Context Based Detection of Urban Land Use Zones

*J.N. Louw ¹, G. Sithole ²,

¹Research and Development, ComputaMaps, South Africa, jlouw@computamaps.com

²Geomatics Division, University of Cape Town, South Africa, George.Sithole@uct.ac.za

*Research undertaken at the University of Cape Town

Abstract

In this paper an automated land use classification system is proposed, based on the context of an urban scene. Land use identification is an important step toward improving object extraction in an urban scene. We characterize urban context by analyzing different land use scenes in an aerial image dataset of the Cape Town region. A set of high-level features (e.g. building density, road-building distance etc) is measured off manually labelled building and road objects in these scenes. A multivariate statistical visualization test shows that land use scenes can be discriminated based solely on the high-level feature set. A feature selection routine is used to compute an optimum subset of features that causes maximum separation of land use classes. The subset gives an indication as to which features are important in defining urban context. An automated land use segmentation and classification routine is then proposed. The land use segmentation routine works by extracting 'road block' regions from bottom-up road classification data using a novel technique. Blocks are classified to a land use type based on high-level features of bottom-up object classification data. Experiments conducted on 3 aerial image test datasets generated land use block classification accuracies of over 75%. This shows that a recognition of context can be achieved through incomplete bottom-up data, which is an encouraging finding within the framework of aerial image understanding.
1 Introduction and Related Work

Because of the increased pace of urbanization in recent years, automated mapping of the urban environment has become a topical research subject. Numerous environmental and socio-economic decision making applications benefit from automated urban mapping, such as urban sprawl analysis, transportation infrastructure management, and architectural evaluation. Remote sensing observation data in the form of aerial or space-borne imagery has shown to be an effective data source for urban mapping.

Traditionally, pixel-based image classification methods have been used to extract information from imagery. These methods work by associating each image pixel with an object class based on the spectral properties of that pixel. In the last two decades high spatial resolution imagery has become readily available, resulting in increased intraclass spectral variation and decreased interclass variation (Aksoy and Akçay, 2005). This is especially evident in urban scenes where different man-made objects are made up of materials with similar spectral signatures (Aplin and Smith, 2008). A modern trend is to classify groups of neighbouring homogeneous pixels, to avoid mis-classification of individual pixels. This is known as Object Based Image Analysis (OBIA) (Benz et al., 2004; Aplin and Smith, 2008; Blaschke, 2009). In addition to spectral properties, shape and texture measurements of the pixel groups have been shown to be effective in improving classification results.

The performance of both the former 'pixel-based' and the later 'object-based' image classification techniques is inherently limited by the total amount of available sensory content, whether it be colour, shape or texture. Many different object types in urban scenes can be similar in shape and colour (e.g. road and pavement), and thus even with object-based techniques, misclassification of regions is possible.

Human visual perception relies to a large extent on a recognition of context. Sensory information is used to build gradually and select from an internal repertoire of 'perceptual hypotheses' (Gregory, 1970). When detecting a building in an industrial scene, for instance, decisions are heavily influenced by an inherent, pre-built understanding of the general description of an industrial settlement. Researchers in Content Based Image Retrieval (CBIR) and remote sensing fields have developed Image Understanding (IU) systems to emulate this perception. Matsuyama and Hwang (1990) presents a general framework for aerial IU. The concept is that a complete and ide-
alized description of a scene is constructed, even if it is partially depicted by initially observed image features. This idealized description is induced by a contextual model. A contextual model is a template for what a scene should look like. In more detail, a two step bottom-up, top-down analysis is performed. Bottom-up analysis refers to object extraction based on image features. An IU system uses bottom-up results, which are normally incomplete, to choose one of a predefined set of contextual models. This can be referred to as 'scene-matching'. In the top-down phase, the chosen contextual model is used to improve the initial bottom-up results. The final scene description is consistent with the contextual model, so that a high-quality semantic depiction is produced.

Liu, Guo and Kelly (2008) presents a top-down urban scene object extraction methodology based on high-level fuzzy spatial relation features (*surround, invade, meet* etc). These features are formulated for areal and non-overlapping objects such as those found in a typical image segmentation. A case study shows that the inclusion of a few of these spatial relation features improves the extraction of car objects from a high resolution image of an urban scene.

Porway et al. (2008) proposes a contextual model as a hierarchical model of an urban scene. The model starts at scene level (the entire scene). Scene level decomposes into groups of objects, such as blocks of buildings or rows of cars. Object groups decompose into single objects, such as a building, which decomposes further into parts and primitives. To learn the model, various spatial relation features of objects, such as *relative position, relative orientation, percentage overlap* etc, are extracted from manually labeled sample images. In a top-down analysis, the bottom-up results are improved in a Markov Random Field (MRF) framework.

An issue with the IU systems in (Liu, Guo and Kelly, 2008; Porway et al., 2008) is that a contextual model for only one scene type is defined. Urban scenes while similar are significantly different due to land use planning policies. There are significant structural differences between e.g. CBD, low cost housing development, industrial etc. A generalized contextual model requires a characterization of different land use types. Before performing top-down analysis, appropriate land use context should be chosen through scene-matching. In other words, automated land use classification is required for a full aerial image understanding system.

A study related to South African urban context can be seen in (Busgeeth, Brits and Whisken, 2008). A hierarchical, rule-based land use classification
typology is proposed. Quickbird imagery of the Soweto region in South Africa was analysed. Informal settlements were distinguished from formal settlements based on the following high-level features: *Average building size, building size variety, formalized / informalized street pattern, and tarred / gravel roads.* These features were, however, manually extracted from the imagery. For automation purposes, *formalized / informalized street pattern*, for instance, needs to be defined. Furthermore, only four features were used based on expert knowledge. Ideally, from all possible discriminatory features, an optimum subset needs to be established that causes greatest land use separation.

In this paper an urban contextual model is proposed that characterizes different land use types (refer to figure 1). Similarly to the above studies, we propose context in the form of a set of high-level features (depicted as \([f_1, f_2, \ldots, f_n]\) in the figure). These are spatial relations and geometric properties of objects (*average road-building distance, building density, average building size* etc). A unique feature space should exist for each land use class. To train the model, these features were extracted from manually labelled images of different land use types. Multivariate statistical visualization tools are used to assess whether land use classes can be discriminated based on the high-level feature set. A feature selection technique is then used to establish a subset of high-level features that optimally separates land-use types. This subset can be regarded as a 'scene descriptor'. It gives an indication as to which features are important in defining urban context.

An automated land use classification routine is also proposed (refer to figure 3). The system works by initially performing bottom-up object classification. Land use regions are then segmented and classified based on these bottom-up results. Land use segmentation is accomplished by extracting 'road block' regions from the incomplete bottom-up road data using a novel routine. A 'road block' is that region within a closed loop of road segments. Each block is classified to a land use type based on a set of high-level features extracted from the bottom-up objects within that block.

This paper is organized in 4 sections. Section 2 presents a detailed description of the proposed contextual model. Section 3 addresses the automated land use classification approach, along with results obtained from experimentation on 3 different datasets, and a discussion of the proposed system. In section 4 conclusions are drawn on the study.
2 An Urban Contextual Model

To train our contextual model, sample scenes were selected for each of the following land use types: medium-density residential, high-density residential, informal, industrial, and commercial. The samples were chosen from well distributed spatial locations and different suburbs in a 2007 high resolution RGB aerial imagery dataset of the greater Cape Town region. Building and road objects were manually labelled in each scene sample. The following set of high-level features are proposed as an initial set for urban contextual analysis.

1. **Building area**: The 2D area of a building footprint.

2. **Building compactness**: Building compactness = \( \frac{4 \times A_{\text{building}}}{P_{\text{building}}} \), where \( A_{\text{building}} \) = building area and \( P_{\text{building}} \) = building perimeter.

3. **Road width**: The shortest distance between two road edges (an edge is defined at a road curb).

4. **Road-building distance**: The shortest distance between a building polygon and the nearest road edge. The distances between all buildings in a scene and their corresponding nearest road edge are measured.

5. **Building-building distance**: The shortest distance between a building polygon and the nearest building polygon. The distances between all buildings in a scene and their corresponding nearest building are measured.
6. **Building density**: Building density $= \frac{A_{\text{buildings}}}{A_{\text{scene}}}$, where $A_{\text{buildings}}$ = sum of all building areas in a scene and $A_{\text{scene}}$ = total scene area.

7. **road-building distance to road width**: Ratio of road-building distance to road width.

8. **building-building distance to building area**: Ratio of building-building distance to building area.

The methodology for extracting these high-level features from the sample scenes is as follows. For a feature such as building area, the areas of all buildings in a scene are measured. Both the mean and the standard deviation of these area measurements are taken into account for further analysis. For other features such as building density, one measurement exists for each scene. Taking into account standard deviation, we end up with a total of 13 high-level features. This complete set was extracted from each sample scene. Thus a 13 dimensional feature space exists. This can be expressed as an observation matrix $X(i, j)$, where rows $i$ correspond to observations (scene samples), and columns $j$ to features.

### 2.1 Multivariate Visualization

The *glyph* plot was used to assess whether land use classes can be separated in the high-level feature space, and to determine which high-level features are significant in separating classes. The plot was generated using the Matlab statistical toolbox, from the observation matrix $X$. The glyph plot represents each observation as a star, where the $j$th spoke in a star is proportional in length to the $j$th feature of that observation. Similarities in observation data can be established based on the shapes of the stars.

The glyph plot results are shown in figure 2. The y axis denotes land use type and the x axis denotes the various sample scenes that were observed. Thus each star in the plot corresponds to a single sample scene (about 5 sample scenes were observed for each land use scene). In the figure we can see that scene samples of the same land use exhibit a similarly shaped star. This shows that a unique signature exists for a particular scene type, which proves that land use types can be discriminated based solely on high-level features. The plot also shows which combinations of features are significant in separating land-use classes. Each star is made up of 13 spokes, which correspond to the 13 high-level features that were measured. On the right
Figure 2: Glyph plot of sample scenes. Consistencies can be observed in the shapes of stars of the same land-use class. Thus land-use classes can be discriminated based on the high-level feature set.

Hand side of the plot a map shows the positions of the features on a star. If we observe the commercial plots for instance, we can see a prominent fan at the 3 o’clock position. This tells us that the \( \text{stdev building compactness} \) and \( \text{mean road width} \) feature pair are prominent in characterizing commercial scenes. This same feature pair is prominent in industrial scenes, but not prominent in all the other scene types, which means that this feature pair can be used to separate industrial / commercial from the rest. Informal settlements are strongly characterized by \( \text{mean building compactness} \). The \( \text{stdev building area} \) feature is fairly consistent in characterizing scene types. This feature (spoke at the 1 o’clock position) is prominent especially for the medium density residential, industrial and commercial classes, but not prominent for the informal class. This is because informal buildings (shacks) are normally consistent in size.

Thus the glyph plot gives an indication as to what significantly characterizes a particular land use type. This is useful to identify or single out a
particular land use with respect to all the others, which can find application in land use detection systems (informal settlement monitoring), or hierarchical land use classification strategies.

2.2 High-level feature subset

The proposed set of 13 features are potentially redundant. It is desirable to establish a minimum subset of features that optimally separates land use classes. To compute an optimum subset, an SVM feature ranking estimate based on a Random Forest (RF), proposed by Chen and Lin (2006), was carried out using the LIBSVM (Chang and Lin, 2001) software package. The procedure works by computing the difference in RF accuracy measures after elimination of features. The procedure was applied to the same observation matrix $X$ that was used for the multivariate visualization. Table 1 lists the feature importance results.

From these results we may choose the first 6 features to be a reasonable subset. The final feature subset is as follows:

\[
< \text{mean road width, stdev building area, mean building area, mean building compactness, stdev road width, stdev building-building distance} >
\]

The results are consistent with the glyph plot results (analysis in section 2.1). Road width is the most highly ranked. This makes sense as there are particular urban planning road width specifications for different land use zones. Thus consistent road width values should exist for scenes of the same land use. Mean building area is highly ranked probably because building type and thus building size is more or less consistent according to land use. Informal settlements buildings, for instance, are on average significantly smaller than buildings in most other zones. Residential buildings are normally slightly larger because of the income bracket difference. Industrial and commercial buildings can be large due to their functionality.

Land use functionality may also be correlated with building compactness, which causes its high ranking. Informal and high-density residential buildings are generally compact because of their size, low cost method of construction, and basic usage. Medium density residential buildings are less compact due to their more complex construction, to accommodate aesthetics and luxury. Ironically, building density (ratio of total building area to total scene area) is not highly ranked, which means that, contrary to common terminology ('medium density residential'), building density is not a major discriminating feature. This may be because high-density suburbs seem 'high density' due to
Table 1: SVM high-level feature importance results

<table>
<thead>
<tr>
<th>Feature</th>
<th>Feature importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean road width</td>
<td>4.75</td>
</tr>
<tr>
<td>Stdev building area</td>
<td>4.09</td>
</tr>
<tr>
<td>Mean building area</td>
<td>3.7</td>
</tr>
<tr>
<td>Mean building compactness</td>
<td>2.16</td>
</tr>
<tr>
<td>Stdev road width</td>
<td>2.1</td>
</tr>
<tr>
<td>Stdev building-building distance</td>
<td>2.06</td>
</tr>
<tr>
<td>Stdev building compactness</td>
<td>1.88</td>
</tr>
<tr>
<td>Building density</td>
<td>1.86</td>
</tr>
<tr>
<td>Mean building-building distance</td>
<td>1.032</td>
</tr>
<tr>
<td>Mean road-building distance / mean road width</td>
<td>0.94</td>
</tr>
<tr>
<td>Mean building-building distance / mean building area</td>
<td>0.80</td>
</tr>
<tr>
<td>Stdev road-building distance</td>
<td>0.65</td>
</tr>
<tr>
<td>Mean road-building distance</td>
<td>0.54</td>
</tr>
</tbody>
</table>

the grid arrangement and close proximity of buildings. However, the average building size is also smaller so building density can be similar.

The subset feature space can be regarded as a scene descriptor, or a contextual model within the image understanding framework. It can be considered as general parameter constraints to the ideal description of particular scene types. Thus the feature space can be utilized to constrain and improve bottom-up object classification results (top-down analysis). For example, if we wish to improve object classification results of a scene that was classified as industrial, we would use the industrial feature space as a constraint to produce a higher-quality final scene description.
3 Automated Land-Use Classification

The concept of image understanding is that higher quality object classification results will be produced if the semantics of a scene is recognized. An urban scene can decompose into land use regions, which further decomposes into urban objects (see Porway et al., 2008). Instead of analyzing each object independently within a scene (as with standard image classifiers), groups of objects are analyzed within sub-scenes of the scene. With the knowledge of land use (context) for a sub-scene, top-down methods can be employed to produce a high quality scene description for that sub-scene. Each sub-scene is treated differently and independently according to its land use type. Thus a required stage in an urban aerial image understanding system is automated land use classification.

We propose a segmentation and classification of land use regions, based on bottom-up object features. An automated segmentation of land use regions is problematic, as there are no well-defined generic land use border discriminatory features. The proposed approach is to segment 'block' regions (those regions that fall within a closed loop of road features). Figure 3 illustrates the concept of the proposed approach. Bottom-up object classification is initially performed. Blocks are then extracted based on the road extraction results using a novel routine described below. Each block is then classified to a land use type based on bottom-up features of objects within that block. The assumption with this approach is that an individual block falls within a closed loop of roads. Thus the approach will have limited effectiveness for scenes where this isn’t the case. This model can be extended to produce maps desirable for certain land use applications, by merging blocks with the same land-use.

The novel routine to extract block regions from incomplete bottom-up road data will now be described. By 'incomplete road data' we refer to a typical automated road extraction result that may contain errors. Although the routine can handle 'incomplete' data, the data must still be of high enough quality for successful block extraction. Thus the routine will be ineffective for scenes with roads that are not easy to locate (e.g. gravel roads).

The routine works by extracting those regions from the road data that have a low road point density. The assumption is that even with incomplete road data, block regions will be sparsely populated with road points. The algorithm works as follows:
Figure 3: Automated land-use classification concept. 1) Bottom-up object segmentation and classification is performed on a raw image. 2) The road classification results are considered for further analysis. 3) Block regions are extracted from the road data. 4) Each block is classified to a land use based on high-level features of the object classification results in step 1).

1. Road points are obtained from the initial bottom-up road classification.

2. A Delaunay triangulation (Lattuada and Raper, 1996) is generated over these points.

3. All 'long edges' in the triangulation are identified. A 'long edge' is any edge of a triangle in the triangulation that has a length above a certain threshold.

4. In a connected component analysis, connect triangles under following criterion. Two triangles are connected if they are:

   (a) adjacent, and
   (b) share a 'long' edge.

5. We then end up with a set of regions (the connected triangle components), which we can refer to as 'block regions'. Each block region is made up of a set of triangles. Thus each block region is essentially a set of points (the triangle vertices). To obtain final block estimates, generate a convex hull around the points of each block region. The convex hull of a set of points $P$ is the minimum convex set containing $P$ (Brown, 1979). These convex hulls are the final block estimates.
3.1 Experimentation

To test the effectiveness of the proposed approach, 3 sample scenes were chosen from the same Cape Town high resolution aerial imagery used to construct the contextual model (section 2). The 3 scenes are of the Landsdowne, N1 city and Elfindale suburbs. The results of just the Landsdowne dataset will be discussed due to space limitations. The Landsdowne scene is shown in figure 4 a). It consists of an industrial area and a medium density residential area.

Bottom-up object classification

Building, road, vegetation and pavement (vegetation and pavement included to provide a more reliable overall scene classification) objects were classified by object-oriented segmentation and classification in the Definiens eCognition software package (Benz et al., 2004). The segmentation routine was developed by Baatz and Schape (2000). Classification was performed using various colour, shape, and textural features and eCognition’s fuzzy Nearest Neighbour classifier.

Object classification results for the Landsdowne dataset are shown in figure 4 b). By visual analysis, building and road features are in general correctly extracted. There are, however classification errors. Buildings tend to be merged due to misclassified vegetation areas, and certain building segments have been misclassified as road. This is understandable since a roof top segment can be identical to a road segment in terms of colour, shape and texture. It is this limited competence of current automated object classifiers that has motivated this research. If context (land use) can be recognized based on this incomplete data, top-down methods can be employed to improve the results. For instance, considering the Landsdowne scene in figure 4 b), through scene-matching we gain the knowledge that the portion of the scene to the upper right is industrial. We can then use external industrial context knowledge (such as the industrial feature space proposed in section 2.2) to improve object classification results in that area. A building segment classified as road, for instance, should cause statistical inconsistencies in the feature space, and would thus be changed to building. The final scene description should cause as much consistency as possible with the model, and will thus be semantically correct.
Figure 4: Landsdowne dataset. a) The scene consists of an industrial region in the upper right, and a residential region elsewhere. The two regions can be separated by the road network. b) Bottom-up object classification results: the scene is incomplete as there are misclassified segments.

Figure 5: Landsdowne dataset. a) Block segmentation results: block regions are more or less delineated. The industrial and residential regions are clearly separated by the segmentation, which is the desired result since we’re interested in land-use classification. b) Block classification results: almost all residential blocks have been correctly classified, and a few industrial blocks have been misclassified as residential. A few regions have not been classified due to the incomplete block segmentation.
Block extraction

The block extraction routine was then used to estimate block regions from the incomplete road data. The original scenes were segmented again in eCognition based on the block estimates. Results for the Landsdowne scene are shown in figure 5 a). The block estimates consist of a set of rounded polygons, with large gaps in between. The gaps are caused by the block extraction algorithm’s estimate of road regions. The results demonstrate the effectiveness of the block extraction algorithm in performing land use segmentation.

Block classification

The block segments now need to be classified to a land use type based on high-level features extracted from the incomplete bottom-up object classification results. The first step was to formulate a ruleset in eCognition to automatically eliminate the gaps in between the block regions. This ruleset simply classifies each segment as a 'block' if its compactness is above a certain threshold.

High-level features similar to those in section 2 were then formulated in eCognition. They were formulated for each block segment based on the bottom-up data within that block, i.e., to compute e.g. the mean building area feature for a given block, the average building area within that block region is computed. The high-level features were measured off representative block samples in all three datasets. Based on these measurements, eCognition’s ‘Feature Space Optimization’ tool was used to compute a subset of features that causes maximum land use class separation. The feature subset results are as follows.

\(<\text{Mean building area, Stdev building area, Stdev road-building distance, Stdev building compactness, Stdev building-building distance}>\>

These results are fairly similar to the feature importance results in table 1 (which were generated from manual object extraction), with the exception of stdev road-building distance. This could be because mis-classification of building objects often occur adjacent to roads, resulting in inconsistent road-building distances. Nevertheless, the feature subset similarities are a promising finding, as it means that current state-of-the-art automated urban object classifiers can produce results of high-enough quality to perform true semantic analysis.

Based on this high-level feature subset, blocks were classified to a land
use. Note that the objective with the block classification is simply to demonstrate that land use scenes can be discriminated based on high-level features, and not to test the performance of a classification algorithm. Thus a simple crisp classification rule set was formulated in eCognition. The rule-set works as follows: For a block region \( b \) to classify to a land-use type \( l \), based on a high-level feature subset \( (f_1, f_2, ..., f_n) \), high-level feature values \( b_{f_1}, b_{f_2}, ..., b_{f_n} \) are measured off \( b \). The rule-set has the following form:

\[
\text{IF } x_{f_1} > b_{f_1} > y_{f_1} \text{ AND } x_{f_2} > b_{f_2} > y_{f_2} \text{ ... AND } x_{f_n} > b_{f_n} > y_{f_n} \text{ THEN } \]

\[ b = l \]

where \( x_{f_1,f_2,...,f_n} \) and \( y_{f_1,f_2,...,f_n} \) are the decision border values of the high-level feature space of sample blocks.

An industrial-residential classifier was formulated for the Landsdowne dataset. Results are shown in figure 5 b).

Accuracy assessment

The block classification results for the Landsdowne dataset show that there are large gaps where no classification exists. This is due to the incomplete block segmentation results. This is considered a limitation to a full aerial image understanding system, since when attempting to improve bottom-up object classification results, the objects within these unclassified regions will not be treated upon and improved. Nevertheless, a large portion of the imagery has been classified. A potential solution to fill the gaps would be to merge block segments of the same land use, and then classify those unclassified regions to the land use of the adjacent segment with the largest percentage shared border. This would result in a final land use classification map useful for top-down analysis or other land use applications.

From a visual inspection of the land use classification results (figure 5 b)) a few industrial segments have been mis-classified as residential in the upper right. This is probably because these industrial blocks have similar features to the residential blocks. This is to be expected since within an industrial area we can expect certain sub-regions to be non-industrial like.

Table 2 shows the error matrix of block classification results for the Landsdowne dataset. An overall block classification accuracy of 82% was achieved. The entire routine discussed above was applied to an additional two datasets of the N1 city and Eifendale suburbs, generating block classification accuracies of 100% and 75% respectively.
### Table 2: Error matrix for Landsdowne dataset block classification results.

<table>
<thead>
<tr>
<th>Class</th>
<th>industrial</th>
<th>residential</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>industrial</td>
<td>22</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>residential</td>
<td>17</td>
<td>62</td>
<td>79</td>
</tr>
<tr>
<td>sum</td>
<td>39</td>
<td>63</td>
<td>102</td>
</tr>
<tr>
<td>producer accuracy</td>
<td>0.56</td>
<td>0.98</td>
<td>overall</td>
</tr>
<tr>
<td>user accuracy</td>
<td>0.96</td>
<td>0.78</td>
<td>0.82</td>
</tr>
</tbody>
</table>

## 4 Conclusions and Future Work

In this paper we have asked the question of what is urban context, with the objective of improving urban object extraction. A contextual model was described that consists of a set of high-level features. With this model it was proved that different urban scene types can be discriminated based solely on a high-level feature set. A feature selection routine established a feature subset that causes maximum separation of land use classes. This feature space is the final definition of a contextual model.

An automated land use classification system was proposed as a contribution to a full urban aerial image understanding system. The system works by segmenting and classifying road block regions based on bottom-up object classification results. The block segmentation algorithm is a novel contribution. The land use classification routine was tested on experimental urban scene datasets each consisting of various land uses. Block segmentation results show a clear separation of land use regions, which demonstrates the effectiveness of the block extraction method. Blocks were then classified to a land use based on a set of high-level features measured off the bottom-up data. The high-level features are similar to those features defined for the contextual model.

An overall block classification accuracy of over 75% for the three datasets proves two points. Firstly, it demonstrates the potential to perform automated zoning based on only a limited number of geometric features of building and road objects. Automated zoning is an important requirement for many land use applications. Secondly, it demonstrates that a recognition of context can be achieved albeit with incomplete bottom-up results. This is
an encouraging result within the framework of image understanding. Future work would be to improve these bottom-up results based on a contextual model such as that proposed in this paper. Each block region will be treated independently depending on its land use. Thus a hierarchical, semantic analysis is performed. The contextual model chosen (by scene-matching) for each block will be used to constrain the final description of the scene in that block, by testing bottom-up object extraction results against the model. The goal is to obtain a final scene description that is consistent as possible with the contextual model. The Markov Random Field, as used in (Porway et al., 2008; Hernandez-Gracidas and Sucar, 2007), might be useful for this purpose.

References


Busgeeth, K., A. Brits and J.B. Whisken. 2008. Potential application of remote sensing in monitoring informal settlements in developing countries


