ROBUST DETECTION OF BUILDINGS
FROM A SINGLE COLOR AERIAL IMAGE

Ali Özgün Ok

Dept. of Geodetic and Geographic Information Technologies,
Middle East Technical University, Ankara, Turkey 06531 - oozgun@metu.edu.tr

Commission VI, WG VI/4

KEY WORDS: Color Aerial Imagery, Building Detection, Segmentation, Mean-shift, Photometric Quasi-invariants

ABSTRACT:

In this study, a robust methodology for the detection of buildings from a single color aerial image is proposed. The methodology is initialized with the mean-shift segmentation algorithm. Next, a vector-valued canny edge detection algorithm which relies on the photometric quasi-invariant gradients is utilized to detect edges from the color segmented image. Morphological operations are applied to the edge image and two raster datasets are generated from the morphologically reconstructed edge image, (i) the edge pixels that form closed-boundary shapes, and (ii) the edge pixels that do not form closed-boundary shapes. The first dataset, the edge pixels that form closed-boundary shapes, are vectorized using boundary tracing followed with the douglas-peucker simplification algorithm. On the other hand, a minimum bounding convex-hull algorithm followed with gradient vector flow (GVF) snake is used to generate polygons from the second dataset. The polygon results of both datasets are joined together in a unification step. In a final verification stage, two vegetation indices are used to mask out the polygons that belong to healthy vegetated areas. One color (RGB) aerial ortho-image with a resolution of 30 cm is utilized to test the performance of the proposed methodology. Based on the results computed, the edge detector suppressed the shadows and highlight edges with accuracy around 99%. Among the available 251 apartment buildings in the site, 231 of them were almost completely detected. The algorithm provided %73 accuracy for the buildings that are in a neighboring house condition (810 out of 1104 are detected).

1. INTRODUCTION

Aerial imaging is one of the most common and versatile way of obtaining information from the objects on the Earth surface. The information obtained can be quite various and high-quality due to the advanced capabilities of aerial imaging such as the availability of ultra high resolution cameras (film & digital), geometric fidelity involved, broadened spectral sensitivity, permanent recording etc. These special characteristics turned out the aerial imaging one of the most important data input techniques to be used for the object extraction task.

Object extraction process from aerial images is generally interested with objects that are most related with applications such as mapping, site selection and management, military etc. Due to the fact that more than about 50% of the world population lives in urban and sub-urban environments (Fischer et. al., 1998), the extraction of building objects is an issue of great importance for those applications. Until now, a large number of studies have been performed within the context of building extraction. The previous work in the context of aerial imagery until the mid-1999 was reported in an exceptional review performed by Mayer (1999). Later, an extended version of Mayer's review (until the late-2003) bounded in his format was conducted by Ünsalan and Boyer (2005). The knowledge-based image analysis for object extraction and the different aspects of knowledge that can be used for building extraction were reviewed by Baltsavias (2004). The trends followed within the state of art of building extraction can be found in elsewhere (Grün et. al., 1995; Grün et. al., 1998; Baltsavias et. al., 2001). Besides those reviews, the literature survey in this section do not aim to perform a complete overview of the domain but rather to discuss the approaches developed for the extraction of the buildings from single (monocular) aerial images.

The work in this context were mostly preferred the data-driven approach, which relies on the extraction of low-level features so called "lines" and/or "corners". Generally, the extracted lines were the main source to a sequential grouping process that was used to achieve higher level features such as parallels and U structures. The grouping process was concluded after the generation of the highest level features, the rectangles or parallelograms, which were always served as potential building hypotheses. Exceptional attempts to perform this strategy on monocular aerial images can be found in Huertas and Nevatia (1988) and Irvin and McKeown (1989). Shadow analysis has always been considered to be one of the most important clues to find elevated structures, especially buildings. In general, this type of analysis requires the knowledge of the exact sun direction and so the sun angular elevation on a per image basis. With the aid of three parameters, (i) the latitude of the image being investigated, (ii) the knowledge of the time of day that the image was acquired, and (iii) the sun's declination corrected to Greenwich Mean Time, it is possible to compute the required sun angular elevation (Irvin and McKeown, 1989) to infer the shadow parts of the buildings. In an early study, Huertas and Nevatia (1988) utilized shadow casts of buildings to interpret the sides and corners of buildings. In a different study, Irvin and McKeown (1989) used the shadow information not only predict the location and shape of the buildings but also the heights of buildings, whereas Liow and Pavlidis (1990) used them in order to complete the boundary grouping process. To achieve a complete set of building boundaries, Shufelt and McKeown (1993) propose the cooperative-methods paradigm which mainly based on the processing using several separate systems,

which also relied on the shadow information extracted. Regardless of any form of usage, the shadow information derived from the images was extensively used during the verification phase of the potential building hypotheses (Huertas and Nevatia, 1988; Irvin and McKeown, 1989; Liow and Pavlidis, 1990; Shufelt and McKeown, 1993; McGlone and Shufelt, 1994; Lin and Nevatia, 1998; Stassopoulou and Caelli, 2000; Turker and San, 2004; Peng and Liu, 2005). However, the attempts were not limited to the shadow information extracted, for example; McGlone and Shufelt (1994) proposed the vanishing point method, which relies on the projective geometry to extract the buildings. They also utilized a photogrammetric search for vertical lines in image space at roof corner points to obtain height estimates of the building roofs. Five years later, in a comparative study, Shufelt (1999) presented extensive performance evaluation results of four different standalone systems (Pivot (Shufelt, 1996), Build, Build+Shave (Irvin and McKeown, 1989), Vhbuild (McGlone and Shufelt, 1994)) that utilize only single gray-scale aerial images. Lin and Nevatia's (1998) monocular-image based system has substantially increased the type and number of constraints and evidences used for the extraction of buildings. The increase in the number of evidences used, raised the problem of combining evidences from the various sources. For this purpose, Kim and Nevatia (1999) tested a number of methods including Neural and Bayesian networks whereas Stassopoulou and Caelli (2000) utilized probabilistic Bayesian networks to combine a number of different sources of evidences found for potential buildings. However, the results of those studies proved that the extraction of buildings from only a single image were quite problematic. Despite those mentioned significant efforts, achieving acceptable results from a single-image based system is still believed to be limited to certain imaging and testing conditions.

A simple observation from the survey of the literature is that, as it is certain, while aerial images have been rich of color information, little use of this color information was utilized during the detection of edges. The availability of color information in aerial images would give systems ability to find the edges that can not be described in gray-scale images due to too-low contrast between ground, resolution problems, accidental alignments etc. Therefore, this study presents as automated as possible methodology for the detection of building objects from a single color (RGB) aerial image. The utilized vector-valued canny edge detection algorithm which relies on the photometric quasi-invariant gradients has ability to classify the edges into three classes in the color image, (i) shadow-shading geometry, (ii) highlight, and (iii) material transitions. Thus, the non-relevant artificial edges, shadows and highlights, are suppressed without spending any extra effort and the remaining edges are inherently suitable for the formation of the two-dimensional polygon objects.

2. METHODOLOGY

The main steps followed in the proposed methodology are given in Figure 1. The approach is initialized with the mean-shift segmentation algorithm in which irrelevant objects such as cars, small trees, chimneys, solar energy panels etc. for the building detection task are removed. Next, a vector-valued canny edge detection algorithm that relies on the photometric quasi-invariant gradients is utilized to detect edges from the segmented color image. The edge detection algorithm used has ability to classify edges into three classes, (i) shadow-shading geometry, (ii) highlight, and (iii) material transitions, using
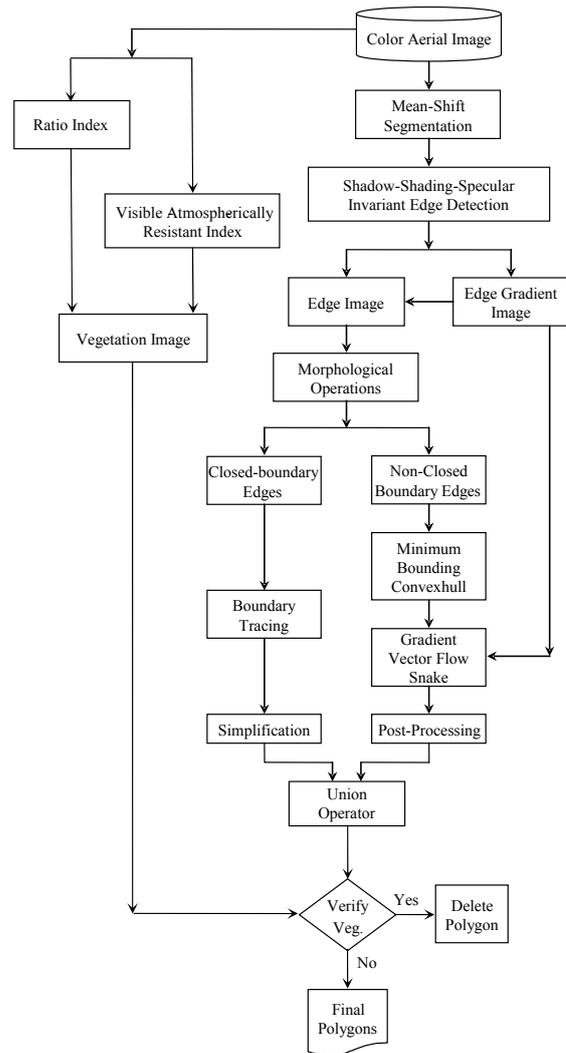
Figure 1. The proposed building detection methodology

photometric quasi-invariants and to suppress the non-relevant artificial edges, i.e. shadows and highlights, existing in the segmented color image. After the detection of the edges, morphological operations such as thinning, closing, and bridging are applied. The morphologically reconstructed edges were then separated into two raster datasets, (i) the edge pixels that form closed-boundary shapes, and (ii) the edge pixels that do not form closed-boundary shapes. The first dataset, the edge pixels that form closed-boundary shapes, are vectorized using boundary tracing followed with the douglas-peucker simplification algorithm. On the other hand, for the second dataset, a minimum bounding convex-hull algorithm followed with gradient vector flow (GVF) snake is used to generate vector polygons. The results of both datasets are unified using a union operator. Finally, the vector polygons that belong to vegetation areas within the unified dataset are removed through a vegetation image which is computed from the original color aerial image using two vegetation indices, (i) a Ratio Index (RI), and (ii) Visible Atmospherically Resistant Index (VARI).

2.1 Mean-Shift Segmentation

In general, the mean-shift based image segmentation is composed of two successive steps that are namely (i) the discontinuity preserving smoothing of the original image, and

(ii) performing the segmentation based on the smoothed image (Comaniciu and Meer, 2002). Discontinuity preserving smoothing technique adjusts the degree of smoothing; i.e., the amount of smoothing significantly reduced in close proximity of the edges. On the other hand, the segmentation is an extension of the discontinuity preserving smoothing algorithm; each pixel is associated with a significant mode of the joint domain density located in its neighborhood, after nearby modes were pruned as in the generic feature space analysis technique. Since an image is composed of pixels with proper gray level, color, or spectral information the incorporation of the spatial coordinates of a pixel into its feature space representation is required. This is accomplished by a joint domain representation which takes into account both the spatial and spectral information of an image. Thus, the multivariate kernel is defined as the product of two radially symmetric kernels and a single bandwidth parameter is allowed for each domain

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \qquad (1)$$

In equation (1) x_s is the spatial part, x_r is the range part of a feature vector, $k(x)$ the common profile used in both two domains, h_s and h_r the employed kernel bandwidths, and C the corresponding normalization constant. In practice, an Epanechnikov or a normal kernel types always provides satisfactory performance, so the user only has to set the bandwidth parameter $h = (h_s, h_r)$ in equation (1), and the minimum number of pixels (area) that allowed within a single region M.

2.2 Color Edge Detection

The low level object extraction performances of the detectors used so far are always limited and mostly defective since they work on a single band or component. To correct this shortcoming, a vector-valued technique which relies on the photometric quasi-invariants is utilized to detect edges from color aerial images. The method was developed by Gevers and Smeulders (1999), and relies on the classification of edges into meaningful classes by means of the derivatives of different color spaces. The classification used is based on the generation of different color space derivatives which are quasi-invariant to specific type of edges, such as shadow-shading, specular, or both shadow-shading and specular. Once these derivatives are accurately found, they are combined using a structure tensor in a color version of the Canny algorithm (Weijer et. al., 2006). In this study, two minor improvements are made to their algorithm, (i) the output of the final gradient map is scaled between zero-and-one before further processing which significantly reduced the remaining noise edges and (ii) a two level hysteresis thresholding is designed to have a better control on the final edge contours. After performing the color Canny algorithm with shadow-shading-specular quasi-invariance, a single gradient image is obtained. The hysteresis thresholding is used to convert the gradient image to a binary edge image which consists of white pixels forming the edges and the black pixels representing the background.

2.3 Morphological Operations

The first step of this part is to bridge the unconnected edge pixels after the color edge detection. This is performed using morphological operations in which each edge pixel in the output image is generated based on a comparison of the corresponding edge pixel in the input image with its very close neighbors. Three morphological operations are successively applied to the original edge image. First, morphological closing (dilation followed with erosion) is applied using a square structuring element of a 3-by-3 matrix. Next, morphological bridging operator is utilized. This operator sets 0-valued pixels to 1 if they have two non-zero neighbors that are not connected. Finally, a thinning operator is used to generate one-pixel wide edges in the morphologically reconstructed edge image.

Although the morphological operations linked the broken edges that are in a very close neighbor, many edges still remain unconnected (non-closed boundary shapes). However, some of the edges in the morphologically reconstructed image inherently resulted in closed-boundary shapes which do not require a particular processing as compared to their non-closed counterparts. To identify and separate these two shape formations, a morphological shrinking operator followed with a cleaning operator is used. The shrinking operator removes pixels so that objects that do not form closed-boundaries shrink to an isolated edge point, and the cleaning operator removes those isolated points from the edge image. Thus, the cleaned edge image is only composed of the edge pixels that are closed-boundary shapes. Finally, the cleaned edge image is subtracted from the previously reconstructed edge image to find the edge pixels that do not form closed-boundary shapes.

2.4 The Construction of the Vector Polygons

The aim of this part is to generate accurate vector polygons from the generated two binary raster datasets, (i) the edge pixels that form closed-boundary shapes, and (ii) the edge pixels that do not form closed-boundary shapes. First, the processing of the closed-boundary shapes is carried out. As previously stated, closed-boundary shapes do not require a particular processing and inherently ready to be an input to a well-known vectorization algorithm, boundary tracing. During the boundary tracing processing, the exterior boundary of each closed-boundary shape is traced and converted to vector polygons. A douglas-peucker simplification algorithm with a threshold ε is used to reduce the number of vertices generated from boundary tracing.

A more complicated processing is involved during the generation of vector polygons from the edge pixels that do not form closed-boundary shapes. Each non-closed-boundary shape in the binary edge image is converted to objects using a method called connected-component labeling. This method assigns a unique integer value to each object based on a selected connectivity measure. Next, for each object, a Convex-Hull algorithm is applied. This algorithm computes the smallest convex polygon that contains the minimum bounding region defined by each object. The convex polygons generated using the convex-hull algorithm for each object is used to initialize the Gradient Vector Flow (GVF) snake. GVF is a dense vector field derived from images by minimizing certain energy functions in a variational framework and a GVF snake is an active contour that uses the GVF field as its external force (Xu and Prince, 1998). The initialization of the GVF snake is performed using the convex polygons; however, the initialization may not be always accurate enough to recover the correct polygons from the binary edge image. Moreover, the hysteresis thresholding applied to the gradient image lost some of the edges available in the gradient image. Although very-close-neighbor information is reconstructed using

morphological processing, there are still many gaps and missing parts in the processed edge image. However, these potential problems are minimized by the nature of the GVF snake itself. First, the GVF snake is insensitive (or less sensitive) to the initialization and has a large capture range. Therefore, the polygons computed with the convex-hull algorithm are mostly sufficient to initialize the GVF snake. Moreover, the GVF field of the snake is computed by using the original gradient image to enhance the progression of the GVF snake. As a result, all the information available in the original gradient image influenced the final configurations of the polygons for each object, and the ability to move the objects' boundary concavities of the GVF snake has resulted in better final configurations. Once the final configurations of all object polygons are computed, a post-processing operation is performed to remove the incorrectly configured parts. This operation mainly includes two criteria; (i) multiple-buffering (margins at a pixel distance) created around each polygon object and the size of the (ii) minimum and (iii) maximum area allowed by a single polygon object. A douglas-peucker simplification algorithm with the same threshold (ε) is used to reduce the number of vertices of the post-processed polygons.

In the last step, the generated vector polygons from the two binary raster datasets are combined to produce a final polygon dataset. This is performed using a simple union operator.

2.5 Verification of the Polygons

In this part, the generated vector polygons that belong to vegetated areas in the color (RGB) aerial image are masked out. It is a well-known fact that a common way to extract vegetated areas is to use various indices which are principally dimensionless radiometric measures that point towards the vegetation. To eliminate healthy green vegetation, many of the indices utilize the apparent inverse relationship between the red and near-infrared bands (Jensen, 2005). However, the aerial image used is only composed of visible bands in which many available indices become useless. Therefore, two different indices each of which utilize only the information obtained from the visible bands are used. The first index used is computed as the ratio between the green reflected radiant flux (ρ_{green}) and the blue radiant flux (ρ_{blue}):

$$RI = \frac{\rho_{green}}{\rho_{blue}} \quad (2)$$

The second index used is proposed by Gitelson et. al. (2002) and called *Visible Atmospherically Resistant Index* (VARI):

$$VARI = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}} \quad (3)$$

The first index expresses the relationship between the green and the blue bands; however, the second index is mostly devoted to the green and red band interaction. The results of both indices are scale to 0-and-255 and converted to binary vegetation images (0-negative evidence and 1-positive evidence) using a certain threshold. Next, two vegetation images a combined using a logical array operator *AND* (&) in a final vegetation image. The generated polygon dataset is overlaid to the vegetation image and the number of positive evidences under each polygon is counted and divided by the total number of pixels (ratio) under that polygon. During the verification, if the ratio is found to be higher than a certain value for a specific polygon, it is automatically labeled as a vegetation object and deleted from the dataset.

3. STUDY AREA AND DATA SETS

The study area selected is over a residential part of Konya, Turkey. The set of raw color (RGB) images was acquired with Zeiss RMK TOP 30 camera with typical photogrammetric overlaps (60% end - 20% side) for mapping. The calibrated focal length of the camera was 305.536 mm and the flying height was approximately 1200 m above the ground. The images were scanned at 14-μm resolution with a photogrammetric film scanner and it corresponds to a final ground sampling distance of approximately 6-cm. The camera was accurately pre-calibrated and passed through a rigorous simultaneous bundle block adjustment procedure. Around 350 3-D vector layers were generated from the available stereo-pairs in a digital photogrammetric workstation environment by qualified human operators. The ortho-image used in this study was generated using a digital terrain model (DTM) which was produced from the specific layers (contours, roads etc.) of the 3-D vector dataset. The area covered by the ortho-image was around 1.5 km^2, and the spatial resolution was determined to be 30 cm during the orthorectification process.

The ortho-image includes buildings with different shapes, sizes, and orientation. In the area, the total number of buildings is 1403. 96.6% of those buildings are residential and the remaining is a mixture of under-construction, government, educational, and religious buildings. Among the available 1403 buildings, 251 of them are residential apartment buildings and mostly built in a regular pattern. Other residential buildings are mostly have heights less than 6 m and are densely neighboring house settlements (detached, semi-detached or terraced buildings).

4. IMPLEMENTATION OF THE METHODOLOGY

The mean-shift segmentation is performed using a stand-alone system (EDISON) developed by Comaniciu and Meer (2002). The remaining processes were all completed in Matlab ® v. 7.5 environment. The parameters used and their values are given in Table 1.

Table 1. Parameter Used and Their Values

Processing Step	Parameter	Value
Segmentation	Spatial-Bandwidth	13x13 pixels
	Range-Bandwidth	14
	Minimum Area	≥ 27 m^2
Shadow-Shading-Specular Edge Detection	Sigma-1	4
	Sigma-2	4
	Threshold-High	≥ 40
	Threshold-Low	≥ 30
Morp. Processing	Structuring Element	3x3 pixels
	Min. Object Length	6 m
GVF Snake	Elasticity	0.05
	Rigidity	0
	Viscosity	1
Post-Processing	Multiple-Buffering	0.3 m
	Minimum Area	≥ 27 m^2
	Maximum Area	≤ 2830 m^2
	Simplification	0.3 m
Vegetation	Threshold	≥ 140
	Ratio Value	≥ 0.75

5. RESULTS AND DISCUSSION

The following accuracy measures proposed in (Lin and Nevatia, 1998) are used to test the quality of the results of the proposed methodology. Five different measures are computed:

- Detection Percentage = 100 x TP / (TP + TN)
- Branch Factor = 100 x FP / (TP + FP)
- Correct Building Pixels Percentage
- Incorrect Building Pixels Percentage
- Correct Non-building Pixels Percentage

In the first two measures, TP (True Positive) denotes the number of buildings that exist in the reference vector dataset and detected by the method, TN (True Negative) denotes the number of buildings that only exist in the reference vector dataset but not detected by the method, and FP (False Positive) denotes the number of buildings detected by the method but not exist in the reference vector dataset. The last three measures are determined by counting the correct building and non-building pixels (Lin and Nevatia, 1998).

In this study, the percentages of the detection are also computed based on four different building types, (i) apartment, (ii) detached or terraced, (iii) under-construction, and (iv) other. It is important to emphasize that a building is considered to be detected if any part of the building is detected; however, for each building, the detected portion is calculated and classified into two detection categories, (i) complete or (ii) partial detection. A building is classified as complete detection if the detected region pixels cover more than 75% of its area (Fradkin et. al., 2001); otherwise it is classified as partial detection. In order to give an idea about how the buildings are partially detected, they are further classified into three sub-classes (see Table 2 for the portions of the sub-classes).

Table 2 summarizes the results of the proposed methodology based on four different types of buildings. The detection rates among the different types of buildings ranged between 14% and 100%. The best detection results (100%) are achieved for the apartment buildings. The quality of the detection of the apartment buildings is very high since most of them (231 out of 251) are completely detected and only a few of them (9 out of 251) are detected with a portion of less than 25%. Figure 2 demonstrates a small part of the study area used. It is very important to state that the locations of the apartment buildings are successfully found independent from their size, shape and orientation. In addition, note that the shadow edges did not affect the results (except for a small polygon generated close to the bottom-right). Most of the buildings in the study area (1104 out of 1403) are densely neighboring house settlements

(detached, semi-detached or terraced buildings). For those buildings, the algorithm provided 73% detection rate. Among the detected dense buildings, approximately 55% percent of them are completely detected. The worst results are achieved for the buildings that are under-construction at the time of image acquisition. For those buildings, the detection rate was computed to be 14%. In fact, this is an *expected* result due to two explicit reasons. First, some of those buildings are in their very early construction stage. Therefore, there is not enough color difference between the building object and its background. Thus, an over-segmentation was observed for those buildings. The second reason is due to the color Canny edge detection used. Most of the under-construction buildings are in a condition of very high reflectance. Since the color edge detection used suppresses the specular edges, the bright nature of those buildings is treated as specular edges and inherently suppressed by the edge detector.

Table 3 shows the overall results computed for the whole ortho-image. The overall detection percentage (76.8 %) is slightly

Table 3. Overall results of the proposed methodology

	Detection Percentage tp / (tp+tn)	Branch Factor fp / (tp+fp)	Correct building pixels	Incorrect building pixels	Correct non-building pixels
Ortho-Image	76.8%	9.19%	71.46%	24.75%	96.11%

Figure 2. A small part of the ortho-image. The red polygons show the available reference data, the yellow polygons shows the results of the proposed methodology.

Table 2. The detection results of the proposed methodology based on the types of buildings

Detection	Detected Portion of Buildings	Building Type			
		Apartment	Detached or Terraced	Under-Construction	Other
Complete	($\geq 75\%$)	231	602	3	8
Partial	($50\% \leq x < 75\%$)	6	30	-	3
	($25\% \leq x < 50\%$)	5	24	-	1
	($< 25\%$)	9	154	1	-
Non-Detected		-	294	24	8
Detected Total		251	810	4	12
Total Number of Buildings		251	1104	28	20
Detection Percentage (%)		*100%*	*73%*	*14%*	*60%*

higher than the detection percentage computed for the densely neighboring house settlements (detached, semi-detached or terraced buildings). This is not surprising because the buildings in the study area are mostly dominated by the detached, semi-detached or terraced buildings. It is also evident that the computed percentage for the incorrect building pixels is relatively higher than expected (24.75%). Actually, this result is mostly related with the verification step. Only a single verification strategy that aims to remove the polygons belonging to the *healthy* vegetated areas is performed. However, more rigorous verification strategies which take into account different parameters (color information, shadows, texture etc.) can be used to improve the percentage computed for the incorrect building pixels. Finally, as can be expected to be high as most pixels in an image are non-building pixels, the percentage for the correct non-building pixels are computed to be 96.11%.

6. CONCLUSIONS AND FUTURE WORK

In this study, an automated and robust methodology for detecting buildings from a single color aerial image is presented. The methodology is tested for a large number of buildings with different shapes, sizes, and types in a complex environment. If the *number* and *diversity* of the buildings in the study area are taken into consideration, it is believed that the 2-D detection results obtained from the proposed methodology is similar to or better than the results stated from other researchers. As an immediate future work, the methodology requires more sophisticated verification strategies to decrease the number of building pixels that are incorrectly labeled. The modification of the methodology to handle stereo and/or multiple color aerial images remains a topic for future research and may further improve the detection rates.

7. REFERENCES

Baltsavias, E. P., Gruen, A., and Van Gool, L., 2001, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, Balkema.

Baltsavias, E. P., 2004, Object Extraction and Revision by Image Analysis Using Existing Geodata and Knowledge: Current Status and Steps Towards Operational Systems, *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, pp. 129– 151

Comaniciu, D., and Meer, P., 2002, Mean Shift: A Robust Approach toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24(5), pp. 603-619.

Fischer, A., Kolbe, T. H., Lang, F., Cremers A. B., Förstner, W., Plümer, L., and Steinhage, V., 1998, Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D, *Computer Vision and Image Understanding*, 72(2), pp. 185-203.

Fradkin, M., Maitre, H., Roux, M., 2001, Building Detection from Multiple Aerial Images in Dense Urban Areas, *Computer Vision and Image Understanding*, 82, pp. 181-207.

Gevers, T., and Smeulders A. W. M., 1999, Color-based Object Recognition, *Pattern Recognition*, 32, pp. 453-464.

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D., 2002, Novel Algorithms for Remote Estimation of Vegetation Fraction, *Remote Sensing of Environment*, 80, pp.76-87.

Grün, A., Kübler, O., and Agouris, P., 1995, Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhauser, Basel, Switzerland.

Grün, A., Baltsavias, E. P., and Henricsson, O., 1998, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhauser, Basel, Switzerland.

Huertas, A., and Nevatia, R., 1988, Detecting Buildings in Aerial Images, *Computer Vision, Graphics, and Image Processing*, 41, pp. 131-152.

Irvin, R. B, and McKeown, D. M., 1989, Methods for Exploiting Relationship between Buildings and Shadows, *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6), pp. 1564-1575.

Jensen, J. R., 2005, *Introductory Digital Image Processing: A Remote Sensing Perspective - 3rd ed.*, Prentice Hall, pp. 310-323.

Kim, Z., and Nevatia, R., 1999, Uncertain Reasoning and Learning for Feature Grouping, *Computer Vision and Image Understanding*, 76(3), pp. 278-288.

Lin, C., and Nevatia, R., 1998, Building Detection and Description from a Single Intensity Image, *Computer Vision and Image Understanding*, 72(2), pp. 101-121.

Liow, Y.-T., and Pavlidis, T., 1990, Use of Shadows for Extracting Buildings in Aerial Images, *Computer Vision, Graphics, and Image Processing*, 49, pp. 242-277.

Mayer, H., 1999, Automatic Object Extraction from Aerial Imagery - A Survey Focusing on Buildings, *Computer Vision and Image Understanding*, 74(2), pp. 138-149.

McGlone, J. C., and Shufelt, J. A., 1994, Projective and Object Space Geometry for Monocular Building Extraction, Proceedings IEEE Conference Computer Vision and Pattern Recognition, Vol. 2486, pp. 25-36.

Peng, J., and Liu, Y. C., 2005, Model and Context-driven Building Extraction in Dense Urban Aerial Images, *International Journal of Remote Sensing*, 26(7), pp. 1289-1307.

Shufelt, J. A., and McKeown, D., 1993, Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery, *Computer Vision, Graphics, and Image Processing*, 57(3), pp. 307-330.

Shufelt, J. A., 1996, Exploiting Photogrammetric Methods for Building Extraction in Aerial Images, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, B6/S, pp: 74-79.

Shufelt, J. A., 1999, Performance Evaluation and Analysis of Monocular Building Extraction from Aerial Imagery, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21(4), pp. 311-326.

Stassopoulou, A., and Caelli, T., 2000, Building Detection using Bayesian Networks, *International Journal of Pattern Recognition and Artificial Intelligence*, 14(6), pp. 715-733.

Turker, M., and San, B. T., 2004, Detection of collapsed buildings caused by the 1999 Izmit, Turkey earthquake through digital analysis of post-event aerial photographs, *International Journal of Remote Sensing*, 25(21), pp. 4701-4714.

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*, 98, pp. 423-461.

Weijer, J., Gevers, T., and Smeulders, A.W.M., 2006, Robust Photometric Invariant Features from the Color Tensor, *IEEE Transactions on Image Processing*, 15(1) pp. 118-127.

Xu, C., and Prince, J. L., 1998, Snakes, Shapes, and Gradient Vector Flow, *IEEE Transactions on Image Processing*, 7(3), pp. 359-369.

8. ACKNOWLEDGEMENTS