OBJECT-ORIENTED ANALYSIS OF VERY HIGH RESOLUTION ORTHOPHOTOS FOR ESTIMATING THE POPULATION OF SLUM AREAS, CASE OF DAR-ES-SALAAM, TANZANIA

M. Aminipouri *, R. Sliuzas [a], M. Kuffer [a]

[a] ITC, Dept. of Urban Planning and Management, Enschede, The Netherlands, (sliuzas, kuffer)@itc.nl

KEY WORDS: Object-Oriented, Very High Resolution, Orthophotos, Slum, *e*Cognition, Segmentation

ABSTRACT:

Obtaining up-to-date spatial information about slum settlements is of great importance for decision making related to slum and housing policy. The availability of Very High Resolution (VHR) satellite imagery in meter or sub-meter level is helpful and promising in the information extraction of slum areas based on object-oriented techniques.
This paper aims to determine the feasibility of using VHR orthophotos (0.6 meter pixels) to create an accurate inventory of buildings in order to estimate the slum population. *e*Cogniton software is used for the image segmentation and classification of the objects of interest (building roofs) for three different slum areas in Dar-Es-Salaam called “Charambe”, “Manzese” and “Tandale”. Roof extraction accuracy is computed on the study wards. In total, out of 1650 reference buildings 1504 buildings are extracted (91% accuracy). The number of inhabitants living in a dwelling unit is obtained from household surveys and used to calculate the Roof Area per Person (RApP) rate. The estimated population by the applied model represents 82.2%, 72.5% and 68.3% for Charambe, Manzese and Tandale wards, respectively.

1. INTRODUCTION

Unplanned development of urban areas and the creation of slum settlements are responses to rapid population growth in fast growing cities like Dar-Es-Salaam, the former capital of Tanzania. The importance of global slum development is reflected in Millennium Development Goal (MDG 7) target 11 “the improvement of the quality of at least 100 million slum dwellers by the year 2020” (UN-Habitat, 2003). One of the substantial steps in meeting this goal is to find reliable procedures for detecting and monitoring slum areas. Obtaining up-to-date spatial information about slum settlements is of great importance for decision making related to slum and housing policy.

The availability of Very High Resolution (VHR) satellite imagery in meter or sub-meter level is helpful and promising in the information extraction of slum areas based on object-oriented techniques. Along with the increasing availability of new high resolution satellite and airborne digital imagery, precise extraction of ground objects, instead of regions of certain land cover classes, has become increasingly important for a variety of remote sensing and GIS applications (Zhang and Maxwell, 2006). Hence, this paper aims to determine the feasibility of using VHR orthophotos (0.6 meter pixels) to create an accurate inventory of buildings in order to estimate the slum population.

In comparison with traditional pixel-based methods, object-oriented image analysis has the advantage of using a hierarchical network of image objects, so-called image segments. Advantages of object-oriented analysis are meaningful statistic and texture calculation, an increased uncorrelated feature space using shape (e.g. length, asymmetry, etc.) and topological features (neighbour, super-object, etc.) and the close relation between real-world objects and image objects. This relation improves the value of the final classification and cannot be fulfilled by common, pixel-based approaches (Benz *et al*., 2004).

In this study, *e*Cogniton software v7.0 developed by Definiens Imaging (Definiens, 2007a) is used for the image segmentation and classification of the objects of interest (building roofs) for three different slum areas in Dar-Es-Salaam called “Charambe”, “Manzese” and “Tandale”.

2. IMAGE SEGMENTATION PROCESS

Segmenting an image into meaningful objects is a fundamental step in object-oriented image analysis; in that the total accuracy of object-oriented classification is highly dependent on the quality of image segmentation. With *e*Cognition, objects are extracted from the image in a number of hierarchical segmentation levels. The term “segmentation” means: “an operation that creates new image objects or alters the morphology of existing image objects according to a given criteria” (Definiens, 2007b).

From the point view of remote sensing, most often a building object is actually visible as the building’s roof (Tian and Chen, 2007). Owing to the fact that the main goal of this study is to estimate the number of slum dwellers based on object-oriented image analysis, the extraction of the slum building roofs will be examined as the preferred type of object.

First several objects representing the land parcels are generated for the study sites based on a “Chessboard Segmentation” (CS). A “Multi-Resolution Segmentation” (MRS) technique is then applied to identify the roofs of houses within each parcel. After the segmentation process, classification of the objects is carried

* Corresponding author. Email: mehdiaminipour@gmail.com

out by using fuzzy membership and then nearest neighbour classifiers.

A detailed description of each step is provided in the following sections. Also, the analysis plan for meeting the needs of this study is a step-wise process, showed as a flowchart in Figure 1.

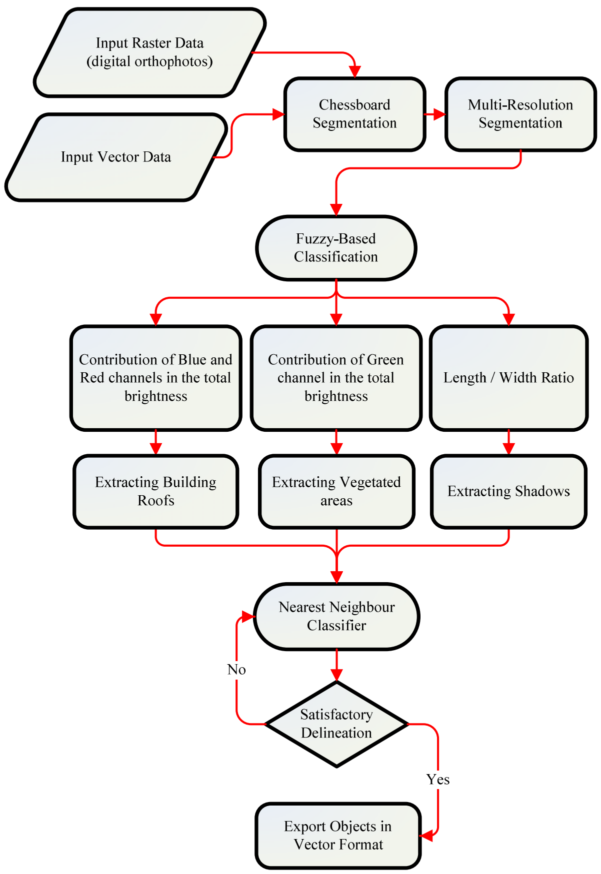


Figure 1. Object-oriented image analysis workflow

2.1 Chessboard Segmentation

First several objects representing the land parcels are generated for the study sites based on an approach called "Chessboard Segmentation" (CS). CS is a top-down segmentation strategy that cuts the scene or the dedicated image objects into smaller objects (Definiens, 2007b). Delineating the land parcels by this technique drastically reduces the number of units to be handled for the upcoming image classification; consequently, it lessens the processing time.

The most important parameter in CS which has to be defined by the user is the object size. In order to produce image objects based exclusively on land parcel (imported thematic layer), the object size must be selected larger than the image size. In other words, the object size must be set to a number which guarantees that one tile is generated only for the whole scene. Since the study areas are hardly larger than 2000 by 2000 pixels, a setting of 5000 pixels for the CS object size is on the safe side. Figure 2 illustrates the delineation of the residential land parcels in different views.

Figure 2. Delineation of land parcels in a sample study site

2.2 Multi-Resolution Segmentation

The next step after mapping out the residential land parcels is to perform a multi-scale image analysis method. This method is well-known as "Multi-Resolution Segmentation" (MRS) technique embedded in *e*Cognition (Baatz and Schape, 2000).

MRS is a bottom-up (assembling objects to create larger objects) segmentation strategy, where the smallest object contains one pixel. In subsequent steps, smaller image objects are merged into larger ones based on the chosen scale, colour and shape parameters. Each parameter can be weighted from 0 to 1. Scale parameter is the most determining factor of MRS with regard to the size of objects (Su *et al.*, 2008) which determines the average image object size; the higher the scale parameter value, the larger the image object becomes.

The MRS approach allows for segmenting an image at different scales, which is used to construct a hierarchical network of image objects. The image objects know their horizontal neighbours (adjacent objects on the same level) as well as their vertical neighbours (objects on different hierarchical levels); the latter also termed sub-objects and super-objects (Laliberte *et al.*, 2004).

The MRS parameters for the study sites are assigned to achieve a realistic segmentation of the land parcels super-objects derived from the CS, such that the roofs of the smallest dwelling units are delineated. Characterisation of an image object based on its super-objects, e.g. dwelling units belonging to a super-object residential land parcel will be classified as residential house, whereas the rest of the objects will remain unclassified.

All parameters of MRS are assigned through a trial-and-error but systematic experimentation. It has to be noticed that the changes in selecting different MRS parameters have taken place simultaneously. In other words, several combinations of the parameters are tested and one optimum set of parameters is selected for each study area based on the visual inspection of the resulting objects, as also described by Im *et al*. (2008) and Nobrega *et al*. (2005).

The scale parameter of 15, 10 and 10 gave the best results for Charambe, Manzese and Tandale respectively.

3. EXTRACTION AND CLASSIFICATION OF IMAGE OBJECTS

The real strength of object-oriented analysis lies in a combination of multi-scale segmentation with subsequent contextual analysis, whereby the spatial, spectral, and contextual properties of extracted segments at different spatial scales are used in conjunction with spatial rules in a subsequent classification (Sliuzas *et al*., 2008).

The object features provided in *e*Cognition supplies a large number of shape, spectral and textural attributes that are used as sources of information to define the inclusion or exclusion parameters used to classify image objects. Object features are obtained by evaluating image objects themselves as well as their embedding in the image object hierarchy. The classification of the objects in *e*Cognition software can be done by two methods: fuzzy membership functions and nearest neighbour classifier.

3.1 Fuzzy Membership Functions

Firstly, the fuzzy membership thresholds are implemented for the object features to describe and classify different classes. The ratio of red, blue and green bands is used for the extraction of the red roofs, common roofs in the region and vegetated areas, respectively. Moreover, length to width ratio is used exclusively for the separation of the shadows casted by the buildings.

For most of the roofs in the study areas, the ratio of blue channel is comparatively high. However, the ratio of red channel is also used to include the roofs in red colour in the extraction procedure. Although the contribution of vegetation and green areas in slum urban environment is low, they must be masked out as a different class in the process. In some cases the trees are misclassified as buildings and this is because of its reflectance also in blue channel which has been applied for the extraction of the rooftops. Hence, the ratio of the green channel is used exclusively for the separation of the vegetated areas inside each land parcel.

The length to width ratio function is also applied for the separation of the shadows, although the ratio varies with the sun azimuth and elevation angles.

Fuzzy membership functions are incorporated at the first level of object class categorisation process. Building rooftops, vegetated areas and shadows are filtered out corresponding to the fuzzy thresholds that have been set for each class. However, misclassification of the objects is inevitable (see examples in Figure 3).

The purpose now is to extract part of the roofs which are ignored or misclassified by the fuzzy membership classifier. Hereafter, "Nearest Neighbour" (NN) classifier is used to examine whether it is possible to refine and improve the classification results derived from the fuzzy membership classifier.

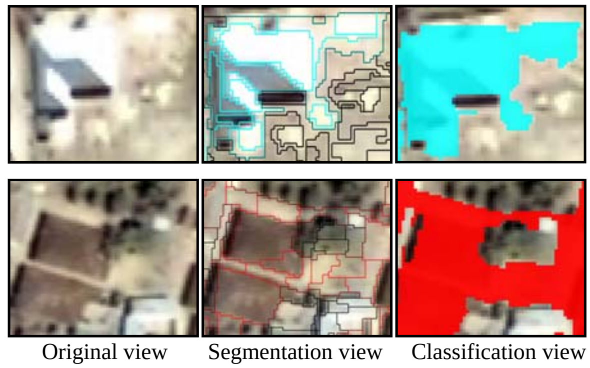


Figure 3. Common missclassification examples of roofs and bare ground

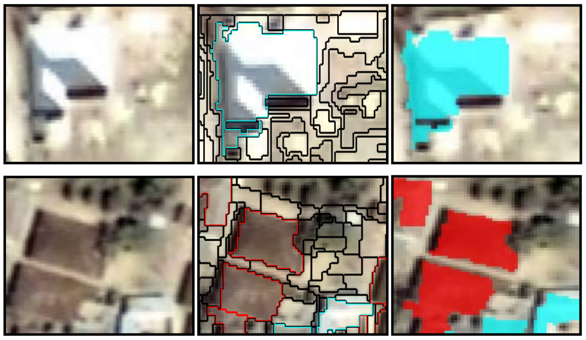
3.2 Classification Refinement Based on NN Classifier

Classification with membership functions are based on user-defined thresholds of object features. In contrast, nearest neighbour classification uses a set of samples for different classes in order to assign membership values. After a representative set of sample image objects has been defined for each class, the algorithm searches for the closest sample image object in the feature space for each image object.

Based on the image object's feature space distance to its nearest neighbour, a membership value between 0 and 1 is assigned by the classifier. The closer the potential degree of membership to 1 is, the higher the probability of belonging to the same class as the user-defined samples.

The "Feature Space Optimisation" (FSO) function in *e*Cognition offers a mathematical method to calculate the best combination of features in the feature space. Determining an initial set of feature objects for the FSO function is just based on the purpose of the classification and the objects to be classified. It is important to notice that the best selection of samples and object features for running the NN classifier is achieved through an iterative procedure.

Ultimately, the classification algorithm in *e*Cognition uses the class descriptor (here is the nearest neighbour descriptor) to assign a class. It evaluates the class description and determines on this basis whether an image object can be a member of this class. According to the aim of this study, the objects of interest which have to be extracted and classified are building roofs. By the time that the entire classification procedure is ended, the rest of the objects (e.g., shadow, vegetated area, bare ground) are considered as unclassified objects category. Figure 4 shows how the classification results are refined by the NN classifier.



Original view Segmentation view Classification view

Figure 4. Classification refinement based on NN classifier

3.3 Export of Classification Result

The first step for exporting the classification results is to reduce the number of classified segments derived from the MRS method. “Merge Region” function is used to merge all the neighbouring image objects of a specific class to one large object without changing the classification. Thereby, the image segments on top of each roof are merged into a single object making the building roofs.

The second step is to export the roof objects as polygons together with their attributes. A large range of feature values may also be exported as the attributes of the roof polygons (vector format). Nevertheless, the only feature value that is of great importance for the next phase of the analysis (estimating the number of slum inhabitants) is the “Area”. Accordingly, all the merged roofs of the three study wards and their attributes (including their area) are then exported.

4. ACCURACY ASSESSMENT

The accuracy of the results is computed on three different study wards, based on a sample size of 550 reference buildings. Using both extracted roofs and reference polygons in vector format, different parameters are assessed and the results are reported for the number of detected buildings, the total extracted roof area and the ratio of roof area coverage (intersection area between extracted and reference buildings). Such parameters are important for slum urban monitoring, planning and management.

4.1 Extraction Rate

The number of detected buildings derived from semi-automatic extraction process, plays a significant role in calculating the extraction rate. Mathematically, extraction rate can be expressed by Equation 1 (Avrahami *et al*., 2004; Mayunga *et al*., 2007; Stassopoulou *et al*., 2000):

$$BER = \frac{BCE}{RBP} \times 100 \tag{1}$$

Where:
BER: Building Extraction Rate
BCE: (number of) Building Correctly Extracted
RBP: (number of) Reference Building Polygon

The only assumption defined for calculating the BER is that if a reference building is covered by more than 25% of an extracted building (see Equation 2), then it is considered as a correctly identified building. This arbitrary threshold is fixed to take into account the co-registration errors and the fact that semi-automated extraction process may not delineate entire buildings as manual photo-interpretation does (Durieux *et al*., 2008).

4.2 Roof Area Coverage

In the roof area coverage metric, the intersected area between reference polygons and extracted buildings is computed in percentage. The “Roof Area Coverage” (RAC) is characterised as the roof area coverage of the extracted buildings with respect to the reference polygons, as shown by Equation 2:

$$RAC = \frac{EBA}{RPA} \times 100 \tag{2}$$

Where:
RAC: Roof Area Coverage
EBA: Extracted Building Areas (m^2)
RPA: Reference Polygon Areas (m^2)

4.3 Estimation of Slum Population

The model designed for the population estimation is called “Roof Area” (RA) which uses a combination of different factors. In RA model, population has a direct relation with roof area and also with roof area per person. The constitutive components of this model are shown as a mathematical function in Equation 3:

$$TEP = EBA \times RApP \tag{3}$$

Where

$$RApP = \frac{RPA}{TRP} \tag{4}$$

In Equation 3, TEP is the “Total Estimated Population”, EBA is the total “Extracted Building Areas” and RApP is the “Roof Area per Person”. In Equation 4, RPA is the total “Reference Polygon Areas” and TRP is the “Total Reference Population”.
The number of inhabitants living in a dwelling unit is obtained from household surveys and used to calculate to Roof Area per Person (RApP).

Accuracy assessment of the estimation of slum population process is of great importance, since it is the main objective of this research. In addition to R^2 which is often used to evaluate the performance of a model based on the modelling dataset, “Relative Error” (RE) is used to assess the model performance based on the validation dataset. The relative or proportional error of population estimation is defined in Equation 5:

$$RE = \frac{(TEP - TRP)}{TRP} \times 100 \quad (5)$$

Where:
RE: Relative Error
TEP: Total Estimated Population
TRP: Total Reference Population

Since the error of estimation for individual study areas may be positive or negative, an average of the absolute values of relative error is further calculated (see Equation 6) which is an indicator of overall accuracy:

$$MRE = \frac{\sum_{k=1}^{n} RE_k}{n} \quad (6)$$

Where MRE is the “Mean Relative Error” and k indexes the number of study areas that is *n* (Lu *et al*., 2006; Harvey, 2002).

5. RESULTS AND DISCUSSIONS

5.1 Roof Extraction Process

Visual inspections overlaid with the reference data revealed that the shape and the orientation of the extracted buildings are different and that there are some small positional errors. It is important to mention that the resulted roof objects from our methodology differ from the reference polygons. In that there can be more than one reference building in one extracted roof with consecutive impacts on comparative area (see Figure 5). Furthermore, the reference polygons are formed by straight lines and they delineate the entire buildings. Whereas, our method detects only part of the roofs (with not very smoothed and orthogonal borders) that match the extraction rule base in *e*Cognition. Thus, only the number of correctly extracted buildings, the total roof area and the intersected area between extracted and reference polygons are considered for the accuracy assessment.

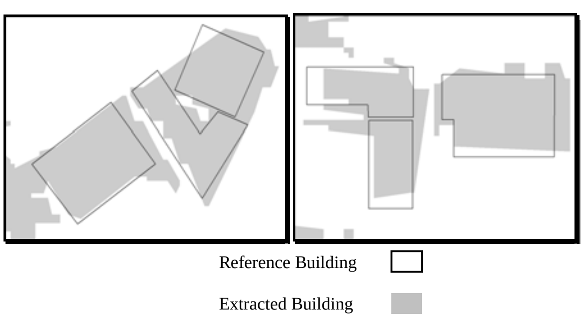


Figure 5. Examples of extracted and reference polygons, showing the positional error

Roof extraction accuracy is computed on three different study wards, based on a sample size of 550 reference buildings. Considering the 25% coverage threshold, 1504 buildings are extracted out of 1650 reference buildings (91% accuracy).

Across all three study sample sites, the total roof area coverage (see Equation 2) extracted from object-oriented method is estimated at 136720 m^2 and from reference data 183195 m^2. This amount shows that 25.7% of the reference polygons are not covered by the extracted roofs. This last result must be interpreted with caution, since it could express the fact that the missing parts of the buildings are not extracted because they are covered by trees or by different roof materials. Nevertheless, this error could be counterbalanced by the merged adjacent objects like cars or bright bare ground (Durieux *et al*., 2008).

Owing to the fact that each study area has its specific physical characteristics, they must be studied separately. The BER and RAC for each sample study site are explained in more details in Table 1.

Ward	*BCE*	*BER*	*EBA/RPA (m2)*	*RAC*
Charambe	534	**97.1%**	53852 / 65518	**82.2%**
Manzese	496	**90.2%**	43025 / 59332	**72.5%**
Tandale	474	**86.2%**	39843 / 58345	**68.3%**
Average		**91.1%**		**74.3%**

Table 1. The resulted extraction rate and roof area coverage for the study areas

5.2 Slum Population Estimation

The overall TEP calculated for the three test sites shows that 74% of the population is estimated by the RA model. Nevertheless, the estimation of population varies in different wards representing 82.2%, 72.5% and 68.3% for Charambe, Manzese and Tandale, respectively. The data shows that building occupancy (RApP) in Charambe, a relatively new area, is 20 m^2 per person, less than half of that in the two older wards.

Statistically, the model performance is assessed using the Relative Error (RE) function (see Equation 5). The RE results ranged from -17.8% for Charambe, -27.5% for Manzese and -31.7% for Tandale. The negative values indicate that the population for all the three study sites are under-estimated and below the reference population data. The derived MRE value determines that the estimation of population in general, is 25.7% less than the real number of slum inhabitants. The overall TEP (74.3%) conforms to the MRE value (25.7%).

6. CONCLUSIONS AND RECOMMENDATIONS

This work has explored an object-oriented method to count the slum buildings from which the number of inhabitants can be estimated. The results presented in this study proved to be satisfactory both for the roof extraction and estimation of slum population. Simplicity is an advantage of the methodology outlined in this study. The proposed approach is simple, effective and can be applied by researchers and non-professional users.

Finding a well suitable method for the building extraction process is highly depending on the user-defined scale parameter, shape and colour criteria. As the results of the image segmentation strongly depend on the image data and the assessment of the segmentation results depends on the classification task, it is almost impossible to suggest well-suited segmentation parameters in general (Hofmann, 2001). Nevertheless, it would be worthwhile to examine the effect of assigning different factors on the classification accuracy results.

VHR digital orthophotos (0.6 meter pixels) is well-suited for the extraction of slum buildings. However, it is expected that the results can be further improved if multi-band imagery including the NIR is available.

Further research should investigate the incorporation of ancillary data (e.g. DSM) in object-based methods to support the identification and extraction of slum buildings.

REFERENCES

Avrahami, Y., Raizman, Y. & Doytsher, Y. (2004) Semiautomatic 3D of building from medium scale (1:40000) aerial photographs. *International Society for Photogrammetry and Remote Sensing Commission III* Istanbul, Turkey.

Baatz, M. & Schape, A. (2000) Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation. IN Strobl, J., Blaschke, T. & Griesebner, G. (Eds.) *Angewandte Geographische Informations-Verarbeitung XII.* Karlsruhe, Germany.

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I. & Heynen, M. (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing,* 58**,** 239-258.

Definiens (2007a) Definiens Imaging Developer 7. *eCognition Software*. *EII Earth.* Munich, Germany.

Definiens (2007b) Definiens Imaging Developer 7. *eCognition User Guide*. Definiens AG.

Durieux, L., Lagabrielle, E. & Nelson, A. (2008) A method for monitoring building construction in urban sprawl areas using object-based analysis of Spot 5 images and existing GIS data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 63**,** 399-408.

Harvey, J. T. (2002) Estimating census district populations from satellite imagery: some approaches and limitations. *International Journal of Remote Sensing,* 23**,** 2071 - 2095.

Hofmann, P. (2001) Detecting urban features from IKONOS data using an object-oriented approach. *First Annual Conference of the Remote Sensing & Photogrammetry Society.* Munich, Germany.

Im, J., Jensen, J. R. & Tullis, J. A. (2008) Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing,* 29**,** 399 - 423.

Laliberte, A. S., Rango, A., Havstad, K. M., Paris, J. F., Beck, R. F., Mcneely, R. & Gonzalez, A. L. (2004) Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sensing of Environment,* 93**,** 198-210.

Lu, D., Weng, Q. & Li, G. (2006) Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing,* 27**,** 3553 - 3570.

Mayunga, S. D., Coleman, D. J. & Zhang, Y. (2007) A semi-automated approach for extracting buildings from QuickBird imagery applied to informal settlement mapping. *International Journal of Remote Sensing,* 28**,** 2343 - 2357.

Nobrega, R. A. A., O'hara, C. G. & Quintanilha, J. A. (2005) Detecting roads in informal settlements surrounding Sao Paulo city by using object-based classification. *OBAI'06 - 1st International Conference on Object-Based Image Analysis.* Slazburg, Austria.

Sliuzas, R. V., Kerle, N. & Kuffer, M. (2008) Object - oriented mapping of urban poverty and deprivation. *Proceedings of the 4th EARSeL (European Association of Remote Sensing Laboratories) workshop on remote sensing for developing countries in conjunction with GISDECO 8.* Istanbul, Turkey.

Stassopoulou, A., Caelli, T. & Ramirez, R. (2000) Automatic extraction of building statistics from digital orthophotos. *International Journal of Geographical Information Science,* 14**,** 795 - 814.

Su, W., Li, J., Chen, Y., Liu, Z., Zhang, J., Low, T. M., Suppiah, I. & Hashim, S. A. M. (2008) Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International Journal of Remote Sensing,* 29**,** 3105 - 3117.

Tian, J. & Chen, D. M. (2007) Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *International Journal of Remote Sensing,* 28**,** 4625 - 4644.

Un-Habitat (2003) Guide to monitoring Target 11: Improving the lives of 100 million slum dwellers.

Zhang, Y. & Maxwell, T. (2006) A trained segmentation technique for optimization of object-oriented classification. *ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processes".* Enschede, The Netherlands.