A DECISION LEVEL FUSION METHOD FOR OBJECT RECOGNITION USING MULTI-ANGULAR IMAGERY

F. Tabib Mahmoudi [a, *], F. Samadzadegan [a], P. Reinartz [b]

[a] Dept. of Surveying and Geomatics, College of Engineering, University of Tehran, Tehran, Iran-
(fmahmoudi,samadz)@ut.ac.ir
[b] Dept. of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, German Aerospace Centre
(DLR), Oberpfaffenhofen 82234 Weßling, Germany- peter.reinartz@dlr.de

KEY WORDS: Object Recognition, Decision Level Fusion, Visibility Map, Shadow Recovery, Texture, Weighting Strategy

ABSTRACT:

Spectral similarity and spatial adjacency between various kinds of objects, shadow and occluded areas behind high rise objects as well as complex relationships lead to object recognition difficulties and ambiguities in complex urban areas. Using new multi-angular satellite imagery, higher levels of analysis and developing a context aware system may improve object recognition results in these situations. In this paper, the capability of multi-angular satellite imagery is used in order to solve object recognition difficulties in complex urban areas based on decision level fusion of Object Based Image Analysis (OBIA). The proposed methodology has two main stages. In the first stage, object based image analysis is performed independently on each of the multi-angular images. Then, in the second stage, the initial classified regions of each individual multi-angular image are fused through a decision level fusion based on the definition of scene context. Evaluation of the capabilities of the proposed methodology is performed on multi-angular WorldView-2 satellite imagery over Rio de Janeiro (Brazil).The obtained results represent several advantages of multi-angular imagery with respect to a single shot dataset. Together with the capabilities of the proposed decision level fusion method, most of the object recognition difficulties and ambiguities are decreased and the overall accuracy and the kappa values are improved.

1. INTRODUCTION

The complex nature and diverse composition of land cover types found within urban areas exhibit several difficulties in producing large scale topographic maps from *VHR* satellite imagery (Shackelford and Davis, 2003). This situation also potentially leads to lower accuracy in pixel based image classification approaches which only use spectral characteristics of remotely sensed data (Myint et al., 2011; Shackelford and Davis, 2003; Zhou and Troy, 2008; Blaschke, 2010; Duro et al., 2012).

As many land cover classes in an urban environment have similar spectral signatures, spatial information such as height and topological relationships must be exploited to produce accurate classification maps. Already many researchers have investigated the potential of the object based image analysis (OBIA) approaches for dealing with *VHR* images and complexities in urban areas (Myint et al., 2011; Zhou and Troy, 2008; Blaschke, 2010; Peets and Etzion, 2010; Jacquin et al., 2008; Laliberte et al., 2012).

As it is depicted in previous researches in the field of OBIA, the accuracy of object recognition results in complex urban areas directly depends on the segmentation and rule based classification processes (Ivits et al., 2005; Platt and Rapoza, 2008; Zhou and Troy, 2008; Myint et al., 2011; Laliberte et al., 2012). Moreover, depending on the viewing angle of the sensor, some parts of urban objects may be occluded by high rise objects such as buildings or trees (Habib et al., 2007; Bang et al., 2007). Therefore, using only single shot imagery, it is very difficult to obtain valuable object recognition results. On the other hand, stereo imagery (from just two viewing directions)

may partially improve object recognition results, but still there might be missing information for filling up occluded areas.

The unique agility and collection capacity of some modern satellite sensors such as WorldView-2 provide image sequences of a single target from many observation angles within one satellite orbit. Fusion of the information coming from multi-angular imagery is valuable for filling up occluded areas and obtaining reliable object recognition results. Therefore, using multi-angular imagery together with contextual information and higher levels of modification should be utilized in order to improve object recognition results in complex urban areas based on *VHR* satellite images. In this paper a context aware system is proposed for decision level fusion of the object recognition results and solving problems arising from occlusions and shadow areas, based on the multi-angular *VHR* remotely sensed data.
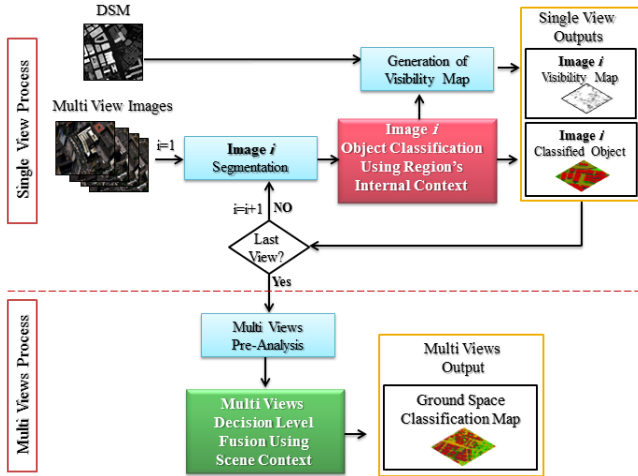
2. PROPOSED METHOD

As depicted in figure 1, the proposed object recognition method is composed of two main stages: single view and multi views processes.

In the first stage, region's internal contextual information is utilized for object classification on each of the individual images. Per segment spectral and textural pixels' interactions together with structural features based on size, shape and height of a segmented region, generate region's internal context. This kind of information is utilized for object based image analysis composed of image segmentation and object classification on each of the individual images. Second stage of the proposed

---

* Corresponding author.

methodology performs decision level fusion on the multi view classified regions based on scene context in order to reduce ambiguities and uncertainties in the generated classification map.

Figure 1. General structure of the proposed object recognition method

2.1 Processing Based on Single View

Object based image analysis requires generating segmented regions as classification units. In this research, multi-resolution segmentation technique is applied on the content of each of the individual images in order to segment it into regions. The multi-resolution segmentation procedure starts with single image objects of one pixel and repeatedly merges a pair of image objects into larger ones. The merging decision is based on the local homogeneity criterion, describing the similarity between adjacent image objects (Baatz and Schape, 2000).

After performing segmentation, a knowledge based classification process should be performed on each of the segmented regions. Therefore, it is necessary to gather proper knowledge composed of per-segment interactions between pixels and structural characteristics of each segmented region in order to provide region's internal context. Following items represent some sample of these features with their basic mathematics:

2.1.1 Spectral Features: the ratios between reflectance values of various spectral bands generate normalized difference indices (NDI) and simple ratios (SR).

$$NDI_{k,l} = (Band_k - Band_l)/(Band_k + Band_l) \qquad (1)$$

$$SR_{k,l} = Band_k/Band_l \qquad (2)$$

where $NDI_{k,l}$ = normalized difference index between bands k & l

$SR_{k,l}$ =simple ratio between bands k & l

Band=image channel

2.1.2 Textural Features: the grey value relationships between each pixel and its neighbours in the pre-identified segmented regions.

$$Energy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} I(i, j)^2 \qquad (3)$$

$$Entropy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} I(i, j)Log(I(i, j)) \qquad (4)$$

where Energy & Entropy = Samples of textural features

I(i, j) =value of image pixel (i,j)

N=neighbour size in the segmented region

2.1.3 Structural Features: Calculating suitable structural features based on spatial characteristics and heights of segmented regions provide another part of internal context for using in the object classification process. In this paper, 2D structural features such as area, elongation and solidity are used together with relief and surface roughness as 3D structural features.

$$Elongation = MajorLength/MinorLength \qquad (5)$$

$$Solidity = Area/ConvexArea \qquad (6)$$

where Major & Minor Lengths = Major & minor axes of the bounding polygon

Convex area= Area of the smallest convex polygon that can contain the region

After generation of the above mentioned spectral, textural and structural features based on the image data and height products, an optimum selection of features for recognizing each individual object classes has the most effective role in generating a rich knowledge base of region's internal context.

The object classification can be performed by encapsulating the knowledge base containing region's internal context, into a rule set and definition of a strategy for object recognition. The proposed strategy is a multi-process classification model that is a progressive process composed of multiple steps. In the first step, the entire segmented image dataset is classified based on spectral reasoning rules. Then, in the next steps, classified objects are modified based on textural and structural reasoning rules, respectively (Tabib Mahmoudi et al., 2013).

Despite the high potential of the strategies in OBIA, wrong classified regions due to occlusion and shadow through high rise 3D objects, decrease the reliability of the object recognition results. Therefore, using the proposed decision level fusion of the multi views object based image analysis can handle these difficulties in an efficient way.

2.2 Processing Based on Multi Views

High rise objects together with viewing angle of the sensor make some uncertainties in the object classification results which cannot be solved using only single shot imagery. Therefore, in the proposed object recognition methodology another processing stage is defined based on the context aware decision level fusion of the object based image analysis on multi views. This stage is composed of two main operations; pre-analysis on object classification results for all images and the decision level fusion of the multi views.

2.2.1 Pre-Analysis: Total visibility map generated from all of the individual visibility maps provides a main tool for performing visibility analysis. Visibility analysis determines the number of views for all, pixel (i,j) of the ground space are visible. Performing visibility analysis as a pre-analysis operation, one can categorize all of the ground space pixels into three groups: visible in all images, visible in some images and visible in none of the (occluded in all) images. This categorization is useful for the definition of scene context in the proposed decision level fusion algorithm.

2.2.2 Decision level fusion of multi views: The first step of the decision level fusion is the back projection from each ground space pixel to its pre-identified visible images (all images for the group of visible in none) based on the results of the pre-analysis operation. These back projections find the objects and sub-objects in multi views those pixels belong to. Therefore, there are three groups of areas containing various object types and their sub objects in multi views.

As scene context of an object can be defined in terms of its co-occurrences with other objects and its occurrences in the whole scene, information regarding sensor's look angle, distance from occluded areas, heights and areas of sub-objects those pixels belong to, are the fundamentals for scene context definition.

In this research, scene contextual information is utilized for weighting all class objects in order to assign them to the ground space pixels. So, this proposed decision fusion strategy takes place in the level of the pre-defined object classes based on the object classification results. Therefore, if there are *n* various recognizable object classes in multi views, also *n* different weights should be calculated for assigning each of the object classes to the ground space pixels. Weighting class objects highly depends on the scene context and ground space categorization as a result of visibility analysis. Following items represent the proposed weighting strategy based on scene context:

1. Off-nadir weight depends on the sensor information such as off-nadir viewing angle for each of the multi view images. Classified regions in the nearest to nadir views have largest off-nadir weights.

2. Weight of occlusion, depends on the longest distances between each object region and its neighbouring occluded areas in the multi view images. More distance from occluded areas leads to a larger weight of occlusion.

3. Weight of structural features, depends on the spatial information such as the area of the object region. For instance, if pixel (x,y) belongs to a large sub-object, the weight of structural features becomes larger.

4. Weight of topological relationships, depends on the neighbouring relationships between occluded regions and their nearest visible object regions. Determining the weight of topological relationship for an occluded sub-object in all views depends on its neighbouring visible objects with smallest height differences. In other word, if there is small amount of height difference between two neighbouring object regions, class label of visible region assigns to the occluded one.

For ground space pixels those are categorized in the groups of visible in all or visible in some images, calculating the weight of scene contextual information for each object class is based on the summation of off-nadir, occlusion and structural weights in their visible views. For ground space pixels those are visible in none of the images, the summation of off-nadir weight and

weight of topological relationships in all views is calculated as the weight of scene context for each object class.

In addition, classification accuracies of various object types in each view also affect on the decision fusion results. Weights of object classes are determined based on the user and producer accuracies in each of the multi views object based image analysis.

$$W_k(Class_i) = \frac{2A_{U(Class_i)} \times A_{P(Class_i)}}{A_{U(Class_i)} + A_{P(Class_i)}} \qquad (7)$$

where

$W_k(Class_i)$ = weight of classification for object class *i* in view *k*

$A_{U(Class_i)}$ = User accuracy of object class *i* in view *k*

$A_{P(Class_i)}$ = Producer accuracy of object class *i* in view *k*

Therefore, total weight of each object class composed of weights of scene context and classification in various views. The object class with the largest weight should be selected as the winner class label for each of the ground space pixels. If the winner class is shadow, structural and height based relations are used in order to determine true object types instead of shadow area.

3. EXPERIMENTS AND RESULTS

The potential of the proposed object recognition methodology is evaluated for automatic object recognition in multi-angular WorldView-2 satellite imagery over Rio de Janeiro (Brazil) which were collected in January 2010 within a three minutes time frame with satellite elevation angles of 44.7°and 56.0°in the forward direction, and 59.8° and 44.6° in the backward direction, as shown in Figure 2. The multi-angular sequence contains the downtown area of the city, including a number of large and high buildings, commercial and industrial structures, and a mixture of community parks and private housing.

Moreover, using multi-angular WorldView-2 imagery the DSM is generated from multiple pairs of panchromatic stereo images – in epipolar geometry – using the Semi-Global Matching (SGM) algorithm (Hirschmüller 2008; d'Angelo et al., 2010; Sirmacek et al., 2012).

In order to obtain OBIA results based on multi-angular images, multi-resolution segmentation algorithm is applied on each of the independent satellite images using eCognition software with the values 90, 0.2 and 0.1 for the scale parameter, compactness and shape parameters, respectively. Then, various spectral, textural and structural features are measured on image regions for the generation of the knowledge base containing internal context and performing object level classification on segmented regions.

Building, road, tree, grass land and shadow area are the pre-identified object classes based on visual inspections. Despite shadow is not a real object class, detecting real objects under shadow areas based on spectral responses is a difficult task dealing with *VHR* imagery. Therefore, in this paper shadow is first recognized as separate object class and in a later step, we are going to recover shadow areas based on decision level fusion of topological relationships in multi-angular views.

Before decision level fusion of multi-angular images, by performing visibility analysis on total visibility map, ground space categorized into three groups: 924914 pixels are visible in
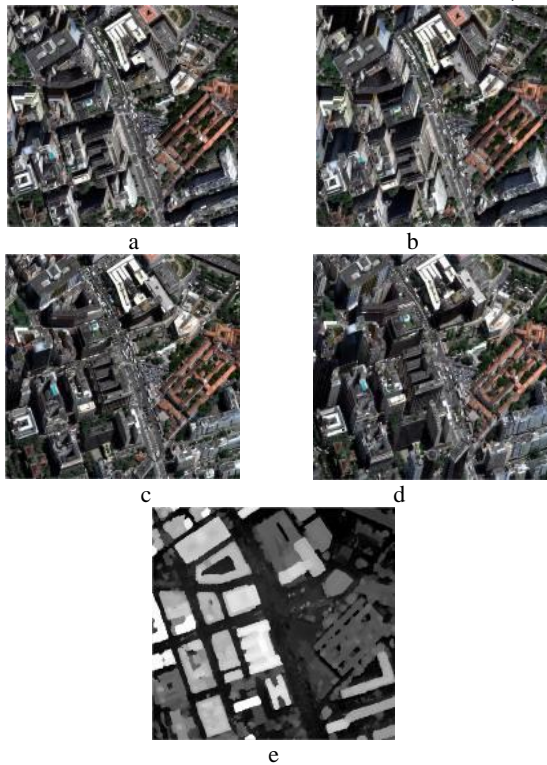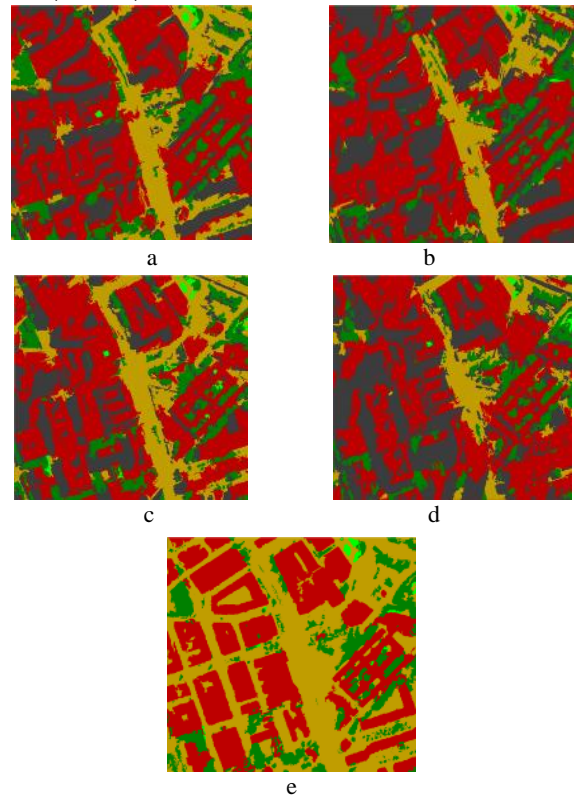
Figure 2. Dataset, a) WV-2 imagery with 56º satellite elevation angle, b) WV-2 imagery with 44.7º satellite elevation angle, c) WV-2 imagery with 59.8ºsatellite elevation angle, d) WV-2 imagery with 44.6º satellite elevation angle, e) Digital Surface Model (Ground space)

Figure 3. Results of the object based image analysis and decision level fusion of them, *a*) OBIA of WV-2 imagery with56º satellite elevation angle, *b*) OBIA of WV-2 imagery with 44.7º satellite elevation angle, *c*) OBIA of WV-2 imagery with 59.8ºsatellite elevation angle, *d*) OBIA of WV-2 imagery with 44.6º satellite elevation angle, *e*) decision level fusion result

all images, 552286 pixels are visible in some images and just 25200 pixels are visible in none of the (occluded in all) images. After performing decision level fusion on the OBIAs based on the proposed context aware strategy, analysis represents that decision level fusion removes shadow from object recognition results and detects road regions occluded by high rise buildings especially in pixels which are visible in all or some images.

Figure 3 compares the object based image analysis from each of the multi-angular images with decision level fusion of them. According to the obtained results, despite the high potential of the utilized regions' internal context, high rise buildings together with small elevation angles of the sensor increase difficulties in object recognition in urban areas. However, using scene context and classification accuracies in the proposed decision level fusion algorithm can improve the classification results.

Moreover, for the quantitative evaluation of the results from decision fusion system, some areas of the pre-defined object classes are manually selected by an expert operator on 3D model generated from multi-angular scenes (Figure 4). Sample areas are then compared with their corresponding results of different steps of the proposed object recognition algorithm.

As depicted in table 1, the comparison is based on the number of correctly detected pixels (true positive), wrongly detected pixels (false positive), and the not correctly recognized pixels (false negative), determined after performing the object recognition algorithm. After that, using quantitative values for each object class, completeness, correctness and quality criteria are determined for the obtained results (Tabib Mahmoudi et al. 2013).
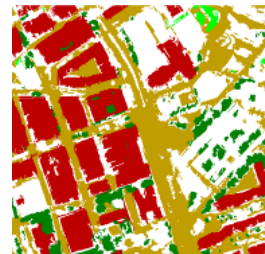
Figure 4. Manually selected sample regions

Table 1. Accuracy assessment of the obtained results from the proposed decision fusion algorithm

Decision Level Fusion Results	Object Classes			
	Building	*Road*	*Tree*	*Grass*
True Positive	91.6	94.8	69.5	43.4
False Positive	25.6	25.3	12.7	0.08
False Negative	8.4	5.2	30.5	56.6
Completeness	0.91598	0.9480	0.6952	0.4337
Correctness	0.59726	0.6259	0.4125	0.8526
Quality	0.56624	0.6051	0.3494	0.4035
Overall Accuracy	89.699			
Kappa	0.8295			

In order to perform more quantitative analysis on the results, object based image analysis of each of the multi-angular images are compared within sample regions and then, their overall accuracies and kappa are compared with their decision level fusion results. As it is depicted in table 2, the largest values of overall accuracy and kappa belong to the views with largest elevation angles and by decreasing the elevation angles; overall accuracies and kappa are decreasing. Comparing tables 1 & 2 with each other shows that using proposed algorithm for decision level fusion of multi-angular images increase the amount of overall accuracy and kappa values in classification results.

Table 2. Accuracies of multi-angular object based image analysis

	View 1	View 2	View 3	View 4
Elevation Angles	56°	44.7°	59.8°	44.6°
Overall Accuracy	72.118	62.606	72.449	55.732
Kappa	0.5648	0.422	0.5641	0.3258

Figure 5 illustrates considerable improvement in quality of context aware decision level fusion results with respect to the object based image analysis of each of the multi-angular images.
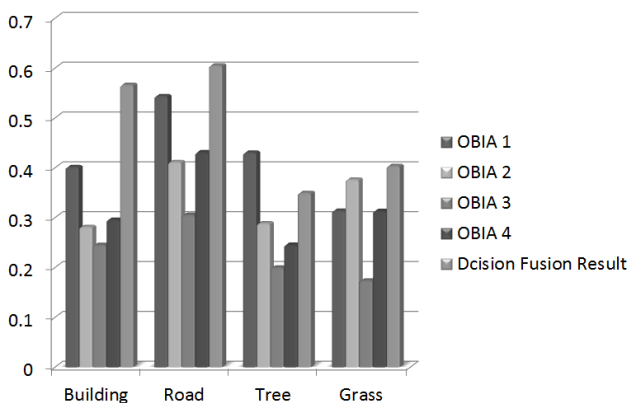

Figure 5. Comparison between the quality of Multi-angular OBIAs and their Decision level fusion results

4. CONCLUSION

A context aware strategy is proposed for decision level fusion of the object based image analysis results based on multi-angular WorldView-2 satellite imagery.
According to the various elevation angles of the sensors, high rise 3D objects such as buildings may cause occlusion and shadow areas in the *VHR* remotely sensed imagery. In such a situation, true class labels of some parts of object regions cannot be detected. So, large number of false positive and false negative pixels decreases the classification accuracies. Using the proposed decision level fusion method of multi-angular imagery based on the developed context aware system improves the classification accuracy.

Results are shown that visibility analysis on class labels before and after fusion and performing decision level fusion on the OBIA results of multi-angular images reduce the amount of wrongly classified pixels for 3D object classes and reveals improved class labelling of the occluded areas.
This method still needs further modifications in the field of definition of contextual information. Moreover, incorporating a Lidar DSM and multi-angular satellite or aerial images can be tested for improvement of object recognition results in complex urban areas.

5. REFERENCES

Baatz, M.,Schape, A., 2000. Multi-resolution segmentation: an optimization approach for high quality multi-scale image segmentation.In Angewandte Geographische Informationsverabeitung. XII. Beitragezum AGIT-Symp. Salzburg, T. Strobl, T. Blaschke and G. Griesebner (Eds), 12–23.

Bang, K.I., Habib, A.F., Kim, C., Shin, S., 2007.Comprehensive Analysis of Alternative Methodologies for True Orthophoto Generation From High Resolution Satellite and Aerial Imagery.American Society for Photogrammetry and Remote Sensing, Annual Conference, Tampa, Florida, USA.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS International Journal of Photogrammetry and Remote Sensing 65(1), 2-16.

d'Angelo, P., 2010. Image matching and outlier removal for large scale DSM generation. Convergence in Geomatics. CGC & ISPRS . ISPRS Symposium Commission I, 15.-18. June 2010, Calgary, Canada, on CD-ROM.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sensing of Environment 118 (2012) 259–272.

Habib, A.F., Kim, E., Kim, C., 2007. New methodologies for true ortho-photo generation. Photogrammetric Engineering and Remote Sensing 73(1), 25–36.

Hirschmüller, H., 2008. Stereo Processing by Semi global Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2), 328-341.

Ivits, E., Koch, B., Blaschke, T., Jochum, M., Adler, P., 2005. Landscape structure assessment with image grey-values and object-based classification at three spatial resolutions. International Journal of Remote Sensing 26(14), 2975–2993.

Jacquin, A., Misakova, L., Gay, M. 2008.A hybrid object-based classification approach for mapping urban sprawl in periurban environment. Landscape and Urban Planning 84 (2008), 152–165.

Laliberte, A.S., Browning, D.M., Rango, A., 2012. A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. International Journal of Applied Earth Observation and Geoinformation 15 (2012) 70–78.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sensing of Environment 115(5), 1145-1161.

Peets, A., Etzion, Y., 2010. Automated recognition of urban objects and their morphological attributes using GIS.ISPRS Archive XXXVIII (4-8-2-W9), "Core Spatial Databases-Updating, Maintenance and services-from theory to practice" Haifa, Israel.

Platt, R.V., Rapoza, L., 2008. An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification. The Professional Geographer60(1), 87-100.

Shackelford, A.K., Davis, C.H., 2003. A Hierarchical Fuzzy Classification Approach for High-Resolution Multispectral Data over Urban Areas. IEEE Transaction on Geosciences and Remote Sensing 41(9), 1920-1932.

Sirmacek, B., Taubenböck, H., Reinartz, P., Ehlers, M. 2012. Evaluation of automatically generated 3-D city models based on six different DSMs from airborne and space-borne sensors. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1), 59-70.

TabibMahmoudi, F., Samadzadegan, F., Reinartz, P., 2013. Object oriented image analysis based on multi-agent recognition system.Can.J.Computers & Geosciences 54 (2013) 219–230.

Zhou, W., Troy, A., 2008. An Object-Oriented Approach for Analyzing and Characterizing Urban Landscape at the Parcel Level. International Journal of Remote Sensing 29, 3119-3135.