A SHADOW BASED TRAINABLE METHOD FOR BUILDING DETECTION IN SATELLITE IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHMET DİKMEN

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

AUGUST 2014

Approval of the thesis:

A SHADOW BASED TRAINABLE METHOD FOR BUILDING DETECTION IN SATELLITE IMAGES

submitted by **MEHMET DİKMEN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, **Graduate School of Natural and Applied Sciences** _____

Prof. Dr. Ahmet Coşar
Head of Department, **Computer Engineering Dept.** _____

Prof. Dr. Uğur Halıcı
Supervisor, **Electrical and Electronics Engineering Dept., METU** _____

Examining Committee Members:

Prof. Dr. Mehmet Lütfi Süzen _____
Geological Engineering Dept., METU

Prof. Dr. Uğur Halıcı _____
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. İlkay Ulusoy _____
Electrical and Electronics Engineering Dept., METU

Assist. Prof. Dr. Emre Sümer _____
Computer Engineering Dept., Baskent Univ.

Assist. Prof. Dr. Mustafa Sert _____
Computer Engineering Dept., Baskent Univ.

Date: 21 . 08 . 2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Mehmet Dikmen

Signature :

ABSTRACT

A SHADOW BASED TRAINABLE METHOD FOR BUILDING DETECTION IN SATELLITE IMAGES

Dikmen, Mehmet
Ph.D., Department of Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. Uğur Halıcı

August 2014, 90 pages

The purpose of this thesis is to develop a supervised building detection and extraction algorithm with a shadow based learning method for high-resolution satellite images. First, shadow segments are identified on an over-segmented image, and then neighboring shadow segments are merged by assuming that they are cast by a single building. Next, these shadow regions are used to detect the candidate regions where buildings most likely occur. Together with this information, distance to shadows towards illumination direction and spectral properties of segments are used to classify them as belonging to a building or not. Then, a resegmentation is performed to extract building patches by merging only the neighboring segments, which are classified as building. Next, a postprocessing step is implemented to eliminate some false building patches. Finally, a one class modeling approach was introduced to refine extracted building patches.

The approach was tested on several Google Earth images of varying characteristics in order to examine the effects of the change in illumination direction, shadow amount and building variety (size, shape, density, etc.). The results were examined by both pixel and object based performance evaluation methods. Best results were obtained on images having relatively shorter shadows and captured almost at the nadir. Best quality for the extracted patches and the least false detections were also observed in the same case.

Keywords: Building detection, image classification, image segmentation, one class modeling

ÖZ

UYDU GÖRÜNTÜLERİNDE BİNA TESPİTİ İÇİN GÖLGE TABANLI EĞİTİLEBİLİR BİR YÖNTEM

Dikmen, Mehmet
Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü
Tez Yöneticisi: Prof. Dr. Uğur Halıcı

Ağustos 2014, 90 sayfa

Bu tezin amacı, yüksek çözünürlüklü uydu görüntüleri için gölge tabanlı bir öğrenme yöntemi kullanan güdümlü bir bina tespit ve çıkarım yöntemi geliştirmektir. İlk olarak, aşırı-bölütlenmiş bir görüntü üzerindeki gölge bölütleri belirlenir, sonra da tek bir binaya ait olduğu kabul edilen ve birbirine komşu olan gölge bölütleri birleştirilir. Daha sonra, bu gölge bölütleri kullanılarak binaların olası yerleri tespit edilir. Bu bilginin yanında, bölütler, aydınlatma doğrultusundaki gölgeye uzaklıkları ve spektral özellikleri de kullanılarak binaya ait ya da değil olarak sınıflandırılır. Sonra, bina parçalarını çıkarmak için yeniden bölütleme gerçekleştirilerek, sadece bina olarak sınıflandırılmış birbirine komşu bölütler birleştirilir. Devamında, hatalı olarak bina şeklinde sınıflandırılmış yerlerin elenmesi için bir son işlem aşaması gerçekleştirilir. Son olarak, tek sınıflı bir modelleme yaklaşımı getirilerek çıkarılan bina parçaları düzeltilir.

Yaklaşım, aydınlatma yönündeki değişim, gölge miktarı ve bina çeşitliliğinin etkilerinin (büyüklük, şekil, yoğunluk, vb.) incelenebilmesi için farklı karakteristiklerdeki Google Earth görüntülerinde test edilmiştir. Sonuçlar, hem piksel hem de nesne tabanlı performans değerlendirme yöntemleri ile incelenmiştir. En doğru sonuçlara, diğerlerine göre daha kısa gölgelere sahip ve tam tepeden çekilmiş görüntüler üzerinde ulaşılmıştır. Yine aynı durum için, elde edilen bina parçalarının en düzgün olduğu ve en az hata ile tespit edilebildiği gözlemlenmiştir.

Anahtar Kelimeler: Bina tespiti, görüntü sınıflandırma, görüntü bölütleme, tek sınıflı modelleme

ACKNOWLEDGEMENT

I would like to express my deepest gratitude and appreciation to my supervisor Prof. Dr. Uğur Halıcı who inspired, encouraged, guided and supported me at all levels of this study. Her intellect, attitude to research and clarity of knowledge were a great source of inspiration for me. Special thanks are also due for all the profitable brainstorming sessions and critical remarks she provided.

I would also like to thank my supervising and examining committee members, Prof. Dr. Mehmet Lütfi SÜZEN and Assoc. Prof. Dr. İlkay ULUSOY whose support and suggestions made great contributions to this work. Their insightful comments helped me to improve my PhD research and finish my thesis.

The greatest thanks go to my family members for their infinite support and patience.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVATIONS

CART: Classification and Regression Tree

DEM: Digital Elevation Model

DMP: Differential Morphological Profile

DSM: Digital Surface Model

DTM: Digital Terrain Model

FAST: Features from Accelerated Segment Test

FN: False Negative

FP: False Positive

GA: Genetic Algorithm

GIS: Geographical Information System

GMSR: Gradient Magnitude based Support Region

HSI: Hue Saturation Intensity

HSV: Hue Saturation Value

LDA: Linear Discriminant Analysis

LIDAR: Light Detection and Ranging

MPP: Marked Point Process

MRF: Markov Random Field

MUSM: Motion based Unsharp Masking

nDSM: Normalized Digital Surface Model

NDVI: Normalized Difference Vegetation Index

NIR: Near Infrared

PAN: Panchromatic

PCA: Principal Component Analysis

QDA: Quadratic Discriminant Analysis

RBF: Radial Basis Function

RGB: Red Green Blue

RoI: Region of Interest

SAR: Synthetic Aperture Radar

SEaTH: SEparability and THreshold

SIFT: Scale Invariant Feature Transform

SVM: Support Vector Machine

TN: True Negative

TP: True Positive

USM: Unsharp Masking

VHR: Very High Resolution

CHAPTER 1

INTRODUCTION

This thesis deals with the detection of buildings and delineation of their boundaries for three-band (i.e., RGB) satellite images by the use of the shadow cue. In the first section, problem definition and motivations are given. The following section includes the objectives and contributions of the study. Finally, chapters of the thesis are summarized as an outline.

1.1 Problem Definition and Motivations

Due to its wide usage on many different applications such as urban planning, urban growth monitoring, change detection, digital map production; the detection of buildings from aerial and satellite imagery has been an important research problem over the past decades. In addition, it is vital for the assessments for disaster response and early recovery, especially when a large geographic region is affected. Therefore, required time for obtaining the information needed is also an important criterion. Although manual extraction of buildings from remotely sensed images can be done very accurately with high reliability, automated methods have emerged to obtain results in a much quicker way and as accurate as possible without the need of qualified people. Thus, all these needs and necessities inspired the researchers to develop many different solutions which made this topic very popular for remote sensing applications.

Although there has been made a great progress recently in automated detection of buildings (Dahiya et al., 2013), it is still a difficult problem due to its limitations and uncertainties. Even at the start of remote sensing there are uncertainties due to the noisy sensor measurements with limited accuracy originating from the radiometric resolution of the sensor. Choosing right methods for data compression, filtering and feature extraction are the other steps to deal with those uncertainties. The reason is that the wrong choice of a method, for example using an average filter with a large kernel size, may modify the input data greatly so that some of the vital information may be lost, thus, reducing the classification potential dramatically.

In addition to uncertainties and limitations explained previously, remotely sensed images are generally very complex and contain a large variety of objects which makes the detection of a specific target a very challenging task. On the other hand, detection of buildings has its own difficulties since there exists a large variety of building shapes to be considered and buildings may have inconsistent spectral properties due to the materials used in their construction. Moreover, in Very High Resolution (VHR) imagery, the increase in the spatial resolution causes more complex structure appearances since the resolution is high enough to detect small distractions both inside and outside of the structures. Nonetheless, the increase in spatial resolution provides finer detail and gives the ability to detect smaller structures as well which allows the extraction of the boundary of a target more precisely. Additionally, buildings in a scene may differ greatly in size. Therefore, a reliable method is needed to extract any building which is also manually detectable regardless of the spatial resolution of the input imagery.

1.2 Objectives and Contributions

The main objective of this study is to detect and extract the boundaries of buildings from a three-band satellite imagery. Supported objectives are listed as follows:

- To present a methodology solely depending on RGB information.

- To develop a trainable resegmentation approach specifically designed for the detection of building patches.

- To present an approach to detect possible locations of buildings with the aid of the shadow cue.

- To develop a relaxation algorithm for the extracted building patches.

Regarding those objectives listed above, the proposed methodology involves following contributions:

- No any other supportive information other than RGB, such as height and infra-red was used. Although those two data sources make it possible to present more robust solutions, using only RGB information allows the usage of freely available Google Earth images. Due to their easy availability and widespread usage, an algorithm which is capable of running on Google Earth images improves its usability and applicability.

- The resegmentation approach is adapted to work with a supervised classification algorithm, thus making it trainable and usable with any other classification algorithm. In this study, the algorithm was trained to merge regions in an over-segmented image to extract building patches. However, it can be easily modified to extract any other target by training the features of the selected target object in the image.

- An automated method was developed to identify possible locations of buildings by using only shadow cue. Even though those locations are used as a feature to train resegmentation model, they can be used as a supporting input for any other solution designed for building detection. Since the method does not require RGB information, it can also be used for single band images.

- A fully automated one-class modeling procedure was introduced to refine the extracted building patches.

1.3 Organization

The rest of this thesis is organized as follows: The second chapter provides background information and the literature review about building detection. The third chapter describes the overall methodology, including image segmentation, shadow detection, feature extraction, resegmentation, post-processing and relaxation algorithm used to refine the results of the resegmentation step. Chapter 4 covers the experimental results and performance evaluation. In addition, discussions of the results are given in detail by specifying where the approach fails and where it is most successful. Finally, in the last chapter, the study is concluded and possible future works are explained.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

In this chapter, past studies about building detection problem are introduced and necessary background information is described. In the first section, an introduction covering available data sources and the choice of processing unit (i.e., pixel or object) are discussed. Then, background information about the methods used for image segmentation and classification are described. Finally, past studies in the literature are presented.

2.1 Introduction

In the literature, for the building detection problem, many data sources were used. Although satellite images are the cheapest; airborne images and height data appeared to be important data sources. Airborne images can be said as one of the oldest data source to detect buildings so that it was widely used by many researchers (Lin et al, 1994; Lin and Nevatia, 1998; Gerke et al, 2001; Tseng and Wang, 2003; Peng and Liu, 2005). One of the biggest advantage of airborne images is that they can have higher spatial resolutions when compared to satellite images. Additionally, there is no need to wait for a specific time to acquire a new imagery unlike satellites which revisits the same scene periodically. However, using airborne images has some drawbacks like high cost when compared to satellite images and long logistics preparation. In addition, airborne data is usually analogue which makes it to process harder.

Using height information became more popular after laser scanner technologies emerged. The typical height data used in the literature are Digital Terrain Models (DTM), Digital Surface Models (DSM), and Digital Elevation Models (DEM) which are usually generated from either stereo airborne, satellite, Synthetic Aperture Radar (SAR) images or Light Detection and Ranging (LIDAR) data (Rottensteiner and Briese, 2002; Jaynes et al, 2003; Simonetto et al, 2005; Zhang et al, 2006; Lee et al, 2008; San, 2009; Lafarge et al, 2010; Sumer, 2011). Although using height data significantly improves detection rate of buildings, this technology has disadvantages of high cost and poor data availability.

Building detection problem can be thought as a binary image classification problem. A traditional solution is to use pixel-based methods in which the processing unit is the pixel itself. Since, these methods make use of only the spectral information (i.e., pixel values) and some limited texture information (i.e., the statistics derived from the immediate neighborhood of a pixel), more indicative spatial information such as texture, shape and context can not be used. It is also shown that traditional pixel-based classification fails to produce accurate results due to the increased variability introduced by VHR images (Hay and Castilla, 2006). As a consequence of these, object-based methods have emerged which work on objects rather than single pixels.

Objects are generally extracted by image segmentation as homogenous, and contiguous group of pixels. The main advantage of this approach is that it offers the use of much more additional information such as object size, range of size, shape, texture, neighborhood, directionality, mean distance, minimum distance and maximum distance between objects to characterize objects (Blaschke, 2003). A more detailed information and a comparison between pixel and object based image classification can be found in the study of Gao and Mas, 2008. In that study, the authors compared the performance of these two different classification approaches on images with varying spatial resolutions and showed that object-based methods had better results, especially in higher resolutions. Moreover, for

larger images, processing each individual pixel becomes computationally more expensive than processing individual objects. Despite these advantages, image segmentation is a challenging task. This is because, image objects are usually segmented in more than one region (i.e., over-segmentation) or a single region may contain more than one object (i.e., under-segmentation) in the process. Resegmentation is one of the techniques to minimize these problems which is applied on a over-segmented image by applying some merging rules specifically for the object to be extracted.

2.2 Background Information

2.2.1 Over-segmentation with Mean Shift

This study introduces an object based solution; therefore, the processing unit is image objects rather than pixels. A general way is to employ an image segmentation. However, an ideal segmentation is almost impossible due to the over and under segmentation problems so that a post-processing is required afterwards, since some segments will need splitting and others will need merging. One way to deal with these problems is first to perform an over-segmentation and then, apply only merging where necessary. To that end, initially over-segmentation is employed in this study using the mean-shift (MS) algorithm (Comaniciu and Meer, 2002).

MS is a general non-parametric feature space analysis technique to delineate the clusters in it. The procedure is based on the kernel density estimation. For n data points x_i (i=1,2,...,n) in d-dimensional space R^d, the multivariate kernel density estimator $f(x)$ with kernel $K(x)$ and bandwidth parameter h is defined by:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right) \tag{2.1}$$

The procedure aims to achieve only a radially symmetric kernel to define kernel profile k(x) for $x \geq 0$:

$$K(x) = c_{k,d} \; k(\|x\|^2) \tag{2.2}$$

where $c_{k,d}$ is the normalization constant which integrates $K(x)$ to one. Introducing this profile notation, Equation 2.1 can be rewritten as:

$$\hat{f}_{h,K}(x) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \tag{2.3}$$

Next, by taking gradient of the density estimator in Equation 2.3, the density gradient estimator is obtained as:

$$\hat{\nabla} f_{h,K}(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) k'\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \tag{2.4}$$

Assuming that the derivative of the kernel profile k exists for all $x \geq 0$ except for a finite number of points, the new profile g(x) can be defined as:

$$g(x) = -k'(x) \tag{2.5}$$

Then, using this profile, the kernel can be redefined as G(x):

$$G(x) = c_{g,d} \; g(\|x\|^2) \tag{2.6}$$

Now, rearranging the terms after introducing g(x) into Equation 2.4, the gradient of the density estimator becomes:

$$\hat{\nabla} f_{h,K}(x) = \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \right] \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right] \tag{2.7}$$

From Equation 2.7, the first term in brackets is proportional to the density estimate at x computed with the kernel G:

$$\hat{f}_{h,G}(x) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right) \tag{2.8}$$

In Equation 2.7, the second term in brackets is the mean-shift that is the difference between the weighted mean (kernel G is used for weights) and the center of kernel window x:

$$m_{h,G}(x) = \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right] = \frac{1}{2} h^2 c \frac{\hat{\nabla} f_{h,K}(x)}{\hat{f}_{h,G}(x)} \tag{2.9}$$

Examining Equation 2.9, it can be concluded that the mean shift vector computed with kernel G is proportional to the normalized density gradient estimate obtained with kernel K (Comaniciu and Meer 2002). Therefore, its direction is always towards the stationary points to obtain the maximum increase in the density. In addition, the normalization term in the formulation makes the procedure a gradient ascent method, since the steps of the mean shift are large in the regions of low density values and small near the local maxima. For image segmentation, the procedure starts with a pixel and defines a window derived from the kernel with the pixel in the center. Then, mean shift vector is computed by calculating the mean inside the window. Next, the center of the window is shifted to the

mean. The procedure repeats until it converges by getting a stationary point where the gradient becomes zero.

To apply the procedure to the image segmentation, the kernel function should be adapted accordingly. An image consists of two independent domains, namely the spatial (2D lattice of pixels) and spectral domain. Therefore, the kernel function should be rewritten as the product of two radially symmetric kernels, one for each of the domains:

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{2.10}$$

where k(x) is the common profile for both domains, C is the normalization constant, x^s and x^r are the spatial and the range (color) parts of the feature vector, respectively. h_s and h_r are the kernel bandwidths which are the only parameters needed to be specified when normal kernel is used. These parameters control the resolution of the mode detection, thus, they are closely related to the size of the objects to be detected. As a result, small values of these parameters with respect to the object size lead to over-segmentation, while larger values cause under-segmentation.

2.2.2 Logistic Regression

Logistic Regression is a probabilistic classification model which creates a non-linear relationship between a dependent variable and several explanatory variables (James et al., 2013). The model is binomial, thus, it is used to produce binary outcomes. On the other hand, the explanatory variables can be either continuous or binary. Quantitatively, the relationship between the occurrence of the outcome and the explanatory variables can be defined as:

$$P = \frac{e^Z}{1+e^Z} \tag{2.11}$$

where P is the estimated probability of an event to occur or the estimated probability of a segment to belong a building. Since Z value varies from -∞ to +∞, the probability varies from 0 to 1 on an S-shaped curve. Z is defined as:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 ..... \beta_n X_n \tag{2.12}$$

where β_0 is the constant, β_i's (i = 1, 2, ..., n) are the coefficients and X_i's (i = 1, 2, ..., n) are the explanatory variables which are the features for the problem addressed in this thesis.

2.2.3 Discriminant Analysis

Discriminant analyses find a group membership by maximizing the ratio of the separation between the two groups. Two variations were used in this thesis. The first one is the Linear Discriminant Analysis (LDA) which searches for a linear combination of the explanatory variables by maximizing the ratio of between-class variance to the within-class variance (Venables and Ripley, 2002). LDA projects all data onto a line w and from all possible lines; it selects the one that maximizes the separability. The objective function is given as:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{2.13}$$

where S_B is the *between classes scatter matrix* and S_W is the *within classes scatter matrix* which are defined as covariance matrices using:

$$S_B = \sum_{k=1}^{c} (m_k - x)(m_k - x)^T \tag{2.14}$$

$$S_W = \sum_{k=1}^{c} \sum_{i \in k} (x_i - m_k)(x_i - m_k)^T \tag{2.15}$$

where c is the number of classes, m_k is the mean of the class k and x represent the observations. For two classes, S_B is defined as:

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \tag{2.16}$$

where m_1 and m_2 are the means of the class 1 and class 2, respectively. The solution w^* is given as:

$$w^* = \arg\max J(w) = S_W{}^{-1}(m_1 - m_2) \tag{2.17}$$

In general, LDA uses a linear classifier to separate two classes of objects by a linear surface whereas in QDA, a quadratic surface is used that best separates the classes (Hastie et al., 2008). QDA finds a group membership consisting of a square n x n matrix (n: number of explanatory variables) and a linear combination of these variables such that:

$$Q = x^T Ax + \beta^T x + c \tag{2.18}$$

where A is the n x n coefficient matrix, β is the linear coefficient vector and c is the constant.

2.2.4 Support Vector Machine

Support Vector Machine (SVM) is a supervised model used for binary classification which relies on finding the optimum separation surface between classes which maximizes the margin between them (Cortes and Vapnik, 1995). The separating surface (also called the hyperplane) is usually obtained after projecting data into an higher dimension feature space by using a kernel function. Many different kernels exist but the two commonly used kernels are the polynomial and Radial Basis Function (RBF) kernels. The polynomial kernel is defined as:

$$K(u,v) = (u.v+1)^p \tag{2.19}$$

where p indicates the degree of polynomial for a point (u, v). The kernel used in this analysis is the RBF which can handle solving non-linear problems and works well in most cases (ENVI Manual, 2006). The definition of RBF is given by:

$$K(u,v) = \exp(-\|u-v\|^2) \tag{2.20}$$

For a set of points *x,* the hyperplane is defined by:

$$w.\cdot x + b = 0 \tag{2.21}$$

where w is the normal vector to the hyperplane and b is the constant. The support vectors, which are described as the subset of training samples on the margin, lie on two hyperplanes parallel to the one in Equation 2.21:

$$w.\cdot x + b = 1 \text{ and } w.\cdot x + b = -1 \tag{2.22}$$

Maximization of the margin leads to the following optimization problem:

$$\min\left\{\frac{1}{2}\|w\|^2\right\} \quad \text{with } y_i\left(w.x+b\right)\geq 1, \quad i=1,2,...,N \tag{2.23}$$

where N is the number of training samples and y_i is the class label which takes the value of 1 or -1.

2.2.5 AdaBoost

Boosting is a machine learning technique which combines a set of weak classifiers to create a stronger one. AdaBoost (Freund and Schapire, 1997) is the first practical algorithm which is developed for this purpose. In this thesis, Modest AdaBoost algorithm (Vezhnevets and Vezhnevets, 2005) is used which has a better generalization capability when compared with the earliest variations Gentle and Real AdaBoost. The description of the algorithm is given by Vezhnevets and Vezhnevets (2005) as:

1. Given the training data (x_i,y_i), x_i: input vector and y_i: its class label (-1 or +1) for $i=1,...,N$; initialize data weights as $D_0(i) = 1/N$.
2. Starting with iteration $m = 1$, while $m \leq M$ and $f_m \neq 0$
   a. Train weak classifier hm(x) using distribution D(i) by weighted least squares:

      $$h_m(x) = \arg\min_h\left(\sum_{i=1}^{N} D_m(i)\cdot(y_i - h(x_i))^2\right)$$

   b. Compute inverted distribution with normalizing constant $\overline{Z}_m$:

      $$\overline{D}_m = (1-D_m(i))\cdot\overline{Z}_m$$

   c. Compute the accuracy of predicting class labels:

      $$P_m^{+1} = P_{D_m}\left(y=+1\cap h_m(x)\right)$$
      $$P_m^{-1} = P_{D_m}\left(y=-1\cap h_m(x)\right)$$

      accuracy on data which are correctly classified in previous steps

$$\overline{P}_m^{+1} = P_{\overline{D}_m}\left(y = +1 \cap h_m(x)\right)$$

$$\overline{P}_m^{-1} = P_{\overline{D}_m}\left(y = -1 \cap h_m(x)\right)$$

d. Set

$$f_m(x) = \left(P_m^{+1}(1-\overline{P}_m^{+1}) - P_m^{-1}(1-\overline{P}_m^{-1})\right)(x)$$

and update the distribution

$$D_{m+1}(i) = \left(D_m(i)\exp(-y_i f_m(x_i))\right)/Z_m$$

3. Construct the final classifier: $sign\left[\sum_{i=1}^{M} f_m(x)\right]$

For the implementation, GML AdaBoost Matlab Toolbox v.0.3 (2011) is used which is developed by Graphics and Media Lab in Department of Computational Mathematics and Cybernetics of Lomonosov Moscow State University. The toolbox uses Classification and Regression Trees (CART) as weak learners. CART is a decision tree analysis introduced by Breiman et al. (1984). The analysis utilizes a binary tree graph in which the leaves represent the classification result and nodes represent some predicate. Each branch of the tree is marked as either true or false. The classification is performed by traversing the tree to reach leaves. In the analysis, the default values for number of tree splits and maximum boosting iterations are used, which are 16 and 100, respectively.

2.2.6 One-class SVM

One-class SVM is a distribution estimation technique which was proposed by Scholkopf et al. (2001). Suppose that $x_1, x_2, \ldots, x_n$ are the training vectors of a single class, the technique attempts to determine whether a new sample x_{n+1} belongs to the same class. Let w be the weight (normal) vector and p be the offset parameter of the hyperplane in the feature space, the primal problem of one-class SVM is defined by:

$$\min_{w,\xi,\rho}\left\{\frac{1}{2}w^T w-\rho+\frac{1}{\nu n}\sum_{i=1}^{n}\xi_i\right\} \tag{2.24}$$

Subject to:

$$w^T\phi(x_i)\geq\rho-\xi_i,\quad \xi_i\geq 0,\ i=1,2,...,n \tag{2.25}$$

where $\phi(x_i)$ maps x_i into a higher-dimensional space and ξ_i are the slack variables to penalize the outliers. $\nu\in(0,1]$ is a parameter introduced by Scholkopf et al. (2001) which is proved to be a lower bound of the fraction of support vectors and an upper bound on the number of training errors. In other words, it controls the tradeoff between incorporating outliers against maximizing the margin of separation from the origin in feature space.

Suppose that the kernel function is:

$$K(x_i,x_j)=\phi(x_i)^T\phi(x_j) \tag{2.26}$$

and α is a parameter satisfying $e^T\alpha = v$, where $e = [1,\ldots,1]^T$ is the vector of all ones. Then the decision function is given as:

$$f^{m}_{\nu,\rho}(x)=\mathrm{sgn}\left(\sum_{i=1}^{m}\alpha_i K(x_i,x)-\rho\right) \tag{2.27}$$

2.2.7 Morphological Operations for Binary Images

Morphological image processing describes a set of operations based on the set theory to extract image components which are useful in the representation and description of region shape (Gonzales and Woods, 2008). Image objects are represented as sets in this terminology. The elements of these sets are the (x, y)

pixel coordinates in 2D integer space Z^2. In binary images, sets are assumed to be a group of connected white pixels. To define morphological operations, following concepts in the set theory are used.

An *empty set* is expressed with the symbol *Ø*. When an element a is the *member of* the set A, it is written as $a \in A$; otherwise, it is denoted as $a \notin A$. Sets are specified as the elements in brackets: $\{\cdot\}$. For instance, the expression $A = \{w \mid w = -d, \ for\ d \in B\}$ defines the *A* as the set of elements *w* such that *w* is the negated elements of the set *B*.

Set A is said to be a subset of set B ($A \subseteq B$) when every element of A is also a member of B. The union of these sets, denoted by A∪B, contains the set of all elements either in A, B or both. The intersection is defined as the elements both belong to these sets and expressed as A∩B.

The *complement* of a set *A* is defined as the elements which are not member of A and denoted by:

$$A^c = \{w \mid w \notin A\} \tag{2.28}$$

The *reflection* of a set *A* is expressed as:

$$\hat{A} = \{w \mid w = -a, \ for\ a \in A\} \tag{2.29}$$

Finally, the *translation* of a set *A* by a point $z = (z_1, z_2)$ is defined as:

$$(A)_z = \{c \mid c = a + z, \ for\ a \in A\} \tag{2.30}$$

Using these definitions two fundamental operations for binary morphological processing can be defined. The first one is called the *dilation* operation which is

used to expand the region defined by a set. The mathematical definition of the dilation operation is given as:

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \varnothing\} \tag{2.31}$$

where A and B are sets in Z^2. A corresponds to the image object to be expanded and B refers to the structuring element (SE). The operation reflects B about its origin and then translates it by z. The result is the all elements (pixels), z, so that the intersection of $\hat{B}$ and A have at least one element.

The second operation, *erosion*, has the reverse effect on the object and shrinks it with respect to shape of the given SE. The operation is defined as:

$$A \Theta B = \{z \mid (B)_z \subseteq A\} \tag{2.32}$$

Where A is the object, B is SE in Z^2. The result contains a set of elements z, such that translation of B by z is a subset of A.

2.3 Past studies

Studies in the literature are given in three sub-sections. First, past studies which make use of image segmentation are given. This is followed by the shadow based studies and the literature on resegmentation. Finally, a general review of the other studies consisting of building detection and urban land cover classification are presented.

2.3.1 Studies based on object-based methods

Image segmentation is the main step of the object-based classification methods. One of the earliest works utilizing image segmentation is the one proposed by Shackelford and Davis (2003). In this study, the authors presented an object-based land cover classification approach which uses the preliminary results of a pixel-based fuzzy classification on pan-sharpened multispectral Ikonos images. Image segmentation was used to refine the initial results of the pixel-based classification and improve it by adding a new class for the impervious surface other than road and building. The algorithm used for the segmentation was based on region merging which starts with single pixels as separate segments and merges them according to the increase in the heterogeneity. Lastly, a final classification was performed in which mean, variance, morphological shape information and the location of potential building segments with respect to shadow segments were used as features. To identify potential building locations, segments identified as Impervious Surface and also adjacent to nearby shadow segments were determined. The algorithm managed to classify buildings with 76% accuracy on a sample image of a dense urban area.

A subsequent study that uses clustering and edge detection for building extraction from panchromatic (PAN) Quickbird images was proposed by Wei et al. (2004). In the clustering step, an image segmentation based on histogram peak selection was performed to partition the image into a number of classes. Their method identified shadows as the lowest grey-valued class. Then, candidate building objects were extracted by using the shadow evidence determined from the shadow cast. Next, the edges of the candidate buildings were extracted with the Canny edge detector. Finally, Hough Transform was applied to the detected edges for polygon delineation of buildings. Comparing with the manually delineated results, the proposed method was found to be quite efficient, although its performance strongly relies on the image segmentation results performed at the clustering step.

Liu et al. (2005) developed a building extraction system for high resolution remotely sensed images based on multi-scale object-oriented classification followed by the probabilistic Hough transform. Their algorithm starts with generating object primitives with a multiresolution segmentation algorithm proposed by Baatz and Schäpe (2000) in which smaller objects are merged with respect to the spectral, tone, texture, shape and context information. These information were evaluated by a fuzzy rule decision classifier provided by the commercial software, eCognition 3.0, to extract building roofs. Finally, those roof shapes are reconfigured by applying first the probabilistic Hough transform to obtain the dominant lines and then rectilinear fitting according to these dominant lines.

An automated building extraction strategy from 1-meter resolution Ikonos imagery by using structural, contextual and spectral information was demonstrated by Jin and Davis (2005). The method is primarily based on constructing a differential morphological profile (DMP) by performing a series of morphological opening and closing operations on the input image to reveal the structural information. DMP was also used to obtain contextual information which is derived by shadow extraction. Shadows were used to hypothesize the position and size of the buildings. Those two types of information were combined with the spectral information in which bright building regions are detected to obtain the final extraction. The method was tested with the Ikonos satellite imagery of City of Columbia, Missouri and was able to identify buildings with 72% accuracy and 58% quality.

In a more recent study, Lefèvre et al. (2007) proposed an approach of three steps for building extraction from VHR remotely sensed images based on the morphological operations. First, PAN images were segmented by binarization with a histogram-based clustering method to obtain image objects. Then, an automatic morphological filtering was performed in order to eliminate the objects

with sizes lower than the desired minimum building size and to determine the structuring element size. In the last step, buildings were extracted by applying an adaptive Hit-or-Miss transform. According to the experiments made on a PAN Quickbird image from Strasbourg, buildings were detected with 63% accuracy. The method generally failed on the cases where building shapes were not rectangular and spectrally similar objects other than buildings existed.

A semi-automatic approach to extract buildings from Quickbird imagery in urban settlement areas was developed in a work proposed by Mayunga et al. (2007). In the first step, some image processing was performed to enhance edges by preserving the geometry of objects at the same time. Then, snake contours were initialized by the radial casting algorithm. The initialization was followed by the fine measurements of the building outlines in which the user marks the center of each building to automatically generate a snake contour. If the generated snake contour was accepted, a minimization procedure was applied to make the snake contours fit the building outlines. It is shown that the method was capable of extracting buildings with different shapes and orientation with reliable accuracy. In addition, the proposed approach was found to be 32% faster than manual extraction process. The study provided competitive results and was found superior to the general boosting method.

Another semi-automatic approach for building rooftop extraction from high resolution satellite imagery was presented by Liu et al. (2008). The method integrated a region-based approach with a feature based one to extract precise roof boundaries. In the region-based approach, a multi-scale object oriented image segmentation based on edge confidence and mean shift followed by a seeded region growth algorithm was implemented to discriminate simple rectilinear rooftops from the background. The feature-based approach made use of features such as corners, edges, lines and orthogonal corners to delineate the precise positions of roofs. In this step, pose clustering and model matching techniques were employed to locate buildings and find their correct shapes. The

results showed that integration of these two methods provided a successful extraction for the 75% of the rooftops of varying shapes.

Further, Sun et al. (2008) introduced an object-based boosting method for automatic building extraction from high resolution remote sensing images. The first step of the method was building a hierarchical network of objects created by the Pyramid-cut segmentation algorithm. Then, a feature vector was formed for each object by calculating four types of features; color, texture, shape and location. Next, those feature vectors were used in the Modest AdaBoost algorithm to train a strong classifier. Finally, buildings were extracted by labeling each object using three contextual models (i.e., topological, inherited and semantic model).

Karantzalos and Paragios (2009) introduced a novel recognition-driven variational framework to solve the problem of automatic building extraction from high resolution PAN satellite and aerial images. The method primarily integrates eight prior building templates into a level-set segmentation and introduces variational formulations to estimate the number and pose of buildings. The experimental results on various satellite and aerial images indicated that the algorithm was able to overcome misleading low level information derived from shadows and occlusions. Additionally, the results demonstrated that the algorithm has the potential of detecting any type of building when more sample templates are incorporated into the database.

Akcay and Aksoy (2010) proposed an algorithm which uses spectral, structural and contextual information to detect buildings with complex shapes in VHR images. Initially, a watershed segmentation was performed to obtain over-segmented image regions. Among all, the regions whose average brightness and average Normalized Difference Vegetation Index (NDVI) values are lower than specified thresholds were identified as shadow regions. Then, directional relationships with respect to these shadow regions along the sun azimuth angle

were used to detect candidate building regions. Since the buildings were not represented as single regions in the over-segmented image, the authors observed that regions forming a building were densely located whereas others separating buildings were far from their neighbors. For this reason, a graph was constructed where the nodes of the graph are determined with respect to the centroids of regions and edges are created between each neighbor. Edge weights were assigned as the spatial distance between neighboring regions. Finally, minimum spanning trees were constructed to cluster these candidate regions to identify buildings by using edge weights. The approach was evaluated on sample test images of Antalya, Turkey and provided promising results for buildings having different shapes and color. However, it failed on cases where shadows could not be detected and shadows of objects other than buildings existed.

Yang et al. (2010) introduced a technique based on texture enhancing of aerial images for building extraction. In the first step, edge detection, Principal Component Analysis (PCA) and texture filtering of second-order probability statistics were applied to obtain a gray image of contrast texture. The second order statistics were calculated by using a gray spatial dependency matrix. In the next step, a false color composite image was produced with the combination of contrast gray image, original blue and green bands. Authors implied that using this false color composite image instead of the original aerial image improved the accuracy of classification, since the former enhanced the edge features of the building and reduced spectral differences. In the last step, building extraction was completed by performing a fuzzy classification on image objects obtained after a multi-scale segmentation. The main drawback of the technique is explained as it needs constant experimentation to determine parameters and requires qualified users. Nevertheless, the proposed method managed to classify 92% of the buildings correctly.

In a different study, Tan et al. (2010) presented an object-oriented classification method to classify building roofs of pan-sharpened Ikonos multi-spectral images

with LIDAR data. First, image objects were obtained by a segmentation performed on pan-sharpened image which was initialized by a preliminary segmentation performed on LIDAR data. Then, those objects were classified in a rule-based manner by considering elevation, mean NDVI value, mean value in PAN band and the ratio of near-infrared (NIR) value to the sum of values in all multispectral bands. Lastly, the results were exported by polygon generalization and refined by geometrical regularization. The algorithm performed well in areas where building density is rather low and achieved 95% overall classification accuracy.

In a more recent study, Hong and Yang (2011) proposed a segmentation method based on multi-feature object-oriented Markov Random Fields Model to classify high resolution satellite images. In the first step, an object adjacency graph was constructed for the objects obtained after performing an over-segmentation using mean-shift segmentation. Then, spectral, texture and shape features were extracted for each object. Finally, a Markov Random Field (MRF) was defined on the graph where Expectation-Maximization algorithm was used to determine model parameters. The performance of the algorithm was evaluated with a sample GeoEye satellite image and achieved better results when compared with the generic MRF and the commercial software, eCognition.

Attarzadeh and Momeni (2012) presented an object based approach using stable and variable features to extract buildings in high resolution satellite images. First, image objects were created by using a bottom-up region merging technique known as multi-resolution segmentation algorithm. Then, stable features were derived from shape, strong edge and shadow adjacency characteristics of objects. Next, variable features were extracted from training objects using SEaTH (SEparability and THresholds) analysis tool. Finally, objects satisfying conditions derived from stable and variable features were identified as buildings. The performance of the algorithm was found to be relied on choose of parameters and threshold values in the segmentation step and extraction of stable features,

respectively. For the sample Quickbird image, over 80% of buildings could be extracted successfully. However, the algorithm failed especially on extracting buildings with light roofs.

An automatic building extraction approach was proposed by Aytekin et al. (2012) for urban areas. The approach initially aimed to extract man-made regions in high resolution satellite images. To that end, first, shadow and vegetated regions were extracted using chromaticity to intensity ratio in the YIQ color space and calculating NDVI. After masking out these regions, remaining parts of the image were segmented with the mean-shift algorithm. Then, thin and long segments were eliminated PCA and then, small segments were removed by morphological operations. The approach was tested on eight small sub sets of a Quickbird imagery of Ankara. The algorithm achieved varying detection rates, ranging from 64% to 91%, for these test sets. Best results were observed for the set where the complexity of the building shapes was moderate. The authors also revealed that existence of bare soil and other man-made structures were the main reasons of false detections.

In a study conducted by Dahiya et al. (2013), the building extraction methodology was initialized by a pixel-based classification. Then, image objects were extracted using a split and merge segmentation. The results were first refined by filtering and then converted into a vector image. Using the Imagine Objective tool of ERDAS 2011, building objects in the vector image were smoothened with a series of cleanup operators. The results of the method were compared with the manually delineated buildings for three sub images, where accuracies were computed as 85%, 73% and 70%. The authors criticized that their method tends to fail where other objects spectrally similar to buildings exist and shadows are needed to be removed for better performance.

Another object based methodology was developed by Benarchid and Raissouni (2013) in order to classify buildings in VHR images. First, image objects were

extracted using mean-shift segmentation. Then, each object was classified by Support Vector Machine (SVM) to identify buildings. The performance of the method was evaluated on a suburban are of 0.5 km^2 where buildings were detected with 81% accuracy.

Guo et al. (2013) presented a novel approach which integrates Geographical Information Systems (GIS) data for automatic extraction of buildings in high resolution satellite images. In the first step, input image was segmented using multi-resolution segmentation in the commercial software eCognition Developer 8.7. Then a graph was formed to represent the connectivity between adjacent segments. Next, a seed region was chosen using the position information derived from the GIS data and then its boundary was expanded by adding adjacent segments with respect to the constraints derived from the shape of GIS building. The experiments on Beijing dataset showed that the method was able to detect 89% of buildings with 79% quality. The main cause of the errors was found to be related with the segmentation step and GIS building map.

In a very recent study, Jabari and Zhang (2014) proposed a novel approach which utilizes Quickbird satellite image and the height information from LIDAR data together. First, a proper segmentation was performed using Hue-Saturation-Intensity (HSI) color transform. Assuming that buildings generally have rectangular shapes and constant colored roofs, a rule-based methodology was used to detect them by considering elevation, shape and color information of segments. All analyses were made by eCognition software on sample test image. The approach was managed to detect all buildings in the scene with a 93% border matching quality.

2.3.2 Resegmentation and shadow based studies

Resegmentation is a recently developed technique which was introduced by Korting et al. (2008) to deal with the over and under segmentation problems in generation of image objects. It can be defined as a set of procedures applied on an over-segmented image in order to generate more adequate objects to the application of interest.

Korting et al. (2008) proposed the resegmentation technique to detect rectangular regions in urban imagery. Their method, first classified image objects into one of the three classes of roof, street and tree using the Self Organizing Maps. Then, a weighted Region Adjacency Graph was constructed for the adjacent objects of same class. Then, image objects were classified. Finally, relevant nodes were cut or merged to obtain building regions by using minimal-cost paths and a rectangularity measure. The authors presented two experiments on sub-images of 300 x 250 and 256 x 256. The results were examined visually and showed the potential of the technique, although the algorithm was found to be computationally expensive.

In a subsequent study, Korting et al (2011) improved their technique by introducing an objective function which considers the rectangularity and area criterions. The authors described their method as more flexible, since objects having different shapes can also be detected by replacing the objective function with an appropriate one. The method was tested with four different portions of a Quickbird image. The root mean square error calculated for those four sub images varied from 0.446 to 1.357.

In a more recent study, Michel et al. (2012) introduced a methodology which applies the resegmentation process in a supervised way. In the first step, over-segments were created using mean-shift algorithm and an adjacency graph was formed using those segments. Then, edges of the same object were removed by

applying a SVM classification. Finally, neighboring segments whose edges were removed were merged. The method was tested on Quickbird and WorldView-2 imageries in which OpenStreetMap was used as a ground truth to evaluate the segmentation performance on buildings. The results were also evaluated by using Hoover scores (Hoover et al., 1996) and was found to reduce fragmentation.

In a master degree thesis, Yuksel (2012) proposed an automated building detection technique by using shadow and its direction. Using satellite image meta-data, illumination direction and detected shadow regions were used to detect potential building landscapes. Landscapes were then bipartioned using Grabcut in order to obtain building and background pixels. The main drawback of the technique was that it is solely based on a single evidence (i.e., shadow). Nevertheless, it was proven that shadow is an invariant for buildings and promising results could still be obtained by using only a single invariant.

In a recent study conducted by Ok (2013), another automated building detection approach was introduced using shadow information. First, a vegetation mask was built by histogram thresholding on the generated NDVI image. Shadows were detected by calculating a ratio image in which saturation and intensity components of HSI color space were used. The HSI image was form as a false color composite image (NIR, R, G). Then, a two-step post processing was applied on detected shadows in order to improve their boundaries and estimate height differences compared with the terrain height. Next, an approach based on a two-level graph theory was implemented. In the first level, building regions were detected separately for each shadow region by an iterative binary graph cut partitioning. In the second level, a multi-level graph optimization was performed to obtain final building patches. According to the experiments performed on sixteen VHR GeoEye-1 images, the proposed method was found to be robust and reliable.

2.3.3 Other studies

Most of the early studies used aerial images for building detection. One of them was a study proposed by Huertas and Nevatia (1988) which didn't only detect buildings but also constructed 3D shape descriptions of them. Their method used linear features extracted from a single image and shadow as evidence to detect 2D rooftops which are assumed to be rectilinear in shape (i.e., 'L', 'T' and 'E' shapes). Shadows with associated walls were also considered as evidence while generating 3D structures. Their study showed that generic shape models can be used to extract buildings.

Zhang (1999) proposed a method which uses texture filtering to improve the classification performance of the multispectral classification which is done by the unsupervised ISODATA clustering method. For texture analysis, grey value co-occurrence matrix was used in which the grey value relationships in the neighborhood of a pixel is described. Unlike normal texture analysis, these relationships were defined in co-occurrence matrix space by applying a transformation from the original image space. The results showed that the proposed technique improved the classification accuracy much better than the normal texture analysis.

In a different study, Perkins et al. (2000) developed a hybrid learning system called GENIE (GENetic Image Exploitation) which uses Genetic Programming to find a set of image processing operations that can produce proper feature planes. Those feature planes are then used to produce the final output by using a conventional supervised classifier. Later, Perkins et al. (2005) improved their system and developed a new one called GENIE Pro which is capable of integrating spectral and spatial information such as texture, local morphology and large scale shape information.

Further, Chen and Blong (2002) proposed a method by performing a pixel-based classification followed by a region and edge analysis to refine the results of the classification. At the first stage, a supervised artificial neural network classifier was performed using RGB bands and texture images (i.e., mean and contrast) to identify the building roofs. The resultant image was thresholded to obtain a binary image containing buildings in black and background in white. Then, morphological operations were applied to this binary image to extract preliminary roof boundaries. Finally, these boundaries were refined by applying a Canny edge detector to each band. The method was tested with a AUSIMAGE^{TM} image having 0.2 m spatial resolution. The study showed some potential for the use of VHR images in risk assessment of natural hazards.

Lee et al. (2003) presented a building extraction approach which first classified multispectral Ikonos images to obtain approximate location and shape for candidate buildings. Results of this classification were then used on the corresponding PAN image for a finer extraction. Next, image binarization was performed by ISODATA segmentation. Finally, boundaries of buildings were squared by an approach based on Hough transform which were assumed to be rectilinear. Experimental results showed that 64% of the buildings in the scene could be successfully detected and extracted.

In a more recent study, Kim et al. (2006) proposed a semi-automatic algorithm for extracting building lines from high-resolution satellite images. The approach utilized line matching and voting to extract lines of relatively large rectangular-shaped building roofs. The procedure was initiated with a manually selected point on a roof. Then, a region of interest was defined around this point so that the lines bounded by the region could be extracted. Next, initial orientation and position of a building line was estimated by a line voting process. Finally, a least squares matching process was applied to refine the orientation and position of each extracted line. The method failed on cases where buildings were relatively high, strong lines from building facades exist and buildings were densely located.

Despite these failures, 83% of the building lines in two Ikonos images were extracted accurately.

San and Turker (2007) developed an approach for building extraction from high resolution pan-sharpened multi-spectral satellite images and the Normalized Digital Surface Model (nDSM). The method first detected the candidate building regions from multi-spectral bands by Maximum Likelihood Classification. nDSM was calculated next by subtracting DTM from DSM and regions above some user threshold were identified as 3D objects. Among those objects, trees were eliminated which were above an empirically determined NDVI threshold value of 0.137. The remaining objects were considered as buildings. Finally, the boundaries of detected buildings were delineated by a series of building extraction techniques such as edge detection, vectorization, line simplification and vector filters. The method achieved a 80% building detection rate with 70% quality on several test samples of Ikonos imagery of Batıkent, Ankara.

Inglada (2007) presented an image processing system for the recognition of man-made structures from SPOT 5 THR images with 2.5 m spatial resolution. An SVM based supervised learning approach was used in order to detect eleven classes of objects. For training, examples were selected as 100 x 100 pixel sub-images with the object centered for each class. The objects were characterized by using a high number of geometric features defined by geometric moments, Fourier–Mellin descriptors and some high level geometry features such as perimeter, surface, compacity and barycenter. The method was able to classify isolated buildings with accuracies higher than 80% on the sample SPOT 5 images.

Further, Song and Shan (2008) proposed a framework for building extraction from high resolution color imagery based on the identification of building boundary and roofs. Initially, a color space conversion was performed from RGB to CIE L* a*b* and then, anisotropic diffusion was applied to enhance the color

information. Next, the J-image was generated from the bands of a* and b* using k-means clustering. Using J-image, active contours with the edge flows were computed to extract the building boundaries. At the last stage, JSEG algorithm was applied for delineation of the building polygons. The framework's performance was examined on small portions of an aerial imagery with 15 cm spatial resolution. Although the results were mostly satisfactory and reliable, there existed some inconsistent results when adjacent buildings have similar reflectance.

In a study proposed by Sirmacek and Unsalan (2009), an approach based on scale invariant feature transform (SIFT) keypoints, multiple subgraph matching, and graph-cut methods was applied to detect urban areas and buildings from panchromatic VHR Ikonos images. SIFT keypoints are extracted from the two building templates (one for dark and the other for bright buildings) and the test image after applying a nonlinear bilateral filtering to smooth out unwanted noises. Next, multiple subgraph matching between SIFT keypoints of template and the test image was applied to detect the urban area. SIFT keypoints were represented as vertices in the subgraph. Then, a novel graph-cut method was used to detect the buildings from the urban area. The algorithm was applied on 28 small test images and building centers were identified with a detection rate of 88% on average.

Turker and San (2010) presented an approach which uses a binary SVM to detect buildings from a pan-sharpened Ikonos imagery. The features used in classification were four spectral bands (B, G, R and NIR), nDSM, NDVI and four principal components (PCs). The study showed that including additional information apart from the original four spectral bands increased the classification accuracy of SVM. Authors also justified the reason of using PCs by revealing the fact that white or blue concrete roofs on PC2 and tile-roofs on PC3 were more clearly differentiated. The approach was tested with three sample

images obtained from Ikonos imagery of Batıkent, Ankara. Overall classification accuracies were computed over 90% for all test samples.

In a study conducted by Cetin et al. (2010), an approach using textural features and Adaboost was introduced to detect buildings in satellite images. A total of 137 features were calculated in a moving 7x7 window. Only 12 of these features were selected by Adaboost which were mean of Hue-Saturation-Value (HSV) components, mean and variances of the gradient magnitudes for each HSV component, Haralick correlation at 90 degree, energy of wavelet output, average magnitude of Fourier Power Spectrum and mean of Gabor filter output. After performing a pixel based classification using these features, some post-processing steps were applied on output in order to refine the results. The performance of the method was evaluated both by pixel and object based metrics on a data set containing 5 subset of satellite images. The worst detection rate was observed as 31% on the data where the complexity and the irregularity of the buildings were highest. On the contrary, the best detection rate was observed as 75% on the sample image where complexity and texture effects were lowest.

Sirmacek and Unsalan (2011) proposed a probabilistic framework based on local features to detected building centers on satellite and aerial images. The methods used for extracting local features were Harris corner detector, Gradient Magnitude based Support Regions (GMSR), Gabor filtering and Features from Accelerated Segment Test (FAST). In order to detect building centers, a probability density function (pdf) of these features were estimated so that the modes of the estimated pdf and their probabilities revealed center locations. Experiments were conducted both on aerial and Ikonos sample images. The proposed approach was found to be more successful for satellite images where 93% of the buildings were correctly detected with a 18% false alarm (FA) rate. For aerial images, however, detection rate decreased to 82% and FA rate increased to 39%.

Cretu and Payeur (2012) presented a building detection approach based on visual attention models. The model uses the features based on color, texture and orientation in order to highlight the regions of interest. Authors introduced a set of weighting schemes based on the contributions of these features to detect buildings. Next, the best weighting scheme was selected by an original adaptive algorithm to highlight the desired regions. In the final step, weights of the selected scheme were used to train a neural network in order to learn the association between the weights and the desired output. The approach was tested with 50 satellite images of 256 x 256 pixels. Although, the overall detection rate was 55%, the approach was found superior to the other salient detectors in the visual attention literature.

In a different study, Mirhassani et al. (2012) introduced an approach which utilizes Unsharp Masking (USM) and Motion based Unsharp Masking (MUSM) for the extraction of buildings in urban areas. The idea was that increasing the local contrast of the image with USM and MUSM improves the performance of Bayesian classifier. The algorithm provided promising results on Google Earth images used for testing. However, the algorithm lacked parameter adjustment for different scenes and adaptivity for the synthetic problems occurred due to the low and high pass filter behaviors.

In a more recent study, an algorithm utilizing DSM and GIS data together with high resolution satellite images was developed by Dini et al. (2013) for the delineation of building patches. The method aimed to extract newly constructed buildings based on an older GIS database by identifying blobs derived from DSM which do not exist in the GIS data. For this purpose, a 3D line matching procedure was applied on these blobs and the lines extracted from the stereo satellite images. Finally, building footprints were determined by using the actual blob information and derived 3D edges using a box-fitting approach. Experiments on GeoEye-1 stereo images demonstrated that the approach succeeded for the

buildings sufficiently large, have simple rectangular shapes and high contrast compared to the surrounding regions.

Sumer and Turker (2013) combined genetic programming concepts with generic image processing operators for building detection in high resolution satellite images. The algorithm, first constructed a candidate solution by applying a Fisher's linear discriminant analysis on the spectral and textural features derived from the image bands. Then, genetic algorithm (GA) operations were applied after computing the fitness value of this solution by comparing it with the reference data. Next, a new solution was generated by updating the parameters of GA by an adaptive fuzzy logic controller and repeating this update for a pre-defined number of iterations. Finally, morphological opening/closing, artifact removal, and hole-filling were performed in order to refine the results. For ten different test scenes obtained from Ikonos imagery of Batıkent, Ankara; kappa statistics were computed between 0.55 and 0.88. According to these experiments, the algorithm had trouble for the detection of buildings in rural test scenes and dense buildings in urban areas.

In a very recent study, Youssef et al. (2014) combined per-pixel supervised classification with Marked Point Processes (MPP) for building detection. Low level primitives derived after pixel based classification were used on the second step to support the generation of the probability map of buildings. To generate such a map, several hypotheses were formulated based on the size and rectangularity information. Finally, the results of the pixel based classification and the probability map were fused together using a MRF formulation to form the final output. The method was evaluated on three test images covering almost 1.5 km^2 in total. Although the detection rates varied from 67% to 81%, the completeness of the detected buildings were generally high ranging from 89% to 94%.

CHAPTER 3

PROPOSED APPROACH

3.1 Overview of the Proposed Solution

A general view of the proposed solution is given in Figure 3.1. The only input of the system is an RGB image. The algorithm starts with performing an over-segmentation on the input image in order to obtain image objects. It should be noted that these image objects do not correspond to real word objects in the image due to the over-segmentation. They will be merged by resegmentation however, which is the main objective of the algorithm.

After generating image objects, a shadow detection procedure is performed. The adjacent objects identified as shadow are merged afterwards. This is followed by the feature extraction step in which, extracted shadows are used to identify meaningful features in order to determine which segments should be merged to form buildings. The following step is called the resegmentation step where all image objects expect the shadows are classified either as building or not, with respect to the features calculated previously. Building patches are obtained by merging neighboring objects which are all classified as building.

The results of the classification are refined in the following post-processing step. In addition to applying some heuristics, building patches which are not adjacent to any shadow are removed from the results. A further refinement is also

performed in the relaxation step where one-class modeling (OCM) is used for each of the extracted building patches in order to obtain the final output.

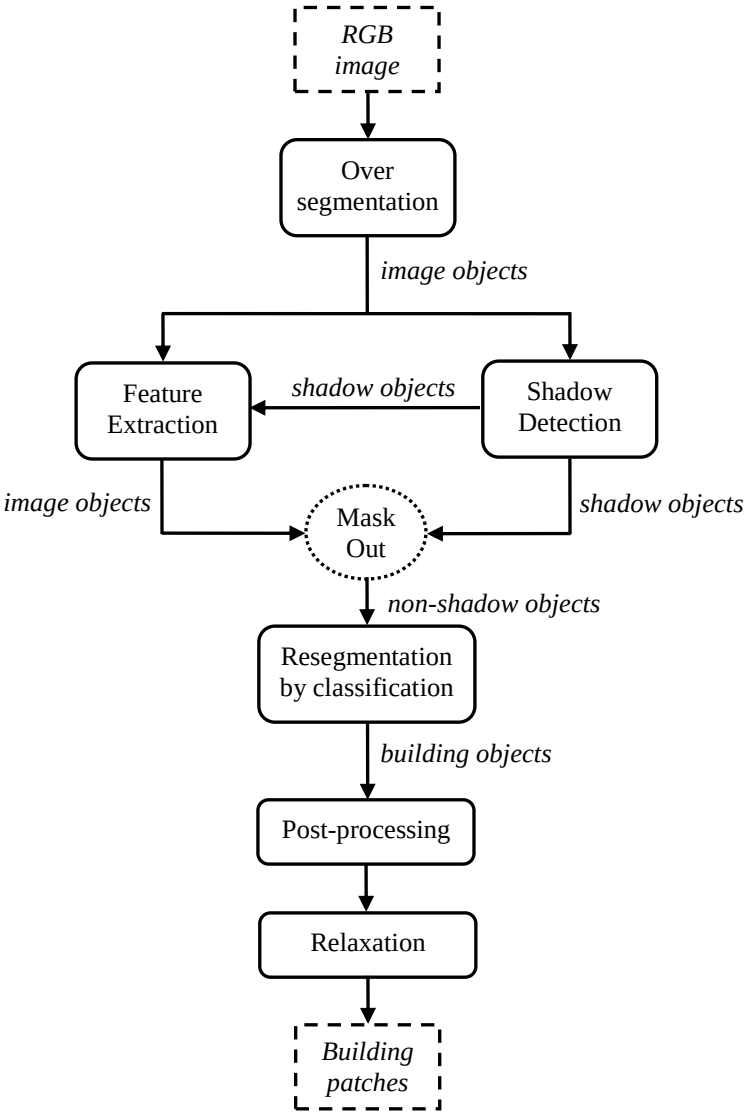


Figure 3.1: Overview of the system

3.2 Over-segmentation with Mean Shift

To perform an over-segmentation using Mean Shift, values of the spatial and color bandwidth should be chosen as small as possible. The smallest possible value for these bandwidths is 1. However, for spatial bandwidth, the value 2 is chosen in order to avoid segments having only a single pixel. Figure 3.2 shows an illustration of segmentation performed on the sample image in Figure 3.2 (a). Over-segmentation (Figure 3.2 (b)) was obtained using spatial and color bandwidth values of 2 and 1, respectively. In Figure 3.2 (b), it can be observed that the building with red roof was partitioned into several fragments after the process. On the other hand, under-segmentation occurs for higher values of these bandwidths. For instance, choosing spatial bandwidth value of 8 and color bandwidth value of 4 resulted an under-segmented image as shown in Figure 3.2 (c). A large segment in the center of the image is clearly visible which combines the building with some of its surroundings.

(a) (b) (c)

Figure 3.2: Segmentation example. (a) Original image. (b) Over-segmentation. (c) Under-segmentation.

For mean-shift algorithm, the implementation of Edge Detection and Image SegmentatiON (EDISON, 2010) system developed by the Robust Image Understanding Laboratory in Rutgers University was used. The code used in the

study is the MEX implementation for MATLAB (Shai Bagon's Matlab Code, 2010).

3.3 Shadow Detection

In general, shadows are known to be dark regions where the source of light is partially or completely blocked. Additionally, they have increased hue property in remotely sensed images (Tsai, 2006). Combining these two properties together, hue/intensity ratio was used to identify shadow regions, since this ratio tends to have higher values in shadow regions than other regions. To compute such a ratio, first, a color space conversion is needed. In a study presented by Tsai (2006), several invariant color spaces were compared with respect to their shadow detection performances using this ratio and YIQ model was found to be the best one in this respect. Therefore, in this study, a color space conversion from RGB to YIQ model was calculated using:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.1}$$

This conversion was applied on the over-segmented image where each segment is replaced with their mean RGB values of its pixels. To form the ratio image, the components of I and Y were used which correspond to hue and luminance (i.e., intensity), respectively. First, these components were normalized between the values of 0 and 1, and then the ratio image was calculated by:

$$r_{shadow} = \frac{I - Y}{I + Y} \tag{3.2}$$

Equation 3.2 calculates the ratio r_{shadow} in a normalized way so that the values are bounded between -1 and 1. Next, a threshold must be determined to discriminate

shadow from the background. It is evident that the threshold changes from scene to scene, since different images may have different contrasts and illumination. To determine a threshold automatically, Otsu's method was applied to the histogram of the ratio image which maximizes between-class variation while minimizing within-class variation (Otsu, 1979). The regions having higher values than this threshold were identified as shadows. Figure 3.3 shows some examples of the detected shadows for the given sample images.

(a) (b)

Figure 3.3: Shadow detection example. (a) Over-segmented images. (b) Detected shadows.

3.4 Feature Extraction

Up to now, the input image is over-segmented and among these segments, the ones belonging to shadows are identified. In this section, features of the image objects which are used in classification to obtain a better segmentation are explained in detail. There are many available object based features, such as shape, texture, area, etc. However, due to the over-segmentation, very few of them are usable for detecting building segments since all image objects are small and have a high degree of homogeneity. More explicitly, for small and homogenous segments, intuitively, there exists almost no texture. The same reason is also valid for the shape related features, since all objects are small which does not lead enough variety for shape. Therefore, most of these features cannot discriminate segments inside buildings from other objects.

Due to the reasons explained above, more sophisticated features are needed. With no other information available, such as height and infra-red, shadow becomes a more important cue in this case. For this purpose, two new features based on shadow information are introduced. In the following sub-sections, a ratio feature derived from color information and these two shadow based features are given.

3.4.1 Red-Green Ratio

This feature was introduced to include color information into the analysis. In this study, instead of using each spectral band (i.e., R, G, and B) separately, a ratio image of red and green bands was preferred. The ratio was calculated in a normalized way using the following equation:

$$r_{color} = \frac{R - G}{R + G} \tag{3.3}$$

Figure 3.4 shows an example for ratio image generation. The Equation 3.3 is applied on the segmented image in Figure 3.4 (a). Figure 3.4 (b) shows the result where the discrimination of red roofs and green vegetation is clearly noticeable. Reddish segments are brighter since the ratio generates higher values and greenish segments are darker due to lower values produced from the formulation.

(a) (b)

Figure 3.4: Color feature example. (a) Over-segmented image. (b) Ratio image.

Using a ratio instead of using R, G and B values together is more effective due to the two main reasons. First, most of the spectral information is represented using just a single value, although blue band is not included in the formulation. The logic behind this is that blue color does not exist in nature as much as green and red. On the other hand, red is a favorite color for roof tiles and green is the main color of tree, shrub and grass.

The second reason is that the value obtained using a ratio makes it independent from image contrast and illumination changes (Riaño 2003, Liu 2009). This property is essential, especially for the cases where one half of a rooftop is shadowed or partly illuminated while the other half gets direct sunlight. In this

situation, RGB bands have lower values in the shadowed half than the illuminated half. However, the ratio does not change unlike RGB. These two halves get similar values if they are made of same material and look same in the absence of shadow.

3.4.2 Closeness to Shadow

Intuitively, the closer the segments to the shadows, the more likely they will belong to a building. Based on this assumption, a feature which measures the degree of closeness of image segments in order to favor closer ones was introduced. However, it is clear that not all segments around a shadow are necessarily building candidates for the cause of the shadow cast. In fact, a building is located in the opposite direction of illumination with respect to its shadow, thus, only segments in this direction should be considered. Therefore, illumination direction should be determined first. For satellite images, it can be derived from the metadata. However, in Google Earth images, there is no such information available. Hence, in this study, the direction is specified manually by the user as one of the eight main directions; north (N), northeast (NE), east(E), southeast (SE), south (S), southwest (SW), west (W), and northwest (NW).

Given the direction of illumination, the closeness degree of each segment to the nearest shadow object is determined in three steps. First, pixels in the opposite direction are identified. For this purpose, shadows are expanded towards to the this direction using binary morphological dilation. The structuring element (SE) used in dilation can be in form of any shape; however, square is used in the process for simplicity. The size of the SE should also be large enough to contain the building entirely. Therefore, it was chosen as the image width to guarantee this requirement. The elements in the SE were chosen in such a way that 1's are placed in the desired direction and all other places are left as 0. Figure 3.5 shows

example 3 x 3 SEs for all eight possible illumination directions. The algorithm picks the proper one regarding to the user's direction choice.

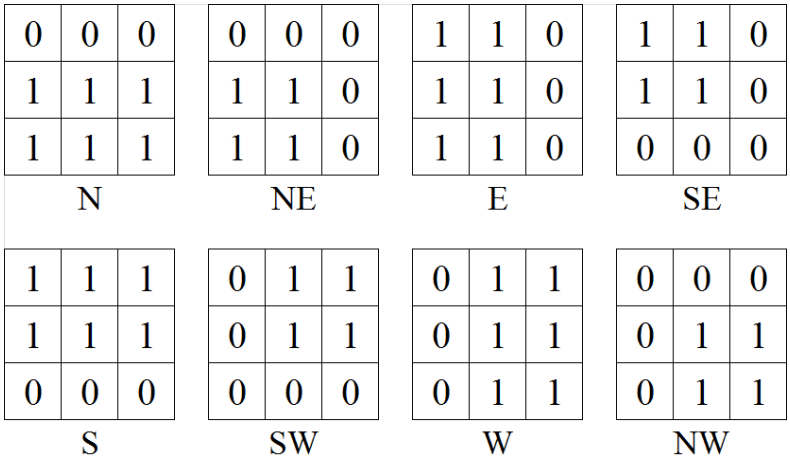
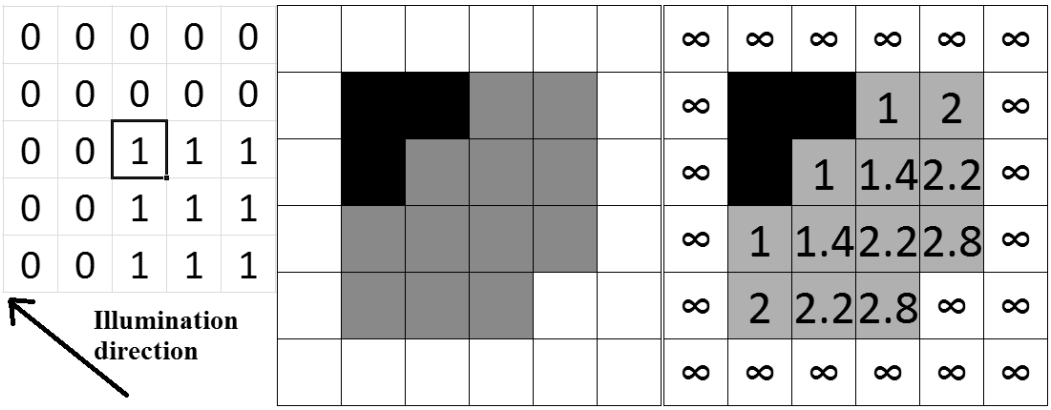


Figure 3.5: SEs for each illumination direction

Another illustration is presented in Figure 3.6 that shows how a 5 x 5 SE is used to delineate the expanded region for the given illumination direction (Figure 3.6 (a)). The center of the SE is shown as the value 1 with a thick black square. The result of the dilation is shown in Figure 3.6 (b), where the black pixels represent the shadow and grey pixels denote the expanded region.

In the second step, for each shadow segment, Euclidean distances of the pixels in its expanded region are calculated. The distances are computed in terms of pixels as illustrated in Figure 3.6 (c). For the pixels outside of this region, values are set to infinity. The result of this process may produce multiple values for pixels, thus, the minimum value is chosen for each pixel to ensure that every pixel gets a single value representing the distance to the closest shadow. Finally, the distance value of each segment is determined by averaging its pixels.



(a) (b) (c)

Figure 3.6: An illustration for distance calculation. (a) SE. (b) Expanded region. (c) Calculated distances.

Figure 3.7 shows an example by presenting the result of this process. The procedure is applied on the over-segmented image in Figure 3.7 (a). As can be seen in the distance image (Figure 3.7 (b)), segments closer to the shadows (red) are brighter than the ones which are far away.

(a) (b)

Figure 3.7: Result of the closeness to shadow feature. (a) Over-segmented image. (b) Distance image.

3.4.3 Presence in Region of Interest

The idea behind this feature is that a region of interest (RoI), where a building is most likely be, can be roughly determined using the shadow cue. To that end, a procedure was implemented in two steps to rotate each shadow object 180^0 to form a closed shape. First, rotated shadow object which is symmetric with respect to the image origin (i.e., upper left corner) is obtained. In other words, the coordinate of each pixel (x, y) in the shadow object is reflected (rotated) in this operation (i.e., (-x, -y)).

The result of the first step gives an 180^0 rotated shadow but with negative pixel coordinates which is meaningless and cannot be showed in an image. Therefore, in the second step, rotated shadow should be moved or translated to its original position. In addition, their end points should touch to form a closed shape so that the RoI can be determined as the region inside this shape. The end points are selected from the extrema points in the opposite illumination direction. Extrema points are the eight main end points as illustrated in Figure 3.8.

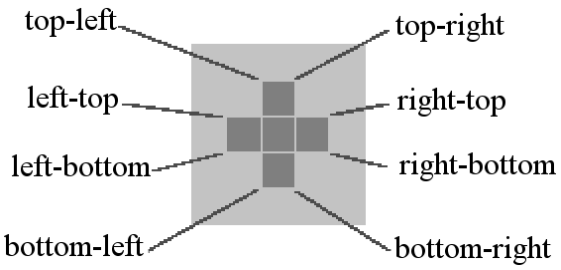


Figure 3.8: Extrema points.

Given the illumination direction, the far two extrema points in the opposite direction are chosen as the end points of the shadow object. For instance, suppose that the illumination direction is towards N which means *top* position in image plane. Then, the points in bottom-left and bottom-right are selected (Figure 3.8).

As in the *closeness to shadow* feature, the algorithm picks proper end points regarding to the user's choice of illumination direction.

Once the end points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ are determined, the rotated shadow object is translated so that the rotated end points $P_1'(-x_1,-y_1)$ and $P_2'(-x_2,-y_2)$ touch one another (i.e., P_1' touches P_2 and P_2' touches P_1) using:

$$\begin{bmatrix} x_r \\ y_r \end{bmatrix} = \begin{bmatrix} x_r \\ y_r \end{bmatrix} + \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 + 1 \end{bmatrix} \tag{3.4}$$

where (x_r, y_r) are the pixel coordinates of rotated shadow. Figure 3.9 shows an illustration for the determination of the RoI. The original shadow is given in Figure 3.9 (a). Figure 3.9 (b) shows how the rotated shadow is aligned with the original one after the process. The resultant RoI is shown as grey in Figure 3.9 (c). The result of the procedure described in this illustration is shown in Figure 3.10. The sample image is given in Figure 3.10 (a). The RoIs identified after the process are shown as grey in Figure 3.10 (b) where the black regions in the image represent the corresponding shadows. It is observed that the procedure works on non L-shaped shadows, too. However, for thin and I-shaped shadows, the area of the detected RoI becomes smaller than the desired.

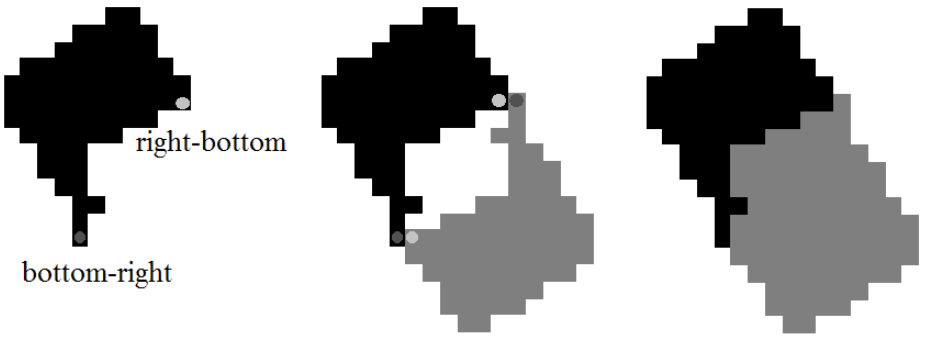


Figure 3.9: Determination of the RoI. (a) Shadow. (b) Alignment after rotation. (c) The RoI in grey.

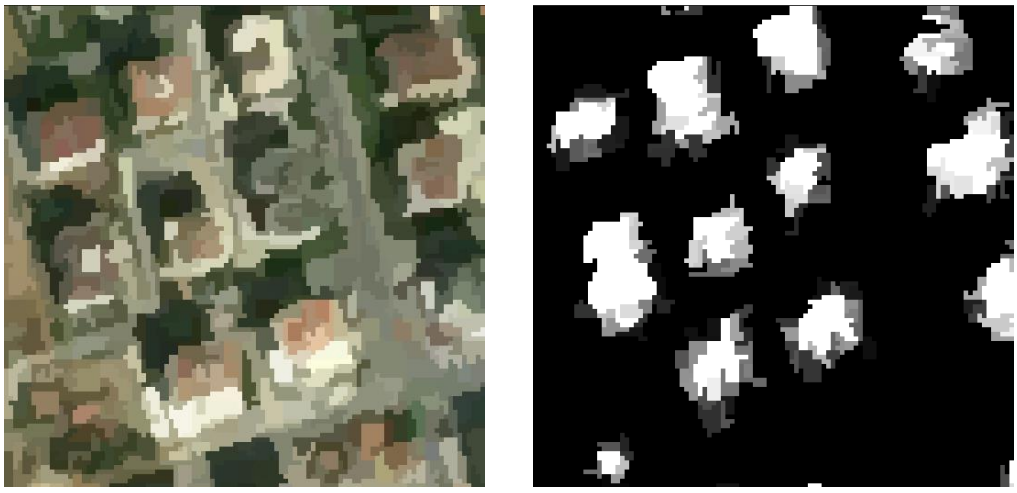
(a) (b)

Figure 3.10: RoI generation example. (a) Over-segmented image. (b) Identified RoIs.

Finally, to determine the presence of a segment inside the RoI, percentage of its area inside RoI is calculated. Area computation is performed as the number of pixels. Let S be the image segment and R be the RoI, the percentage value p_i for each segment i is calculated by:

$$p_i = \frac{Area(S \cap R)}{Area(S)} \tag{3.5}$$

Figure 3.11 shows an example for percentage (p_i from Equation 3.5) calculation. In the result image given in Figure 3.11 (b), brighter segments have higher percentage values when compared to the darker ones.



(a) (b)

Figure 3.11: Result of the presence in RoI feature generation. (a) Over-segmented image. (b) Percentage image.

3.5 Resegmentation

The resegmentation procedure intends to merge over-segmented image objects by classifying the features calculated previously. The classification is binary, since the process aims to determine whether a particular segment belongs to a building or not. Any binary classifier can be used in this step. To choose a proper one, the performances of some popular classifiers such that Logistic Regression, Discriminant Analysis, Support Vector Machine and AdaBoost which are explained in sections 2.2.2, 2.2.3, 2.2.4 and 2.2.5 were examined. The results are given in section 4.3. All these classifiers were implemented using built-in functions of MATLAB except AdaBoost, for which GML AdaBoost Matlab Toolbox v.0.3 (2011) is used. This toolbox is developed by Graphics and Media Lab in Department of Computational Mathematics and Cybernetics of Lomonosov Moscow State University that uses Classification and Regression Trees (CART) as weak learners. CART is a decision tree analysis introduced by Breiman et al. (1984) that utilizes a binary tree graph in which the leaves

represent the classification result and nodes represent some predicate. Each branch of the tree is marked as either true or false. The classification is performed by traversing the tree to reach leaves.

3.6 Post-Processing

This step refines the resegmentation results in three steps. First, holes inside the extracted building patches are filled. This step is performed by a set of morphological dilations, complementation and intersections (Gonzales and Woods, 2008). The procedure begins with a point *p* in the hole of the object A and fills the region inside by:

$$X_k = (X_{k-1} \oplus B) \cap A^c \tag{3.6}$$

where X_k is the filled region in k^{th} iteration which is initiated as $X_0=p$ and B is the 3 x 3 symmetric structuring element (SE) shown in Fig 3.12.

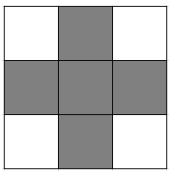


Figure 3.12: Structuring element.

The algorithm stops at step k if $X_k=X_{k-1}$. The result of the filling process is returned as the set union of X_k and A. An example is shown in Figure 3.13 where the hole in the extracted building patch is filled successfully.

(a) (b) (c)

Figure 3.13: Hole filling example. (a) Sample image. (b) Extracted building patch. (c) Result after filling.

In the second step, building patches which are not adjacent any shadow are eliminated since the method is shadow based and could only find buildings near shadows. To perform such an operation, each building patch is dilated with a 3 x 3 square SE and removed if its intersection with the shadow image is empty. Let R be the removed patches, then the operation is defined by:

$$R = \{A \mid (A \oplus B) \cap S = \varnothing\} \tag{3.7}$$

where A is the building patch extracted by the resegmentation (classification) step, B is the SE and S is the binary shadow image.

Figure 3.14 (b) shows the result of this operation performed on the sample image given in Figure 3.14 (a). Shadows are indicated as red and the boundaries of the extracted building patches are shown as green. It can be observed that the operation successfully removed falsely extracted patches in the center and lower-left part of the image in Figure 3.14 (a).

(a) (b)

Figure 3.14: Patch elimination example. (a) Result after resegmentation. (b) Result after elimination.

Final step is the area thresholding. Assuming that the area of a building cannot be smaller than a certain value, such patches are removed from the results. The area for the threshold is determined automatically when the spatial resolution of the image is given. In all experiments, the area of $100m^2$ is chosen as the minimum detectable building, thus, the threshold is determined as the number of pixels corresponding to this criterion given the spatial resolution.

3.7 Relaxation

This section describes an automated approach introduced to refine the building extraction results. The method is based on constructing separate models for the regions precisely classified for each extracted building patch. To construct such a model, One-Class SVM technique described in section 2.2.6 is utilized. To integrate necessary functions of this technique into the implementation, LIBSVM library and its MEX implementation for MATLAB which is developed by Chang and Lin (2011) is used. The kernel used in One-Class SVM is the Radial Basis Function given in Equation 2.20 which can handle solving non-linear problems.

To construct a separate model for each extracted building patch, a sub-region inside this patch is selected to be used as the training set. Then, nearby pixels are classified as whether they are similar to the training set or different as illustrated in Figure 3.15. The main assumption is that the extracted patch is under-segmented. In other words, in addition to buildings, it contains some surrounding regions as well. Therefore, the process will eliminate some of these regions if they are not found to be similar. For the extracted patches where this assumption fails (i.e., when a building could not be detected as a whole), it is better to skip the process for that patch since the method may shrink it further and fails to improve the result. The decision for skipping can be made after analyzing the previous classification. For the experiments done in this study, some recommendations are given with respect to the results in section 4.2.

Figure 3.15: An illustration of applying One-class SVM on the extracted patch.

To determine the regions for training, the classification results of the resegmentation step is used. The classification procedure assigns each segment a probability value between 0 and 1. Values closer to 1 indicate that those segments are more likely to belong a building whereas the segments with values closer to 0 are less likely. Therefore, for each extracted building patch, regions with the highest values can be used to identify the regions for training. However, this region may be too small; thus, it may contain very few segments so that their number may not be enough for training. For this reason, the classification in

relaxation step is performed as pixel based. In addition, pixels inside each segment are assigned with the same value produced by resegmentation.

The algorithm given below is composed of four main steps:

Inputs:

I: RGB satellite image

B: set of building patches

P: probabilities regarding previous classification result

t_{tr}: threshold for patch area percentage

v: parameter of One-class SVM, (0,1]

Output:

R: output binary image initially assigned to all zeros

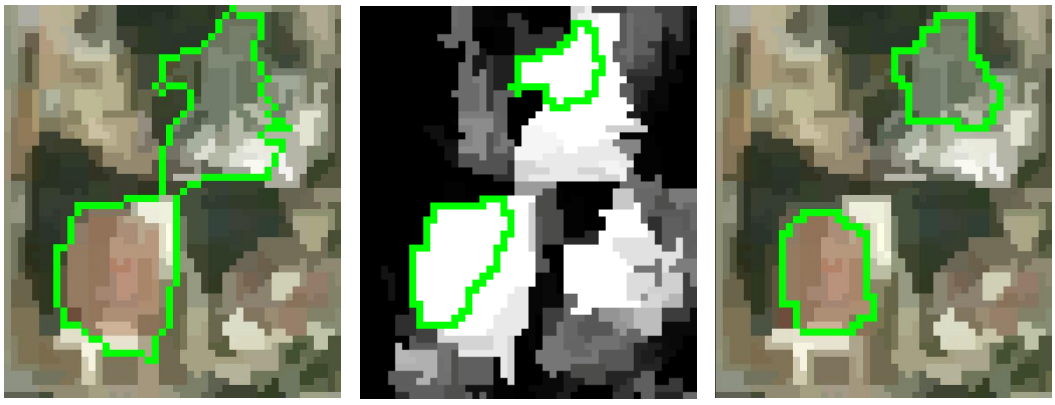
Begin:

For each b_i in B (i=1,2,…,n)

1. Sort pixel probabilities of b_i in descending order:
   p_i = sort(P(b_i), "descend").
2. Select top t_{tr} pixels inside b_i:
   $x_i = p_i(1,2,\ldots,t_{tr}*\text{Area}(b_i))$
3. Find connected components formed by selected pixels:
   CC = cc_analysis(x_i)
4. For each connected component cc_j
   - 4.a. Train One-class SVM model using RGB values of pixels inside:
     M = ocsvm_train(I(cc_j), v)
   - 4.b. Dilate b_i to obtain the region to be used for testing the model:
     x_i = dilate(b_i, squareSE(3,3))
   - 4.c. Classify pixels in the region using the model M:
     R(x_i) = ovsvm_classify(I(x_i), M)

End

Given the probability values obtained after resegmentation, the pixels inside the building patch are sorted in descending order with respect to these probabilities in the first step. Second, some percentage (t_{tr}) of these pixels is selected for training. The value of t_{tr} is selected as 0.5 so that maximum number of pixels inside ground truth objects (buildings) is selected. For higher values of t_{tr}, it is observed that some pixels around buildings become selected which is not preferable since the training set should contain pixels belonging to buildings. If selected pixels do not form a single object, there may be more than one building in the patch. When buildings are densely distributed in the scene, such a case is more likely to happen. As a solution, all connected components are identified and remaining steps are executed for each component in the fourth step. Figure 3.16 illustrates such a situation where the extracted patch contains two buildings (Figure 3.16 (a)). The probability values are shown in Figure 3.16 (b) where brighter pixels indicate higher probabilities. In this example, selecting t_{tr} percentage of pixels inside the extracted patch in Figure 3.16 (a) yielded two connected components (with green boundaries) as shown in Figure 3.16 (b). The algorithm is then applied on each of them separately to obtain two results (Figure 3.16 (c)).

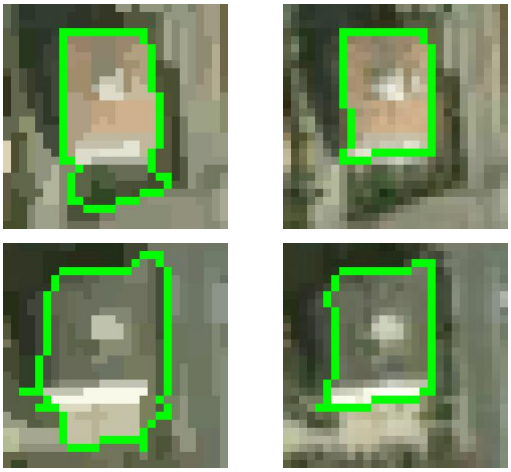


(a) (b) (c)

Figure 3.16: Example case when two buildings are extracted as a whole. (a) Extracted patch after resegmentation. (b) Probability image. (c) Final result.

In any case, RGB values of the selected pixels are used for training the One-class SVM model using parameter *v*. To identify the region to be classified, the building patch is dilated with a square structuring element (SE) of size 3 x 3 to acquire a little larger but smoother region. The reason is to fill small indentations in the boundary and to compensate small errors that may arise when the patch does not fit the building boundary well. Finally, pixels in this region are classified using the model acquired in the training step. The procedure continues with the next building patch and iterates these steps until all patches are used. The result is a binary image where 1s in the image correspond to the pixels belonging to a building. An additional step is also performed to identify the segments touched by the 1s in this image.

Figure 3.17 shows some examples where the results of the relaxation are demonstrated for the sample building patches given in Figure 3.17 (a). The results in Figure 3.17 (b) show that final patches fit building boundaries better.



(a) (b)

Figure 3.17: Relaxation example. (a) Building patches before relaxation. (b) Final patches.

CHAPTER 4

RESULTS AND DISCUSSION

In this chapter, the results of the proposed approach are given by evaluating and discussing its performance on several test data. Performance evaluation approaches are described in the first section. Next, sample images used for testing are given in detail. Results of the experimentation done for the classifier selection are presented right after. Finally, results of the resegmentation and relaxation steps are given.

4.1 Performance Evaluation

To evaluate the performance of the proposed approach, both pixel and object based approaches were used. Following sub sections describe these approaches and their metrics. The ground truths were prepared manually for the test images.

4.1.1 Pixel Based Approach

This approach uses four primitive labels calculated for each pixel:

1. True Positive (TP): It represents the number of pixels which are correctly classified as building with respect to the ground truth.
2. True Negative (TN): It is the number of pixels which are not classified as building and also not labeled as building in the ground truth. Since the

main aim is to evaluate the building detection performance, this measure is not used.

3. False Positive (FP): It corresponds to the number of pixels which are classified as building but do not exist in the ground truth.
4. False Negative (FN): It is the number of building pixels which could not be detected although they exist in the ground truth.

Using these primitives, recall, precision and quality metrics are calculated. Recall is the true positive rate describing the proportion of the actual buildings which is calculated using:

$$recall = \frac{TP}{TP + FN} \tag{4.1}$$

Precision means how many of the pixels classified as buildings are correct. It is calculated in terms of percentage:

$$precision = \frac{TP}{TP + FP} \tag{4.2}$$

Finally, the quality defines the ratio of correctly classified buildings to the actual buildings and all positive results. It is calculated by:

$$quality = \frac{TP}{TP + FP + FN} \tag{4.3}$$

4.1.2 Object Based Approach

For object based performance evaluation, a measure similar to overlapping area matrix (Beauchemin and Thomson, 1997; Cetin et al., 2010) approach is used in this study. The objective is to measure the correct detections, false alarms and

missed detections by using the overlapping area between the ground truth objects and extracted patches.

Suppose that $GT=\{GT_1, GT_2, ..., GT_N\}$ is set of building objects in the ground truth, $O=\{O_1, O_2, ..., O_M\}$ is the set of building objects (patches) in the output map, N is the number of ground truth objects and M is the number of output objects. Then, the number of pixels in the ground truth object GT_i that overlap with the output object O_j is defined as:

$$C_{ij} = Area\left(GT_i \cap O_j\right) \tag{4.4}$$

A ratio between the number of pixels in the overlapping area C_{ij} and the number of pixels in the ground truth object GT_i is calculated using:

$$\alpha = C_{ij} / Area\left(GT_i\right) \tag{4.5}$$

If the ratio α is greater than or equal to a threshold T, then the output object Oj is labeled as a correct detection (i.e., TP). Otherwise, it is labeled as a false alarm (i.e., FP). Furthermore, the ground truth object GT_i is classified as a missed detection (i.e., FN) if none of the output objects overlap with it by a ratio of at least T.

After the values of TP, FP and FN are determined; precision and recall measures are calculated using Equation 4.1 and 4.2, respectively.

4.2 Test Data

The proposed approach was tested on 11 images captured from Google Earth application. Four of these images belong to the same scene captured at different

dates. Information about capture date, image size, spatial resolution, latitude and longitude of the center pixel are shown in Table 4.1:

Table 4.1: Data sets used in the study

Name	Date	Size	Spatial Resolution	Latitude	Longitude
Ankara-1	17.09.2010	240 x 240	2.04 m	39°54'0.03"N	32°48'16.36"E
Ankara-2	17.09.2010	240 x 240	2.04 m	39°54'19.29"N	32°48'14.88"E
Ankara-3	17.09.2010	244 x 244	2.01 m	39°49'50.59"N	32°39'42.12"E
Russia-1	08.06.2008	512 x 512	1.04 m	50°48'11.24"N	40°36'14.14"E
Russia-2	08.06.2008	512 x 512	1.04 m	50°48'27.37"N	40°36'40.11"E
Cyprus	25.09.2010	512 x 512	0.96 m	34°34'56.70"N	32°58'16.41"E
Australia	29.07.2013	512 x 512	0.98 m	37°41'44.74"S	144°52'26.13"E
Canakkale-1	25.04.2009	300 x 300	1 m	40° 9'52.82"N	26°24'38.32"E
Canakkale-2	19.07.2011	300 x 300	1 m	40° 9'52.82"N	26°24'38.32"E
Canakkale-3	26.08.2013	300 x 300	1 m	40° 9'52.82"N	26°24'38.32"E
Canakkale-4	02.09.2013	300 x 300	1 m	40° 9'52.82"N	26°24'38.32"E

The spectral resolution of all test images is 8 bits per pixel, thus, pixel values are between 0 and 255. However, these images have different characteristics which provide opportunity to examine the effects of the change in building size, density, shape and shadow amount.

In order to determine the spatial resolution of test images, the scale bar inherent in Google Earth application is used, since the application does not provide such information. For this purpose, the actual length of the bar (shown in meters) is divided to the number of pixels wide. For instance, the length of the scale bar for the image given in Figure 4.1 is 274 m and is measured as 274 pixels. Therefore, the spatial resolution is roughly calculated as 274 / 274 = 1 m.

Figure 4.1: The scale bar in Google Earth application.

4.3 Determination of the Classifier

In this section, a preliminary experimentation which was conducted to decide which classifier should be used in the resegmentation step is explained. To that end, five different classifiers described in Chapter 3 were employed on Ankara-1 image (Figure 4.2) to classify buildings.

Figure 4.2: Ankara-1 test image.

Performances are assessed by the pixel based approach explained in section 4.1.1. Table 4.2 shows the results which were obtained without applying the relaxation

step. QDA was the best classifier in terms of building detection rate (i.e., recall) and quality, although the precision value is not the highest. A more precise result was obtained by AdaBoost which also has the second highest recall and quality values. However, the difference between the recall values of these two classifiers is so high that AdaBoost could not be considered. High recall values are crucial for the relaxation step since relaxation is performed on the regions classified as 'Positive' which should cover as much building as possible. In addition, QDA is easier to utilize since it does not require parameter tuning or any preliminary analysis to setup. However, in AdaBoost, a proper weak classifier should be selected and its parameters must be specified.

Table 4.2: Performances of classifiers on Ankara-1 test image

Classifier	**True Positive**	**False Positive**	**False Negative**	**Recall %**	**Precision %**	**Quality %**
QDA	9052	3536	2413	78.95	71.57	60.10
LDA	8151	4212	3314	71.09	65.93	51.99
LR	8046	4197	3419	70.18	65.72	51.37
SVM	7346	2437	4119	64.07	75.09	52.84
AdaBoost	8293	2858	3172	72.33	74.37	57.90

4.4 Determination of parameter v

v is a parameter of Once-Class SVM which takes values between 0 and 1 (Scholkopf et al., 2001). For instance, when $v = 0.05$, at least 5% of training data are selected as support vectors and at most 5% of training examples are misclassified at the cost of a small margin. With higher values, the model generalizes the training samples better. However, this causes higher training errors, since potentially more samples are misclassified in training.

To determine a proper value for *v*, the relaxation procedure was executed with its different values on Ankara-1 test image (Figure 4.2). The results were evaluated with the pixel based approach by calculating the recall, precision and quality measures. Figure 4.3 shows a graph to demonstrate how these measures change with the value of *v*.

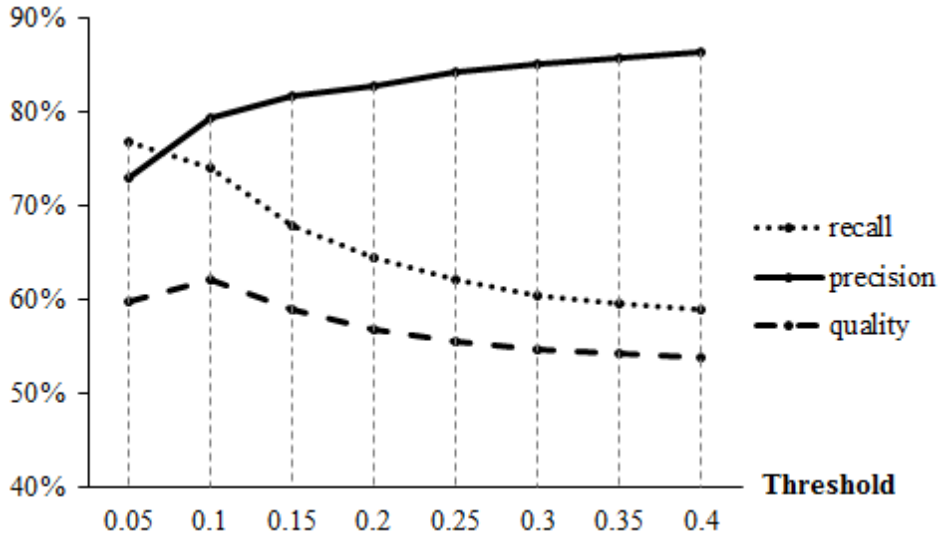


Figure 4.3: Effect of parameter v.

As shown in Figure 4.3, recall values decrease as the value of *v* increases. This is not surprising since Once-Class SVM is trained on the regions somehow known to be buildings. Thus, the classification accuracy tends to decrease due to the higher training errors brought by larger values of *v*. On the other hand, the increase in precision is related to the decrease in recall; since fewer regions are classified as building after the procedure, less false alarms occur. To determine the value of *v*, quality measure is considered since making a decision by examining other two measures is difficult due to the inverse proportion they have. As a result, the value which corresponds to the highest quality value is selected (i.e., $v = 0.1$). The quality measure can also be used to decide whether the

relaxation step improves the result of the resegmentation or not, by comparing the quality values before and after the relaxation.

4.5 Results and Discussion

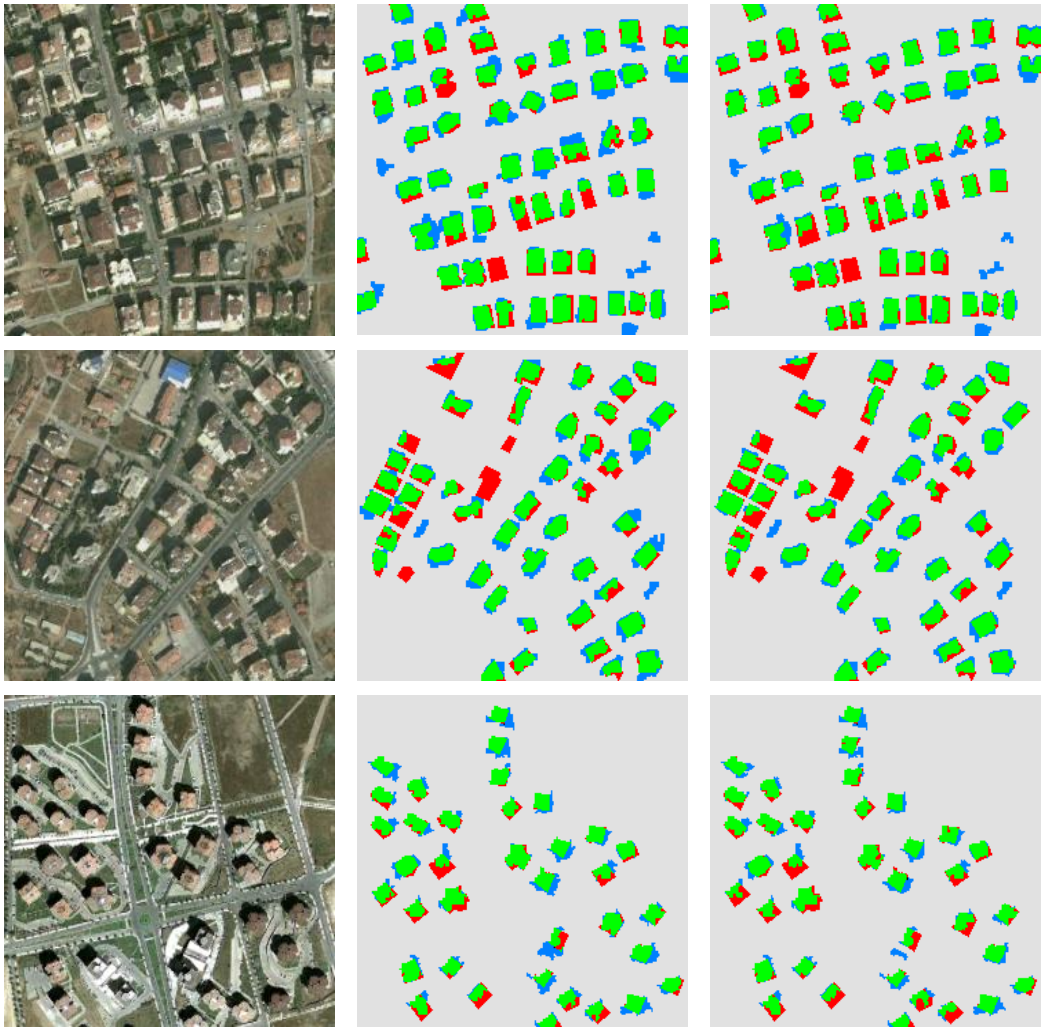
All steps of the proposed algorithm were implemented with MATLAB R2010a software. For each test image, training samples were selected in such a way that they cover approximately 25% of the total number of building polygons in the ground truth image. First, a number of building polygons in ground truth are selected manually. Then, image segments to be used in training are determined automatically as the segments inside and outside of these polygons. Figure 4.4 illustrates this process for the ground truths given in Figure 4.4 (a). Manually selected building polygons are shown as green in Figure 4.4 (b). Corresponding image segments identified for training are shown in Figure 4.4 (c) where green and red segments refer to 'Positive' and 'Negative' samples, respectively.

(a) (b) (c)

Figure 4.4: An example for determination of the training samples. (a) Ground Truths. (b) Buildings selected for training. (c) Segments identified for training.

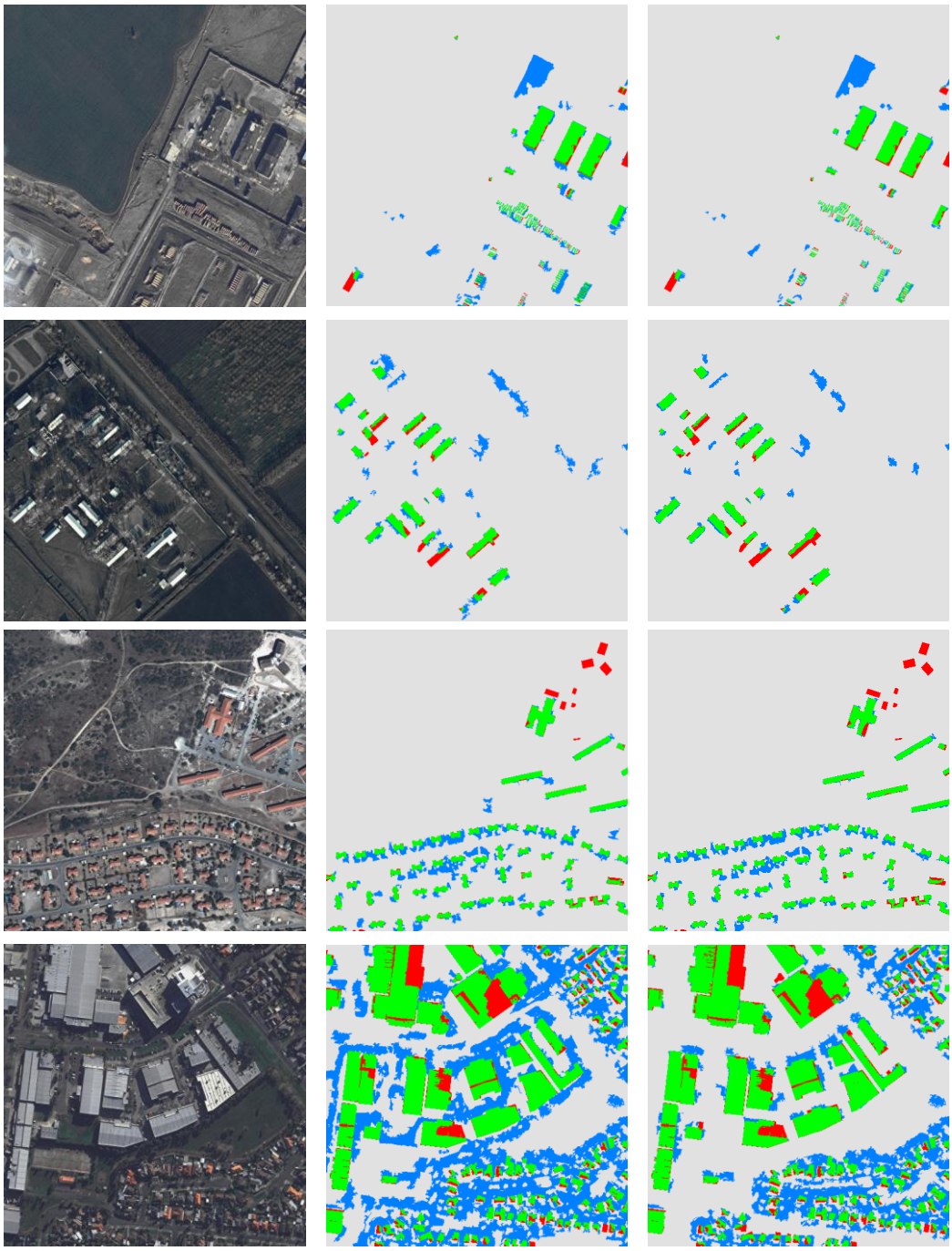
Visual results of the resegmentation and relaxation procedures are given in Figure 4.5, 4.6 and 4.7. Figure 4.5 shows the results of Ankara sets and Figure 4.7 demonstrates the results of the images taken from the same scene in Canakkale at

different times. Figure 4.6 presents the results of the other images which are not taken from Turkey. In all figures; green, blue and red pixels in the resultant images correspond to TP, FP and FN, respectively.



(a) (b) (c)

Figure 4.5: Visual results of Ankara sets. (a) Original images: Ankara-1, 2 and 3 from top to bottom. (b) Results of resegmentation. (c) Results after relaxation.



(a) (b) (c)

Figure 4.6: Visual results of images from other countries. (a) Original images: Russia-1 and 2, Cyprus and Australia from top to bottom. (b) Results of resegmentation. (c) Results after relaxation.

(a) (b) (c)

Figure 4.7: Visual results of Canakkale sets. (a) Original images: Canakkale-1, 2, 3 and 4 from top to bottom. (b) Results of resegmentation. (c) Results after relaxation.

In Table 4.3, pixel based performance evaluation results of the proposed method on different test images are given. In addition, object based performances of the resegmentation and relaxation procedures are presented in Table 4.4 and 4.5, respectively. Object based measures are calculated for varying overlapping thresholds (i.e., T).

Table 4.3: Pixel based performance evaluation results

	Initial Results			**After Relaxation**		
	Recall %	**Precision %**	**Quality %**	**Recall %**	**Precision %**	**Quality %**
Ankara-1	78.95	71.57	60.10	74.06	79.31	62.07
Ankara-2	73.19	71.58	56.71	71.16	75.96	58.08
Ankara-3	81.22	75.02	63.93	77.08	77.30	62.85
Russia-1	84.50	53.25	48.51	82.26	63.74	56.04
Russia-2	78.66	56.86	49.26	73.76	65.41	53.06
Cyprus	86.61	59.71	54.66	84.51	65.20	58.24
Australia	83.27	42.95	39.54	81.89	58.86	52.08
Canakkale-1	78.61	82.22	67.19	74.26	86.63	66.62
Canakkale-2	67.35	65.89	49.94	65.07	70.02	50.89
Canakkale-3	65.87	66.87	49.66	63.73	70.58	50.35
Canakkale-4	77.49	72.00	59.54	75.11	74.76	59.92

Examining Figure 4.5 and Tables 4.3, 4.4 and 4.5 together; it could be said that the proposed algorithm performed well on test images of Ankara. In these images, building shapes are rather regular, sizes do not vary much and shadows are visually apparent in common. Still, there exist some false detections originating from the shadows of other objects like trees. Shadows are the reasons of the missed detections, too. The algorithm failed to detect buildings having low shadow amounts, possibly due to low building heights.

Table 4.4: Object based performance evaluation results before relaxation

	T	0.4	0.5	0.6	0.7	0.8
Ankara-1	Recall %	94.64	92.86	89.29	78.57	57.14
	Precision %	89.83	88.14	84.75	74.58	54.24
Ankara-2	Recall %	87.50	83.33	81.25	77.08	41.67
	Precision %	89.36	85.11	82.98	78.72	42.55
Ankara-3	Recall %	97.22	97.22	91.67	80.56	66.67
	Precision %	97.22	97.22	91.67	80.56	66.67
Russia-1	Recall %	84.30	83.47	81.82	80.17	76.86
	Precision %	82.93	82.11	80.49	78.86	75.61
Russia-2	Recall %	86.96	78.26	73.91	65.22	65.22
	Precision %	50.00	45.00	42.50	37.50	37.50
Cyprus	Recall %	86.36	86.36	84.85	84.85	81.82
	Precision %	74.03	74.03	72.73	72.73	70.13
Australia	Recall %	95.31	91.41	82.81	69.53	51.56
	Precision %	50.21	48.15	43.62	36.48	28.21
Canakkale-1	Recall %	92.31	92.31	89.74	82.05	66.67
	Precision %	90.00	90.00	87.50	80.00	65.00
Canakkale-2	Recall %	76.92	71.79	71.79	66.67	61.54
	Precision %	81.08	75.68	75.68	70.27	64.86
Canakkale-3	Recall %	82.05	79.49	79.49	64.10	43.59
	Precision %	74.42	72.09	72.09	58.14	39.53
Canakkale-4	Recall %	92.31	87.18	82.05	82.05	71.79
	Precision %	87.80	82.93	78.05	78.05	68.29

Table 4.5: Object based performance evaluation results after relaxation

	T	0.4	0.5	0.6	0.7	0.8
Ankara-1	Recall %	92.86	92.86	87.50	69.64	48.21
	Precision %	88.14	88.14	83.05	66.10	45.76
Ankara-2	Recall %	85.42	83.33	79.17	68.75	39.58
	Precision %	87.23	85.11	80.85	70.21	40.43
Ankara-3	Recall %	97.22	91.67	80.56	69.44	63.89
	Precision %	94.59	89.19	78.38	67.57	62.16
Russia-1	Recall %	80.17	79.34	77.69	75.21	72.73
	Precision %	86.61	85.71	83.93	81.98	77.88
Russia-2	Recall %	86.96	73.91	69.57	60.87	52.17
	Precision %	66.67	56.67	53.33	46.67	40.00
Cyprus	Recall %	86.36	84.85	84.85	81.82	77.27
	Precision %	81.43	80.00	80.00	77.14	72.86
Australia	Recall %	93.75	88.28	78.91	65.08	47.66
	Precision %	62.50	59.47	54.59	44.57	34.66
Canakkale-1	Recall %	94.74	92.31	85.37	74.42	60.47
	Precision %	92.31	90.00	83.33	72.73	59.09
Canakkale-2	Recall %	75.00	70.00	70.00	66.67	60.00
	Precision %	85.71	80.00	80.00	76.47	68.57
Canakkale-3	Recall %	84.21	81.58	79.49	62.50	36.96
	Precision %	78.05	75.61	73.81	58.14	34.69
Canakkale-4	Recall %	92.31	87.18	82.05	74.36	67.50
	Precision %	87.80	82.93	78.05	70.73	64.29

Speaking for the pixel based evaluation on Ankara sets (Table 4.3), relaxation procedure did not improve the results much, even reduced the quality for the

results of Ankara-3 image which has the highest recall and precision values. Considering object based results (Tables 4.4 and 4.5), even precision values are decreased a little, in addition to recall. The main reason should be the high precision values initially obtained after resegmentation, since the results look improved for the images in Figure 4.6 where the initial precisions are lower. As a conclusion, it may not worth using the relaxation step when a certain degree of satisfactory results are achieved after resegmentation, such as when precisions are higher than 72-73%. Thus, relaxation can be left as optional for this case. In overall, algorithm was able to detect at least 70% of each building polygon with a detection rate of 77% or higher without relaxation (Table 4.4).

Test images in Figure 4.6 represent more complex cases when compared to the ones in Figure 4.5. In general, building sizes vary much and shadow of other objects exist more in these images. The latter is found to be the main reason of the large falsely detected regions in Russia images. An interesting result is observed on Cyprus set that several buildings in top right corner could not be detected although their shadows are clearly visible. This might originate from the fact that red color is very dominant for the roofs of the buildings in this image. Thus, the color feature might have been more discriminative when compared to other features. Highest false detections are observed in the test image of Australia where building size and shapes differ more than any other set. In this set, missed detections occurred more on large buildings, whereas false detections are located more around small buildings. The main reason is the distance of pixels on buildings to the shadows differs greatly in image depending on the size of the buildings. Small buildings in this set are also dense. As a result, shadows of some of the adjacent buildings look merged. Therefore, the algorithm detected these buildings as a single object.

Despite the false detections and lowest precision values observed on images in Figure 4.6, the proposed method managed to achieve highest correct detection rates (Tables 4.3, 4.4 and 4.5) compared to other sets. High amount of ‘Positive’

results make these sets appropriate for relaxation which is supported by both pixel and object based results. As a result, pixel based precision values are improved higher than the decrease in recall values. Highest increase in precision is observed in Australia set (from 42.95% to 58.86%). Same observation is also valid for object based results. Thus, as oppose to Ankara sets, relaxation step proves to be useful in these images. Although the relaxation step was not able to achieve 70% precision in pixel based evaluation, its use is mandatory after running resegmentation on the images having similar characteristics to the ones in Figure 4.6.

Images in Figure 4.7 were used to evaluate the performance of the algorithm with varying amount of shadows and off-nadir angles. For all images in Figure 4.7, same buildings (Figure 4.4 (b)) were used for training in order to get a meaningful comparison. Two of these images (Canakkale-2 and 3) were not taken at nadir. It is observed that FPs around buildings are mostly occurred in these images and pixel based results support this statement (Table 4.3). Apart from that, shadow amount seems to play a role, too. This is because, although both Canakkale-1 and 4 are taken at nadir, FPs in Canakkale-4 are visibly more due to the larger regions of shadows. Since larger shadows produce larger RoIs, more segments fell inside those RoIs and as a result, more FPs occurred. Back to the 2nd and 3rd images, white buildings in top-right could not be detected or partially detected although their shadows are visible. However, for the images taken at nadir (Canakkale-1 and 4), these buildings could be detected.

When object based performances on the images in Figure 4.7 are considered (Table 4.4), the results are similar. 82% of the buildings were detected with an areal coverage of at least 70% for the images taken at nadir (i.e., Canakkale-1 and 4). However, detection rates were dropped to 74% after relaxation for these images (Table 4.5). As in Ankara images, relaxation fails to improve the initial results with high precision. However, the results on other two images (Canakkale-2 and 3) are improved as pixel based precision values are increased

from 65%s to 70%s (Table 4.3). A similar improvement can also be observed in object based results (Table 4.4 and 4.5).

Test images in Figure 4.7 revealed that off-nadir angle and shadow amount affects the performance of the proposed approach more than the shape and size of the buildings. The main cause for the change in shadow amount is the sun angle which is affected by the time of the day, the season of the year and also the latitude of the place. Earth observation satellites are generally sun-synchronous and they pass over a place at the same local solar time. Therefore, the surface illumination angle is nearly the same each time they obtain an imagery of that place. However, for different seasons in a year, the height of the sun in sky should vary even though the local time is same. Therefore, shadow heights can differ throughout a year although the local time is same. On the other hand, time of the day becomes an important factor, especially when working on airborne images. Unlike satellite images, they can be captured at any time of the day. However, off-nadir angle for imaging the same swath width is generally higher when compared to satellite images. The reason is that the angular field of view for the same swath width is smaller for satellites, since they are further away from the Earth's surface. Finally, due to the Earth’s curvature, sun angle varies for the places at different latitudes. Thus, shadow heights for these places differ even they are at the same longitude and have same local time.

CHAPTER 5

CONCLUSION

In this thesis, a trainable method is proposed for detecting buildings in RGB satellite images. The method mainly relies on the shadow cue, thus, it tries to explain how much and well buildings can be extracted using shadow-derived features in general. To that end, first, input image is partitioned by over-segmentation and then, shadow segments are detected. Using these shadow objects, two new features are introduced representing closeness to shadows and presence in the region of interest derived from each shadow object. Together with the color information calculated as a normalized ratio between the red and green components, these features are utilized to classify each segment using Quadratic Discriminant Analysis. Finally, a fully automated one-class modeling algorithm is proposed to refine the results of this classification which is found to be useful if the initial results are not precise enough.

Examining the results in Chapter 4.5, the method produces promising results for the buildings having different complexity and irregularities, although it only makes use of RGB information. Experiments showed that the algorithm can work satisfactorily on Google Earth images which are publicly and freely available. In addition, installing the application provides to access temporal images from almost anywhere throughout the world, thus, it makes a building extraction algorithm which uses only RGB information valuable.

One of the advantages of the proposed method is that its steps are independent from each other so that algorithms in each of these steps can be replaced by other appropriate ones. For instance, feature extraction step takes shadow regions and image segments as its inputs. Therefore, any image segmentation and shadow detection algorithm could be used to replace the ones existed in this study. Moreover, the resegmentation step can work with any binary classifier which is already shown in section 4.3 where several classifiers were tested on a sample image.

The main drawback of the proposed method is that it relies too much on amount of the shadows detected. The method could not detect buildings when their shadows could not be found, as expected. In addition, the method could partly detect a building if the area of its shadow is not greater than a certain amount. The reverse case is also problematic; most of the false alarms around buildings occur due to the high volume of shadows. Furthermore, the performance of the method is affected negatively when the size of the buildings in the scene varies much, which causes big changes in distance to shadow from building to building. Another problem occurs when images are not captured at nadir, thus, the methods tends to detect sides of the buildings in addition to their roofs.

The method can be improved in the areas explained below:

- Shadow detection: Significant improvements can be achieved with a different algorithm specifically designed to detect the shadows of buildings, only. This is because, shadows of objects other than buildings are observed to be the main reasons of the false alarms. As a suggestion, the pattern and shape of the cast shadow can be examined to introduce a novel algorithm. Also, it is possible to eliminate some thin and long shadow regions which occur around walls using PCA. Furthermore, if the near-infrared band is available, most of the shadows originating from trees and shrubs can be eliminated.

- Feature detection: The proposed method utilizes only three features for the classification process. Thus, introducing more features which are specifically created to identify buildings may improve the classification results, noticeably. Also, presenting more features that are not based on shadow cue will give the method the ability to detect buildings when their shadow could not be detected.

- Relaxation: The relaxation procedure is not capable of detecting buildings which are missed in the previous classification, since it works on the detected regions, only. Some of these regions inside missed buildings might have been removed by area thresholding in the post-processing step. Therefore, instead of removing them, creating a buffer zone around these small regions can make the relaxation algorithm run on these regions, too.

- One-Class modeling: In the relaxation step, One-Class SVM is used to determine similar pixels around the ones selected for training. Instead, Gaussian mixture models or sparse linear representations can be utilized in the process.

REFERENCES

Akcay, H.G. and Aksoy, S., (2010), Building Detection Using Directional Spatial Constraints, IEEE International Geoscience and Remote Sensing Symposium, Honolulu, Hawaii, July 25-30.

Attarzadeh, R. and Momeni, M., (2012), Object-Based Building Extraction from High Resolution Satellite Imagery, ISPRS: Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol.XXXIX-B4, pp.57-60, doi:10.5194/isprsarchives-XXXIX-B4-57-2012.

Aytekin, O., Erener, A., Ulusoy, I. and Duzgun, S., (2012), Unsupervised building detection in complex urban environments from multispectral satellite imagery, International Journal of Remote Sensing, 33(7):2152-2177, doi:10.1080/01431161.2011.606852

Baatz, M., and Schäpe, A., (2000), Multiresolution segmentation - an optimization approach for high quality multi-scale image segmentation, In Strobl, J. et al. (eds.): Angewandte Geographische Informationsverarbeitung XII. Wichmann, Heidelberg, pp.12-23.

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

Benarchid, O., Raissouni, N., (2013), Support Vector Machines for Object Based Building Extraction in Suburban Area using Very High Resolution Satellite Images, a Case Study: Tetuan, Morocco, IAES International Journal of Artificial Intelligence, 2(1):43-50.

Blaschke, T., (2003), Object-based contextual image classification built on image segmentation, IEEE Advances in Techniques for Analysis of Remotely Sensed Data, ISBN:0-7803-8350-8, pp.113-119.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., (1984), Classification and Regression Trees, Monterey, CA: Wadsworth & Brooks / Cole Advanced Books & Software, ISBN 978-0-412-04841-8.

Cetin, M., Halici, U. and Aytekin O., (2010), Building detection in satellite images by textural features and Adaboost, 2010 IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS).

Chang, C.-C. and Lin, C.-J., (2011) LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chen, K., & Blong, R. (2002), Extracting building features from high resolution aerial imagery for natural hazards risk assessment, IGARSS'02, IEEE International Geoscience and Remote Sensing Symposium, vol.4, pp.2039-2041.

Comaniciu, D. and Meer, P., (2002), Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, pp.603-619.

Cortes, C. and Vapnik, V., (1995), Support-vector networks, Machine learning, 20(3):273-297.

Cretu, A.M. and Payeur, P., (2012), Visual Attention Model with Adaptive Weighting of Conspicuity Maps for Building Detection in Satellite Images, International Journal on Smart Sensing and Intelligent Systems, 5(4):742-766.

Dahiya, S., Garg, P.K. and Jat, M.K. (2013), Object oriented approach for building extraction from high resolution satellite images, 3rd International Advance Computing Conference (IACC), pp.1300-1305.

Dini, G.R., Jacobsen, K. and Heipke, C., (2013), Delineation of Building Footprints from High Resolution Satellite Stereo Imagery Using Image Matching and a GIS Database, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1(1):81-85.

EDISON, Edge Detection and Image SegmentatiON system web page, *Retrieved 12 August 2010*,
[Url: http://coewww.rutgers.edu/riul/research/code/EDISON/index.html]

ENVI Manual, (2006), Version 4.3.

Freund, Y. and Schapire, R.E., (1997), A Decision-Theoretic Generalization of on-line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1):119–139.

Gao, Y. and Mas, J.F., (2008), A comparison of the performance of pixel based and object based classifications over images with various spatial resolutions, Online J. Earth Sci., 2(1):27-35.

Gerke, M., Heipke, C. and Straub, B.M., (2001), Building Extraction from Aerial Imagery Using a Generic Scene Model and Invariant Geometric Moments, IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp.85-89, Rome, Italy.

GML AdaBoost Matlab Toolbox v.0.3, *Retrieved 17 September 2011,* [Url: http://graphics.cs.msu.ru/en/science/research/machinelearning/adaboosttoolbox]

Gonzales, R.C. and Woods, R.E., (2008), Digital Image Processing, 3rd Ed., chapter 9, pp.627-635, Prentice-Hall, Upper Saddle River, New Jersey, USA.

Guo, Z., Luo, L., Wang, W. and Du, S. (2013), Data fusion of high-resolution satellite imagery and GIS data for automatic building extraction, ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1(1):23-28.

Hastie, T., Tibshirani, R. and J. Friedman, (2008), Linear methods for classification, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer-Verlag, 2008, pp. 106–113.

Hay, G.J., and Gastilla, G., (2006), Object-based image analysis: strength, weakness, opportunities, and threats (SWOT), 1st International Conference on Object-based Image Analysis (OBIA 2006), 4-5 July, Salzburg, Austria.

Hong, L., Yang, K., (2011), A novel segmentation method of high resolution remote sensing image based on multi-feature object-oriented Markov random fields model, RSETE, pp.8019-8024.

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, Pattern Analysis and Machine Intelligence, IEEE Transactions, vol.18, no.7, pp.673-689.

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

Inglada, J., (2007), Automatic Recognition of Man-made Objects in High Resolution Optical Remote Sensing Images by SVM Classification of Geometric Image Features, ISPRS Journal of Photogrammetry and Remote Sensing, 62(3), pp.236-248.

Jabari, S. and Zhang, Y., (2014), Building Detection in Very High Resolution Satellite Image using HIS Model, ASPRS 2014 Annual Conference, Louisville, Kentucky, 23-28 March.

James, G., Witten, D., Hastie, T. and Tibshirani, R., (2013), An Introduction to Statistical Learning, chapter 4, pp.130-133, Springer Texts in Statistics, vol.103, ISBN 978-1-4614-7138-7, New York.

Jaynes, C., Riseman, E. and Hanson, A., (2003), Recognition and Reconstruction of Buildings from Multiple Aerial Images, Computer Vision and Image Understanding, vol.90, pp.68-98.

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.

Karantzalos, K. and Paragios, N., (2009), Recognition-Driven Two-Dimensional Competing Priors Toward Automatic and Accurate Building Detection, IEEE Transactions on Geoscience and Remote Sensing, 47(1):133-144.

Kim, T., Lee, T.Y. and Kim, K.O., (2006), Semiautomatic Building Line Extraction from Ikonos Images Through Monoscopic Line Analysis, Photogrammetric Engineering and Remote Sensing, 72(5):541-549.

Korting, T.S., Dutra L.V. and Fonseca L.M.G., (2011), A resegmentation approach for detecting rectangular objects in high-resolution imagery, IEEE Geosci. Remote Sens. Lett., vol.8, no.4, pp.621-625.

Korting, T.S., Fonseca, L.M., Dutra L.V. and Silva, (2008), F. C., Image resegmentation applied to urban imagery, Int. Archives Photogram. Remote Sens. Spatial Inf. Sci., vol.37, pt.B3b, pp.393–398.

Lafarge, F., Descombes, X., Zerubia, J. and Pierrot-Deseilligny, M., (2010), Structural Approach for Building Reconstruction from a Single DSM, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(1):135-146.

Lee, D.H., Lee, K.M. and Lee, S.U., (2008), Fusion of Lidar and Imagery for Reliable Building Extraction, Photogrammetric Engineering and Remote Sensing, 74(2):215-225.

Lee, D.S., Shan, J. and Bethel, J.S., (2003), Class-Guided Building Extraction from Ikonos Imagery, Photogrammetric Engineering and Remote Sensing, 69(2):143-150.

Lefèvre, S., Weber, J., Sheeren, D., (2007), Automatic building extraction in VHR images using advanced morphological operators, in: IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN), Paris.

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

Lin, C., Huertas, A. and Nevatia, R., (1994), Detection of Buildings Using Perceptual Grouping and Shadows, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'94, pp.62-69, Seattle, USA.

Liu, H., Xu, Z. and Abraham, A., (2005), Hybrid Fuzzy-Genetic Algorithm Approach for Crew Grouping, Proceedings of 5th International Conference on Intelligent Systems Design and Applications, ISDA'05, pp.332-337, Wroclaw, Poland.

Liu, J.G. and Mason, P.J., (2009), Algebraic operations (multi-image point operations), Essential Image Processing and GIS for Remote Sensing, Hoboken, NJ, USA: Wiley.

Liu, Z., Cui, S. and Yan, Q., (2008), Building Extraction from High Resolution Satellite Imagery Based on Multi-scale Image Segmentation and Model Matching, Proceedings of 2008 International Workshop on Earth Observation and Remote Sensing Applications, Beijing, China.

Mayunga, S.D., Coleman, D.J. and Zhang, Y., (2007), A Semi-automated Approach for Extracting Buildings from Quickbird Imagery Applied to Informal Settlement Mapping, International Journal of Remote Sensing, 28(10):2343-2357.

Michel, J., Grizonnet, M. and Canevet, O., (2012), Supervised re-segmentation for very high-resolution satellite images, Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2012, pp.68–71.

Mirhassani, S.M., Yousefi, B. and Moghaddamjoo, A., (2012), Spectral Information Adjustment Using Unsharp Masking and Bayesian Classifier for Automatic Building Extraction from Urban Satellite Imagery, Journal of American Science, 8(1):554-564.

Ok, A.O., (2013), Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts, ISPRS Journal of Photogrammetry and Remote Sensing, 86, pp.21-40.

Otsu N., (1979), A threshold selection method from graylevel histograms, IEEE Trans. Syst. Man Cybern., vol. SMC-9, no. 1, pp. 62–66.

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):1289-1307.

Perkins, S., Edlund, K., Esch-Mosher, D., Eads, D., Harvey, N. and Brumby, S., (2005), Genie Pro: Robust Image Classification Using Shape, Texture and Spectral Information, Proceedings of SPIE, vol.5806. pp.139-148.

Perkins, S., Theiler, J., Brumby, S.P., Harvey, N.R., Porter, R., Szymanski, J.J. and Bloch, J.J., (2000), GENIE: A Hybrid Genetic Algorithm for Feature Classification in Multi-Spectral Images, Proceedings of SPIE, vol.4120, pp.52-62.

Riaño, D., Chuvieco, E., Salas, J. and Aguado, I., (2003), Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types, IEEE Trans. Geosci. Remote Sens., vol.41, no.5, pp.1056–1061.

Rottensteiner, F. and Briese, C., (2002), A New Method for Building Extraction in Urban Areas from High-Resolution Lidar Data, Proceedings of the Photogrammetric Computer Vision, ISPRS Commission III Symposium, PCV'02, Graz, Austria.

San, D.K., (2009), Approaches for Automatic Urban Building Extraction and Updating from High Resolution Satellite Imagery, PhD Thesis submitted to the Graduate School of Natural and Applied Sciences, Middle East Technical University, Turkey.

San, D.K. and Turker, M., (2007), Automatic Building Extraction from High Resolution Stereo Satellite Images, Conference on Information Extraction from SAR and Optical Data with Emphasis on Developing Countries, May 16-18, Istanbul, Turkey, CD.

Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C., (2001), Estimating the support of a high-dimensional distribution, Neural Computation, 13(7):1443-1471.

Shackelford, A.K. and Davis, C.H., (2003), A Combined Fuzzy Pixel-Based and Object-Based Approach for Classification of High-Resolution Multispectral Data Over Urban Areas, IEEE Transactions on Geoscience and Remote Sensing, 41(10):2354-2363.

Shai Bagon's Matlab Code, *Retrieved 12 August 2010*, [Url: http://www.wisdom.weizmann.ac.il/~bagon/matlab_code/edison_matlab_interface.tar.gz]

Simonetto, E., Oriot, H. and Garello, R., (2005), Rectangular Building Extraction from Stereoscopic Airborne Radar Images, IEEE Transactions on Geoscience and Remote Sensing, 43(10):2386-2395.

Sirmacek, B. and Unsalan, C., (2009), Urban-Area and Building Detection Using SIFT Keypoints and Graph Theory, IEEE Transactions on Geoscience and Remote Sensing, 47(4):1156-1167.

Sirmacek, B. and Unsalan, C., (2011), A probabilistic framework to detect buildings in aerial and satellite images, IEEE Transactions on Geoscience and Remote Sensing, 49(1):211-221.

Song, Y.H., Shan, J., (2008), Building Extraction from High Resolution Color Imagery Based on Edge Flow Driven Active Contour And JSEG, Proceedings of International Society for Photogrammetry and Remote Sensing (ISPRS'08) Congress, Beijing, China.

Sun, X., Fu, K., Long, H., Hu, Y., Cai, L. and Wang, H., (2008), Contextual models for automatic building extraction in high resolution remote sensing image using object-based boosting method, Proc. 2008 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), vol.2, pp.437-440.

Sumer, E., (2011), Automatic reconstruction of photorealistic 3-D building models from satellite and ground-level images, PhD Thesis submitted to the Graduate School of Natural and Applied Sciences, Middle East Technical University, Turkey.

Sumer, E. and Turker, M., (2013), An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images, Computers, Environment and Urban Systems, vol.39, pp.48-62.

Tan, Q., Wei, Q. and Liang, F., (2010), Building extraction from VHR multi-spectral images using rule-based object-oriented method: A case study, Geoscience and Remote Sensing Symposium, IGARSS'2010, pp.2754-2756.

Tsai, V.J.D., (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, pp.1661-1671.

Tseng, Y. and Wang, S., (2003), Semiautomated Building Extraction Based on CSG Model-Image Fitting, Photogrammetric Engineering and Remote Sensing, 69(2):171-180.

Turker, M. and San, D.K., (2010), Building detection from pan-sharpened IKONOS imagery through support vector machines classification, Proc. of ISPRS Technical Com. VIII Symposium, Volume XXXVIII - Part 8, Kyoto, Japan.

Wei, Y., Zhao, Z. and Song, J., (2004), Urban Building Extraction from High-Resolution Satellite Panchromatic Image Using Clustering and Edge Detection, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, pp.2008-2010, Anchorage, Alaska, USA.

Venables, W.N. and Ripley, B.D., (2002), Modern applied statistics with S, Fourth edition. Springer, Berlin, ISBN: 0-387-95457-0, p.495.

Vezhnevets, A. and Vezhnevets, V., (2005), Modest AdaBoost - Teaching AdaBoost to Generalize Better, Graphicon, vol.12, no.5, pp.987-997.

Yang, G., Duan, F., Zhao, W., Zhao, W., Zhang, L., (2010), Building extraction in towns and villages based on Digital Aerial Image by texture enhancing, 18th International Conference on Geoinformatics: GIScience in Change, Beijing, China, pp. 1-6.

Youssef, M.M.S., Mallet, C., Chehata, N., Le Bris, A., Gressin, A., (2014), Combining top-down and bottom-up approaches for building detection in a single very high resolution satellite image, Proc. of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 13-18 July, Québec City, Canada.

Yuksel, B., (2012), Automated Building Detection from Satellite Images by using only Shadow Information as an object Invariant, MSc Thesis submitted to the Graduate School of Natural and Applied Sciences, Middle East Technical University, Turkey.

Zhang, Y. (1999), Optimisation of building detection in satellite images by combining multispectral classification and texture filtering, ISPRS Journal of Photogrammetry and Remote Sensing, 54(1):50-60.

Zhang, K., Yan, J. and Chen, S.C., (2006), Automatic Construction of Building Footprints from Airborne LIDAR Data, IEEE Transactions on Geoscience and Remote Sensing, 44(9):2523-2533.

CURRICULUM VITAE

PERSONEL INFORMATION

Surname, Name	: Dikmen, Mehmet
Nationality	: Turkish (TC)
Date and Place of Birth	: 10 July 1981, İskenderun
Marital Status	: Single
Phone	: +90 312 246 66 66 / 1306
Fax	: +90 312 246 66 60
Email	: mdikmen@baskent.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
Ph.D.	METU GGIT	2014
M.Sc.	METU EE	2006
B.Sc.	Baskent University Computer Engineering	2003
High School	Gaziantep Vehbi Dinçerler Fen Lisesi	1999

WORK EXPERIENCE

Year	Place	Enrollment
2003 – Present	Baskent University – Dept. of Computer Eng.	Research Asst.
2013 – 2014	Netcad Yazılım A.Ş. – Dept. of R&D	Sr. Specialist

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Dikmen, M. and Halici, U., 2014, "A Learning-Based Resegmentation Method for Extraction of Buildings in Satellite Images", IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 12, pp. 2150-2153.

Mutlu, G., Guzel, M.S., Haciomeroglu, M., **Dikmen, M.** and Sever, H., 2014, “Silhouette Extraction from Street View Images”, International Journal of Advanced Robotic Systems, 11:114. doi: 10.5772/58740.

Dikmen, M., Eker, A.M., Canbazoğlu, S. and Akgün, H., 2014, "Heyelan Duyarlılık Haritalanmasında Mekansal Çözünürlüğün Etkisi", 3. Jeolojik Uzaktan Algılama Sempozyumu (JEOUZAL 2014), 13-14 Mart, Antalya.

Eker, A.M., **Dikmen, M.**, Cambazoğlu, S., Düzgün, Ş. and Akgün, H., 2012, "Bartın, Ulus İlçesi için Yapay Sinir Ağı ve Lojistik Regresyon Yöntemlerinin Heyelan Duyarlılık Çalışmasına Uygulanması ve Karşılaştırılması", Gazi Üniv. Müh. Mim. Fak. Der., Cilt 27, No 1, 163-173.

Eker, A.M., **Dikmen, M.**, Cambazoğlu, S., Düzgün, Ş. and Akgün, H., 2011, "Heyelan Duyarlılık Haritalamasında Veri Seti Hazırlanması ve Analizi", 64. Türkiye Jeoloji Kurultayı, 25-29 Mart, Ankara.

Ergezer, H., **Dikmen, M.** and Özdemir, E., 2013, "Artificial Neural Networks based Face Recognition using Discrete Cosine Transform", TAINN '03: Turkish Artificial Intelligence and Neural Networks Symposium, 2-4 July.

HOBBIES

Computer Technologies, Table Tennis, Basketball, Movies, Formula 1.