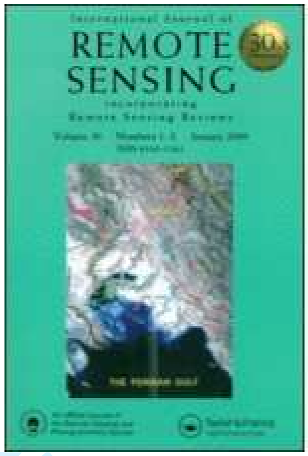
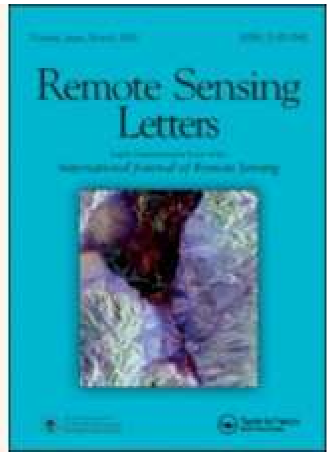




Unsupervised Building Detection in Complex Urban Environments from Multi Spectral Satellite Imagery

Journal:	*International Journal of Remote Sensing*
Manuscript ID:	TRES-PAP-2010-0099.R1
Manuscript Type:	IJRS Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Aytekin, Orsan; METU, Electrical and Electronics Eng. Department Erener, Arzu; SU, Geodesy and Photogrammetry Engineering Department Ulusoy, İlkay; METU, Electrical and Electronics Eng. Department Duzgun, Sebnem; METU, Mining Engineering Department
Keywords:	IMAGE ENHANCEMENT, IMAGE PROCESSING
Keywords (user defined):	Building extraction, shadow detection, morphological operations

Unsupervised Building Detection in Complex Urban Environments from Multi Spectral Satellite Imagery

Aytekin, Ö. [1], Erener, A. [2], Ulusoy, İ. [1], and Düzgün, HSB. [3]

[1] Electrical and Electronics Engineering Department, Middle East Technical University, METU Ankara, Turkey, aytekin@eee.metu.edu.tr, ilkay@metu.edu.tr
[2]Geodesy and Photogrammetry Engineering Department, Selcuk University, Konya, Turkey, erener@metu.edu.tr
[3]Geodetic and Geographic Information Technologies, Middle East Technical University Ankara, Turkey, duzgun@metu.edu.tr

Abstract: A generic algorithm is presented for automatic extraction of buildings and roads from complex urban environments in high resolution satellite images where the extraction of both object types at the same time enhances the performance. The proposed approach exploits spectral properties in conjunction with spatial properties, both of which actually provide complementary information to each other. First, high resolution pan-sharpened color image is obtained via merging the high resolution panchromatic and the low resolution multispectral images yielding a color image at the resolution of the panchromatic band. Natural and man-made regions are classified and segmented by using Normalized Difference Vegetation Index (NDVI). Shadow regions are detected by using chromaticity to intensity ratio in YIQ color space. After the classification of the vegetation and the shadow areas, the rest of the image consists of man-made areas only. The manmade areas are partitioned by mean shift segmentation where some resulting segments are irrelevant to buildings in terms of shape. These artifacts are eliminated in two steps: First, each segment is thinned using morphological operations and its length is compared to a threshold which is specified according to the empirical length of the buildings. As a result, long segments which most probably represent roads are masked out. Second, the erroneous thin artifacts which are classified by principle component analysis (PCA) are removed. In parallel to PCA, small artifacts are wiped out based on morphological processes as well. The resultant manmade mask image is overlaid on the ground truth image, where the buildings are previously labeled, for the accuracy assessment of the methodology. The method is applied to Quickbird images of eight different urban regions each of which includes different properties of surface objects. The images are extending from simple to complex urban area. The simple image type includes a regular urban area with low density and regular building pattern. The complex image type involves almost all kinds of challenges such as small and large buildings, regions with bare soil, vegetation areas, shadows etc. Although the performance of the algorithm slightly changes for various urban complexity levels, it performs well for all types of urban areas.

Keywords- Building extraction, road extraction, mean shift segmentation, shadow detection, morphological operations

1. Introduction

Identification and characterization of urban objects such as buildings, roads, utilities and recreational areas is crucial for city planning, disaster management, map making, military target detection etc. Since the manual extraction of urban objects from high resolution images requires qualified domain experts and a large amount of effort in terms of time and cost, researchers have been working on automatic urban object detection methods to increase the speed of this process for many years. However, due to the required accuracy and the involved complexity in the high resolution satellite images, semi or fully automated building extraction methods have still need to be improved (Wilkinson 2005).

Extraction of urban objects from high resolution satellite data has mainly two different aspects. The first aspect is related to the object properties. Inherently, man made structures are composed of different sizes and different surface materials such as concrete, brick, asphalt, metal, plastic, glass, shingles, soil, etc. Hence, there is a high spatial and spectral diversity. However, the existing methodologies in the literature are mostly restricted to specific types of shapes or surface features. A complex urban environment involves various shapes and surface materials and buildings may appear indistinguishable from roads and pavements. Also rooftops may reflect fragmented characteristics due to shading or they may be occluded by other buildings or vegetation. The second aspect is related to the image properties. Images differ in resolution, sensor type, orientation, quality, dynamic range, illumination conditions, weather conditions and seasons, etc. Thus, it is hardly possible to use a certain algorithm for all kinds of images. As a result, due to the complexity of the problem, it is complicated to develop generic methods for building extraction (i.e., detection and delineation of buildings) from numerous types of images.

The early works in building detection were based on line extraction, edge detection and building polygon generation. These methods mostly use a large set of heuristic rules and are computationally expensive. Also, they were content dependent. Typical examples of these methods are Herman and Kanade (1986), Huertas and Nevatia (1988), Irvin and McKeown (1989), Matsuyama and Hwang (1990), Venkateswar and Chellappa (1991), Krishnamachari and Chellappa (1996), Lin and Nevatia (1998), Kim and Nevatia (1999), Mayer (1999), Gereke et al. (2001), Persson et al. (2005), Peng and Jin (2007). Recently, with the availability of high spatial and spectral resolution satellite images, most of the studies focus on the use of spectral reflectance values or features extracted from spectral information. The automatic feature extraction techniques from high spatial and spectral resolution satellite images can be divided into two main categories. The first category relays on the classification of the objects by using multi spectral reflectance values (e.g. Segl and Kaufmann 2001, Shan and Lee 2002, Lee et al. 2003, Benediktsson et al. 2003, Ünsalan and Boyer 2005, Sohn et al. 2005, Katartzis and Sahli 2008). The second category is mainly based on feature extraction techniques from panchromatic images (e.g. Lin and Nevatia 1998, Wei et al. 2004, Wei and Prinet 2005). Moreover, the studies of Muller et al. (1997), Baltsavias et al. (2001), Sohn and Dowman (2001) discuss the effect of resolution on the building extraction extensively.

It seems to be promising that, with the availability of wide range of data diversity, feature level fusion incorporated into the structural information improves the performance of manmade structure detection. Besides, this would increase the generic characteristic of the methods. In general, the problem of building extraction can be considered in two phase tasks, namely low level and high level tasks. First, low level tasks concentrate on determining the region of interest. Then, high level tasks (feature extraction and classification) are performed. In the literature, different kinds of features were defined and feature spaces were created (Pesaresi 2000, Benediktsson et al. 2001, Tatem et al. 2001, Haverkamp 2004, Zhen et al. 2004). These features were either classified if supervision is available or clustered if supervision is not possible. Features which are widely used in the literature can be grouped as geometric, photometric and structural. Geometric features define basic geometrical properties such as area, circumference, roundness, right angles, corners, straight lines etc. Photometric features are related to color information. Structural features refer to connectedness of neighbors according to some similarity measures.

Classification of the content was generally performed by the rule-based and the context-driven approaches and the content was classified into several types such as buildings, vegetation, roads and water areas. In doing this, density (rural, suburban, urban), object complexity (residential, industrial, military), architecture (elaborate, plain, none), terrain defined by Digital Elevation Model (DEM) (flat, hilly, mountainous), or vegetation defined by NDVI (none, moderate, heavy) were taken into account. Some studies concentrated on extracting low level features for model based context driven hypothesis and subsequently set relations among them in favor of supporting the building hypothesis (Haverkamp 2004, Zhen et al. 2004, Peng and Liu 2005, Katartzis and Sahli 2008, Lizarazo and Elsner 2009). On the other hand, multi-scale analyses are also studied in the literature (Huang et al. 2007, Chen et al. 2009).

Besides buildings, road extraction is also considered in semi or fully automated object detection from satellite images. Mostly snakes, higher order active contours, dynamic programming or probabilistic approaches have been proposed for road detection. For example, Klang (1998), Laptev et al. (2000), Peteri and Ranchin

(2003) used the most common snake's algorithm for the detection of road. Mena and Malpica (2003) and Guo et al. (2004) focused on segmenting road areas. Guo et al. (2004) dealt with the investigating on how to build geo-specific road databases from aerial images for driving simulation. Mena and Malpica (2003) used the Dempster-Shafer theory of evidence for the fusion of texture to extract linear features. Amini et al. (2002) proposed a fuzzy logic algorithm for road extraction from multispectral imagery. Barzohar and Cooper (1996) used dynamic programming and Bicego et al. (2003) proposed probabilistic approaches for road detection. Bacher and Mayer (2005) introduced an approach for automatic road extraction from high resolution multispectral imagery, Christophe and Inglada (2007) proposed a robust geometric method to provide a first step extraction level of road and Yang and Wang (2007) proposed an improved model for road detection based on the principles of perceptual organization and classification fusion in human vision system (HVS).

Most of the works in the literature concentrates only on the extraction of a single object such as only buildings or only roads. They do not consider road extraction and building extraction together. The main aim of this paper is to present a generic algorithm for automatic extraction of both buildings and roads from complex urban environments by using high resolution satellite images, as the extraction of both features at the same time enhances the performance of object detection. Moreover, majority of the studies, which propose semi/fully automated building extraction algorithms, implemented their algorithm for a limited number of cases. In this study, eight images from various urban environments are tested. The images can be extended from simple to complex. The performance of the algorithm is evaluated for these eight urban areas having different properties.

The paper is organized as follows: The study region and the data sets are described in Section 2. The methodology and the algorithm are provided in Section 3. In Section 4, the experimental results and the accuracy assessment are discussed. The results and the discussions are presented in Section 5 before concluding the paper with final remarks in Section 6.

2. The study region and data sets

The data is composed of medium resolution (2.4m) multi-spectral (R, G, B, NIR) bands and high resolution (0.6m) panchromatic band of Quickbird image of Ankara city acquired in year 2002. Eight different small test regions are selected from the image to test the proposed approach. The test regions are located in different districts of Ankara region such as: Yenimahalle-Emniyet (1) Çankaya-Ortadoğu (2), Altındağ-Karapürçek (3), Yenimahalle-25 Mart (4), Yenimahalle-Mehmet Akif Ersoy (5), Yenimahalle-Ormançiftliği (6-7), and Çankaya-Karakusunlar (8) (see Figure 1). The regions are numbered starting from simple (1) to complex (8). Each region is formed by different properties of surface objects. The test regions from 1 to 3 involve simple surface types which mean that the buildings in the images have brick rooftops which provide high contrast with the background. Additionally, low density built up area are present in these regions and this may reduce the over and the under estimations in accuracy assessment. On the other hand the test regions from 4 to 8 contain denser buildings with different kinds of challenges. For example, some buildings have different building rooftops such as concrete, brick, and metal. Also, some buildings have similar intensity reflectance to the roads and this causes interference of the roads and the buildings.

(Figure 1)

3. The methodology

The proposed method mainly consists of three steps: First, vegetation and shadow areas are masked and man made segments are obtained. Next, main roads are detected. Finally, thin and long artifacts are filtered by PCA and small segments are eliminated by morphological operations. The main steps of the method are given in Figure 2.

(Figure 2)

3.1. Masking vegetation and shadow regions

The characteristics of urban objects are formed not only by their spectra but also through their structure (Zhang 1999). Therefore, it is important, in land mapping or urban applications, both spectral and spatial resolution to be high. In order to produce multispectral images having the highest spatial resolution available within the data set, many methods have been proposed, namely IHS (Intensity, Hue, Saturation), PCS (Principal Component Substitution), Multiplicative, Brovey, High Pass Filter, NN (Neural Networks), Wavelet transforms (WT) and PANSHARP fusion methods (e.g., Cliche et al. 1985, Tom, 1987; Ranchin and Wald 1993,Wald et al. 1997, Zhang and Albertz 1997, Zhou et al. 1998,Zang, 1999). Intuitively, the selected image fusion algorithm may be thought to have an effect to the quality of posterior analysis, because the complexity of a scene increases with the resolution. However, as Wald et al., (1997) discussed, many of the studies such as: Woodcock and Strahler (1987), Welch et al. (1989), Rowe (1992), Raffy (1993) demonstrate that the quality of the assessment of a parameter is an unpredictable function of the resolution. Among the fusion methods, the most frequently used methods, i.e. the IHS and the PCS usually distort the spectral characteristics of the original multispectral images to different extents (Shettigara 1992, Zhang 1999). In this study, PANSHARP algorithm is considered because it is one of the best merging techniques that give the best results without changing the statistical parameters of the original images (Nikolakopoulos, 2004) at all. Therefore, initially, medium resolution multi-spectral imagery (MS) and high resolution panchromatic imagery (PAN) of Quickbird data are fused by using PANSHARP algorithm (Yun 2002). As a result, a color image at the resolution of pan is obtained. NDVI is calculated by using the near infrared (NIR) and the Red (R) bands of the pan-sharpened image from the ratio (NIR-R) / (NIR+R). High index values indicate vegetation regions whereas low values represent manmade regions. The histogram of the index image has two peaks (one for the vegetation and other for the other regions) and a suitable threshold is determined according to Otsu's method (Otsu 1979). Otsu's method is used to automatically determine the threshold that effectively separates two-mode histogram image into two classes.

In Otsu's method, a threshold is exhaustively searched that minimizes the within-class variance, defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \tag{1}$$

where weights w_1 and w_2 are the probabilities of the two classes, σ_1^2 and σ_2^2 are variances of these classes separated by the threshold t. Computing this within-class variance for each of the two classes for each possible threshold involves a lot of computation. Otsu shows that minimizing the within-class variance is the same as maximizing between-class variance by subtracting the within-class variance from the total variance of the combined distribution:

$$\sigma_b^2(t) = \sigma^2 - \sigma_1^2(t) = w_1(t)[\mu_1(t) - \mu]^2 + w_2(t)[\mu_2(t) - \mu]^2 \tag{2}$$

where σ^2 is the combined variance and μ is the combined mean. This method takes an image and computes its normalized histogram which is treated as the discrete probability density function. After that, the desired threshold is found by maximizing the between-class variance.

To remove the shadow regions, ratio of the chromaticity to the intensity is used and the best performance is obtained in YIQ color space (Tsai 2006). The shadow regions have higher ratio of I to Y. A suitable threshold is determined by using Otsu's method as explained above. At this point, vegetation and shadow areas are masked out leaving the manmade structures to be segmented in the next step (Figure 3).

(Figure 3)

3.2. Mean shift segmentation

After masking out vegetation and shadow regions, the image is then segmented by mean shift segmentation algorithm (Comaniciu and Meer 2002).

The mean shift is a general nonparametric analysis method to delineate the clusters in multi-modal feature space. It is not based on a priori model assumption for the clusters. A feature space is a transform domain of input obtained through processing of sensor outputs. On the other hand, the nature of the feature space and the analysis of the feature space is application independent. Methods which rely on a priori knowledge of the number of clusters and implicit assumption of the shape for clusters, are not able to delineate the clusters as expected. The mean shift is applied to image analysis as presented in this paper although the applicability of the mean shift is not restricted to image analysis rather being a general technique.

Mean shift procedure originates from the kernel density estimation, known also as Parzen window method, and the density gradient estimation based on a kernel. Given n data points $x_i = 1,\ldots,n$ in d dimensional space R^d, the kernel density estimator of the underlying density f(x) is defined as follows with the kernel K(x) and the bandwidth parameter h

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3}$$

The main aim is to achieve radially symetric kernel which are often more suitable for generic density estimation (Comaniciu and Meer 2002). Radially symetric kernels can be defined by using so called kernel profile, k(x) for $x \geq 0$ as follows:

$$K(x) = c_{k,d}\, k(\|x\|^2) \tag{4}$$

where $c_{k,d}$ is the normalization constant that makes the integral of K(x) equal to one. Introducing the profile notation, density estimator can be written as

$$\hat{f}_{h,K}(x) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \tag{5}$$

Next, density gradient estimator is obtained as the gradient of the density estimator which is defined as follows:

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}_{h,K}(x) = \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(x - x_i)k'\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \tag{6}$$

If g(x) is defined with the assumption that the derivative of the kernel profile exists for all $x \geq 0$, except for a finite number of points, then Equation (7) is obtained

$$g(x) = -k'(x) \tag{7}$$

and can be used for the profile, the kernel G(x) can be defined as follows:

$$G(x) = c_{g,d}\, g(\|x\|^2) \tag{8}$$

where $c_{g,d}$ is the corresponding normalization constant. Next, putting g(x) into Equation (6), Equation 9 is obtained.

$$\nabla \hat{f}_{h,K}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x_i - x)\, g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \tag{9}$$

Rearranging the terms, the following equation is obtained

$$\nabla \hat{f}_{h,K}(x) = \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \right] \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right] \tag{10}$$

The first term on the right side of the equation is proportional to the density estimate at x computed with the kernel G

$$\hat{f}_{h,G}(x) = \frac{c_{g,d}}{nh^{d}} \sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \tag{11}$$

and the second term is the mean shift

$$m_{h,G}(x) = \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right] \tag{12}$$

which is the differece between the weighted mean based on kernel G for the weights and x for the center of the kernel window. Then equation (10) can be expressed as

$$\nabla \hat{f}_{h,K}(x) = \hat{f}_{h,G}(x) \frac{2c_{k,d}}{h^2 c_{g,d}} m_{h,G}(x) \tag{13}$$

where

$$m_{h,G}(x) = \frac{1}{2} h^2 c \frac{\hat{\nabla} f_{h,K}(x)}{\hat{f}_{h,G}(x)} \tag{14}$$

As seen from Equation (14), the mean shift vector computed with kernel G is proportional to the normalized density gradient estimate computed with kernel K (Comaniciu and Meer 2002). A significant property of the mean shift vector is that it always points in the direction of the gradient of the density. Therefore, it indicates a path towards the stationary points in the density. Another favorable property is that the vector is actually computed without explicit density estimation and density gradient estimation. Lastly, the presence of the normalization by the density estimate enables adaptive gradient descent method. That is, the mean shift steps are large in the regions of low-density values and small near the local maxima, which is obviously desirable. The procedure is initiated with a pixel and continues through computation of the mean shift vector followed by a translation of the kernel by the vector. This procedure is applied iteratively until getting to the stationary point where the gradient is zero.

An image is represented as a combination of two dimensional lattice of pixels and their spectral information. The space of the lattice is known as spatial domain while spectral information is known as range domain. Range domain and spatial domain have different nature so they should be normalized accordingly. Therefore, a multivariate kernel is defined as the product of two radially symmetric kernels as follows:

$$K_{h_s,h_r}(x) = \frac{C}{{h_s}^2 {h_r}^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{15}$$

where x^s is the spatial and x^r is the range component of a feature vector, k(x) is the common profile for both domains, h_s and h_r the kernel bandwidths, and C is the normalizing constant. In general, normal kernel provides satisfactory results so only the bandwidth parameters (h_s and h_r) need to be specified.

There are two important parameters to be specified: The spatial bandwidth and the bandwidth range of the kernel. These parameters control the resolution of feature space analysis and are closely related to the size and the saliency of underlying objects. If these parameters are specified small relative to object sizes in the image this brings oversegmentation and if specified large then this leads to undersegmentation and loss of salient features.

Mean shift can be used for detecting modes, smoothing by preserving edges and segmentation. As for image segmentation, the aim is to cluster pixels sharing a similarity in pixel values. For this purpose, the filtering procedure is run and all convergence points are stored. The set of all pixels converging to the same mode, basin of attraction of that mode, are delineated by grouping the converged pixels which are closer than the spatial and the range bandwidth. One should be aware of the fact that on flat plateaus the graident is close to zero and the procedure could stop. Therefore, this may lead to oversegmentation and unreal modes. To overcome these artifacts, postprocessing should be done through merging mode candidates at a distance less than the kernel bandwidths and the segments smaller than a prespecified area threshold.

The resultant image obtained by mean shift segmentation includes only the building rooftops along with some irrelevant segments generated by side effects of the previous masking processes. For example, pavements, roads and bared soil regions are highly correlated with rooftops. These kinds of problems appear frequently in the literature also. To get rid of road segments, a methodology is proposed based on the hypothesis that road segments are longer and thinner than buildings.

3.3. Main road detection

Ideally, it is expected that road segments are different from building structures in length and width. Road segments are longer and thinner than building segments and they usually have undefined branches. With this motivation, all segments are processed one by one to assess their shape characteristics in terms of length. As a first step, each segment is filled to cover the holes which may be caused by small objects such as cars on the road segments and which are possibly labeled as distinct segments than road. Consequently, closing and opening morphological operations are applied. Then, a modified version of the thinning algorithm (Lam et al. 1992) is applied to obtain representative one-pixel wide skeletons of the segments. The thinning algorithm used in this study is summarized as follows:

- Divide the image into two distinct subfields in a checkerboard pattern.
- In the first subiteration, delete pixel p from the first subfield if and only if the conditions 1, 2, and 3 are all satisfied.
- In the second subiteration, delete pixel p from the second subfield if and only if the conditions 1, 2, and 4 are all satisfied.

Condition 1:

$$X_H(p) = 1$$

where

$$X_H(p) = \sum_{i=1}^{4} b_i$$

$$b_i = \begin{cases} 1, & (x_{2i-1} = 0) \wedge ((x_{2i} = 1) \vee (x_{2i+1} = 1)) \\ 0, & otherwise \end{cases}$$

$x_1, x_2, ..., x_8$ are the values of the eight neighbors of p, starting with the east neighbor and numbered in counter-clockwise order. $\wedge$ and $\vee$ are used for logical AND and OR operations, respectively.

Condition 2:

$$2 \le \min\{n_1(p), n_2(p)\} \le 3$$

where

$$n_1(p) = \sum_{k=1}^{4} x_{2k-1} \vee x_{2k}$$

$$n_2(p) = \sum_{k=1}^{4} x_{2k} \vee x_{2k+1}$$

Condition 3:

$$(x_2 \vee x_3 \vee \overline{x_8}) \wedge x_1 = 0$$

Condition 4:

$$(x_6 \vee x_7 \vee \overline{x_4}) \wedge x_5 = 0$$

The two sub-iterations made up of one iteration of the thinning algorithm. The iterations are repeated until there is no more pixel deletion and the remaining pixels form the single pixel wide skeleton.

Skeletons may contain erroneous protrusions outgoing from the main body due to boundary imperfections of segments. Thus, to mitigate the effect of undesirable protrusions, end points of the skeletons, which have only one neighbor, are removed iteratively. Finally, we end up with single pixel wide skeleton of the corresponding segment, the length of which is equal to the number of pixels on the skeleton (Figure 4).

(Figure 4)

The distribution of segment lengths may be regarded as evaluation criteria of labeling the segments as road or building. When a threshold, which is automatically estimated from this distribution by Otsu's method (Otsu 1979), is applied, the main road segments are eliminated. Eliminated main road segments are shown in Figure 5 for all test images.

(Figure 5)

After this elimination step, there still remain some artifacts of road segments, lengths of which are smaller than the determined threshold and comparable to the lengths of the buildings. In addition to these road artifacts, some artifacts which have smaller sizes than buildings may exist. In the following section, the method to handle these two types of artifacts is explained.

3.4. Filtering the artifacts

There are two types of artifacts. One of them is building residuals which are unreasonably small in area. The other one is road residuals which are unreasonably thin and short. In order to decide on whether a given segment is an artifact or not, principle component analysis (PCA) is applied to each segment to estimate the spatial extend of the segment. Considering thin artifacts, they show large variances along the first principle component whereas small variance along the second principle component. Therefore, the ratio of the corresponding eigenvalues provides the variances along the corresponding eigenvectors and offers a measure of how thin the segment is. Higher ratios represent thin segments. This ratio is thresholded to detect the artifact segments where the threshold value is automatically estimated by Otsu's method (Otsu 1979) from the distribution of the ratios. Eliminated artifact segments are shown in Figure 6.

(Figure 6)

After PCA elimination, remaining regions which are small in area (i.e. less than 5 pixels) are also removed. Figure 7 shows the delineated candidate buildings detected with the proposed algorithm. These candidates are then overlaid with the manually labeled ground truth in order to assess the accuracy.

(Figure 7)

4. Accuracy assessment

As for the accuracy assessment of the proposed method, pixel-based and object-based evaluation metrics are applied. Basically, the ground truth, which is produced by manually labeling the building boundaries in the GIS environment, is compared with the output image obtained by the algorithm.

In pixel-based evaluation (Shufelt and Mckeown 1993), the accuracy assessment involves computation of True Positive (TP), False Positive (FP) and False Negative (FN) pixel numbers. TP refers to the regions detected correctly as building. FP refers to the false alarm detected as buildings. FN refers to the regions, which could not be detected as buildings although they exist in the ground truth. Based on these components the split factor, SF, missing factor, MF, percent of building detection, PBD, and quality percent, QP, are calculated as follows:

- $SF = FP/(TP + FP)$
- $MF = FN/(TP + FP)$
- $PBD = 100 * TP/(TP + FN)$
- $QP = 100 * TP/(TP + FP + FN)$

For the object-based error measure, the overlapping area matrix (OAM) (Beauchemin and Thomson 1997), is used to measure the performance of the algorithm.

The ι'th ground truth object is shown as GT_i while the φ'th output object is denoted as O_j . The set of objects in the ground truth is denoted as:

$GT_r = \{ GT_0, GT_1, ..., GT_{N_r} \}$ and the set of objects in the output image is denoted as: $O_o = \{ O_0, O_1, ..., O_{N_o} \}$ [10].

Here

GT_0 : is the background in the ground truth,
O_0 : is the background in the algorithm output image,
N_r : is the number of objects in the GT, and
N_o : is the number of objects in the output image.

The sizes of the areas covered by the objects GT_i and O_j and the size of the whole image I can be calculated from the OAM as follows:

$$n(GT_i) = \sum_{j=0}^{N_0} C_{ij}$$

$$n(O_j) = \sum_{i=0}^{N_r} C_{ij}$$

$$n(I) = \sum_{i=0}^{N_r} n(GT_i) = \sum_{j=0}^{N_0} n(O_j)$$

Here, C_{ij} ισ the number of pixels in the ι'th object in the ground truth that overlap with the φ'th object in the output image.

By using OAM every pair of ground truth GT_i and output O_j objects are classified as correct detections, over detections, under detections, missed detections or false alarms (Hoover et al. 1996) by a given threshold where $T = 0.5$ is used in this study as follows:

Correct detection: A pair of objects GT_i and O_j is classified as correct detection if

- $C_{ij} \geq T \times n(O_j)$ and
- $C_{ij} \geq T \times n(GT_i)$.

Over detection: An object GT_i and a set of objects $O_{j_1}, ..., O_{j_k}$, $2 \leq k \leq N_o$, are classified as over detection if

- $C_{ij_t} \geq T \times n(O_{j_t}),\ \forall t \in \{1, ..., k\}$, and
- $\sum_{t=1}^{k} C_{ij_t} \geq T \times n(GT_i)$.

Under detection: A set of objects $GT_{i_1}, ..., GT_{i_k}$, $2 \leq k \leq N_r$, and an object O_j are classified as under detection if

- $\sum_{t=1}^{k} C_{i_t j} \geq T \times n(O_j)$, and
- $C_{i_t j} \geq T \times n(GT_{i_t}),\ \forall t \in \{1, ..., k\}$.

Missed detection: A ground truth **o**bject GT_i is classified as a missed detection if it is not included in any instance of correct detection, over detection or under detection.

False alarm: An output object O_j is classified as a false alarm if it is not included in any instance of correct detection, over detection or under detection.

5. Results and discussions

The algorithm performance is tested for eight different region types (Figure 2). The ground truth data and the output map of the method are analyzed in the Matlab environment. The performances are computed by evaluating pixel-based and object-based criteria.

The pixel-based performance evaluation results are given in Table 1. These results show that the Percent of Building Detection and Quality Percent depend on the complexity and texture of the region. The best Percent of Building Detection is observed as 96.11%, which is obtained for data from medium complex urban area. The main problem is that high False Positive values are obtained due to the detection of bare soil and irrelevant man-made structures such as pavement as building.

(Table 1)

The object-based correct detection, false alarm, missed detection, over detection and under detection rates are computed as described in the previous section. In order to evaluate and compare the algorithm performance, the percent rates of measurements are computed for different test regions (I1:Image 1, I2:Image 2, I3:Image 3, I4:Image 4, I5:Image 5, I6:Image 6, I7:Image 7, I8:Image 8) and presented in Figure 8.

(Figure 8 a)
(a)

(Figure 8 b)
(b)

(Figure 8 c)
(c)

(Figure 8 d)
(d)

(Figure 8 e)
(e)

(Figure 8)

Higher values for correct detection, over detection and under detection in Figure 8.a, Figure 8.d and Figure 8.e represent better performance (Aksoy et al. 2008). The algorithm results higher rates of correct detection, in the range of 64% to 91%, for most of the images (I1, I2, I3, I4, I5, I6, and I7) for the threshold values less than 0.7. For images I2 and I5, high and constant rate of correct detection is observed for the threshold values less than 0.8. The algorithm provides higher performances for I8 for the threshold values less than 0.4 however its performance of correct detection reduces just after this value when compared to the other test images. This might be due to the presence of complex structure types, larger region sizes and higher number of buildings with different rooftops. The over detection and the under detection are present only for I6 and I6, I7, I4 respectively (Figure 8.d and Figure 8.e).

Lower values of false alarm and missed detection rates indicate better performance (Aksoy et al. 2008). In Fig 8b, it is clear that I1 provide the lowest false alarm rates for threshold values less than 0.7 and for the rest of the regions the false alarm rates range between 26% and 67% for I2, I3, I4, I5, I6, I7 for threshold values less than 0.8. The false alarm rate is around 50% for I8 for the threshold values less than 0.4, however false alarm rate increases after this threshold value. Similarly, I8 present missed detection around 30% for the threshold values less than 0.4 (Figure8c) and the missed detection increases up to 80% for 0.9 threshold values. I7 provides the lowest percents of missed detection compared to the other images. Additionally, I2, I3, I4, I5, I6 indicate missed detection rate between 10% and 30% and these detections are approximately constant up to 0.8 threshold value.

In addition to the graphical presentation of the algorithm performance (Figure 8), the performance is also illustrated in spatial domain by overlaying OAM rates for each object in each test image (Figure 9). The OAM results in Figure 9 are computed with the threshold value of 0.5. As a result of this presentation, the correct, missed, over and under detected buildings can be assessed and evaluated. In overall results, it can be concluded that the buildings with brick rooftops can be correctly detected, however such buildings as shopping centre or trade centers (A, B, C, D) are mostly missed detections and some bare ground surfaces (E, F, G, H) of images are mostly false detected by the algorithm. The algorithm doesn't provide any over detection with 0.5 threshold value due to the distant locations of the buildings and it detects a few (I) buildings as over detection. This situation can be also confirmed by analyzing Figure 8.d and Fig 8.e. In Fig 8.d there is no over detection present in any image at 0.5 threshold value on the other hand, when Fig 8.e is analyzed it is clear that for 0.5 threshold value provide only I7 over detection as presented in Figure 9 at image 7 (I).

(Figure 9)

As a result it can be conclude that depending on the surface materials and spectral reflectance similarity of the test regions, the algorithm performance changes.

6. Conclusion

As the proposed approach exploits both spectral and spatial properties of the high resolution satellite images together, it has the ability to extract the rooftops completely and thus it provides considerably high performance for automatic extraction of buildings. A new method, which is based on PCA analysis, is proposed to evaluate the shapes of the segmented regions and to decide if the segment is a building or not. Moreover, the illustrated methodology considers extracting the roads prior to building extraction to improve the performance of building detection.

New criteria are proposed to evaluate building detection performances such that correct detection, over detection, under detection, missed detection and false alarm can be computed, evaluated and visualized clearly. Although the performance of the algorithm changes depending on the selected urban environment, the overall performance is well for different complexity levels of urban areas.

During the application of the algorithm to various urban environments with different complexities, it is observed that overall building detection is highly sensitive to segmentation performance as well as the selected thresholds. For this reason the authors plan to improve the performance of the proposed methodology based on adaptive segmentation. In addition to that, improvements of the algorithm performance are possible by implementing rule-based approaches, where certain rules related to spatial and spectral features are developed.

In this paper only the spatial and spectral properties of the images are taken into account and certain detection improvements are achieved. As the certain urban environment elements differs in terms of having 3D feature (e.g. buildings) or not (e.g. road), they can be differentiated by using this property as well. As another way of improvement, fusing the 3D information in the forms of digital surface and terrain models would be considered. Thus, the authors are also planning to evaluate the effect of fusing 3D information on the performance of the proposed algorithm.

References

Aksoy S., Özdemir B., Eckert S., Kayitakire F., Pesarasi M., Aytekin O., Borel C. C., Cech J., Christophe E., Düzgün Ş., Erener A., Ertugay K., Hussain E., Inglada J., Lefevre S., Ok Ö., Koc San D., Sara R., Shan J., Soman J., Ulusoy I., Witz R., 2008, Performance Evaluation of Building Detection and Digital Surface Model Extraction Algorithms: Outcomes of the PRRS 2008 Algorithm Performance Contest. 978-1-4244-2653-9/08/$25.00 ©2008 IEEE

Amini J., Lucas C., Saradjian M., Azizi A., and Sadeghian S., 2002, Fuzzy Logic System for Road Identification Using IKONOS Images. Photogrammetric Record vol.17, no.99, pp. 493–503, 2002.

Bacher U. and Mayer H., 2005, Automatic Road Extraction From Multispectral High Resolution Satellite Images. In: Stilla U, Rottensteiner F, Hinz S (Eds) CMRT05. IAPRS, vol. XXXVI, part 3/W24. CMRT05: Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring - Concepts, Algorithms, and Evaluation. Vienna, Austria, August 29-30, 2005.

Baltsavias E., Pateraki M., and Zhang L., 2001, Radiometric and geometric evaluation of Ikonos geo images and their use for 3D building modelling. Proc. Joint ISPRS Workshop on High Resolution Mapping from Space, Hannover, Germany, September 2001.

Barzohar M. and Cooper D.B., 1996, Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, pp. 707–721, July 1996.

Beauchemin M. and Thomson K.P.B., 1997, The evaluation of segmentation results and the overlapping area matrix. International Journal of Remote Sensing, 18, pp. 3895– 3899, December 1997.

Benediktsson A., Arnason K. and Peraresi M., 2001, The use of morphological profiles in classification of data from urban areas. In IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp. 30–34, 2001.

Benediktsson J.A., Pesaresi M., and.Arnason K., 2003, Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sensing, vol. 41, no. 9, pp. 1940–1949, 2003.

Bicego M., Dalfini S., Vernazza G., and Murino V., 2003, Automatic road extraction from aerial images by probabilistic contour tracking. IEEE International Conference on Image Processing, ICIP'03, vol. 3, pp. 585–588, 2003.

Chen J., Pan D. and Mao Z., 2009, Image-object detectable in multiscale analysis on high-resolution remotely sensed imagery. International Journal of Remote Sensing Vol. 30, No. 14, 20 July 2009, pp. 3585–3602

Christophe E. and Inglada J., 2007, Robust Road Extraction For High Resolution Satellite Images. ICIP 2007. IEEE International Conference on Image Processing, 2007.

Cliche, G., Bonn, F., and Teillert, P., 1985, Integration of the SPOT panchromatic channel into its multispectral mode for image sharpness enhancement. Photogrammetric Engineering and Remote Sensing, 51, pp. 311-316.

Comaniciu D. and Meer P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Machine Intell., 24, pp.603–619, May 2002.

Gereke M., Straub B.-M.and Koch A., 2001, Automatic Detection of Buildings and Trees from Aerial Imagery, Using Different Levels of Abstraction. Publications of the German Society for Photogrammetry and Remote Sensing, vol X.E. Seyfert, Ed., pp. 273-280, 2001.

Guo D., Weeks A., Klee H., 2004, Segmentations Of Road Area in High Resolution Images. IGARSS 2004 : Science for society: exploring and managing a changing planet 2004 IEEE International Geoscience and Remote Sensing Symposium (proceedings) (20-24 September, 2004, Anchorage, Alaska), 2004.

Haverkamp D., 2004, Automatic building extraction from IKONOS imagery. In Proceedings of ASPRS 2004 Conference, Denver, CO.

Herman M. and Kanade T.,1986, Incremental reconstruction of 3D scenes from multiple, complex images. Arti$cial Intell., vol. 30, pp. 289-341, 1986.

Hoover A., Jean-Baptiste G., Jiang, X. Flynn P. J., Bunke H., Goldgof D. B., Bowyer K., Eggert D. W., Fitzgibbon A., and Fisher R. B., 1996, An experimental comparison of range image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 7, pp. 673–689, July 1996.

Huang X., Zhang L. and Li P., 2007, An Adaptive Multiscale Information Fusion Approach for Feature Extraction and Classification of IKONOS Multispectral Imagery Over Urban Areas. IEEE Geoscience and Remote Sensing Letters, Vol. 4, No. 4, October 2007, pp. 654-658.

Huertas A. and Nevatia R., 1988, Detecting buildings in aerial images. Comput. Vision, Graphics Image Processing, vol. 41, pp. 131-152, 1988.

Irvin R.B. and McKeown D.M., 1989, Method for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Trans. Syst., Man Cybem., vol. 19, pp. 1564-1575, 1989.

Jin X. and Davis C.H., 2005, Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information. EURASIP Journal on Applied Signal Processing vol.14, pp. 2196–2206, 2005.

Katartzis A. and Sahli H., 2008, A Stochastic Framework for the Identification of Building Rooftops Using a Single Remote Sensing Image. IEEE Transactions On GeoScience And Remote Sensing, vol. 46, no. 1, January 2008

Kim W.Z. and Nevatia R., 1999, Uncertain Reasoning and Learning for Fearum Grouping. Computer Virion orid Inrage Underslunding, vo1.76. no:3. pp. 278-288, 1999.

Klang D., 1998, Automatic Detection of Changes in Road Databases Using Satellite Imagery. In: International Archives of Photogrammetry and Remote Sensing, vol. 32 no.4/1, pp. 293– 298, 1998.

Krishnamachari S. and Chellappa R.., 1996, Delineating Buildings by Grouping Lines with MRFs. Image Processing, on IEEE Trans. vol. 5, no.1, pp. 164-168, 1996.

Lam L., Lee S.W., and Wuen C. Y., 1992, Thinning Methodologies-A Comprehensive Survey. IEEE TrPAMI, Vol. 14, No.9, pp. 869-885, 1992.

Laptev I., Mayer H., Lindeberg T., Eckstein W., Steger C., and Baumgartner A., 2000, Automatic extraction of roads from aerial images based on scale space and snakes. Machine Vision and Applications, vol. 12, pp. 23–31, 2000.

Lee D.S., Shan J., and Bethel J.S., 2003, Class-guided building extraction from Ikonos imagery. Photogrammetric Engineering and Remote Sensing, vol. 69, no. 2, pp. 143–150, 2003.

Lin C. and Nevatia R., 1998, Building detection and description from a single intensity image. Computer Vision and Image Understranding, vol. 72, no.2, pp. 102-121, 1998.

Lizarazo I. and Elsner P., 2009, Fuzzy segmentation for object-based image classification. International Journal of Remote Sensing, Vol. 30, No. 6, 20 March 2009, pp. 1643–1649

Matsuyama T. and Hwang V.S.S.,1990, SIGMA: A Knowledge Based Aerial Image Understanding System. New York: Plenum, 1990.

Mayer H., 1999, Automation Object Extraction from Aerial Imagery–A Survey Focusing on Buildings. Computer vision and image understanding , vol. 74, no.2, pp. 138-149, 1999.

Mena J. and Malpica J., 2003, Color Image Segmentation Using the Dempster-Shafer Theory of Evidence for the Fusion of Texture. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, no. 3/W8, pp. 139–144, 2003.

Muller J.P., Ourzik C., Kim T., and Dowman I.J., 1997, Assessment of the effects of resolution on automated DEM and building extraction. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), pp. 245–256, Birkhauser Verlag, Basel, Switzerland, 1997.

Nikolakopoulos K.G., 2004. Pansharp vs. wavelet vs. PCA fusion technique for use with Landsat ETM panchromatic and multispectral data. Image and Signal Processing for Remote Sensing. Proceedings of the SPIE, Volume 5573, pp. 30-40 (2004); doi:10.1117/12.565726

Otsu N., 1979, A threshold selection method from graylevel histograms. IEEE Trans. Sys. Man Cyber., vol. 9, no. 1, pp. 62–66, 1979.

Peng J. and Jin Y.Q., 2007, An unbiased algorithm for detection of curvilinear structures in urban remote sensing images. International Journal of Remote Sensing, Vol. 28, No. 23, 10 December 2007, pp. 5377–5395

Peng J. and Liu Y.C., 2005, Model and context-driven building extraction in dense urban aerial images. International Journal of Remote Sensing, Vol. 26, No. 7, 10 April 2005, 1289–1307

Persson M., Sandvall M., and Duckett T., 2005, Automatic Building Detection from Aerial Images for Mobile Robot Mapping Proceedings. 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation June 27-30, 2005, Espoo, Finland, 2005.

Pesaresi M., 2000, Texture analysis for urban pattern recognition using fine-resolution panchromatic satellite imagery. Geographical and Environmental Modelling, 4, pp. 43–63, 2000.

Peteri R., and Ranchin T., 2003, Multiresolution Snakes for Urban Road Extraction from Ikonos and Qickbird. In: EARSeL Symposium, 2003.

Raffy M., 1993. Remotely-sensed quantification of covered areas and spatial resolution, International Journal of Remote Sensing, Vol.14(1), pp:135-159.

Ranchin T., and Wald L., 1993. The wavelet transform for the analysis of remotely sensed images, International Journal of Remote Sensing, Vol.14(3), pp:615-619.

Rowe C.M., 1992. Incorporating landscape heterogeneity in land surface albedo models, Journal of Geophysical Research, 98(3), pp:5037-5044.

Segl K. and Kaufmann H., 2001, Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation. IEEE Trans. Geosci. Remote Sensing, vol. 39, no. 9, pp. 2080–2083, 2001.

Shan J. and Lee S., 2002, Generalization of building polygons extracted from IKONOS imagery. Symposium on Geospatial Theory, Processing and Applications, Working Group IV/3, Ottowa, 2002.

Shettigara V.K., 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. Photogrammetric Engineering and Remote Sensing, Vol. 58, pp: 561-567.

Shufelt A.A. and Mckeown D.M., 1993, Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery. CVGIP: Image Understanding, 57 (3), pp.307-330.

Sohn G. and Dowman I.J., 2001, Extraction of buildings from high-resolution satellite data. Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), pp. 345–355, Swets & Zeitlinger B.V., Lisse, The Netherlands, 2001.

Sohn H. G., Park C.H., Kim H.S., and Heo J., 2005, 3-D building extraction using IKONOS multispectral images. Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International, vol: 2, pp. 1432- 1434. ISBN: 0-7803-9050-4, 2005.

Tatem A.J., Lewis H.G., Atkinson P.M. and Nixon M.S., 2001, Super-resolution mapping of urban scenes from IKONOS imagery using a hopfield neural network. In Proceedings of IGARSS, 2001, Southampton, UK, pp. 3203–3205.

Tom V.T., 1987. System for and Method of Enhancing Images Using Multiband Information, USA Patent 4,683,496, 28 July 1987.

Tsai, 2006, A comparative study on shadow compensation of color aerial images in invariant color models. IEEE Transactions On Geoscience And Remote Sensing, vol. 44, no. 6, June 2006.

Ünsalan C. and Boyer K.L., 2005, A system to detect houses and residential street networks in multispectral satellite images. Computer Vision and Image Understanding, vol. 98, pp. 423–461, 2005.

Venkateswar V. and Chellappa R., 1991, A hierarchical approach to detection of buildings in aerial images. Tech. Rep. CAR-TR 567, Univ. of Maryland, 1991.

Wald L., Ranchin T., and Mangolini M., 1997. Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images.Photogrammetric Engineering and Remote Sensing, Vol. 63, No. 6, June 1997, pp. 691-699.

Wei L. and Prinet V., 2005, Building Detection from High-resolution Satellite Image Using Probability Model. 0-7803-9050-4/05/$20.00 ©2005 IEEE. 2005.

Wei Y., Zhao Z., and Song J., 2004, Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection. Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International, vol. 3, pp. 2008 – 2010, 20-24 Sept. 2004.

Welch R.M., Navar M.S., and Sengupta S.K., 1989. The effect of spatial resolution upon the texture-based cloud field classifications, Journal of Geophysical Research, 94(D12):14,pp:767-14,781.
Woodcock C.E., and Strahler A.H., 1987. The factor of scale in remote sensing, Remote Sensing of Environment, Vol.21, pp:311-332
Wilkinson G.G., 2005, Results and implications of a study of fifteen years of satellite image classification experiments. IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 3, pp. 433–440, March 2005.
Yang J. and Wang R.S., 2007, Classified road detection from satellite images based on perceptual organization. International Journal of Remote Sensing, Vol. 28, No. 20, 20 October 2007, 4653–4669
Yun Z., 2002, A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images. IEEE/IGARSS'02, Toronto, Canada, June 24-28, 2002.
Zhang Y., and Albertz J., 1997. Comparison of four di€ erent methods to merge multisensor and multiresolution satellite data for the purpose of mapping. Proceedings of the ISPRS Joint Workshop `Sensors and Mapping f rom Space' of Working Groups I/1, I/3 and IV/ 4, Hannover, Germany, 29 September± 2 October 1997, pp. 275- 287.
Zhang Y, 1999. A new merging method and its spectral and spatial effects, International Journal of Remote Sensing, Vol. 20 (10), pp: 2003-2014.
Zhen G.Q., Shi, T.Z., Chun, L.Z. and Jin, Y.F., 2004, Automatic building detection from high reolution images based on multiples features. In IEEE Commission II, ICWG II/IV.
Zhou J, Civco D.L, ve Silander J.A, 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data, International Journal of Remote Sensing, Vol. 19 (4), pp: 743–757

Figure legends

Figure 1. Study region and test images.

Figure 2. Flow chart of the method

Figure 3. Vegetation mask (a) and shadow mask (b) overlaid onto the test images.

Figure 4. Eastern part of test image 8 (a). Road mask overlaid on the image (b). Single pixel wide skeleton of the road segment (c). Focused view of a piece of the road segment (d).

Figure 5. Main road mask overlaid onto the test images

Figure 6. The artifact segments identified by PCA algorithm and overlaid on the test images

Figure 7. The candidate buildings extracted at the end of the algorithm overlaid on the ground truth

Figure 8. Object-based a. correct detection, b. false alarm, c. missed detection, d. over detection and e. under detection, rates in percent for the eight different test regions.

Figure 9. The accuracy assessment of test images by the object based measure OAM for the 0.5 threshold value.

Table 1: Pixel-based performance evaluation

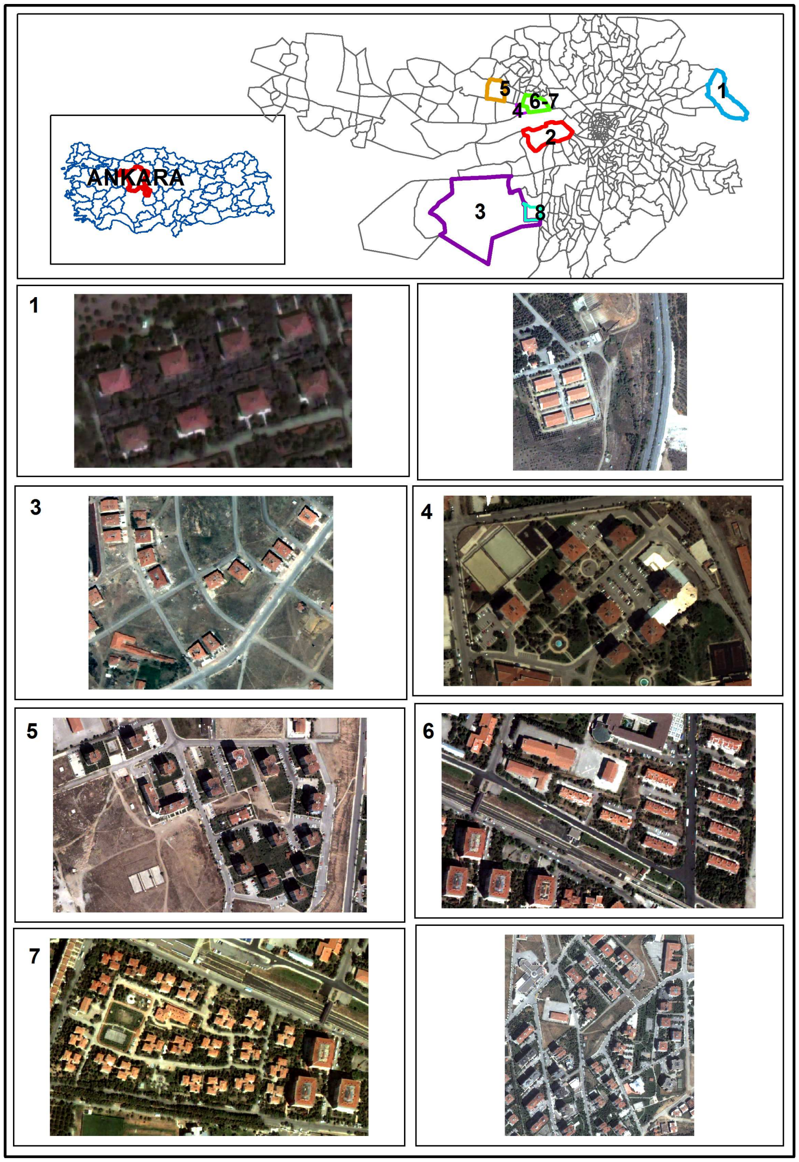


Figure 1
210x297mm (300 x 300 DPI)

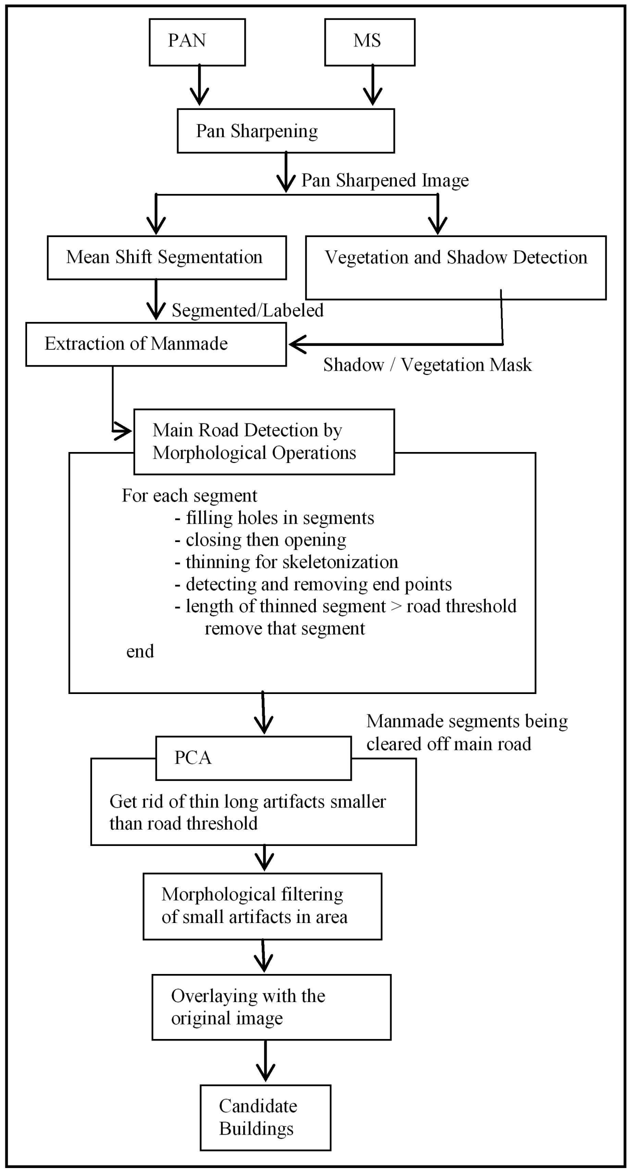


Figure 2
210x296mm (300 x 300 DPI)

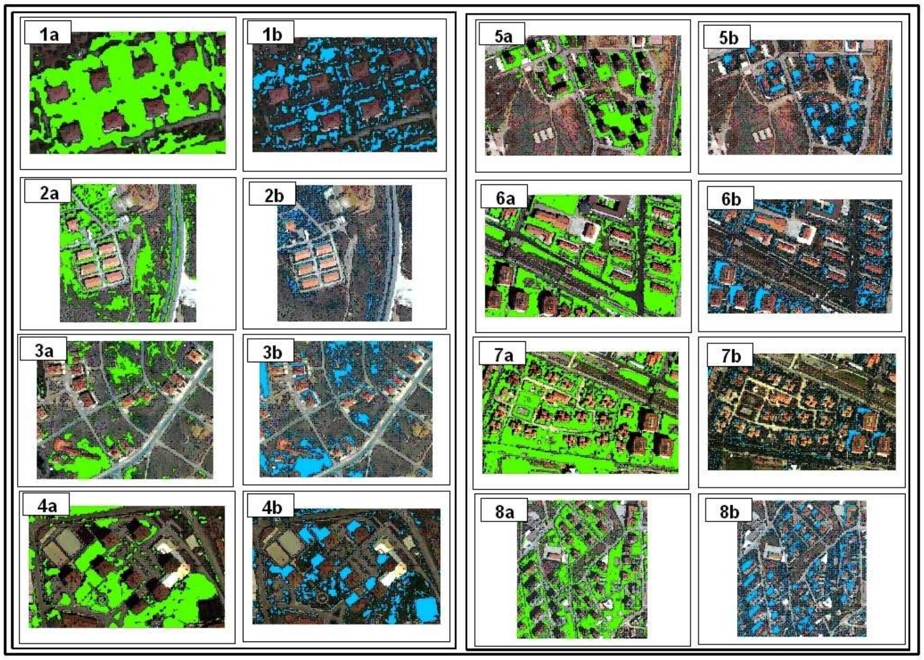


Figure 3
296x210mm (96 x 96 DPI)

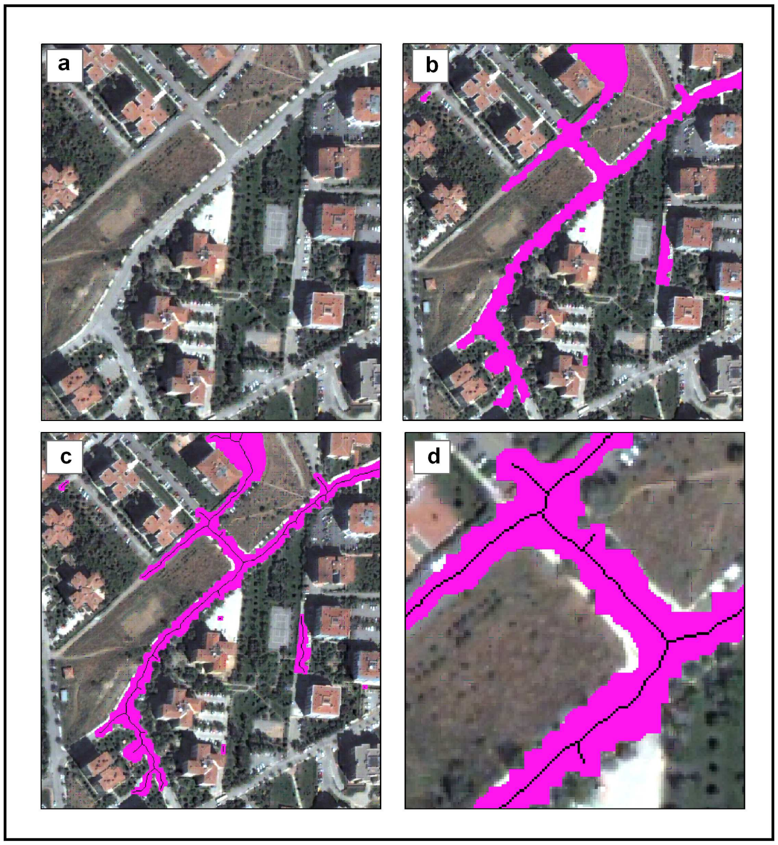


Figure 4
210x296mm (300 x 300 DPI)

Figure 5
210x297mm (96 x 96 DPI)

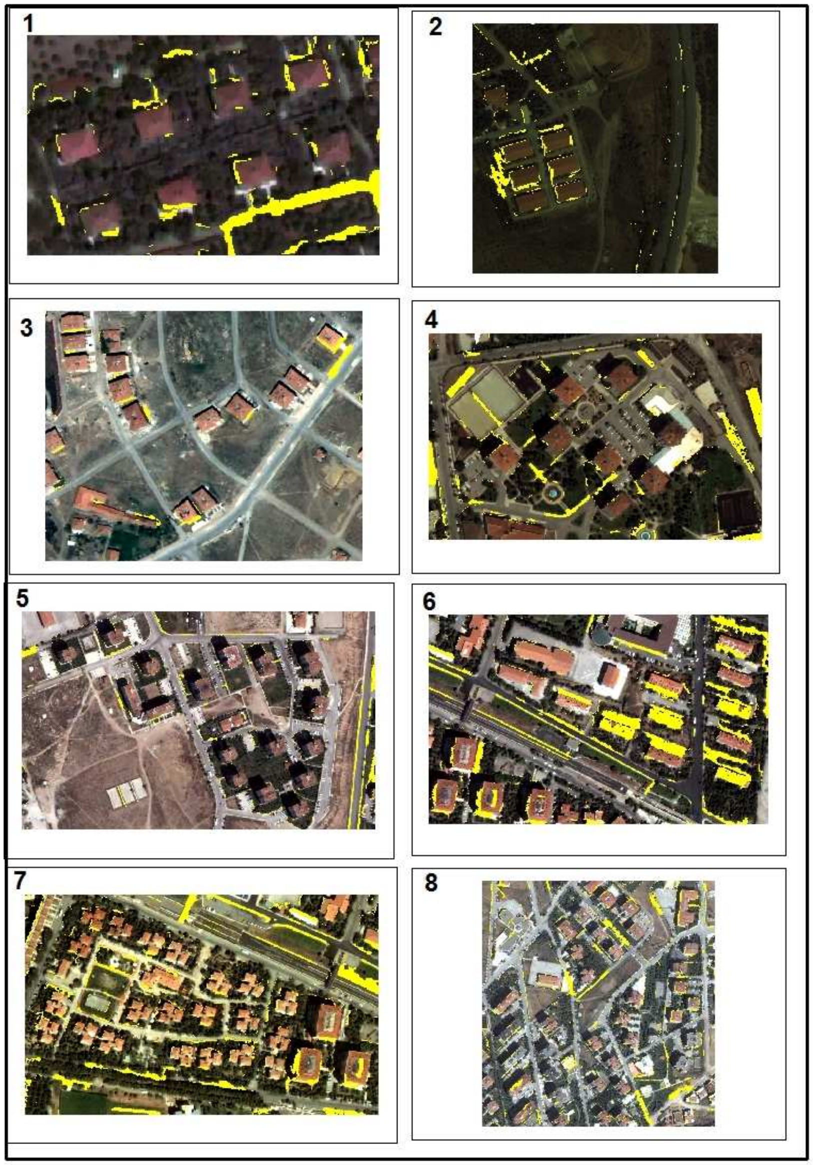


Figure 6
210x297mm (96 x 96 DPI)

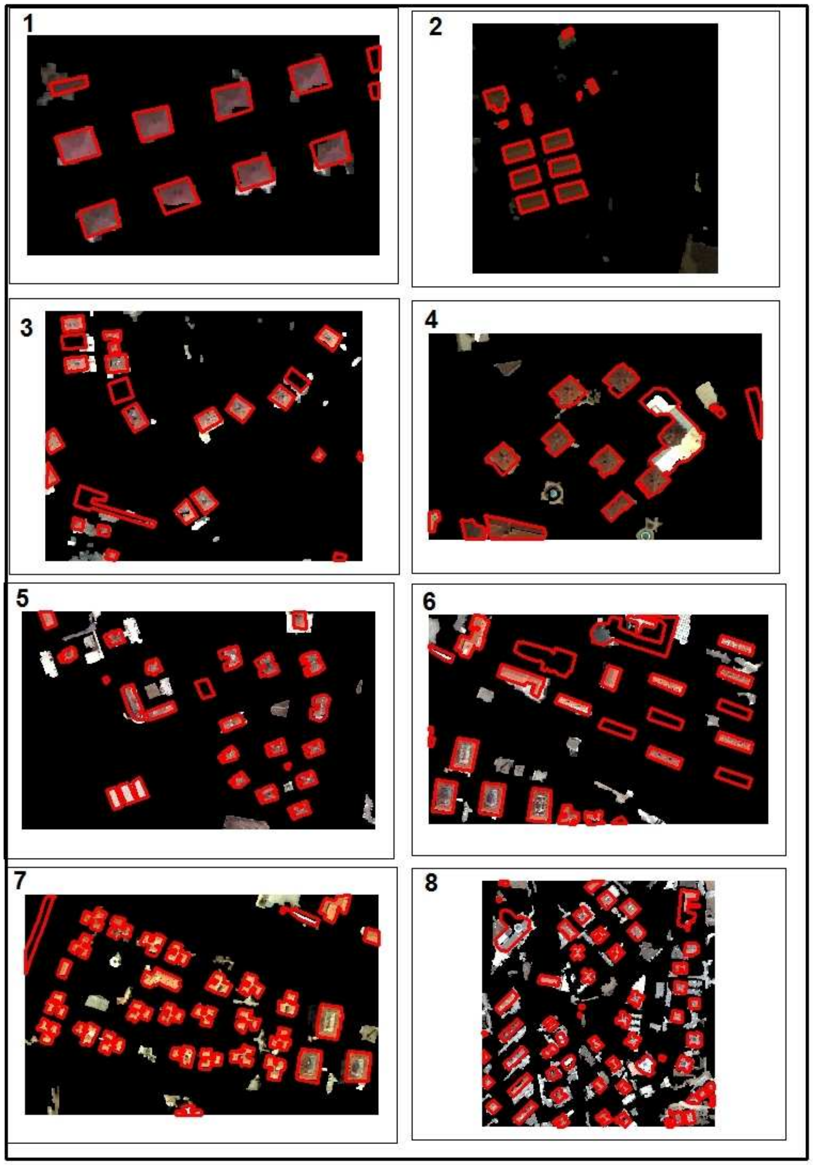
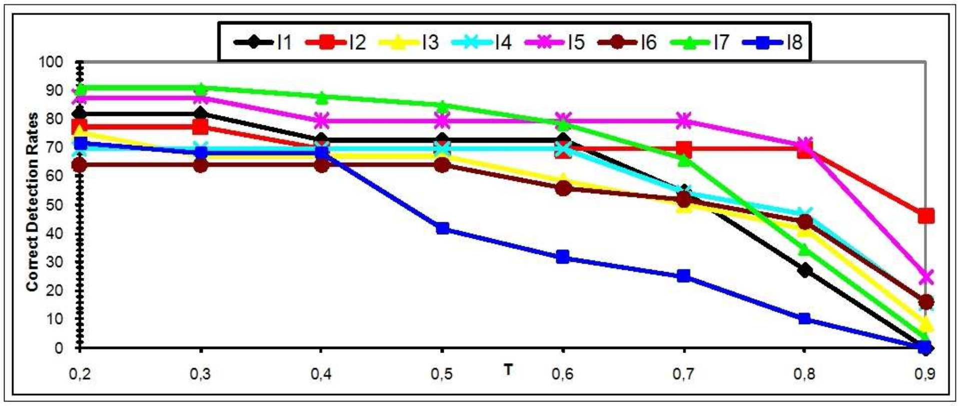
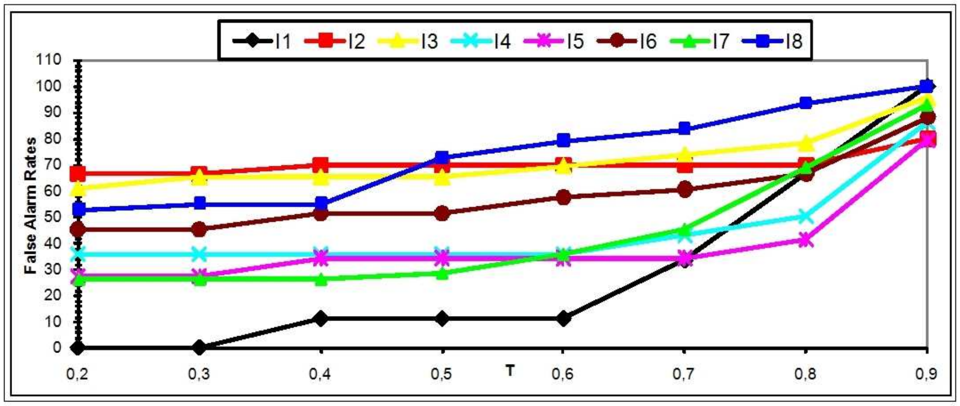


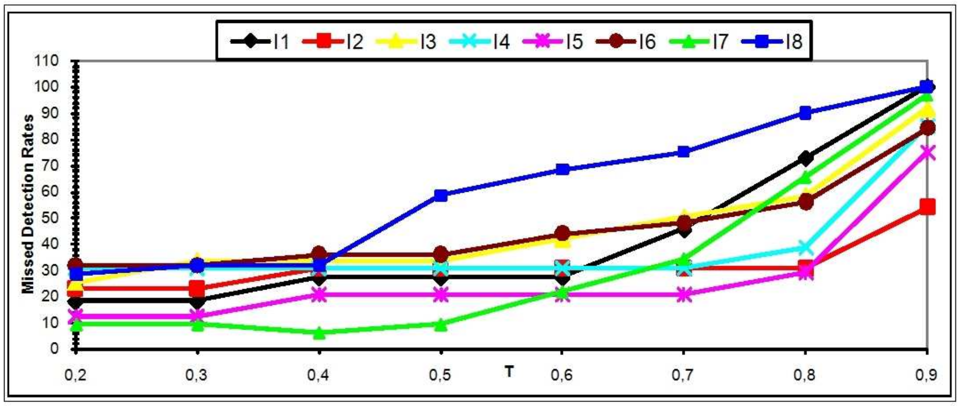
Figure 7
210x297mm (96 x 96 DPI)



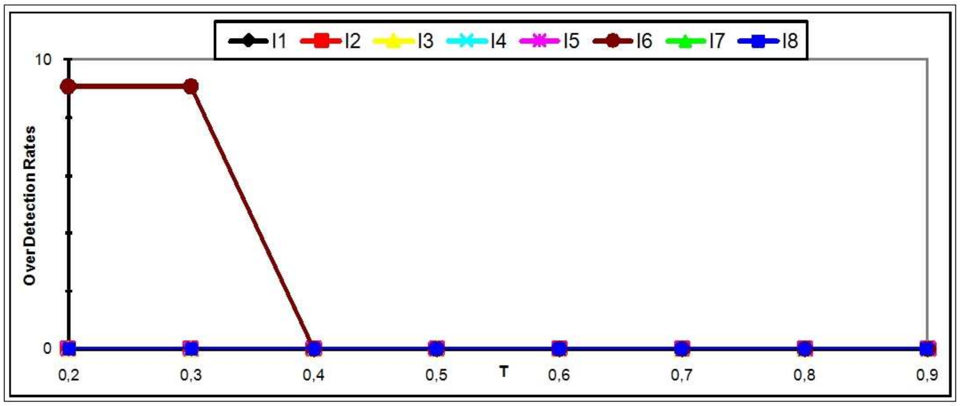
297x210mm (96 x 96 DPI)



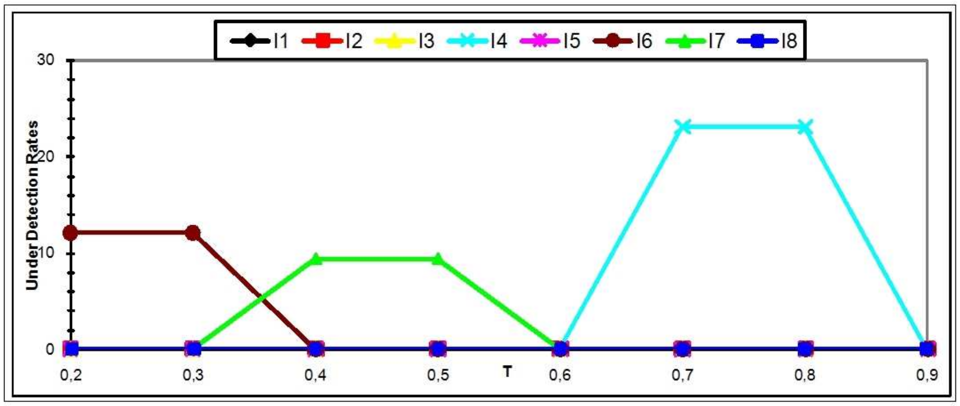
297x210mm (96 x 96 DPI)



297x210mm (96 x 96 DPI)



297x210mm (96 x 96 DPI)



297x210mm (96 x 96 DPI)

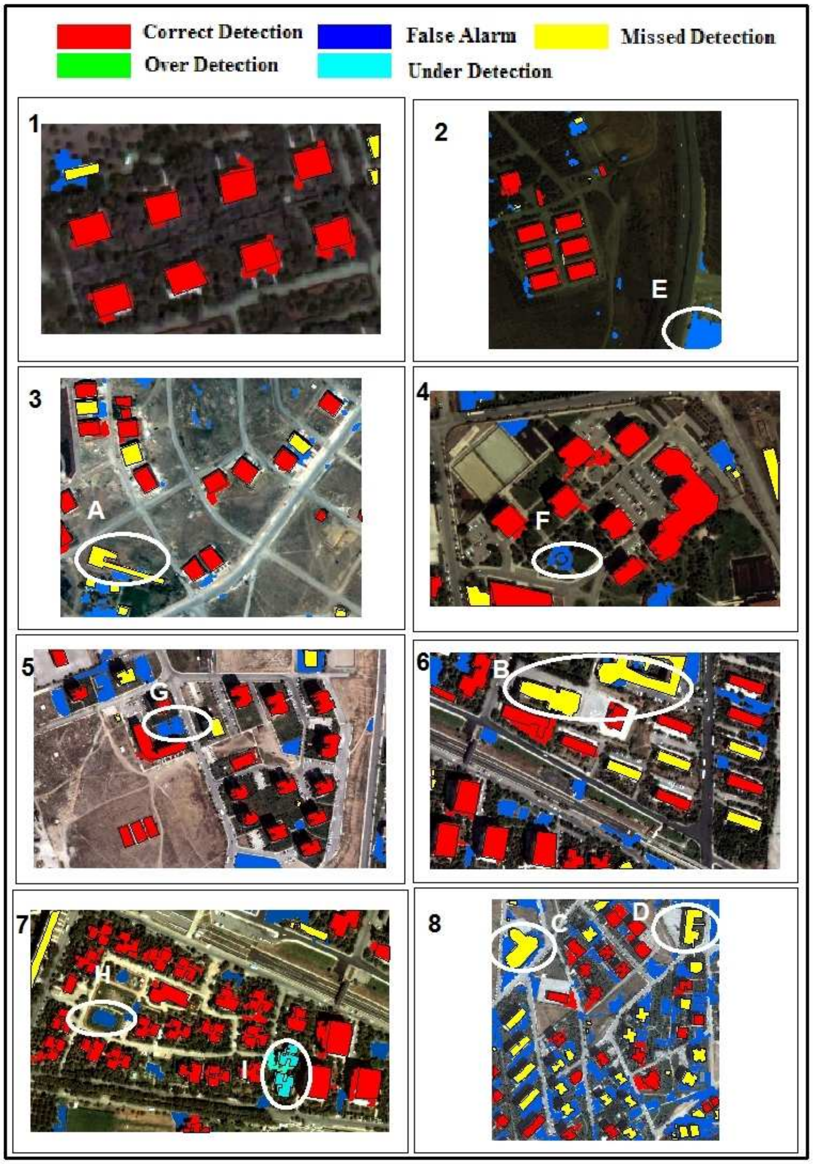


Figure 9
210x297mm (96 x 96 DPI)

Table 1: Pixel-based performance evaluation

Data Set	Number of Pixels			Ratio		%	
	True Positive	False Positive	False Negative	Split Factor	Missing Factor	Percent of Building Detection	Quality Percent
Data1	3820	1386	574	0,36	0,15	86.94	66.09
Data2	13282	9406	538	0.71	0.04	96.11	57.18
Data3	13068	8383	7643	0.64	0.58	63.09	44.92
Data4	21532	8447	3987	0.39	0.19	84.38	63.39
Data5	32156	24882	5310	0.77	0.16	85.83	51.57
Data6	33660	21810	20661	0.65	0.61	61.96	44.21
Data7	42283	24297	4976	0.57	0.12	89.47	59.09
Data8	90257	156270	16712	1.73	0.18	84.37	34.29