The development of a morphological unplanned settlement index using very-high-resolution (VHR) imagery

Kuffer, Monika

Barros, Joana

Sliuzas, Richard

Abstract

Information about unplanned settlements in cities in developing countries is often unavailable or incomplete, mainly due to a combination of their informal development and capacity constraints of planning authorities. Despite the extent of unplanned areas in many countries, which at times can dominate residential land-use, very few tools exist for their identification and monitoring. Therefore, there is a clear need for such tools to support timely updating of spatial databases. The present research aims to contribute to the development of such tools, by using spatial metrics to characterise the morphology of unplanned urban settlements in VHR images. The methodology was tested in two case study areas: Dar es Salaam (Tanzania) and New Delhi (India). The research methodology is built on using image segmentation and on the assumption that segments representing homogenous urban patches are different in planned and unplanned areas. Homogenous urban patches were extracted using multi-resolution image segmentation. The morphological aspects (size, density and layout pattern) of planned and unplanned patches were then analysed using spatial metrics. A set of metrics that reflected morphological characteristics of unplanned settlements was identified. This final set was used to build an 'unplanned settlement index' (USI) using spatial multi-criteria evaluation methods. Comparison between results and available land use data showed that the index can assist in the identification of unplanned settlements, with an accuracy of 73% for five selected parts of New Delhi and 75% for Dar es Salaam%.

Keywords: Unplanned settlement index; urban morphology, VHR imagery, image segmentation; spatial metrics

1. Introduction

The morphology of planned and unplanned built-up areas in cities of developing countries often shows distinct differences. An unplanned area is usually developed without planning provision and is often associated with informality, overcrowding, insufficient infrastructure provision, and poor housing quality and haphazard layout (UN-Habitat, 2008). Information on the extent and nature of such areas is commonly unavailable and frequently of poor temporal accuracy and consistency (M. Herold, Goldstein, & Clarke, 2003). Thus mapping such areas, as well as understanding their heterogeneity and development dynamics, is not only a concern but also a challenge for local authorities. The amount of unplanned areas differs between cities and (developing) countries. However, they are an issue for the majority of large cities in developing countries, where the extent of unplanned areas commonly ranges between 30 and 60% of total areas (Busgeeth, Brits, & Whisken, 2008). In fact, in some Sub-Saharan African cities the amount of unplanned urban land can dramatically exceed the amount of planned (Kombe, 2005). Estimates show that this is likely to increase in future (UN-Habitat, 2003).
The need to map and monitor unplanned settlements requires the development of methods and tools that are both low cost and effective. As such, good and reliable data and derived information that is easily accessible and timely is important to better manage unplanned urban development (Turkstra & Raithelhuber, 2004). When aiming at detecting and/or analysing unplanned settlements, it is important to understand clearly what spatial characteristics make these areas differ from the planned ones. In VHR images (such as the one shown in Figure 1), unplanned settlements can be easily visually identified because of their organic layout and densely clustered buildings, while transition areas or areas that underwent regularisation or new unplanned developments in the outskirt will display more complex and ambiguous spatial characteristics.

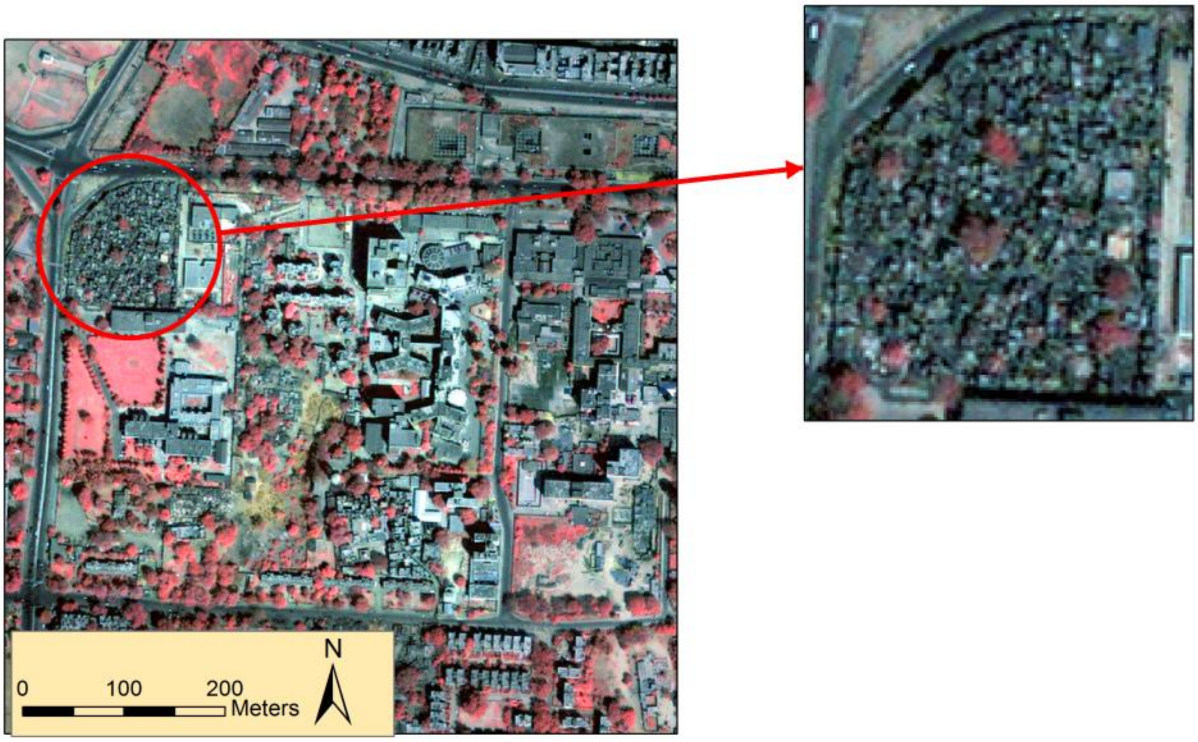


Figure 1: Example of false colour VHR image (Delhi), unplanned settlement (right), (Ikonos 2001)

For this research the terms unplanned area was selected as it captures best its development process, where development happens without zoning, site planning and service provision, resulting commonly in more irregular layout patterns without obeying planning standards. Other related terms such as informal, spontaneous, illegal, slum and squatter settlements have different connotations that relate e.g. to the tenure status (which is very difficult to capture from remotely sensed images). The morphology of unplanned areas presents an organic pattern, meaning the areas are more irregular, complex, diverse and often denser built than the planned ones (Kit, Lüdeke, & Reckien, 2012; Kostof, 1991; Weeks, Hill, Stow, Getis, & Fugate, 2007). In addition, such settlements tend to be irregular shaped and contain significantly smaller size of dwellings than planned areas (Kohli, Sliuzas, Kerle, & Stein, 2012). Planned areas, by contrast, normally have regular street layouts, larger buildings and planned open spaces (Kostof, 1991). In the present study, three of these inherent morphological characteristics were used for the identification of unplanned settlements on VHR. They are:

a) Lack of road infrastructure which causes a more organic **pattern**: the lack of space for urban infrastructure left by the development of individual dwellings results in an unclear or non-existent road infrastructure (Lemma, Sliuzas, & Kuffer, 2006), clearly differentiating these areas from planned ones.
b) Noncompliance with planning standards causes high built-up **densities**: e.g. non-compliance with set-back rules leads to insufficient space between the individual buildings (see example Figure 2-left). Yet, high densities can be also found in low-moderate income (formal) areas (Amato, 1970). However, Taubenböck and Kraff (2013) found in slums (using the case of Mumbai) significantly higher densities than in formal settlements (besides density also size, height of buildings, distance between buildings and an heterogeneity index allowed separating slum from non-slum areas). Also Rakodi and Lloyd-Jones (2002) found a relation between high density settlements and lack of physical capitals of households in such areas. As well as planned public open spaces, in particular green areas (e.g. parks) are absent (Weeks, et al., 2007).
c) The rather small building **sizes** are commonly found in unplanned areas. The role of buildings size as indicator for delineating slum areas was confirmed by expert interview conducted by Kohli et al. (2012) (size was one of eleven basic slum indicators – other indicators listed where roof material, absent or irregular roads, lack of vegetation and open spaces, density, irregular shape of settlement, association with neighbouring areas, texture and locality).

Figure 2: Unplanned areas in Dar es Salaam, *setback between adjacent buildings (left) and street view of area (right)*

The presence of these three morphological features (Baud, Kuffer, Pfeffer, Sliuzas, & Karuppannan, 2010) was used as a criteria for the identification of unplanned areas in remotely sensed imagery, as demonstrated in Table 1. The role of these three basic features to detect unplanned areas has been confirmed by several previous studies (Baud, et al., 2010; Kohli, et al., 2012; Lemma, et al., 2006; Stewart & Kuffer, 2007) with cases of Asian and African cities. These three indicators have been appearing as the most comprehensive while others are more locations specific, such as building height (in some Asian cities unplanned areas can have several stories while in many African cities single story buildings are more common). The main limitations of these three selected morphological features are that they focus on developed areas while areas in the outskirt of the city which are just starting to growth would not be detected by the combination of the three features (having e.g. commonly lower densities). However, as pointed out by Rüther, Martine, & Mtalo (2002), these same features cause limitations to the use of remote sensing methods as the lower accuracies often obtained when extracting buildings in unplanned areas is caused by their unstructured characteristics. As a consequence, very high spatial resolution is needed (at times below ½ m) to allow the identification of small and densely clustered buildings (Sliuzas, Kerle, & Kuffer, 2008).

Table 1 Common morphological features of unplanned areas

Morphological features	Unplanned areas	Planned areas
Size	• Small (substandard) building sizes	• Generally larger building sizes
Density	• High densities (roof coverage densities at least 80% and more) • Lack of public (green) spaces within or in the vicinity of residential areas	• Low – moderate density areas • Provision of public (green spaces) within or in vicinity of residential areas
Pattern	• Organic layout structure (no orderly road arrangement and noncompliance with set-back standards)	• Regular layout pattern (showing planned regular roads and compliance with set-back rules)

In the past decades remote sensing has become an increasingly important data source to support urban planning and management, especially since the availability of very-high resolution (VHR) images with resolution of 1m and below. This new generation of VHR imagery "has created spatiotemporally continuous and politically less biased sources of data" (Kit, et al., 2012, p. 661).

Similarly, the use of spatial metrics for urban applications has also grown in importance. Spatial metrics originate from the discipline of landscape ecology where they are used to analyse environmental patterns of ecological processes. Analysis of patterns plays an important role for urban applications, where homogenous patches of land use types, for an example, can be analysed through the “physical arrangement and interaction of the patches” (Chin, 2006, p. 217), providing a “global summary descriptor of the pattern of the landscape” (Chin, 2006, p.218). Herold, Couclelis, & Clarke, (2005, p. 371) stressed that spatial metrics used in combination with remote sensing “can provide more spatially consistent and detailed information on urban structure and change than either of these approaches”.

In the past years a large number of publications used spatial metrics to quantify patterns of urban land-cover/use (e.g. Frohn 2006, Herold et al. 2005, Chin, 2006, Taubenböck, 2009, Schwarz, 2010, Peng et al. 2010). Many studies have been analysing spatiotemporal dynamics on metropolitan scale (e.g.M. Herold, Goldstein, et al., 2003; Pham, Yamaguchi, & Bui, 2011; Taubenböck, Wegmann, Roth, Mehl, & Dech, 2009). While only few studies used spatial metrics to differentiate between urban spatial categories e.g. extracting areas of different population densities (Liu, Clarke, & Herold, 2006), areas of different urban land use categories (M. Herold, Liu, & Clarke, 2003) or focus on the use of spatial metrics to indicate physically deprived areas (at a settlement scale) (Baud, et al., 2010; Kit, et al., 2012). Despite spatial metrics being one of the most growing applications of remotely sensed data (Saura & Castro, 2007), they have been not much used in the image classification phase, as it was believed that only classified images (categorical maps) can be analysed by spatial metrics (Frohn, 2006). Frohn has also suggested that the only prerequisite for the use of spatial metrics is the presence of homogeneous regions (patches). For this purpose, Herold, Liu, & Clarke (2003) established ‘land-use-regions’ which are ‘homogenous urban patches’ of a specific land-use type having similar texture, sufficient sizes and being bounded by streets. Yet, there seems to be no consensus on the set of spatial metrics most suitable for urban applications (M. Herold, et al., 2005).

This paper demonstrates how to analyse the morphological character of unplanned areas by capturing their main spatial characteristics through a set of spatial metrics derived from VHR remotely sensed images. Therefore it explores the use of spatial metrics to produce comprehensive information on settlement scale with the specific focus on unplanned areas to add to the larger body on literature on metropolitan scale. The paper is structured as follows. Section 2 highlights the two case study areas used for developing and testing the methodology. Section 3 describes the data and methodology to extract homogenous patches, the selection of a set of spatial metrics and the aggregation to an unplanned settlement index. Section 4 presents and discusses the respective results. Section 5 provides the final conclusion, limitations and future research directions.

2. Case Studies

Two case study areas were selected to contrast to two different urban environments: an Asian example, New Delhi, India and an African example, Dar es Salaam Tanzania. New Delhi is a very densely built-up urban environment that presents many different types of unplanned settlements, ranging from the very small squatter pockets to large unplanned settlements (e.g. unauthorized colonies). Dar es Salaam (Tanzania) was selected as example of a sub-Saharan African city with large areas of unplanned settlements dominating the planned areas. By comparison, around 38% of New Delhi’s population lives in such unplanned areas (Risbud, 2002) and 70% of the residential land in Dar es Salaam is unplanned (Kombe, 2005). VHR imageries were available for both cities (presented in Table 2). For Delhi, Ikonos scenes (multi-spectral and panchromatic bands) of the years 2001 and 2002 were pan-sharpened to 1 m spatial resolution. In the case of Dar es Salaam the multi-spectral QuickBird image mosaic of 2007 was also pan-sharpened resulting in 0.6 m spatial resolution.

Table 2: Overview of data sets for both case study areas

Dataset Dar	Year	Description
QuickBird mosaic	2007	Covering most part of the city
Land use data (location of unplanned areas)	2002	Polygons (boundaries) of unplanned areas
Dataset Delhi	Year	Description
Ikonos mosaic	2001/2	Covering the entire city
Land use data of 12 wards (location unplanned areas)	2001/2	Polygons (boundaries) of unplanned areas

For each city, ten different areas were selected: five test areas (Figure 3) for evaluating the spatial metrics, and five assessment areas (Figure 9) for testing the performance of the unplanned settlement index. The areas selected in Dar es Salaam cover different types of planned and unplanned settlements, including older and consolidated settlements and unplanned areas which show lower densities (having still some development scope). Areas selected in Delhi included small pockets of squatter settlements and large areas of unauthorized colonies.

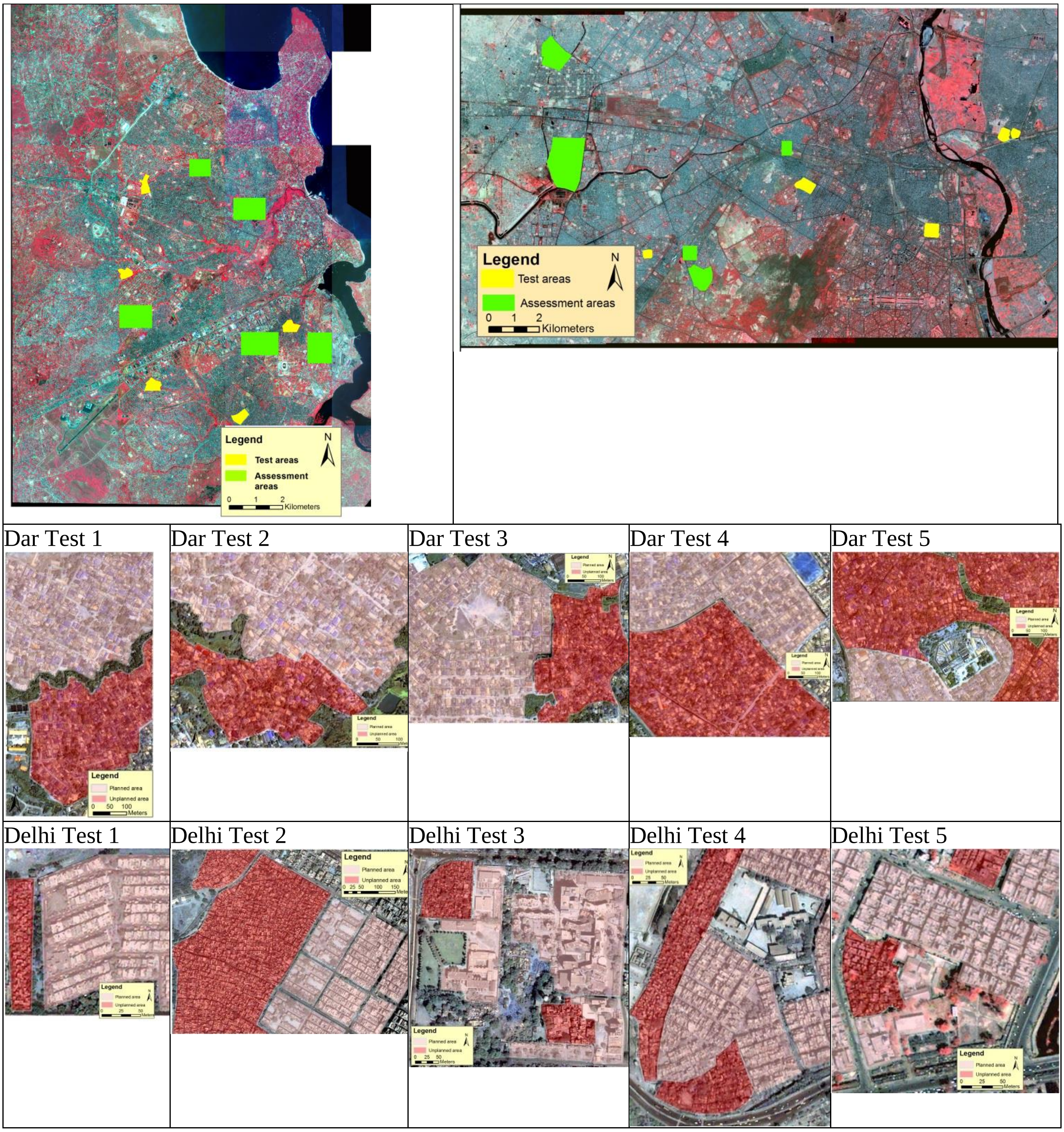


Figure 3: On the top, the location of test and assessment areas in Dar es Salaam – *QuickBird image 2007 (left) and Delhi – Ikonos image 2001-2 (right). Below, an overview of the individual test areas*

The selection of the test and assessment areas was bound to the available reference data.. For the city of New Delhi the result of a visual image interpretation validated by ground truth data was available from a previous research (Baud et al. 2010). The ground validation done in 2008 covered 12 wards (covering a range from highly deprived to low deprived wards). To reduce problems of temporal consistency (between the image and ground truth data) the data collection was only including areas of stable land use. From the 12 wards available in the reference data set of Delhi 10 were selected, one ward of Delhi was excluded because of its manly institutional characteristics as well as a second ward because it did not have a significant unplanned area.. The 10 available wards were spit into five test and five assessment areas. For the city of Dar es Salaam land use information of the year 2002 was available for this research. The data was generated by manual digitising of VHR images, this data set was also used in previous researches and showed good accuracy (Sliuzas, 2004). The data set of Dar es Salaam is covering the entire urban area, while the data set available for Delhi includes only 10 wards. In order to have a comparative data base for both cities the a selection of areas was spilt into five test and five assessment areas. .

3. Methodology

The proposed methodology consists of three steps, as shown in Figure 4. First, homogenous urban patches (HUPs) at object (building) level were extracted using image segmentation. Second, a set of spatial metrics was selected that has potential for analysing the morphology of the selected test areas. Third, the unplanned-settlement-index was calculated using significant spatial metrics and aggregated for HUPs (at area level). The final index is then assessed using reference data.

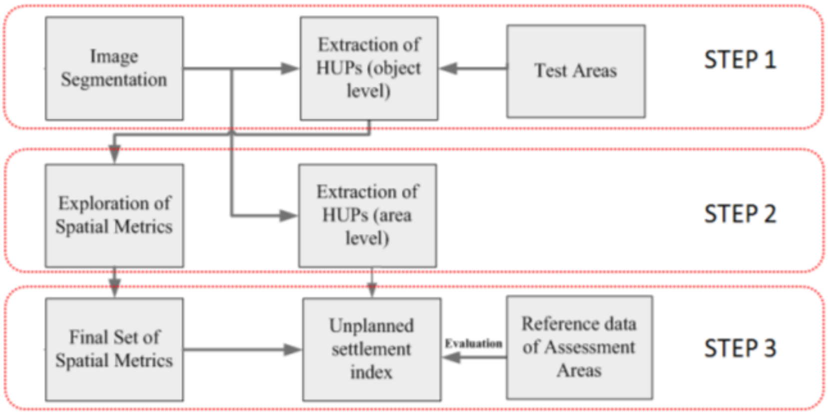
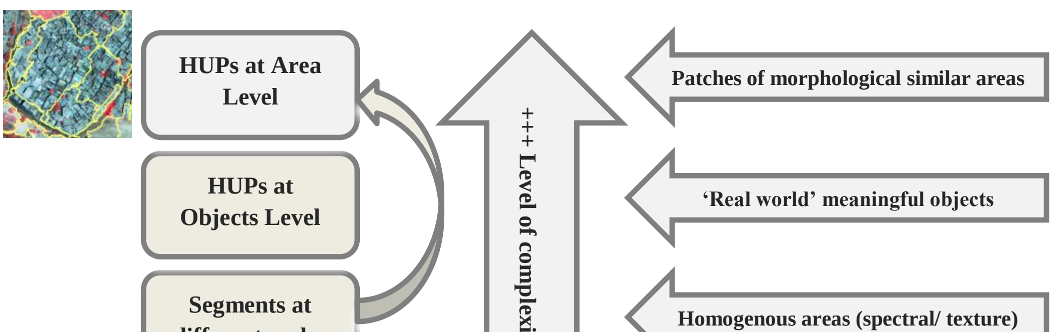


Figure 4: Overview of the analysis procedure

2.1. Step 1: Extraction of homogenous urban patches (HUPs) using image segmentation from VHR imagery

The first step consists of segmenting the VHR imageries in order to extract HUPs at object level (see Figure 4). Image segmentation is used for partitioning the image into non-intersecting regions, thus each region is homogenous while neighbouring regions are kept heterogeneous (Wang, Jensen, & Im, 2010). The result of the image segmentation process can be either a complete or partial segmentation. While the first extracts real-world objects, the latter is commonly used as input into further remote sensing techniques to ultimately extract real-world objects (Wang, et al., 2010). The advantage of image segmentation is that various information levels can be extracted from an image depending upon scale. The most basic level in an image are individual pixels containing spectral information while homogenous neighbouring pixels form segments using e.g. spectral and spatial threshold values (see Figure 5). Homogenous patches representing building objects (HUPs at object level) can be extracted depending on the object characteristics and the spectral and spatial resolution. Larger areas with homogenous physical characteristics, e.g. street blocks or homogenous settlements (HUPs at area level) can be extracted using spatial characteristics like density or texture. Figure 4 shows the increase of complexity from pixel to HUPs at area level for extracting different information levels from an image.



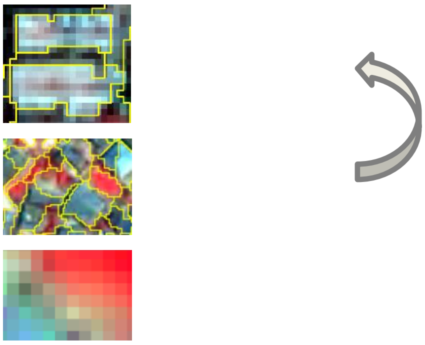


Figure 5: Scale dependent information levels of image using segmentation (from pixels to higher level objects)

Image segmentation and the selection of a suitable segmentation algorithm for an application are important as well as difficult steps in image analysis (Cheng, Jiang, Sung, & Wang, 2001). Zhang, Fritts & Goldman (2008) differentiate between 3 main classes a) pixel-based, e.g. histogram thresholding, b) region based, e.g. region-growing methods and c) boundary based methods e.g. edge-detection. A large number of commercial as well as public domain segmentation algorithm are available, which have been implementing the different algorithms. For this research three different segmentation algorithm where considered. The selection included the multi-resolution algorithm of Definiens (Ecognition) a very popular image segmentation algorithm (Wan et. al 2010) which is a bottom-up, region-merging algorithm based on Fractal Net Evolution Approach (Baatz et al. 2004) which is a combination of a histogram-based method and the homogeneity measurement. The second segmentation algorithm used in Prabat was a region growing segmentation. And the third algorithm was Erdas Imagine segmentation which is an edge detect algorithm. Thus the selected algorithm covered different types classes of segmentation algorithms. All algorithm required optimising the parameters. Wuest and Zhang (2009) stated that it is useful to experiment with segmentation parameters as no standard rules exists for selecting good segmentation parameters.

Extracting individual buildings using image segmentation of VHR images in an urban area is a complex process. In order to successfully extract individual buildings, it is necessary to have a minimum number of pixels per building (according to Welch (1982) these were usually four in moderate resolution images – while in VHR images more than four pixel will be required to identify a building object) as well as a clear set-back between buildings. However, in unplanned settlements, individual buildings often cannot be easily extracted because there are not enough pixels per building but also the organic layout with high densities does not present enough spacing between buildings. As the complete segmentation of buildings was not always possible as anticipated, the aim for this procedure was reformulated to extract segments that reflect the morphological structure (approximately at building object level) for the case study areas. In order to achieve this objective, three different segmentation algorithm (Erdas edge detection, Definiens multi-resolution and Prabat region growing) representing different classes of segmentation algorithm were compared after optimizing the segmentation parameters to best fit the objects on ground.

Once the segments were extracted, the result of image segmentation was assessed. According to Wang et al. (2010), there is no standard approach available to quantify the accuracy of image segmentation. The criteria to define a good segmentation is very much application dependent and often requires a comparison between the results of several segmentation algorithms (Zhang, et al., 2008). The performance of the three segmentation algorithms was compared using the area fit index (AFI) suggested by Lucieer (2004), which compares the area of a reference object (here selected building objects which were manually digitised in the test areas) with the area of the largest segment.

$$AFI = \frac{A\, reference\ object - A\, largest\ segment}{A\, reference\ object}$$

(Where A is the area in pixels, the result of a perfect match is 0, if AFI is greater than 0 over-segmentation occurred when AFI is less than 0 under-segmentation happened.)

2.2 Step 2: Selection of a set spatial metrics

In the next step a set of spatial metrics were selected with the potential to analyse the morphology of urban areas, in respect to size/density and pattern. Based on the review of existing research (Barros Filho & Sobreira, 2005; Barros, 2004; Chin, 2006; M. Herold, et al., 2005; M. Herold, Goldstein, et al., 2003; M. Herold, Liu, et al., 2003; Peng, et al., 2010; Schwarz, 2010; Taubenböck, et al., 2009; Yu & Ng, 2007), all listed metrics (presented in Table 2) were analysed for their potential to differentiate planned and unplanned areas. A common problem for selecting metrics is their high correlation (Huang et al. 2007). In a first step all metrics in Table 2 were analysed on their potential to show clear separation between planned and unplanned areas. For this purpose the five test areas were used and all metrics were calculated for the unplanned as well as for the planned areas. Metrics that did not show a clear separation between planned and unplanned areas were excluded. In a second step metrics that were highly correlating within one dimension were excluded. All spatial metrics were calculated using the software FRAGSTATS (McGarigal, Cushman, Neel, & Ene, 2002).

Table 3: Considered spatial metrics for the different dimension of 'unplannedness'

Main spatial characteristics	**Spatial metrics with the potential of measuring spatial characteristics**	**Formula**	**Definitions**
Size *The mean patch size is expected to be higher in unplanned areas when groups of buildings are clustered into one patches because of little space between building objects.*	• Mean patch size	$MN = \frac{\sum_{j=1}^{n} x_{ij}}{n_i}$	x_{ij} = patch size, n_i = total number of patches
	• Area standard deviation	$SD = \sqrt{\frac{\sum_{j=1}^{n}\left[x_{ij} - \left(\frac{\sum_{j=1}^{n} x_{ij}}{n_i}\right)\right]^2}{n_i}}$	
	• Effective mesh size	$MESH = \frac{\sum_{j=1}^{n} a_{ij}^2}{A} \times \left(\frac{1}{10000}\right)$	aij = area (m^2) of patch ij, A = total landscape area (m^2).
	• Splitting index	$SPLIT = \frac{A^2}{\sum_{j=1}^{n} a_{ij}^2}$	
	• Landscape division index	$DIVISION = \left[1 - \sum_{j=1}^{n}\left(\frac{a_{ij}}{A}\right)^2\right]$	
Density *In unplanned areas densities of patches will be different compared to planned*	• Patch density	$PD = \frac{n_i}{A}(10000)(100)$	
	• Edge density	$ED = \frac{\sum_{k=1}^{m} e_{ik}}{A}(10000)$	e_{ik} = total length (m) of edge in landscape involving patch type (class) i; A = total landscape area (m^2).
	• Patch richness density	$PRD = \frac{m}{A}(10000)(100)$	m = number of patch types (classes) present in the landscape, A = total landscape area (m^2).
Layout structure (pattern/shape) *Unplanned areas are commonly more aggregated with complex patterns and less diverse.*	• Aggregation index	$AI = \left(\frac{g_i}{max \rightarrow g_{ij}}\right)(100)$	g_{ii} = number of like adjacencies (joins) between pixels of patch type i based on the single-count method, max->g_{ii} = maximum number of like adjacencies (joins)
	• Fractal dimension	$FRAC = \frac{2\ln(.25p)}{\ln a_{ij}}$	p_{ij} = perimeter (m) of patch ij, a_{ij} = area (m^2) of patch ij.
	• Shape index	$SHAPE = \frac{p_{ij}}{\min p_{ij}}$	p_{ij} = perimeter of patch ij in terms of number of cell surfaces, min p_{ij} = minimum perimeter of patch ij
	• Perimeter area ration	$PARA = \frac{p_{ij}}{a_{ij}}$	pij = perimeter (m) of patch ij, aij = area (m2) of patch ij.
	• Shannons diversity index	$SHDI = \sum_{i=1}^{m} P_i * lnP_i$	P_i = proportion of the landscape occupied by patch type (class) i.
	• Simpson diversity	$SIDI = 1 - \sum_{i=1}^{m} P_i^2$	

	index		
	• Shannons evenness index	$SHEI = \frac{-\sum_{i=1}^{m}(P_i * lnP_i)}{lnm}$	P_i = proportion of the landscape occupied by patch type (class) i. m = number of patch types (classes) present in the landscape
	• Simpson evenness index	$SIEI = \frac{1-\sum_{i=1}^{m} P_i^2}{1-\left(\frac{1}{m}\right)}$	
	• Contagion index	$CONTAG = \left[1 + \frac{\sum_{i=1}^{m}\sum_{k=1}^{m}[P_i]\left(\frac{g_{ik}}{\sum_{k=1}^{m} g_{ik}}\right) * \left(lnP_{i\left(\frac{g_{ik}}{\sum_{k=1}^{m} g_{ik}}\right)}\right)}{2ln(m)}\right] * (100)$	Pi =proportion of the landscape occupied by patch type (class) I, g_{ik} =number of adjacencies (joins) between pixels of patch types (classes) i and k based on the double-count method. m =number of patch types (classes) present in the landscape t

2.3 Step 3: Calculation of an unplanned-settlement-index (USI)

An unplanned-settlement-index (USI) was designed by aggregating morphological aspects of unplanned areas, following the slum-index developed by Weeks et al. (2007) the concept of operationalizing several dimensions of deprived areas into a combined index. For the index selected those spatial metrics that best characterize the morphological differences between planned and unplanned areas were selected. and The spatial metrics that showed high correlating within one dimension were excluded. The selected metrics were combined into a spatial multi-criteria evaluation (SMCE) approach using Ilwis SMCE (ILWIS3.08-Open, 2012), an open source software tool, to build a composite index of several standardized, weighted indicator maps. Equal weights were assigned to the different dimensions (i.e. 1/3 size, 1/3 density and 1/3 pattern) as well as to the metrics within each dimension.

The index was the result of a weighted summation of all standardized maps of the 3 dimensions (size-density-pattern). The values of the USI maps ranged from 0 to 1, where values closer to 1 indicated higher likelihood of ‘unplannedness’ and lower values higher likelihood of ‘plannedness’. The final USI showed the value range of ‘unplannedness’ similar to the slum-index of Weeks et al. (2007). However, unlike the index developed by Weeks et al. (2007) which also used census data, the USI was built using only physical data from imagery. Once the final map was produced, the USI was aggregated by a second segmentation process extracting larger segments (HUPs at area level). In this step the segmentation parameters were optimized to extract larger areas of similar morphology. The final result was evaluated using available land use data as reference as detailed in the section above (case studies). This second aggregation level was needed to classify homogenous urban patches (neighbourhoods – area level) using the USI values as likely to planned or unplanned.

4. Results and Discussion

In the following sections the results for the three methodological steps as well as for both case studies are presented and discussed.

4.1 Results of image segmentation – extracting homogenous urban patches (HUPs)

Three different segmentation algorithm (Erdas edge detection, Definiens multi-resolution and Prabat region growing) were compared and the extracted segments were compared with a random selection of digitized building footprints. After optimization the best fitted scale parameter for the image segmentation at objected level resulted in a scale factor 15 (Figure 6), which also shows the AFI index output. Lower scale factors (e.g. 10) led to more over-segmentation while higher scale factors (e.g. 20) led to more under-segmentation (showing weaker performance to capture the object level).

As shown in Figure 6, results indicated that Definiens segmentation had a better performance in all cases but for the case of Delhi where Erdas edge detection had a similar good performance. Under-segmentation happened in all cases but the Definiens multi-resolution segmentation case for Delhi, meaning that segments were larger than the digitized roof areas. Thus, as segments of the test area tended to be larger than the average roof areas (average roof area of test sample was for Dar es Salaam 103 m^2 and Delhi 85 m^2), a complete segmentation was not achieved. This was attributed to the strong clustering of small buildings particularly in unplanned areas.

Only for the case of Delhi (multi-resolution segmentation) the average segments size was smaller than the average building size. This occurred in particular in planned areas with larger buildings and spectrally heterogeneous roof surfaces (roofs are frequently used to store objects). Unplanned areas in Delhi were also under-segmented.

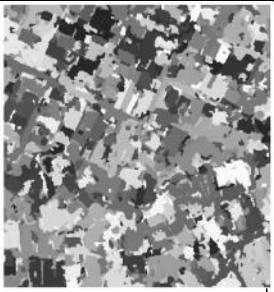
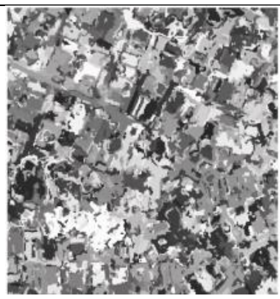
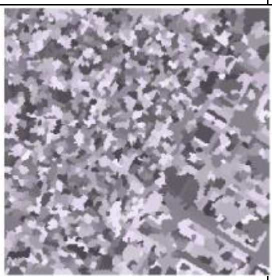
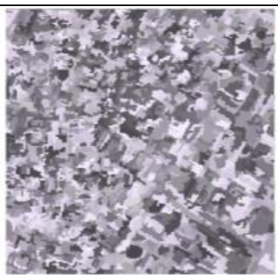
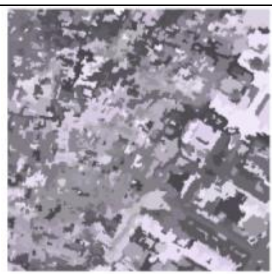
	Erdas segmentation (edge detection)	Definiens segmentation (multi-resolution)	Prabat segmentation (region growing)	Image approx. 250 by 250 meter
Subset of QuickBird images (Dar es Salaam				
AFI Dar	-3.60	-0.65	-2.90	Reference data
Subset of Ikonos image (Dehli)				
AFI Delhi	-0.15	0.15	-2.19	Reference data

Figure 6: Selection of a suitable segmentation algorithm using Area Fit Index (AFI) values

An issue preventing the achievement of a complete segmentation is illustrated in Figure 7; which shows the comparison of the average roof-coverage density in the test areas for the two cities. The roof-coverage density was extracted by using a maximum-likelihood classifier (achieving for the image of Delhi a classification accuracy of 77% and for Dar es Salaam 83%). The reason for using such a standard classification approach was to illustrate difference in the overall urban morphology of unplanned areas in the two cities. Unplanned areas in Delhi had a roof coverage density of more than 80% while the values in Dar es Salaam were lower, ranging between 50 and 75%. Thus buildings in unplanned areas of Delhi are more clustered which results in grouping of several buildings into one segment. This results in larger patches sizes and lower patch densities for the city of Delhi. While in Dar es Salaam individual buildings are captured better by the segmentation which results in smaller patches sizes of unplanned areas and higher patch densities (compared to planned areas).

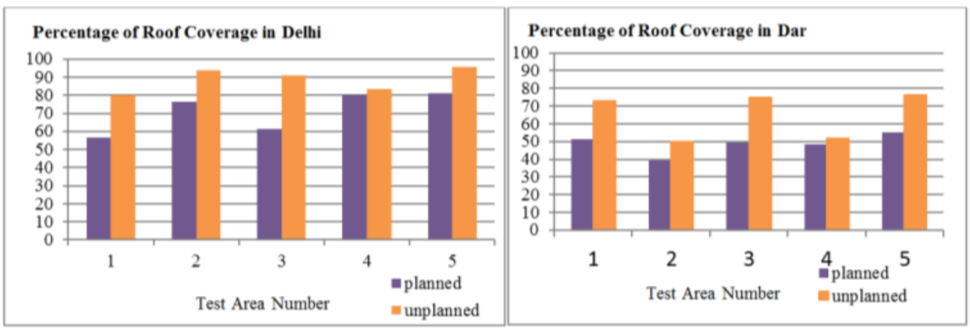


Figure 7: Roof cover percentages, Delhi (left) and Dar es Salaam *(right)*

Although image segmentation was employed to extract HUPs at object (building) level, this was showing limitations in very dense unplanned areas where extracted segments were frequently an aggregate of several building objects. This issue was more apparent for the case of Delhi due to its lower image resolution, smaller object (building) sizes, and higher density of buildings in unplanned areas. Also, Delhi presents a more heterogeneous urban morphology than Dar es Salaam, making the result of the image segmentation more complex. As a consequence, large multi-family houses (in planned areas) were over-segmented while buildings in unplanned areas were under-segmented. Although a similar problem occurred in Dar es Salaam, the impact was minimized because most planned areas (within the test areas) consisted of relatively large, detached, single family houses, the so-called Swahili house which is predominant in Dar es Salaam's residential areas. The decision was made to keep the scale parameter constant (within a city and between the cities) to ensure the spatial metrics calculated based on the segments produced comparable results.

As a result, Definiens multi-resolution segmentation was selected due to its better performance for the AFI. Despite the problems of extracting building object HUPs, spatial metrics were then used to analyse whether the HUPs of planned and unplanned areas show a quantifiable differences.

4.2 Results of spatial metrics – analysing the building HUP morphology

The final set of spatial metrics after excluding the insignificant and/or highly correlating metrics (from the initial list is Table 2) is presented in Figure 8. As a consequence of the different urban morphologies and data sources used, a different set of metrics was found suitable for each city. It was not possible to select the same spatial metrics for both cities as the metrics suitable for Dar es Salaam did not differentiate well unplanned from planned areas in Delhi. This limits the comparability of the individual metrics values, while the final value of the USI can be compared as it uses a standardisation procedure for calculating the three dimensions (size, density and pattern) within the spatial multi-criteria framework.

For the case of Dar es Salaam, the most significant metrics were: mean area (size related), patch density (density related), aggregation index (AI) and the Shannon diversity index (SDI) – the latter two both indicating pattern. Tests with AI showed close, but clearly distinguishable, values for planned and unplanned areas. Tests with SDI were controversial as results for test area 5 were not compatible with results for all other areas. In general, planned areas have higher SDI values except for test area 5, which is a rather small planned area with less regular arrangement patterns and where the mean area of building footprints is smaller than other planned areas and rather similar to unplanned areas. However, overall it can be concluded that values indicated that unplanned areas (all 5 test areas) in Dar es Salaam tended to have smaller patches, higher (segment) densities and were more aggregated and less diverse than planned areas.

For the case of Delhi, several of the metrics were highly correlated. As a result, all highly correlated metrics were excluded in order to allow modelling of the three different spatial dimensions (size/density/pattern). All remaining significant metrics were selected, with the final list including landscape division index and effective mesh size (size), patch density (density) and aggregation index, Shannon diversity index and contagion index (pattern). Mesh metric results indicated unplanned areas in Delhi tend to present smaller mesh size. Patch density results suggest density in unplanned areas was smaller than in planned ones, which can be explained by the under-segmentation that occur in unplanned areas (many clustered houses are contained in single patches). In terms of patterns, planned areas were in general more aggregated, even and orderly arranged. Only test areas 4 and 5 have unplanned areas values close to planned ones. Both areas comprise of high density planned and unplanned areas thus it makes sense the arrangement patterns have similarities.

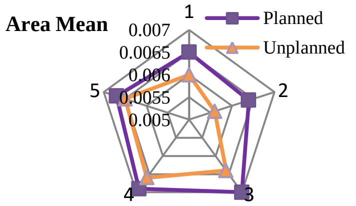
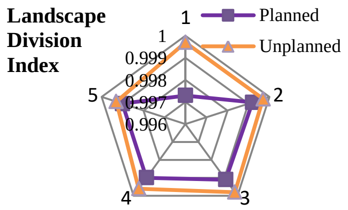
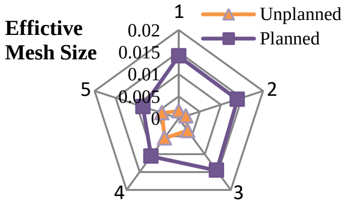
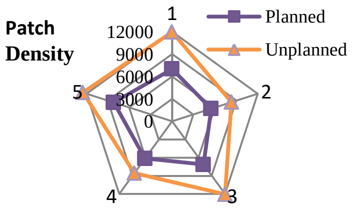
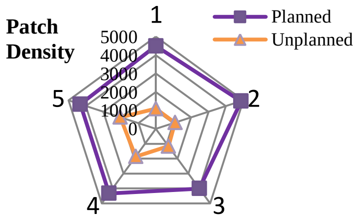
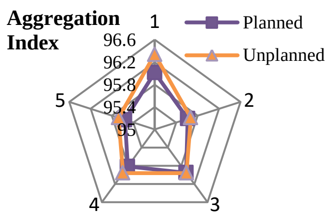
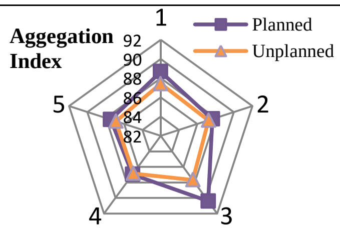
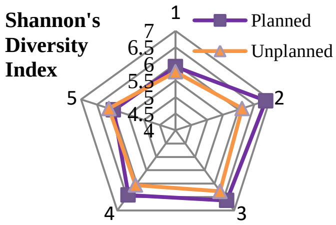
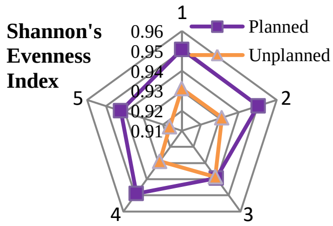
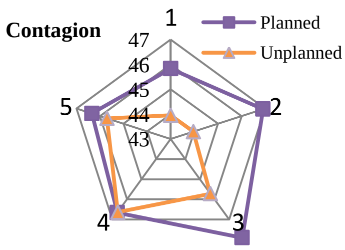
<table>
<tr><th colspan="3">Dar es Salam result of selected metrics</th><th colspan="2">Dehli result of selected metrics</th></tr>
<tr><th>Dimension</th><th>Selected metrics</th><th>Overal result of test areas Dar es Salaam</th><th>Selected metrics</th><th>Overal result of test areas Delhi</th></tr>
<tr><td>Size</td><td>Mean area</td><td></td><td>Division</td><td></td></tr>
<tr><td></td><td></td><td></td><td>Mesh</td><td></td></tr>
<tr><td>Density</td><td>Patch density</td><td></td><td>Patch Density</td><td></td></tr>
<tr><td>Pattern</td><td>Aggregation index</td><td></td><td>Aggregation index</td><td></td></tr>
<tr><td></td><td>Shannon’s diversity index</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Figure 8: Selected spatial metrics for Dar es Salaam *(left) and Delhi (right)*

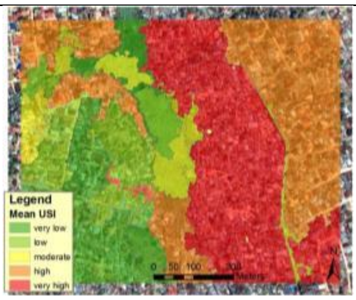
4.3. Unplanned Settlement Index

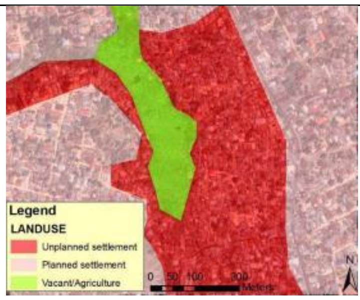
For calculating the unplanned settlement index, homogenous patches were generated by image segmentation using lager scale values to create HUPs at settlement area level (not building objects), using an approach similar to the 'land-use-regions' (M. Herold, Liu, et al., 2003) and 'homogenous urban patches' (M. Herold, Scepan, & Clarke, 2002). The scale parameter that gave the best results was 80 for Delhi and 160 for Dar es Salaam. This second segmentation step was necessary to generate HUPs at area level (Figure 4) extracting larger areas of homogenous morphology. The scale parameter was selected by using an optimization process to compare segments with ground reference data. The difference in scale between the two case study areas can be explained by the fact unplanned areas in Delhi tend to be small pockets (average HUP is 0.2 ha) while in Dar es Salaam they tend to be larger settlements (average HUP is 1.1 ha). This decision deviates from the decision made at object level where the scale parameter was kept the same (for Delhi and Dar es Salaam). Buildings in both cities are of relatively similar dimension, which is different to the HUPs at area level. The extracted segments were used to aggregate the composite maps of the USI using the mean of the metrics values. The rational for aggregation is that the concept of 'unplannedness' is normally captured on the level of an areas, e.g. neighbourhood not on an individual building object (an 'unplanned building' would refer to an illegal structure without permit which is difficult to capture with remote sensing) or smaller groups of buildings. The results were compared with existing land use data for the in total 10 assessment areas (five for each city), the results of four assessment areas are presented in Figure 9.

For a visual inspection the results of the USI were classified in 5 different classes (very high, high, moderate, low and very low) using equal frequency (Figure 9 displays four example assessment areas). In general, 'very high' and 'high' USI values indicate areas that showed characteristics of unplanned areas. As shown in Figure 9, 'high' or 'very high' USI values visually matched the areas of unplanned settlements in the reference data reasonably well. For assessment area 1 in Dar es Salaam, the main unplanned area fell into the category 'very high'. In this case, some small areas along the edges of unplanned areas were not detected by the USI. Also the planned area in the East fell into the category 'high' because of the presence of high density settlement. The assessment area 2 of Dar es Salaam showed a similar result in which the large unplanned area fell mainly into the category 'high' while other parts were within the category 'very high' and 'moderate'. Smaller unplanned areas along the edge of the settlement fell into various categories. Some small areas within planned settlements had 'high' values mainly due to high built-up densities.

Similarly, the USI values in Delhi also indicate unplanned areas. Yet, the results for some specific areas are rather complex, such as assessment area 1 for Delhi. In this area, two unplanned areas in the Centre and South (mainly slum areas) were classified as 'high', while the large unplanned area in the North-East fell into various categories (ranging from low to very high). This area is a typical example of unplanned development (unauthorized colony) but not a slum, meaning that the built-up structure has larger roof areas and lower density. Thus such areas are morphologically rather similar to high-density planned areas. Another issue in assessment area 1 is some small segments in planned areas have '(very) high' USI values. This showed that the smaller scale used for the segmentation of the Delhi image (which was necessary extracting small slum pockets) caused problems (outliers). The results for the assessment area 2 of Delhi showed the location of the unplanned areas more clearly within the category 'very high', with the exception of the two smaller slum pockets (of 900 and 1800 m^2) that were within the class 'low' and 'moderate'. This is due to scale issues in the extraction phase (segmentation). A smaller scale parameter would be necessary to extract such small settlements.

	Unplanned Settlement Index (USI)	Land use reference data
Assessment Area 1 Dar		





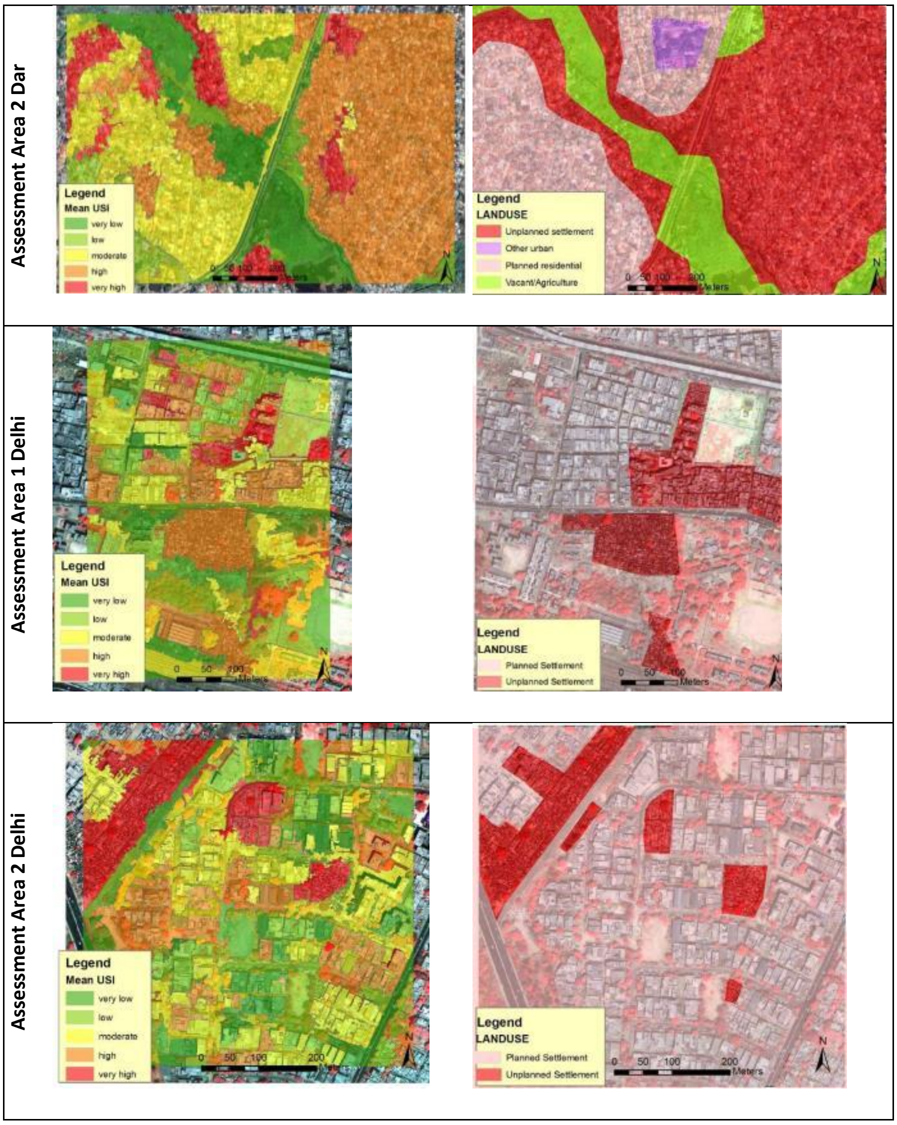


Figure 9: Unplanned settlement index for homogenous patches

For a quantitative comparison of the extracted USI values with the reference data 300 samples within the 5 assessment were selected for each city. With the help of boxplots (Figure 10 and 11) the mean and deviation of the samples were plotted for the planned and unplanned areas (extracted from the reference data). For both cities the samples showed higher values for the unplanned areas as compared to the planned areas. Interestingly in Dar es Salaam a number of outliers which are in the reference unplanned had low USI values (Figure 10). These are example of relatively low density developments (typical for the outskirts of Dar es Salaam). While in Delhi the outliers are found within the group of planned areas showing high USI values (Figure 11). These areas are of high built-up density and show very similar morphological characteristics than unplanned areas.

In order to assess the utility of the USI values to indicated unplanned areas a threshold was set to classify the HUPs into a binary map of planned versus unplanned. USI values of more than 0.5 were classified as unplanned and 0.5 and below as planned. When comparing the cross tables for the 300 samples of both locations it can be stated that the agreement of reference and classified sampled locations was 73% for Delhi and 75% for Dar es Salaam. Analysing the results in more detail, in Dar es Salaam out of 151 planned areas 58 have been classified as unplanned, mostly rather high density planned areas. While only 16 out of 149 unplanned areas went into the class planned (several of them are displayed as outliers in Figure 10). A similar trend can be observed in Delhi 75 planned areas have been classified as unplanned. Most of these areas are again highly densely built-up. Also in Delhi the majority of unplanned areas have been classified correctly (only 6 have been classified as planned).

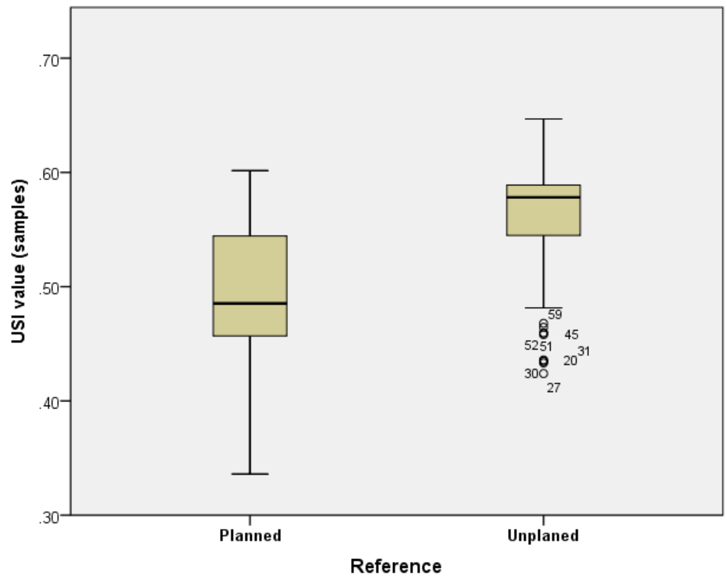


Figure 10: Comparison of reference data with USI values for Dar es Salaam

Table 4: Comparison of reference data with USI values for Dar es Salaam *(overall accuracy 75%)*

		Reference	
		Planned	Unplanned
		Count	Count
Classified Data	Planned	**93**	16
	Unplanned	58	**133**

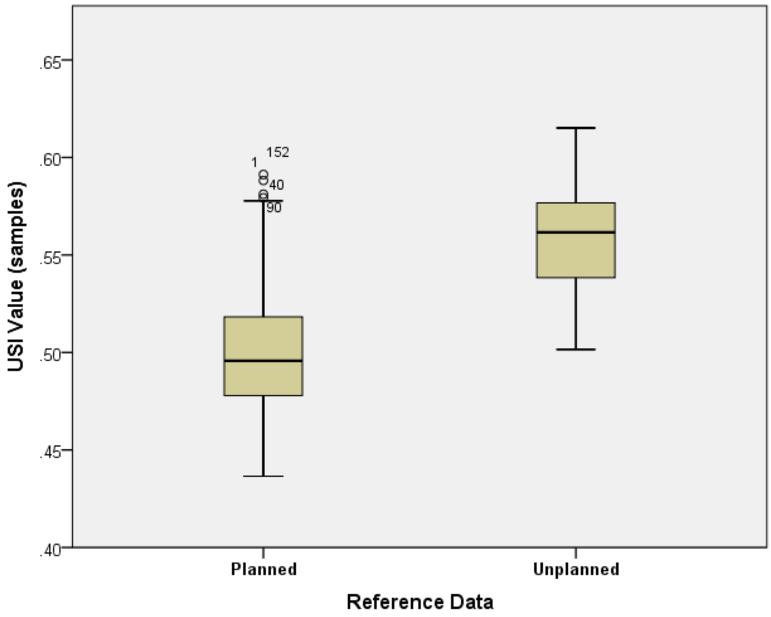


Figure 11: Comparison of reference data with USI values for Delhi

Table 5: Comparison of reference data with USI values for Delhi (overall accuracy 73%)

		Reference	
		Planned	Unplanned
		Count	Count
Classified Data	Planned	149	6
	Unplanned	75	70

General issues found with USI were: smaller clusters were often not detected; high density planned settlement produced erroneous high USI values; unplanned areas with lower densities and larger buildings sizes had often low USI values and were not detected. These limitations are a consequence of employing morphological only criteria. Still, the USI could distinguish between planned and unplanned areas with an accuracy of more than 70% in both cities.

5. Conclusions

The presented research focused on quantifying differences on morphological features of unplanned areas in terms of size, density and pattern. The methodology used a rather simple image processing technique, namely segmentation, to generate input data for performing an analysis with spatial metrics. Unlike object based feature extraction, image segmentation is a processing technique which does not require customizing rules and allows processing of larger image data sets relatively quickly. This research has demonstrated that segmented images can be analysed by spatial metrics producing meaningful information about the urban morphology of cities in developing countries. Yet, the application of this methodology can be hindered by the fact that planned areas presenting similar morphological characteristics may fall within the same class as unplanned areas. This research also showed that the two rather different cities have differences in their morphological features. The city of Dar es Salaam shows in general lower built-up densities in unplanned areas than unplanned areas in Delhi. In Delhi buildings of unplanned areas are commonly very clustered that even a visual delineation is difficult using images of 1 m spatial resolution, while building outlines in Dar es Salaam can be in general better identified. The high clustering of buildings in Delhi caused limitations to extract object level information using image segmentation.

Although the segmentation did not achieve a complete segmentation (extracting accurately building objects), the obtained results provide enough quantifiable differences between planned and unplanned areas. These differences were analysed by a set of spatial metrics and combined into an unplanned settlement index composed by all three dimensions of morphological aspects of 'unplannedness' (size, density and pattern). This composite index was aggregated to settlement level HUPs using image segmentation to extract 'homogenous neighbourhoods', thereby avoiding a manual delineation as done by M. Herold, Scepan, & Clarke (M. Herold, et al., 2002) also used in Herold and Liu et al. (2003). The case studies demonstrate context is important and parameter settings cannot be used universally, as the set of spatial metrics required adjustment for different cities and types of VHR imagery in order to best extract the morphology. Ultimately, for both case studies the set of metrics that indicated best planned versus unplanned were quite different, even presenting opposite values. The main limitation of the approach is segmentation in unplanned areas did not succeed in extracting single roof objects well. Thus, a possible next step would be to repeat the analysis with even higher resolution imagery (below 0.5 meter) in an attempt to achieve a more complete segmentation.

This study also demonstrated that spatial metrics can successfully support the detection of unplanned areas, and that a set of local meaningful metrics has the potential to identify the areas with unplanned morphological characteristics. Spatial metrics were able to quantify differences between planned and unplanned areas in segmented images using size, density and pattern, which were identified as the three key spatial dimensions of 'unplannedness'. The presented unplanned settlement index, which combined the three dimensions into an index coupled with a spatial multi-criteria framework, identifies areas with morphological characteristics of 'unplannedness' with relative success (of more than 70% for both cities). Thus, the proposed methodology provides a step towards a low cost and effective method for mapping unplanned areas.

References

Amato, P. W. (1970). A Comparison: Population Densities, Land Values and Socioeconomic Class in Four Latin American Cities. *Land Economics, 46*, 447-455.

Barros Filho, M., & Sobreira, F. (2005). Assessing texture pattern in slum across scales an unsupervised approach., *CASA Working Papers Series*.

Barros, J. X. (2004). *Urban Growth in Latin American Cities Exploring urban dynamics through agent-based simulation.* University College London.

Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R. V., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan India: the added value of remote sensing data for analyzing sub - standard residential areas. *International Journal of Applied Earth Observation and Geoinformation, 12*, 359-374.

Busgeeth, K., Brits, A., & Whisken, J. (2008). Potential Application of Remote Sensing in Monitoring Informal Settlements in Developing Countries where Complimentary Data Dos not Exist. *Planning Africa Conference*. Sandton Convention Centre, Johannesburg, South Africa, April 14-16, 2008.

Cheng, H. D., Jiang, X. H., Sung, Y., & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition, 34(12)*, 2259-2281.

Chin, N. N. G. (2006). *Spatial Analysis and the measurement of urban sprawl.* University College London, University of London.

Frohn, R. C. (2006). The use of landscape pattern metrics in remote sensing image classification. *International Journal of Remote Sensing, 27(10)*, 2025–2032.

Herold, M., Couclelis, H., & Clarke, K. C. (2005). The role of spatial metrics in the analysis and modeling of urban land use change. *Computers, Environment and Urban Systems, 29(4)*, 369–399.

Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment, 86(3)*, 286–302.

Herold, M., Liu, X. H., & Clarke, K. C. (2003). Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering and Remote Sensing, 69(9)*, 991–1001.

Herold, M., Scepan, J., & Clarke, K. (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A, 34*, 1443–1458.

ILWIS3.08-Open (2012). 52°North, http://52north.org/communities/ilwis/about.

Kit, O., Lüdeke, M., & Reckien, D. (2012). Texture-based identification of urban slums in Hyderabad, India using remote sensing data. *Applied Geography, 32*, 660-667.

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems, 36*, 154-163.

Kombe, W. J. (2005). Land use dynamics in peri-urban aeas and their implications on the urban growth and form: The case of Dar es Salaam, Tanzania. *Habitat International, 29(1)*, 113-135.

Kostof, S. (1991). *The City Shaped; Urban patterns and Meanings through history*. London: Thames & Hudson.

Lemma, T., Sliuzas, R. V., & Kuffer, M. (2006). participatory approach to monitoring slum conditions : an example from Ethiopia. *In: Participatory learning and action (2006)54. Special issue on: Mapping for change : practice, technologies and communication : proceedings of the international conference on participatory spatial information management and communication PGIS '05 - KCCT, Nairobi, Kenya, 7-10 Sept 2005. / ed. by Giacomo Rambaldi; Jon Corbett; Mike McCall; ...[et al.] London : IEED, 2005. pp. 54-58*.

Liu, X. H., Clarke, K., & Herold, M. (2006). Population density and image texture: A comparison study. *Photogrammetric Engineering and Remote Sensing, 72*, 187-196.

Lucieer, A. (2004). *Uncertainties in Segmentation and their Visualisation.* University of Utrecht, The Netherlands.

McGarigal, K., Cushman, S. A., Neel, M. C., & Ene, E. (2002). FRAGSTATS v3: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html.

Peng, J., Wang, Y., Zhang, Y., Wua, J., Li, W., & Li, Y. (2010). Evaluating the effectiveness of landscape metrics in quantifying spatial patterns. *Ecological Indicators, 10(2)*, 217–223.

Pham, H. M., Yamaguchi, Y., & Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. *Landscape and Urban Planning, 100*, 223-230.

Rakodi, C., & Lloyd-Jones, T. (2002). *Urban livelihoods. A people-centred approach to reducing poverty* (1st ed. ed.). London: Earthscan.

Risbud, N. (2002). Policies for Tenure security in Delhi. In A. Durand-Lasserve & R. Royston (Eds.), *Holding Their Ground-Secure Land tenure for the Urban poor in developing countries* pp. 59-74). London: Earthscan.

Rüther, H., Martine, H. M., & Mtalo, E. G. (2002). Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas. *ISPRS Journal of Photogrammetry and Remote Sensing, 56*, 269-282.

Saura, S., & Castro, S. (2007). Scaling functions for landscape pattern metrics derived from remotely sensed data: Are their subpixel estimates really accurate? *ISPRS Journal of Photogrammetry & Remote Sensing, 62(3)*, 201–216.

Schwarz, N. (2010). Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning, 96(1)*, 29-47.

Sliuzas, R. V., Kerle, N., & Kuffer, M. (2008). Object - oriented mapping of urban poverty and deprivation. In E. A. o. R. S. L. (EARSeL) (Ed.), *Proceedings of the 4th EARSeL workshop on remote sensing for developing countries in conjunction with GISDECO 8*. Istanbul, Turkey.

Stewart, G., & Kuffer, M. (2007). Where are the poor? A disaggregation approach of mapping urban poverty *12th international conference on urban planning and spatial development in the information society and 2nd international Vienna real estate conference : REAL CORP 2007*. Vienna.

Taubenböck, H., & Kraff, N. J. (2013). The physical face of slums: a structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment, 2013*, 1-24.

Taubenböck, H., Wegmann, M., Roth, A., Mehl, H., & Dech, S. (2009). Urbanization in India – Spatiotemporal analysis using remote sensing data. *Computers, Environment and Urban Systems, 33(3)*, 179–188.

Turkstra, J., & Raithelhuber, M. (2004). Urban slum monitoring. *Proceedings of the 24th Annual Esri International User Conference*. San Diego, California,.

UN-Habitat (2003). The Challenge of Slums - Global Report on Human Settlements London and Sterling, VA.

UN-Habitat (2008). State of the world's cities 2010/2011. Nairobi.

Wang, Z., Jensen, J. R., & Im, J. (2010). An automatic region-based image segmentation algorithm for remote sensing applications. *Environmental Modelling & Software, 25(10)*, 1149-1165.

Weeks, J., Hill, A., Stow, D., Getis, A., & Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana *GeoJournal, 69(1-2)*, 9-22.

Welch, R. (1982). Spatial resolution requirements for urban studies. *International Journal of Remote Sensing, 3(2)*, 139-146.

Yu, Z. J., & Ng, C. N. (2007). Spatial and temporal dynamics of urban sprawl along two urban-rural transects: A case study of Guangzhou, China *Landscape and Urban Planning, 79(1)*, 96-109.

Zhang, H., Fritts, J. E., & Goldman, S. A. (2008). Image Segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding, 110*, 260-280.