Fine-grained population mapping using coarse census data and satellite imagery

John E. Vargas-Muñoz[1], Nando Metzger[2], Rodrigo C. Daudt[2], Benjamin Kellenberger[1], Thao Ton-That Whelan[3], Brenda Ayo[3], Muhammad Imran[4], Ferda Ofli[4], Konrad Schindler[2], and Devis Tuia[1]

[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[2]Photogrammetry and Remote Sensing, ETH Zurich, Switzerland
[3]International Committee of the Red Cross (ICRC), Switzerland
[4]Qatar Computing Research Institute, HBKU, Qatar

ABSTRACT

Humanitarian organizations need population information for efficient crisis response planning. However, obtaining up-to-date population data in countries in the Global South is difficult due to infrequent censuses and high urban growth rates. The granularity of the available population data is another major issue, as censuses typically provide coarse population data (e.g, at the level of provinces or counties) in developing countries. However, these countries are typically also the regions where humanitarian organizations operate and require fine-grained population maps for applications such as disaster response.

In order to generate fine-grained population maps (e.g., 100m x 100m resolution), available coarse census data are often spatially disaggregated, typically using building maps and assuming direct proportionality between the two[1]. Other approaches make use of several other useful features, such as night light image bands and estimations of distances to the closest road[2]. These features are aggregated for each coarse administrative region, with known population census data, and then used to train a model for performing fine-grained population mapping, therefore assuming that a model predicting at regional scale could be used as such to provide fine-grained predictions. This assumption has the disadvantage that, when the original regions used for training are very coarsely mapped, the difference in resolution in the prediction domain makes the domain shift between predictors increase, thus reducing the model's performance.

In this work, we aim to produce fine-grained population maps using multiple available features and avoiding the aforementioned problem of feature aggregation. We propose a method based on Markov Random Field (MRF) that iteratively improves the initial estimations of a dasymetric disaggregation method. During the iterations, the MRF-based method minimizes an energy function that encourages locations (e.g., regions of 100m x 100m area) with similar features to have similar population predictions while at the same time ensuring that the predictions sum up to a value close to the available regional census data.

We evaluated our proposed method in a scenario with very coarse census data available, in the country of Tanzania, following the validation approach presented by Stevens and colleagues[2]. In the task of population disaggregation, the proposed method improves the accuracy obtained by the aforementioned baselines. Given that the census data used for training is very coarse (170 administrative regions for the whole country), the method proposed by Stevens and colleagues only obtains a R^2 of 0.41 and Mean absolute error (MAE) of 5188[2]. Direct dasymetric disaggregation using building counts obtains R^2 of 0.65 and MAE of 3713[1]. Whereas, our proposed MRF method produces the best estimations, with a R^2 of 0.78 and MAE of 3314.

In our future work, we plan to build a method that can also estimate population solely based on features without using census data for a given region in order to generalize to other countries where census data are unavailable or inaccurate. Furthermore, we aim to use additional features extracted from social media data to improve accuracy.

References

1. Huang, X., Wang, C., Li, Z. & Ning, H. A 100 m population grid in the CONUS by disaggregating census data with open-source microsoft building footprints. *Big earth data* **5**, 112–133 (2021).

2. Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS one* **10**, 1–22 (2015).