximately 2 months) and required an experienced operator, which was not always available.

Although the above problems were frequently occurring in the first period of this research work, they have been reduced significantly towards the completion of the project.

[1] Recently the LPS photogrammetric software advanced and can process ADS40 images as well.

1.3 Outline of dissertation

This thesis consists of seven chapters. Chapter 2 provides an insight into current developments with respect to digital photogrammetric cameras. Aspects related to CCD technology are briefly presented and an overview of existing frame and linear array cameras is given, with a focus on the ADS40 sensor. In Chapter 3, the radiometric analysis and preprocessing that has been performed on ADS40 data is presented. The importance of the material in this chapter should not be underestimated, as preprocessing can significantly influence matching. Moreover, almost all developed methods can be also employed with any images. Developments with respect to algorithms are presented in Chapter 4, followed by the general description of the AIM algorithm. Chapter 5 analyzes in more detail the individual modules of AIM and justifies the selected approaches, whereas in Chapter 6 the experimental matching results for DSM generation are analyzed with respect to manual measurements (reference data) of different point classes and further compared with the results derived from the commercial system SS. Finally, Chapter 7 summarizes the main findings, problems, and advantages and disadvantages of the method and provides an outlook for further research.

Chapter 2

Airborne digital photogrammetric cameras and ADS40

2.1 CCD Technology

Developments in the semiconductor industry have made possible the widespread use of semiconductor sensors in a vast number of systems serving different applications. This also applies to image sensors which form the operational core of the digital cameras. However, although digital cameras have in many ways exceeded the capabilities of film cameras, the human eye remains the ultimate standard for comparison. The eye has the ability to sense light but also process light information, before even the converted signal is sent to the brain. By genetic design, the eye encodes the vast visual input in such a way that the limited neural output retains the most significant descriptors of the scene while the rest are discarded (Atick and Redlich, 1990). In this context commercial efforts in image sensing seek to reproduce the biological function, more than the biological structure and merely because of the differences between semiconductor physics and cellular biology, information processing structures must be tailored to the medium (Giles, 2001; Dileepan, 2002).

The two types of imaging sensors mostly used are the Charge-Coupled Device (*CCD*) and the Complementary Metal Oxide Semiconductor (*CMOS*). These sensors have similar light sensitivity over the visible and near-IR spectrum, however, their electronic design differs. The CCD sensor was invented in 1970 by Boyle (Boyle and Smith, 1970) and consists of a discrete array of photosensitive pixels that produce an electrical charge proportional to the amount of light they receive (photoelectric effect, Figure 2.1). Typically, pixels are arranged in either a single line (linear array CCDs) or in a two-dimensional grid (frame array CCDs), as shown in Figure 2.1. Once the integration time has elapsed, the pixel charge is further transferred across the surface of the array into an adjacent readout register, where a time discrete signal is generated before being converted into an equivalent digital value from an analog-to-digital converter and stored in memory. This generic description of the readout process for frame arrays will vary depending on the overall architecture of the device (full frame, frame transfer, split frame transfer and interline transfer), for which a detailed description can be found in Taylor (1998). The image collection mode of a frame array sensor may vary and can be either *interlaced*, where the frame consists of an odd and an even field corresponding to odd and even numbered rows respectively, which are recorded at different instances of time, or *progressive*, where the complete frame is recorded at one instance of time. A major drawback of the interlaced method is encountered when capturing moving objects. Since the two fields are separated in time, the position of the object will have changed, resulting in blur when combining the two fields to produce the final image. It is worthy of note also that CCD technology offers many advantages (pixel-based information, high dynamic range, low or no geometric distortions, high fill factor, rapid response, etc.) over the electron beam scanning vacuum tube technology[1], such as the vidicons traditionally used in TV cameras. Imaging tubes have limited resolution and are geometrically unstable due to the difficulties in precisely controlling the beam deflection.

In CMOS image sensors each pixel consists of a photodetector, and one or more transistors and capacitors depending on the design. A CMOS sensor converts the charge to voltage within each pixel (Figure 2.1) and the way in which the signal is produced and the type of signal depends on the pixel design. This difference in the readout technique between CMOS and CCD has significant implications for sensor architecture, capabilities and limitations (Blanc, 2001; Litwiller, 2001).

However, in order to characterize the performance of the different types of semiconductor sensors several criteria are used. These are explained in detail in the existing literature (e.g. Baltsavias, 1991; Beyer, 1992). Briefly summarized, they include:

1. *Dynamic range*, the ratio of the pixel's saturation level to its signal threshold. It generally refers to the ability of the sensor to detect small grey level differences even in high contrast.
2. *Responsivity*, the amount of signal the sensor delivers per unit of input optical energy.
3. *Uniformity*, the consistency of response for different pixels under identical illumination conditions.
4. *Fill factor*, the percentage of each pixel that is sensitive to light.
5. *Quantum efficiency*, the measure of the efficiency with which incident photons are detected or the ratio of the number of detected electrons to the product of the number of incident photons times the number of electrons each photon can be expected to generate.
6. *Shuttering*, the ability to start and stop exposure arbitrarily.
7. *Resolution*, described mostly by the Rayleigh criterion, the Modulation Transfer Function (MTF) or the Point Spread Function (PSF).
8. *Linearity* of the incident signal in relation to the output grey values.
9. *Signal-to-Noise Ratio*, given by the ratio of the number of the incident electrons within the integration time to the square root of the sum of the squares of each individual noise component and measured in decibel.
10. *Antiblooming*, the ability to drain localized overexposure without affecting neighboring pixels. Blooming occurs when the dynamic range of the scene is much higher than the dynamic range of the sensor. Then, an excess of electrons is generated in overexposed pixels, which is spilled to neighboring pixels, generating white spots.

[1] Imaging tube technology: Incident light forms an electric charge pattern on a photoconductive surface. Scanning by a low-velocity electron beam discharges the surface, producing a current in an adjacent conductive layer.

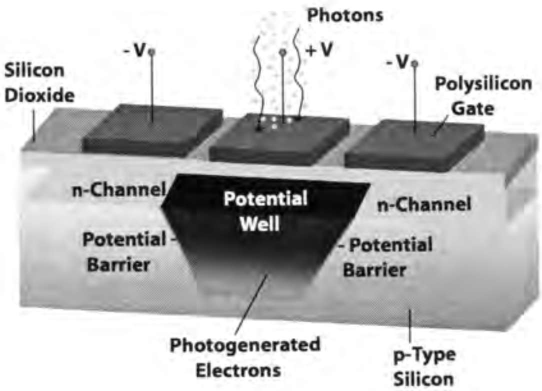


Figure 2.1: Photoelectric effect and Metal Oxide Semiconductor (MOS) capacitor. The MOS capacitor is the basis of the CCD sensor. The generated electrons from incident light are collected in the potential well until the integration time is finished.

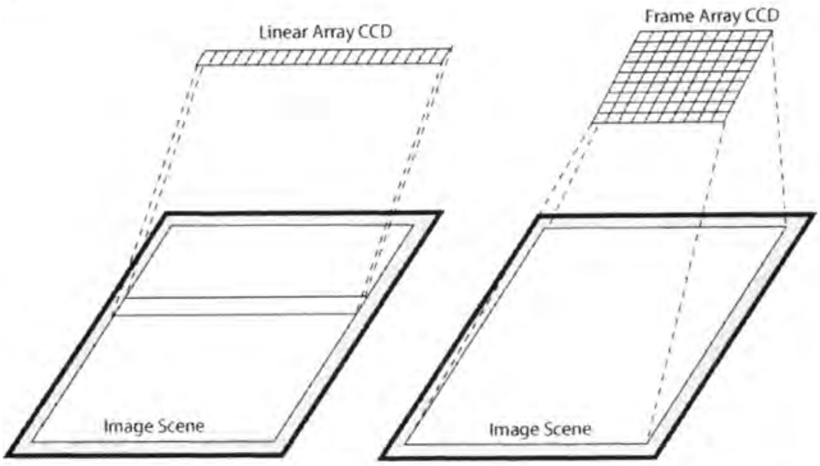


Figure 2.2: Linear and Frame Array CCDs. The linear array CCD (left) captures continuously lines of imagery, whereas the frame array CCD (right) captures the scene in a single frame image.

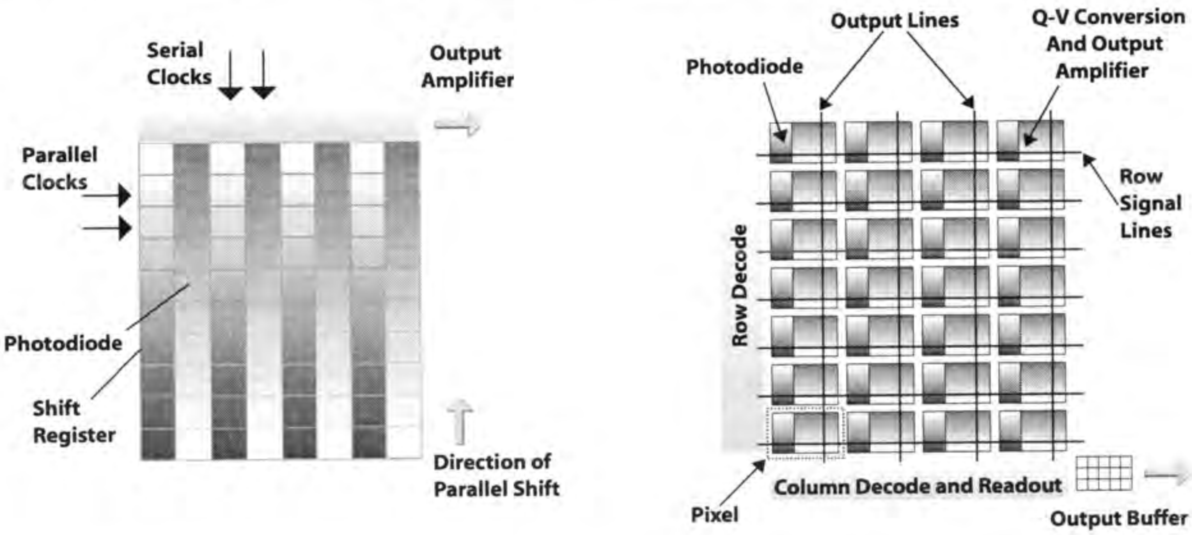


Figure 2.3: Electronic design of interline transfer CCD (left) and CMOS sensor (right).

With respect to the above criteria the advantages and disadvantages for CMOS and CCD technology are listed in Table 2.1. Although CCD image sensor technology has evolved over the past twenty five years it is difficult to manufacture when compared to CMOS image sensors. CCD technology faces many challenges among others being the difficulties in the CCD integration with complex circuits, the requirements for specialized processes for high charge transfer efficiency, the decrease of noise performance at higher speeds, the later appear as demand of the high quality video market, and most important the susceptibility to radiation damage (Diefendorff, 1998). If a defect appears in a single pixel of a CCD sensor then it interrupts the charge transfer process of the column, rendering most of the columns useless. CMOS technology allows the integration of analog-to-digital converters, on-chip clock drivers and other signal processing functions on a single sensor, whereas CCD systems require an additional chip to provide the same functions. Therefore, CMOS vs. CCD technology: (a) is more cost-efficient, (b) has a faster frame rate due to the shorter signal and power trace distances, (c) has a low power consumption and single voltage power supply, due to the functional integration, unlike the CCDs that often require 5 or more supply voltages at different clock speeds with significantly higher power consumption. CMOS imagers are also superior to CCDs in terms of responsivity, because gain elements are easier to place on a CMOS sensor. Blooming effects are also reduced, since active transistors are used for charge transfer between pixels. In addition higher dynamic range can be achieved ($>$ 120 db) by exploiting the logarithmic response of the transistor or by using a lateral overflow gate (Stoppa et al., 2002).

The drawbacks of CMOS are seen in the relatively low pixel sensitivity (low quantum efficiency and non-uniformity), the reduced calibration capabilities of the array (Kramer, 2002) and the relatively high noise levels of the small arrays of pixels. The high temporal noise with CMOS sensors occurs due to the fact that signals are transferred via multiple transistor stages and the variations in device characteristics (feature dimensions and silicon doping levels) lead to substantial fixed pattern noise. Both fixed and temporal noise is smaller with CCD sensors because the charge is transferred almost perfectly within the sensor and is passed to the outside world via a single output stage (Dileepan, 2002). Further, the lower fill factors of CMOS result in longer exposure times for the same scene as compared to CCD sensors. In general, CMOS imagers offer superior integration, power dissipation and system size at the expense of image quality (especially in low light) and therefore they have mostly used in applications where low and medium quality images suffice (e.g. consumer cameras). CMOS expect to scale better with technology not because the pixels are any better than CCDs but because more and more additional circuity can be placed in each pixel without affecting pixel size, fill factor and sensitivity (Diefendorff, 1998). In the last years significant developments in CMOS technology targeting high quantum efficiency and low dark current shot noise by using different approaches, i.e. microlenses, backside thinning, etc. (more details in Blanc (2001); Magnan (2003)) indicate a general trend to replacement of the CCD technology in a wide range of applications. However, currently CCD still remains the first choice for high-end applications, as it offers large size arrays and with minimum pixel size. The smallest pixel size for CCD and CMOS sensors, recently demonstrated, is 2.4 μm and 3.3 μm respectively (Nicolas Blanc, personal commumination).

Several techniques have been developed to produce multispectral images and these apply to both CCD and CMOS image sensors. In general, four techniques can be distinguished:

Criterion		CCD Sensor		CMOS Sensor
Dynamic Range	-	50-70 dB linear	+	50-70 dB > 120 dB logarithmic
Responsivity	-	amplification at significant power penalty	+	due to easier placement of gain elements
SNR	+	less noise due to less on-chip circuity	-	high noise levels, dark current noise increasing with long exposure times
Uniformity	+	considerably better in dark uniformity than CMOS	-	worse than CCD under normal illumination and near dark, due to variations in the offset and gain amplifier of each pixel
Shuttering	+	electronic shuttering with little fill-factor compromise	-	uniform electronic shuttering at the expense of fill factor, due to the extra transistors for the shutter
Speed	-	lower than CMOS	+	higher than CCD due to shorter power trace distances
Antiblooming	-	CCD require specific engineering to suppress blooming	+	no to less blooming occurs
Fill factor	+	high (100% for Frame Transfer CCD)	-	low (25-65%)
Quantum efficiency	+	typically higher than CMOS	-	lower than CCD
Relative power consumption	-	10	+	1
Typical voltage requirements	-	10 V	+	3 V
Smallest pixel size	+	2.4 μm	-	3.3 μm
Additional pixel functionality	-	NO	+	YES

Table 2.1: Advantages and disadvantages of CMOS vs. CCD techology.

1. Mosaic filters, such as the Bayer filter, are used in many consumer grade digital cameras such as the Kodak's DCS series and Dicomed products (small and medium format frame array cameras; more details in section 2.2). In order to capture RGB images, a mosaic of pixel-sized spectral filters is placed over the CCD or CMOS detectors (pixels). Each pixel can capture radiation from one spectral band and the values for the remaining bands are interpolated from the values of the surrounding pixels .
2. Rotating filter wheels and one image sensor array were employed in late 1990s. Due to the time delay of the wheel, postprocessing of the image was needed for the registration of the spectral bands.
3. A beam splitter together with a color separating prism, located behind the lens, allows the available image sensor arrays to capture simultaneously the corresponding spectral bands. This technique overcomes the limitations of the two previous options.
4. Multiple camera systems instead of a single camera can be used especially in the case of frame CCD

arrays. The cameras are usually arranged with parallel optical axes and with the appropriate filter, placed in front of the CCD, defining the spectral band. The final multispectral image is composed through the co-registration and fusion of the individual images recorded by each camera. Systems utilizing this technique are described in section 2.2.

All the above four techniques are used in airborne imaging (section 2.2). Apart from these, a new digital imaging technology exists, which has not been yet adopted for imaging from airborne platforms, and is employed in the X3 sensor chip by Foveon (Lyon and Hubel, 2002). The chip has three layers of CMOS arrays, corresponding to the red, green and blue channels, capturing the different colors.

2.2 Overview of digital photogrammetric cameras

CCD sensors are employed in satellite- and airborne based cameras and are used to acquire panchromatic and multispectral imagery for photogrammetric and remote sensing applications. Spaceborne CCD technology was introduced in 1980 with MSU-E on Meteor-Priroda-5 (Soviet Union). Other early CCD sensor instruments include: MOMS-01 of DLR on Shuttle flights (1983), HRV on SPOT of CNES (1986), LISS on IRS-1 series of ISRO (1988). To date, a number of spaceborne systems have been launched for which more information can be found in the existing literature (Jacobsen, 1998; Kramer, 2002; Petrie, 2003b). This section is focused on airborne cameras since it is worth describing and categorizing existing airborne digital systems in terms of architecture and characteristics.

Airborne platforms range from kites and model aircraft operating at altitudes of a few hundred feet, through conventional small single and twin-engined propeller-driven aircraft and helicopters operating at low to medium altitudes, through business jets operating at high altitudes and finally to ultra-high-flying military aircraft such as the piloted U-2 and the SR-71 or the robotic Global Hawk that can operate at altitudes of up to 70.000 feet (21 kilometers). Depending on the operating altitude of the aircraft and the constraints for images of a certain scale or ground resolution different kinds of imager can be mounted on airborne platforms. For large scale mapping the quality standards have been set by metric film cameras, e.g. Wild RC-30 from LGGM and RMK-TOP from Z/I imaging (more details in Petrie, 2003a). Assuming these cameras are able to produce images with film resolutions of 50 to 60 lp/mm (20 to 16 $\mu m/lp$) on the film plane, the equivalent pixel size is of 7.0 to 5.7 μm, given a Kell factor value of 2.8. For this comparison the following relation is used to convert the resolution of optical (film-based) to electro-optical sensors:

Width of a linepair (resolution of film-based) = 2.8 * Pixel size (resolution of electro-optical)

Moreover, if the film is scanned with 7 μm pixel size an image of 32857 x 32857 pixels will be produced for the standard format 23cm x 23cm. In airborne mapping, both main types of CCD detectors are utilized (frame and linear array CCD sensors). However, achieving simultaneously a ground coverage and a geometric resolution similar to that of metric film cameras with a single frame array CCD sensor is difficult from the manufacturing point of view, plus with larger arrays the probability of defect pixels increases. As far as geometrical modelling is concerned, frame CCD-based systems work with the standard frame camera model. On the contrary, linear array CCD systems have a more complicated imaging geometry.

which requires the application of different sensor models used in classical frame images (see also section 2.3.4). Another characteristic of these systems is that they use DGPS/INS systems on board for direct georeferencing.

2.2.1 Frame array CCD cameras

Petrie (2003a) categorizes digital frame cameras in three main types according to the size of the frame array detector, which is the factor that controls their usage in various applications:

1. Small format cameras with typical formats of 1000 x 1000 to 2000 x 3000 pixels. Developed systems are used mainly from light aircraft for rapid response, e.g. flooding or forest fires, or for crop estimation or generally agricultural applications. Since the ground coverage of each CCD image is small, a very large number of images needs to be acquired and processed to cover any substantial area of the terrain. For example the IGN camera system (Thom and Souchon, 1999), used in the Rennes survey in 1998, acquired 1068 images with ground sample distance of 30 cm over an area of 11 x 10 km with 70% and 20% along and across track overlap, respectively, due to the small size of the sensor (2k x 3k). Lutes (2002) also mentions for the DAIS-1 system of Space Imaging that in order to minimize the processing effort of a large number of images, the accuracy specifications of the data products should correspond to mapping scales of 1:4800 or smaller to permit fully automation in the processing workflow. Other multi and single camera systems have been developed either within a collaboration between industry and research institutions, e.g. ADPS (Koh et al., 1996), AEROCam (Hulst et al., 2002; AEROCam, 2004), or only by the industry, e.g. ADAR (Positive Systems, 2004), AA497 (AMDC, 2004), ISAACS (ISAACS, 2004), TerraSim-3 (TerraSim-3, 2004), Spectra-View & Agri-View (ADS Inc., 2004), AirCam (AirCam, 2004). Multispectral images can be generated with all four techniques mentioned in the previous section as can be seen from Table 2.2.
2. Medium format cameras with image formats around 4000 x 4000 pixels. The concept for most of these cameras was based on fitting the CCD array into existing camera bodies (Hasselblad, Rollei, Mamiya, Contax). During the late 1990s, with the availability of larger frame CCD arrays, the severe limitations in ground coverage of a single image were reduced. One of the first systems developed for scientific purposes was the AIMS system developed in the Center of Mapping at Ohio State University in 1995 (Toth, 1997). Its camera component consisted of a Hasselblad 553ELX camera body equipped with a Zeiss lens and a 4K by 4K pixel frame CCD array with $15\mu m$ pixel size, manufactured by Lockheed Martin Fairchild Semiconductors. The digital camera produced only panchromatic images and test results (collected data of image scale 1:6000 and GSD $\approx 10cm$) till now have demonstrated that this experimental system has fulfilled the minimum requirements for large scale mapping. However, its CCD resolution was still a constraint to process large areas and high accuracy of the georeferencing system could not be realized (more details in Toth, 1999). Based on this concept, several systems have been developed utilising digital camera backs produced by Creo Inc. (Leaf Valeo 22), Dicomed (BigShot 3000, 4000), Imacon (Ixpress), Fuji (Super CCD, Luma II), Kodak (DSC Pro Back), Linhof, PhaseOne (H25, H20) and Sinar (Sinarback 44,54). As

System (Developer)	No Lenses	Camera or Sensor chip	Sensor dimensions (pix.)	RGB method
ADPS (Geotechnologies)	1	Kodak DCS-460[a]	2036 x 3060	Bayer Filter
ADAR (Positive Systems)	1	Kodak DCS-460[a]	2036 x 3060	Bayer Filter
AA497 (SensyTech)	1	Kodak MegaPlus 4.2i	2020 x 2041	Rot. Filter Wheels
Agri-View (Airb. Data Systems)	1	Redlake MS4100	1080 x 1920	Beam Splitter & 3 CCDs
Spectra-View (Airb. Data Systems)	4	SV4S2	2048 x 2048	Interf. Filter
ISAACS (Integr. Spectronics)	1	Redlake MS4100	1080 x 1920	Beam Splitter & 3 CCDs
TerraSim-3 (STI Services)	4	N/A	2048 x 3072	Interf. Filter
DAIS-1 (Space Imaging)	4	Dalsa CA-D7-1024	1024 x 1024	Interf. Filter
AirCam (Kestrel Corp.)	4	N/A	2048 x 2048	Interf. Filter
AEROCam (Univ. of Minnesota)	4	Dalsa CA-D4	1024 x 1024	Interf. Filter
IGN camera (IGN)	3	Kodak KAF-6303[b]	2048 x 3072	Interf. Filter
	2	Kodak KAF-6300[b]	2048 x 3072	Bayer Filter
	1	Kodak KAF-16801E[b]	4096 x 4096	None (PAN)

Table 2.2: Small Format Frame Array CCD Cameras. N/A indicates that the digital camera was built by the company or institution and no name is given. ([a]) indicates that the Kodak -DCS 420 (1012 x 1024 pixels) has been also used in the system. ([b]) indicates sensor chip. All systems apart from the AEROCam and IGN camera systems have been developed for commercial purposes.

examples, the Digital Sensor System by Emerge and Applanix (DSS, 2004), the GigaCAM 22R by IGI, the Rolleimetric 6008/AIC by Rollei, the HeliMap by EPFL and Riegl (HeliMap, 2004) and the Medium Format Digital Mapping Camera by Geotechnologies are quoted here (Table 2.3). The only system in this category that can combine up to four medium format camera modules (5440 x 4080 pixels each)is the Digital Modular Aerial Camera system (DiMAC, 2004) by AeroPhoto that was introduced in the market in 2004. The system can be configured with different focal lengths and viewing angles, based on the user requirements for orthophoto and DTM production. Various medium format frame cameras have also been designed to complement airborne laser scanners [2] and so form a complete mapping system.

3. Large format cameras having a format of larger than 6000 x 6000 pixels. One can distinguish two main directions in current developments. Multiple medium format cameras that are synthesized into a single large-format image, and single cameras fitted with very large CCD frame arrays (Table 2.4). In the first category belong the Digital Modular Camera System (DMCS) from Z/I Imaging and the UltraCam-D from Vexcel Austria. The DMCS (Hinz, 1999; Heier and Hinz, 2002) comprises

[2] e.g. DATIS, RAMS, Earthdata systems

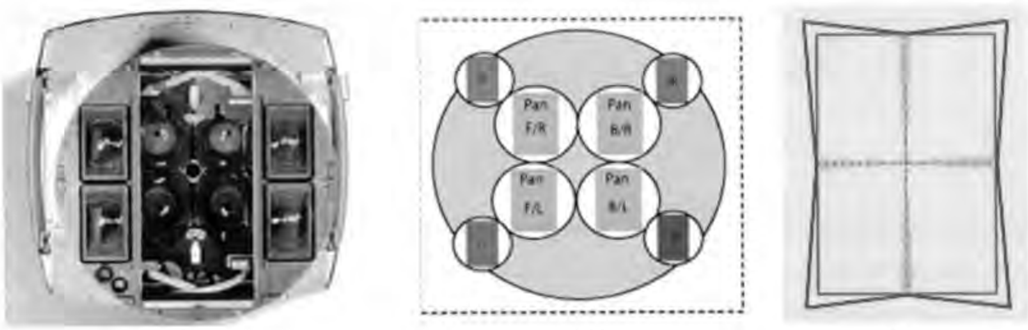
System	(Developer)	No Lenses	Camera, Sensor chip	Sensor dimensions (pix.)	RGB method
AIMS	(CFM at OSU)	1	Hasselblad 553ELX, LMF Semiconductor	4096 x 4096	None (PAN)
MF-DMC	(Geotechnologies)	1	Hasselblad 555ELD, Kodak ProBack	4080 x 4080	None (PAN)
DSS	(Emerge Applanix)	1	Contax 645, Kodak MegaVision	4092 x 4079	Bayer Filter
GigaCAM 22R	(IGI)	1	Rollei	4096 x 4096	None (PAN)
Rolleimetric 6008	(Rollei)	1	Rollei 6008, PhaseOne H20	4080 x 4080	Bayer Filter
HeliMap	(EPFL - Riegl)	1	Hasselblad SWCE 903, Kodak ProBack	4080 x 4080	None (PAN)
DiMAC	(Aerophoto)	1-4	PhaseOne, Kodak True Color	5440 x 4080	Bayer Filter

Table 2.3: Medium Format Frame Array CCD Cameras.

two camera systems, one to produce panchromatic images and the second multispectral images of lower ground resolution (Figure 2.4(a), 2.4(b)). Four individual panchromatic cameras are arranged in star-type configuration, tilted outwards, with small overlap. Images are simultaneously exposed and rectified to form a single perspective or the so-called *virtual image* (Figure 2.4(c)). As images are simultaneously exposed, the offset of the projection centers causes deformations to the *virtual image* (Tang et al., 2000), which are influenced by height differences in object space and by the flying height above ground, given that the rectification utilizes the projective transformation to a planar surface instead of being a more accurate differential rectification. The size of each of the four frame array CCDs is 7k x 4k pixels. The true or false colour image is formed through co-registration and fusion of the multispectral channels (2k x 3k), which are arranged with parallel optical axes (nadir viewing angles) and with the appropriate filters defining the spectral bands. Similarly to the DMCS, UltraCam-D (Leberl et al., 2003) comprises two camera systems of panchromatic and multispectral sensors, but with a different configuration; see Figure 2.5. Instead of the four tilted cameras of the DMCS, the UltraCam-D has the four cameras in-line pointing in the flight direction and with parallel optical axes. The advantage over the DMCS is that errors arising from the mosaicking of the sub-images are minimized since the panchromatic sub-images are acquired almost from the same exposure station, due to a time delay of $1 - 2msec$. In the second category, systems with a single CCD frame array were developed initially for military purposes. A 9.2k x 9.2k pixel CCD with $8.75\mu m$ pixel size is used in the Ultra High Resolution Reconnaissance Camera of BAE Systems (Gorin et al., 2002) and a 10k x 5k pixel CCD with $10\mu m$ pixel size is used in the CA-260/50 by Recon/Optical.

System (Developer)	No Lenses	Sensor dimensions (pix.)	Type	RGB method
DMCS (Z/I Imaging)	4	7680 x 13824*	PAN	-
	4	2000 x 3000	MS	Interf. Filter
UltraCam-D (Vexcel)	4	7500 x 11500*	PAN	-
	4	2672 x 4008	MS	Interf. Filter
UHRRC (BAE)	1	9216 x 9216	PAN	-
CA-260/50 (Recon/Optical)	1	5000 x 10000	PAN	-

Table 2.4: Large Format Frame Array CCD Cameras. (*) indicates the dimensions of the final composite image.



(a) (b) (c)

Figure 2.4: DMCS configuration. (a), (b) arrangement of the eight cameras. (c) the ground coverage of the four tilted pan cameras marked with red, the overlapping areas marked with yellow and the ground coverage of the virtual image marked with blue (Source: Z/I Imaging).

(a) (b)

Figure 2.5: UltraCam-D configuration. (a) The multiple frame camera unit with eight cones, four of them located in-line to produce the large format pan image. (b)Acquisition concept and ground coverage with overlapping areas of the sub-images (Source: Vexcel).

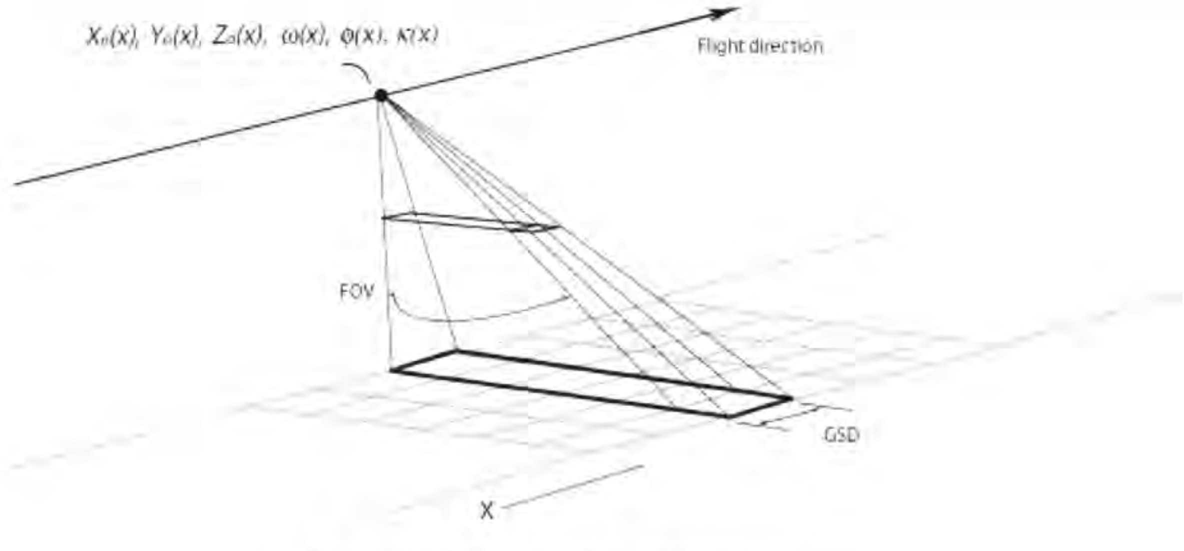


Figure 2.6: Schematic of linear pushbroom sensor.

2.2.2 Linear array CCD cameras

Linear CCD sensors have been mostly operated on spaceborne platforms. Space provides a non-turbulent medium for the orbiting platform. From a photogrammetric data processing point of view, the use of linear array CCD sensors has a higher degree of complexity since the orientation of every image line requires robust modelling of the data acquisition trajectory. This is due to the fact that unpredictable platform motions, caused by the turbulence of the Earth's atmosphere, result in large image distortions in the linescan imagery. However, with the development of modern DGPS/INS systems to provide positional and attitude information on a line-by-line basis for georeferencing the above mentioned drawback is resolved and the flight trajectory can be accurately acquired. Most of existing airborne linear array CCD sensors utilize the pushbroom scanning technique (imaging occurs in parallel for all line detector cells in the cross-track direction, Figure 2.6) with three or more CCD lines on the focal plane for simultaneous stereo acquisition.

Initially, the idea to substitute film-based data acquisition by linear array CCDs was pioneered by Dr. Otto Hofmann in the '70s. The first sensor with stereo capabilities to be developed for experimental purposes was EOS (Electro Optical System). Since then, the airborne systems that have been developed either for experimental or commercial purposes are the HRSC-A, WAAC, DPA, TLS and ADS40. The latter is described in detail in section 2.3. DLR (German Aerospace Agency) developed the High Resolution Stereo Camera (HRSC) and the Wide Angle Optoelectronic Stereo Scanner (WAOSS) for the exploration of Mars in 1996 (Albertz, 1996; Hauber et al., 1996). Their airborne derivates were the HRSC-A (Neukum, 1999), later HRSC-AX and HRSC-AXW, and WAAC (Sandau and Eckardt, 1996) respectively. Characteristics of these sensor systems are summarized as follows:

- HRSC-A comprises nine CCD lines, which are mounted in parallel behind a single lens. Since 1997 the sensor has been used in several airborne experiments and applications in telecommunications (Renouard and Lehmann, 1999), volcanology (Gwinner et al., 1999), forestry (Hese and Lehmann, 2000) and for commercial DSM production over urban areas (Scholten et al., 2003). In parallel,

HRSC	-A	-AX	-AXW
focal length	175 mm	151 mm	47 mm
FOV [along x across]	38° x 12°	41° x 29°	30° x 79°
No. CCD lines	5 PAN, 4 MS	5 PAN, 4 MS	3 PAN, 2 MS
pixels per line	5184	12000	12000
pixel size	7 μm	6.5 μm	6.5 μm
radiometric resolution	8 bit	12 bit	12 bit
PAN spectral range	585-765 nm	520-760 nm	515-750 nm
RED spectral range	730-770 nm	635-685 nm	570-680 nm
GREEN spectral range	485-575 nm	530-570 nm	475-575 nm
BLUE spectral range	395-485 nm	450-510 nm	-
NIR spectral range	925-1015 nm	770-810 nm	-
PAN stereo angles	$\pm18.9°, \pm12.8°, 0°$	$\pm20.5°, \pm12.0°, 0°$	$\pm14.4°, 0°$
MS stereo angles[a]	$+15.9°, -3.3°, +3.3°, -15.9°$	$+2.3°, -2.3°, +4.6°, -4.6°$	$+7.1°, -7.1°$

Table 2.5: HRSC-A/ -AX/ -AXW technical data. ([a]) The order of MS channels is Red, Green, Blue, Near infrared.

	WAAC	DPA	SI100	SI250
focal length	21.7 mm	80 mm PAN 40 mm MS	60 mm	65 mm
FOV [along x across]	50° x 80°	50° x 74°	42° x 61.5°	51° x 58°
No. CCD lines	3 PAN	2 x 3 PAN 4 MS	9 MS	10 MS
pixels per line	5184	2*6000 PAN 6000 MS	10200	14400
pixel size	7 μm	10 μm	7 μm	5 μm
radiometric resolution	8 bit	8 bit	11 bit	11 bit
PAN spectral range	470-670 nm (NAD) 580-770 nm (BW, FW)	515-780 nm	-	-
RED spectral range	-	610-685 nm	598-700 nm ([a])	598-700 nm ([a])
GREEN spectral range	-	520-600 nm	506-577 nm	506-577 nm
BLUE spectral range	-	440-525 nm	413-505 nm	413-505 nm
NIR spectral range	-	770-890 nm	-	760-860 nm
PAN stereo angles	$\pm25.0°, 0°$	$\pm25.0°, 0°$	-	-
MS stereo angles	-	0°	$\pm21.0°, 0°$	$-30.0°, 0°, +21°$

Table 2.6: WAAC, DPA and STARIMAGER® technical data technical data. According to the technical specifications of KLI sensors (Kodak, 1999), the spectral ranges for the SI100 and SI250 are given.([a]) The red filter exhibits a high level transmission beyond 700 nm (filter becomes transparent).

several tests for the geometric validation of the sensor have been performed (Wewel et al., 1998; Wewel and Brand, 1999). Based on the HRSC-A, a follow-on project of DLR that started in 1998 has resulted in the development of the narrow angle HRSC-AX and the wide angle HRSC-AXW imagers (Neukum et al., 2001). The design, and geometric and radiometric characteristics of these new versions are listed in Table 2.5.

- WAAC was developed in 1993 in order to test and verify the processing algorithms of the WAOSS camera for airborne applications. In contrast to HRSC-A, WAAC captures only three panchromatic images and has a larger FOV (image products of 1m ground resolution from an altitude of 3000m). In Table 2.6 the technical parameters of the system are listed.
- The Digital Photogrammetry Assembly (DPA) was built under the initiative of the German Ministry of Defense and in 1995 the Institute of Photogrammetry at the University of Stuttgart began to evaluate the system. The purpose of DPA (Fritsch, 1997; Hofmann et al., 1993) was directed to topographic mapping at scales of 1:25000 and 1:50000. The DPA system utilized two lenses with three panchromatic CCD lines each and four additional lenses for the multispectral channels. The concept of the double lens for the stereo channels aimed to increase the size of the three stereo channels by a factor of two. In total, ten CCD line detectors of 6000 pixels were used. The number of pixels per image line is different for the panchromatic and multispectral modules, however the field of view is kept constant due to the different focal lengths of the two camera types. The system was mainly used for research and evaluation purposes.
- The helicopter-based Three-Line-Scanner (TLS) system was developed by Core Co., Japan, in cooperation with the Murai Lab of the University of Tokyo (Murai et al., 1995). The system has evolved into a commercial product, managed by Starlabo Corp., Japan, under the name STARIMAGER®. Three versions of the system exist to date (SI100, SI250 and SI290), both aiming at high-resolution images of 5-10 cm ground resolution from 500m flying altitude, acquired by a three by three configuration of RGB CCD lines (Kodak KLI series) on the focal plane of a single lens (for SI250 one near infrared line has been added between Nadir and Backward). This set results to that the individual linear arrays are offset from each other and the lines may have different attitude values due to the different acquisition time. This further requires that the scanner is mounted on a very stable platform. A stabilizing platform with a highest attitude angle resolution of 1 sec is employed in the system to minimize distortions in the raw images. In Table 2.6 the SI100 and SI250 versions are listed. The SI290 is based on the SI250 version but with a different focal length (93 mm) and a different configuration of viewing angles ($-23.0°, 0°, +15°$ degrees for Backward, Nadir, Forward lines respectively). Chen et al. (2003) reports accuracy results for ground point determination of 0.03-0.08 m in planimetry and 0.06-0.15 m in height. Similarly the accuracies that are presented by Grün and Zhang (2003) are in the range of 0.05-0.06 m in planimetry and of 0.08-0.10 m in height. TLS has been used for different applications in GIS, 3D city modelling, generation of DSMs and orthophotos (Grün and Zhang, 2002; Grün et al., 2003; Murai et al., 2003).
- Similarly to the TLS system, the DAS-1 system by Wehrli Ass., USA, uses RGB CCD lines in a three by three configuration (Kodak KLI-8023) with $9\mu m$ pixel size (DAS-1, 2004). The viewing angles of the Backward and the Forward lines are $-16.0°$ and $+16°$ degrees respectively. The system

was introduced in the market in 2004 and to date the system has not been evaluated thoroughly. As one can observe, several systems, with similar or different architecture focused on different mapping fields have been developed. As far as large scale mapping is concerned, the employment of larger CCDs is favorable due to the reduction of data volumes. The utilization of either large frame or linear array CCDs has both advantages and disadvantages, especially when high accuracy standards must be fulfilled (see Table 2.7).

Parameter		Frame Array CCD Sensor		Linear Array CCD Sensor
Technical Parameters	- - -	long integration time (e.g. Philips CCD) mechanical shutter is usually required larger probability of defect pixels	+ + +	short integration time (e.g. Fairchild CCD) electronic shutter smaller probability of defect pixels
Orientation	+	per frame	-	per line
Stereo capability	- - -	perspective distortions complicated implementation of simultaneous stereo from different CCDs (not realized) typically 60% overlap	+ + +	smaller perspective distortions (parallel projection) easier implementation of different stereo angles 100% overlap
Spectral resolution	- -	multiple cameras, difficult calibration typically resolution of spectral channels lower than pan channels (Ultracam-D, DMCS)	+ +	single camera for spectral and pan lines similar for all channels (excl. DPA)
Swath width	-	smaller than linear for a single frame CCD	+	larger than frame CCDs
Data volume	-	each scene in many image files (difficult handling of data)	+	each scene in a single image file (easier handling of data)
Hardware	-	complicated implementation, expensive read out for real-time corrections (FMC is required)	+	simple implementation
Software	+	existing photogrammetric software can be used	-	require special processing software

Table 2.7: Comparison between Frame and Linear Array CCD Sensors.

2.3 ADS40

The ADS40 camera system is described in more detail in this section. The design of the sensor concerning the geometrical configuration of CCD lines on the focal plate, the employment of staggered lines, the selection of the spectral channels and the optical system are reviewed. The radiometric and geometric parameters of the system are investigated in the later section on calibration, followed by the description of the sensor model. Moreover, aspects of its implementation are mentioned. Last but not least, the basic steps in ground processing are explained.

2.3.1 Design

ADS40 is the Airborne Digital Sensor (Sandau et al., 2000; Tempelmann et al., 2000; Fricker et al., 2000; Fricker, 2001) developed through a cooperation of Leica Geosystems (former LH Systems) and DLR. It is among the sensors that employ multi-line CCD technology in conjunction with GPS and INS technology (Applanix) for sensor orientation and new developments in sensor technology, optics, electronics, data transfer and storage. The camera system consists of ten parallel sensor lines in the focal plane of a single lens system: panchromatic, red, green, and blue lines, along with one infrared line, each with 12000 pixels of $6\mu m$ pixel size. Eckardt (2002) in his dissertation describes in detail several aspects that were taken into account in the design of the ADS40 system. Among them, methods have been investigated to increase image dimension across track by a factor of two, in order the image size to be similar to the dimension of scanned analog imagery (23000^2 or 32857^2 image size, assuming $10\mu m$ or $7\mu m$ pixel size respectively, see also section 2.2). The method to compose one 24k line from the two 12k lines by optical means (concept similar to DPA), was discarded because a) the expected registration accuracy of 1/3 pixel could not be guaranteed and b) the calibration of two lenses (instead of one) was more complicated. Out of the three alternatives illustrated in Figure 2.7, the concept of staggered lines (lines shifted with respect to each other by 0.5 pixels) was selected and used (more information in Eckardt (2002)). The concept of staggered lines does not double the size of the acquired image line but gives two times better geometric resolution of $3.25\mu m$ and an increase in radiometric resolution by 2 bits while avoiding aliasing effects. These were considered advantageous for the use of staggered lines. The parallel configuration of CCD lines on the focal plane was chosen so that the 12k CCD line could operate either in single or staggered mode. The processing of staggered arrays and the proposed method for restoration and generation of the "new"image is described in Reulke et al. (2004).

To date four different sensor plate designs for the ADS40 have been built, each with a different configuration of CCD lines. In Figure 2.8 the configuration of the CCD lines is shown and in Table 2.8 the viewing angles are given for each focal plate design (FP1, FP2, FP3, FP4, where FP is used as abbreviation of Focal Plate). Usually the forward panchromatic CCD lines are rotated 180° degrees due to mounting restrictions of the CCD lines in the focal plate. FP2 is used by Pasco Corp., Japan in one of their three ADS40 systems and FP3 is employed in the system that has been delivered to Northwest Geomatics, USA. The standard and most known configuration applies to the FP1 design with staggered panchromatic lines in the backward, the forward and the nadir position, the multispectral red, green and blue lines between the forward and the nadir lines and the near infrared close to nadir. An RGB triplet of CCDs in the nadir is preferred for true orthophoto generation (FP2 and FP4 versions). Positional errors due to the relief, which are apparent in non-vertical images, are at a minimum in the nadir view. The acquisition of the red, green and near infrared images with similar viewing angles (FP3 and FP4), translates into small time differences in capturing the same surface area. This improves co-registration of the images when they are merged to form a false color image[3]. Artifacts occurring through moving objects due to the time delay in acquisition of the separate channels are minimized (the FP1, FP2 are not optimal therefore for false color image). The forward looking view of these three lines has been employed

[3]The FP3 design was used by Northwest Geomatics to generate true- and false-color orthoimages for updating US-statewide the Digital Ortho Quarter Quads (DOQQs)

Figure 2.7: Alternative concepts to staggered lines. Versions illustrated in left and middle image were disregarded. The adopted version of staggered lines is illustrated in the right image.

(a) (b) (c) (d)

Figure 2.8: ADS40 focal plate layouts. (a) FP1, (b) FP2, (c) FP3 and (d) FP4.

CCD lines	FP1	FP2	FP3	FP4
PAN	$+28°, 0°, -14°$	$+28°, +16°, -14°$	$+28°, -14°$	$+28°, -14°$
RED	$+16°$	$0°$	$+16°, +2°$	$+14°, 0°$
GREEN	$+16°$	$0°$	$+16°, 0°$	$+16°, 0°$
BLUE	$+16°$	$0°$	$+16°$	$0°$
NIR	$-2°$	$+14°$	$-2°$	$+18°$

Table 2.8: Viewing angles of the CCDs for the different ADS40 focal plate designs.

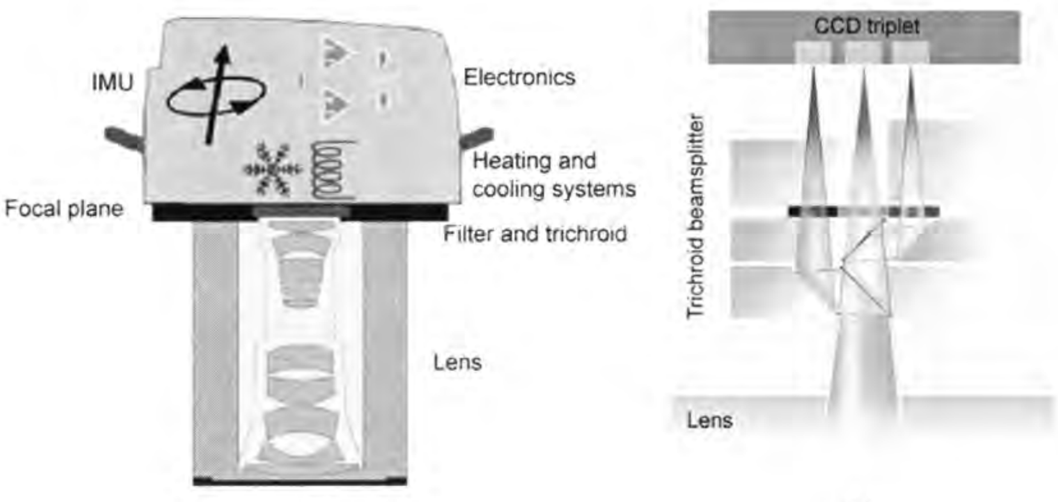
in FP4 for visual purposes, i.e. better perspective view of the terrain. Moreover, the asymmetrical configuration of viewing angles offers more options for stereo and for incorporating views with large incident angles (large base between sensor positions), thus avoiding singularities in the mathematical model of adjustment (Müller , 1991).

The selection of the spectral channels has been based on requirements derived from land use applications, surface observation and monitoring of water and vegetation areas. For example, the blue channel was included as it could receive information from shadowed areas, due to the high scattering of particles in its wavelength range; information on water quality and algae are acquired through combination of the green and blue channels; and red and near infrared channels can be used to discriminate vegetation and monitor temporal changes in the biosphere (absorbed active radiation, crop yield estimation etc.). Each CCD sensor applies to a certain wavelength range within the spectral region from 400 nm to 1000 nm (see Table 2.9). In order to avoid reduction of the surface signal due to atmospheric absorption, the channels are positioned within atmospheric windows, namely regions where only minor absorption by the atmospheric constituents occurs (Reulke et al., 2000).

Band	Spectral range(nm)
Panchromatic	465-680
Red	610-660
Green	535-585
Blue	430-490
NIR	835-885

Table 2.9: Spectral characteristics of ADS40 CCD lines.

The optical system of the ADS40 consists of a telecentric lens of 62.5 mm focal length and a trichroid beam splitter (a cross-section of the camera and the trichroid beam splitter are illustrated in Figures 2.9(a) and 2.9(b) respectively). The requirement for a telecentric lens resulted from the use of interference filters. The filters, which are placed on a glass in front of the CCD, should not change the spectral response across the wide angular field of view. The light should impinge on the focal plate with right (90°) angles in order to achieve for each adjacent line of pixels illumination by light of common angular incidence and avoid wavelength shifts. The telecentric lens is able to compensate wavelength shifts that occur due to large incident angles, and guarantees ideal transmission. The radial distortion is about 3%. The trichroid beam splitter divides the light into different color components with small loss of energy (average transmittance > 80%) compared to classical beam splitters or rotating mirrors, and it reduces color line offsets from 2 degrees to 1 arc minute. As a result, color offsets especially visible on building edges are reduced and R, G and B channels are co-registered. In the ADS40, systems that ensure pressure- and temperature stabilization while compensating for condensation effects are employed.



(a) (b)

Figure 2.9: Cross-section of camera (a) and trichroid beamsplitter (b)(Source: Leica Geosystems).

2.3.2 Calibration

For the verification of the radiometric and geometric characteristics of the ADS40 sensor a radiometric (laboratory) and a geometric calibration (either in the laboratory or by photogrammetric means) are performed. In this section the laboratory calibration is handled and the refinement of the calibration parameters by means of self-calibration is discussed in section 2.3.4.3.

2.3.2.1 Radiometric Calibration

Regarding radiometric calibration, the calibration sphere analyzes the radiometric accuracy of the ADS40, with an accuracy of 1% (approximately 7 bit). This resolution is related to the homogeneity of the measured radiation on the aperture plane as well as to the absolute radiance value. The integrating sphere provides a highly lambertian surface, therefore any radiation emitted in the sphere's interior generates a uniform illuminance through multiple reflections. A detector is used to measure the radiant power, generated by the halogen lamps. When the illumination level varies, the linear array's response is measured, thus the responsivity, linearity and dynamic range can be characterized. The calibration procedure is divided into relative and absolute radiometric calibration. The first aims at maintaining data quality and removing system artifacts at the recording stage of raw imagery (Lev0, see also section 2.3.3), whereas the second aims in adapting each band to the spectral properties of the human eye at the postprocessing stage (Lev1 images).

In the relative radiometric calibration the following corrections are determined:

- Dark Signal Non-Uniformity (DSNU) correction. The digital response DN for each pixel without light input is measured once during radiometric calibration. The normalized signal is uploaded into the camera head and used as a weighting for the dark current subtraction.
- Dark Current Subtraction. The dark current is determined by averaging over 3 to 5 dark pixels on each side of the sensor line.
- Photo Response Non-Uniformity (PRNU) correction. The different gain and lens vignetting for each pixel is determined by registering the complete line at a bright level which is still in the linear response level of the CCD. A normalization factor for each pixel is calculated by dividing the brightest pixel to each pixel and is uploaded into the camera head.
- Normalization procedure in the compression mode. In order to shift 16 bit data into an 8 bit range a gain-offset transformation is applied to the signal. For sudden bright objects in the image signal clipping is applied.

Eckardt (2002) also gives details of the methods used to determine the DSNU and PRNU, the additive and multiplicative corrections respectively. The computed corrections are stored in the camera head by means of look up tables and are applied in real time, after digitization of the analog signal (A/D conversion 14 bit). PRNU corrections are applied after DSNU correction and dark current subtraction. The remaining PRNU effects after the corrections are smaller than 1 grey value (Börner and Reulke, 2001).

In the absolute calibration, since the DSNU and PRNU have corrected for line uniformity in the previous calibration stage, radiometric coefficients, derived using Equation 2.1 (see also Mendenhall et al. (1999) for more details), are applied as spectral weights to account for sensitivity of the detector and there can be used when a color or a pansharpened image is generated.

$$G_i = \frac{L_i I}{DN} \tag{2.1}$$

where $L_i(Wm^{-2}sr^{-1}\mu m^{-1})$ is the averaged spectral radiance for band i, DN the recorded digital number for integration time $I(s)$ and $G_i(Wsm^{-2}sr^{-1}DN^{-1}\mu m^{-1})$ is the derived radiometric coefficient for band i. The band averaged spectral radiance of the source radiation reaching each sensor band is calculated according to Equation 2.2.

$$L_i = \int L(\lambda)R(\lambda)d\lambda \tag{2.2}$$

where L_i the averaged spectral radiance for band i, $L(\lambda)$ the spectral radiance of the calibration sphere, $R(\lambda)$ the normalized spectral response and λ the wavelength. Relative spectral response functions are calculated using the CCD responsitivity, the lens transmission, the beam-splitter and the filter transmission specifications. From the above method, the extracted radiometric coefficients are then used to compute absolute spectral response functions, whose integral corresponds to the measured sensor response. Figure 2.10 shows the radiance curve of the calibration sphere and the wavelength range of the bands. In table 2.10 the digital response, the integration time and the radiometric coefficients of each band (Udo Tempelmann and Ullrich Beisl, personal communication) are illustrated.

2.3.2.2 Geometric Calibration

Regarding geometric calibration, the ADS40 uses the same set of interior parameters as the classical frame camera: principle distance (c), principal point of auto-collimation (PPA) and radial-symmetric distortion. Also, the estimation of the principal point of symmetry (PPS) and polynomial coefficients for the even powers of radius (r^2, r^4, r^6), for the non-symmetrical radial and tangential distortion, need to be estimated, as large distortions have to be accepted due to the telecentricity of the lens system. Additional offsets and distortions result from the composition of the focal plate from several CCD lines, beam splitter and filter assemblies. For the laboratory calibration complex and costly equipments are generally used since the accuracy of camera calibration depends on the quality of known geometry of targets being viewed from the camera. For ADS40 (see also Cramer, 2004), the laboratory geometric calibration is performed using the Coded Vertical Goniometer (CVG), illustrated in Figure 2.11(a). The CVG was developed in 1998 by upgrading an existing vertical goniometer used for film-based analogue lenses. The upgrade consisted of hardware and software modifications which aimed in calibration of the lens cone and the digital camera head and the calibration facilities are described in detail in Pacey et al. (1999). The camera is mounted in the goniometer and an inverted light path projects two different code patterns on the CCD lines. The code patterns consists of a series of stochastically arranged black and white bars of width 3 mm pointing in two orthogonal directions and one of the code patterns is shown in Figure 2.11(b). In order to allow measurements in off-nadir lines an additional mirror scanner is mounted on top of the goniometer arm.

The CVG measures MTF, best image plane and derives approximate values of distortions. To

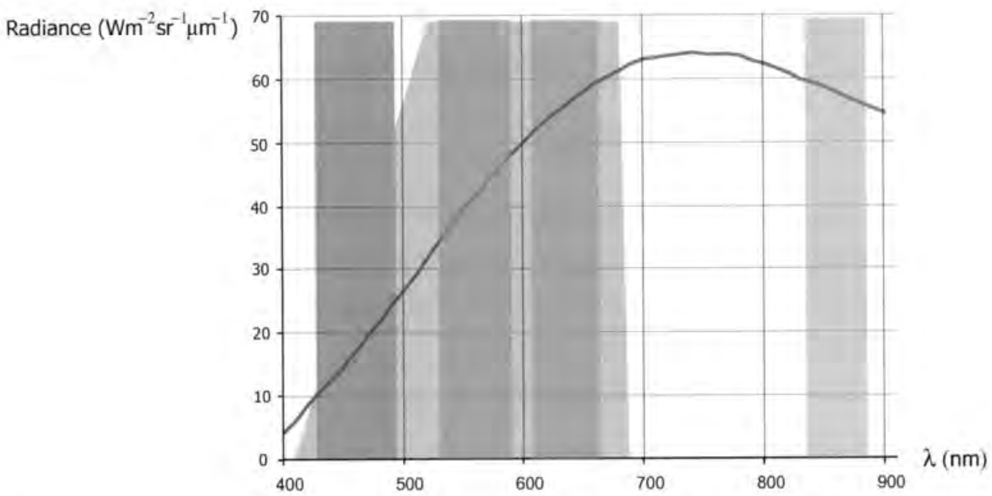
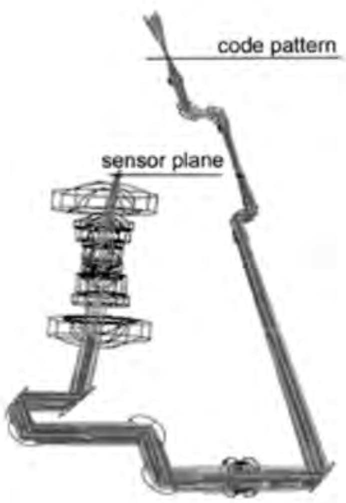


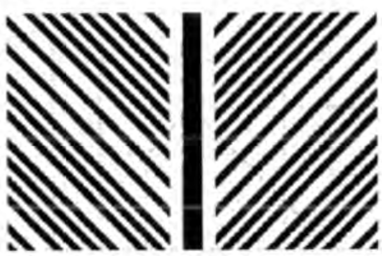
Figure 2.10: Plot of integrating sphere's radiance versus wavelength (red curve). The positions of the bands in the spectral range are indicated with the different colors corresponding to the separate channels.

Band	Digital Response (DN)	Integration time (s)	Radiometric coefficient G_i
Pan	8500	0.0075	$3.969 \cdot 10^{-5}$
Red	3500	0.0175	$2.809 \cdot 10^{-4}$
Green	2300	0.0175	$3.178 \cdot 10^{-4}$
Blue	1100	0.0175	$2.875 \cdot 10^{-4}$
NIR	1100	0.0175	$1.407 \cdot 10^{-4}$

Table 2.10: Digital response, integration time and radiometric coefficients of ADS40 bands for a specific calibration.



(a)



(b)

Figure 2.11: Coded Vertical Goniometer (a) and code pattern (b) (Source: Leica Geosystems).

each measured pixel two polar angular co-ordinates are assigned, though according to Schuster and Braunecker (2000) it is sufficient to measure pixels every 2-5 degrees within the field of view and interpolate the values of the intermediate pixels. The approximate value of the principal point is computed, after the best image plane has been determined, with polynomial approximation based on more than 23 points per line. However, the CVG does not determine the principal point and its precision potential is only about 1 pixel. The σ_0 a-posteriori (standard deviation of unit weight a-posteriori) derived from the CVG measurement is $6.5\mu m$ (Tempelmann et al., 2003). Due to the above limitations, the values derived from the calibration in the laboratory are used as approximations for the later calibration by photogrammetric means over a test field (section 2.3.4.3).

2.3.3 Sensor model

Apart from the hardware aspects regarding system components (CCD, optics, cooling system, GPS and IMU system etc.) that were discussed in sections 2.3.1 and 2.3.2, the mathematical model, formulated to derive 3D information, has to be considered as well. As the trajectory of the sensor is not easily predictable and not smooth, the orientation of each line has to be used in the mathematical model. A strict sensor model is required, taking into account the calibration parameters and the orientation and attitude information of each image line. The sensor model for ADS40 has been developed and implemented by LGGM and here a short description of the coordinate systems and the transformations from object to image space and backwards is given. Moreover, aspects related to accuracy and speed of transformations have been investigated for the processed datasets and several problems are outlined below.

The classical image space becomes the focal plane coordinate system (FPS), while the acquired image is a series of single line images with the same line position on the focal plane system. The focal plane coordinate system is the physical position of the pixels in the focal plane. The x axis of the FPS is perpendicular to the CCD lines and 180° degrees rotated relative to flight direction and the y axis is perpendicular to x (Figure 2.12). The origin of FPS is the principal point (PPA) and the rotational reference is established as the best fit of three panchromatic stereo lines (Forward, Backward and Nadir). For the recorded raw imagery (Lev0)(see also section 2.3.4) each pixel is described with a line (y_{L0}) and position (x_{L0}) value, as illustrated in Figure 2.13(a). The Lev0 image coordinate system has its origin in the upper left corner of the first pixel in the first scan line. The X axis is aligned with the CCD line, where x_{L0} is the position of the pixel in the scanline (from 0 to (number of calibrated pixels - 1)) and the y_{L0} is the index of the scan line in the image. The collinearity equations are introduced for each line on Lev0 to relate Lev0 image- and object coordinate system.

$$
\begin{aligned}
x_f &= -c \cdot \frac{r_{11} \cdot (X_G - X_o) + r_{12} \cdot (Y_G - Y_o) + r_{13} \cdot (Z_G - Z_o)}{r_{31} \cdot (X_G - X_o) + r_{32} \cdot (Y_G - Y_o) + r_{33} \cdot (Z_G - Z_o)} \\
y_f &= -c \cdot \frac{r_{21} \cdot (X_G - X_o) + r_{22} \cdot (Y_G - Y_o) + r_{23} \cdot (Z_G - Z_o)}{r_{31} \cdot (X_G - X_o) + r_{32} \cdot (Y_G - Y_o) + r_{33} \cdot (Z_G - Z_o)}
\end{aligned}
\tag{2.3}
$$

In Equation 2.3 the left-hand side elements x_f, y_f refer to the focal plane coordinates (FPS), while the

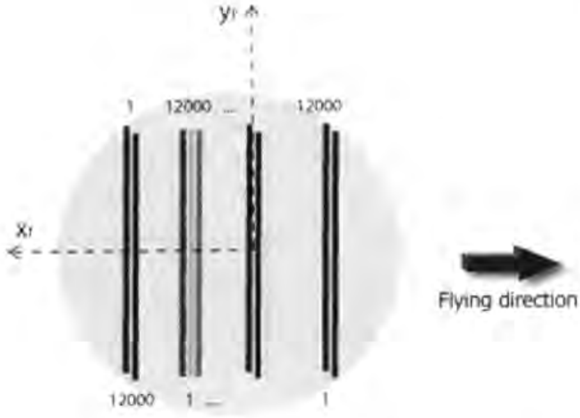


Figure 2.12: Focal plane coordinate system of ADS40. As an example, the FP1 sensor plate is illustrated. The numbering of the pixels on the CCD is from bottom to top (1...12000), except for the forward staggered lines which are rotated by 180° degrees on the focal plate.

rotation matrix elements (r_{kk}, $k = 1\ldots3$) refer to the transformation from object space to focal plane coordinate system, the X_o, Y_o, Z_o to the position of the line with respect to the object coordinate system and the X_G, Y_G, Z_G are the object coordinates. To retrieve the pixel position on the scan line from known focal plane coordinates (x_f, y_f), the pixel coordinates are interpolated from the coordinates of the pixels on the CCD with respect to the FPS system, obtained from the calibration file[4]. The sequence of transformations from Lev0 to object coordinate system is (Figure 2.13(a)):

Lev0 image (pixel) → FPS → Object space

For example, in order to obtain the planimetric coordinates of the object point P_A (X_G^A, Y_G^A) for which the height component (Z_G^A), and the point position on Lev0 (x_{L0}^A, y_{L0}^A) are known, first the respective point position on the FPS (x_f^A, y_f^A) is computed. The solution for X_G^A and Y_G^A is straightforward, using the inverse form of collinearity equations (Equation 2.4). Assuming subpixel accuracy of the point position on Lev0, namely y_{L0}^A as a non integer number, the orientation elements $(X_o, Y_o, Z_o, r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33})$ are computed by linear interpolation from the orientation elements of the two nearest lines.

$$
\begin{aligned}
X_G^A &= X_o + (Z_G^A - Zo) \cdot \frac{r_{11} \cdot x_f^A + r_{21} \cdot y_f^A + r_{31} \cdot (-c)}{r_{13} \cdot x_f^A + r_{23} \cdot y_f^A + r_{33} \cdot (-c)} \\
Y_G^A &= Y_o + (Z_G^A - Zo) \cdot \frac{r_{12} \cdot x_f^A + r_{22} \cdot y_f^A + r_{32} \cdot (-c)}{r_{13} \cdot x_f^A + r_{23} \cdot y_f^A + r_{33} \cdot (-c)}
\end{aligned}
\tag{2.4}
$$

For the inverse transformation from the object coordinate system to FPS, an iterative approach is utilized to determine the corresponding scan line and to obtain the orientation parameters that are used in the collinearity equations. The average ground height, three image points in triangular distribution

[4]The calibration file for each sensor line contains the nominal focal length, the x and y coordinate for the center of each pixel on the CCD with respect to the focal plate coordinate (FPS) system and the IMU misalignment (description in section 2.3.4.3)

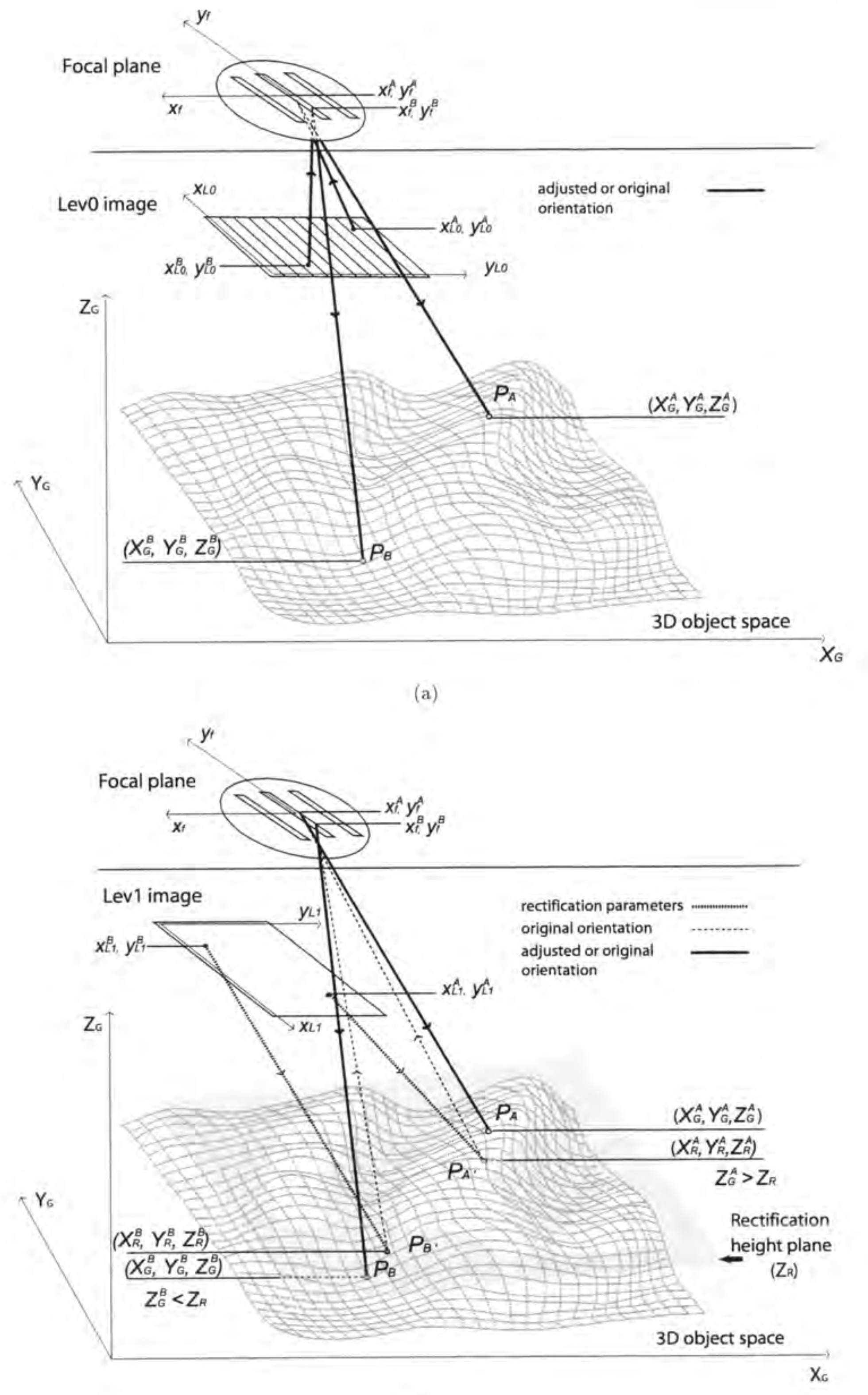


Figure 2.13: Sensor model transformations for Lev0 (a) and Lev1 (b) images. Two selected points on Lev0 (a) or Lev1 (b) images are indicated by i_A and i_B. Their position on the focal plane by f_A and f_B. The ground points are indicated by G_A and G_B. $G_{A'}$ and $G_{B'}$ are the projected image points on the rectification height plane.

and their respective object points are used as input. The image points are selected on Lev0 in a triangular distribution (two on the first and last pixel position of the first scan line and one in the middle of the last scan line) and the respective object points are coplanar to an anchor ground point of average ground height. Initial transformation parameters are derived by means of a 2D affine transformation using the three image and object points and an approximate scan line is determined for a given ground point. The approximate value for the line is refined through a number of iterations and until the solution converges (convergence limit is fixed at 0.1 of pixel size). The main disadvantage of the above described implementation is the time performance. Since the number of iterations (in most cases exceed four) depends on the initial accuracy of the computed scan line, the performance of the sensor model is affected.

In case rectified images onto a height plane or Lev1 image products (section 2.3.4) are used, the sensor model is extended as explained below. An additional transformation, taking into account the rectification parameters, is inserted in the sequence of transformations from Lev1 to object coordinate system (Figure 2.13(b)):

Lev1 image (pixel) $\rightarrow$ height plane (rectification) $\rightarrow$ FPS $\rightarrow$ Object space

In the implementation of the sensor model that we used, the coordinate system of Lev1 images is rotated with respect to Lev0 images. x_{L1} coincides with the y_{L0} axis (position of the pixel in the flight direction) and the y_{L1} with the x_{L0} axis (position in the scanline). To project a point (x_{L1}^A, y_{L1}^A) of Lev1 image onto the rectification height plane (Z_R) Equation 2.5 is used.

$$
\begin{aligned}
X_R^A &= \frac{(x_{L1}^A + tx)}{s_{rect}} \cdot \cos a + \frac{(lines - y_{L1}^A + ty)}{s_{rect}} \cdot \sin a \\
Y_R^A &= -\frac{(x_{L1}^A + tx)}{s_{rect}} \cdot \sin a + \frac{(lines - y_{L1}^A + ty)}{s_{rect}} \cdot \cos a \qquad (2.5) \\
Z_R^A &= Z_R
\end{aligned}
$$

where X_R^A, Y_R^A, Z_R^A are the coordinates of the object point P_A on the rectification height plane, tx and ty are the rectification offsets between the origin of the Lev1 image (upper left corner) and the origin of the reference coordinate system in pixels, a is the rotation angle between the two coordinate systems in radians and s_{rect} is the rectification scale. Furthermore to retrieve the object coordinates of the respective point at height Z_G^A, the pass through the Lev0 image is inevitable with the current implementation, as calibration and orientation data refer to the raw image data. The model becomes more complex if the Lev1 image product has been derived using the original orientation data and the data have been further triangulated and the orientation parameters have been refined. The Lev1 images have to use both orientation data (adjusted and original) for Lev1 image $\leftrightarrow$ object coordinate system transformation, or re-rectify Lev0 image using the adjusted orientation (Figure 2.3.3). To overcome the complexity of transformations, especially when using Lev1 images, each pixel of the Lev1 image should be directly linked to its orientation parameters, or its position on the Lev0 image, by means of a look-up table. However, the only drawback is the increase of the amount of data, as the table would have a similar size to the image in order to store information of each pixel. In most processed datasets, errors in the implementation of the sensor model were the reason for different problems, but which have been fine-tuned over the last years.

2.3.4 Ground Processing

Ground processing is composed of the series of steps, shown in Figure 2.14, which are described in more detail in the sections below.

2.3.4.1 Applanix Post-Processing - Generation of orientation data from GPS/IMU

During flight raw imagery (Lev0) and GPS/IMU data are recorded and stored to the mass memory system (MMS). In the first processing level, post-processing by Applanix, coordinate system conversion and time tagging are applied to the GPS/IMU data, collected for each line separately using the POS system by Applanix, to generate the orientation data file (ODF). The IMU unit (Litton 200) generates a relative position, velocity and orientation of the sensor at 200 Hz, the differential GPS the absolute position at 2 Hz, and the frequency of acquired imagery is 800 Hz and 400 Hz for the panchromatic and multispectral lines, respectively (typical integration time 1.25 ms or 2.5 ms). Errors in the accelerometers and gyros are integrated over time into increasing positional and orientation errors and these are corrected via Kalman Filtering using the GPS measurements. As the frequencies of the GPS/IMU and image differ, the trajectory is interpolated, using a cubic spline, to generate position and orientation at 800 Hz. The accuracies in real-time (navigation solution) and after post-processing (positioning solution) are given in Table 2.11 (Tauno Saks, personal communication). These values are further corrected during the aerial triangulation process, described below, and they are quoted in Table 2.11 for comparison purposes.

2.3.4.2 Image rectification - Generation of Lev1 images

In the second processing level, Lev0 imagery is rectified onto the average ground height plane by using the raw GPS/IMU and the sensor calibration data to produce the Lev1 imagery (using GPro software). Aircraft motions are corrected for Lev1 (Figure 2.15) images and these are used for stereo viewing and feature extraction (semi- or automatic) as geometric distortions (scales, rotations, shears) are reduced, compared to Lev0 images, between the different scenes. A second form of more accurate rectification can be applied, using an approximate DTM. Errors due to the displacement of the relief, apparent with the first mentioned rectification to a height plane, especially visible in areas with generally large height differences, are removed with the second approach. However, since in most cases a DTM does not exist beforehand or is difficult to retrieve from local authorities the first approach is generally employed.

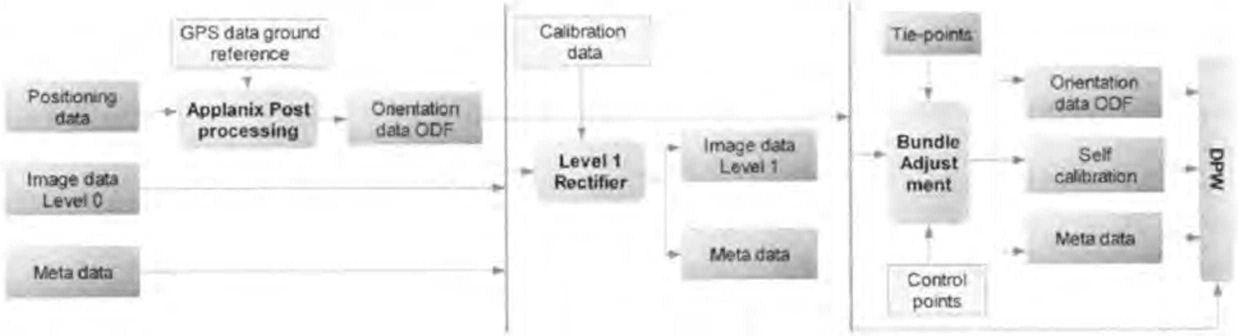


Figure 2.14: Ground processing steps, input and output data. The processes employed are illustrated in the rounded boxes, whereas the input and output data in every process are shown in the regular boxes.

Parameter Accuracy (RMSE)	Real-time	Post-processed	AT
Position	3-10 m	0.1-0.3 m	0.035 m
Velocity	0.5 m/s	0.05 m/s	
Roll - Pitch	$3'$	$2'$	$4''$
Heading	$10'$-$20'$	$3'$	$6''$

Table 2.11: Accuracies of the GPS/IMU system of the ADS40, as compared to the results of aerial triangulation.

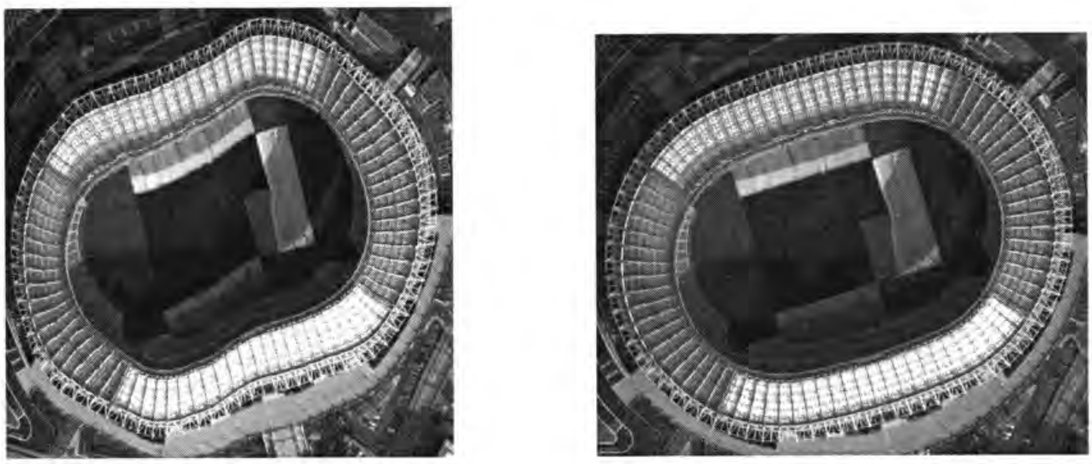


Figure 2.15: Raw (Lev0) and rectified (Lev1) image.

2.3.4.3 Bundle adjustment - Refinement of sensor model, including calibration parameters

In the third step aerial triangulation is carried out using ORIMA software to improve the overall georeferencing accuracy (IMU misalignment, GPS offset, datum differences) and additionally to refine calibration parameters. Prior to bundle adjustment, tie points are generated automatically[5] of using the Lev1 images, as distortions due to perturbations of the aircraft have been corrected and therefore matching is less difficult. The method of orientation fixes is employed in order to replace the projection centers of

[5]Tie points are extracted with the commercial system Socet Set and more recently with LPS.

classical photogrammetry by a series of orientation points in regular line intervals along the flight path (Müller , 1991). The image coordinates are expressed as a function of the ground point coordinates and the orientation parameters of the two neighboring orientation fixes (k) and $(k+1)$(Equation 2.6).

$$\begin{aligned} x_f^{ij} &= F_{ij}(X_G^i, Y_G^i, Z_G^i, X_o^k, Y_o^k, Z_o^k, \omega^k, \varphi^k, \kappa^k, X_o^{k+1}, Y_o^{k+1}, Z_o^{k+1}, \omega^{k+1}, \varphi^{k+1}, \kappa^{k+1}) \\ y_f^{ij} &= G_{ij}(X_G^i, Y_G^i, Z_G^i, X_o^k, Y_o^k, Z_o^k, \omega^k, \varphi^k, \kappa^k, X_o^{k+1}, Y_o^{k+1}, Z_o^{k+1}, \omega^{k+1}, \varphi^{k+1}, \kappa^{k+1}) \end{aligned} \tag{2.6}$$

In Equation 2.6, the image coordinates x_f^{ij}, y_f^{ij} are introduced as observations and the unknowns are the 3 n coordinates (X_G^i, Y_G^i, Z_G^i) of the ground points $P_i(i = 1 \ldots n)$ and the 6 m parameters of the orientation fixes $I^k(k = 1 \ldots m)$. In contrast to classical bundle adjustment, the collinearity equations contain as unknowns the orientation parameters from two orientation fixes k and $k+1$.

$$\begin{aligned} X_o^j &= c_j X_o^k + (1 - c_j) X_o^{k+1} - \delta X^j \\ &\ldots \\ \kappa^j &= c_j \kappa^k + (1 - c_j)\kappa^{k+1} - \delta\kappa^j \end{aligned} \tag{2.7}$$

$$\begin{aligned} \delta X^j &= c_j X_{GPS}^k + (1 - c_j) X_{GPS}^{k+1} - X_{GPS}^j \\ &\ldots \\ \delta\kappa^j &= c_j \kappa_{IMU}^k + (1 - c_j)\kappa_{IMU}^{k+1} - \kappa_{IMU}^j \end{aligned} \tag{2.8}$$

$$c_j = \frac{t_{k+1} - t_j}{t_{k+1} - t_k} \tag{2.9}$$

Orientation of each line is computed by piecewise polynomial interpolation of the adjusted orientation fix parameters and observed GPS/IMU values (see Figure 2.16). The orientation parameters $(X_o^j \ldots \kappa^j)$ for a line j (Equation 2.7) are computed from the nearest orientation fixes $(k, k+1)$ and a correction $(\delta X^j \ldots \delta\kappa^j)$ from GPS/IMU observations (Equation 2.8). Interpolation coefficients (c_j) are a function of the time differences of the nearest orientation fixes (Equation 2.9). Results are used to update the ODF files with the adjusted orientation data. The model used for the adjustment requires a high density of tie points, however a small number of ground control points is needed. Hinsken et al. (2002) justifies with several examples the selection of the spacing of orientation fixes, which depends on the quality of the IMU system. More than 8 and usually 15 tie points are required between neighboring orientation fixes for a stable bundle solution.

The standard configuration of images used in aerial triangulation comprises 4 parallel strips and optionally 2 additional cross strips. The σ_0 a-posteriori for blocks with parallel strips of 20% sidelap is $1.5 - 2.5\mu m$ and with additional cross-strips is $2.0 - 3.5\mu m$. Without any ground control, the accuracy ranges from 0.5-1 pixel. Following triangulation, the Lev1 images can be regenerated with the adjusted orientation data and used for classification, mapping, revision of databases, automatic extraction of DSMs/DTMs and orthophotos, 3D city modelling and fly-throughs.

The calibration by photogrammetric means can verify the quality of the sensor calibration over a test field with precisely measured control and check points. Although a complete measurement and process flow was established for lab calibration, the self-calibration is a system driven approach, which

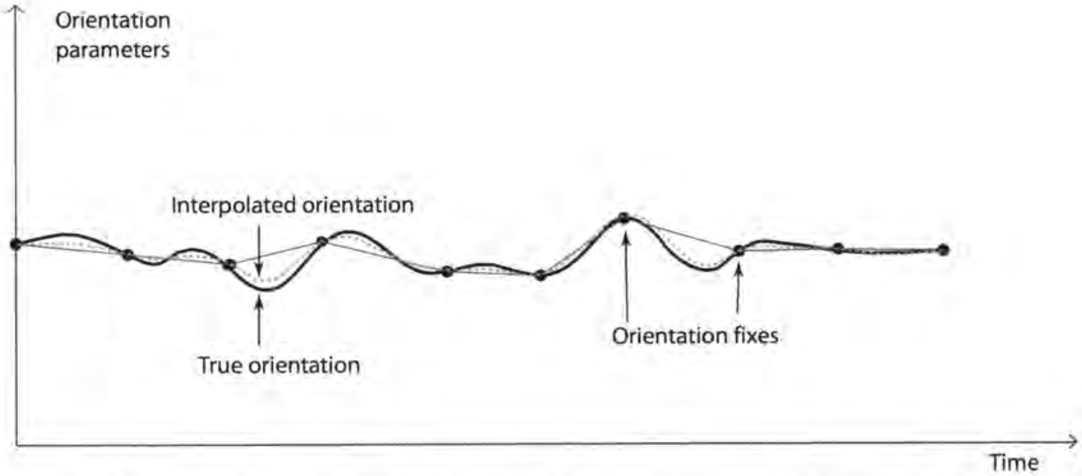


Figure 2.16: Orientation for each line of Level0. True orientation is illustrated with the solid thick polyline. Orientation fixes used in the bundle adjustment are marked with black dots and the thin line defines their coarse trajectory. The interpolated orientation is computed and the dotted line shows the corresponding trajectory.

enables that all image-relevant system components are calibrated. Especially the relative orientation between IMU and FPS system (IMU misalignment) can be only determined via self-calibration or during adjustment of a normal block (usually 4 parallel- and 2 cross strips), as angles of the IMU have to be determined in the range of 10 arc sec or less. To date the self-calibration is performed by the ORIMA software, which uses the set of additional parameters of Brown (1971) to model the system effects and three additional unknowns to model the misalignment angles. However, the Brown set does not suffice for ADS40 geometry and residual analyses showed that the remaining systematic errors can be modelled by a 6^{th} degree of order polynomial for x and y components of each sensor line. This modelling is currently applied with a post-processing package. Usually two iterations[6], are performed using Brown parameters and 3 additional iterations with remaining parameters. Tempelmann et al. (2003) mentions that for an unbiased misalignment estimation and elimination of the correlation of parameters between principal point and IMU misalignment parameters, the calibration block is flown in a bi-directional cross pattern. Moreover, the focal length (or principal distance) can be estimated when a scale is introduced into the block either by using GCPs and one bi-directional cross or without ground control and two bi-directional crosses flown at different heights. Practical tests have shown that from the self-calibration, after five iterations, the σ_0 a-posteriori (standard deviation of unit weight a-posteriori) is $2.5 - 2.9\mu m$. The additional polynomial modelling decreases residuals to less than $1.0\mu m$ for 90% percent of the line length and less than $2.0\mu m$ at the line ends. Since the accuracy potential from the CVG measurement is $6.5\mu m$ and only one single iteration step can be saved, the laboratory calibration will not be used in the future (personal communication, Udo Tempelmann). It is worth to mention that the classical approach for calibration of analogue cameras has to be extended. More generalized approaches have to be adopted for airborne digital imaging systems, as the calibration must not only restrict to the optical part but also include additional system components. Towards this, a network has been established aiming in the

[6]each iteration consists of rectification with the updated calibration files, automatic tie point measurement and bundle adjustment (see section 2.3.4)

development of commonly accepted procedures for calibration and verification of airborne digital system. More information can be found in Cramer (2004).

Chapter 3

ADS40 radiometric data analysis and preprocessing

Investigations on the radiometric accuracy of the ADS40, performed by DLR and LGGM, have been already mentioned in section 2.3.2. However differences in the radiometric quality of Lev0 and Lev1 images have not been reported to date. In this research, analysis has been performed on panchromatic and multispectral images in order to quantify radiometric characteristics, e.g. dynamic range and noise, evaluate overall image quality and to use extracted information to enhance and optimize images for further processing.

3.1 Radiometric quality

The factors upon which the radiometric quality of images depends are on one side the CCD line, the optical system, the integration time and the spectral characteristics of the channels and on the other side the configuration and viewing angles of the CCD lines, the sun azimuth, season and atmospheric conditions. In addition, the ground processing, especially the rectification process, was found to have an impact on the image quality of Lev1 images. Most of the factors are general and apply to both airborne and spaceborne sensors and may lead to radiometric problems, like saturated roofs, edge jitter, strong reflections, color purity, artefacts, which have an impact on image interpretability and use for automatic object extraction. As various image datasets were acquired over different epochs and different areas (Switzerland, Japan, Italy), and the ADS40 system was under constant fine-tuning, the identified problems varied. However, some were common for the majority of the datasets.

An analysis on the distribution of grey values of ADS40 images verified that their dynamic range reaches in general 14 bits (A/D conversion is also 14 bits), however the number of grey values having a substantial frequency was less than 16384. The peak of the histogram is typically towards the darker values with the right part of the histogram decreasing smoothly towards the higher grey values with a large tail as shown in Figure 3.1. Thus, the effective grey level range was 13-14 bits for the panchromatic and 12-13 for the multispectral images, excluding grey values at the histogram ends with frequency less

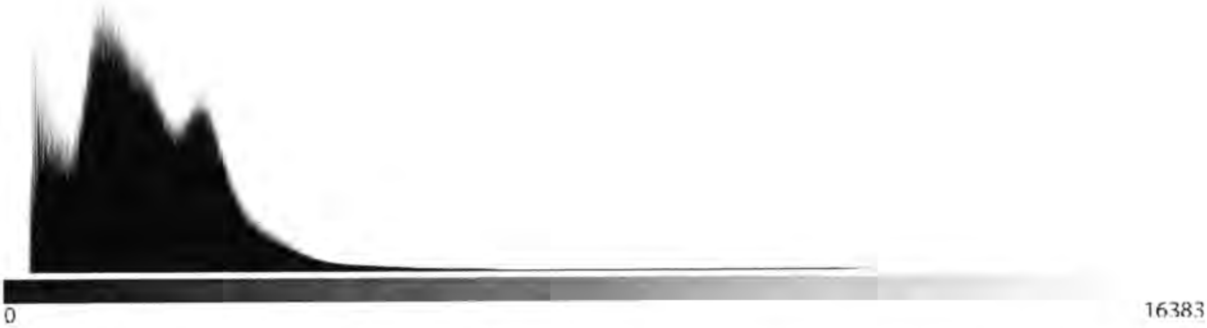


Figure 3.1: Typical histogram of ADS40 Lev0 imagery. As an example the histogram of the Nadir image of SP1_020516 dataset is illustrated. The maximum grey value is less than 16383 (2^{14} bits), being 14644. Frequencies have been scaled by 50 to be viewed as an image, and frequencies towards the brighter values are suppressed due to the scaling.

than ca. 0.0012% (5 times less than the frequency occurring if all 14-bit values were equally occupied). E.g for the Pan Nadir and Green Lev0 image (SP1_020516 dataset over Waldkirch, Switzerland) the maximum grey values are 14644 and 7804, corresponding to 14 and 13 bits, respectively. Deviation from this generic form of histogram applies for the Near-infrared images, as the peak of the histogram is shifted slightly towards the brighter values and the frequencies occupy more values in grey value range with uniform frequencies (see Figure 3.2).

The peak and grey value distribution of the histograms of the stereo panchromatic and multi-spectral channels depends on both the spectral response of the channels and the integration time. Among all channels the Blue channel has a lower sensitivity compared to Red and Green and tends easily to saturation due to atmospheric moisture. Apart from the characteristics of the camera system, image radiometry is also influenced in general by imaging conditions and the sun-sensor-surface geometries. The best collection condition occurs when the sun is behind the sensor and as high in the sky as possible. On the contrary, lower sun elevation can result in heavily shadowed imagery and decreased contrast, which has an even greater effect in the blue channel. Atmospheric absorption and scattering give rise to an increasing blueshift towards the borders of the image. Haze close to the ground can also affect the image quality, decreasing the contrast and interpretability of objects. In the previously mentioned dataset (SP1_020516), the maximum frequencies of the panchromatic channels of one strip are all towards the darker values, i.e. at 1884, 494 and 370 grey value for the Forward, Nadir and Backward channel, respectively, and differences are due to the different viewing angles of the lines and the sun elevation and azimuth at the time of acquisition. Differences in contrast and both quantity and quality of shadows, could be also verified for two datasets over the same area, the first acquired at the end of autumn and the second at the end of spring.

3.1.1 Noise estimation

The radiometric accuracy of an image sensor is affected by different noise sources (dark current shot noise, readout noise, thermal noise, noise and quantization errors from AD converters, noise in the

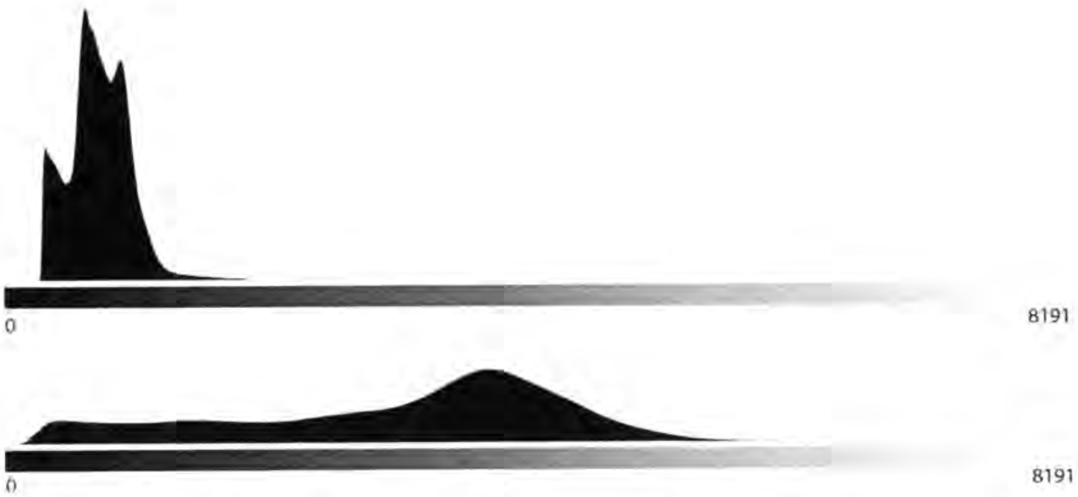


Figure 3.2: Histograms of multispectral channels. Differences of the Green (top) and Near-Infrared (bottom) images are due to spectral characteristics of the filters and the viewing angles of the CCD lines. The spectral radiance of NIR is higher than Green (section 2.3.2). The channels reach the dynamic range of 13 bits (2^{13}=8191).

signal processing circuity, etc.). As it is difficult to separate the different noise sources, we examined the uniformity of the photo response by analyzing the overall noise characteristics in both homogenous (lakes, river surfaces - in general large water surfaces) and nonhomogeneous areas (e.g. whole image, excluding large homogeneous areas). Noise is especially visible in the darker homogeneous areas of the image, where fewer photons are being collected. The use of non-homogeneous areas is justified as large homogeneous areas do not always exist. Plus, this allows an analysis of the noise variation as a function of intensity, as noise for CCD-imagers is not additive but intensity dependent. Both areas are selected manually and should be as large as possible. Water surfaces with apparent specular reflections or waves should not be used as homogeneous areas. A small window, usually of spacing 3x3 or 5x5, is selected and shifted within the area with a freely defined spacing and the standard deviation, as an indication of noise, is calculated. Small windows are justified since homogeneous areas often show low frequency grey value variations, which would lead to higher standard deviation if it were computed from the whole area. For non-homogeneous areas, a small window is necessary in order to get small homogeneous areas between edges. Noise is estimated as follows:

- In *homogeneous areas*, the computed standard deviations are sorted and the N% of smallest ones is used to calculate an average standard deviation, which indicates the noise. Typical values for N are 80-95.
- In *non-homogeneous areas* the grey level range is divided in bins, and the standard deviations are assigned to a bin according to the mean grey value of each window. In each bin, the standard deviations are sorted, and the noise is estimated as the mean of the N% smallest standard deviations. Typical values for N are 3-8 and usually 5, using the reasonable assumption that at least some windows will be homogeneous, even in highly textured areas. The estimation of noise in each bin is reliable only if the N% sample number is sufficiently large (e.g. number of samples $>$ 100). Since

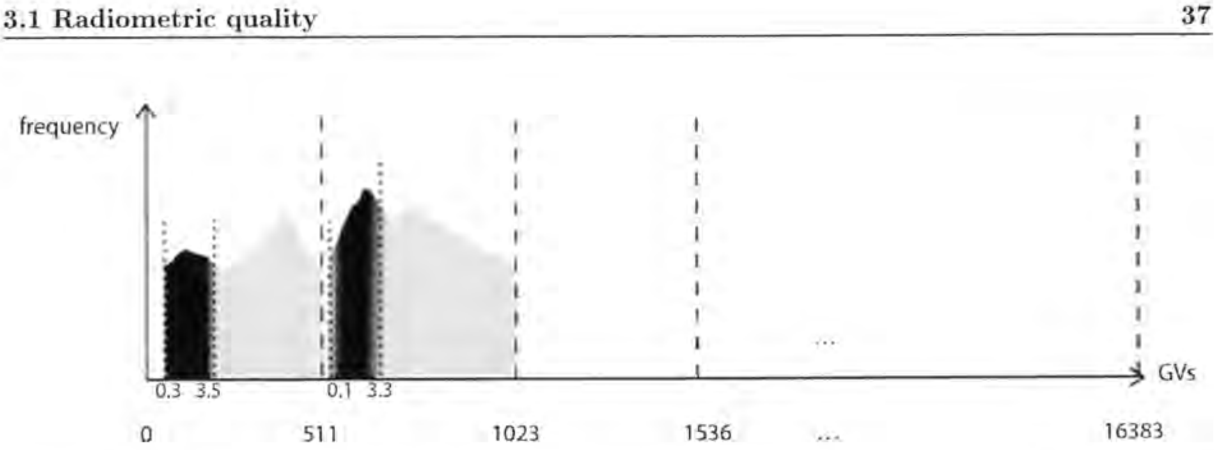


Figure 3.3: Method for noise estimation in non-homogeneous areas. The grey value range is divided in bins (16 usually used). The range of each bin is defined with the vertical dashed lines. The standard deviation used as input range is 3.2 and the minimum non-zero standard deviation is 0.1 grey level. In the two first bins the selected frequencies of standard deviations are illustrated.

the N% has a fixed value for all bins, the equivalent number of samples will be different in each bin, as frequencies do not follow a uniform distribution. Further modifications have been made to the current approach, the reason being twofold: To perform a noise estimation based on the number of significant samples in each bin (derive the percentage in each bin) and to compare the noise level at different processing stages (after noise reduction, signal enhancement, etc.) with respect to a common percentage. In order to compute the number of significant samples in each bin, a minimum standard deviation std_{min} and a range of standard deviations std_R are used as input.

$$\overline{std_i} = \frac{\sum\limits_{k=std_{min}}^{\mathrm{std_{min}+std_R}} k_i \cdot j_{i,k}}{\sum\limits_{k=std_{min}}^{\mathrm{std_{min}+std_R}} j_{i,k}}, \qquad i = 0 \ldots nbin \tag{3.1}$$

According to Equation 3.1, we compute the mean standard deviation $\overline{std_i}$ in each bin i out of the samples j with standard deviation k in the range $[std_{min}, \quad std_R + std_{min}]$. The smallest value ($> 0$) in each bin with non-zero samples is used for std_{min} (Figure 3.3). For std_R the mean standard deviation estimated beforehand in homogeneous areas is used[1]. Usually the mean standard deviation is computed in two ranges std_{R1} and std_{R2}, where $std_{R2} = 2 \cdot std_{R1}$, in order to analyze the reliability of the result, namely to examine the number of selected samples, if they are significant and how they can influence the calculation of the mean standard deviation.

Our results from three panchromatic (Lev0 and Lev1) and four multispectral images (Lev1) from different datasets were found to be consistent with the noise analysis performed by DLR. According to Börner and Reulke (2001) the dynamic, time depended noise is about 2 grey values. As verified, the noise is about 1-2 grey values for panchromatic images and slightly less for multispectral images due to the differences in their dynamic range ($2^{12} - 2^{13}$ bits). Compared to Lev0, noise is less for Lev1 images

[1]If homogeneous areas do not exist in the image, an initial value is taken from another image with common spectral characteristics, assuming that their noise levels would be approximately similar. The std_R is refined in 2-3 iterations, analyzing the reliability of the results (selected samples in each bin) after each iteration.

Image	1 - 511	511 - 1023	1024 - 1535	1536 - 2047	2048 - 2559	2560 - 3071
Pan Lev1	1.3	2.8	3.2	3.6	3.7	5.0
Pan Lev0	2.3	4.4	5.0	5.6	5.5	7.2
RGB & NIR Lev1	1.3	1.9				

Table 3.1: Noise estimation of Waldkirch data using non-homogeneous areas. The average of Backward, Nadir and Forward is used for the panchromatic channels (differences between the three views are small, more information is given in Table 3.2), and similarly the values of standard deviation are taken from the average of the four multispectral channels. In cases where the number of samples was small the standard deviation was unreliable and no value is given.

Image	1 - 511	511 - 1023	1024 - 1535	1536 - 2047	2048 - 2559	2560 - 3071
Pan Bwd Lev1	1.2/1.3	3.0/3.4	3.1/3.6	3.8/4.4	3.7/4.1	5.1/5.8
Pan Nad Lev1	1.3/1.4	3.0/3.4	3.6/4.0	4.0/4.5	4.1/4.5	4.8/5.4
Pan Fwd Lev1	1.3/1.4	2.3/2.6	2.9/3.3	3.2/3.7	3.5/4.0	5.2/6.2
Pan Bwd Lev0	2.0/2.2	4.7/5.1	5.0/5.5	5.7/6.4	5.5/6.2	7.0/7.8
Pan Nad Lev0	2.5/2.6	4.8/5.3	5.3/5.9	6.0/6.7	5.9/6.6	6.8/7.5
Pan Fwd Lev0	2.3/2.6	3.8/4.2	4.6/5.0	5.0/5.5	5.0/5.6	8.0/8.6
Red Lev1	1.0/1.2	1.8/2.0				
Green Lev1	1.0/1.2	2.0/2.3				
Blue Lev1	1.0/1.0	1.2/1.4				
NIR Lev1	1.2/1.3	2.4/2.8	3.3/3.6			

Table 3.2: for all channels of one strip of Waldkirch data using non-homogeneous areas. Modified version with two ranges. The mean standard deviation for the 1^{st} and 2^{nd} range is given for each bin. For unreliable estimation due to small number of samples no value is given.

due to the resampling of pixels during rectification and besides this, as can be seen in Table 3.1, noise is intensity depended for all images. A more detailed analysis for each channel individually is shown in Table 3.2. It is worthy of note that noise estimation in homogeneous areas could not be performed for a number of images, since large water surfaces did not exist in the images, or water areas were avoided due to reflections from their surface.

The developed method has also been applied to image products of the IKONOS-2 high resolution satellite, which also utilizes CCD technology for image acquisition. Baltsavias et al. (2001b); Fraser et al. (2001) in their assessment on IKONOS Geo products were able to quantify the low noise levels of different IKONOS images, apart from the visual verification.

3.1.2 Artefacts

ADS40 images were found to exhibit artefacts. Black and white stripes on roofs (Figure 3.4(a)) in Lev0 images occurred as the grey value exceeds the upper limit (white) in the dynamic range and the value is zeroed (black). However, this stripe-effect is corrected in the Lev1 images (Figure 3.4(b)) during rectification. Saturation due to overexposure is visible in both panchromatic and multispectral (Figure 3.4(c)) images. It increases with longer integration time (tests performed by LGGM showed higher saturation with 5 ms integration time compared to the 2.5 ms used in most cases) and it is especially strong for vertical walls or roofs with surface normals approximately in the middle of the illumination-to-sensor angle. Visible edge jitter appears in some parts of Lev1 images. This results from large distortions in the Lev0 images that can be corrected with the bilinear interpolation used in the rectification process. Moreover, identified black dots (Figure 3.4(d)) and stripes across Lev1 images (Figure 3.4(e)), due to errors in the interpolation of grey values, that existed in early datasets have now been corrected in the later ones by using large masks (5x5 pixels or more) for interpolation of the grey values. The spectral channels were more subject to radiometric problems: the misregistration due to problems with the calibration of the color channels, which has been improved with a new algorithm described in Tempelmann et al. (2003); the apparent ghosting effect (Figure 3.4(f), 3.4(g)) characterized by shifted green, blue and red stripes due to reflections of light in the beam splitter and which are of 5 pixels or more; and the radiometric or illumination uniformities, namely a general brightening and a blue hue towards the border pixels of the CCD (Figure 3.4(h)) due to atmospheric absorption and scattering[2], which occur especially at short wavelengths, increase proportionally with the flying height, are highly angle dependent and therefore the impact is different within the image. Due to the atmospheric moisture, the blue channel also tends to be easily saturated. These atmospheric effects can be corrected by means of atmospheric filters, i.e. Beisl and Woodhouse (2004) uses a dark pixel subtraction algorithm, which accounts for the largest effects of the atmosphere and it takes into account the view angle dependence of the path radiance by calculating column statistics.

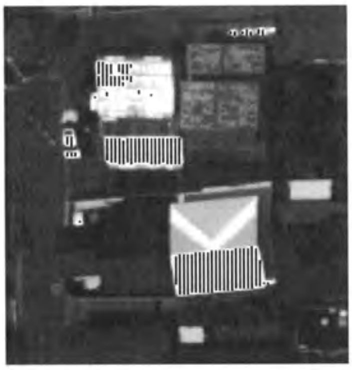
Additional artefacts are visible in pseudo-color images, due to the different viewing angles of near-infrared and red and green channels (FP1 configuration). Moving objects appear in different positions on the ground in each acquired channel, leaving their color pattern in the composite image. In the FP3 and FP4 configuration (section 2.3.1), co-registration of the above channels is realized as they are placed in nearby positions on the focal plane.

3.2 Radiometric data preprocessing

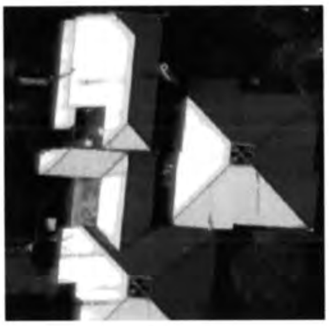
We have developed and tested various processing methods in order to both reduce the effects of the radiometric problems and optimize the images for subsequent preprocessing:

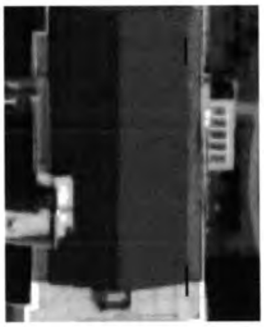
1. Noise reduction filters, in order to reduce even more the noise level and minimize subsequent noise amplification when contrast and edge enhancement are applied.
2. Carry out image enhancement to improve definition of features, and in parallel to radiometrically balance the images, namely minimize radiometric differences among panchromatic and multispectral channels.
3. Reduction to 8-bit imagery by non-linear methods in order to preserve as much information as possible in the 8-bit grey level range.

[2]gaseous absorption of the directly transmitted light and the gaseous (Rayleigh-) and aerosol (Mie-) scattering of indirect components reaching the sensor

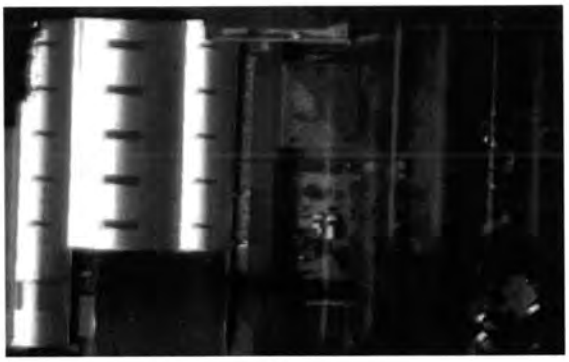



(a) (b)



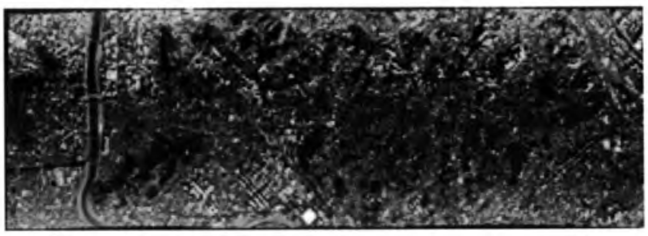


(c) (d) (e)



(f)

(g)



(h)

Figure 3.4: Artefacts in ADS40 images(explanations in text).

4. Optimize image quality for visual purposes, mainly of color images, by merging the separate bands using a weighting scheme, so aiming to reduce the effect of the blue channel in the composite image.

The methods are used prior to matching and are combined according to the scheme shown in Figure 3.5. First noise is estimated in homogeneous and non-homogeneous areas. Then, the noise reduction filters are applied and the noise estimation is repeated (dashed line) to quantitatively evaluate the factor of noise reduction. Wallis filter is applied for contrast enhancement, followed by an optional reduction to 8-bit.

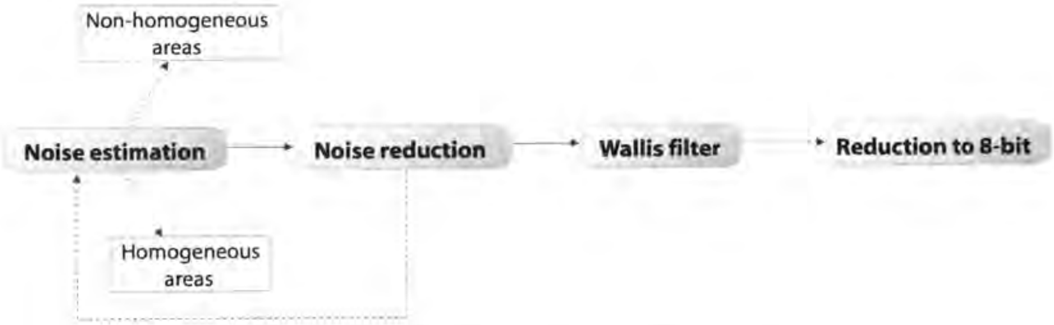


Figure 3.5: Scheme of radiometric data preprocessing.

3.2.1 Noise reduction

Several algorithms that employ spatial filters for noise reduction exist in the literature. Local average, median and Gaussian are the most frequently used. In general these methods cannot remove spikes in homogeneous areas and in the neighborhood of sharp changes in intensity, details are blurred. The median filter rounds corners in cases of random noise, but is ideal for elimination of salt and pepper noise. Other filters, e.g. sigma (Lee, 1983), inverse weighted smoothing and edge preserving smoothing (Nagao and Matsuyama, 1979), avoid smoothing of edges to a certain extent, however smoothing is weak in homogeneous areas. Out of a series of tests that we performed, edge preserving smoothing was proved to be computational demanding and the input parameters are empirically estimated. Uniform grey level regions are created even in regions of slow signal variation, plus all details smaller than the mask size are eliminated.

We developed two adaptive local filters, aimed at reducing noise, while preserving even fine detail such as one-pixel wide lines, corners and line end-points. The effect of the two filters is quite similar as they are both based upon the concept of sub-masks employed in edge preserving smoothing, and one utilizes a fuzzy method. Besides, both were implemented to be used also for images higher than 8-bit, thus utilizing a larger range of grey value differences. Look-up tables were also used in order to improve the time performance. Both filters require as input an estimate of the noise, which may be known or estimated beforehand by the methods mentioned in section 3.1.1. The first filter, called adaptive edge preserving smoothing, computes the homogeneity by means of absolute differences of neighboring pixels to the central one D_i in a central 3x3 mask and compares the differences to a threshold T (usually T is set greater than $3 \cdot s$, where s the noise level). Differences larger than T receive zero weights, otherwise

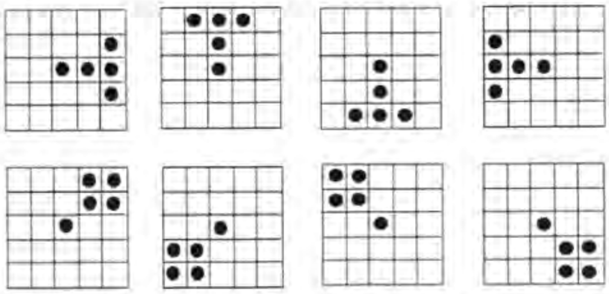


Figure 3.6: Directive masks (5 x 5) used in Adaptive Edge Preserving Smoothing.

the weight equals one and the number of M neighbors with non-zero weights is stored (M decreases with decreasing T). Four options are handled, according to the number of M neighbors:

- $M = 1$, point can be an end point of a 1-pixel thick line (flag F=1).
- $M = 0$, point could be a one pixel spike, which could be faint if T is small. The grey value remains unchanged or replaced by the median of the eight neighbors.
- $M > 4$, the point is in a homogeneous area with random noise. The grey value is replaced by the average of central pixel and M neighbors.
- $M > 0$ and $M \leq 4$, the mask size is increased to 5x5. We use the eight directive masks (see Figure 3.6) to compute the number of non-zero weights (m_i, $i = 0 \ldots 4$) in each mask. m_i is used for the eight sub-masks, whereas M indicates the non-zero weights in the 3x3 mask. To reduce computations the non-zero weights m_i are calculated only for the masks, in the direction of the eight neighbors with non-zero weight (M) in the 3x3 mask. The number of masks (n) with maximum number of m_i are found, and can be more than one. If the sum of max m_i in the n masks $\geq n \cdot 2$, the grey value is replaced by the average of the central pixel and its neighbors with non-zero weight in masks of max m_i. Else if F=1, then the case M=1 is valid, the positions correspond to two isolated pixels, and the grey value is replaced by the average of the two pixels or the median of the seven neighbors with zero weights in the 3 x 3 mask. For the last case, where M would have 2 or 3 non-zero weights, the grey value is computed similarly to the previous case (F=1) (average of the 2 or 3 pixels, or the median of the 6 or 5 neighbors with zero weights in the 3x3 mask.

If s is too small, then smoothing of homogeneous areas is less, whereas edges remain, but sharpening is less. Otherwise if s is too large, small grey value differences in homogeneous areas are disappeared as the number of M increases ($M > 4$) and results are similar to those from a simple median or average filter (depending on whether the average or median option is used in the programm) and therefore edges are less preserved.

The differences D_i are also treated in a fuzzy way. The averaging is replaced by a weighted averaging. To compute the weights two input thresholds (T_1, T_2) are required, namely the lower threshold below which the weight becomes 1 and the upper threshold above which the weight becomes 0. T_1 is computed as T in the previous approach, i.e. $3 \cdot s$, whereas T_2 is given a large number, e.g. $T_2 = 40$, and

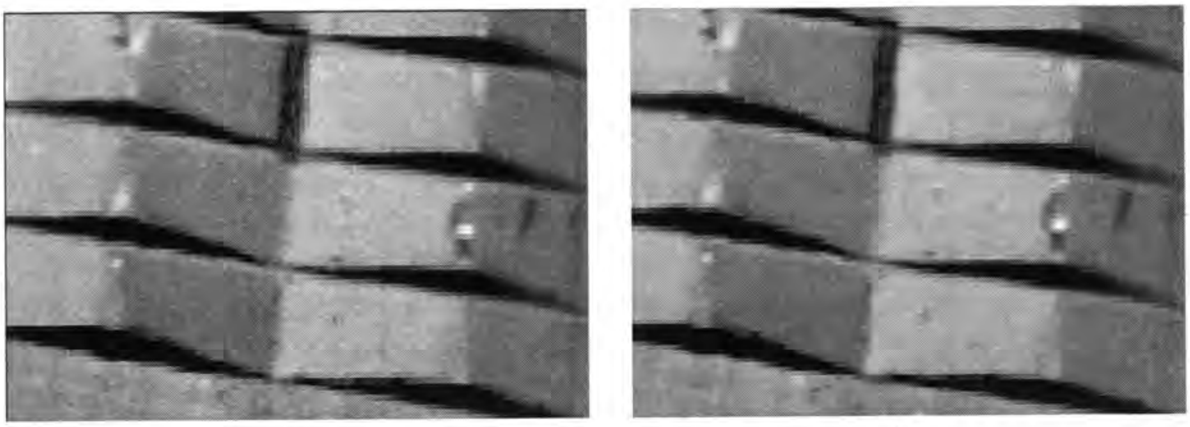


Figure 3.7: Noise reduction in panchromatic ADS40 imagery. Roof detail in the nadir channel of ADS40 before (a) and after (b) noise reduction. In the left image noise is amplified through strong contrast enhancement. In the right image noise reduction using the fuzzy method is employed with visible sharpening of the edges and reduction of the noise.

Image	Noise Reduction	1 - 511	511 - 1023	1024 - 1535	1536 - 2047
Backward	no	1.2	3.0	3.1	3.8
Backward	yes	0.4	1.3	1.4	1.7
Nadir	no	1.3	3.0	3.6	4.0
Nadir	yes	0.5	1.3	1.5	1.7
Forward	no	1.3	2.3	2.9	3.2
Forward	yes	0.5	1.1	1.4	1.6

Table 3.3: Noise estimation results from the Waldkirch data before and after noise reduction (NR) using non-homogeneous areas.

weights are computed according to Equation 3.2.

$$w_i = \begin{cases} 1 & ,\forall \quad D_i \in [0, T_1] \\ \frac{T_2 - D_i}{T_2 - T_1} & ,\forall \quad D_i \in (T_1, T_2) \\ 0 & ,\forall \quad D_i \in [T_2, \infty) \end{cases} \tag{3.2}$$

The effect of both filters is fairly similar, however the fuzzy approach leads to slightly more edge sharpening and is more favourable for visual purposes. When replacing the grey value by the median, instead of the average, similar results are obtained. However when the thresholds are set too large, the smoothing of homogeneous areas appears artificial when the median is used. Aside from via visual verification (Figure 3.7), a reduction of noise can be evaluated through the analysis of noise levels (see 3.1.1) in the filtered image. As it can been seen from Table 3.3, noise is reduced for the panchromatic channels by a factor of about 2-3 after filtering is employed. Similar reduction applies for the multispectral channels with larger effect on the Blue channel due to the lower SNR.

3.2.2 Contrast and edge enhancement, radiometric balancing

Contrast enhancement is applied to increase signal variation in parts of the image with less texture (e.g. shadowed areas) and improve feature definition. The grey values are distributed so that significant frequencies occur for a larger number of grey values. Besides, in order to minimize radiometric differences between channels[3], it is important that all channels are radiometrically balanced. Thus, noise level has to be reduced prior to contrast enhancement to avoid noise amplification, occurring parallel to signal amplification. The Wallis filter (Wallis, 1976) was used for image enhancement (see Figure 3.8) and radiometric balancing among channels by forcing the filter parameters to be identical for all processed channels. A thorough description of the filter and its advantages over other radiometric corrections (utilizing polynomials) can be found in Baltsavias (1991). The filter was re-implemented to encompass processing of images with a dynamic range higher than 8-bit, plus to permit automatic estimation of mean and standard deviation at run-time by an iterative non-linear method, which enforces similar frequencies for the grey values. The frequency for each grey value is set as $p = \frac{1}{2^n}$, where n is the number of significant bits. The initial frequencies are sequentially accumulated till the p frequency is reached for each grey value in increasing order. The mean and standard deviation are then computed out of the updated distribution of grey value frequencies. Due to the general form of histograms (see section 3.1), the accumulated frequencies, are still towards the darker values, however more uniform frequencies occur. If images are processed in the original bit range, the mean value is adjusted to half of the range (8192 for 2^{14}) to increase brightness.

With the Wallis filter grey value differences were locally amplified and definition of details increased. Texture enhancement could be observed in agricultural fields, whereas shadowed areas did not have a significant signal variation, and thus in spite of the contrast enhancement, details in these areas were not often easily visible. In images with variations in landcover and man made objects, different windows in the Wallis filter should be used according to the percentage of bare ground areas, shape and size of buildings and road types and the degree of required contrast. The filter would then enhance feature definition more uniformly, strong contrast enhancement over large homogeneous areas would decrease and visual interpretation of images would improve. This approach is more time consuming as in order to distinguish the different area types, prior to the contrast enhancement with Wallis, segmentation of the image is required.

3.2.3 Reduction to 8-bit

To date most commercial systems although they are able to import images higher than 8-bit, their internal algorithms are designed for processing 8-bit images. The most common approach to compress the range of grey values from 16 to 8 bits uses a linear transformation, having as main disadvantage that frequently occurring grey values are less preserved and more information is lost. Considering also the distribution of frequencies in the histograms, mentioned in section 3.1, a linear transformation does not suffice to scale in an optimal way grey values in the 8-bit range. As observed in Figures 3.9(a) and 3.9(d), the grey values of the output histogram, using linear scaling, tend to be dark, plus many grey value

[3] Area based matching methods (e.g. cross-correlation, least squares matching) are sensitive to radiometric differences

(a) (b)

Figure 3.8: Contrast enhancement by Wallis filtering. Results before (a) and after (b) Wallis filtering.

differences are merged together, resulting in loss of information. To overcome this negative aspect and thus preserve more grey values that are more frequently occurring, histogram equalization is performed by a non-linear iterative method (a similar concept to that presented in section 3.2.1), with the aim to occupy all 8 bits with similar frequencies and preserve as much information as possible from the 14-bit grey value range. Therefore, for each grey value the frequency threshold p_i is set according to Equation 3.3.

$$p_i = \frac{1}{2^8}, \qquad i = 0 \ldots 255 \tag{3.3}$$

The histogram equalization may lead to too strongly bright and dark regions for visual interpretation as shown in Figures 3.9(b) and 3.9(e). The peaks in the histogram (Figure 3.9(e)) occur either due to grey value frequencies $\gg p_i$ in the 14-bit range that cannot be divided, or due to the iterative combination of frequencies in order to occupy the final 256 grey values. Alternatively the non-linear reduction with uniform frequencies can be replaced by a non-linear reduction of Gaussian type and the results are illustrated in Figures 3.9(c) and 3.9(f). The distribution of frequencies depends on the input standard deviation σ, which determines the slope of the Gaussian histogram, and the normalized frequency threshold for each grey value in the 8-bit range is set according to Equation 3.4.

$$p_i = \frac{\dfrac{e^{(i-m)^2}}{2 \cdot \sigma^2}}{\sum p_i}, \qquad i = 0 \ldots 255 \tag{3.4}$$

where m is the midpoint in the Gaussian distribution with the maximum value or frequency and usually is set at the 127 grey value. This transformation has similar results to linear 8-bit reduction, if Wallis has been applied beforehand. The reason is twofold: the grey value frequencies follow the Gaussian distribution, and more significant frequencies occur compared to the original image. In this case a linear map-down also suffices, since contrast has been enhanced, resulting in minimal loss of information.

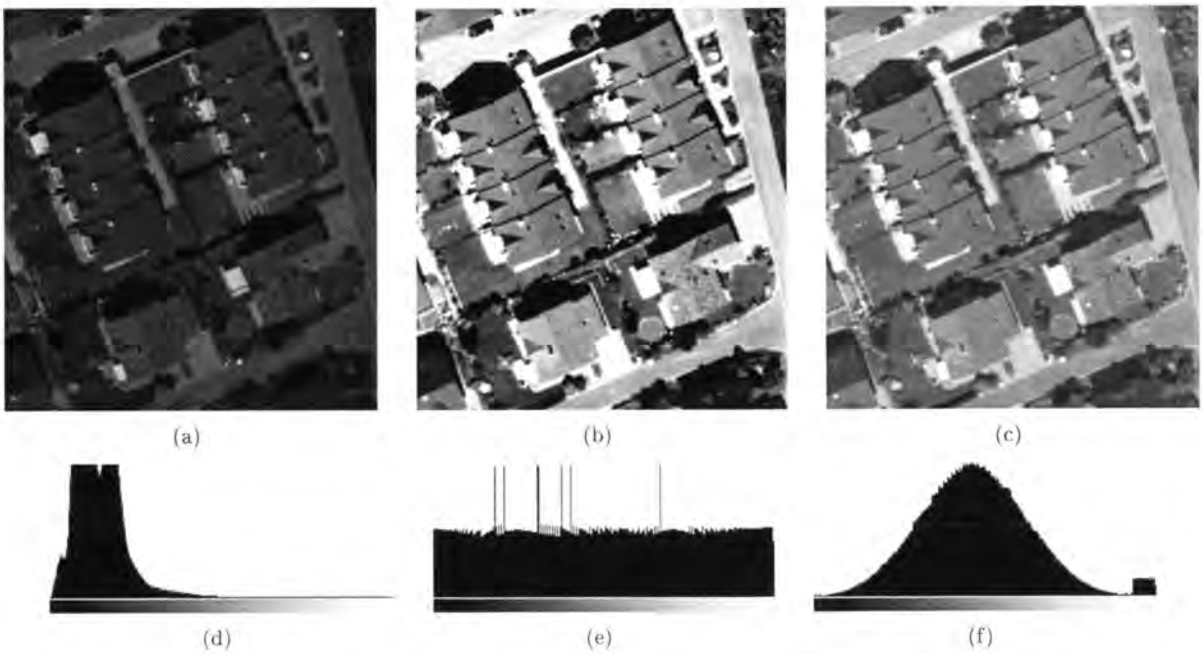


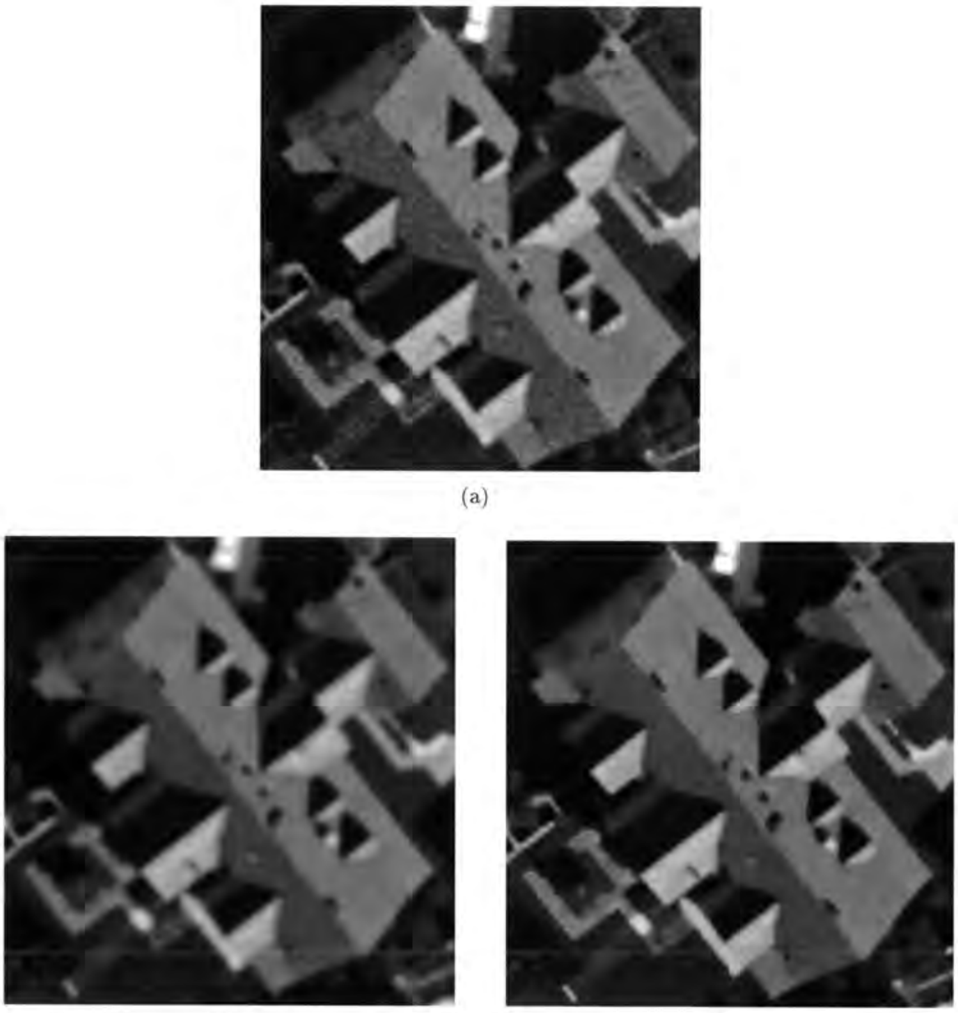
Figure 3.9: 8-bit reduction methods and corresponding histograms.

3.2.4 Processing of multispectral channels

As already mentioned in sections 2.3.2.1 and 3.1, the spectral response and noise level of the multispectral channels differ. To reduce the blueshift effect and improve image quality in true or false color images for visual purposes, the individual bands (red, green and blue or near infrared) can be composed by different weights, which are determined according to noise level of the multispectral channels and computed beforehand. The average standard deviations, estimated in a homogeneous area (see section 3.1.1), are used as color balance factors for the channels, and further normalized according to Equation 3.5:

$$w_i + w_j + w_k = 3.0 \tag{3.5}$$

where w_i, w_j, w_k are the weights for any three of the available multispectral channels (i,j,k). It should be considered that relative higher noise level in one channel compared to the other channels will result in very small weight. Tests showed that the weight cannot be less than 0.7 and larger than 1.3, otherwise the colors in the composed image are distorted. Moreover, these weights depend on the imaging conditions as well as on the spectral response of the channels, in contrast to the radiometric coefficients computed as part of the absolute radiometric calibration (see section 2.3.2.1). For improved results, prior to fusion of channels, noise reduction can be applied to minimize the noise level. As the preprocessing techniques of noise reduction and contrast enhancement presented in this chapter are general, they can be applied to different types of images. With scanned color film the noise level in the blue channel is generally high and the results obtained by combining noise reduction and weighting for the individual channels are illustrated in Figure 3.10. However, problems in the radiometric correction of color images may appear when contrast enhancement or/and 8-bit reduction are applied. If each of the R, G and B channels is



(a) (b) (c)

Figure 3.10: Noise reduction in color scanned frame imagery. The color film has been scanned with the Zeiss SCAI scanner with $14\mu m$ resolution. The visible noise on the roof part of the image (a) has been reduced with Gaussian (b) and fuzzy filter (c). Noise reduction has been employed for each channel individually, prior to generation of the RGB image.

processed separately, then when the channels are combined new unnatural colors appear (color shift). In 8-bit reduction of color channels by non-linear methods, even if the transformation matrix (LUT) is computed out of one band, or from the average of the histograms of all bands, color disturbances may still occur, especially if the histograms of the separate bands have significant differences. In general, a more efficient approach would be to apply the transformation to the intensity component of an HSI image, or the luminance component of a YIQ image, thus leaving the chromaticity unaltered. This separation of the luminance component from chrominance information is stated to have advantages in applications such as image processing (Ford and Roberts, 1998). Processing images with the RGB models makes it difficult to maintain hue and makes the RGB model not desirable in case of non-linear transformations or filters, whereas linear filters can be used to process an RGB image because hue is preserved.

The Gaussian iterative method, used for 8-bit reduction, is applied in both grey scale and color imagery in RGB space and examples, illustrated in Figure 3.11, correspond to the same part of the image. In Figures 3.11(a) and 3.11(b) the Green channel has been reduced to 8-bit using 20 and 40 as input values for standard deviation (σ), respectively. With higher standard deviation, contrast increases due to the distribution grey value frequencies, which are assigned to the Gaussian type distribution of smaller slope. For color images, when the transformation matrix is computed out of one reference band (the one with the largest number of significant bits, in this case the Green channel) and applied to all bands, colors in the final image are enhanced for the reference band. Examples of this method using a σ of 20 and 40 as input are illustrated in Figures 3.11(c) and 3.11(d), respectively. Computing separate transformation matrices out of the histograms of each band will change the image chromaticity even more (Figure 3.11(e)) compared to a common transformation for all bands, computed from the average of the Red, Green and Blue histograms (Figure 3.11(f)).

The problems indicated in the previous paragraph could be reduced by performing the processing in HSI color space and evaluating visually the results. Examples from the tests for both the non-linear 8-bit reduction and the Wallis filtering on color images are illustrated in Figure 3.12. The non-linear 8-bit reduction was applied according to the following scheme: (a) first the image is converted from RGB to HSI color space using the transformation of Gonzalez and Woods (2002); (b) the histogram and the transformation parameters for the 8-bit reduction are computed from the intensity component; (c) the intensity values are converted to 8-bit by keeping unaltered the saturation and hue components; (d) the reverse conversion is applied from HSI to RGB color space.

Part of the final image (same as in the examples of Figure 3.11) is illustrated in Figure 3.12(a) and the respective intensity component is illustrated in Figure 3.12(b). The image was reduced to 8-bit using the Gaussian iterative non linear approach (σ=20) and as expected and compared to the previous examples the balance of colors in the final RGB has been improved. Similar approach was also followed in the case of the Wallis filter, namely processing of the intensity component (Figure 3.12(c)) and reverse transform in the RGB color space. With respect to this, results from the standard approach, namely processing each individual channel in the RGB color space (Figure 3.12(d)), are compared with the intensity-based approach of the HSI image (Figure 3.12(e)) using common parameters for the Wallis filter. The advantage of the second approach is that color information is more preserved and less problems (color shifts) occur, especially in highly textured areas. In the two above examples the mask size was set to 50 x 50 pixels, whereas it is interesting to observe that with a larger mask size (Figure 3.12(f)) the color shifts are less apparent.

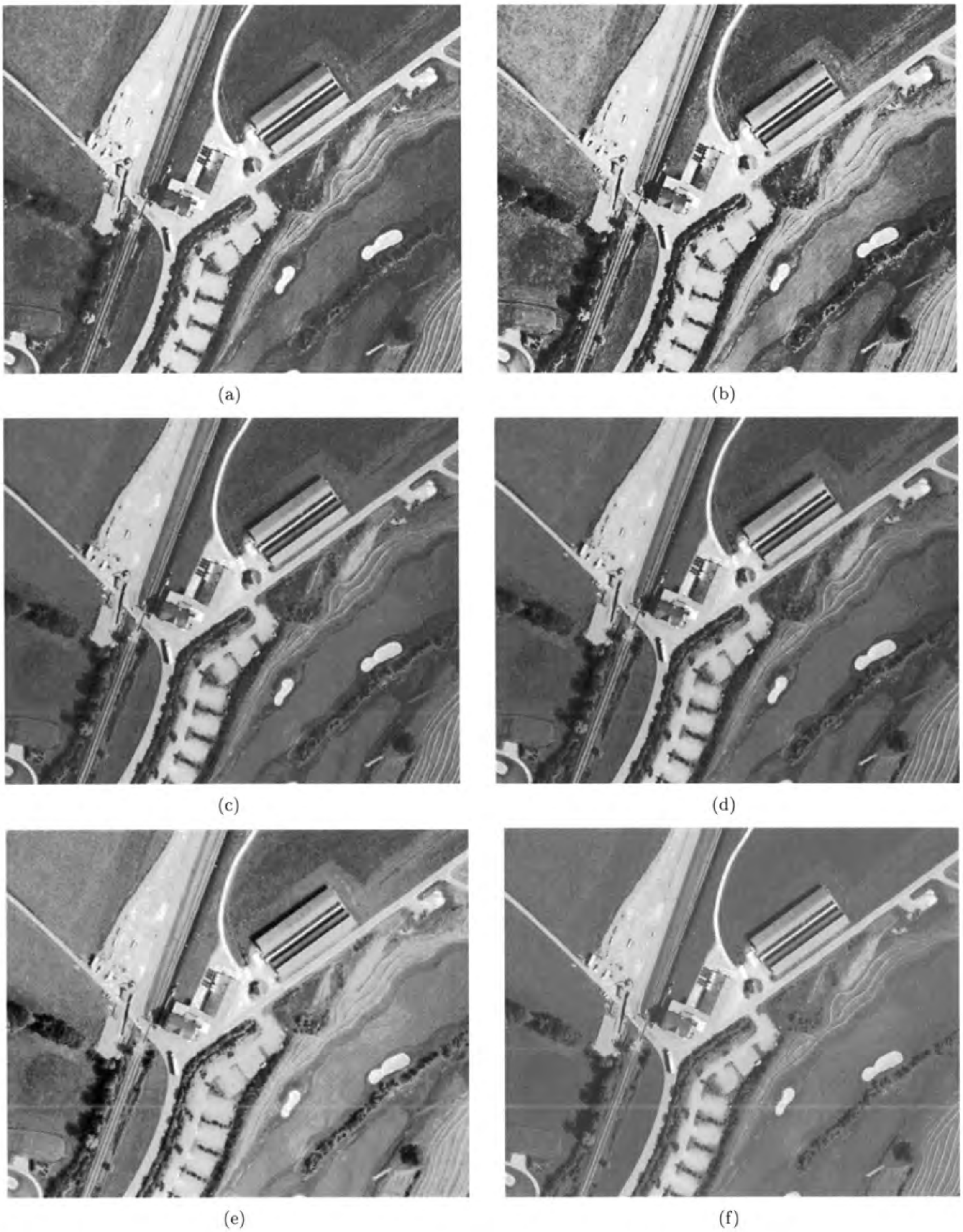
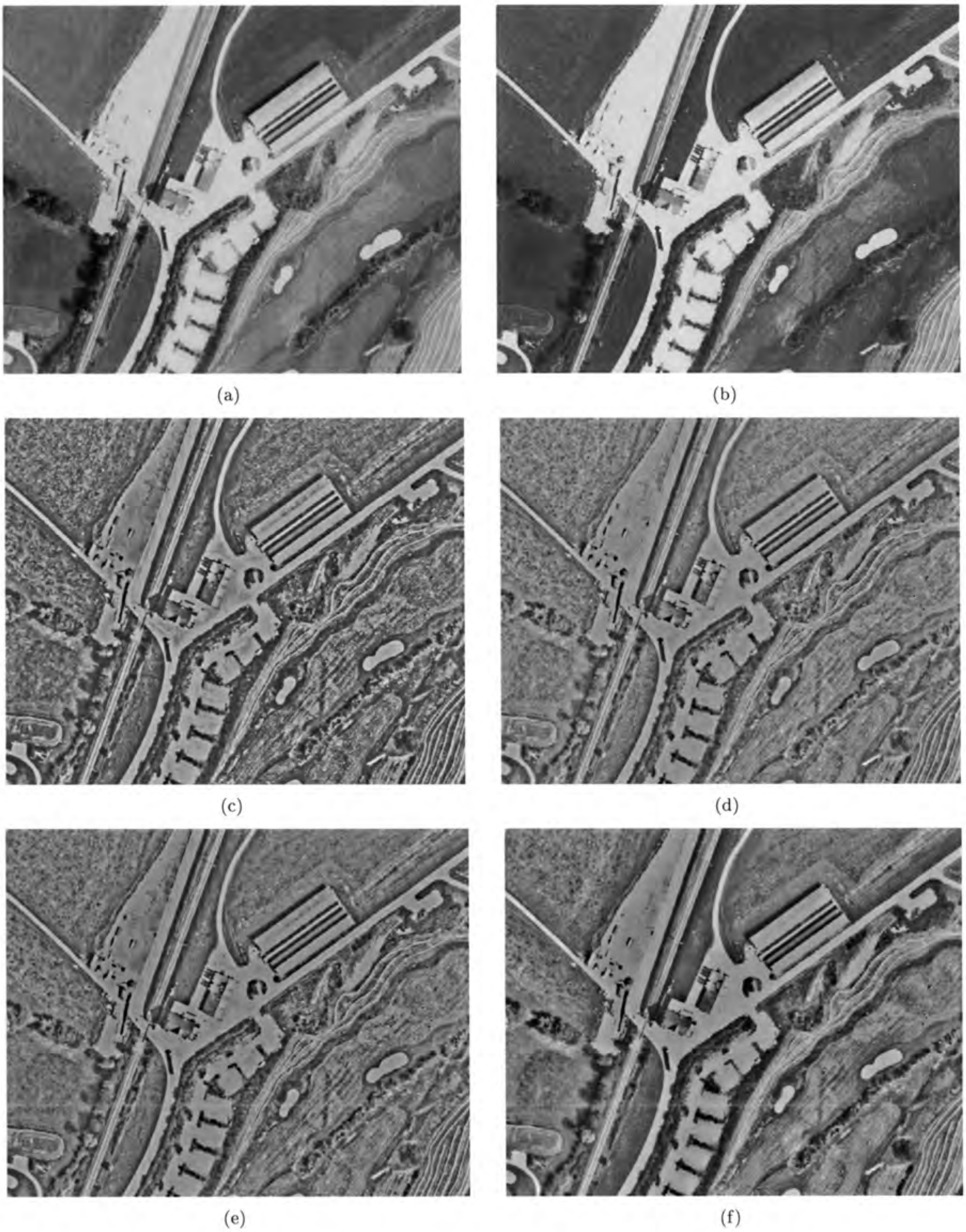


Figure 3.11: Variations of 8-bit reduction in grey scale and color images (explanations in text).



(a) (b) (c) (d) (e) (f)

Figure 3.12: Results from radiometric processing using HSI color space (explanations in text).

Chapter 4

Image matching

4.1 Overview of image matching algorithms

Matching has long been and still is one of the most challenging tasks in photogrammetric research, since many steps in the photogrammetric chain (interior and relative orientation, tie point extraction, image registration and DSM generation) are linked to matching in one way or another. However, matching itself has expanded into various disciplines, serving different applications but without a unifying theoretical base, as each discipline has different theorems and notations. Considering these facts, it is almost impossible to give a complete overview of such a wide spread subject as matching. This chapter thus gives an insight into the techniques that have been developed and used within the last years. Only matching related to photogrammetric applications is handled, since there is a direct link to the subject of this thesis, plus the focus is on the latest developments, as several review publications (some listed below), outline or describe in detail the older methods used in photogrammetry.

Matching is usually categorized to area- or signal-, feature-based and relational matching (Hannah, 1988; Lemmens, 1988; Lang and Förstner, 1995; Schenk, 1999). Pilgrim (1991) divides conducted research into two- and three-dimensional matching. Baltsavias (1991) gives an exhaustive overview and classifies the matching methods according to the two components of data (original sensor data, processed image data, features and associated attributes) and algorithm used to solve the correspondence problem. The latter is subdivided into the matching entities, the criteria (similarity measures, energy criteria, feature attributes), the models, the strategy and the parameter estimation. Nevatia (1996) proposes a taxonomy based on the following issues: the representation level (pixel intensity values, point features, grouped features like curves, high-level features like surfaces and volumes); whether matching is performed in 2-D vs. 3-D space; whether local vs. global matching is applied; and the method of matching (area-, feature-, structure- based matching) that is being employed. Heipke (1996) follows a similar classification pattern as Nevatia, but one which is more generalized, and Julien (1999) focuses on the aspects of area-based correlation. The article of Dhond and Aggarwal (1989) gives an overview on stereo matching from the computer vision point of view. Faugeras (1993) similarly discusses the correspondence problem in terms of selected features (points, line segments, curves and regions), constraints (epipolar, uniqueness, continuity, ordering, disparity gradient, geometric shape), correlation and relaxation techniques, as well as dynamic programming. Scharstein and Szeliski (2002) adopt a different classification scheme based on the observation that stereo algorithms generally perform the following four steps: matching cost computation, cost aggregation, disparity computation and disparity refinement. As the factors, upon which the matching strategy depends, are interrelated, it is reasonable to build our taxonomy upon the two distinctive and elementary approaches of area- (or photometric) and feature- (or geometric) based matching.

4.1.1 Area-based matching

Area-based matching uses image intensities as input to solve the correspondence problem between two or more images utilizing a geometric transformation, a similarity measure and an algorithmic solution. The geometric transformation implicitly refers to the type of the object surface: translation; translation and rotation; translation, rotation and scale; affine (locally planar surface); projective (globally planar surface); smooth (smooth surface without occlusions); piecewise smooth (possibly with occlusions). Chambon and Crouzil (2003) present a taxonomy for fifty similarity measures covering five categories (cross-correlation-based, classical statistics-based, derivative-based, ordinal and robust measures) and focus on the properties of robust measures. However, the most frequently applied similarity measures are: distance (Hausdorff, Euclidean), the sum of products (2nd moments, covariance); sum of squared differences (least squares, L_2-norm); sum of absolute differences (L_1-norm); and normalized cross-correlation. The algorithmic solution refers to whether the search is sequential, heuristic, iterative or provided by dynamic programming. In Table 4.1 selected methods utilizing area-based matching are listed. The majority of algorithms utilize cross-correlation as a similarity measure in 2-D (Kölbl et al., 1994; Zhang et al., 1995; Belli et al., 2001) or in 3-D (Miller and De Venezia, 1992; Zhang and Miller, 1997; Elaksher and Bethel, 2002). Cross-correlation is based on the assumption that geometric differences are modelled only by translation, and radiometric differences exist only due to brightness and contrast. Thus, its precision is limited, decreases rapidly if the geometric model is violated (rotations greater than 20° and scale differences between images greater than 30%, more details in Förstner (1984)) and cannot handle occlusions in a straight-forward manner. Partial correlation (Lan and Mohr, 1995), and offset windowing (Okutomi et al., 2002) have been proposed to overcome the problem of occluded area parts. Moreover, no technique is known how to generalize it for multi-image matching.

Least squares matching (Förstner, 1982; Ackermann, 1984) is a generalization of cross correlation. Several approaches exist in the current literature using least squares for surface reconstruction, employing different geometric and radiometric models. Assuming that the local surface patch is a plane to sufficient approximation, the projective transformation (8-parameter transformation) can be employed. Furthermore, the projection can be approximated by an affine transformation (6-parameter transformation), if the surface patch is formed by a very narrow bundle of rays (Grün, 1985). The model used to approximate the surface depends on the complexity of the surface. Rosenholm (1987) employs bilinear finite elements to represent the true surface in multi-point matching. Representations of this type have also been adopted by other researchers (i.e. Terzopoulos, 1988; Li, 1991; Hsia, 2001) for surface generation.

Method	Geometric transformation	Similarity measure	Refinement
Grün (1985)	affine	least squared differences L_2-norm	various geometrical constraints, multi-image
Helava (1988)	affine	least squared differences L_2-norm	hierarchical
Miller and De Venezia (1992)	translation	cross-correlation	hierarchical
Norvelle (1992)	translation & window shaping	cross-correlation	iterative orthophoto refinement
Kölbl et al. (1994)	translation	cross-correlation	filtering by finite elements method
Bösemann (1994)	affine	least squared differences L_2-norm	-
Zhang et al. (1995)	translation	cross-correlation	relaxation, epipolar constraint
Calitz and Ruether (1996)	affine, projective	L_1-norm	-
Zhang and Miller (1997)	translation	cross-correlation	hierarchical, epipolar resampling on the fly
Wiman (1998)	translation	cross-correlation	iterative orthophoto refinement
Olson (2002)	translation	maximum likelihood of L_1-norm	-

Table 4.1: Examples of area based matching methods.

Models based on surface gradients (planar, curvilinear) potentially improve the conventional least squares approach (Mustaffar and Mitchell, 2001), but have proved to be more complex and more demanding in terms of a-priori information. In order to strengthen matching and improve computational stability and convergency, constraints are included in the least squares adjustment using the known interior and exterior orientation parameters and its effectiveness is demonstrated in Baltsavias (1991). Similar stabilizing effects from constraints are realized by 1-D image matching along epipolar lines (i.e. Baltsavias and Stallmann, 1992; D'Apuzzo, 2002). In the first approach, constraints are introduced through the collinearity equations, whereas in the second approach through the epipolar line equation. An object-space least squares matching method (Ebner and Heipke, 1988; Helava, 1988; Wrobel, 1991) has been introduced to relate information from image/s directly to an object space model (instead of an image-to-image mapping) by including height-, optical-, density-, sensor- and illumination model parameters. The model has been used by several researchers (i.e. Lo and Mulder, 1992; Diehl and Heipke, 1992; Holm, 1994; Bösemann, 1994), but has proved to be computational expensive and complex as more parameters have to be estimated, and thus also less stable. In addition, common problems for other methods arising from geometric disturbances (e.g. occlusions) can also not be compensated, even though the model is more complete.

Least squares matching has been extended for edge tracking (Grün and Agouris, 1994) and

Property		Cross-correlation		LSM
pull in range	+	large	-	small
scale/rotation sensitivity	-	high	0	medium
occlusion sensitivity	-	high	-	high
accuracy	+	high	+	high
pair-wise matching	+	yes	+	yes
multi-image matching	-	not available	+	yes

Table 4.2: Properties of normalized cross-correlation and LSM. Positive and negative properties are indicated with + and - respectively.

matching (Grün and Stallmann, 1992; Tseng and Schenk, 1992), as well as being applied to warped images for surface reconstruction (Schenk et al., 1990) and tie point extraction (Krupnik and Schenk, 1997). Warped images (or orthophotos) have also been used by others in an iterative approach (Norvelle, 1992; Wiman, 1998) to improve correlation of images. Least squares matching is considered to be the most precise among existing algorithms (sub-pixel accuracy in the range of 0.01-0.05 pixels for ideal targets), but it is sensitive to extreme radiometric differences, is computationally expensive compared to cross correlation, cannot fully model excessive geometric differences, and 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. Other approaches, combining area matching and phase images (produced by 2D Fourier kernels) are described in Weng (1993) and Jenkin and Jepson (1994). A summary of some properties of the two widely used approaches of cross correlation and LSM are listed in Table 4.2.

4.1.2 Feature-based matching

Feature based matching uses symbolic description of images for establishing correspondences. Using features instead of the original intensities allows a representation to be selected which is more invariant with respect to distortions, caused by e.g. illumination, reflectance and geometry. This makes feature based algorithms, in general, more robust than intensity based procedures. These advantages are paid for by a loss in spatial resolution as features do not cover the whole image with sufficient density and therefore are combined in most cases with area based methods to improve surface representation. Extracted features can be interest points, edges and higher level features. Interest points are usually corners (L-corners, T-junctions and Y-junctions) or centers of circular features and the operators can be divided into three different categories: contour-based (Asada and Brady, 1986; Medioni and Yasumoto, 1987; Mokhtarian and Suomela, 1998; Pikaz and Dinstein, 1994; Shilat et al., 1997), intensity-based (Förstner, 1994; Harris and Stephens, 1988; Heitger et al., 1992; Kitchen and Rosenfeld, 1982; Laganière, 1998; Luhmann and Ehlers, 1984; Moravec, 1980; Reisfeld et al., 1995; Smith and Brady, 1997; Tomasi and Kanade, 1991; Sojka, 2002) and parametric model-based (Brand and Mohr, 1994; Deriche and Blaszka, 1993; Gamble and Poggio, 1987; Parida et al., 1998; Rohr, 1992) methods. A number of these approaches have been evaluated in various publications (Deriche and Giraudon, 1991; Schmid et al., 2000), which examine their behavior based on a set of criteria and under changes in imaging conditions, i.e. image rotation, scale change, illumination, viewpoint change and noise. More attention is given to intensity based methods as they can be easily implemented, are less complex and can provide better localization accuracy compared to parametric model-based methods. As far as edge extraction is concerned, algorithms usually perform detection of edgels, which are later aggregated to form curved or straight edges. Detection of edges can be categorized in five groups: gradient-based (i.e. *Sobel*, *Prewitt* or the *Compass* operator (Ruzon and Tomasi, 2001), which based on color descriptors), surface-fitting algorithms (Haralick and Shapiro, 1992), model matching algorithms (Fua and Leclerc, 1990; Fua et al., 1998; Nalwa and Binford, 1986), Laplacian of Gaussian (Canny, 1983, 1986; Hildreth, 1985) and moment-based algorithms (Lyvers et al., 1989). Shan (1996) presents a qualitative comparison of edge detection algorithms, where the algorithms based on the zero crossings of the Gaussian are more reliable compared to the other groups. Iverson and Zucker (1995) included logical checks for the existence of edges in order to reduce the number of false positive edges detected with linear edge detectors (i.e. Canny). Other assessments (Heath et al., 1997; Baker and Nayar, 1999) on edge detection demonstrate that the Canny algorithm generally has the highest performance.

Wavelet theory has been also applied for feature extraction but has not been extensively used. Contour following, grouping, linking and tracking can be performed by Hough transform, salient edge point linking, dynamic programming and graph searching, search trees, relaxation linking, contour tracking and neural techniques. The paper of Gilson and Damper (1997) gives an insight into neural techniques. For further reading, the book of Ballard and Brown (1982) and the dissertations of Baltsavias (1991) and Henricsson (1996) can be used as comprehensive references for edge extraction and aggregation. As far as the matching of features is concerned, least squares has been used for edge matching (Grün and Stallmann, 1992; Maas, 1996; Tseng and Schenk, 1992). Alternatively, pose clustering (i.e. Stockman, 1987) and robust estimation[1] (i.e Kubik et al., 1987; Krzystek and Wild, 1992; Pilgrim, 1996) are used to estimate the set of transformation parameters between images or object descriptions by clustering in parameter space and maximum likelihood estimators (Huber, 1981), respectively. String matching (i.e Ohta and Kanade, 1987) is favorable when a sorted list of attributed features has to be compared with a sorted list of a reference object. Also, relational (or structural) matching (i.e Vosselman, 1992; Wang, 1998) is used when the symbolic description of the image contains their mutual relations in addition to the primitives and their attributes. Both the method proposed by Zhang based on feature bridging and global relaxation (i.e Zhang, 1990; Zhang et al., 1992, 2000) and the work that has been carried out by different groups introducing invariants to image matching (i.e Schmid and Mohr, 1995) is worthy of note. Förstner (1995) draws some conclusions with respect to some of the most important properties of feature-based matching with points, edges and blobs, and relational matching. Feature-based matching with points combined with the cross-correlation coefficient of neighboring points for measuring similarity, can not handle large rotations. Feature-based matching with edges although able to handle occlusions,

[1]The field of robust statistics has developed methods to handle outliers and these techniques have been applied to common problems in computer vision and photogrammetry. Robust approaches are applied for local image smoothing, classification surface reconstruction, segmentation, edge detection etc., for which an overview is presented in the article of Black and Rangarajan (1996).

is sensitive to rotations. In addition, feature-based matching with blobs or segments in general does not lead to the desired precision. Among all, relational matching is highlighted as the most powerful technique for handling the general setup, however, it is time consuming and no multi-image version is known.

4.2 Multi-image matching

The advantages of performing matching across multiple images, especially for 3D reconstruction in urban areas, have been pointed out in different studies. These relate to the growing robustness of the automatic matching procedure against different sources of problems through the complementarity of the surface local visibility in object space and redundancy of multiple view points (Baltsavias and Stallmann, 1992; Tahir and Schenk, 1992; Henricsson and Baltsavias, 1997; Paparoditis et al., 2001; Roux and Maître, 2001; Schlueter and Wrobel, 1996; Schlueter, 1998), and is valid for both area and feature based reconstruction strategies. It worths noting that the term multi-image matching is utilized in the literature in different approaches. Although under this term, matching is generalized to any number of images greater than two, these images can be matched in pairwise mode or simultaneously, plus matching can be performed from and to all images, namely all images can be used both as template/reference and search images. Collins (1996) introduced the term *true* multi-image matching and three conditions that should be satisfied for a matching method to be called a *true* multi-image approach. In addition to the number of images, the method must be of linear algorithmic complexity (O(n)) and all images should be used in an equal manner.

Okutomi and Kanade (1993) describe a multi-baseline stereo method for producing a dense depth map from multiple images by performing pairwise stereo matching on all pairs of images and combining the results. Although they show convincingly that integrating information from multiple images is effective in reducing matching ambiguity, the main disadvantage of the approach is the time efficiency. This strategy was later changed to combining matching results between one reference image and all other images (Kanade , 1994), but since only one image is used as a reference view any areas of the scene that are occluded in that image can not be reconstructed. Similarly, many other approaches attempt to solve the multi-image matching problem by using one image as reference and by splitting the set into pairs of images that are processed pairwise. One of the main reasons is the straightforward integration of the well-known matching constraints (see also section 5.1.1) based on the epipolar geometry of two the views. But there are also methods that have used extensions of the geometrical constraints to multiple images. For example Shashua (1994) presents a trilinear constraint where points in three images can be the projections of a single 3D scene point if and only if an algebraic function vanishes. Hartley (1994) devised a similar constraint for lines in three views, whereas Triggs (1995) provided a quadrilinear relation for points in four views. However, in the above approaches the number of images is limited and more flexible methods with respect to the number of images exist, such as the multiphoto geometrically constraint matching approach of Baltsavias (1991) and the methods based on object-space least squares that were discussed in section 4.1.1. Fua and Leclerc (1995) describe an approach for object-centered reconstruction via image energy minimization, where 3D surface mesh representations are directly reconstructed from multiple intensity images. Triangular surface elements are adjusted so that their projected appearance in all the images is as similar as possible to the observed image intensities, while still maintaining a consistent shape in object-space. One thing that the true multi-image matching/reconstruction methods above have in common is the explicit reconstruction of a surface or features in object space, simultaneous with the determination of image correspondences. Though, object space approaches, as already mentioned in section 4.1.1, involve optimization problems with a large number of parameters, and require good initial estimates to reliably reach convergence. Maas (1996) proposed in addition to the use of many images also the use of each one as a fixed reference in the search of conjugate points. The proposed methodology extracts interest points from the reference image and uses the intersection of epipolar lines in the remaining images to extract their exact corresponding point. Ambiguous points are traced through the number of successful verified matched points and the range constraints of their respective ground coordinates. More recently, Brown et. al. (2004) developed a feature space outlier rejection strategy that uses in pairwise mode all of the available images in an n-image matching problem based on the noise statistics of correct/ incorrect matches give background distribution for incorrect matches.

4.3 Image Matching for Airborne Linear CCD Sensors

The majority of matching methods are geared towards frame imagery. These methods can be also employed for images acquired by linear CCD sensors, but without exploiting necessarily their special characteristics. Linescan imagery is handled similarly to frame imagery, whereas in most cases, matching is performed in a simplified pair-wise mode. Less investigations and even less publications exist on the matching of images from airborne linear array CCD sensors. Wiedemann et al. (1996) focus mainly on image matching for tie-point extraction utilizing images from the WAOSS and HRSC cameras. The approach is based on combining existing methods of feature- and area- based matching, namely correlation of interest points extracted with the Moravec operator, and subsequently refinement by least squares matching in conjunction with a hierarchical matching strategy. Haala et al. (2000) uses images from the DPA, HRSC and WAAC in order to derive conclusions on their accuracy potential in photogrammetric processes. Rectified images were used for DSM generation, but the information provided is restricted to geometric transformations for point determination and does not cover the matching algorithm. ? used the concept of object space-based image matching, described in Schlueter and Wrobel (1996) and applied for scanned large-scale aerial images, with data from the DPA sensor. The approach requires that pixels are projected into object space and the mathematical model integrates pixel and surface grey values, radiometric parameters and interior and exterior orientation parameters. As a result the model is relative complex and cannot be used straightforward for rectified images. For the later case, an iterative procedure has to be utilized to establish the relation between image space and object space and thus the process is computationally expensive considering the derivation of dense point clouds for surface reconstruction. His approach was only experimentally used and due to deficiencies in the absolute exterior orientations it was not possible to fully explore the matching accuracies. For HRSC-A images, the method used by Wewel (1996) in the initial tests with HRSC images for DSM generation was applied for the data from the airborne camera slightly modified (Hese et al., 1999). Grid point matching based on cross-correlation and LSM was applied to four image pairs, utilizing in parallel multi-resolution levels to improve initial approximations, but no information is given on other matching aspects, i.e. concerning geometrical constraints, quality control, etc. More recently, Elberink and Vosselman (2002) used ADS40 images for 3D line reconstruction and Zhang and Grün (2003) developed a matcher for the Starimager sensor of Starlabo based on features and grid points (area patches), utilizing three or more stereo images (the Green channel is usually selected from the 3x3 configuration of RGB CCD lines). Geometrical constraints are enforced through both the known sensor model and orientation parameters, and these are implemented in the multi-photo geometrically constrained least squares matching (MPGC). The relaxation technique, used in a grid point matching module, has been also employed by Wu and Murai (1997) for the predecessor of Starimager, the experimental system TLS. Moreover, region matching based on area correlation and straight edge matching has been used for building extraction, by fusing laser and TLS image data (Nakagawa et al., 2002). The ISTAR company has also performed several tests on automatic extraction of DSMs from airborne digital cameras (Nonin, 2003) using their in-house developed algorithm, based on grid matching by normalized cross correlation. But, little information on this has so far been published. The paper of Gabet et al. (1997) can be used as reference for the general, not particularly exploiting the characteristics of airborne linear CCDs, algorithmic approach, used in the ISTAR system. For the ADS40, to date an algorithm used mainly for the processing of frame imagery, included in the Socet Set photogrammetric workstation, is being routinely used for tie-point and DSM extraction with slight improvements in terms of speed. Pasco Corp., is currently performing investigations on DSM extraction from ADS40 images. The adopted approach uses elements of AIM and to date the system has not been extensively evaluated. Currently, the import of image data from airborne linear arrays to existing DPWs, is still in an ongoing development phase, however, these images are processed by already existing built-in algorithms of the systems. In other words, advancements in matching do not particularly follow developments and investigations on sensor modelling.

4.4 General approach of AIM

In this section, a general description of the newly developed AIM algorithm is presented and the individual modules are analyzed in more detail in the next chapter. The aim of our research is to develop a matching procedure that can lead to reliable, precise and complete results, while being flexible enough to be applied for other sensors, i.e. frame imagery. Moreover, since the raw image data (Lev0) are less suitable for DSM extraction (section 2.3.3), matching is applied on the rectified images (Lev1) and the matching aspects that we have investigated include:

1. Incorporation of a priori known geometric information about the sensor orientation and image rectification in matching, so that the search space for corresponding features in the three or more images, which is normally 2D, is reduced. For frame cameras and line sensors on satellites, implementation of geometrical constraints is a solved problem and the search space is reduced from 2D to 1D. For airborne linear CCDs, implementation of geometrical constraints is more complex and different approaches need to be investigated (section 5.1).

2. Provision of approximate values by hierarchical methods, considering the fact that minimum information about the terrain is available (i.e average terrain height). Although hierarchical methods are routinely used in matching algorithms, problems such as propagation of mismatches and the selection of the filter used to generate the pyramid levels (section 5.2) are further investigated and a improved hierarchical matching strategy is presented.
3. Performing investigations on the selection of optimal features and on the combination of different feature primitives in each pyramid level also taking into consideration the surface characteristics of the processed area. Our focus is on retrieving a more complete description of the scene and reducing smoothing of discontinuities. Apart from the density, the reliability and the localization of extracted features are analyzed (section 5.3).
4. Combination of different matching methods and parameters, in relation to the matching entities, the accuracy of approximations and the surface form. Matching techniques, especially those which have been commercialized have in most cases[2] restricted themselves to one of the two common matching classes, ABM or FBM. ABM techniques smooth surface discontinuities (e.g. buildings), while FBM techniques tend to perform worse when the image texture is weak or strong but with many small, dense or repetitive patterns. In the reported approach, such techniques are combined (hybrid approach), using also information on the image texture and terrain roughness from previous coarser matching. For each matched feature, critical parameters such as patch size, shape and orientation, number of parameters for the geometric transformation between template and patch are allowed to vary. The decision on these matching parameters is through an analysis of the local image texture and surface form (section 5.4).
5. For the 3D reconstruction, especially in urban areas, in addition to the inclusion matching of features, matching across multiple images has the advantage of complementarity and redundancy of multiple view points (Baltsavias and Stallmann, 1992; Tahir and Schenk, 1992; Henricsson and Baltsavias, 1997; Paparoditis et al., 2001; Roux and Maître, 2001). In AIM, the simultaneous use of three or more available images in a multi-photo approach takes into consideration the FP design and the viewing angles of the CCDs. Colour lines are combined with 3-fold panchromatic imagery in order to disambiguate difficult matching cases. In most matching systems, one image is selected as a template (usually the one with less geometric differences to other images) but in cases of occluded areas, the surface reconstruction can be partly incomplete. Therefore, we extend the matching strategy by including more than one template image at run time, especially in the case of rough terrain and large occlusions. The multi-template approach is described and justified and compared with the single template strategy (section 5.4.3).
6. Self-diagnosis and error detection. Each matching algorithm provides mathematically and statistically well-founded criteria to judge the matching quality. Since no single criterion is reliable, a combination of multiple criteria is necessary. The criteria that are derived during run time are used to identify possible mismatches that belong to different error types, like multiple solutions and occlusions (section 5.5).

[2]There are some approaches, in which ABM and FBM techniques are combined (e.g Baillard et al., 1998), but they have not been extensively analyzed and usually the raw matching results are post-processed to improve the surface description and completness.

7. Consideration of computational aspects. Although they are not important from a scientific point of view, they still play a role in a production environment. However, since computing power is nearly doubling every year, while time requirements for manual editing are nearly constant, the main effort is on reliable, highly successful matching, with the aim being to cut editing time.

The above described features of AIM, are embedded in the matching procedure as shown in Figure 4.1. Edgels and grid points are extracted and the first are further aggregated to form edges (section 5.3.2). Edges are then assigned to two different classes, straight edges longer than 11 pixels and all other edges (straight edges shorter than 11 pixels, curved and closed edges). The above two types of edges, along with the grid points and edgels form the four different classes/cases, as illustrated in Figure 4.2, for which the matching procedure varies (Figure 4.3). Multi-patch size matching (MP matching) and LSM are combined, while the quality control is applied after each matching stage. In the case of edge features two different approaches of LSM are used, the Edgel LSM and Edge LSM (section 5.4.4). Geometrical constraints are enforced by means of *quasi*-epipolar curves to strengthen matching (section 5.1). The results are iteratively refined using the image pyramids and the doublet strategy (section 5.2.3). The overall matching approach can also be embedded in a single- (ST) or multi-template (MT) strategy.

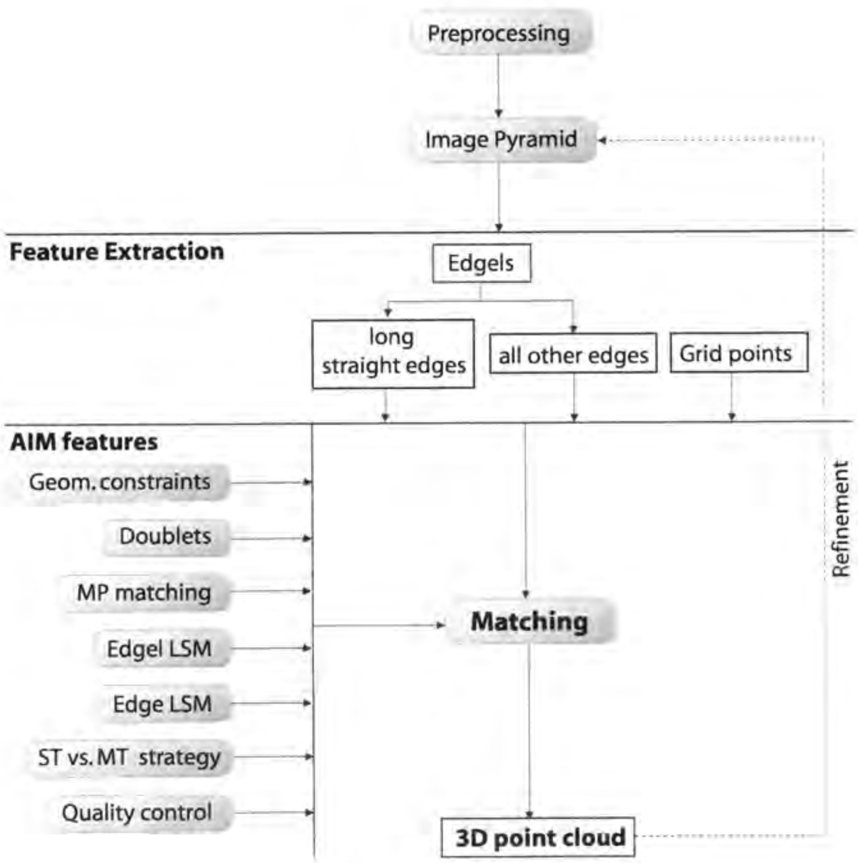


Figure 4.1: Matching procedure and AIM features.

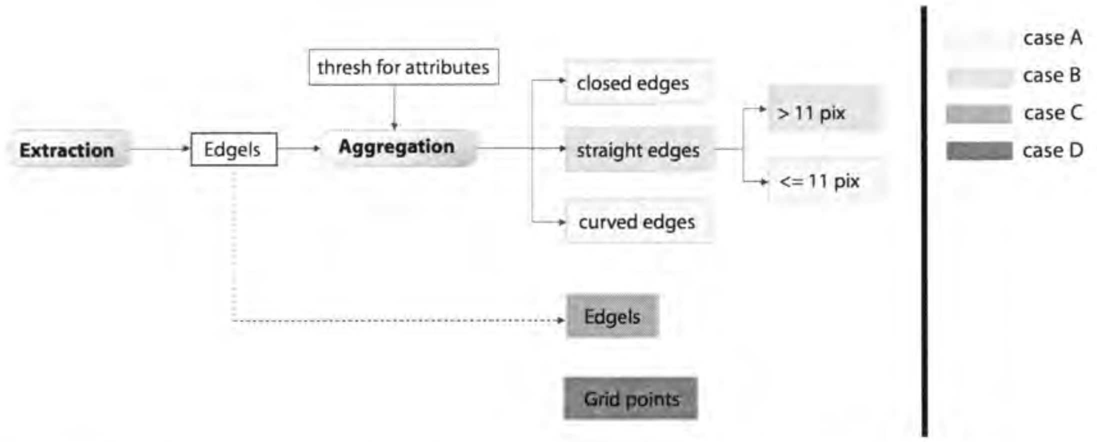


Figure 4.2: Extraction procedure and selected features. For each case a different matching scheme is used.

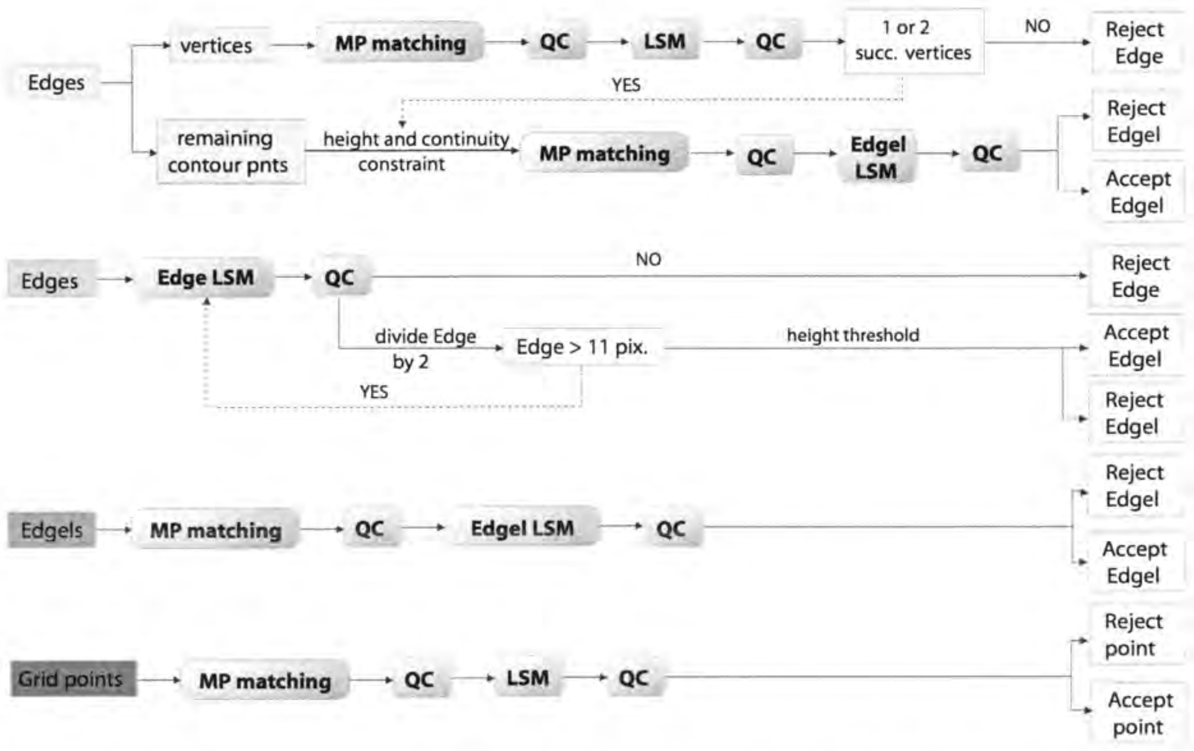


Figure 4.3: Matching options according to features. The abbreviations QC, MP matching and LSM are used as abbreviations for quality control, multi patch size matching and least squares matching respectively.

Chapter 5

Adaptive Image Matching (AIM)

5.1 Geometrical constraints

The exploitation of the a priori known geometric information to constrain the matching solution, along with its benefits, has been recognized by various researchers, as mentioned in section 4.1. Geometrical constraints have been proved to strengthen matching in terms of precision and reliability and these aspects have been extensively analyzed for frame imagery. Unlike frame-based imagery, linear arrays have a number of perspective centers equal to the number of acquired scan lines, since lines are collected at different instances of time. Epipolar lines do not really exist as every scan line has its own position and attitude. Orientation cannot be easily modelled for the entire length of the image as the trajectory in airborne linear arrays is not easily predictable. Thus, implementation of geometrical constraints in the algorithm has a higher degree of complexity, as will be shown in the following sections.

5.1.1 From epipolar lines to quasi-epipolar curves

For frame-based imagery, geometrical constraints can be introduced e.g. by means of epipolar lines. In Figure 5.1(a), given a point P on the template image the corresponding points on the search image must lie on the *epipolar line*. This is determined as the intersection of the search image plane with the plane formed by $\overrightarrow{O_1P}, \overrightarrow{O_1O_2}$, the *epipolar plane*, where O_1, O_2 are the projection centers of the two images and $\overrightarrow{O_1O_2}$ is the base vector. By intersecting $\overrightarrow{O_1P}$ with different height planes (at least two) and back projecting onto the search image, the epipolar line on the search image is defined. Considering the fact that images acquired from linear array CCD sensors consist of a series of lines, the points on the search image will not share the same projection center, as shown in Figure 5.1(b). Points P', P'', P''' are imaged through the O_{2a}, O_{2b}, O_{2c} projection centers respectively. In this sense, for each object point epipolar lines will exist for each separate scan line and the trajectory,of these short epipolar lines, can be approximated and modelled by a polynomial, or a curve. The term *quasi*-epipolar curves is therefore used.

In sections 5.1.2 and 5.1.3, *quasi*- epipolar curves are analyzed for both Lev0 and Lev1 images and different approaches are adopted regarding the implementation of geometrical constraints. As access

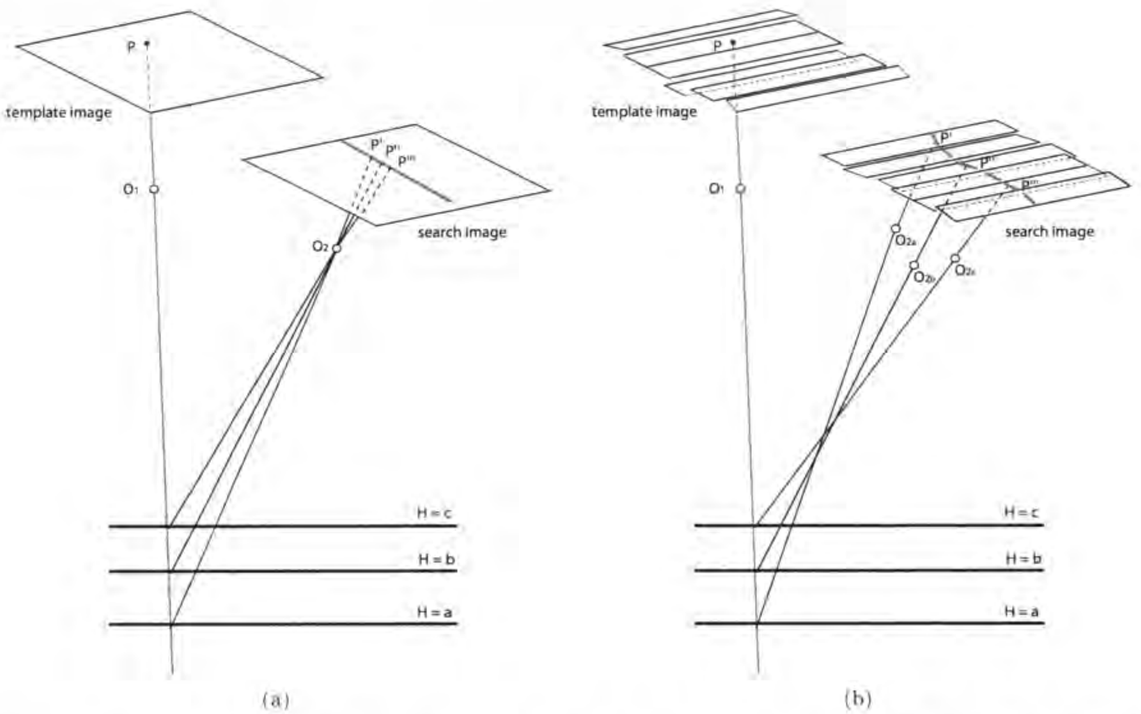


Figure 5.1: Epipolar and *quasi*-epipolar geometry. On the left (a), the epipolar line for frame imagery is the straight line on the search image and on the right (b) the *quasi*-epipolar curve for scanline imagery is indicated with the dashed line on the search image. For (b) epipolar lines are finite and defined for each scanline and when combined the *quasi*-epipolar curve is defined.

to the sensor model is limited to transformations between coordinate systems and other sensor model data are regarded as proprietary information, geometrical constraints can only be enforced by the indirect methods explained below. In order to analyze the *quasi*-epipolar curves the following approach is used: a number of well distributed ground control points is projected onto the template image, and by iterative alteration of the height along the ray of the template, the trajectories of *quasi*-epipolar curves are simulated. Usually a step of 0.5 or 1 meter is used between the different height planes.

5.1.2 Lev0 images

Nine ground points are projected on the Lev0 Pan Nadir image[1] and their distribution is illustrated in Figure 5.2. In Figure 5.4(a) the trajectories of the nine points are shown. The points have been selected on the Nadir image and projected onto the Backward image. Height planes are selected with a step of 1 m and with maximum absolute difference relative to the reference height (Z coordinate of GCP) of 100 m. The horizontal axis coincides with the flight direction and the vertical with the pixel position along the scanline. The computed x and y coordinates are relative to the minimum x and y coordinates for each trajectory. As it can be seen, the forms of the *quasi*-epipolar curves for Lev0 images follow irregular and random forms, due to the unpredictable motions of the aircraft. When smooth flights occur or when the stabilizing platform can compensate for these motions to a large extent, smoother trajectories

[1] In these tests the images from strip 1052 of the Waldkirch block are used (section 6.2)

in the raw images are visible. It worths noting here that these aircraft motions or also terrain undulation can deteriorate the image quality and therefore the line readout rate, namely the integration time, was adjusted to produce a double number of sampling points in flight direction, which are then resampled in Lev1 image (compare number of lines in Lev0 and Lev1 images in Figures 5.4(a) and 5.4(b) respectively).

With an increasing trajectory length, its geometrical modelling becomes more complex and consequently introducing a global model for long trajectories becomes a difficult task. In AIM, we enforce the geometrical constraints for Lev0 by utilizing the two approaches described below.

5.1.2.1 Iterative approach

The iterative approach is quite general and can be applied to all types of image data. An approximate height, if not available, is computed by a forward intersection for the initial points on the template and search images in pairwise mode (using the template and each available search image). The search range of matching in image space is transformed to a height search range on the ray of the template. The selected points for matching are back projections of 3D points along this ray, within the height search range, and with a height step that corresponds to one pixel step in image space. This approach has the advantage that no geometrical modelling of the *quasi*-epipolar curve on each search image is required. However, as the ADS40 sensor model is an iterative model (section 2.3.4) with relatively complex transformations, time performance decreases proportionally with the increase of the search range. This is justified, as for every point on the template image, an image to ground and a ground to image transformation is performed to retrieve the point position on the search image. This approach can be used when no geometrical modelling has to be introduced in the matching algorithms, such as cross-correlation, but not in the case of least squares matching with additional geometrical constraints (see section 5.4.3.2).

5.1.2.2 Polynomial model

To overcome the above mentioned drawbacks and approximate the trajectories by means of analytical functions, polynomial models are utilized. As illustrated in Figures 5.3(a), 5.3(c) and 5.3(e), the epipolar trajectory of an object point is approximated by a first- (red line) and a second degree polynomial (green curve) using an increasing number of points (9, 15 and 20 points, corresponding to a search range in line direction of about 17, 30 and 50 pixels). The residual graphs in Figures 5.3(b), 5.3(d) and 5.3(f) show the absolute differences between modelled and true position for the selected points. With an increasing number of points, the modelling of the trajectory can be less accurate. For the trajectory of Figure 5.3 the maximum error for the first and the second degree polynomial approximation is around 0.8 and 0.1 pixels, respectively. In general, these errors may increase or decrease depending on the stability of the sensor platform and the number of points (length) used in the curve modelling. As shown in Figure 5.3(b) the modelling errors of the epipolar curve for this object point and a search range of ca. 20 pixels is less than 0.21 and 0.05 pixels for the first and second degree polynomial respectively. Therefore, the search range should be kept small (less than 10 pixels), especially at the original resolution level. As also Figure 5.4(a) shows within 10 lines the trajectory can be modelled well, at least by a second degree polynomial. In least squares matching, less strict weights can be used for the constraint equation in order to permit slight deviation of the solution from the approximated *quasi*-epipolar curve (see section 5.4.3).

5.1.3 Lev1 images

In Figure 5.4(b) the trajectories of the nine points, previously projected on Lev0 images, are computed for the Lev1 images. The settings for the selection of height planes and the maximum absolute difference relative to the reference height remain unchanged. It is observed from Figure 5.4(c) that the *quasi*-epipolar curves of Figure 5.4(b) can be fitted to a straight line and the resulting residuals are equal or less than 0.25 pixel for the whole length of the trajectory of ca. 250 lines. Over shorter lengths of ca. 40 lines the residuals are less than 0.05 pixels. When Lev1 images are utilized in matching both the iterative approach and the polynomials can be used. However, the first method is disregarded as it is time consuming, whereas the second can sufficiently model the epipolar curve even with increasing number of points.

5.1.3.1 Straight line approximation

Several trajectories have been investigated and it has been proved that *quasi*-epipolar curves can be approximated by straight lines. For a search range of 20 and 40 pixels, the maximum error is less than 0.01 and less than 0.05 pixels. A first order polynomial fit is utilized in order to derive the line parameters for the trajectory. As deviations are small, only a minimum number of points, at the two extremes of the search range and in the middle, is used for the line fitting. As a result, the trajectories in Lev1 images can be modelled more easily and the images can be processed faster than Lev0 images.

5.1.4 Practical aspects and comparison

We analyzed different methods that utilize geometric information and constrain the matching solution both for the Lev0 and Lev1 images. The iterative approach has the advantage of being the more general and accurate method compared to the polynomial modelling. However, it is a time consuming process as for each point on the search image an image to ground and a ground to image transformation has to be performed, and in addition it can not be used in a straightforward manner in algorithms that incorporate additional modelling of the geometric information. Polynomial modelling is favored for both image types, Lev0 and Lev1, as it can sufficiently model the trajectory.

In the case of Lev0 images more than three points are needed along the trajectory to retrieve the polynomial parameters. Additionally, the weights of the constraints have to be more relaxed (see section 5.4.3) in order to permit convergence of the solution to the true position and to compensate for small residuals. For most parts of the trajectory and for a length of 25 pixels the residuals are less than 0.1 pixels, but these small errors can be of a maximum of 0.25 pixels for certain parts. Especially for Lev0 images one should take also into account whether the original or the adjusted orientation is used, i.e. for the first case the deviation of the *quasi*-epipolar line from its true position. The weights should be adjusted accordingly to compensate residuals in the direction perpendicular to the *quasi*-epipolar curve of a maximum of 5 pixels. When multi-resolution levels are used to derive approximate values (see section

5.2), a first degree polynomial suffices for the upper levels and second order is utilized in the two lower levels. For the simpler case of Lev1 images, the straight line approximation is used in all levels of the pyramid.

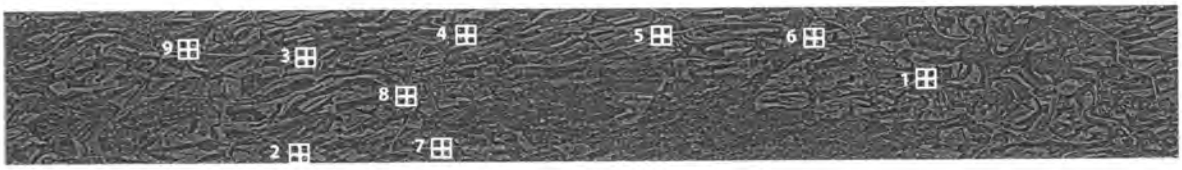
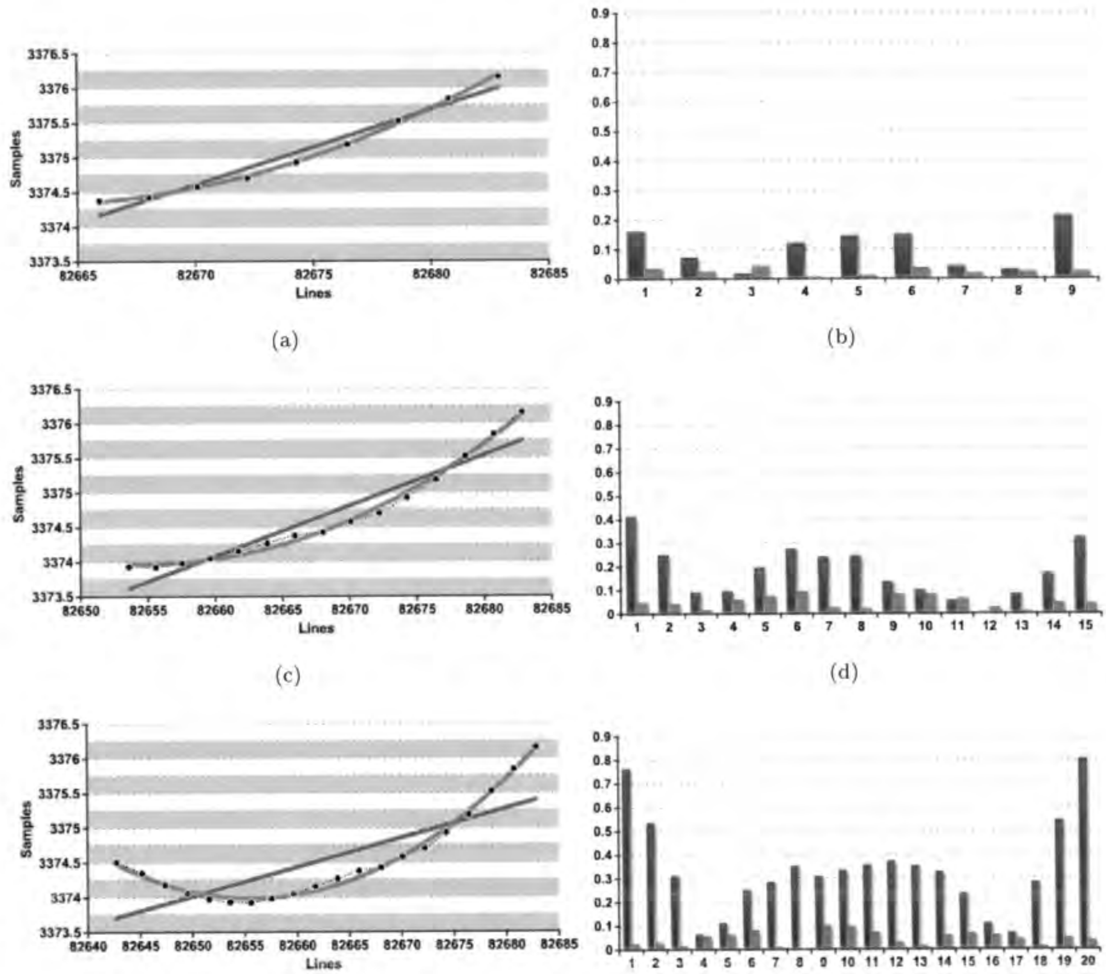


Figure 5.2: Distribution of GCPs on the Lev0 Pan Nadir image.

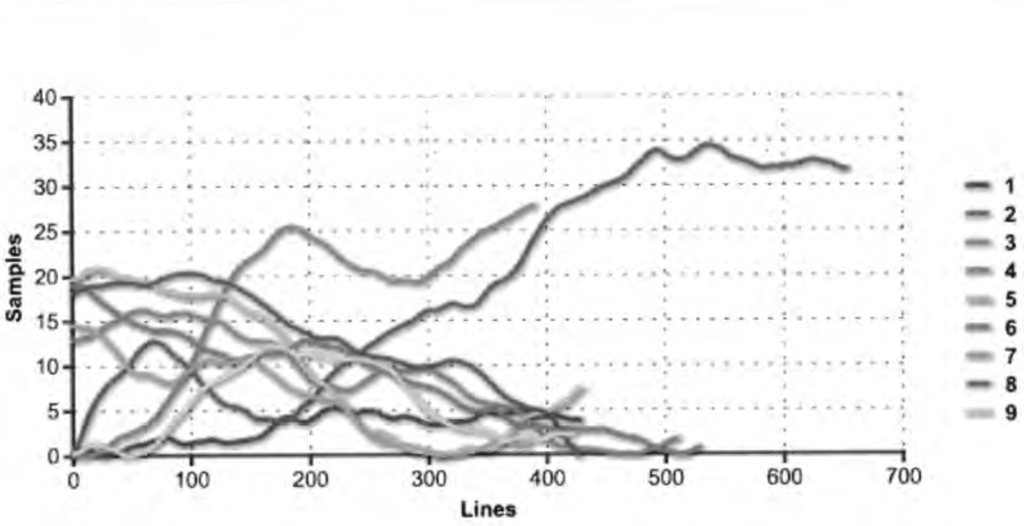


(a) (b)

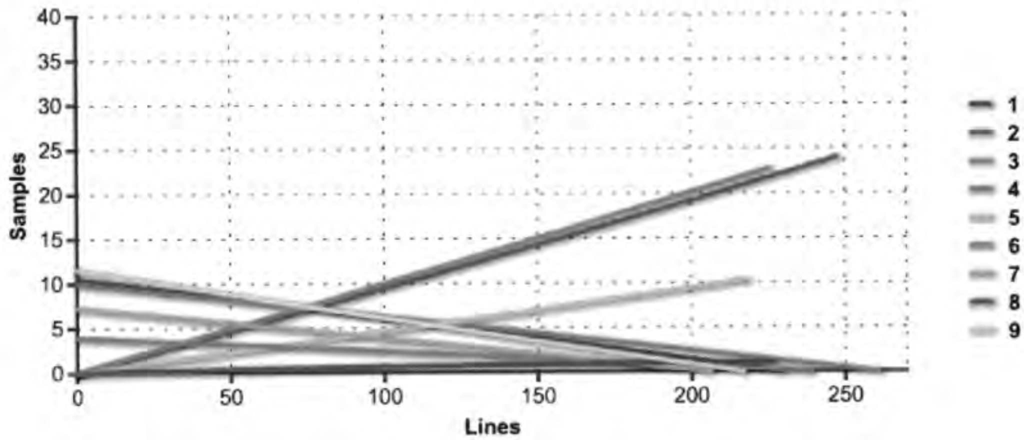
(c) (d)

(e) (f)

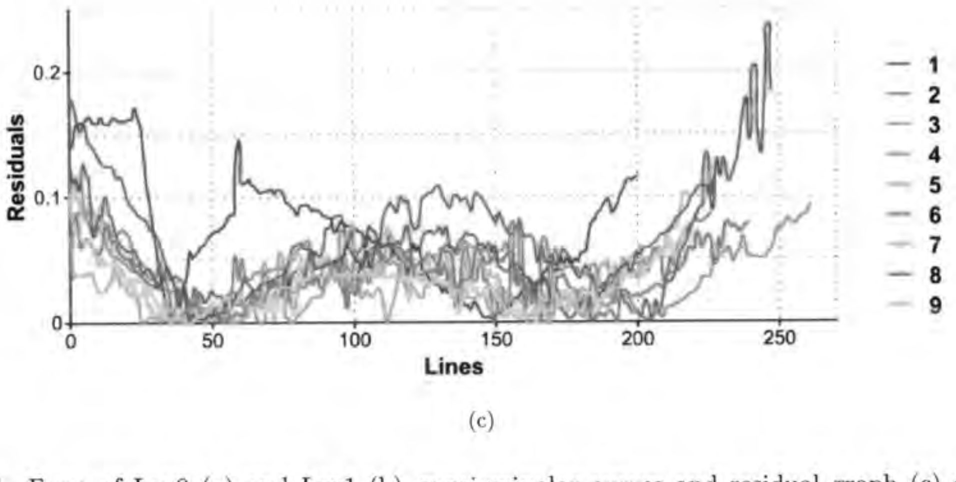
Figure 5.3: Polynomial approximation of *quasi*-epipolar curves. For the graphs on the left (a, c and e) the horizontal axis coincides with the flying direction and vertical one with the direction along the scan line. The approximations of the epipolar trajectory by a first- and second degree polynomial are indicated by the line and the curve. The corresponding residual graphs (b, d and f) show the absolute differences of the modelled from the real pixel coordinates of the epipolar curve. For each point, the left and right bar indicate the residuals from the first and second degree polynomial.



(a)



(b)



(c)

Figure 5.4: Form of Lev0 (a) and Lev1 (b) *quasi*-epipolar curves and residual graph (c) of the Lev1 *quasi*-epipolar curves. The *quasi*-epipolar curves were computed from one image pair (Pan Nadir and Backward). The nine points were selected on the Pan Nadir and projected onto the Pan Backward. The Lev1 *quasi*-epipolar curves (b) have been modelled by a first degree polynomial, computed for the whole length of ca. 250 lines. The vertical axis of Figure (c) shows the absolute differences of the modelled from the real pixel coordinates of the Lev1 *quasi*-epipolar curve.

5.2 Derivation of approximate values using image pyramids

For the derivation of approximate values for image matching, different approaches exist. Among them, the use of a-priori information of the surface, if available, is extensively used, for example manually measured seed points or a coarse DSM or DTM. Moreover, most matching methods use the concept of image pyramids or, more generally, hierarchical techniques[2].

5.2.1 Introduction

Hierarchical techniques are explained in a large number of reports in the current literature, from which the reader can find detailed information on this topic. The publications of Ackermann and Hahn (1991) and Baltsavias (1991), for example, can be used as reference for further reading. The main idea of hierarchical techniques in matching and object reconstruction is to generate a series of images of coarser resolution, or an image pyramid, and apply matching in one pyramid level after the other, starting from the upper one. A coarser resolution is equivalent to a smaller image scale and a larger pixel size. Therefore, as resolution decreases, local disturbances such as occlusions become less of a problem. The approximate surface, generated from the matching results in the upper pyramid level, is used to reduce the search space and improve the approximations in the lower level. This coarse-to-fine strategy is iteratively applied until the level at full resolution is reached. In addition, another coarse to fine approach, currently used in the DPWs of SS and Virtuozo, is based on matching of object points, or points that are extracted in the template image of full resolution, in all levels from top to bottom. The main drawback of the later, is that the respective point positions in higher levels of the image pyramid may not correspond to a feature or may lie in areas that are less suitable for matching (e.g. textureless areas). Even though image pyramids are widely known and are frequently used, there are certain interrelated aspects that have to be considered for the derivation of approximations:

1. The main disadvantage of image pyramids is that texture may disappear in the upper levels. This is essentially harmful when it happens in the low pyramid levels, and the approximations in the upper levels come from distant features.
2. The number of pyramid levels and the reduction factor, if not set optimally, may increase matching error propagation from the previous pyramid levels and in parallel may result in increase of processing time.
3. Approximations also depend on the density and type of the selected features to be matched in each level and the size of the search window (see also section 5.3).
4. The quality of approximations is also related to the matching algorithm employed. For example area based matching, in general requires better initial approximations compared to feature based matching. Matching algorithms that are not very accurate but are robust and do not need very accurate approximations, can be used in the upper levels and with less pyramid levels.

According to the above, three major issues in hierarchical matching are distinguished and further analyzed: a) generation of image pyramids, b) the hierarchical matching strategy and c) the interpolation

[2]other known terms for hierarchical techniques are: multi-resolution representation, scale-space representation, image decomposition, frequency analysis, etc.

algorithm for approximate values in the lower levels.

5.2.2 Image pyramid generation

For the generation of image pyramids, an appropriate filter has to be selected, in order to avoid smoothing and/or loss of texture in the upper levels. Radiometric differences among the available images at each pyramid level should be minimized to improve matching performance and, most important, the geometry should not be distorted by the filters. Shown in Figure 5.5 are four pyramid levels, generated using four different reduction filters with a decimation step of 2 (same for all image pyramids). Apart from the 3x3 local average (Figure 5.5(a)) and the 3x3 Gaussian (Figure 5.5(b)) filters, the adaptive filters (Figures 5.5(c) and 5.5(d)), mentioned in section 3.2.1, were also examined. As the last two filters, employed for noise reduction, aim at preserving details, with less smoothing than conventional filters, they are used in image pyramid generation and evaluated. For up to three pyramid levels, edges are sharper, finer details of the object shape are visible and contrast is higher. However, in the upper pyramid levels although contrast is higher, the image content and the objects become very discontinuous and start "breaking apart"(see right image in Figure 5.5(c) and 5.5(d)). For ADS40 images, measured parallax differences in urban areas justified the selection of five or more pyramid levels in hierarchical matching, and as a result the adaptive edge preserving and fuzzy filter have been disregarded due to this effect in the upper pyramid levels. Comparing the 3x3 Average and Gaussian filters, although small differences were observed, Gaussian led to less smoothing and blurring of objects.

The processing of images with a Wallis filter for contrast enhancement and radiometric balancing has already been discussed in section 3.2.2. In hierarchical matching, the Wallis filter can be applied in each pyramid level with common parameters for all available channels to be matched to minimize radiometric differences. If the Wallis filter is applied only in the full resolution level, prior to the generation of pyramid levels, texture tends to disappear in the upper levels and the image layers are more homogeneous (Figure 5.6). Therefore, first the image pyramid is generated and then the Wallis filter is applied in each level individually. Filter parameters are not kept constant for all pyramid levels due to the different image resolution and size of objects in each pyramid level, but not all filter parameters have to be adapted for each level. It suffices if the Wallis parameters have been set properly at full resolution. Then, we only adapt the window size in each pyramid level reducing it from level to level, but less than the decimation step and partly the standard deviation and to approximately the size of imaged objects.

Summarizing on image pyramids, we use the 3x3 Gaussian filter and a decimation step of 2 to generate levels of coarser resolution after we have employed the noise reduction processing for the full resolution image. Then, we process each individual image level with the Wallis filter using different window sizes. Other aspects related to the strategy of hierarchical matching are handled in the next section (5.2.3). The number of pyramid levels used is determined by the user. Measured parallax differences in all processed datasets justified the following conclusions with respect to the number of pyramid levels that should be utilized in matching: For relative flat areas six pyramid levels suffice, whereas for urban areas seven pyramid levels should be used. The aim is to avoid propagation of mismatches to the lower levels, to retrieve a surface which suffices as an initial approximation at each matching level, and to

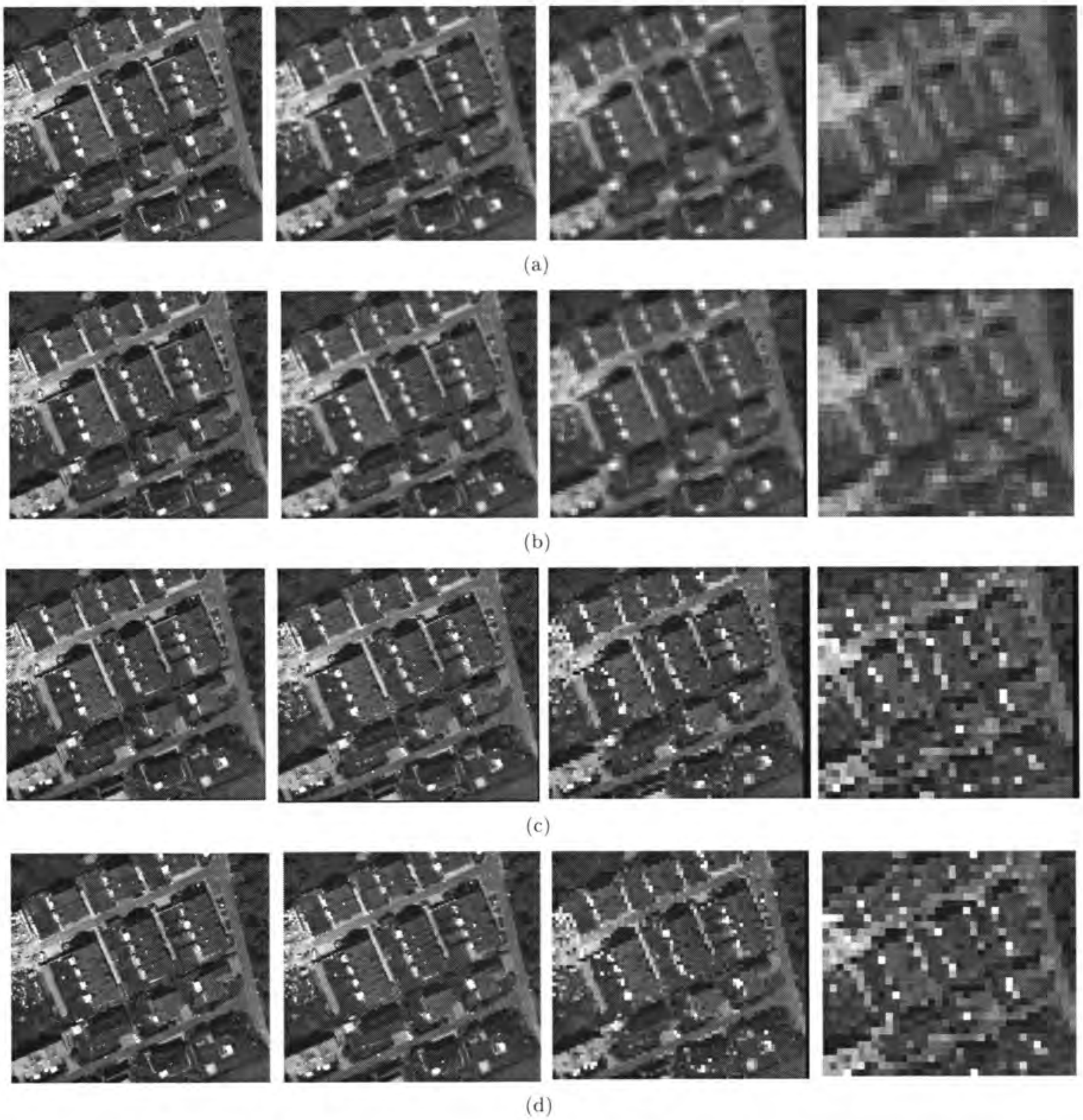
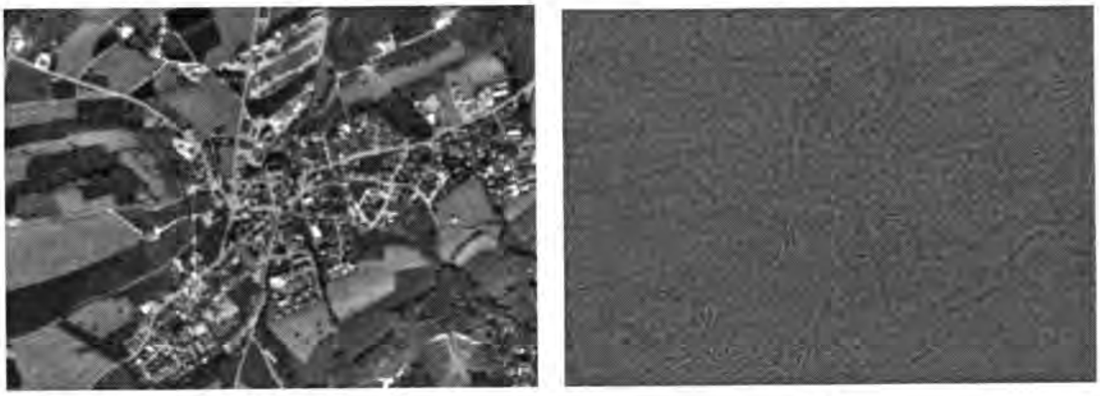


Figure 5.5: Image pyramids generated by different filters. Part of a Nadir ADS40 image is processed by (a) 3x3 local average, (b) 3x3 Gaussian, (c) adaptive edge preserving smoothing and (d) fuzzy filter. Resolution decreases by a factor of two (2) from left to right, starting from the second pyramid level.

reduce interpolation time between subsequent levels.

5.2.3 Doublets

Among the two general coarse-to-fine strategies (section 5.2.1), the strategy based on the consecutive matching of selected object positions (or template positions) in all pyramid levels is less optimal. The final number of matched points can be significantly less to the number of initially selected points, e.g. if the point is not matched in the levels of coarser resolution and approximate values are not derived from neighboring points, by means of interpolation, the object point is rejected. Moreover, processing time



(a) (b)

Figure 5.6: Wallis filter and image pyramids. In both images (a) and (b) the 5th pyramid level with Wallis filtering is illustrated. For the left image the Wallis filter has been employed on the 5th level, whereas for the right image the Wallis filter has been employed only on the full resolution level and prior to pyramid generation.

and memory can increase significantly since the positions in the coarser levels must be traced and linked to the remaining levels for every individual point. If features (interest points or edges) are used, then the extracted features in all levels must be stored at significant memory cost. However, most matching methods utilize the second approach, namely they employ a time consuming interpolation to pass coarse matching results to the lower levels. To overcome this, we utilize the doublet strategy in hierarchical matching with the aim to reduce processing time and interpolation of matching results from each pyramid level to the next, but most important to reduce propagation of matching errors to lower levels. Doublets are formed from two consecutive pyramid levels A and B, illustrated in Figure 5.7. Extraction of features is performed in the lower level B of a doublet. These features are then transferred one level up (A), and are retained only if an extracted feature also exists on level A (in the case of interest points or edges). Equation 5.1 is used to retrieve the point position in coarser levels of the pyramid and is further simplified for the doublet according to Equation 5.2. Connection is established between points on levels A and B.

$$x_{upper} = \frac{x_{lower}}{s^m} - \sum_{i=0}^{m-1} \frac{x_k}{s^{(m-i)}} + \sum_{i=1}^{m} \frac{x_o}{s^{(m-i)}} \tag{5.1}$$

$$x_A = \frac{x_B}{s} - \frac{x_k}{s} + x_o \tag{5.2}$$

The y coordinates are computed similarly to the x coordinates, according to the above equations. x_{upper}, y_{upper} and x_{lower}, y_{lower} are the pixel coordinates in the upper and lower level respectively. m is the level difference between upper and lower, s is the decimation factor, x_o, y_o are the pixel coordinates of the center of the upper left pixel and x_k, y_k are the pixel coordinates of the filter mask center on the parent level. In Equation 5.2, x_A, y_A and x_B, y_B are the pixel coordinates in levels A and B, m is set to 1 and s is set to 2.

Matching in all pairs of levels from top to bottom is then performed. Propagation of matching errors in level A to neighboring points in level B is avoided. For example, when 6 pyramid levels are

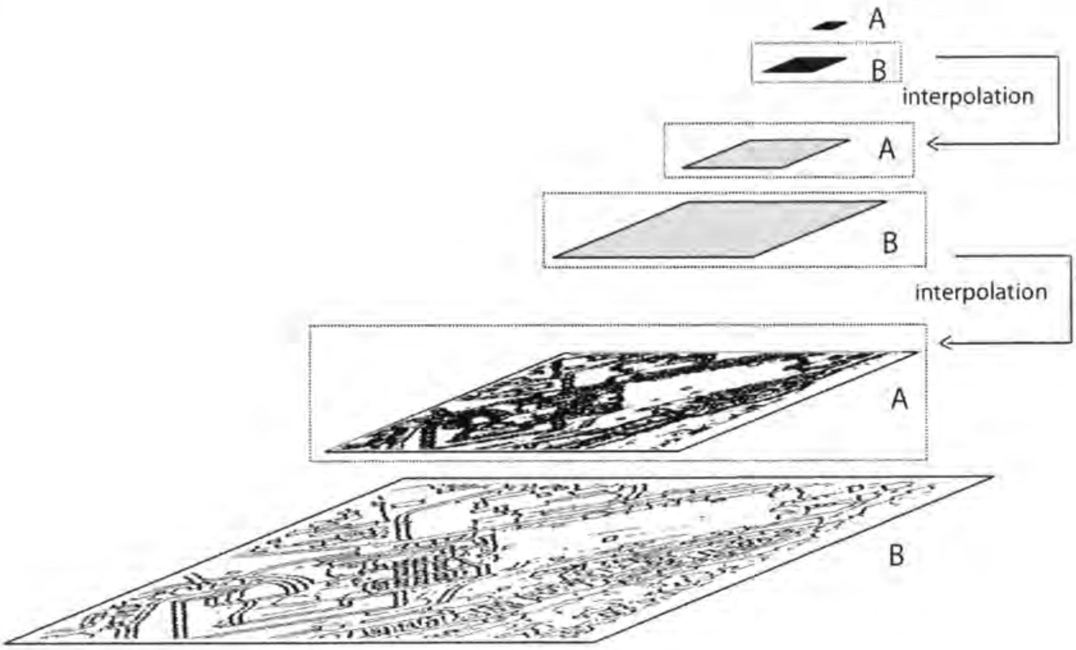


Figure 5.7: Doublet method (explanations in text).

used, three doublets are formed, and the number of interpolations decreases from 5 to 2, as interpolation occurs between doublets instead of between subsequent image pyramid levels. The type of features to be matched in each pyramid level varies and more details are discussed in section 5.3. A feature extracted on level B would not always correspond to a feature on level A, or two features of level B may be mapped to the same position on level A, due to the coarser resolution and the generalization of objects. Moreover, only successful matched points on level A would be matched on level B. As a result, a lower number of points will be matched in level B. If grid points are selected for matching in level B, their grid spacing will become half in level A. In the case of features extracted by an operator such as Canny (one-pixel wide edgels), again one point on level A may correspond to two feature points of level B but also less features would be extracted, due to the loss in detail. This problem is resolved to a certain extent by modifying the thresholds of the operator in order to extract more features on A or by using different operators on A and B. The latter approach is the one employed in the developed algorithm, using operators extracting one-pixel and more than one pixel wide edges for levels B and A, respectively (see more details in section 5.3). The matching strategy within the doublets, as described, is illustrated in Figure 5.8. The selection of the start points on the search image on level A of the doublets varies. For the search images of the top doublet, since the parallaxes are small, the initial pixel coordinates are equal to the pixel coordinates x and y on the template image minus an offset value[3] for x and y, respectively. For the search images of the lower doublets, approximations are derived by projection on the search image of the interpolated object point coordinates. After matching, the approximate DSM generated from level B is used then as approximation for matching on level A of the next doublet.

[3]The offset values for Lev1 images are computed during Ground Processing and are given in the meta data files. For Lev0 images the offset value for x is computed based on the convergent angle and for y is set to 0

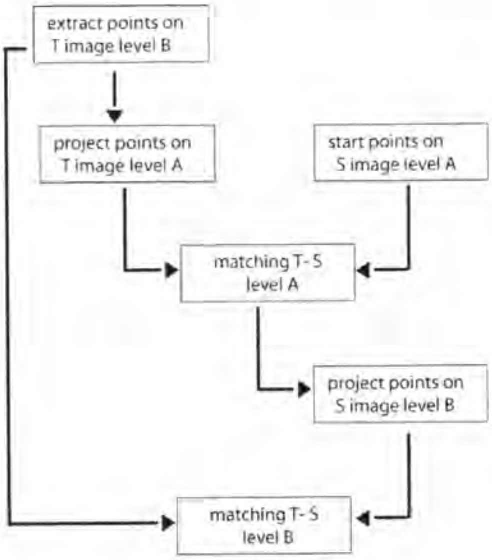


Figure 5.8: Hierarchical matching for a doublet. T and S indicate template and search image respectively.

5.2.4 Interpolation of height approximations

Interpolation procedures are used in different research areas/disciplines and a large number of interpolation schemes exist (Schut, 1976). Many of the techniques of spatial interpolation are two dimensional extensions of the one-dimensional methods originally developed for time series analysis. For 2.5D interpolation, usually local interpolators are applied, restricted to small interpolation support areas. B-splines, inverse distance weighting, kriging, trend surface analysis or approximation by polynomials are the methods widely used in DEM/DSM production (Kim et al., 1999). In commercial software packages for DTM modelling, the following interpolation schemes are often used: linear prediction (SCOP, SURFER), moving average (SURFER), finite elements (HiFi). In AIM, after blunders are excluded at each level, height interpolation is performed for a set of irregularly distributed extracted points from another set of points, defining a coarse surface, in order to obtain approximate values for matching. The height value is estimated from the weighted average of n points within a certain radius in pixels. The minimum and maximum points within the radius is set by the user. The radius is adjusted based on logical checks, namely if the number of points within the initial radius exceeds the maximum number of points or is less than the minimum number of points, the radius is divided or multiplied by a factor (of a maximum value of three), respectively. Weights w_i as a function of distance d_i (in pixels) are applied according to the following schemes:

1. Inverse distance weighting,

$$w_i = \frac{1}{d_i}, \quad i = 1 \ldots n \tag{5.3}$$

2. Gaussian,

$$w_i = e^{\frac{-d_i^2}{\sigma^2}}, \quad i = 1 \ldots n \tag{5.4}$$

where σ defines the width of the Gaussian function and is set to one third of the interpolation radius.

3. Smoothed weighting.

$$w_i = \frac{f^2}{d_i^2 + f^2}, \quad i = 1 \ldots n \tag{5.5}$$

where f defines the smoothing factor, which is assigned a very small value (the default used value is $1e-5$).

According to Kim et al. (1999), between Gaussian and inverse distance weighting, the first showed better performance according to visual and quantitative evaluation. Moreover, the third weighting scheme was used in other investigations for DSM generation (Baltsavias et al., 2001a) and delivered good results. As the focus in this project was not to perform an extensive analysis of interpolation methods, the above described methods that have been proven in previous assessments to deliver good results and are computationally inexpensive were implemented also in AIM. The smoothed weighting scheme was selected and used finally in our tests. However, these methods have certain disadvantages as for all point classes (e.g. points that belong to breaklines and single points on the ground) the same interpolation method is used, and less accurate values are derived for breakline points, e.g. for ground points close to buildings.

To decrease processing time, height values are interpolated only for the extracted points on the template image, planimetric coordinates are computed by the intersection of the ray of the template with the height plane and the approximate pixel positions on the remaining images are derived by projection of the ground points.

5.3 Selection of optimal features

In this section, the selection of matching entities, in combination with the hierarchical matching approach, described in the previous section is discussed. Optimal are the features which permit a dense, precise and reliable matching. Features can be area patches, interest points (corners or centers of circular features), points along edges (unlinked edgels), edges (curves or straight lines), or more complex features (polygons). However, several criteria are taken into account for the selection and combination of matching features and different operators are evaluated in terms of precision, reliability and completeness in describing the scene, in order to select those with better performance.

5.3.1 Criteria for selection of optimal features

- Areas with lack of texture should be avoided. Commercial systems select matching points either in a grid (image or object) or as point features. The first approach leads to matching even at areas with no texture, obviously leading to wrong results.
- Advantages and disadvantages of feature- and area based matching, mentioned already in section 4.1.2 should be considered. Both types of methods should be utilized or combined to a hybrid method, depending on the accuracy of the initial approximations and the landcover type of the area to be matched (see also section 5.4).
- Geometric considerations and texture information should be taken into account. In general, features with good texture in two directions should be matched. In the case of features selected in partly

occluded areas, additional parameters (e.g. optimal shape of mask) should be included or the features should be disregarded. If features lie along lines parallel to epipolar lines or in areas with repeated texture patterns, multiple solutions will arise. However, geometric constraints can reduced the problems arising from repeated texture patterns. In AIM multiple solutions for features parallel are handled with the edge LSM approach (see section 5.4.4.2).

- Feature selection parameters should be adapted to ensure sufficient density in all image regions, excluding textureless areas.
- Knowledge about the surface form should be used, in order to select denser features along discontinuities compared to flat areas. However, if features are extracted by an interest operator, then in highly textured flat areas (e.g. agricultural fields)their number would be redundant.
- In doublets, the density of features should be sufficient in both level A and B. Therefore, the percentage of features that are transferred from level B to level A should be high to account for the subsequent decrease in the number of points to be matched in level B. In other words, not all the features extracted on level B will be finally matched, and only the ones that exist and are successfully matched on level A will be used.
- In the case of large geometric differences, that may also lead to partly occluded areas, feature selection should be extended to more than one image that serves as a template, depending on the configuration of the available images (see section 5.4.5).
- Speed aspects related to hierarchical matching. Selection of grid points suffices to generate a coarse surface and with less cost in time than extraction and matching of interest points or edge features. The approximate surface can later be densified in the lower pyramid levels by additional features (e.g. edges), in order to improve surface definition, especially along discontinuities.

To date, a number of researchers have used different measures to evaluate feature extraction algorithms. In several studies, these measures have been too specific and could not be used for a broader range of extractors or even for images serving different applications, e.g. Demigny and Kamlé (1997) used Canny's criteria, which proved to be less optimal than other algorithms, and López et al. (1999) focused on ridge and valley detection for specific applications. With respect to the previous criteria of optimal features, the selection principle of a feature operator should fulfill the following characteristics:

1. Invariant with respect to radiometric and geometric distortions and noise.
2. Features should be distinct, i.e. different from neighboring features.
3. Completeness in describing the surface form.
4. Precision in localization of extracted points.
5. Reliability of features, i.e. the extracted feature should correspond to an existing feature.
6. Optimal time and memory performance.

5.3.2 Feature extraction

In AIM, the three types of features that were evaluated are corner points, points along edges (edgels) and contours or aggregated edge pixels and the last two were finally used. More complex features than edges are not used as the algorithmic complexity and processing time increases. Moreover, as a large number of operators exist, and a detailed analysis is beyond the scope of this thesis, the most frequently used operators in the photogrammetric and computer vision community are evaluated. In sections 5.3.2.1 and 5.3.2.2 the operators used for point and edge extraction are briefly presented and are later analyzed and evaluated in section 5.3.3.

5.3.2.1 Point extraction

The Förstner operator (Förstner and Gülch, 1987) used also in commercial photogrammetric systems (MATCH-T) was designed as an interest operator for the detection and sub-pixel location of intersections, centers of small circular features, patches of good texture for matching, but it can also be used for edge points. It is an intensity based method, which is based on the autocorrelation function of the grey values. It selects points with small and round intensity ellipses. The algorithm is based on the following steps:

1. The image is divided into tiles of constant size (user defined).
2. Image gradients g_x, g_y are computed with the Sobel or the Roberts operator in each tile. In our implementation we used the Sobel operator.
3. The normal matrix elements N_{xx}, N_{yy}, N_{xy} are derived from the sums of the second grey level gradients $\sum g_x^2$, $\sum g_y^2$ and $\sum g_{xy}$ respectively, for a given mask size. The normal matrix N is a texture descriptor and shows the direction and strength of the local texture within the operator window.
4. The strength w_i and the roundness of the intensity ellipse q_i of each point within the mask is calculated, according to Equations 5.6 and 5.7.

$$w_i = \frac{detN_i}{trN_i} \tag{5.6}$$

$$q_i = \frac{4 \cdot detN_i}{tr^2 N_i} \tag{5.7}$$

   where $detN_i$ and trN_i are the determinant and the trace of the normal matrix N_i, respectively.
5. Points with $w_i > w_{lim}$ and $q_i > q_{lim}$ are kept. The lower bound q_{lim} is a free parameter and Förstner and Gülch (1987) suggest values of greater than 0.5 but less than 0.75. In our experiments we set the lower bound of q_{lim} to 0.4. w_{lim} should be adapted to the image content and the same authors suggest to derive it from the average w_{mean} of w_i (optionally multiplied by a factor in the range between 0.5 and 1.5), however the approach having the disadvantage that the computed w_{lim} depends on the sharpness of features and even more on the noise level.
6. Suppression of all local non-maxima with the suppression distance being equal or greater to the half of the mask size.
7. Localization of the points based on a least squares estimation model, followed by classification of the points into intersections and centers of small circular features.

Based on the image content and the density of the extracted points, the user has to set the optimal window size (minimum size 5x5) and empirically set the thresholds for w_{lim} and q_{lim}. Furthermore, the classification of the points requires two thresholds, namely for intersections and centers of small circular features, and that the user is familiar with statistical analysis and the F distribution to set the two thresholds. Thus, in our work we refined some procedures in the Förstner operator related to the automatic determination of the input threshold w_{lim}, as it is related to the mask size and some experience is required from the user. This part was simplified and the threshold w_{lim} is automatically derived from the average strength w_{mean} computed from the non-zero w_i values and additionally constrained within the range of two min and max values, the w_{min} and w_{max} respectively. The min, max strengths are derived from the minimum and maximum difference, d_{min} and d_{max}, in grey values of adjacent pixels, given by the user. The normal matrix elements are computed with the differences d_{min} and d_{max} being the min, max derivatives in both x and y direction. The sum of the squared differences is taken over half the pixels of the mask and amplified by a factor of three (multiplication instead of sum is used, since all squared differences have the same value). $\sum g^2_{x_{min}}$, $\sum g^2_{y_{min}}$ are computed according to Equation 5.13 and $\sum g_{xy_{min}}$ is ca. half the value of $\sum g^2_{x_{min}}$, or $\sum g^2_{y_{min}}$, (see Equation 5.9). This approach is justified under the assumption that a possible corner a) will be centered in the selected window and b) gradients of the corner will usually occur in a part of the mask, i.e. in approximately a quarter of the mask. However, possible cases of intersections of edge elements are more than the one described, but for the derivation of a global and approximate threshold the above assumption can be made. Based on Equation 5.6, Equation 5.10 is derived and w_{min} is computed.

$$\sum g^2_{x_{min}} = \sum g^2_{y_{min}} = s \cdot d^2_{min} \tag{5.8}$$

$$\sum g_{xy_{min}} = \frac{1}{2} \sum g^2_{x_{min}} \tag{5.9}$$

$$w_{min} = \frac{(s \cdot d^2_{min})^2 - s \cdot d^2_{min}}{2 \cdot s \cdot d^2_{min}} \tag{5.10}$$

with

$$s = (\frac{m-1}{2} + 1) \cdot 3 \tag{5.11}$$

where m is the mask size (odd number) and s the multiplicative factor. The nominator and denominator correspond to the determinant and the trace of the normal matrix respectively. w_{max} is derived similarly to w_{min}. Moreover, the computed mean strength w_{mean} is further compared with w_{min} and w_{max} and the w_{lim} is selected according to Equation 5.12.

$$w_{lim} = \begin{cases} w_{min} & , if \quad w_{mean} \leq w_{min} \\ w_{mean} & , if \quad w_{min} < w_{mean} < w_{max} \\ w_{max} & , if \quad w_{mean} \geq w_{max} \end{cases} \tag{5.12}$$

Other refinements in the implementation include optional use of a smoothing function prior to the derivation of the gradients, i.e. Gaussian filtering by circular or square masks, and employment of LUTs in the derivation of the gradients to improve computational speed. The alternative thresholding method based on the noise variance mentioned in Förstner (1998) has been implemented and evaluated (see section 5.3.3). The noise variance is used as a criterion to segment the image in homogeneous and non-homogeneous areas. The interest points that lie in non-homogeneous areas with $q_i > q_{lim}$ are selected.

Similarly to the Förstner operator, the Harris operator (Harris and Stephens, 1988; Schmid et al., 2000), as a representative of a step edge corner detector, is based on the autocorrelation function. There are certain differences in the computation of the gradients and of the normal matrix elements. As an interest value, the cornerness response defined as $cornerness = detN - \alpha \cdot trN^2$ is used for each pixel. α is set by the user and typical value is 0.04. Even though a general description of the operator exists in the publication referred to, no information on thresholding is given. For this reason, the implementation of Harris (VXL, 2004) has been analyzed, in order to understand the logic behind the derivation of thresholds. The thresholds used are: a) the minimum cornerness value derived from the product of the ratio of min/max corner strength (set by the user) and the maximum value of cornerness of the points and b) the maximum number of points that should be extracted (set by the user). These thresholds are difficult to interpret and less optimal for use in fully automatic processes, especially the second one, as the user has to pre-estimate, depending on the size and content of the scene, the maximum number of points that should be extracted. A common problem with all these operators, is that the corner response varies considerably with image contrast.

In addition, the SUSAN (Smallest Univalue Segment Assimilating Nucleus) operator (Smith and Brady, 1997) used for both corner and edge detection has been evaluated. The SUSAN operator identifies features by determining what fraction of a circular mask has values the same, or similar, to the value at the center point. Thresholds are therefore defined in terms of the size of the mask and no image smoothing is required. However, the SUSAN operator assumes that edges and corners are formed by the intersections of regions having constant, or near constant, intensity, and this limits the types of corners that can be modelled. Apart from the detailed description of the operator, the source code (SUSAN, 2004) was provided by the authors and no further modifications have been applied.

5.3.2.2 Edge extraction

Edges have been extracted using the previously mentioned Förstner and SUSAN operators. In addition, Canny (1986)[4] and a gradient thresholding algorithm, the later developed within AIM, have been evaluated. All methods consider edges as being points of high intensity gradients. The first three detect one pixel-wide edges, whereas the gradient thresholding is used for the extraction of edges of any width. The interest values computed with Förstner operator in the case of edgels are the trace trN_i and the roundness of the signal ellipse q_i. Compared to the thresholds used in the extraction of intersection and centers of small circular features, the upper bound of q_{lim} is set to 0.6, as for edgels the signal ellipse is elongated in edge direction and $q_i < q_{lim}$ and the threshold for the trace tr_{lim} is derived similarly to w_{lim}. The average trace tr_{mean} is computed from the trN_i and should be within the range of a minimum tr_{min} and maximum tr_{max} trace. tr_{min} and tr_{max} are derived from the the input values of minimum and maximum grey value differences d_{min} and d_{max} using Equations 5.13, 5.11, 5.14 and 5.15.

[4]The dissertation of Canny (Canny, 1983) presents more details on the algorithm and the performed investigations.

$$\sum g_{x_{min}}^2 = \sum g_{y_{min}}^2 = s \cdot d_{min}^2 \tag{5.13}$$

$$tr_{min} = 2 \cdot s \cdot d_{min}^2 \tag{5.14}$$

$$tr_{max} = 2 \cdot s \cdot d_{max}^2 \tag{5.15}$$

The gradient thresholding algorithm is relatively simple and consists of a two-step procedure. First, the magnitudes of the gradients are derived from the image and then the threshold T is computed according to Equation 5.16. The statistical values, i.e. *mean* and *standard deviation (std)*, are derived from the gradients and a factor n is defined by the user. The factor should be relatively small, i.e. $-3.0 < n < 3.0$, and the larger the factor the lower the threshold. Edge pixels are marked when its magnitude is greater than the threshold T.

$$T = mean - n \cdot std \tag{5.16}$$

Except the SUSAN operator that derives contours and node points, the other operators extract single points that lie on edges (so called edgels), linking into contours or edges is performed in a later processing stage using the method described in Henricsson (1996). The edgel aggregation method is a sequential process, where the significant edgels are aggregated before the weaker ones and small gaps are bridged based on criteria of proximity and collinearity. After the contour graph is generated, a further post-processing is applied, based on defined attributes to remove weak contours and obtain as long and straight contours as possible, assigned to one specific class. The attributes can be geometrical (i.e. length, curvature of contour), radiometrical (strength of edge), topological (adjacency, common end-points). Contours are assigned after post-processing to three classes, namely curved, straight and closed contours.

5.3.3 Practical aspects and comparison of extracted features

Tests have been performed on synthetic and real image data with apparent noise and results have been qualitatively evaluated. In Figures 5.9 and 5.10, points and edgels have been extracted on synthetic images by using different operators. Interest values have been computed out of the whole image area for all methods (one tile was used). For the Förstner operator, the two different thresholding schemes, based on minimum/maximum difference of grey values and noise variance, have been employed (from this point forward called versions 1 and 2), using the same mask size (5 x 5) and non-maxima suppression distance (2). The number of extracted points derived with the Förstner operator has been set in the Harris operator as parameter value for the maximum number of points required. In Harris operator the default mask size was 3 x 3, and the Gaussian $\sigma = 1$. For SUSAN the default option, suggested by the authors (SUSAN, 2004), was employed.

Among the four operators the time performance of SUSAN and Harris was approximately three times better than the Förstner operator, though for the latter better localization accuracy could be realized. Both these characteristics are related to the least squares adjustment employed in the algorithm.

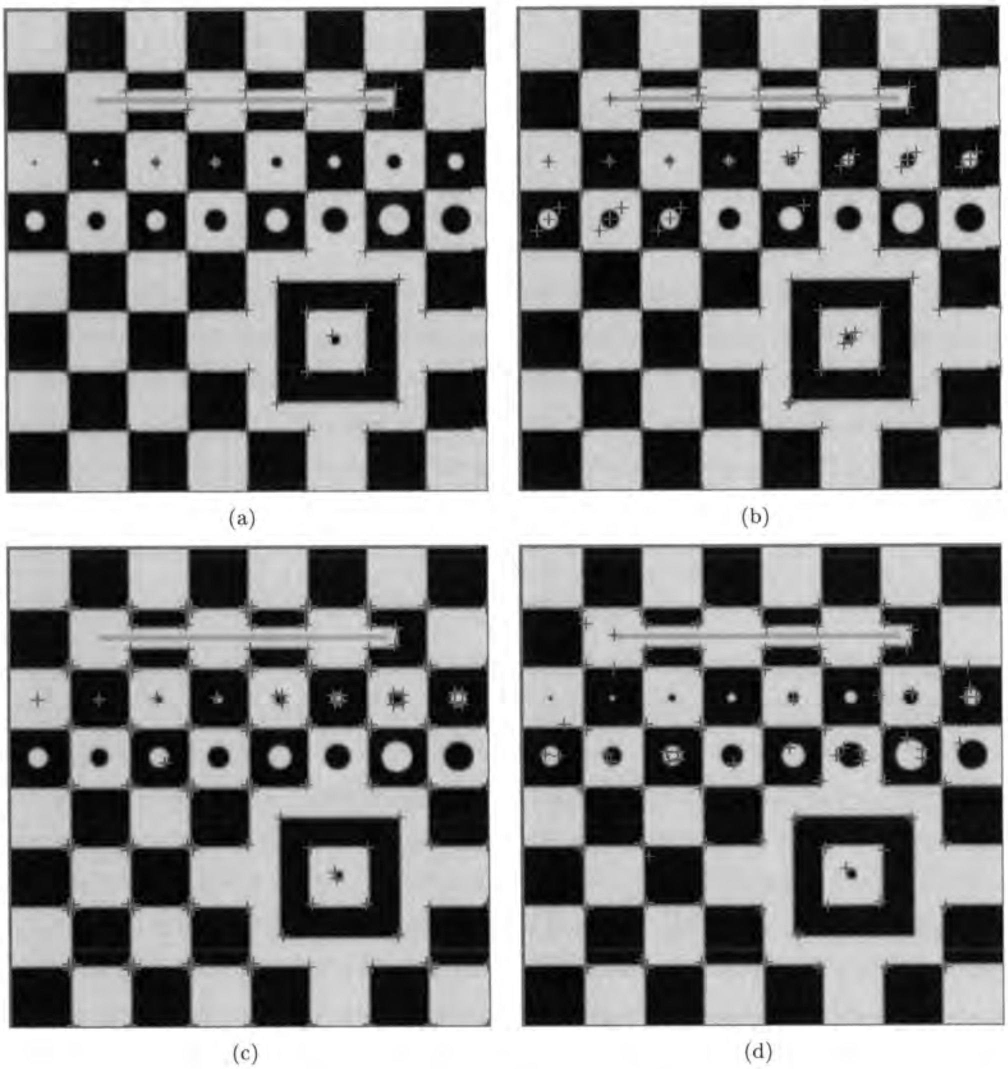


Figure 5.9: Extraction of interest points on synthetic image data. In (a) and (b) points have been extracted with the Förstner operator using the two different thresholding options (1 and 2). In (c) and (d) results from the Harris and SUSAN operators are illustrated.

Moreover, in contrast to corner points, not all centers of circular features have been extracted, i.e. when the circular feature is larger than the mask size, gradients do not exist in the mask and as a result no points are extracted. Between the two versions of the Förstner operator, more points have been extracted with the version 2 (Figure 5.9(b)), i.e. 83 and 165 points for versions 1 and 2 respectively, and more centers of circles have been identified. However, erroneous points existed in homogeneous areas, along with multiple points close to gradient intersections. Erroneous points may appear mainly due to locally false segmentation of non-homogeneous areas in which the q_{lim} does not suffice as a reliable thresholding criterion. Multiple points also occur due to the segmentation of the areas, and the computed shifts from the least squares adjustment, performed at the last processing stage, are larger than 2 pixels. If the noise variance criterion is combined with the criterion used in version 1, namely after the segmentation in homogeneous and non-homogeneous areas the threshold derived from the minimum/maximum difference of grey values is applied, the same number of points is derived from both versions 1 and 2. Regarding the Harris operator, apart from the poorer localization accuracy, centers of circular features are not extracted, multiple points (2-4) exist close to intersections of edges, points along the perimeter of circles have been extracted. Although the user sets the maximum number of points, the number of extracted points is more than the input value (177). With an increasing number of extracted points, even small features are identified. With SUSAN, some false points arise, especially along edges and in homogeneous areas, but compared to Harris in areas of edge intersections one point is extracted.

In terms of precision and reliability, the Förstner operator performs better than the others. Features are distinct, but with a higher cost in time performance. However, certain features are not extracted (small corners of less sharper edge intersections) and as a result the description of the scene is incomplete. In view of the fact that usually large images are processed, thresholds should be adapted for each tile. As the scene may vary locally, generalized thresholds are less optimal[5]. This concept is already used optionally in the Harris operator, but was not used in these tests as it needs to be further investigated, towards higher distinctiveness and precision of features. Extraction of edges leads to a higher density of features compared to interest points (Figures 5.10 and 5.11). There are visible differences in the edge features extracted by the operators. Namely with Förstner (Figures 5.10(a), 5.10(b) and 5.10(c)) a number of edgels have not been extracted, and more gaps along edges are visible due to the non-maxima suppression. Both the mask size and the suppression window was set to 3x3 and in addition a suppression perpendicular to edge direction was used. SUSAN (Figures 5.10(d), 5.10(e) and 5.10(f)) and Canny (Figures 5.10(g), 5.10(h) and 5.10(i)) are more sensitive to noise and tend to extract even small features in low-textured areas compared to gradient thresholding (Figures 5.10(j), 5.10(k), 5.10(l)). As the tests performed on ADS40 images (Figure 5.11), also show the use of edge features is justified as more points are extracted, especially along roof discontinuities, and therefore assist in the later detection and modelling of discontinuities. SUSAN was disregarded due to the partially incomplete extraction of roofs, as seen in Figure 5.11(f). The later edgel aggregation has been employed for one-pixel wide edges, i.e. derived with Canny. Figure 5.12 illustrates post-processed edges (including initial edgel aggregation) overlaid on the original image. Depending on the threshold, related to the minimum edge length, which is defined by the user, less significant edges are omitted. This threshold must be carefully selected with regard to the scene content, i.e. the size of buildings and roof types, existence of highly textured flat areas, and image scale. In flat areas with relatively high texture (i.e. agricultural fields), many edgels will be extracted with Canny, and less with gradient thresholding. However, a lower number can be used to describe the surface. A summary of the performance of the previously discussed feature operators (point and edgel) with respect to the different criteria are listed in Table 5.1.

Since doublets are utilized in AIM, Canny and gradient thresholding are combined for optimal performance. Canny and gradient thresholding are used to extract features in level B and level A respectively. This is justified as due to the dense extraction by gradient thresholding, the number of corresponding features between levels A and B will be high. A large number of edgels, extracted with Canny in highly textured flat areas, are thresholded, as less edgels are derived in these areas with the second operator, namely the gradient thresholding. Finally, only significant edgels will be kept and

[5]except the case of the q_{lim} threshold used in Förstner operator (see section 5.3.2)

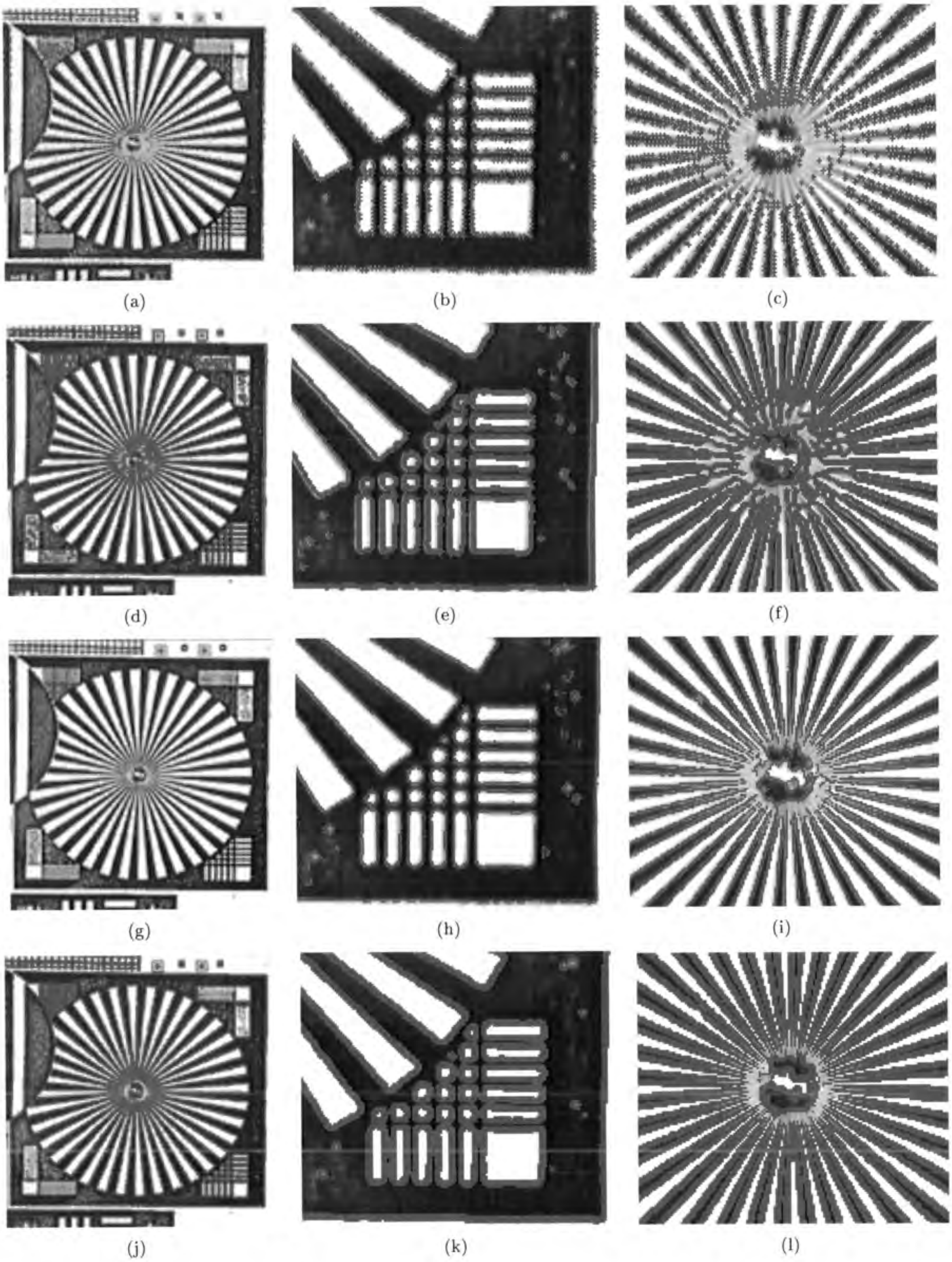


Figure 5.10: Extraction of edgels on an image containing different types of features. In subfigures (a), (d), (g) and (j), edgels extracted with the Förstner, SUSAN, Canny and gradient thresholding operators, respectively, are overlaid on the original image. The subfigures in the middle and on the right columns show enlarged parts of the images on the left column from the centre and the lower right areas.

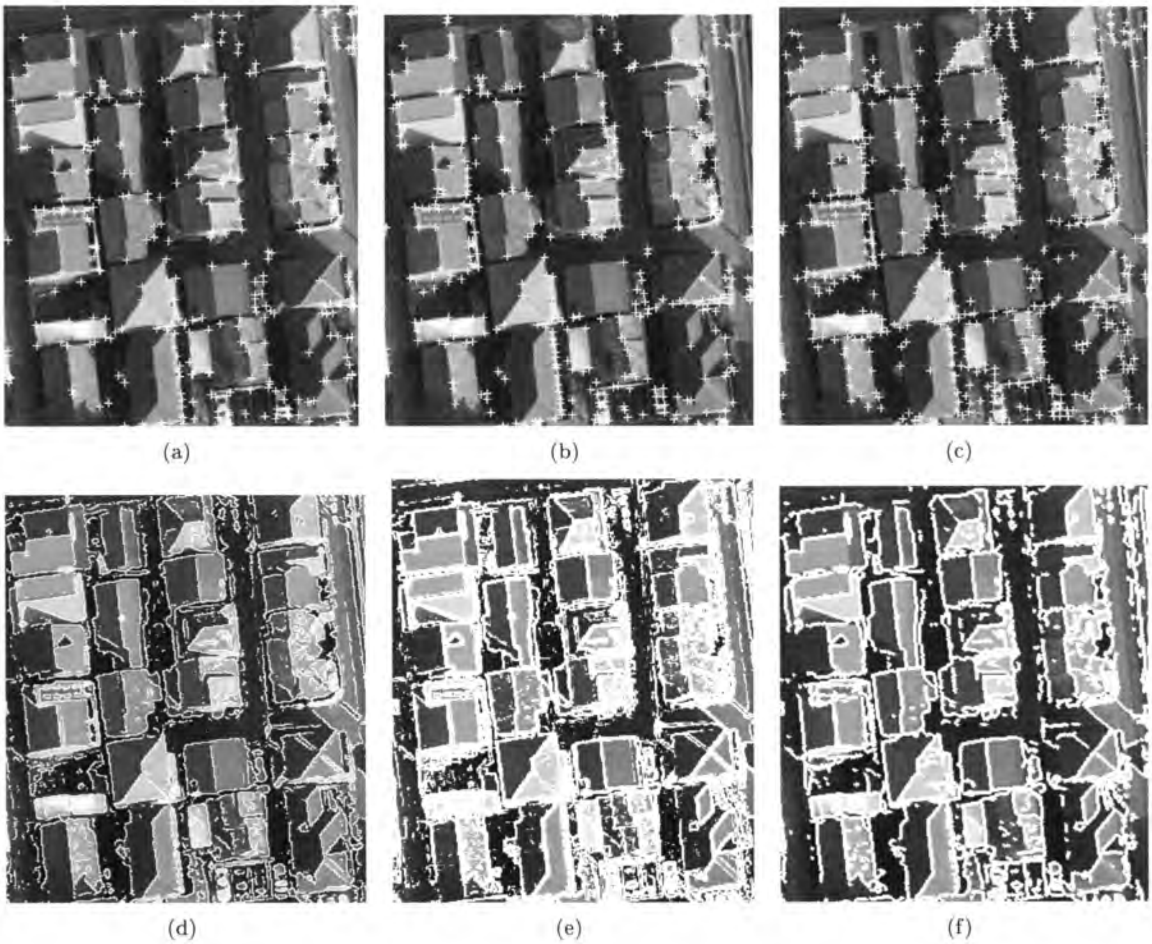


Figure 5.11: Extraction of features on ADS40 image data. On the top row extraction of point features with Förstner (a), Harris (b) and SUSAN (c) operators. On the bottom row extraction of edge features with Canny (d), gradient thresholding (e) and SUSAN (f).

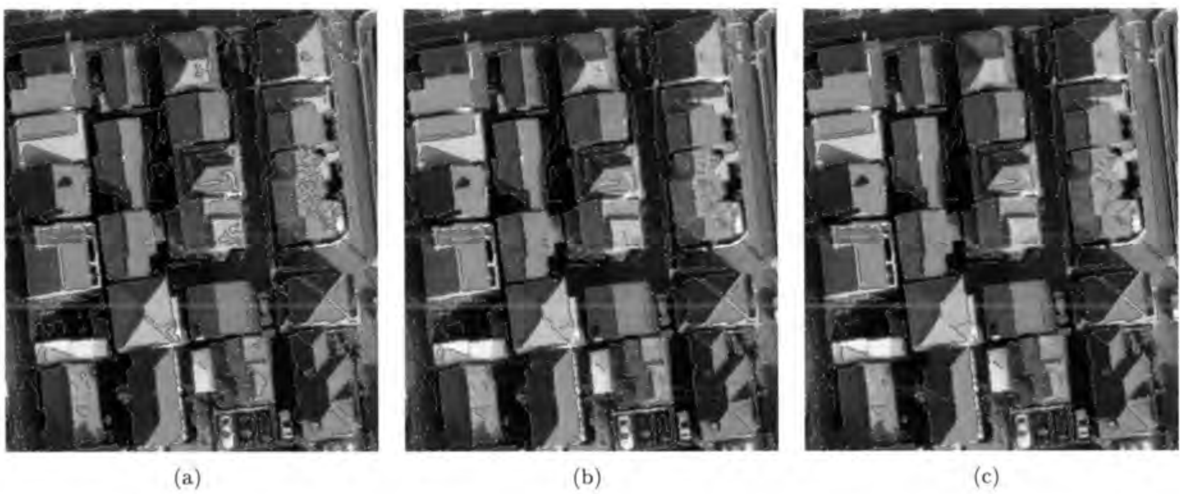


Figure 5.12: Results of edgel aggregation. Vertices and contour points are indicated with yellow and red color respectively. In (a) all classes of edges are overlaid on the image, whereas in (b) only straight edges. In (c) straight edges with length more than 10 pixels are kept.

	Points			Edgels			
Criterion	Förstner	Harris	SUSAN	Förstner	SUSAN([a])	Canny	G. Thresh.
Completeness	-	-	-	-	+	+	+
Localization	+	-	-	+	-	0	-
Distinct features	+	-	0	+	+	+	0
Reliability	+	-	-	+	-	+	+
Time-memory	-	+	+	-	+	+	+

Table 5.1: Performance of point and edge extractors with respect to different criteria. Positive and negative properties are indicated with + and - respectively. 0 indicates average performance. ([a])With SUSAN edges are extracted.

further aggregated to form edge features, reducing in parallel processing time and excluding less robust and reliable features in the edge matching process, which is described in section 5.4. Straight edges, longer than 11 pixels, are assigned to a different class from remaining edges, as Figure 4.2 illustrates. For each class a different matching scheme is used, as explained in the following section.

5.4 Combination of different matching methods

In AIM, we use different matching methods and combine them with the selected matching entities and with the doublet strategy and type of the selected area (i.e. terrain relief and landcover). At the outset, some practical aspects are listed and later the employment of two matching algorithms, multi-patch size matching and extensions of LSM respectively, is analyzed and justified. In addition, the integration of all the above mentioned matching aspects in a multi-image and multi-template matching strategy is presented. This relates to the selection and order of ADS40 channels in the matching strategy.

5.4.1 Practical aspects

- From an operational point of view the algorithm should be well designed and have optimal memory and time performance.
- Algorithms requiring less accurate approximations should be used in the upper image pyramid levels, where approximations usually are very coarse, in order to retrieve better approximations required by other algorithms that are used for refinement and better control of the matching solution (i.e. LSM).
- Even though surfaces are seen from different viewpoints, differences in scale and rotation in Lev1 images are small, thus algorithms that are less optimal in the presence of large geometric differences can be employed, especially in the upper pyramid levels.
- Matching precision on upper doublet level (A) can be restricted to pixel accuracy, since successfully matched points are transferred directly one level lower. Subpixel accuracy is required for matching results derived from the lower doublet level (B), namely higher accuracy in height determination from 3D intersection, in order to minimize subsequent height interpolation errors.

- Employment of more than two images in matching for higher measurement accuracy, elimination of erroneous matches (section 5.5), and reduction of errors due occlusions, image noise and surface discontinuities.
- Partially incomplete surface representation, due to an insufficient number of matched points arising from occluded areas, is handled through the use of multiple template images (section 5.4.5).

5.4.2 Multi-patch size matching

Multi-patch size matching is used in the first matching stages and in order to both improve approximations derived from previous pyramid levels and deliver pixel accurate results. Three passes of matching with different patch dimensions and sets of parameters are performed. The use of the multi-patch size approach can be justified since the larger patches aim at reliability of a coarse solution and the smaller ones at higher precision. The larger patch is less sensitive to noise, occlusion, multiple solutions, etc., while the smaller one is more precise and better preserves height discontinuities. As far as the similarity measure is concerned, a set of different measures, shown in Table 5.2, has been tested.

In cases of changes in illumination (shift or scaling of intensity), the SAD and SSD measures are not invariant. In other words they do not support intensity distortions, compared to the normalized measures like NCC, which is invariant to any linear transformation. As all above measures are not robust with respect to occlusions, they are only used to improve coarse approximations, especially in the upper pyramid levels of Lev1 images where small geometric differences exist. We employed these similarity measures for grey level images (8- and 16-bit) and partially thresholded grey level images for which minimum information is kept, namely either information of the gradient magnitude or additional edge directionality information. The main reason to use partially thresholded rather than grey level images, especially for coarse matching, was to examine possible gains in speed and decrease in memory allocation (only one matrix, holding the information of the extracted edges, is stored). Out of the different combinations of image types and similarity measures, LUTs for SAD and SSD could not be employed on grey level images due to limitations in memory allocation. Moreover, determination of thresholds for SAD and SSD, in order to distinguish between false and correct matches was more difficult compared to NCC, especially when grey level images were used (i.e. higher value range) and not partially thresholded. Our results showed that the use of partially thresholded images can decrease processing time slightly, less than initially expected, and without leading to significant differences in the matching results of the upper pyramid levels. Based on these observations, we followed a common matching scheme for all pyramid levels, namely using the grey level images with NCC as the similarity measure.

Other parameters that we used in multi-patch size matching are the patch dimension, the patch dimension reduction step (from one patch to the smaller one), minimum patch size dimension, search window for large and the two smaller masks, lower threshold for NCC T_1 (minimum value $T_1 = \frac{\sqrt{m_x \cdot m_y}}{m_x \cdot m_y}$, where m_x and m_y the mask dimensions), and threshold T_2 for maximum absolute difference of the NCC of the smallest mask from the maximum NCC. The largest mask is used in the first pass and is sequentially reduced in the second and third pass by the defined reduction step. The search range remains constant for the second and third pass and should not exceed the search range, used in the first pass. Typical values (used in the The starting approximations for each pass are taken from the last updated position,

Name	Abr.	Definition	I	Type
Normalized Cross-Correlation	NCC	$\frac{(g_t - \overline{g_t}) \cdot (g_s - \overline{g_s})}{\|g_t - \overline{g_t}\| \|g_s - \overline{g_s}\|}$	[0;1]	S
Sum of Absolute Differences	SAD	$\|(g_t - g_s)\|$	$[0; I_{max}]$	D
Sum of Squared Differences	SSD	$\|(g_t - g_s)^2\|$	$[0; I^2_{max}]$	D

Table 5.2: Similarity measures tested in multi-patch size matching. The abreviations (Abr.), the measure type: similarity (S) or dissimilarity (D) and the intervals of variation (I) are given. The vectors g_t and g_s contain the grey levels of the pixels in the template and search images and their mean values are $\overline{g_t}$ and $\overline{g_s}$. The L_2 and L_1 norms are denoted by $\| \;\; \|$ and $| \;\; |$.

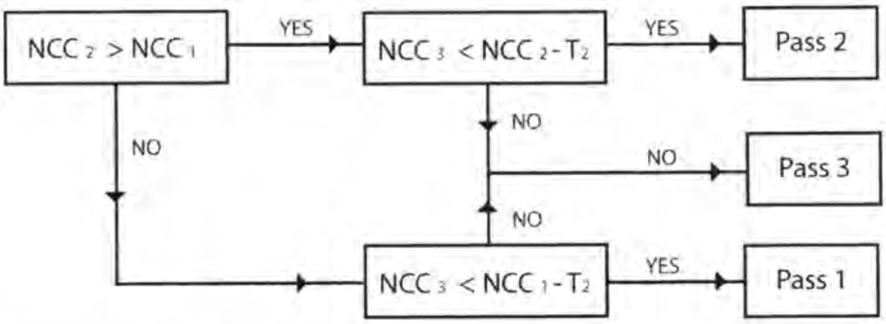


Figure 5.13: Selection of best patch size in multi-patch size matching. NCC_1, NCC_2, NCC_3 are the maximum NCCs from the three passes.

derived from the position of maximum NCC from the previous pass. In this way a linked list is formed, connecting initial and final positions from each pass. The maximum similarity measure is computed for each pass starting from the largest patch dimension, defining the first pass together with a set of quality criteria, which are explained later in this chapter (section 5.5). Next, through the decision making process, illustrated in Figure 5.13, the best point is selected.

The NCC will normally increase from largest to smaller mask as grey value differences between template and patch are sequentially reduced, due to the smaller patch size (e.g. in which varying grey values due to occlusions are reduced). Still, convergence to a false position may arise (i.e. features parallel to epipolar lines) and NCC has to be combined with other criteria (section 5.5) to identify possible errors. However, we used the above best-pass criterion, in which the thresholds T_1 and T_2 and used to derive a first solution, which is re-evaluted at the stage of quality control. In the case that the maximum NCC is found in multiple positions of one pass, starting from the first pass (i.e. largest mask) but except on the last pass, the position in the last pass with the largest NCC sum in all passes is selected. This problem may arise in images where other than grey level values are used (i.e. thresholded images with intensity gradient magnitude information). However, in case multiple positions are found in all passes, the point is disregarded, as being possible multiple solution.

5.4.3 Least squares matching

The LSM method has already been introduced in section 4.1.1. In AIM, LSM aims for both refinement of matching results derived by multi-patch size matching (section 5.4.2), and additional quality control of the matching solution. The usual LSM implementation uses area patches and when applied for single points that lie close to or on edges, does not suffice to model discontinuities, especially when the patch size is large. Since in AIM edge features are primarily used, LSM has been modified towards reducing smoothing of discontinuities. The discontinuity modelling, as part of the DSM extraction, through the exploitation of edge information in LSM, is discussed in the next section (5.4.4). Here, the introduction of geometrical constraints in the general estimation model is briefly presented.

5.4.3.1 General estimation model

Assume that a surface is viewed in i=1, ..n images of different viewpoints. The problem statement is finding the corresponding part of the template image patch $f(x, y)$ in the search images $g_i(x, y)$, $i = 1, ...n - 1$, called patches hereafter.

$$f(x, y) - e_i(x, y) = g_i(x, y) \tag{5.17}$$

Equation 5.17 gives the least squares grey level observation equations, which relate the $f(x, y)$ 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 to be very small with respect to the image size, i.e. formed by a very narrow bundle of image rays. Therefore, a projective transformation, can be approximated by an affine transformation (Grün, 1985). Radiometric parameters are not included in the estimation model. Instead, radiometric corrections (i.e. noise reduction, contrast and edge enhancement and radiometric equalization among the images, as per section 3.2.2) are applied either prior to or in some cases during the LSM.

5.4.3.2 Geometrical constraints imposed as additional observations

Geometrical constraints have proved to strengthen matching in terms of precision and reliability and can be included in the model when a-priori information, e.g. 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, corresponding to a small standard deviation (e.g. s= 0.001 pixel), must be used (Baltsavias, 1991). The least squares solution of the joint system is given by Equation 5.18:

$$\hat{x} = (A^T PA + B^T P_t B)^{-1}(A^T Pl + B^T P_t t) \tag{5.18}$$

where $\hat{x}$ is the vector of unknowns, A and B are the design matrices of grey level and geometric constraint observation equations respectively, P and P_t are the respective weight matrices, and l and t the discrepancy vectors of the observations. The weights are typically diagonal with elements set to unity for all the grey level observation equations. In the current implementation, we introduced the geometric constraint observation equations by means of epipolar line equations for the case of Lev1 images (section

5.1.3.1) or by the more generalized polynomial equation for the case of Lev0 images (section 5.1.2.2):

$$t = a_n(x+dx)^n + a_{n-1}(x+dx)^{n-1} + \ldots + a_1(x+dx) + a_0 - (y+dy) \tag{5.19}$$

where $a_n, \ldots a_0$ are the polynomial parameters, (x, y) the approximate pixel coordinates of the point in the search image, (dx, dy) the unknown shifts and t the discrepancy vector. For ADS40 images, the epipolar polynomial equations are used instead of the collinearity equations, as there is restricted access to the sensor model information (see also section 5.1). An alternative to the use of higher degree polynomials, i.e. to those employed for Lev0 images, is to use less strict weights for the epipolar line constraints (e.g. s = 0.5 -2 pixels). Moreover, as the model is quite general it can be used for other images, acquired by sensors with frame and non-frame geometry, e.g. spaceborne linear array images, for which not all sensor model parameters are available.

5.4.4 Extensions of LSM for surface discontinuity modelling

LSM, as all ABM algorithms, cannot easily handle surface discontinuities. If the match points lie along edges, scales are often non-determinable and estimation of the shifts is uncertain 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 matched 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 (Figure 5.14), especially in the presence of different grey values resulting from occlusions.

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, to exclude occluded rays and to compute the object point from the remaining rays. In this context, the ADS40 facilitates object point determination due to the quasi parallel projection in the flight direction (less perspective distortions) and through the simultaneous use of 3 or more line CCDs

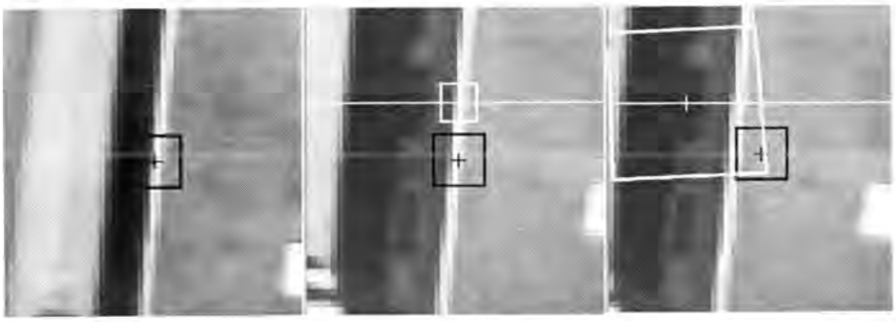


Figure 5.14: Example of correct (middle) and wrong (right) solution, related to the quality approximate values. The template (left) is selected in the Nadir view of ADS40 L1 and the patch in the Backward view. Initial and final positions of the patch are indicated with the black and white rectangle, respectively. The thin white horizontal line is the epipolar line. In the right image the approximations are poor (homogeneous area).

of different viewing angles, i.e. the Nadir, Backward, Forward and possibly a spectral line, e.g. Green. In the 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 through use of gradients instead of grey values. The small patch size reduces both discontinuity smoothing and the danger of including in the patch varying grey values due to occlusions. The use of gradients, eliminates the problem of homogeneous surfaces varying due to occlusions (e.g. in one image a dark grass 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, in the case information from other neighboring edges is included in the patch, while the matching approximations must be very good. The matched edgels must be very dense along the discontinuity to model it comprehensively, while edgels parallel to the epipolar line cannot 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 and speeds up convergence and reduces computing time. In addition, the template patch is rotated and made a thin rectangle, so that in the 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 precision and leading to less smoothing. The edge approach is quite different from the approach used up until now, and is described in section 5.4.4.2. 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 model, due to poor determinability of the affine parameters at edge points and the use of Lev1 images.

5.4.4.1 Edgel least squares matching

We apply matching to edge pixels which are extracted in the template image beforehand by the Canny operator or the gradient thresholding (see section 5.3.2). Points can also be extracted by manual selection, e.g. in the case of semi-automated measurements for building roof reconstruction. Edgels extracted with one of the above operators in the template image should be matched in the other images. To derive initial approximations for the rotation, an intensity analysis in the template and each patch is made. The intensity 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 intensity 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 the following Equations:

$$q_a{}^2 = \frac{\sum g_{xx} + \sum g_{yy}}{2} + \sqrt{\frac{(\sum g_{xx} - \sum g_{yy})^2}{4} + \sum g_{xy}{}^2} \tag{5.20}$$

$$q_b{}^2 = \frac{\sum g_{xx} + \sum g_{yy}}{2} - \sqrt{\frac{(\sum g_{xx} - \sum g_{yy})^2}{4} + \sum g_{xy}{}^2} \tag{5.21}$$

$$\phi = 0.5 \cdot \arctan(2 \cdot \sum g_{xy}, \sum g_{xx} - \sum g_{yy}) \tag{5.22}$$

$$e = \sqrt{1 - \frac{q_b{}^2}{q_a{}^2}} \tag{5.23}$$

where $\sum g_{xx}, \sum g_{yy}, \sum g_{xy}$ 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 intensity ellipse may not be elongated and the procedure is interrupted. Otherwise, the pre-rotation of each patch is determined by the difference of the angle ϕ between the patch and template. Determination of ϕ may be wrong, especially if the patch is large and contains edges in other directions (see Figure 5.15, middle). 5.15 also shows well the quite large perspective differences between the template image (here 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 reduce this problem, smaller patches (e.g. $9^2 - 11^2$) are justified. When pre-rotation is combined with the use of geometrical constraints, the signal ellipse and angle are determined at the result of the first iteration on the epipolar line. If the edge is only partially included in the patch, being otherwise homogeneous, then the estimation of ϕ is not reliable. This problem stresses the necessity of having good shift approximations before computing the signal ellipse. Furthermore we tested the use of gradients instead of grey values combined with pre-rotation, as this often decreases the number of iterations (example in Figure 5.16(a)) and improves success rate (example in Figure 5.16(b)).

In Lev1 images the rotation, scales and shears between the images are small, since the images are rectified. In spite of this, pre-rotation with gradients compared to matching without these two options, and using the 25 test points (for which approximations were manually given), led to 3 times less iterations, to a decrease of oscillations for scales and to an increase of the success rate by 35%. For the

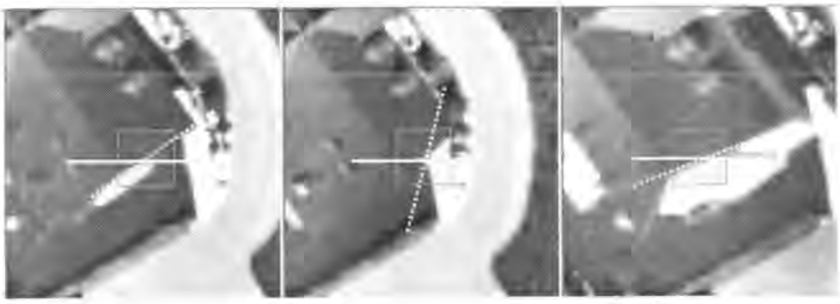


Figure 5.15: Selection of large patches (17 x17) on edge discontinuities. Left: template (Nadir), middle: Backward, right: Forward. The dotted line indicates the direction of the ϕ angle with respect to x-axis (horizontal line).

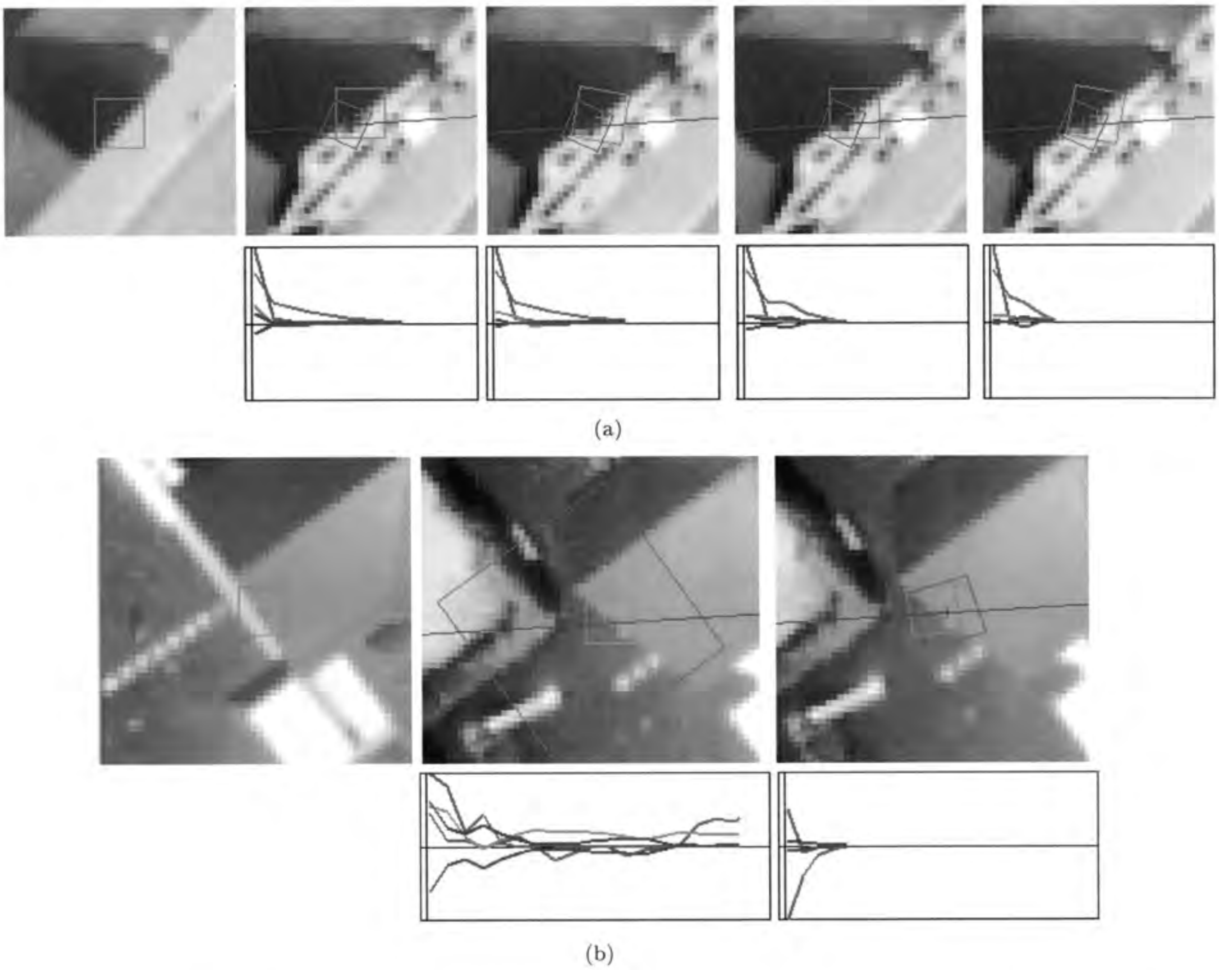


Figure 5.16: Examples of LSM with pre-rotation of patch combined with the use of grey values and gradients. On the left, the selected template (9 x 9) in the Nadir view, in other images the solution in the Backward view. (a) in the second and third image from the left grey values without and with pre-rotation, in the third and fourth image from the left gradients without and with pre-rotation; (b) in the middle the solution without pre-rotation using grey values and on the right using pre-rotation and gradients. Initial (green) and final (red) position and shaping of the patch are shown. At the lower part of (a) and (b), the alterations of shaping parameters are shown, using red, green and blue color for shifts, scales and shears respectively.

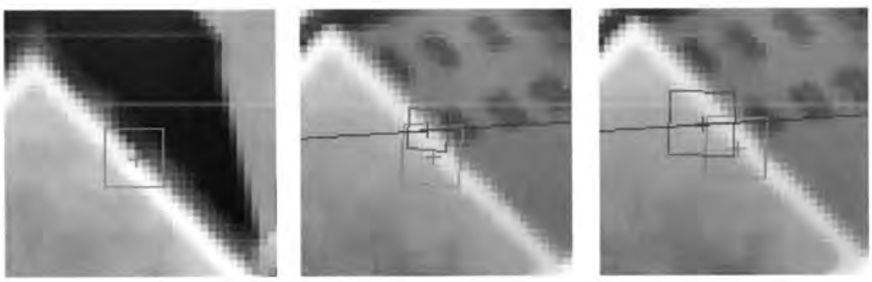
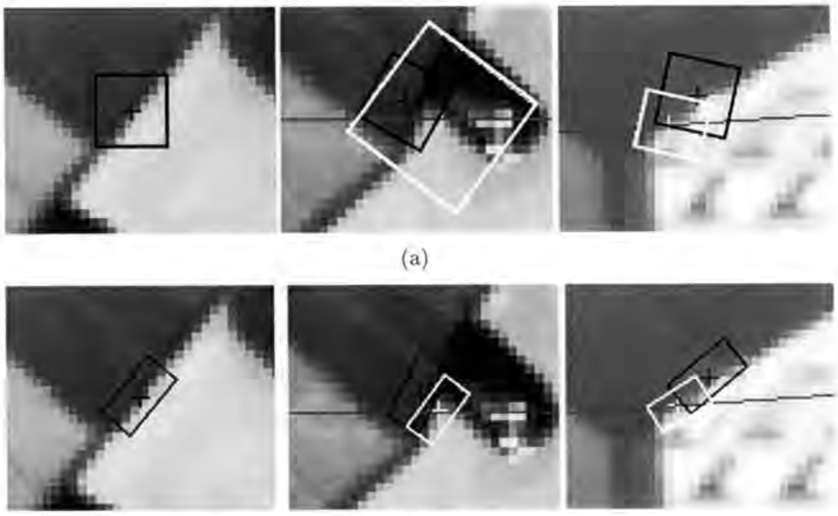


Figure 5.17: Edgel LSM with pre-rotation using grey values (middle) and gradients (right). Using gradients the danger of matching a wrong similar nearby edge increases, depending also on the quality of the initial shifts approximation.



(a)

(b)

Figure 5.18: LSM with (a) pre-rotation of patch and (b) with reshaping of template and patch. Left: template Nadir, middle: Backward, right: Forward. The initial patch and position is shown in black and the final one in white.

25 points matched with grey values and gradients with and without pre-rotation, the points that failed to be matched in each case were: 11 and 9 points when using grey values without and with pre-rotation, respectively and 5 and 3 points when using gradients without and with pre-rotation. In Figure 5.16(a)) an example is illustrated showing the decrease of iterations (compare bottom of Figure 5.16(a)) in case grey values and gradients are used, both with and without pre-rotation. The second and third image from the left shows the results derived by using grey values without and with pre-rotation (decrease of the number of iterations from 10 to 8), respectively and in the fourth and fifth images from the left, the results derived by using gradients without and with pre-rotation (decrease of the number of iterations from 6 to 4). Another example highlighting the improvement in success rate when gradients with pre-rotation instead of grey values without pre-rotation are used is shown in Figure 5.16(b). The use of gradients is favored in case of discontinuities, since it can improve convergence of the solution, though is more sensitive to the quality of the initial shifts approximation . As Figure 5.17 shows, if more than one edge is included in the patch and the approximate position is selected on a nearby edge, the solution based on gradients converges to the position.

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 intensity analysis of the template. The patch dimension across the patch can either be selected as 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 the 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. Figure 5.18(a) shows an example of pre-rotation with wrong matching for the Backward channel, while in Figure 5.18(b) a template reshaping is also included. Tests with the 20 points showed that false matching could be almost eliminated, as long as the derived edge directions in the template and patch are reliable. In all the above procedures, the angle ϕ could also be estimated from the orientation of the edgels, e.g. if the Canny operator is used beforehand. This may have the advantage that small gradients (e.g. noise) do not influence the computation of ϕ, as in Equation 5.22.

5.4.4.2 Edge least squares 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 according to the procedure described in section 5.3.2.2. Edges are extracted only in the template image and for each edge to be matched the patch dimensions of the template are expanded in order to contain the edge points and the end-points (vertices) of the edge (see left patch in Figures 5.20(a), 5.20(b), 5.20(c), 5.20(d), 5.20(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 x 5) to avoid singularities of the system, as the number of observations should be larger than the number of unknowns and the smaller weights must not be zero, especially for short edge lengths (and thus small patch sizes), in order to avoid singularities of the system. 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 adopted from the conformal transformation computed between patch and template for the edge middle point. In Figure 5.19, the $x_b^{'}, y_b^{'}$ coordinates are computed from Equations 5.24 and 5.25. By setting $x_b = x_c + dx$ and $y_b = y_c + dy$ the Equations 5.26 and 5.27 are derived and the distances dx' and dy' are computed from the transformation parameters, according to Equations 5.28 and 5.29.

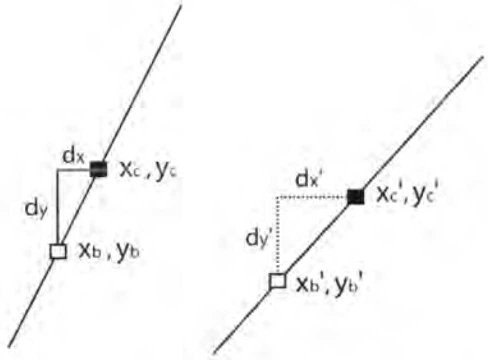


Figure 5.19: Derivation of approximations for edge points. The x_c, y_c are coordinates of the middle point on the edge of the template and the corresponding point with coordinates $x_c^{'}, y_c^{'}$ is in the patch image. Other edge points (x_b, y_b) with distances dx, dy from the middle point, are transformed to $x_b^{'}, y_b^{'}$ with dx', dy' derived from transformation parameters.

$$x_b^{'} = a_0 + a_1 x_b + a_2 y_b \tag{5.24}$$

$$y_b^{'} = b_0 + b_1 x_b + b_2 y_b \tag{5.25}$$

$$\begin{aligned} x_b^{'} &= a_0 + a_1(x_c + dx) + a_2(y_c + dx) \\ &= x_c^{'} + a_1 dx + a_2 dy \end{aligned} \tag{5.26}$$

$$\begin{aligned} y_b^{'} &= b_0 + b_1(x_c + dx) + b_2(y_c + dx) \\ &= y_c^{'} + b_1 dx + b_2 dy \end{aligned} \tag{5.27}$$

$$dx' = a_1 dx + a_2 dy \tag{5.28}$$

$$dy' = b_1 dx + b_2 dy \tag{5.29}$$

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 Figures 5.20(b) and 5.20(c), the dotted edge lines show these approximations. In 5.20(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 into 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. The division by two continues until the edge segments become less than 11 pixels long. At this stage no matching has yet been performed 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 (9x9 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 approximations 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 has been proven to be relatively robust, with fast convergence (3-7 iterations), and due to the often large patch size, it requires less accurate approximations than conventional LSM. This reduces the danger of multiple solutions and leads to a safer determination of rotation. Even the case of edges being parallel to the epipolar line can be partially treated, when at one of the end-points of the edge (assumed included in the template patch), other edges with a large enough angle to the epipolar line are included, e.g. corners on roof tops. However, the problem of edges parallel to the epipolar lines

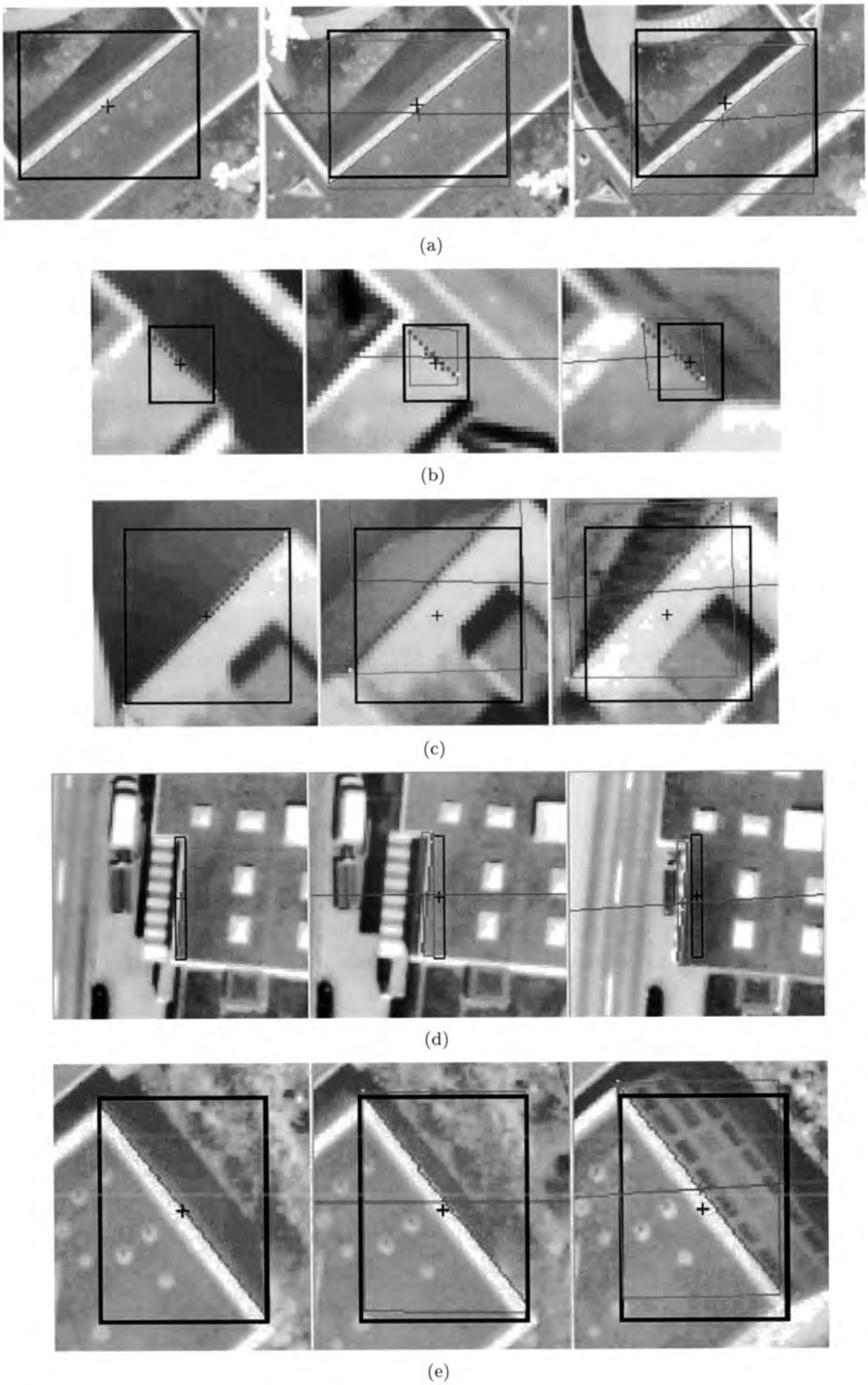


Figure 5.20: Examples of edge LSM. Edges are extracted in the template image, Nadir, (left) and matched in the Backward and Forward images (middle, right). Convergence of the least squares solution is achieved in 3 to 7 iterations. The dots indicate the edge points. Results are illustrated after the LSM solution for the middle of the edge. The black and red rectangles show the initial and final patches respectively.

is not generally solved with this method. Furthermore, occlusions may still cause false matching. Or, even if matching is correct, occlusions may cause wrong scale and rotation estimation in the conformal transformation, and thus lead to wrong approximations for the remaining edge points to be matched.

5.4.5 Single- versus multi-template strategy

The matching strategy consists of different steps or modules that have already been described in the previous sections in this chapter and a general overview exists in section 4.4. The analysis of the strategy starts from the core matching scheme, which is further utilized in more advanced matching combinations as explained below. All images that are used in matching are preprocessed and features are extracted in the template image. Figure 5.21 illustrates the processing steps of the base strategy for single points using Nadir as template and matching to Backward and Forward (here only an example as more images can be used simulteneously). After the initial multi-patch size matching (MP matching), filtering of possible mismatches takes place during the quality control (see section 5.5) and image rays that were not rejected in the quality control are used in LSM simultaneously to recheck the remaining points with stricter criteria.

The above core scheme can be applied to grid points, edgels and aggregated edgels or edges, adapted to each feature class as explained in section 4.4 (see also Figure 4.3) and it can be embedded in a single- or a multi-template (ST or MT) strategy. The MT strategy is justified since one template may not suffice to detect and avoid problems occurring in matching, especially occlusions. The ADS40 permits to use more than one template and facilitates with the given configuration of channels the identification of errors. The following combinations of channels have a different role in matching:

- If best accuracy in Z is sought, then matching between Backward and Forward channels is favored, though is considered to be the most difficult case due to large geometric differences.
- Combination of Backward and Nadir channel is optimal for least 'left' occlusions.
- Red or Green and Forward channel is optimal for least 'right' occlusions.
- Red or Green can be included in matching, due to low noise level and spectral similarity to the panchromatic channels.
- Near-infrared is less favorable for matching as the radiometric differences to the panchromatic channels are large (Figure 5.22).

In the case of the MT strategy, where the three channels are used (Backward, Nadir, Forward), the left- and right-most are set as templates and matching is performed pairwise by starting from one template and matching to all other search images. If the point matching is not successful, the template is changed and the matching is repeated using the second template. Figure 5.23 illustrates the MT strategy by using two templates (Backward and Forward), in a total of four images.

Initially, the general matching scheme followed the concept illustrated in Figure 5.24. Matching was directed to edge features with attributes (aggregated edgels), since contours can support a better representation of discontinuities compared to single points. An extracted edge (section 5.3.2.2) consists of a chain of pixels of a certain length. The end-points of the chain are the vertices, which also serve as linking nodes between neighboring edges. Edge vertices were matched first and then points positioned on

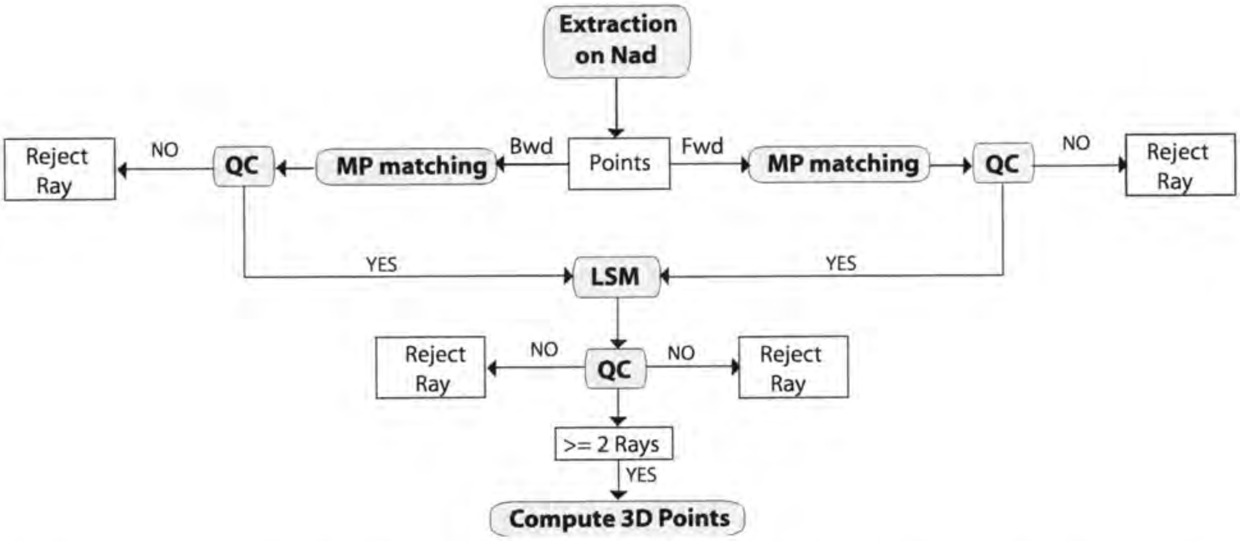
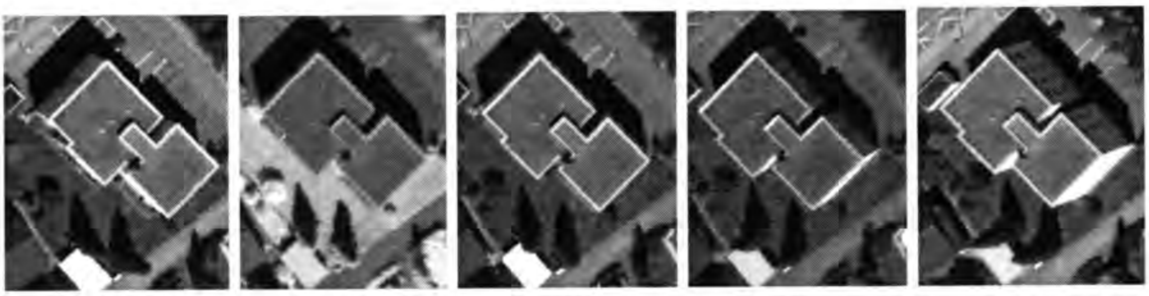


Figure 5.21: AIM strategy for single point matching. The Nadir is used as template and matching is performed on Backward and Forward. First multi-patch size matching (MP matching) is used, followed by quality control (QC). Successful image rays are re-matched simultaneously with LSM.



(a) (b) (c) (d) (e)

Figure 5.22: Example of radiometric and geometric differences among channels. Channel sequence: (a) Backward, Near-Infrared, Nadir, RGB, Forward. Geometric differences are larger between Backward and Forward. Radiometric differences compared to Pan channels and near-infrared are larger, as vegetation is lighter and MMO are darker.

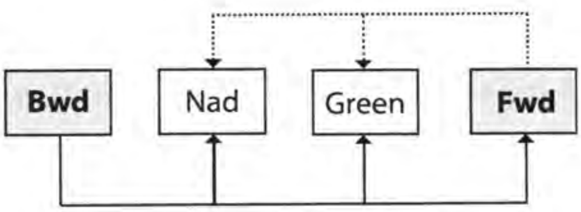


Figure 5.23: Multi-template strategy. Template images are indicated with red color. Pairwise matching starts from the Backward channel to other search images (thick black line). If matching fails, the Forward channel is set as template and pairwise matching is performed with the remaining images (black dotted line), except the first template, i.e. the Backward channel.

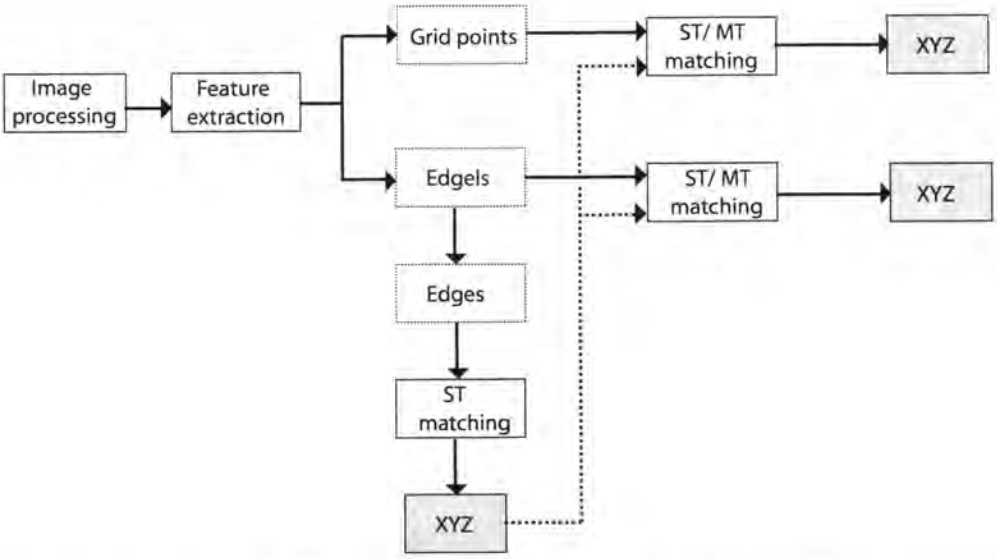


Figure 5.24: Matching scheme A. The linked edgels form the edge features, which are matched first in the upper levels. The derived 3D coordinates are used as approximations (dashed line) in lower pyramid levels, where grid points and non-linked edge pixels are included.

each contour were matched sequentially with an approximation of height, derived from the 3D coordinates of the vertices and the length of the contour (continuity and height constraint). Three possible cases are distinguished:

1. If both vertices at the end-points of the contour are indicated as successful matches, the height approximation for each point on the contour Hc_i is derived from Equation 5.30,

$$Hc_i = Hv_1 + i \cdot \frac{Hv_2 - Hv_1}{l} \tag{5.30}$$

   where Hv_1 and Hv_2 are the computed heights of the vertices, l the length of the contour in pixels and i the pixel number.
2. If the height value has been derived for only one vertex, then Hc_i for all contour points is equal to the height of the respective vertex, Hv_1 or Hv_2.
3. If none of the vertices has passed through the quality control, all the remaining contour points are excluded from matching and the edge is rejected.

Given that doublets (section 5.2.3) are utilized in the matching strategy, only the vertices are matched in level A and the contour matching is employed only in the lower doublet level B, the reason being twofold: the height approximation of contour points is related to the height accuracy derived for the vertices, which increases with higher resolution. Moreover, reduction of processing time is favored. According to the above, only the vertex points are transferred and matched in the upper doublet level A. If they are not rejected by the algorithm (3D coordinates are determined), they are matched also in the lower doublet level B. In case of vertex points are rejected in level B, but not in level A, the edge will be rejected, i.e. approximations from level A will not be used for the remaining contour points. The diagrams in Figure 5.25 illustrate the decision scheme and the derived height approximation for the

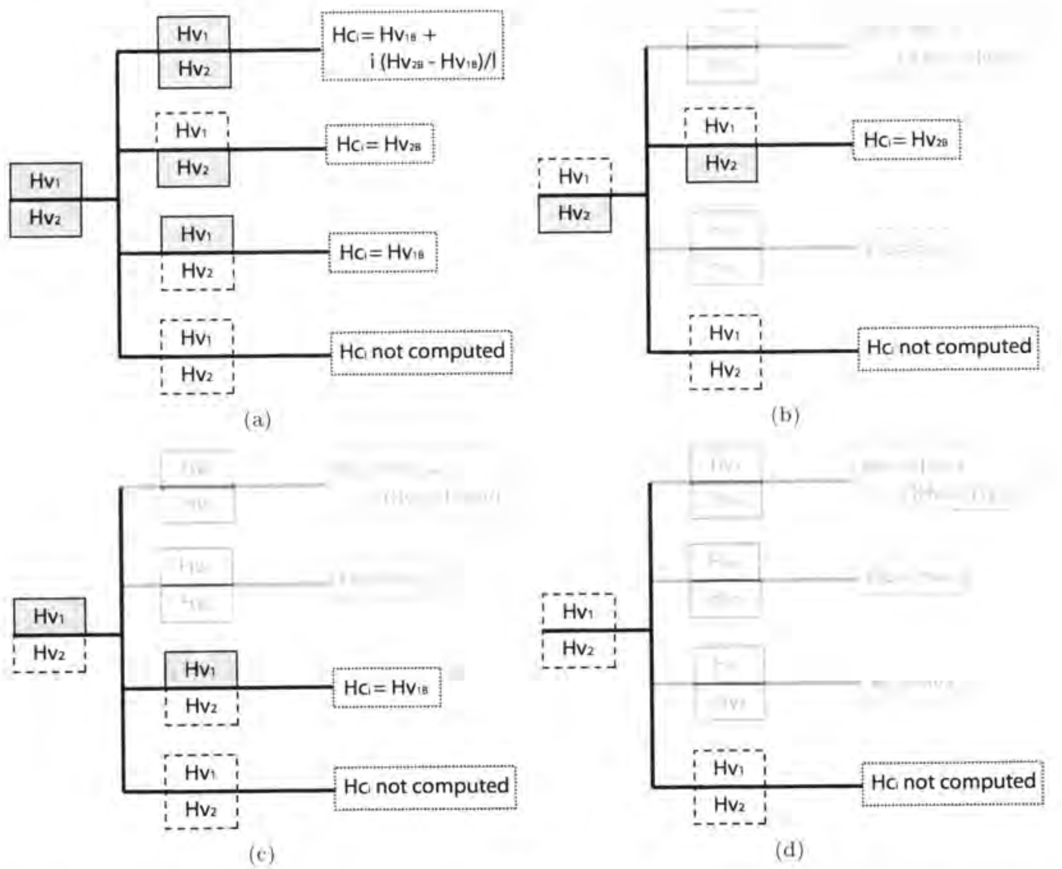


Figure 5.25: Decision scheme of edge matching combined with the doublets. Hv_1 and Hv_2 are the heights of the vertices and Hc_i the computed approximate height value for the remaining contour points. The grey boxes and boxes with dashed line indicate whether the height of the vertex has been computed successfully or not in both doublet levels A and B. In each of the subfigures (a), (b), (c) and (d) the possible cases to occur are not dimed.

contour points for each possible case. Both vertices have to pass through a two-step matching procedure, including quality control, in order to use their height in approximation derivation. Although stricter criteria are used, which increase the possibility of excluding the full edge from matching, the height constraint is derived with higher accuracy, in parallel to a reduction of processing time.

Yet, the matched edge features may not suffice for a complete surface representation and additional points were required to fill possible gaps. Grid points and/or non-linked edge pixels are then introduced in the lower pyramid levels, where an approximation of the surface already exists (derived from the matching of edges). The edgel LSM approach (section 5.4.4.1) has been used for both linked and non-linked edgels in the lower levels, at this stage only with pre-rotation of the rectangular patch in the search images. This strategy was used in test series A (section 6.5.1) and the following conclusions were drawn, justifying further modifications to the strategy:

- The matching of edges in the upper levels was less favorable, as usually building edges were matched and less features existed in open surfaces to ensure surface continuity. Consequently, when additional grid points or non-linked edgels were included in the lower levels, approximations came from

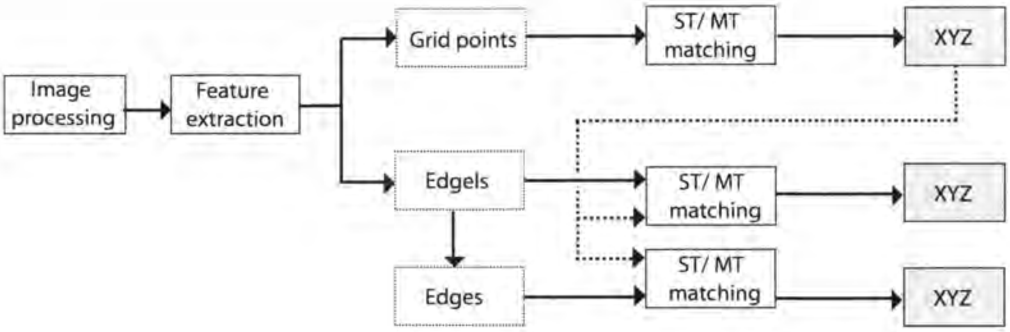


Figure 5.26: Matching scheme B. Grid points are matched first in the upper levels. The derived 3D coordinates are used as approximations (dashed line) in lower pyramid levels, where edges and non-linked edge pixels are added.

features far away. Many points were therefore rejected and modelling of the surface was rather incomplete.

- Edge matching had the disadvantage that although edges were formed, matching was performed point-wise, i.e. first for vertices and then for remaining points along the contour. Moreover, when both vertices failed to match in both levels A and B, the remaining contour points were rejected, and this led in cases of difficult matching areas to an inadequate number of 3D edges. Consequently, poorer approximate values were derived for other features extracted in the same and lower pyramid levels.

In the newest version of the strategy that has been applied in test series B (section 6.5.2), grid points instead of edges are used for the derivation of a coarse surface. Hence, less processing time is needed in the upper levels. Surface continuity and description are also improved. According to Figure 5.26, in which the revised matching strategy is illustrated, edges and edgels are introduced in the lower pyramid levels[6], using as input the coarse surface points generated by the matching of grid points. Furthermore, edge LSM (section 5.4.4.2) is used for straight edges with lengths of greater than 11 pixels, whereas the remaining edges are handled as in the previous strategy, where vertices and contour points are matched separately. When the MT strategy is applied to edges, the extraction and linking is performed in each template image, and matching to all remaining images. Based on matching scheme B, gaps along straight edges are avoided and common problems with multiple solutions are reduced (see also sections 5.4.4.2 and 5.5.2).

5.5 Quality control

The method should be able to provide reliable quality criteria for the evaluation of the matching results and perform an automatic self-diagnosis for error detection. The error detection and exclusion implies that the system is trained to analyze the matched points and detect possible mismatches. The training is based on a set of pre-defined error classes (section 5.5.1) and a combination of quality criteria

[6]the pyramid level in which edges and edgels are introduced, depends on the type of the area to be matched (flat, undulating, urban etc.) and more details are given in section 6.4)

(section 5.5.2). The quality control consists of two parts. In the first part, after multi-patch size matching, the points are assigned to error classes according to their quality measures. In the second part, LSM is used only for those points that were not indicated as errors in the first part and these points are rechecked with stricter quality criteria resulting from the least squares approach.

5.5.1 Error types

A number of different error types exist, which are related to different sources. The characteristics of the surface and the sensor, the illumination conditions, the configuration and scale of images, etc. are possible sources of problems and error types that occur in matching. The list with the possible error types includes:

- noise
- weak texture
- radiometric distortions (saturation, reflections, artefacts)
- multiple solutions
- occlusions
- surface discontinuities, which might also lead to occlussions
- errors in the sensor model, which imply that its transformations will be erroneous, due to the implementation of the sensor model.
- errors in the sensor model parameter values, due to problematic calibration or orientation (GPS/INS recordings).
- errors due to moving objects (e.g. cars, animals) and semi-transparent surfaces occur less frequently.

The first two error types of noise and weak texture can be significantly reduced through the radiometric data preprocessing (section 3.2) by performing noise reduction and contrast enhancement. Weak texture is mostly related to grid points, since extraction of feature points and edges is based on image gradients and few or no features exist in low textured areas, depending on the selected thresholds. With good approximate values for the matching position, optimal patch size, the use of constraints, the simultaneous use of more than 2 images and the quality criteria, most of the above problematic cases can be avoided, as explained in the next section.

5.5.2 Quality criteria and error detection

Since no single criterion is reliable, a combination of multiple criteria is necessary. The quality criteria are used for each image ray individually. Weak rays are excluded and the final 3D intersection is computed using only the good rays. The criteria receive different weights, while the threshold values for acceptance or rejection of a criterion include both absolute (strict, valid for all images) and relative (derived from matching statistics of all rays of each image) thresholds.

The quality criteria are used in the first part of the quality control and are calculated at run time. They include:

- the NCC as the similarity measure, computed during multi-patch size matching. The best NCC and patch are derived from the best-pass criterion, using two absolute thresholds, as explained in section 5.4.2.

- the 2^{nd} best NCC among the three patches used in multi-patch size matching.
- the sequential change of the NCC between patch sizes (also from multi-patch size matching, section 5.4.2).
- the change of position between patch sizes (multi-patch size matching).
- the angle of dominant edge direction with the epipolar line at the vicinity of selected match point.
- compatibility of left-right and right-left matches, i.e. back-matching .
- residuals from forward intersection.
- height consistency in the local neighborhood.
- approximative standard deviations for x and y coordinates (multi-patch size matching), which are computed for the template image according to Equations 5.31 and 5.32.

$$std_x = \frac{\sigma_o}{\sum g_x} \tag{5.31}$$

$$std_y = \frac{\sigma_o}{\sum g_y} \tag{5.32}$$

$$\tag{5.33}$$

where g_x and g_y are the grey level gradients in x and y, and $\sigma_o = \frac{1}{(\frac{r+1}{2})^2 + e}$ with r the best NCC derived from multi-patch size matching and e a small positive number, e.g. 0.001.

The quality criteria are combined according to possible error occurrences. Before deriving any conclusions as to which criteria reveal the above problematic cases, the quality measures for a series of points was examined after the multi-patch size matching was employed. The last criterion (approximative standard deviations for x and y) was not finally used, as it proved to be less reliable and also computationally expensive. Errors in the sensor model can be easily identified when constraints are applied (derivation of the *quasi*-epipolar curve) or when the number of iterations in the forward intersection increases significantly, often leading to non-convergence of the solution. In both cases inconsistencies in the image to ground and the ground to image transformations are observed[7].

Large errors can be detected and excluded based on the NCC and the residual values from forward intersection, for which the thresholds are adapted in each pyramid level. These thresholds can be strict, given that the density of the extracted and initially matched features is high. This mainly applies to points that have been initially matched with multi-patch size matching and for which the computed NCC was higher than the defined lower bound. These points are used to derive a relative threshold for all rays of each individual image by statistical analysis based on the distribution of the NCCs. The cutoff value for excluding the smallest NCCs is usually within the lowest 5-10% and should not exceed a maximum threshold, usually set at 0.8. The statistically determined threshold of the NCC may be higher than 0.8 for the upper pyramid levels, since occlusions and perspective differences are small and therefore points should exhibit higher NCC. In this case, the upper bound enables to avoid exclusion of points with high NCC. For the residual values, absolute thresholds are computed for x (in flight) and y (across flight) directions and are set to 0.2 and 1.0 pixels, respectively. These values are given with respect to the full resolution level and for higher pyramid levels are multiplied by a factor of 2^n, where n is the number

[7]By using as input a set of image coordinates and an approximate height value (average terrain height), the image to ground and ground to image transformations are sequentially employed. Thus, the differences in image coordinates are computed ($2D \rightarrow 3D \rightarrow 2D$) and should be smaller than 1 pixel value.

of the pyramid level. As the residual values are related to the computations and accuracy of the sensor model, we analyzed the residuals of image coordinates for a large number of points, both correct and false matches in different levels of the image pyramid and also residuals from the forward intersection of image coordinates of known GCPs, taken from different datasets prior to deriving absolute thresholds. The first filtering of erroneous points is based on their residual values and the NCC threshold. Based on the behavior of the NCC in multi-patch size matching that is discussed below, and on the remaining criteria, multiple solutions and occlusion can be identified and filtered from the data:

- In the case of an occlusion, the NCC would be small and the change of the similarity measure would be generally decreasing from the largest to the smallest patch, or the NCC of the second patch would deviate from the other two and would be smaller. In the first case, the threshold is derived by statistical analysis based on the differences of the NCCs from the largest and smallest mask. In the second case, an absolute empirically derived threshold is used.
- In the case of a multiple solution, apart from the case already mentioned in multi-patch size matching, the 2^{nd} best NCC would be probably close to the solution found and would be similar to the highest NCC (difference of approximately 0.005). As these measures are just an indication of a multiple solution, the angle of the edge dominant direction to the epipolar line (in case of edge or edgel matching) and the compatibility of left-right, right-left match are included as additional criteria. In the case of multiple solutions along the epipolar lines, the angle to the epipolar line would be less than 15° (case of feature almost parallel to the epipolar line) and the height consistency in the local neighborhood is checked by a median-type filter (4 neighbors at the patch corners are selected and matched) to detect and eliminate spike errors. In the above case, the angle is derived from the operator (Canny or gradient thresholding) during the edgel extraction stage. As this error case of multiple solutions will usually occur for features along edges almost parallel to epipolar lines, it is possible that points off-the-edge nearby would either lie in low-textured areas, e.g. shadowed areas, where less reliable points can be extracted and matched, or on different height planes (i.e. edge discontinuity) and the point would be identified as a spike and be rejected. For all other points with angle to the epipolar line, larger than 15°, if the back-matching results are not compatible with the initial solution (deviation of a maximum of 1.5 pixel), the point is disregarded.

LSM is employed in the second part of the quality control for refinement and verification of the points (good rays) that were not indicated as errors. Matching is repeated using the smallest mask (see also section 5.4.4) from the multi-patch size matching and the conformal set of parameters. In the case of an edge pixel, the patch dimensions are altered and initial approximations for rotation are derived according to the edgel LSM method (section 5.4.4.1). As far as quality and blunder detection are concerned, Baltsavias (1991) refers to several criteria that should be chosen in LSM for error detection and the following have been selected:

- number of iterations
- alteration of the size of parameters in each iteration
- size of parameters
- change of x, y coordinates between LSM and multi-patch size solution

- residuals from forward intersection using the LSM solution

The number of iterations is a rather good criterion, assuming that the approximations are good and that better precision is sought. In parallel, variations in the parameter values (magnitude and sign) in each iteration have to be observed in order to evaluate the stability of the solution, i.e. when geometrical constraints are enforced, oscillations for the x shift will serve as an indication of a multiple solution case. However, the threshold for the iterations should not be set too high (maximum number of 15 iterations), considering that fast convergence should be achieved, since the initial values should be close to the correct solution. The size of parameters, especially the estimated value for scales should not exceed a an upper (> 3.0) and lower value (< 0.3) and the difference to their initial values should be small. The initial values for scales are set to 1 and The variation of x, y coordinates from their initial values (multi-patch size solution) is also checked and if it is above a certain threshold (2.5 pixels) the point is rejected, assuming that the initial values should be close to the solution. The decision as to whether the LSM or the multi-patch size solution should be finally accepted is based on the residual values. If the residuals from LSM are higher than those obtained from the multi-patch size solution, the 3D coordinates of the latter are used.

The above presented scheme for quality control applies to points and is based on the combining two matching techniques, and on a gradual evaluation of the matching results and exclusion of problematic cases. The thresholds for the quality measures are either derived empirically or computed based on quality measures for each image ray individually. In the first stage, the focus is on the analysis of the behavior of NCC, as it is one of the most discriminating criteria. Thus, the danger of rejecting good points cannot be avoided, especially in the second stage of quality control, where LSM is introduced with stricter quality criteria. This can be less of a problem, if the remaining good points suffice for a good surface representation. Gaps can still exist in some cases in areas where feature points have been selected on edges parallel to epipolar lines and these were rejected in the quality control. However, this problem could be solved with the edge LSM approach that has been discussed in section 5.4.4.2.

Chapter 6

Evaluation of matching results

6.1 General description

Two series of tests took place in two different epochs in order to analyze the performance of AIM algorithm and compare it with the results of manual measurements. Based on the results of the first benchmark test (section 6.5.1), the components of edge matching (section 5.4.4.2), error detection (section 5.5) and the overall matching strategy (section 5.4.5) were further improved and integrated into the algorithm by the time of the second benchmark test. During the first tests one dataset was available, whereas in the second test phase a second dataset was also used, this being acquired over a different area (section 6.2). In addition, the acquisition of reference data was an important aspect that had to be analyzed, as the analysis of the results is strongly influenced by the extraction method (manual measurements, acquisition of laser scanner data, GPS measurements) and the accuracy of the reference data (section 6.3). In the second test series (section 6.5.2), the results derived with AIM were also evaluated against the output of the commercial system Socet Set v4.4.1 (SS) (section 6.4). At the time, this was the only commercial system that was able to import and process ADS40 imagery.

6.2 Datasets

The acquisition of datasets with minimum radiometric problems (artefacts) and free of geometric problems (sensor model, calibration and orientation) was a difficult task, as the camera system was under constant fine-tuning. Two ADS40 datasets were used in these investigations. The first was acquired over the rural area of Waldkirch in Switzerland and was used in both benchmark tests. The second was over the densely developed city center of Yokohama, Japan. This was available only in the second benchmark test. Since the radiometric quality and the block geometry of the second dataset was poor, and only the Waldkirch dataset fulfilled the requirements for good geometric accuracy and image quality, the description and evaluation of the Yokohama dataset will not be covered in this chapter, though a summary can be found in the Appendix.

The Waldkirch block was flown with the FP1 camera-type by LGGM in May 2002, and consisted of 4 parallel and 2 cross-strips (see Figure 6.1). It comprised panchromatic and multispectral imagery in Lev0 (raw) and Lev1 products with 0.20 m ground sampling distance. The coordinate system datum that was used was WGS84. The camera was radiometrically and geometrically calibrated (section 2.3.2). Interior orientation and IMU misalignment parameters were estimated by photogrammetric means over a test field, set up by LGGM in Switzerland. The orientation parameters acquired by the on-board GPS/INS systems were adjusted and refined through the AT process. Tie points were automatically measured using the Automatic Point Matching (APM) module of the SS software. Errors could be visually controlled and iteratively eliminated. The GCPs in the Waldkirch dataset were measured by GPS instruments and the measurement accuracy was less than 5 cm (2-3 cm). Non signalized points on road markings (at the corner or the middle of one side) were selected in the field. Their image coordinates were later measured in SS with subpixel accuracy. These GCPs were distributed at the block corners (Figure 6.1), resulting in a more stable block geometry compared as to the Yokohama dataset (section A.1).

The derived geometric accuracy from bundle adjustment in terms of a-posteriori standard error was $2.5\mu m$, and Table 6.1 summarizes the acquisition and bundle adjustment parameters of the dataset. The radiometric quality and interpretability of ground object features was good. All images used for DSM extraction were first preprocessed (section 3.2) to enhance feature definition and the radiometric balancing of channels (Figure 6.3).

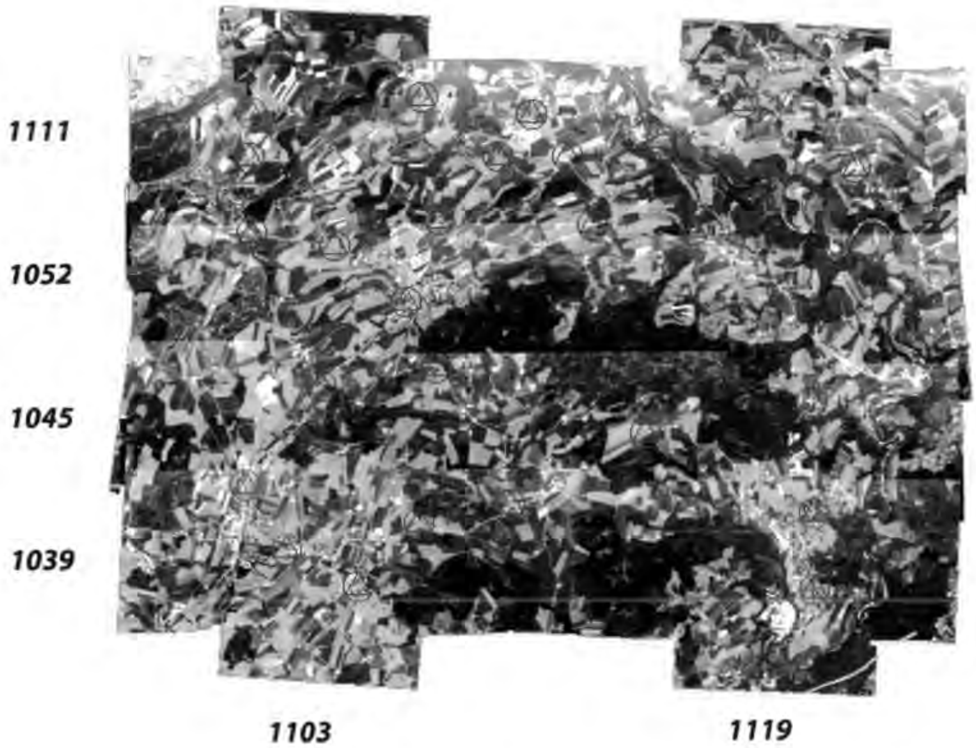
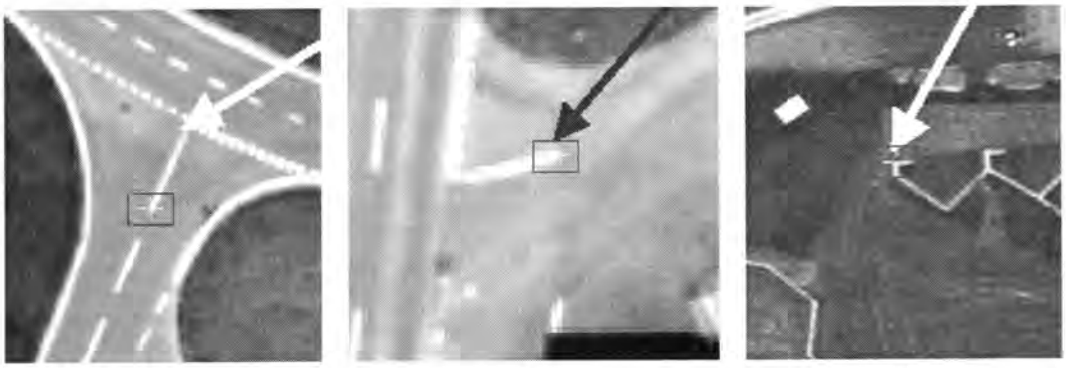


Figure 6.1: Waldkirch block. The strip numbers are labelled by time, the first two digits being the minutes and the last two the seconds of the acquisition time. The GCPs and check points are marked with red and blue color respectively.



(a) (b) (c)

Figure 6.2: GCPs in the Waldkirch dataset. The points are selected within the marked rectangles, usually in the middle of the shorter side of roadmarks.

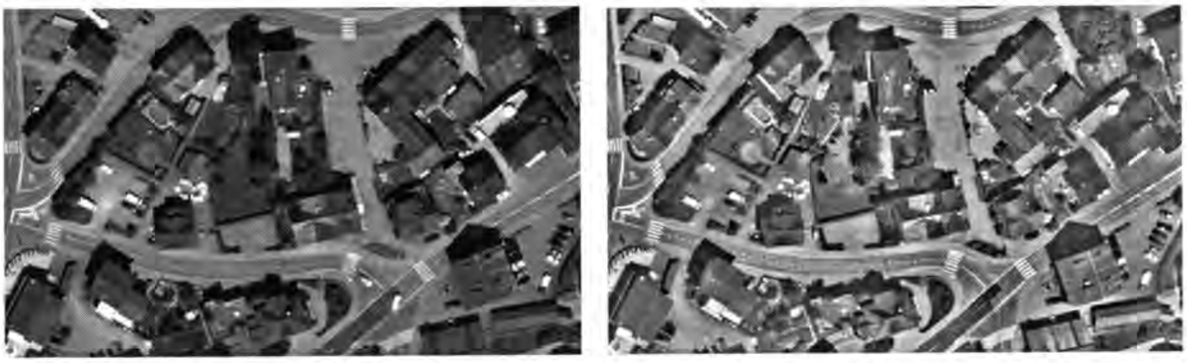
Parameters	Waldkirch
Coord. system	UTM in WGS84
Acquisition date	May 2002
Camera of type	FP1
No. strips [paral.– cross]	4–2
Ground pixel size [m]	0.20
Sensor pixel size [μm]	6.5
Radiometric quality	Good
No. GCPs	8
σ_0 [μm]	2.5
Flying height [m]	∽2000

Table 6.1: Acquisition and bundle adjustment parameters.

6.3 Reference Data

For a reliable analysis and evaluation of the results of the automatic DSM generation, the reference data had to fulfill the following requirements:

- The accuracy should be better (approximately 3-10 times) than the typical accuracy that could be derived from the ADS40 images.
- Multitemporal differences between the reference and automatically derived DSM should be a minimum.
- Errors arising from orientation differences between the ADS40 and data used to derive the reference DSM should be avoided (i.e. when frame imagery is used to derive the reference DSM, existing planimetric errors due to the different orientation should not influence the accuracy of the result).
- Conversion errors between coordinate systems need to be insignificant. These can arise due to datum definitions, applied formulas and transformation between orthometric and ellipsoidal heights and



(a) (b)

Figure 6.3: Original image quality of the Waldkirch dataset (a) and with radiometric preprocessing applied (b).

they can degrade accuracy. Therefore, the optimal approach is to acquire the reference data in the same coordinate system as the ADS40 project.
- The number of reference points should be relatively large to enable a reliable evaluation.

With respect to these considerations, there are alternative means to obtain reference data, but they were not available in the project. These alternatives are:

- Manual measurements in stereo mode using frame images of larger scale were disregarded for three reasons: multitemporal differences will likely exist, as well as errors due to the different sensor models, orientation errors cannot be distinguished from matching errors. Moreover, if the coordinate system differs, a conversion between the different coordinate systems (e.g. Local Coordinate System to WGS84 UTM) should be applied.
- Acquisition of laser scanner data to a height accuracy of better than 0.5 m, especially at a different acquisition time than the ADS40 data, could not be realized.
- GPS measurements in the field on characteristic points can be carried out with high accuracy. However, taking into account that a few points do not suffice to describe the surface, plus points on the tops of buildings are difficult to measure, the method was not applicable in this case.

Moreover, as the aim is to evaluate matching accuracy, treating measurement errors independently from sensor model or other additional error sources, the automatically extracted points have been compared with manual measurements derived from ADS40 images. The same orientation data were used in all measurements, automatic as well as manual. In the first series of tests (test series A), irregularly distributed mass points, collected in stereo mode in SS, have been classified to: bare earth (BE), man-made objects (MMO) and trees (T). Mass points on MMO, have been measured on end points of roof edges. For single trees the points were collected as follows: one point close to the top of the tree and two to three points close to its base. For groups of trees several points were measured on their tops and points around their perimeter and on the ground. A quality analysis has been performed for these three different classes.

In the second series of tests (test series B), mass points were refined by a second more experienced operator and breaklines, were collected via manual stereo measurements in SS. Mass points have been separated into two classes: ground points close to the perimeter of buildings (hPts) and ground points defining the BE surface, excluding points close to buildings (Pts). This classification is justified as measurement errors should be treated separately from modelling errors of the surface in 2.5D. Modelling errors arise due to the surface modelling and usually for point positions close to abrupt surface discontinuities (e.g. close to building outlines). As it can be seen from Figure 6.4, even if the measurements (points at the top and bottom of building at the left and right most) are accurate, the modelled profile does not follow the true profile of the surface and for the selected point at the base of the building a large height difference (Dh) is measured. Therefore, it was important to separate measurement from modelling errors in order to evaluate the internal matching accuracy. If we would analyze in detail also modelling errors, then the computation of 3D orthogonal distance (d) between the surfaces would be used (Poli et al., 2004). In addition to these two mass point classes, breaklines were extracted along both discontinuities on the ground and on buildings or man made objects (MMO), even on small roof features (Figure 6.5). Points along the breaklines were interpolated every 0.20 m, corresponding to approximately 1 pixel in image space, using microstation as add-in component in SS. Forest areas and trees have been excluded from the measurements as it was difficult to manually extract reliable 3D coordinates for imprecise and unclear shape. The estimated height accuracy of the manual measurements was $20 - 40$ cm, however accuracy was degraded for points on the ground in areas surrounded with buildings, due to shadowing and low texture. For the analysis, the reference data were interpolated in the matching DSM, using the "quality statistics"utility of SS to compute the elevation differences.

6.4 Matching parameters of compared software systems

As already mentioned, in the second benchmark test the SS performance on DSM generation was to be compared with AIM and evaluated on a common basis, namely against the reference data. In SS, the adaptive method or AATE (Adaptive Automatic Terrain Extraction) was used. Adaptive matching can import more than two images, can generate regular grids or triangulated irregular networks (TINs), can affect changes to matching parameters based on an "inference"engine, and can compute the mean terrain inclination in small neighborhoods. Based on this inclination and image exterior orientation, the

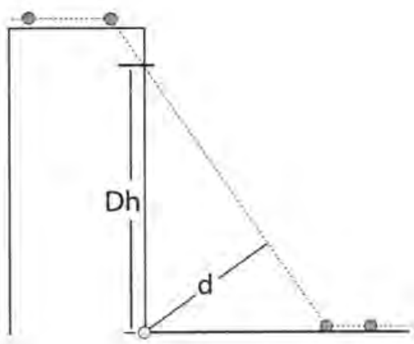


Figure 6.4: Modelling problems. The true profile is the thick black line and the modelled profile is the dashed line.

(a)

(b)

Figure 6.5: Reference data overlayed on part of the Nadir image of 1052 strip of the Waldkirch image block. Shown in the top image (a) are mass point and breaklines on the ground (BE), without points close to the perimeter of buildings, in the bottom image (b) breaklines on MMO are indicated.

best two out of all available images are selected. This selection is preferrable (e.g. Bacher, 1998; Baltsavias et al., 2001a) and can lead to better results compared to the non-adaptive matching as problems due to occlusions and large perspective differences can be reduced through an appropriate choice of images. However, several problems have been reported, apart from the two previous mentioned authors, also from photogrammetric firms, with respect to the matching results in flat areas. These problems can be due to the poor algorithms used in the choice of the two best images and also because of the mean terrain slope seems to be computed over a neighborhood that is too large. In some cases, AATE produces severe errors at image borders, i.e. the terrain is flattened, which is an indication that the implementation of AATE has pitfalls. 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 utilized in SS, namely matching is performed on selected grid positions, though points that fail to be matched are not interpolated from neighboring successfully matched points. The AIM method is based on a combination of area and feature based matching techniques and different types of primitives (area patches, single edgels that belong to contours and edges) are combined based on the type of the terrain (rugged, steep or flat). As automatic matching in each system is based on different strategies, the assessment has been focused on the quality of the final product, namely the DSM itself. Alternatively, an analysis on a different level, namely forcing the system parameters to be relatively similar, would not be realistic for SS as it has certain limitations for full control of the matching strategy and error detection. For AIM, modifications would be feasible in terms of implementation to a certain extent, to adapt some of its parameters to those of SS, by using area-based grid matching for example. However, this would be less favorable as the AIM method takes into consideration several characteristics of the ADS40 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 equal and the same number of images has been used. In SS, the TIN version without additional filtering (elimination of tress/buildings/other objects) has been used and similarly for AIM, no additional smoothing has been applied and the raw matched data have been analyzed (irregularly distributed points).

6.5 Test results and error analysis

6.5.1 Test series A

Due to the landcover variations in the Waldkirch region, three test areas were selected for DSM extraction and a common naming scheme was adopted[1] and used throughout the tests. Considering also the programming effort for optimal tiled processing (handling boundaries of overlapping tiles), which was beyond of the scope of this research, the size of the test areas was constrained based on the computer limitations in memory allocation for processing each area as a single block (i.e. without dividing the area into smaller parts or tiles).

[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. Mainly for comparison purposes with future tests.

(a) (b)

Figure 6.6: Selected areas in test series A. In images (a) and (b) are the respective W1052A and W1052B areas. The images have been processed (brightened) for visualization purposes.

- The W1052A area was relatively flat with good texture in agricultural fields and its size was approximately 2000 x 1000 pixels (Figure 6.6(a)).
- Area W1052B was of size 1500 x 1000 pixels and it was characterized by bare ground and man-made objects (Figure 6.6(b)).

In addition to the areas listed above, a part of a forest was also processed but results have been omitted due to the less reliable results obtained in the analysis. Trees are considered to be one of the most difficult matching cases, plus the accuracy of the manual measurements on the tops of trees was less than required. In addition, many automatically derived points on trees were indicated as errors (poor quality measures, e.g. NCC below threshold, large differences in x,y positions in multi-patch size matching) and were excluded from the final DSM. Additional errors arising from height interpolation of points on trees from nearby points, also adversely influenced the analysis. With respect to this, the RMSE for class T (section 6.3) was higher than 1.5 m, and was thus not a good indicator of the matching accuracy.

In the preliminary results (Pateraki and Baltsavias, 2003), it was shown that the accuracy increased with the number of images used; there was a significant improvement when using three instead of two images, and less of an improvement when using four instead of three. In these tests three images were used, the Backward, the Nadir and the Forward channel. The analysis focused on the comparison of the strategies, single-template (ST) vs. multi-template (MT), as well as on how different selected features for matching influenced the accuracy of the result. In the ST strategy, the Nadir channel was used as the template and in the MT strategy the outer channels were used, in this case, the Backward and Forward panchromatic channels. In all tests, six pyramid levels combined with the doublet strategy were used (section 5.2.3). The raw DSM points without any postprocessing (e.g. filtering, modeling of breaklines) and interpolated elevations of the manually-measured points (section 6.3) were utilized to compute elevation differences and derive statistical quality measures (mean with sign and RMSE). Landcover variation in the test areas justified the use of different strategies in terms of template and feature selection. In the area W1052A, ST and MT strategies were applied on extracted edgels[2] (P) and the ST strategy on grid points (G) with a predefined step of 5 pixels, which approximately corresponds to 1m on the ground. Due to the type of the terrain in this area only the use of the ST strategy would suffice and therefore the MT strategy was applied only for the single edgels (P), simply for comparison purposes.

Area W1052B included complex objects (small and large building blocks, trees close to roofs, etc.). Either a few or no points at all were extracted in shadowed areas close to the bases of buildings. Through the integration of edges (E) with attributes (i.e. chain of edgels), the representation of discontinuities improved. Following the matching scheme A, explained in section 5.4.5 (Figure 5.24), edges with attributes (E) were introduced in matching in addition to unlinked edgels (P) and grid points (G) for this area. Edges (E) with attributes were matched in all pyramid levels and unlinked edgels (P) and grid points (G) were used in the lower levels (1 and 0) to generate a denser surface. The points of each contour (belonging to type E) were not only constrained along the epipolar line but also within the computed height range, which was derived from the successfully matched vertices of the contour, from the same pyramid level (see section 5.25). With respect to the contour matching, 100 manually-measured roof corners were compared with identical automatically derived points. The mean Z difference was ö 0.33 m and the RMSE was 0.65 m. However, not all extracted roof corners were successfully matched due to the fact that the geometrical differences between template and search could not be modelled by the affine transformation in LSM.

The error statistics for areas W1052A and W1052B are shown in Table 6.2. For area W1052A, as the terrain was smooth (no discontinuities) and relatively many features were extracted in the agricultural fields due to contrast enhancement, computation of edges with attributes (E) was meaningless as well as costly in terms of time. Even with higher thresholds for Canny, many features were extracted in this area due to the contrast enhancement that was applied in the whole image. Similar accuracy was obtained via the ST and MT strategies for both grid point- and edgel matching with the RMS error for grid point and edgel matching being 0.43 m and 0.45 m, respectively. However, extraction and matching of grid points compared to edge pixels was more favorable in terms of time, since 6000 grid points gave similar results to 308000 matched edgels. Measurement accuracy decreased slightly for the MT strategy in this area (slightly larger mean) as the 3D coordinates that are derived from matching from the second template, are computed from two image rays (Forward and Nadir) instead of three.

For area W1052B, accuracy degraded since some points along building outlines were rejected by automatic error detection when MMO were included and elevation was interpolated from the nearest points which apparently were on the ground or on near-by edges. Therefore, these interpolated elevations were indicated as errors in the analysis but in reality not enough points were successfully matched. It is also apparent that the difference in landcover compared to test area W1052A decreased the accuracy even when results are only compared with bare earth (BE). This is due to the fact that the MMO object surface is more complex, without many open areas as in test area W1052A. Although these problems are common for both strategies, the performance of the MT strategy is better in the presence of occlusions. Even though in MT strategy only two rays can be used with the second template to compute the 3D

[2]Edgels: single points on edges that are not linked to form edges

Test data	Type of points/number of comparison points	Strategy/number of matched points	RMSE (m)	Mean with sign (m)
W1052A	BE/100	ST - P/308000	0.45	0.18
		ST - G/6000	0.43	0.17
		MT - P/367000	0.45	0.27
W1052B	BE/80	ST - EP/62000	0.68	0.14
		ST - EG/23000	0.67	0.13
		MT - EP/65300	0.64	0.12
		MT - EG/26456	0.62	0.11
	MMO/350	ST - EP/62000	1.17	0.50
		ST - EG/23000	1.14	0.48
		MT - EP/65300	1.07	0.42
		MT - EG/26456	1.02	0.39

Table 6.2: Error statistics of reference data (manually measured), BE and MMO compared to automatically generated DSMs for test series A using single- (ST) and multi-template (MT) strategies. Grid points, edgels and edges are indicated as G, P and E respectively. EP and EG indicate the combination of edges (E) with unlinked edgels (P) and grid points (G) respectively.

coordinates of the point (if matching with the first template fails), more points can be matched and the description of a complex surface improves. Grid points (G) deliver similar but faster results compared to edgels (P) for both MT and ST strategies. Moreover, grid points are used to improve surface continuity and fill possible gaps. Discontinuities are described by aggregated edgels[3] (E).

6.5.2 Test series B

Based on the results of the first tests, the AIM strategy was modified. The coarse DSM in the upper levels was derived by matching of grid points for better time performance, edge LSM was included to improve modelling of discontinuities and LSM error detection was refined (better thresholds for scales and shifts). Moreover, it became evident from the first tests that the manual measurements should be improved: surface discontinuities should be described by more points, namely through better extraction of breaklines for a more reliable analysis, plus points close to building outlines on bare ground should be separated from other points on bare ground. The test areas W1052A and W1052B (Figure 6.7(a)) were extended to approximately 2000 x 2000 and 2500 x 1800 pixels respectively, and in addition the W1052C area, of 2000 x 2000 pixel size was included. This comprised a undulating terrain and included sparse houses, trees and small salt lakes (Figure 6.7(b)).

The AIM and SS systems were compared in terms of DSM extraction accuracy, using the same manually-measured data. The parameters of each system were modified according to the characteristics of each area. In these tests, the three panchromatic and green channels were used. As mentioned already in section 6.4, in SS the AATE method was used in all tests in combination with the TIN option. For the test areas W1052B and W1052C, a 1 m grid interval in object space was used and for W1052A the grid interval was set to 1.5 m as the terrain was less undulating. In AIM, the strategy of each area varied, namely ST strategy was used for areas W1052A and W1052C and and MT strategy was used

[3] or Edges with attributes

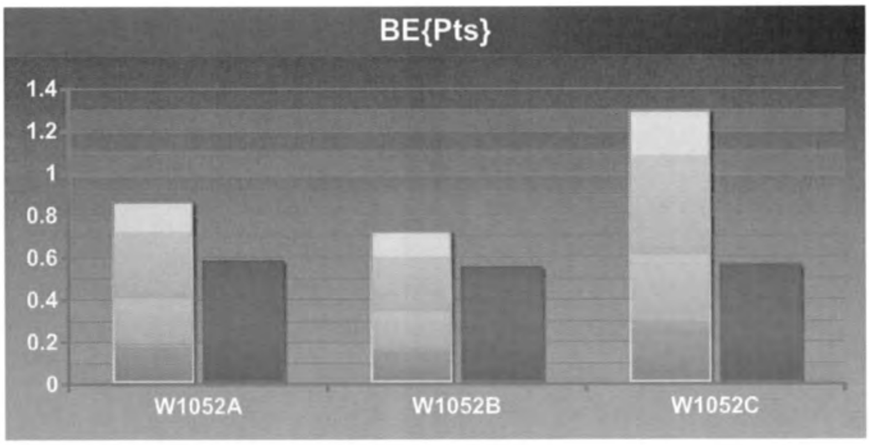
(a)

(b)

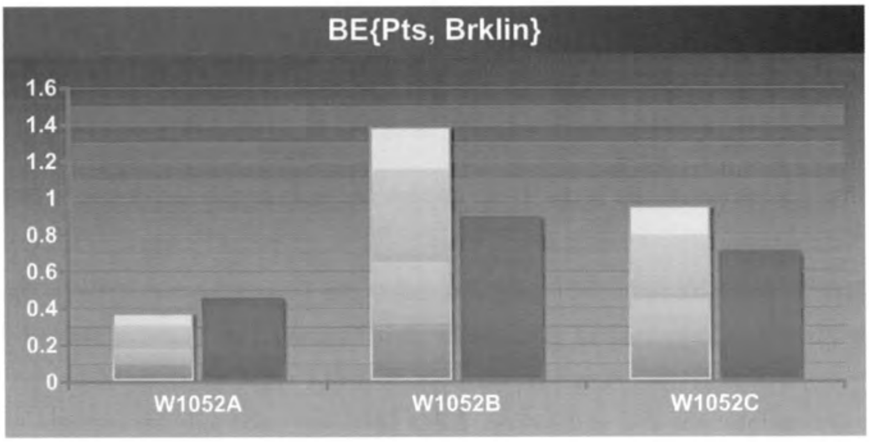
Figure 6.7: Selected areas in test series B. In images (a) and (b) the respective W1052B, W1052C.

for area W1052B. In total six pyramid levels were used. An approximate surface in the upper pyramid levels was generated by a matching of grid points (5 m interval), and further refined by the matching of edges in the lower four levels. However, for area W1052A edge matching was used only in the two lower levels as the surface was less undulating. Edge LSM was used only for straight edges longer than 11 pixels. For the remaining contours, each point on the contour was matched individually by enforcing the continuity and height constraints in multi-patch size matching and using edgel LSM for the points that have passed the first stage of the quality control (section 5.5.2). In these tests, single edgel points (P) that were not aggregated were not used in any of the areas. Moreover, as in test series A, also for these tests the raw DSM points without any postprocessing, and with interpolated elevations of the reference-measured points, were utilized to compute statistics of elevation differences (RMS, mean with sign, absolute maximum and standard deviation). The results of the analysis, performed for the above areas, are shown in Table 6.3 and Figures 6.8(a)- 6.8(c), 6.9(a)- 6.9(f).

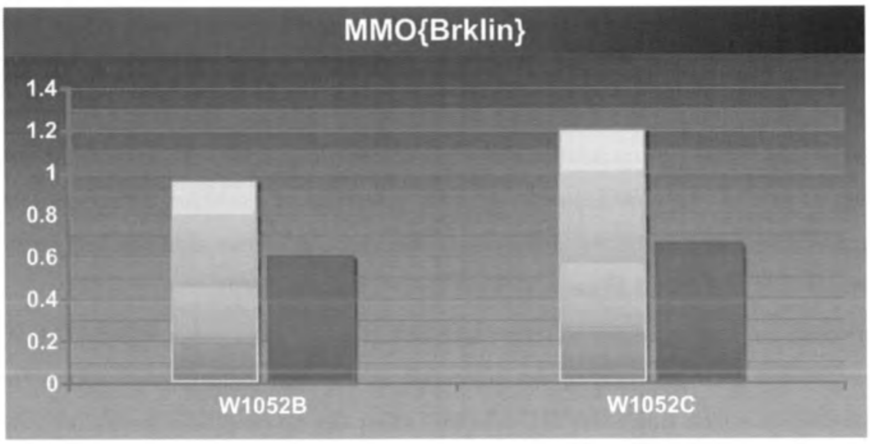
Errors with respect to points on bare earth (BE) and on man made objects (MMO) were analyzed separately (see Table 6.3). Mass points are indicated as Pts and interpolated points from breaklines as Brklin. The number of single points on BE was less than the interpolated breakline points and the first were measured in locally flat areas. Modelling errors are computed separately 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 relatively flat areas (W1052A), less differences in the accuracy could be observed. Errors in DSMs derived with SS were significantly higher, in locally flat areas and in the case of MMO, also at breaklines. AIM shows a better performance in the case of discontinuities, as accuracy increased by approximately two times and errors were less than SS for all areas in the case of



(a)



(b)



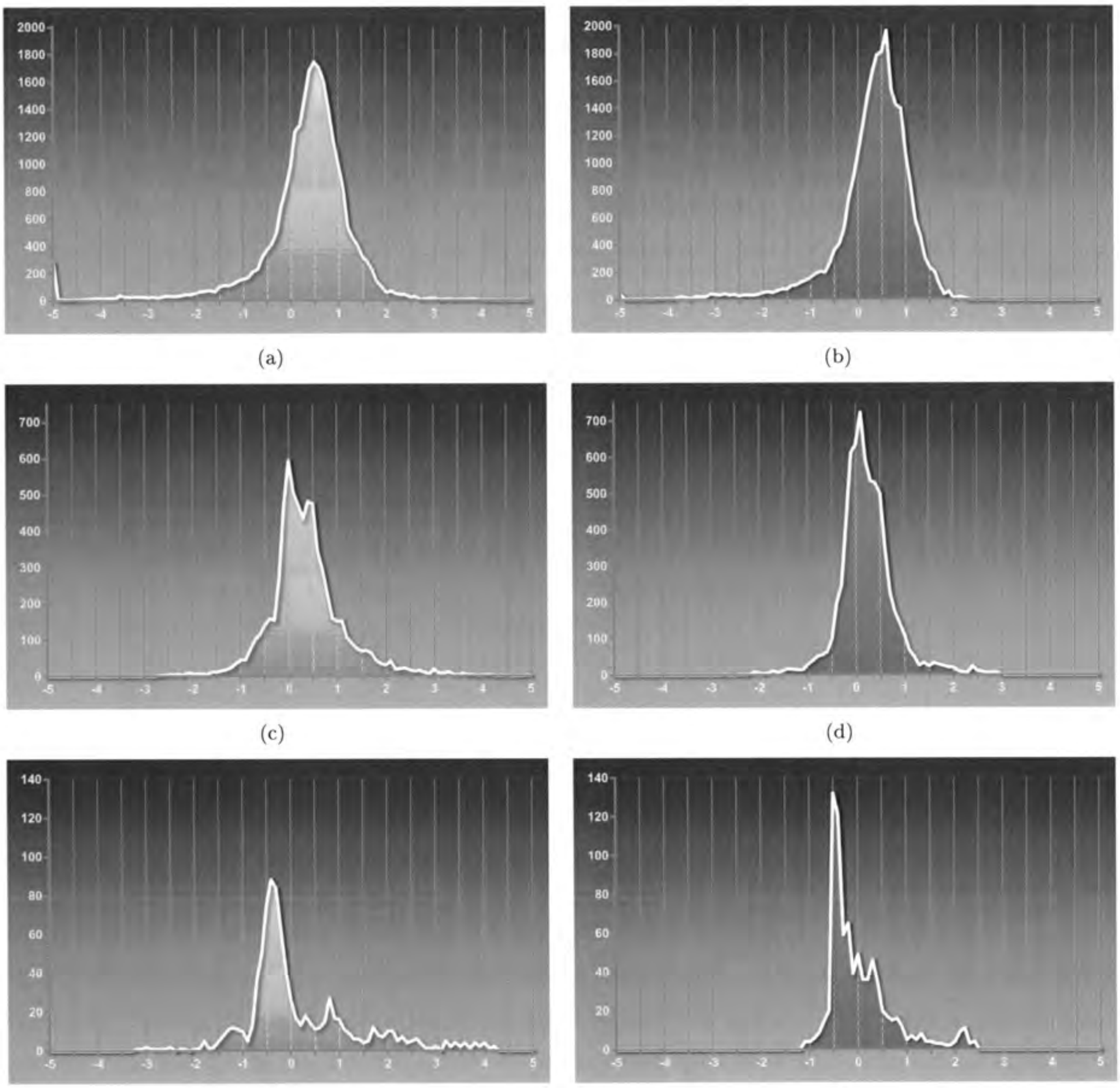
(c)

Figure 6.8: Graphical representation of RMSE values (in m) derived from test series B. In figures (a), (b) and (c) the RMS values (Y axis) with respect to BE{Pts}, BE{Pts, Brklin} and MMO{Brklin} for SS (red) and AIM (green).

Test data/number of matched points {SS/AIM}	Type of points/number of comparison points	System	RMSE (m)	Mean with sign (m)	Maximum absolute (m)	Std. dev. (m)
W1052A/ {141512/342692}	BE{Pts}/35	SS	0.83	0.02	2.65	0.83
		AIM	0.55	-0.04	1.07	0.55
	BE{Pts, Brklin}/1100	SS	0.37	-0.10	1.84	0.33
		AIM	0.45	-0.36	1.27	0.30
W1052B/ {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
W1052C/ {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

Table 6.3: Error statistics of reference data (manually measured) minus automatically generated DSMs for test series B. 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 and are used to discriminate modelling errors.

MMO. For both systems, accuracy degraded for points close to building outlines, as less points could be automatically derived in these areas and interpolation errors, often large, arose. In addition, planimetric errors, even if small, can also cause large height errors at surface discontinuities. SS has generated more points than AIM in all areas except W1052A. The reason for this is the inclusion of edges in the two lower pyramid levels, leading to the extraction of many points in highly-textured areas, e.g. the agricultural fields. The mean values for both methods show that matching generally measures higher elevation than the manual measurements, while for MMO matching results produce lower elevations. For area W1052A, the negative and positive height differences for the 35 points in class BEPts, justify the small mean values, close to zero. When the breakline points Brklin are included in addition to the mass points Pts the accuracy improves slightly for class BE{Pts, Brklin} due to the inclusion of breaklines. However, as most of the measured breaklines in area W1052A follow terrain profiles, which are not significant features in the image, the results are influenced by small interpolation errors. Breaklines of this type are also included in the other areas and they contributed to modelling errors. By observing the graphs of the frequency distributions of the height differences for classes with significant (> 1000) comparison points (Figure 6.9), the maximum errors for ground points have negative values (Figures 6.9(a) and 6.9(b)) and for points on MMO positive (Figures 6.9(c)- 6.9(f)). Moreover, the good performance of AIM on MMO is also shown by the smaller mean values compared to SS and the distribution of height differences. E.g.



(a) (b) (c) (d) (e) (f)

Figure 6.9: Frequency distribution of the height differences in m (horizontal axis) for different point classes in test series B. The graphs that correspond to the results with SS and AIM are illustrated on the left and right (yellow and green color) respectively. In figures (a) and (b) the distribution of height differences for the point class BE{Pts, Brklin} in area W1052B. In (c), (d) and (e) and (f) are results from class MMO{Brklin} in areas W1052B and W1052C respectively.

in Figure 6.9(c), many height differences exist in the ranges -1.0 to -0.4 m and 0.5 to 1.2 m, whereas in Figure 6.9(d) the histogram is more steep. This observation can be also an evidence that SS leads to smoothing of discontinuities, a point that has been pointed out in different studies (e.g. Baltsavias et al., 1995). The fact that the accuracy was less for BE than MMO in W1052B is due to the density of the buildings and the relatively short distances of the measured points from neighboring MMO. BE breaklines in these areas have been extracted along discontinuities on the ground, which were close to MMO. Consequently, it is more difficult to discriminate measurement from modelling errors. More details from the analysis of the Yokohama dataset are discussed in Appendix A.1.3.

Chapter 7

Conclusions and Outlook

In this dissertation, methods and algorithms for automated image matching have been presented, with the focus being upon the exploitation of the radiometric and geometric characteristics of the airborne linear array CCD sensor ADS40. A summary of the research work is now presented (section 7.1), conclusions are drawn (section 7.2) and, finally, suggestions for further research are given (section 7.3).

7.1 Summary

The image matching system developed in this research consists of several processing modules and methods that are embedded in the overall matching strategy of AIM. One could distinguish two major topics in our research. The first relates to the processing prior to image matching, with respect to the methods for radiometric analysis of the data and the enhancement of image quality. The second refers to the individual aspects of matching, namely the geometric constraints, approximate values generation, feature extraction, quality control, combination of different matching algorithms and the matching strategy. Although these two issues can be examined individually, both are equally considered and integrated into the matching procedure, as matching itself can be adversely influenced if the first part is omitted, considering that more features can be extracted in areas in which texture has been enhanced and the images are radiometrically balanced. The radiometric analysis and preprocessing include:

- analysis of the distribution of grey levels
- noise estimation in homogeneous and non-homogeneous areas prior to and after preprocessing
- identification of artefacts
- simultaneous noise reduction and edge sharpening by intelligent filters
- Wallis filtering for contrast enhancement and radiometric balancing among channels
- non-linear methods reduction to 8-bit by (a) equal frequencies and (b) Gaussian distribution of frequencies based on the input slope of the Gaussian histogram
- weighting scheme for the merging of color bands

The individual matching aspects that have been analyzed are:

- the exploitation of geometrical constraints by *quasi* epipolar curves; geometrical constraints for Lev0 images are enforced either by the iterative approach or by the polynomial model approach, and for Lev1 images *quasi* epipolar lines are approximated by straight lines.
- the derivation of approximate values by hierarchical methods, namely (a) the generation of image pyramids, (b) the concept of doublets and (c) the height interpolation by an inverse distance weighting, Gaussian or smoothed weighting scheme.
- the evaluation and comparison of different operators, in both synthetic and real image data, for the selection of optimal features for image matching, including (a) Förstner, Harris and SUSAN operators for point extraction and (b) Förstner, SUSAN, Canny and gradient thresholding for edge extraction.
- hybrid matching by combining area and feature based methods used in (a) multi-patch size matching and (b) LSM.
- extensions of LSM to edge matching to improve surface discontinuity modelling via (a) edgel matching through patch adaptation and (b) edge matching through a modified weighting scheme.
- the selection of the general matching strategy, namely (a) a single- versus multi-template strategy, (b) a combination of different primitives, (c) continuity and height constraints for edge matching.
- a scheme for quality control comprising (a) definition of error types, (b) quality criteria and thresholds, and (c) combination of quality criteria according to possible occurring error.

The various components of AIM have been analyzed and evaluated individually but also on an integrated basis. The results from automatic matching have been compared to manual measurements in order to evaluate the performance of the newly developed AIM system with respect to:

- different terrain cover
- selection of CCD lines (two, three or four)
- selection of matching strategy (single- vs. multiple-templates, selection of features)
- different point type classes (bare earth (BE) and man made objects (MMO))
- measurement and modelling errors

The research conducted has concentrated on the analysis of the matching accuracy of the algorithms involved and not on the accuracy of the whole system. For this analysis, we used manual and automatic measurements from the same dataset (same images and orientation). Other possible errors such as multitemporal differences in the images used for automatic and manual measurements, sensor model differences, and coordinate system differences, have therefore not had an impact in the matching evaluations conducted. Moreover, our interest was to compare the results derived from AIM and the commercial system SS. However, the analysis of the individual matching components and of the complete system was a difficult task, due to several problems in the acquisition of datasets, which were not free of sensor model problems, calibration errors, orientation errors and artefacts, as the ADS40 camera system was under constant development and fine-tuning. Nevertheless, the Waldkirch dataset from 2002 was of relative good quality and a part of it was thoroughly analyzed and processed.

7.2 Conclusions

- **Radiometric analysis**

  The radiometric quality of ADS40 imagery has been analyzed, apart from the visual inspection and identification of artefacts, based on noise characteristics and grey value distribution. For different ADS40 datasets the following has been verified: Although the dynamic range reaches in general 14 bits (panchromatic images), the number of grey values having a substantial frequency is less than 2^{14}, i.e. $2^{12} - 2^{13}$ for the panchromatic and $2^{10} - 2^{11}$ for the multispectral. Images were found to exhibit relatively low noise (about 2 grey values). The number of artefacts related to software problems was reduced in the later datasets. Quantification of the noise level prior to and after preprocessing showed that the noise estimation method and the adaptive filters show a good performance as noise is reduced of about a factor of three and can be used for a broader range of image types. This applies also for the non-linear reduction to 8-bit methods; especially the Gaussian type is preferred for better visual interpretation. Contrast enhancement via Wallis filtering can be employed using common parameters for all images in order to radiometrically adjust them to each other. However, problems are observed in relatively large homogeneous areas, which after the processing with the Wallis filter exhibit high texture, with a number of extracted features being redundant. This would require Wallis to enforce different contrast over different area types in the images. With respect to the processing of the separate channels of RGB images, a color shift may appear in the final composite image, which can be reduced to a certain extent by computing the transformation parameters from only one reference band. For optimal results the image is transformed from RGB to HSI space, with the intensity component from the latter being processed. Then the modified HSI image is transformed back to RGB space.

- **Geometrical constraints**

  Since there is no direct access to the ADS40 sensor model parameters, epipolar constraints are employed by means of *quasi*-epipolar lines instead via the collinearity equations. Although the iterative approach retrieves the exact position on the epipolar trajectory, the polynomial model can be utilized for different images, for which there is no access to the sensor model parameters and in algorithms like LSM that require modelling of the geometrical constraints. For the Lev1 images and a maximum height interval of 100 meters, a first degree polynomial is used to gain a good approximation and the residuals from the true position are only 0.05 pixels. For the Lev0 images the polynomial is approximated from a smaller height interval (length of the epipolar trajectory of about 10 pixels). However, if the orientation parameters are not very accurate the weight factor of the constraints should be reduced (i.e. increase of standard deviation).

- **Image pyramids and doublets**

  For pyramid generation, it has been observed that the adaptive filters of adaptive edge preserving smoothing and the fuzzy filter distort the geometry of the lower resolution images in the upper levels of the image pyramid. Therefore, the 3x3 Gaussian filter is utilized with a reduction step of 2. The Wallis filter is applied in each image pyramid level by reducing the window size of the filter towards the upper levels in the pyramid (decreasing resolution). By applying Wallis filtering only to the full resolution level prior to pyramid generation, loss of texture in the upper levels can be encountered. The maximum number of pyramid levels used in AIM was determined by measuring parallax differences in different datasets and the conclusion was drawn that minimum five pyramid levels should be used. However, in densely built up areas with high buildings more than five, usually six should be used. Doublets are used in the algorithm to reduce processing time and number of interpolations between consecutive pyramid levels. By directly linking the two levels of the doublet, through projection of the selected points from the lower to the upper doublet level and performing matching from top to bottom, propagation of mismatches is reduced (less number of interpolation and direct linking of points between the doublet levels). For an image pyramid with six levels, only two instead of five interpolations are performed. The adjustment of the interpolation radius at run time ensures that a sufficient number of points is used in height interpolation. Height values of points close to height discontinuities might be partly incorrect, since breaklines are not treated.

- **Feature selection**

  During the project, we have analyzed several point and edge operators and further modified and evaluated them. A higher density of features is realized with the use of edges compared to points, plus representation of discontinuities is favored. Edgels are extracted in the template image. In the lower doublet level, the Canny operator is used for one-pixel wide edges and in the upper levels the gradient thresholding is applied. Aggregation of edgels is performed only in the lower doublet level and the edges are projected one level up (upper doublet level). In addition, grid points are selected in all levels with a predefined spacing to fill possible gaps between extracted edges and provide better surface continuity.

- **Multi-patch size matching**

  Multi-patch size matching has been used to improve approximations prior to LSM. The NCC is used as a similarity measure, as it is invariant to linear transformations (shift or scaling of intensity) in contrast to SAD and SSD. The behavior of the NCC in the three passes of matching (different mask sizes) is analyzed based on predetermined criteria and the best pass is selected. Moreover, empirically derived criteria are used to indicate possible multiple solutions or occlusions. Both multi-patch size matching and LSM methods are used in all levels.

- **LSM and extensions to edge features**

  The performance of LSM can be improved to a certain extent using the initial approximations for position and the smallest mask derived from the multi-patch size matching. Moreover, the geometrical constraints are used as additional observations to strengthen matching in terms of precision and reliability. For edge features, the LSM method extended to edgels and edges is utilized and it has been proved quite robust in cases of edge discontinuities. For single edge points convergence improves (less iterations and less oscillations for scales and shears) when using thin rectangular patches and initial approximations for rotation derived from the signal analysis are used. Edge LSM is applied to straight edges of length longer than 11 pixels and has proved to be the most effective method, as it can lead to convergence after 3-5 iterations even for features parallel to epipolar lines. Also it requires less accurate approximations, plus it can model all the points along the edge after one least squares solution. Although this method is quite robust, the modelling of the contour points depends on both the accuracy of the extracted edge and the transformation parameters, and on the change of parallax along the edge. Occlusions may still pose a problem.

- **Quality control and error detection**
  A set of error types is defined and different quality criteria are combined according to the possible occurring errors. For single point matching, the quality control is performed in two stages with the quality measures derived from (a) multi-patch size matching and (b) LSM. If a point corresponds to a single edge point, the edgel LSM approach (patch adaptation) is used in the second stage of the quality control. If a point corresponds to a grid point, the standard multi-image LSM is applied without patch adaptation. Each image ray is checked individually and is rejected or accepted. The NCCs from multi-patch size matching and the residuals from forward intersection are the most discriminating criteria. The angle to the epipolar line and the results from backmatching are used to indicate multiple solutions. Two sets of thresholds exist: relative (computed at run time) and absolute, a number of the latter is derived empirically. Multi-image LSM is employed only for the "good"points and with stricter criteria. The approach sequentially checks and eliminates problematic points. The same scheme is applied for edge features (aggregated edgels), for which vertices are matched first and remaining contour points are matched by enforcing the continuity and height constraints. For straight edge features, for which the Edge LSM method is used, the quality of the matched edge is checked only by the quality criteria of LSM (no multi-patch size matching). From the evaluation of the results it has been observed that errors are significantly reduced, however some criteria are quite strict and good points may be also excluded.

- **General strategy**
  Multi-template matching is favored since one template may not suffice to detect and avoid problems, especially occlusions. The single template strategy suffices in rural areas with relatively flat terrain. In areas with more complex objects, the multi-template approach should be used and features have to be extracted in both templates, leading to an increase in memory allocation. The characterization of the type of the area and selection of the matching strategy is done by the user. Based on the studies performed and on the tests conducted, the following scheme is recommended: Use only grid points in the upper pyramid levels to generate a coarse surface, i.e. if six multi-resolution levels are used, then use only grid points in the upper two pyramid levels (complex objects) or the upper four pyramid levels (relatively flat terrain). Densify the surface in the lower levels by edge features. Long straight edges ($>$ 11 pixels) are matched with Edge LSM. For the remaining edges the vertices are matched first and then the contour points are matched using the height and continuity constraint. The alternative scheme of performing edge matching in the upper levels, followed by refinement of the surface in the lower levels by grid points was disregarded. The surface was partially described in the upper levels and interpolation errors in the lower levels increased.

In this study, we have analyzed existing methods of image matching and developed both algorithmic improvements and new approaches to the image correspondence problem. The whole matching strategy includes the combination and mutual support of different algorithms that enhance the adaptability of matching. Although the current investigations were directed towards exploiting special ADS40 characteristics in the matching, it has been shown that different modules of AIM can also be used for a broader range of image data, namely imagery from linear and frame CCD sensors, scanned frame images, and close-range, aerial and satellite imagery. The potential of AIM in close-range imagery has been already indicated in Pateraki et al. (2002). With respect to commercial systems and the ADS40, the improved matching performance of AIM compared to the commercial system SS has been confirmed and was close to the theoretically expected accuracy of the manual measurements (20-40 cm). The accuracy and reliability of automatic measurements, especially in rural areas and on man-made objects, has been increased compared to SS, while also leading to a significant reduction in the number of gross errors (maximum absolute values and distribution of height differences). The accuracy of AIM, especially on MMO, is close to 60 cm, whereas for SS 100-120 cm. Trees have been excluded from the analysis as the manual and automatic measurements of these were not reliable. For dense urban areas, denser matching and the use of images from neighboring strips is required. Moreover, in order to improve object surface modelling, the matching results should be further post-processed by combining information from the DSM, the images and extracted breaklines.

7.3 Outlook

The advantages of using airborne linear CCD sensors such as the ADS40 for automatic DSM extraction have already been pointed out. From the matching point of view, it is believed that although a detailed analysis on several matching aspects related to these sensors has been carried out, there is still scope for further research in this field. Other important issues worthy of additional investigations include:

- Improvements to object surface modelling to reduce modelling errors, in areas where abrupt changes in elevation occur (e.g. close to building outlines). To automatically detect these areas and include additional points, the derived DSM data, the intensity and the breakline information, extracted from the images, will need to be combined.
- Refinements to the reduction of DSM to DTM, namely detection and exclusion of non-DTM objects based on geometric criteria (e.g. area, shape, slope and height of the object) and additional criteria like shadow information (useful especially for buildings), texture and multispectral information.
- Exploitation of the radiometric characteristics of the near infrared channel, for automatic detection and exclusion of forest areas or single tress. This information can be either used prior to (area masks) or after (point exclusion) DSM generation.
- Inclusion of images from additional cross strips and enforcement of additional geometric constraints. Epipolar constraints in approximately perpendicular directions will disambiguate problems occurring for features parallel to epipolar lines. Although the multiple solution problem for these features has been partly solved by the Edge LSM method presented in this thesis, this option will further strengthen matching. The strategy should be further investigated, namely by performing matching

(a) for all images simultaneously or (b) by first using images of one strip and then using approximations for pairwise matching of common channels between different strips.

- Implementation of collinearity equations as geometric constraints in LSM. This approach is direct, as it is not required to approximate the epipolar curve over a definite height range and the 3D coordinates are simultaneously computed. However, this will require re-implementation of the sensor model.
- Improvements to the Edge LSM method for connected, neighboring edges. The 3D coordinates of the linking nodes should be determined by evaluating the matching results of all connected edges and enforcing continuity constraints.
- Improvements with respect to the interpolation method utilized to transfer the matching results to the lower pyramid level. A different interpolation method should be used for breaklines and for single points to avoid smoothing of discontinuities in the lower levels.
- More datasets of both good radiometric and geometric quality are required and consequently more tests have to be conducted to evaluate the algorithm in mountainous areas and in densely built areas. The acquisition of reference data with better accuracy (3-10 times) than the typical accuracy that could be derived from ADS40 images should be considered in future research.

Apart from the above, the most critical issue that influences several implementation aspects, time performance and accuracy of AIM relates to the sensor model. It has been observed during this project that the current concept and implementation of the ADS40 sensor model is not flexible, is relative complex and cannot be used in a straightforward way by third party users and cannot be integrated in stand alone applications. Worse, the sensor model software implementation has become subverted by a set of requirements linked to implementation inheritance. The sensor model should be based on more simple, efficient and less time consuming approaches with respect to the image-to-ground and the ground-to-image transformations. In the more general view of algorithmic developments, a point that is usually underestimated is related to implementation efforts. The latter should lead to a well-crafted software implementation with less risk of degenerating into 'spaghetti code'.

Bibliography

Abbreviations:

ACRS	Asian Association on Remote Sensing
ASPRS	American Society of Photogrammetry and Remote Sensing
CVPR	Conference on Computer Vision and Pattern Recognition
IAPRS	Interantional Archives of Photogrammetry and Remote Sensing
IEEE	The Institute of Electrical and Electronics Engineers
INRIA	Institut National de Recherche en Informatique et en Automatique
ISPRS	Society of Photogrammetry and Remote Sensing
OEEPE	European Organization for Experimental Photogrammetric Research
PAMI	Transactions on Pattern Analysis and Maschine Intelligence
PERS	Photogrammetric Engineering and Remote Sensing
PFG	Photogrammetrie, Fernerkundung, Geoinformation
SMC	Transactions on Systems, Man and Cybernetics
SPIE	The Society of Photo-Optical Instrumentation Engineers

Ackermann F., 1984. Digital image correlation: Performance and potential application in photogrammetry. Photogrammetric Record 11 (9), 429–439.

Ackermann F., Hahn M., 1991. Image Pyramids in Digital Photogrammetry. In: H. Ebner, D. Fritsch, C. Heipke (Eds.), Digital Photogrammetric Systems, Wichmann, Heidelberg, pp. 43-59.

ADS Inc., 2004. Airborne Data Systems homepage - remote sensing, multispectral digital imaging technology. http://www.airbornedatasystems.com (accessed December 2004).

AEROCam, 2004. The Airborne Enviromental Research Observation Camera. http://www.umac.org/content/sensors/aerocam.shtml (accessed December 2004).

AirCam, 2004. AirCam technical specifications - Kerstel Corp. homepage. http://www.kestrelcorp.com/aircam.html (accessed December 2004).

Albertz J., 1996. The HRSC/WAOSS camera experiment on the Mars96 mission - a photogrammetric and cartographic view of the project. In: IAPRS, Vol. 31, Part B4, pp. 58–63.

AMDC, 2004. Description of AA497 Airborne Multispectral Digital Camera in the Argon ST homepage. http://www.argonst.com/docs/AMDC.pdf (accessed January 2005).

Asada H., Brady M., 1986. The curvature primal sketch. IEEE PAMI Pattern Analysis and Machine Intelligence 8 (1), 2–14.

Atick J.J., Redlich A.N., 1990. Towards a Theory of Early Visual Processing. Neural Computation 2, 308–320.

Bacher U., 1998. Experimental studies into automated DTM generation on the DPW770. In: IAPRS, Vol. 32, Part 4, pp. 35–41.

Baillard C., Dissard O., Jamet O., Maître H., 1998. Extraction and textural characterization of above ground areas from aerial stereo pairs: a quality assessment. ISPRS Journal of Photogrammetry and Remote Sensing 53 (2), 130–141.

Baker S., Nayar S.K., 1999. Global measures of coherence for edge detector evaluation. In: Proc. CVPR, 23-25 June, Ft. Collins, CO, USA, pp. 373–379.

Ballard D.H., Brown C.M., 1982. Computer Vision. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Baltsavias E.P., 1991. Multiphoto geometrically constrained matching. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Mitteilungen 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.

Baltsavias E.P., Mason S.O., Stallmann D., 1995. Use of DTMs/DSMs and orthoimages to support building extraction. In: A. Grün, O. Kübler, P. Agouris (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, Basel, pp. 199–210.

Baltsavias E.P., Käser C., 1998. DTM and orthoimage generation - a thorough analysis and comparison of four digital photogrammetric systems. In: IAPRS, Vol. 32, Part 4, pp. 42–51.

Baltsavias E.P., Favey E., Bauder A., Bösch H., Pateraki M., 2001a. Digital surface modelling by airborne laser scanning and digital photogrammetry for glacier monitoring. Photogrammetric Record 17 (98), 243–273.

Baltsavias E.P., Pateraki M., Zhang L., 2001b. Radiometric and geometric evaluation of IKONOS Geo images and their use for 3D building modelling. In: Proc. Joint ISPRS Workshop "High Resolution Mapping from Space 2001", 19–21 September, Hannover, Germany (on CD-ROM).

Beisl U., Woodhouse N., 2004. Correction of atmospheric and bidirectional effects in multispectral ADS40 images for mapping purposes. In: IAPRS, Vol. 35, Part 7, pp. 1–5 (and on DVD).

Belli T., Cord M., Jordan M., 2001. 3D data reconstruction and modeling for urban scene analysis. In: E. Baltsavias, A. Grün, L. Van Gool (Eds.), Automatic extraction of Man-Made objects from Aerial and Space images (III). A.A. Balkema Publishers, Lisse, pp. 125–134.

Beyer H., 1992. Geometric and Radiometric Analysis of a CCD-Camera Based Photogrammtric Close-Range System. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Mitteilungen No. 51, 186 p.

Black M., Rangarajan A., 1996. On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision. International Journal of Computer Vision 19 (1), 57–91.

Blanc N., 2001. CCD versus CMOS - has CCD imaging come to an end?. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '01. Wichmann, Heidelberg, pp. 131 – 137.

Börner A., Reulke R., 2001. Results of Test Flights witt the Airborne Digital Sensor ADS40. In: R. Klette, S. Peleg, G. Sommer (Eds.), Robot Vision, Lecture Notes in Computer Science, Vol. 1998. Springer, Berlin, pp. 270–277.

Bösemann W., 1994. Geometric models in Object Based Multi-Image Matching. In: IAPRS, Vol. 30, Part 3, pp. 61–68.

Boyle W.S., Smith G.E., 1970. Charge-coupled semiconductor devices. Bell Systems Technical Journal 49, 587–593.

Brand P., Mohr R., 1994. Accuracy in image measure. In: Proc. SPIE, "Videometrics III", Vol. 2350, pp. 218–228.

Brown D.C., 1971. Analytical calibration of close-range cameras. PERS 37 (8), 855–866.

Brown M., Szeliski R., Winder S., 2004. Multi-Image Matching Using Multi-Scale Oriented Patches. Microsoft Research Technical Report MSR-TR-2004-133. http://www.research.microsoft.com/ research/pubs/view.aspx (accessed December 2004)

Brunn A., Guelch E., Lang F., Foerstner W., 1998. A hybrid concept for 3D building acquisition. ISPRS Journal of Photogrammetry and Remote Sensing 53 (2), 119–129.

Calitz M.F., Ruether H., 1996. Least absolute deviation (LAD) image matching. ISPRS Journal of Photogrammetry and Remote Sensing 51 (5), 223–229.

Canny J., 1983. Finding edges and lines in images. MSc Thesis, Artificial Intelligence Lab. Massachusetts Institute of Technology, Cambridge, USA.

Canny J., 1986. A computational approach to edge detection. IEEE PAMI 8 (6), 679–698.

Chambon S., Crouzil A., 2003. Dense matching using correlation: new measures that are robust near occlusions. In: R. Harvey, A. Bangham (Eds.), British Maschine Vision Conference 2003, Norwich, UK (on CD-ROM).

Chen T., Shibasaki R., Murai S., 2003. Development and Calibration of the Airborne Three-Line Scanner (TLS) Imaging System. PERS 69 (1), 71–78.

Collins R., 1996. A space-sweep approach to true multi-image matching. In: Proc. Computer Vision and Pattern Recognition Conference, pp. 358–363.

Cramer M., 2004. EUROSDR Network on Digital Camera Calibration. In: IAPRS and SIS, Vol. 35, Part B6, pp. 204–209.

DAS-1, 2004. Digital Aerial Camera DAS-1. Product homepage. http://www.wehrliassoc.com (accessed December 2004).

D'Apuzzo N., 2002. Surface measurement and tracking of human body parts from multi-image video sequences. ISPRS Journal of Photogrammetry and Remore Sensing 56 (5-6), 360–375.

Demigny D., Kamlé T., 1997. A discrete expression of Canny's criteria for step edge detector performances evaluation. IEEE PAMI 19 (11), 1199–1211.

Deriche R., Giraudon G., 1991. Accurate corner detection: An analytical study. Rapport de recherche no. 1420, Unité de recherche INRIA, Sophia Antipolis, France.

Deriche R., Blaszka T., 1993. Recovering and characterizing image features using an efficient model based approach. In: Proc. CVPR, 15-17 June, New York, NY, USA, pp. 530–535.

Dhond U.R., Aggarwal J.K., 1989. Structure from stereo - a review. IEEE SMC 19(6), 1489–1510.

Diefendorff K., 1998. CMOS Image Sensors Challenge CCDs. Microprocessor Report, MicroDesign Resources, 22 June 1998.

Diehl H., Heipke C., 1992. Surface reconstruction from data of digital line cameras by means of object based image matching. In: IAPRS, Vol. 29, Part B3, pp. 287–294.

Dileepan J., 2002. Modelling and calibration of logarithmic CMOS image sensors. Ph.D. Thesis, Department of Engineering Science, University of Oxford, 188 p.

DiMAC, 2004. Digital Modular Aerial Camera - Product homepage. http://www.dimac-camera.com (accessed December 2004).

DSS, 2004. Digital Sensor System - Product outline, Applanix homepage. http://www.applanix.com/products/dss_index.php (accessed December 2004).

Ebner H., Heipke C., 1988. Integration of digital image analysis and object space reconstruction. In: IAPRS, Vol 27, Part B3, pp. 534–545.

Eckardt A., 2002. Design und Verifikation der ersten digitalen kommerziellen Luftbildkamera. Ph.D. Thesis, Fakultät Verkehrs- und Maschinensysteme, Technische Universität Berlin, 105 p.

Elaksher A., Bethel J., 2002. Object surface reconstruction using highly redundant imagery. In: Proc. ASPRS Annual Conference, Washington DC, USA (on CD-ROM).

Elberink S.O., Vosselman G., 2002. 3D Line reconstruction from Airborne Three-Line Scanner Data. In: IAPRS, Vol. 34, Part B3/A, pp. 108–113.

Faugeras O., 1993. Three-Dimensional Computer Vision - A Geometric Viewpoint. M.I.T. Press, Cambridge, MA, USA.

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

Förstner W., 1984. Quality Assessment of Object Location and Point Transfer Using Digital Image Correlation Techniques. In: IAPRS, Vol. 25, Part B3, pp. 197–219.

Förstner W., 1994. A Framework for Low Level Extraction. In: J.O. Eklundh (Ed.), Computer Vision, Lecture Notes in Computer Science, Vol. 802. Springer, Berlin, pp. 383–394.

Förstner W., 1995. Matching strategies for Point Transfer. In: D. Fritsch, D. Hobbie (Eds.), Photogrammetric Week '95. Wichmann Verlag, Heidelberg, pp. 173–183.

Förstner W., 1998. Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images. In: A. Dermanis, A. Grün, F. Sansò (Eds.), Geomatic Methods for the Analysis of Data in the Earth Sciences, Lecture Notes in Earth Sciences, Vol. 95. Springer, Berlin, pp. 165–189.

Förstner W., Gülch E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In: Proc. Intercommission Conference on Fast Processing of Photogrammetric Data, 2-4 June, Interlaken, Switzerland, pp. 281–305.

Ford A., Roberts A., 1998. Colour space conversions. Technical Report, Westminster University, London, 31 p.

Fraser C.S., Baltsavias E., Grün A., 2001. 3D building reconstruction from high-resolution Ikonos stereo imagery. In: E.P. Baltsavias, A. Grün, L. Van Gool (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (III). A.A. Balkema Publishers, Lisse, pp. 331–344.

Fricker P., 2001. ADS40 - Progress in digital aerial data collection. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '01, Wichmann, Heidelberg, pp. 105 – 116.

Fricker P., Sandau R., Walker A.S, 2000. Development of an airborne digital sensor for photogrammetric and remote sensing applications. In: Proc. ASPRS Annual Conference, 22-26 May, Washington DC, USA (on CD-ROM).

Fritsch D., 1997. Experiences with the airborne three-line photogrammetric image acquisition system DPA. In: D. Fritsch, D. Hobbie. (Eds.), Photogrammetric Week '97. Wichmann, Heidelberg, pp. 63–74.

Fua P., Leclerc Y.G., 1990. Model driven edge detection. Machine Vision and Applications 3, 45–56.

Fua P., Leclerc Y.G., 1995. Object-centered Surface Reconstruction: Combining Multi-Image Stereo and Shading. International Journal of Computer Vision 16 (1), 35-56.

Fua P., Grün A., Li H., 1998. Optimization-based approaches to feature extraction from aerial images. In: A. Dermanis, A. Grün, F. Sansó (Eds.), Geomatic Methods for the Analysis of Data in the Earth Sciences, Lecture Notes in Earth Sciences, Vol. 95. Springer, Berlin, pp. 191–228.

Gabet L., Giradaun G., Renouard L., 1997. Automatic elevation of high resolution urban zone digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing 52(1), 33–47.

Gamble E., Poggio T., 1987. Visual integration and detection of discontinuities: the key role of intensity edges. Tech. Rep., A.I. Memo No. 970, Artifial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, USA.

Giles J., 2001. Think like a bee. Nature 410, 510–512.

Gilson S.J., Damper R.I., 1997. An empirical comparison of neural techniques for edge linking of images. Neural Computing & Applications 6 (2), 64–78.

Gonzalez R.C., Woods R.E., 2002. Digital Image Processing. Addison-Wesley, Second Edition, 2002.

Gorin B., Mathews B., Threadgould A., 2002. An Ultra-High Resolution Electro Optical Framing Camera with a 9216 by 9216 Focal Plane Array and Differential Image Motion Compensation. In: Proc. ASPRS Annual Conference, 19-26 April, Washington DC, USA (on CD-ROM).

Graham R., Koh A., 2001. Remote Sensing: Training for digital aerial survey and mapping. Geoeurope 11. http://www.geoplace.com/ge/2001/0111/0111trn.asp (accessed December 2004).

Grün A., 1985. Adaptive least squares correlation: A powerfull image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography 14 (3), 175–187.

Grün A., Stallmann D., 1991. High accuracy edge matching with an extension of the MPGC matching algorithm. In: Proc. SPIE, "Industrial Vision Metrology", Vol. 1526, pp. 42–45.

Grün A., Stallmann D., 1992. High accuracy matching of object edges. In: Proc. SPIE, "Videometrics", Vol. 1820, pp. 70–82.

Grün A., Agouris P., 1994. Linear feature extraction by least squares tempate matching constrained by internal shape forces. In: IAPRS, Vol. 30, Part 3, pp. 316–323.

Grün A., Zhang L., 2002. Automatic DTM Generation from Three-Line-Scanner (TLS) Images. In: IAPRS, Vol. 34, Part 3A, pp. 131–137.

Grün A., Zhang L., 2003. Sensor Modeling for Aerial Triangulation with Three-Line-Scanner. PFG 2003 (2), 85–98.

Grün A., Zhang L., Wang X., 2003. 3D city modeling with TLS (Three Line Scanner) Data. In: IAPRS, Vol. 34, Part 5/W6 (on CD-ROM).

Gwinner K., Hauber E., Hoffmann H., Scholten F., Jaumann R., Neukum G., Coltelli M., Puglisi G., 1999. The HRSC-A Experiment on High Resolution Imaging and DEM Generation at the Aeolian islands. In: Proc. 13th International Conference Applied Geologic Remote Sensing, Vol. I, 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: IAPRS, Vol. 33, Part B4/1, pp. 324–331.

Hampel F.R., 1973. Robust estimations: a condensed partial survey. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 27, 87–104.

Hannah M. J., 1988. Digital stereo image matching techniques. In: IAPRS, Vol. 27, Part B3, pp. 280–293.

Haralick R.M., Shapiro L.G., 1992. Computer and Robot Vision. Vol. 1. Addison-Wesley, Reading, MA, USA.

Harris C., Stephens M., 1988. A combined corner and edge detector. In: Proc. 4th Alvey Vision Conference, 31 August-2 September, Manchester, England. pp. 147–151.

Hartley R., 1994. Lines and Points in Three Views: an Integrated Approach. In: Proc. ARPA Image Understanding Workshop, 13-16 November, Monterey, CA, USA, pp. 1009–1016.

Hauber E., Oberst J., Flohrer J., Sebastian I., Zhang W., Robinson C., Jaumann R., Neukum G., 1996. The High Resolution Stereo Camera (HRSC) for Mars 96: Results from outdoor tests. In: IAPRS, Vol. 31, Part B4, pp. 349–457.

Heath M.D., Sarkar S., Sanocki T., Bowyer K.W., 1997. A robust visual method for assessing the relative performance of edge-detection algorithms. IEEE PAMI 19 (12), 1338–1359.

Heier H., Hinz A., 2002. Results from the digital modular camera DMC. In: Proc. ASPRS Annual Conference, 19-26 April, Washington DC, USA (on CD-ROM).

Heipke C., 1996. Overview of image matching techniques. In: Proc. OEEPE Workshop on Application of Digital Photogrammetric Workstations, 4-6 March, Lausanne, Switzerland, OEEPE Off. Publ. No. 33, pp. 173–189.

Heitger F., Rosenthaler L., Heydt R., Peterhans E., Kuebler O., 1992. Simulation of neural contour mechanism: from simple to end-stopped cells. Vision Research 32 (5), 963–981.

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

HeliMap, 2004. HeliMap System homepage. http://www.helimap.ch/index_en.html (accessed December 2004).

Henricsson O., 1996. Analysis of image structures using color attributes and similarity relations. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Mitteilungen No. 59, 124 p.

Henricsson O., Baltsavias E., 1997. 3-D building reconstruction with ARUBA: a qualitative and quantitative evaluation. In: A. Grün, E.P. Baltsavias, O. Henricsson (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser, Basel, pp. 65–76.

Hese S., Hoffmann A., Lehmann F., 1999. Vergleich digitaler Waldoberflächenmodelle der HRSC-A (High Resolution Stereo Camera Airborne) mit Laserscanner-Daten zur Bestimmung von bestandesspezifischen Parametern. In: Photogrammetrie und Fernerkundung - Neue Sensoren, neue Anwendungen, 19. Jahrestagung der Deutsche Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF), Band 8, Essen.

Hese S., Lehmann F., 2000. Comparison of digital surface models of HRSC-A and laser scanner for forest stand characteristics. In: IAPRS, Vol. 33, Part B3 (on CD-ROM).

Hildreth E., 1985. Edge detection. Tech. Rep., A.I. Memo No. 858, Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, USA.

Hinsken L., Miller S., Tempelmann U., Uebbing R., Walker S., 2002. Triangulation of LH Systems ADS40 imagery using ORIMA GPS/IMU. In: IAPRS, Vol. 34, Part 3A, pp. 156–162.

Hinz A., 1999. The Z/I Imaging Digital Aerial Camera System. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '99. Wichmann, Heidelberg, pp. 109–115.

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: IAPRS, Vol. 33, Part B2, pp. 295–302.

Hofmann O., Navé P., Ebner H., 1982. DPS A Digital Photogrammetric System for Producing Digital Elevation Models and Orthophotos by Means of Linear Array Scanner Imagery. In: IAPRS, Vol. 24, Part B3, pp. 216–227.

Hofmann O., Kaltenecker A., Müller F., 1993. Das flugzeuggestützte digitale Dreizeilenaufnahme- und Auswertesystem DPA - erste Erprobungsberichte. In: D. Fritsch, D. Hobbie (Eds.), Photogrametric Week '93. Wichmann, Heidelberg, pp. 97–107.

Holm M., 1994. Global Image Matching and surface reconstruction in object space using three aerial images. In: IAPRS, Vol. 30, Part 3/1, pp. 379–386.

Horn B., Schunk B., 1980. Determing optical flow. Tech. Rep., A.I. Memo No. 572, Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, USA.

Hsia J.-S., 2001. A research on the method of simultaneity least squares Multi-Point matching for Producing Digital Terrain Models. In: Proc. ACRS 2001, 5-9 November, Singapore, pp. 1073–1078.

Huber P.J., 1981. Robust Statistics. Willey, New York, USA.

Hulst N., Gullicks J., Johnson J., Lauinger G., Larson D., Lemcke S., 2002. The airborne enviromental research observational camera (Aerocam): A multispectral digital photography system for remote sensing. In: Proc. National Conference of Undergraduate Research, 25-27 April, Whitewater, WI, USA (on Internet).

ISAACS, 2004. Integrated spectronics airborne acquisition camera system homepage. http://www.intspec.com (accessed December 2004).

Iverson L.A., Zucker S.W., 1995. Logical/Linear Operators for Image Curves. IEEE PAMI 17 (10), 982–996.

Jacobsen K., 1998. Status and tendency of sensors for mapping. In: IAPRS, Vol. 32, Part 1, pp. 124–130.

Jenkin M.R.M., Jepson A.D., 1994. Recovering local surface structure through local phase difference methods. Computer Vision, Graphics and Image Processing: Graphical Model and Image Understanding 59 (1), 72–93.

Julien P., 1999. Principles of digital matching. In: OEEPE Workshop on Automation in Digital Photogrammetric Production, 4-6 March, Marne le Vallée, France, OEEPE Off. Publ. No. 37, pp. 161–172.

Kanade T., 1994. Development of a Video-Rate Stereo Machine. In: Proc. Arpa Image Understanding Workshop, 13-16 November, Monterey, CA, USA, pp.549–557.

Kim S., Kim T., Park W., Lee H.-K., 1999. An optimal interpolation scheme for producing a DEM from the automated stero matching of full scale SPOT images. In: Proc. Joint ISPRS Workshop "Sensors and Mapping from Space 1999", 27–30 September, Hannover, Germany (on CD-ROM).

Kitchen L., Rosenfeld A., 1982. Gray-level corner detection. Pattern Recognition Letters 1 (2), 95–102.

Kodak, 1999. KLI series color linear image sensors spectral response characteristics. Application Note DS 03-001, Eastman Kodak Company. http://www.kodac.com/global/ plugins/acrobat/en/digital/ccd/applicationNotes/linear/spectralResponseCharacteristics.pdf (accessed December 2004).

Kölbl O., de Laporte K., Gasoir D., Walker S., 1994. Multi-Templet-Matching, a sensitive Matching Algorithm. In: IAPRS, Vol. 30, Part 3/1, pp. 448–454.

Koh A., Edwards E., Curr R., Strawbridge F., 1996. Techniques in Electronic Imaging for Natoural Resource Monitoring. In: C.H. Power, L.J. Rosenberg, I. Downey (Eds.), Remote Sensing and GIS for Natural Resource Management, University of Greenwich, UK, pp. 47-54.

Kramer H.J., 2002. Observation of the Earth and its Environment - Survey of Missions and Sensors. Springer, Berlin.

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

Krzystek P., Wild D., 1992. Experimental accuracy analysis of automatically measured digital terrain models. In: W. Förstner, W. Ruwiedel (Eds.), Robust Computer Vision: Quality of Vision Algorithms. Wichmann, Heidelberg, pp. 372–389.

Kubik K., Merchant D., Schenk T., 1987. Robust estimation in photogrammetry. PERS 53 (2), 167–169.

Laganiére R., 1998. Morphological corner detection. In: Proc. International Conference on Computer Vision. 4-7 January, Bombay, India, pp. 280–285.

Lan Z.-D., Mohr R., 1995. Robust matching by partial correlation. Rapport de recherche no. 2643, INRIA, Rhône Alps. France.

Lang F., Förstner W., 1995. Matching Techniques. In: Notes of Second Course in Digital Photogrammetry, Bonn, Germany.

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, 5-7 May, Anchorage, AK, USA (on CD-ROM).

Lee J.-S., 1983. Digital image smoothing and the Sigma filter. Computer Vision, Graphics and Image Processing 24 (2), 255–269.

Lemmens M., 1988. A survey on stereo matching techniques. In: IAPRS, Vol. 27, Part B8, pp. 11–23.

Li D., Wang S., 1996. An automatic quality diagnosis method for digital image matching. In: IAPRS, Vol. 31, Part B3, pp. 498–504.

Li M., 1991. Hierarchical multipoint matching. PERS 57 (8), 1039–1047.

Litwiller D., 2001. CCD vs. CMOS: Facts and Fiction. Photonics Spectra, January 2001, 154–158.

Lo K., Mulder N., 1992. High precision DEM generation from SPOT stereo imagery by object least squares matching. In: IAPRS, Vol. 29, Part B3, pp. 133–138.

López A.M., Lumbreras F., Serrat J., Villanueva J.J., 1999. Evaluation of methods for ridge and valley detection. IEEE PAMI 21 (4), 327–335.

Luhmann T., Ehlers M., 1984. AIMS – A System for Automatic Image Matching. In: Proc. International ERIM- Symposium, 1-5 October, Paris, France, pp. 971–979.

Lutes J., 2002. DAIS : A Digital Airborne Imaging System. In: IAPRS, Vol. 34, Part 1, pp. 136–141.

Lyon R., Hubel P., 2002. Eyeing the Camera: into the Next Century. In: Proc. 10th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, 12 November Scottsdale, Arizona, USA, pp. 349–555. http://www.imaging.org/store/epub.cfm?abstrid=8450. (accessed December 2004)

Lyvers E.P., Mitchell O.R., Akey M.L., Reeves A.P., 1989. Subpixel measurements using a moment-based edge operator. IEEE PAMI 11 (12), 1293–1309.

Maas H.-G., 1996. Automatic DEM generation by multi-image feature based matching. In: IAPRS, Vol. 31, Part B3, pp. 484–489.

Magnan P., 2003. Detection of visible photons in CCD and CMOS: A comparative view. Nuclear Instruments and Methods in Physics Research A 504, 199–212.

Medioni G., Yasumoto Y., 1987. Corner detection and curve representation using cubic B-splines. Computer Vision, Graphics and Image Processing 39 (3), 267–278.

Mendenhall J., Lencioni D., Parker A., 1999. Radiometric calibration of the EO-1 Advanced Land Imager, In: Proc. SPIE. "Earth Observing Systems IV", Vol. 3750, pp. 117–131.

Miller S., De Venezia K., 1992. Automatic elevation extraction and the digital photogrammetric workstation. In: Proc. ASPRS Annual Conference, 3-8 March, Albuquerque, NM, USA, pp. 572–580.

Mokhtarian F., Suomela R., 1998. Robust image corner detection through curvature scale space. IEEE PAMI 20 (12), 1376–1381.

Moravec H.P., 1980. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Ph.D. Thesis, Report STAN-CS-80-813, Stanford University.

Müeller F., 1991. Photogrammetrische Punktbestimmung mit Bilddaten digitaler Dreizeilenkameras. Ph.D. Thesis, Deutsche Geodätische Kommission, München, Germany, Series C, No. 372.

Murai S., Matsumoto Y., Li X., 1995. Stereoscopic imaging with an airborne 3-line scanner (TLS). In: IAPRS, Vol. 30, Part 5W1, pp. 22–25.

Murai S., Murakami M., Tsuno K., Morita K., Chen T., 2003. Three-Line Scanner for Large Scale Mapping. In: A. Grün, H. Kahmen (Eds.), Proc. 6th Conference on Optical 3D Measurement Techniques, Vol. 1, published by IGP, ETHZ, Zurich, pp. 1–10.

Mustaffar M., Mitchell H.L., 2001. Improving area-based matching by using surface gradients in the pixel co-ordinate transformation. ISPRS Journal of Photogrammetry and Remote Sensing 56 (1), 42–52.

Nagao M., Matsuyama T., 1979. Edge Preserving Smoothing. Computer Graphics and Image Processing 9 (4), 394–407.

Nakagawa M., Shibasaki R., Kagawa Y., 2002. Fusing stereo linear CCD image and laser range data for building 3D urban model. In: IAPRS, Vol. 34, Part 4 (on CD-ROM).

Nalwa V., Binford T.O., 1986. On detecting edges. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6), 699–714.

Neukum G., 1999. The Airborne HRSC-A: Performance Results and Application Potential. In: D. Fritsch, D. Hobbie (Eds.), Photogrammetric Week '99. Wichmann, pp. 83–88.

Neukum G., HRSC-TEAM, 2001. The airborne HRSC-AX cameras: evaluation of the technical concept and presentation of application results after one year of operations. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '01. Wichmann, Heidelberg, pp. 117–131.

Nevatia R., 1996. Matching in 2-D and 3-D. In: IAPRS, Vol. 31, Part B3, pp. 567–574.

Nonin P., 2003. Automatic extraction of digital surface models from airborne digital cameras. In: A. Grün, H. Kahmen (Eds.), Proc. 6th Conference on Optical 3D Measurement Techniques, Vol. 1, published by IGP, ETHZ, Zurich, pp. 106–113.

Norvelle F.R., 1992. Stereo Correlation: Window Shaping and DEM Corrections. PERS 58 (1), 111–115.

Ohta Y., Kanade T., 1987. Stereo by Intra- and Inter-Scanline search Using Dynamic Programming. IEEE PAMI 7 (2), 139–154.

Okutomi M., Kanade T., 1993. A Multiple-Baseline Stereo. IEEE Pattern Analysis and Machine Intelligence 15 (4), 353–363.

Okutomi M., Katayama Y., Oka S., 2002. A simple stereo algorithm to recover precice object boundaries and smooth surfaces. International Journal of Computer Vision 47 (1-3), 261–273.

Olson C., 2002. Maximum-likelihood image matching. IEEE PAMI 24 (6), 853–857.

Pacey R.E., Scheidt M., Walker S., 1999. Calibration of analog and digital airborne sensors at LH systems. In: Proc. ASPRS Annual Conference, Portland, Oregon, USA (on CD-ROM).

Paparoditis N., Maillet G., Taillandier F., Jibrini H., Jung F., Guigues L., Boldo D., 2001. Multi- image 3D feature and DSM extraction for change detection and building reconstruction . In: E. Baltsavias, A. Grün, L. Van Gool (Eds.), Automatic extraction of Man-Made objects from Aerial and Space images (III). A.A. Balkema Publishers, Lisse, pp. 217–230.

Parida L., Geiger D., Hummel R., 1998. Junctions: Detection, classification and reconstruction. IEEE PAMI 20 (7), 687–698.

Pateraki M., Baltsavias E. P., Patias P., 2002. Image combination into large virtual images for fast 3D modelling of archaeological sites. In: IAPRS, Vol. 34, Part 5, pp. 440–445.

Pateraki M., Baltsavias E., 2003. Analysis and performance of the Adaptive Multi-Image matching algorithm for airborne digital sensor ADS40. In: Proc. ASPRS Annual Conference, Anchorage (AK), USA (on CD-ROM).

Pateraki M., Baltsavias E., 2004. Surface discontinuity modelling by LSM through patch adaptation and use of edges. In: IAPRS and SIS, Vol. 35, Part B3, pp. 522–527.

Pateraki M., Baltsavias E., Recke U., 2004. Experiences on automatic image matching for DSM generation with ADS40 pushbroom sensor data. In: IAPRS and SIS, Vol. 35, Part B2, pp. 402–407.

Petrie G., 2003a. Airborne digital frame cameras. Geoinformatics 6 (7), 18–25.

Petrie G., 2003b. Eyes in the sky. Geoinformatics 3 (1), 42–49.

Pikaz A., Dinstein I., 1994. Using simple decomposition for smoothing and feature point detection of noisy digital curves. IEEE PAMI 16 (8), 808–813.

Pilgrim L.J., 1991. Simultaneous three dimensional object matching and surface difference detection in a minimally restrained enviroment. Ph.D. Thesis, Report No. 066.08.1991, Department of Civil Engineering and Surveying, The University of Newcastle, New South Wales.

Pilgrim L., 1996. Robust estimation applied to surface matching. ISPRS Journal of Photogrammetry and Remote Sensing 51 (5), 243–257.

Poli D., Zhang L., Gruen A., 2004. SPOT-5/HRS stereo images orientation and automated DSM generation. In: IAPRS and SIS, Vol. 35, Part B1, pp. 421–432.

Positive Systems, 2004. Positive Systems webpage. http://www.possys.com (accessed December 2004).

Reisfeld D., Wolfson H., Yeshurun Y., 1995. Context-free attentional operators: The generalized symmetry transform. International Journal of Computer Vision 14 (1), 119–130.

Renouard L., Lehmann F., 1999. High resolution digital surface models and orthoimages for Telecom network planning. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week. Wichmann, Heidelberg, pp. 241–247.

Reulke R., Franke K., Pomierski T., von Schönemark M., Tornoe C., Wiest L., 2000. Target related multispectral and true color optimization of the colour channels of the LH Systems ADS40. In: IAPRS, Vol. 33, Part B1, pp. 244–250.

Reulke R., Tempelmann U., Stallmann D., Cramer M., Haala N., 2004. Improvement of spatial resolution with staggered arrays as used in the airborne optical sensor ADS40. In: IAPRS, Vol. 35, Part B1, pp. 114–119.

Rohr K., 1992. Recognizing corners by fitting parametric models. International Journal of Computer Vision 9 (3), 213–230.

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

Roux M., Maître H., 2001. Some more steps towards 3D reconstruction of urban areas from multiple views. In: E. Baltsavias, A. Grün, L. Van Gool (Eds.), Automatic extraction of Man-Made objects from Aerial and Space images (III). A.A. Balkema Publishers, Lisse, pp. 135–147.

Ruzon M.A., Tomasi C., 2001. Edge, junction and corner detection using color distributions. IEEE PAMI 23(11), 1281–1294.

Sandau R., Bärwald W., 1994. A three-line wide-angle CCD stereo camera for Mars-94 mission. In: IAPRS, Vol. 30, Part 1, pp. 82–86.

Sandau R., Eckardt A., 1996. The stereo camera family WAOSS/WAAC for spaceborne/airborne applications. In: IAPRS, Vol. 31, Part B1, pp. 170–175.

Sandau R., Braunecker B., Driescher H., Eckardt A., Hilbbert S., Hutton J., Kirchhofer W., Lithopoulos E., Reulke R., Wicki S., 2000. Design principles of the LH Systems ADS40 airborne digital sensor. In: IAPRS, Vol. 33, Part B1, pp. 258–265.

Scharstein D., Szeliski R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47 (1/2/3), 7–42.

Schenk T., 1999. Digital Photogrammetry, TerraScience, Ohio, USA, pp. 267–299.

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.

Schlueter M., 1998. Multi image matching in object space of a general 3D surface model instead of common 2.5 D surface models and its applications for urban scenes. In: IAPRS, Vol. 32, Part 4, pp. 545–562.

Schlueter M., 2000. Multi-image matching in object space using airborne CCD-linescanner imagery. In: IAPRS, Vol. 33, Supplement Part B3 (on CD-ROM).

Schlueter M., Wrobel B.P., 1996. High resolution surface reconstruction of a landscape from large scale aerial imagery from Facets Stereo Vision - an extended test. In: IAPRS, Vol. 31, Part B3, pp. 758–763.

Schmid C., Mohr R., 1995. Matching by local invariants. Rapport de recherche no. 2644, Unité de recherche INRIA, Rhône-Alpes, France.

Schmid C., Mohr R., Bauckhage C., 2000. Comparing and evaluating interest points. International Journal of Computer Vision 37 (2), 151–172.

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.

Scholten F., Gwinner K., Tauch R., Boulgakova O., 2003. HRSC-AX - High Resolution Orthoimages and Digital Surface Models for Urban Regions. In: Proc. of 2nd GRSS/ISPRS Joint Workshop "Remote Sensing and Data Fusion over Urban Areas", Berlin, Germany, pp. 225–229.

Schuster R. and Braunecker B., 2000. Calibration of the LH Systems ADS40 airborne digital sensor. In: IAPRS, Vol. 33, Part B1, pp. 288–294.

Schut G.H., 1976. Review of interpolation methods for digital terrain models. The Canadian Surveyor 30 (5), 389–412.

Shan J., 1996. Edge detection algorithms in photogrammetry and computer vision. Tech. Rep. No. 1996.2, Department of Geodetic Sciences, University of Stuttgart.

Shashua A., 1994. Trilinearity in Visual Recognition by Alignment.In: Proc. European Conference on Computer Vision, May 2-6, Stockholm, Sweden. Lecture Notes in Computer Science 800, Springer-Verlag, pp. 479–484.

Shilat E., Werman M., Gdalyahu Y., 1997. Ridge's corner detection and correspondence. In: Proc. CVPR, 17-19 June, San Juan, Puerto Rico, pp. 976–981.

Smith S.M., Brady J.M., 1997. SUSAN - a new approach to low level image processing. International Journal of Computer Vision 23 (1), 45–78.

Sojka E., 2002. A new and efficient algorithm for detecting the corners in digital images. In: L. Van Gool (Ed.), Proc. 24th DAGM Symposium, Lecture Notes in Computer Science 2449. Springer, Berlin, pp. 125–132.

Stockman G.C., 1987. Object Recognition and Localization via Pose Clustering. Computer Vision, Graphics, and Image Processing 40 (3), 361–387.

Stoppa D., Gonzo L., Gottardi M., Dalla Betta G., 2002. Novel CMOS image sensor with 132-db dynamic range. IEEE Journal on Solid State circuits 37 (12), 1846–1852.

SUSAN, 2004. SUSAN, Low Level Image Processing and source code homepage. http://www.fmrib.ox.ac.uk/ steve/susan/index.html (accessed December 2004).

Tahir R., Schenk T., 1992. Some more steps towards 3D reconstruction of urban areas from multiple views. In: IAPRS, Vol. 29, Part B3, pp. 227–232.

Tang L., Dörstel C., Jacobsen K., Heipke C., Hinz A., 2000. Geometric Accuracy Potential of the Digital Modular Camera. In: IAPRS, Vol. 33, Part B4, pp. 1051–1057.

Tang L., Heipke C., 1996. Automatic relative orientation of aerial images. PERS 62 (1), 47–55.

Taylor S., 1998. CCD and CMOS Imaging Array Technologies: Technology Review. Tech. Rep. epc 106, Xerox Research Center Europe.

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 airborne digital sensor. In: IAPRS, Vol. 33, Part B2, pp. 552–559.

Tempelmann U., Hinsken L., Recke U., 2003. ADS40 Calibration & Verification Process. In: A. Grün, H. Kahmen (Eds.), Proc. 6th Conference on Optical 3D Measurement Techniques, Vol. 1, published by IGP, ETHZ, Zurich, pp. 48–54.

Terzopoulos D., 1988. The computation of visible-surface representations. IEEE PAMI 10 (4), 417–438.

Thom C., Souchon J.-P., 1999. The IGN digital camera system. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '99, Wichmann, Heidelberg, pp. 89–94.

Tomasi C., Kanade T., 1991. Detection and tracking of point features. Tech. Rep. CMU-CS-91-132, Carnegie Mellon University.

Toth C., 1997. Direct Sensor Platform Orientation: Airborne Integrated Mapping System (AIMS). In: IAPRS, Vol. 32, Part 3-2W3, pp. 148–155.

Toth C., 1999. Experiences with frame CCD arrays and direct georeferencing. In: D. Fritsch, R. Spiller (Eds.), Photogrammetric Week '99, Wichmann, Heidelberg, pp. 95–109.

Toth C., Schenk T., 1992. On matching image patches under various geometrical constraints. In: IAPRS, Vol. 29, part B3, pp. 400–403.

TerraSim–3, 2004. TerraSim–3, multispectral camera systems. STI Services homepage. http://www.sti-services.com/Products/CameraSystems/Multispectral/TS3.html (accessed December 2004).

Triggs B., 1995. Matching Constraints and the Joint Image. In: Proc. International Conference on Computer Vision, 20-23 June, Cambridge, MA, USA, pp. 338–343.

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.

Vosselman G., 1992. Relational matching. In: Lecture Notes in Computer Science 628. Springer, Berlin.

VXL, 2004. VXL Homepage, C++ Libraries for Computer Vision Research and Implementation. http://vxl.sourceforge.net (accessed June 2004).

Wallis R., 1976. An approach to the space variant restoration and enhancement of images. In: Symposium on Current Mathematical Problems in Image Science, 10-12 November, Monterey, CA, USA, pp. 326-340.

Wang Y., 1998. Principles and applications of structural image matching. ISPRS Journal of Photogrammetry and Remote Sensing 53 (3), 154–165.

Weidner U., 1997. Digital Surface Models for Building Extraction. In: A. Grün, E. Baltsavias, O. Henricsson (Eds.), Automatic extraction of Man-Made objects from Aerial and Space images (II). Birkhäuser, Basel, pp. 193–202.

Weng J., 1993. Image matching using the windowed Fourier Phase. International Journal of Computer Vision 11 (3), 211–236.

Wewel F., 1996. Determination of conjugate points of Stereoscopic Three Line Scanner Data of Mars96 Mission. In: IAPRS, Vol. 31, Part B3, pp. 936–939.

Wewel F., Scholten F., Neukum G., Albertz J., 1998. Digitale Luftbildaufnahme mit der HRSC - Ein Schritt in die Zukunft der Photogrammetrie. PFG 1998 (6), 337–348.

Wewel F., Brand M., 1999. Geometrische Validierung des hochauflösenden multispektralen Mehrzeilen-stereoscanners HRCS-A. PFG 1999 (7), 245–252.

Wiedemann C., Tang L., Ohlhof T., 1996. A new matching approach for three-line scanner imagery. In: IAPRS, Vol. 31, Part B3, pp. 940–945.

Wiman H., 1998. Automatic generation of digital surface models through matching in object space. Photogrammetric Record 16 (91), 83–91.

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

Wu X., Murai S., 1997. Image matching using a three line scanner. ISPRS Journal of Photogrammetry and Remote Sensing 52 (1), 20–32.

Zhang B., Miller S., 1997. Adaptive Automatic Terrain Extraction. In: Proc. SPIE, "Integrating Techniques with Scene Analysis and Machine Vision III", Vol. 3072, pp. 27–36.

Zhang L., Grün A., 2003. Automatic DSM generation from TLS data. In: A. Grün, H. Kahmen (Eds.), Proc. 6th Conference on Optical 3D Measurement Techniques, Vol. 1, published by IGP, ETHZ, Zurich, pp. 93–105.

Zhang Z., 1990. Extension of feature bridging mode method and global image matching. In: IAPRS, Vol. 28, Part 3, pp. 1106–1120.

Zhang Z., Zhang J., Wu X., Zhang H., 1992. Global image matching with relaxation method. In: Proc. International Colloqium on Photogrammetry, Remote Sensing and Geographic Information Systems, 11-14 May, Wuhan, China, pp. 175–188.

Zhang Z., Deriche R., Faugeras O., Luong Q.-T., 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial Intelligence 78 (1-2), 87–119.

Zhang Z., Zhang J., Liao M., Zhang L., 2000. Automatic registration of multi-source imagery based on global image matching. PERS 66 (5), 625–629.

Appendix A

Appendix

A.1 Evaluation of AIM for the Yokohama dataset

A.1.1 Dataset

The block of Yokohama consisted of three parallel strips and was flown by Pasco Corp. of Tokyo. It included panchromatic and multispectral imagery in Lev0 and Lev1 products with 0.20 m ground sampling distance. The Japanese UTM grid was used as the coordinate system. The camera utilized the sensor plate SP2 and had been radiometrically and geometrically calibrated (section 2.3.2) with common procedures as for the camera with SP1 used in the Waldkirch dataset. The interior orientation and IMU misalignment parameters had been estimated by photogrammetric means and the orientation parameters were adjusted and refined through the AT process. The geometric accuracy was low as the block geometry was poor (no cross strips), the radiometric quality was low and more blunders existed in automatic tie point measurement, plus some small errors in the GPS ground control data resulted in $\sigma_o = 7.2\mu m$. The GCPs were distributed at the block center and no check points were selected (Figure A.1).

A summary of the acquisition and the bundle adjustment parameters of the dataset are shown in Table A.1. In terms of radiometric quality, the Yokohama imagery 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 (Figure A.2(a)). Therefore, in order to help matching in shadowed areas and reduce noise and to improve feature definition and minimize radiometric differences among channels (section 3.2), the preprocessing of all images used for DSM extraction was essential (Figure A.2(b)).

A.1.2 Reference Data

Reference data was collected by manual stereo measurements in SS and followed the classification scheme used in Test series B (section 6.3). The estimated height accuracy of the manual measurements was $40 - 60$ cm but due to the poor radiometric quality of the Yokohama dataset, many roofs, especially the smaller ones lying partly in shadowed areas, could not be accurately extracted and for several manually

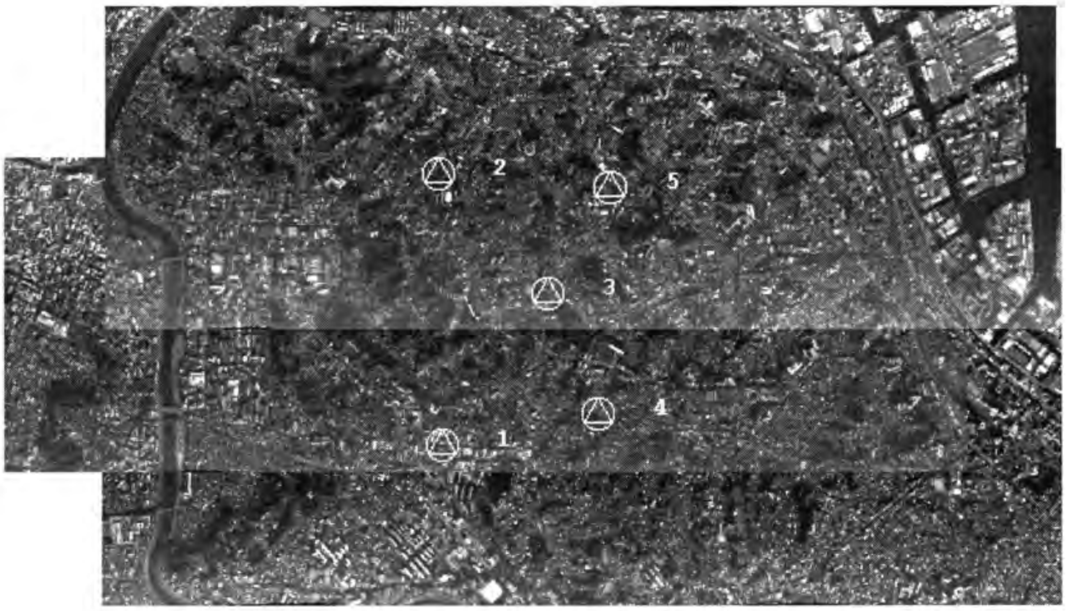
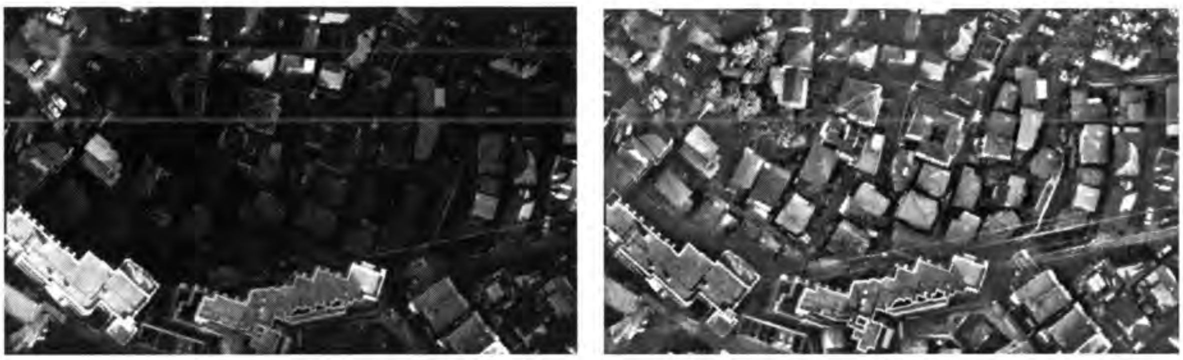


Figure A.1: Yokohama block. On the left the strip numbers are indicated. The GCPs are marked with white color.

measured points the estimated accuracy was degraded (~ 60 cm). Roadmarks were not always easily visible due to shadowing. Less points could be extracted in open surfaces and close to building outlines and therefore the class used to separately analyze modelling errors was not used.

A.1.3 Test results

One representative region of 1000 x 3000 pixels was selected for analysis from the Yokohama dataset (Figure A.3) and it included high buildings (some over 20 m high) and thus large discontinuities and occlusions. Large shadowed areas were excluded. In combination with the low radiometric quality, the degree of difficulty in extracting a reliable DSM of this area increased. For SS a denser matching,



(a) (b)

Figure A.2: Image quality prior to (a) and after preprocessing (b) for the Yokohama dataset.

Parameters	Yokohama
Coord. system	Japan Grid
Acquisition date	October 2003
Camera	SP2
No. strips [paral.– cross]	3–0
Ground pixel size [m]	0.20
Sensor pixel size [μm]	6.5
Radiometric quality	Average-poor
No. GCPs	5
σ_0 [μm]	7.2
Flying height [m]	∽1944

Table A.1: Acquisition and bundle adjustment parameters.

Figure A.3: Selected area from the Yokohama dataset. The total size of the Y0624B area is 1000 x 3000 pixels and here only a part is illustrated.

compared to the tests in Waldkirch, was employed with a 0.5 m grid spacing due to the density of buildings and large parallax differences, while for AIM the MT strategy was used with grid point matching in the upper levels and gradually refinement by edge matching in the lower levels. The statistical measures that have been derived from the elevation differences between manually and automatically-measured points are shown in Table A.2.

Test data/number of matched points {SS/AIM}	Type of points/number of comparison points	System	RMSE (m)	Mean with sign (m)	Maximum absolute (m)	Std. dev. (m)
Y0624B/ {128950/40750}	BE{Pts}/33	SS	4.55	-1.96	14.31	4.16
		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 A.2: Error statistics of reference data (manually measured) minus automatically generated DSM for the Yokohama area. Single and interpolated points from breaklines are indicated as Pts and Brklin respectively.

As one could expect, the accuracy degraded compared to the results from the Waldkirch dataset (poor data quality). Still, AIM delivered better results, especially along discontinuities on MMO, with an RMS error of 1.30 m compared to 7.56 m from SS. In addition, SS exhibited gross errors of over 100 m. The fact that the accuracy was less for BE than MMO for Y0624B is due to the density of the buildings and the relatively short distances of the measured points to neighboring MMO. This made it more difficult to discriminate measurement from modelling errors.