ROAD DETECTION BY MEAN SHIFT SEGMENTATION AND STRUCTURAL ANALYSIS

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

BY

MUSTAFA DURSUN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2012

Approval of the thesis:

ROAD DETECTION BY MEAN SHIFT SEGMENTATION AND STRUCTURAL ANALYSIS

submitted by **MUSTAFA DURSUN** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

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

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering** ________________

Prof. Dr. Uğur Halıcı
Supervisor, **Department of Electrical and Electronics Engineering** ________________

Examining Committee Members:

Prof. Dr. Tolga Çiloğlu
Department of Electrical and Electronics Engineering, METU ________________

Prof. Dr. Uğur Halıcı
Department of Electrical and Electronics Engineering, METU ________________

Prof. Dr. Aydın Alatan
Department of Electrical and Electronics Engineering, METU ________________

Prof. Dr. Gözde Bozdağı Akar
Department of Electrical and Electronics Engineering, METU ________________

Prof. Dr. Yasemin Yardımcı
Informatics Institute, METU ________________

Date: ________________

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: MUSTAFA DURSUN

Signature :

ABSTRACT

ROAD DETECTION BY MEAN SHIFT SEGMENTATION AND STRUCTURAL ANALYSIS

Dursun, Mustafa
M.Sc., Department of Electrical and Electronics Engineering
Supervisor : Prof. Dr. Uğur Halıcı

June 2012, 125 pages

Road extraction from satellite or aerial images is a popular issue in remote sensing. Extracted road maps or networks can be used in various applications. Normally, maps for roads are available in geographic information systems (GIS), however these informations are not being produced automatically. Generally they are formed with the aid of human. Road extraction algorithms are trying to detect the roads from satellite or aerial images with the minimum interaction of human. Aim of this thesis is to analyze a previously defined algorithm about road extraction and to present alternatives and possible improvements to this algorithm. The baseline algorithm and proposed alternative algorithm and steps are based on mean-shift segmentation procedure. Proposed alternative methods are generally based on structural features of the image. Firstly, fundamental definitions of applied algorithms and methods are explained, mathematical definitions and visual examples are given for better understanding. Then, the chosen baseline algorithm and its alternatives are explained in detail. After the presentation of alternative methods, experimental results and inferences which are obtained during the implementation and analysis of mentioned algorithms and methods are presented.

Keywords: road extraction, satellite images, mean-shift segmentation, template-matching filter, structural analysis

ÖZ

KAYAN ORTALAMA BÖLÜTLEME VE YAPISAL ANALİZLERLE YOL TESPİTİ

Dursun, Mustafa
Yüksek Lisans, Elektrik-Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Uğur Halıcı

Haziran 2012, 125 sayfa

Uydu veya hava görüntülerinden yolların çıkarılması uzaktan algılamada gözde olan bir konudur. Çıkarılan yol haritaları veya ağları çeiştili uygulamalarda kullanılabilmektedir. Normalde yol haritaları coğrafi bilgi sistemlerinde (GIS) bulunmaktadır ancak bu bilgiler otomatik olarak üretilmemektedir. Genelde bu bilgi sistemleri insan yardımıyla oluşturulmuşlardır. Yol çıkarım algoritmaları, uydu veya hava görüntülerinde yolları minimum insan etkileşimiyle tespit etmeye çalışmaktadır. Bu tez çalışmasının amacı yol çıkarmayla ilgili daha önceden tanımlanmış bir algoritmayı incelemek, bu algoritmaya alternatifler ve muhtemel iyileştirmeler sunmaktır. Temel seçilen algoritma ve önerilen alternatif algoritma ve adımlar kayan-ortalama bölütleme prosedürü üzerine kuruludur. Önerilen alternatif yöntemeler genelde görüntüde bulunan yapısal özelliklere dayalıdır. İlk olarak uygulanan algoritma ve yöntemlerin temel tanımlamaları açıklanmış, daha iyi bir anlaşılırlık için matematiksel tanımlar ve görsel örnekler verilmiştir. Daha sonra, seçilen temel algoritma ve bu algoritmanın alternatifleri detaylı olarak anlatılmıştır. Alternatif yöntemlerin anlatılmasından sonra belirtilen algoritma ve yöntemlerin çalıştırılması ve incelenmesi ile elde edilen deneysel sonuçlar ve çıkarımlar sunulmuştur.

Anahtar Kelimeler: yol çıkarımı, uydu görüntüleri, kayan-ortalama bölütlemesi, şablon eşleme filtresi, yapısal analiz

To My Family...

ACKNOWLEDGMENTS

I am grateful to my supervisor Prof. Dr. Uğur Halıcı for her valuable guidance and opinions which were helped me to progress and finalize my thesis.

I am also grateful to the ASELSAN INC. company, to my colleagues and to my managers for their valuable support and for the comfort, opportunity and tolerance that they provided during my graduate education.

I thank to Prof. Dr. Yasemin Yardımcı, Mahir Öztürk, Ersin Karaman, Ekin Gedik, Umut Çınar and HAVELSAN INC. company for the exchange considerations they performed with me during my studies.

Finally, I thank to my family and friends on account of their support and trust.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Importance of road extraction from satellite or aerial images increased gradually in recent years. There are various application areas where road maps can be used in. For example, navigation systems that are started to be used commonly in cars, are using road maps to inform the driver. Road extraction is not only intended to be used in civil applications, road coordinates are also crucial information for martial and strategic purposes. Currently available road maps are generally taken from geographic information system(GIS) which is formed with human interaction. The road map database of GIS has to be updated if new roads are constructed or existing structure of roads are modified. By extracting road maps automatically, the work time for the update of the GIS database can be decreased. For these reasons, the topic, detection of roads in satellite images, is investigated in this thesis.

There are several algorithms in the literature. Generally classification is performed by using spectral and structural features that exist on the image. In order to be able to use structural classification, it is decided to work on an algorithm that includes segmentation. On the other hand, the algorithm has to be efficient and simple. The algorithm suggested by (Long and Zhao, 2005) was using mean-shift segmentation method and the performance of the algorithm was good for the test image that authors provided. Also the algorithm was efficient and simple. For this reasons, the algorithm suggested by (Long and Zhao, 2005) is chosen as baseline algorithm.

It has been observed that the performance of the baseline algorithm was not good enough for the data sets used in this thesis and it has been noticed that several improvements can be applied to the baseline algorithm. Hence, for the purpose of performance improvement, an alternative algorithm, pre-segmentation and post-segmentation methods are proposed.

1.2 Scope of Thesis

The scope of this thesis is to investigate road extraction from 3-band satellite/aerial images by using mean-shift segmentation procedure. A previous algorithm in which mean-shift segmentation is used, taken as the baseline algorithm and alternative pre-processing and post-processing steps are proposed to improve the overall performance of it.

1.3 Contribution of Thesis

In this thesis, some alternative pre-classification and post-classification methods are proposed in order to improve overall performance of the baseline algorithm in which mean-shift segmentation used. Two of the alternative methods are based on segment merging and their aim is to post-process the output of the mean-shift segmentation procedure in order to improve the structural features of the road segments. These methods are applied to the segmented image and then output of this steps is used in automatic classification step.

In addition to the segment-merging steps, an automatic classification method is proposed. In this method, firstly, structural features of segments are analyzed to determine seed road segments. Then spectral features of seed road segments are used as the training space of Gaussian Mixture Model which is constructed by using Expectation Maximization algorithm. At the end, the resultant clusters are analyzed and the maximum likelihood estimates of the parameters of some clusters are used for spectral modeling of road pixels.

Furthermore a rotation invariant multi-scale template matching filter is proposed to analyze the structural features of images. Roads in different widths can be selected by this filter using different target road width parameters and the rotation invariability is achieved by using rotated versions of the template filter.

Finally, by using experimental results that are obtained during the implementation and investigation of alternative steps of the baseline algorithm, a resulting road detection algorithm is proposed. Steps of the proposed algorithm are determined experimentally by selecting the best alternative among the others.

1.4 Outline of Thesis

In this thesis, a previously defined road extraction algorithm which is using mean-shift segmentation procedure, is chosen as the baseline algorithm. In chapter 2, background informations on the methods and fundamental operations that are implemented in this thesis are provided together with visual examples. In chapter 3, alternative methods are proposed in order to improve overall performance. In chapter 4, experimental results that are obtained during the implementation and analysis of the chosen baseline algorithm and proposed methods are given. In chapter 5, the study is concluded.

1.5 Literature Survey

There are a lot of methods, algorithms and softwares in literature which are aimed to detect roads from satellite or aerial images. Generally it is being tried to extract road masks or road networks. About 250 different resources on road extraction is reviewed in (Mena, 2003), in which the methods used in reviewed articles are explained briefly. More recently, Hauptfleisch(2010) made an itemization on road extraction articles and he listed and explained several methods and algorithms in detail. In both of these articles, the methods that are generally used in road extraction algorithms are presented without a sequential algorithm flow. In this section, the studies available in literature are examined according to input images, preprocessing, classification and post-processing methods. Firstly, input image types which are commonly used in road extraction literature are explained. Then general structure of road extraction algorithms and their sub-steps are given.

1.5.1 Input Images

In earlier times of detection of roads from images, resolution of the images was not as good as today's images' resolution. Also there were generally single band images rather than multi-spectral images. Today there are several image sources that are used in road detection literature. In his survey paper, Hauptfleisch(2010) categorized images by the sensor types that they are formed. He listed most popular sensor types and gave usage statistics of 62 research groups.(Figure 1.1)

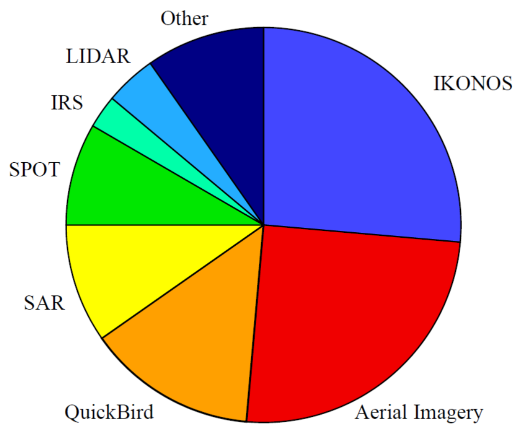


Figure 1.1: Sensor types used by research groups (Hauptfleisch, 2010)

Hauptfleisch(2010) mentioned that commercial satellite IKONOS was capable of taking 1m resolution panchromatic and 4m resolution 4-band spectral(R, G, B and NIR) images. By using the combination of panchromatic and multi-spectral images, 1m resolution pan-sharpened version of the images are obtained. It can be seen from the Figure 1.1 that IKONOS images are the most popular images used in road extraction literature.

Aerial images are the oldest images used in remote sensing and there are a lot of sensor types used in aerial imaging that cover different wavelengths. It is stated by Hauptfleisch(2010) that the best resolution used in remote sensing literature was 20cm and used in (Baltsavias et al., 2004).

SAR images are obtained differently than the other satellite images. SAR is the abbreviation of Synthetic Aperture Radar and images are obtained by using electromagnetic waves rather than optics.

QUICKBIRD is another commercial satellite and it can take 64cm resolution panchromatic and 2.4m resolution multi-spectral images. SPOT is not a single satellite, it is a satellite group in which five different satellites are used.(SPOT-1,2,3,4 and 5). The last satellite SPOT-5 is capable to take 2.5m-5m resolution panchromatic images and 10m resolution multi-spectral images.

LiDAR is used with aircrafts and it uses laser pulses in order to detect objects. LiDAR is generally used for the measurement of the height difference between the roads and surrounding objects.

1.5.2 Automation of Algorithms

Road extraction algorithms can be categorized to two types as automatic and semi-automatic algorithms. In both type of algorithms, the main steps of processing are same which are pre-processing step, classification step and post-processing step. Difference between automatic and semi-automatic algorithms come out whether human interaction or a priori information is needed or not.

An algorithm is called automatic if it is not externally guided and called semi-automatic if external informations used for processing. This external input can be used in any step of semi-automatic algorithms. For example, Chiang(2008) used a raster map rather than satellite images in order to detect road intersection points. Hence it can be said that they are using an already pre-processed image as input to their algorithm, so **in pre-processing step**.

On the other hand most of the semi-automatic algorithms are using external information **in classification step**. Generally a seed data set is taken as external information and then features of it are analyzed to classify the image. For example Gruen and Li(1995) used seed points that are marked by a user, in order to classify images. Also Song and Civco(2004) used external training data to train SVM in order to classify image.

Similar to the external data usage in pre-processing and classification steps, some of the algorithms need human interaction **in post-processing step**. For example, Zhao et al.(2002) defined an algorithm in which external informations that are taken from an operator are used to combine and post-process the classified binary image. In the same way, a road tracker based on profile matching is proposed and applied in (Baumgartner et al., 2002) and then with the aid of an operator, road networks are extracted.

1.5.3 Pre-processing Methods

In this section, pre-processing methods that are used in road extraction literature are explained. Pre-processing operations are used to increase performance of the classification step. Smoothing filters, edge detection, segmentation and color-space transformation operations are **mostly** used before the classification step of algorithms in road extraction literature. In addition to these operations, seed selection can be also defined as a pre-processing operation since seed data is used in classification step.

1.5.3.1 Smoothing Filters

Smoothing filters are generally used to increase performance of spectral classification. When there are spectral noise on an input image, smoothing filters are used to remove noise. For this purpose, Long and Zhao(2005) used an edge-preserving smoothing filter which is called multi-scale morphological cleaning and strengthening algorithm. In (Senthilnath et al., 2009), noises on the high resolution input image are filtered before the classification step. Firstly they applied ceiling operation to obtain a binary image and then they applied grouping and masking operations for the purpose of removal of noises and smoothing the image.

1.5.3.2 Edge Detection

There are several edge detectors defined in literature. Generally edges are detected by investigating local spectral variations of the image. Basically, when there is a big spectral difference between adjacent pixels it is assumed that there is an edge. One of the most popular edge detector among the others is Canny Detector which is defined by Canny(1986). According to the Hauptfleisch(2010), Sobel(Duda and Hart, 1973), (Marr and Hildreth, 1980), (Nalwa and Binford, 1986), (Sarkar and Boyer, 1993) and Laplace(Bovik, 2005) are other popular edge detectors used in road extraction literature.

In (Ruskone and Airault, 1997), edge detection is applied to the input image and then parallel edges are detected. The pixels between parallel edges are considered as road seed. Same method is also used in (Mei et al., 2003) and (Chen et al., 2004).

1.5.3.3 Segmentation

In segmentation, image is divided into sub-pieces which are containing spectrally similar pixels. Several segmentation methods are used in road extraction literature. For example, in (Long and Zhao, 2005), mean-shift segmentation method is used before the classification step. Hinz and Baumgartner(2003) used textural segmentation in their road extraction algorithm for several purposes.

Iterative Self-Organizing Data Analysis Technique(ISODATA) which is described in (Hall and Ball, 1965) is also used in several algorithms. In this algorithm, a cluster whose spectral variation is big can be divided into sub-clusters and clusters which have similar spectral properties can be merged in to a single cluster, dynamically. For example, Zhang(2004) used ISODATA for the purpose of segmentation and classification.

Xu et al.(2009) extracted initial outline of roads by using watershed dual threshold algorithm and then used multi-weighted method to extract exact edges of the roads. Finally they used morphological operations and shape index parameters in order to remove noises and non-road regions.

Yuan et al.(2009) defined an algorithm which is based on segmentation of input image by using Locally Excitatory Globally Inhibitory Oscillator Networks(LEGION). Rather than assigning weights to all of the image pixels they assigned weights to the locally attractive pixels. After segmentation, they investigate medial-axis of segments and select narrow ones. Then again by using LEGION algorithm in the sense of medial-axis alignments, well-aligned segments are connected and mapped as road.

1.5.3.4 Color Space Transformation

Mostly, gray-level images or RGB color space is used in road extraction literature. However alternative color spaces are also being used by transforming RGB color-space. In (Zhang, 2004), instead of RGB space, a different space is defined and used. One of the bands is called as greenness and defined as (G-R)/(G+R). The other band is chosen as the saturation value of the HSI space.

Christophe and Inglada(2007) firstly translated input color space to a spectral angle domain. Number of image bands used with this spectral angle transformation can be any number. All of the band informations are gathering into spectral angle domain value. Then pixels which are darker in the spectral angle domain are labeled as roads.

1.5.3.5 Seed Point Selection

As it is mentioned automation of an algorithm is related to its need to human interaction. In most of the semi-automatic algorithms, human interaction is used before the classification step.

Features that exist in input space are compared with the expected feature values and classification is realized according to the result of this comparison. In some studies, expected feature values are extracted from seeds. Seeds are the regions that are representing the characteristics of roads. While some of the algorithms need human aid in seed selection, some algorithms automatically select seeds. In first type, a seed region on the image is given to the algorithm and the algorithm extracts the features by using the region. Then in classification step, these extracted features are used. Some of the algorithms do not need human interaction for every image. Road regions and non-road regions are used as training set and the algorithm uses features of these regions in the classification.

In automatic algorithms, seed points are determined automatically. There are several methods that are used in seed selection but most popular methods among the others are edge detection and segmentation.

1.5.4 Classification

Classification is the most important step of the road detection algorithms. There are four different types of classification which are spectral, structural, textural and contextual. Structural features are generally extracted by applying edge detection, Hough transform and segmentation. Roads are generally long and narrow straight lines. In edge detection, the most popular usage is to find parallel edges in edge map. These edges are assumed to be road edges and spectral information of the region between that edges is used to spectrally classify the image. Segmentation is also another powerful method to extract structural features from the image. There are several structural features used to define roads in literature. By investigating the structural features of the roads, the segments which are similar in shape to the roads are classified as roads.

There might be 2 or more classes in classification step. The simplest and the most popular classification is two-class classification where classes named as road and non-road. At the output of classification a binary image is obtained. On the other hand, number of classes can be more than two. For example, in (Mohammedzadeh et al., 2008) spectral classification was used and five membership functions were defined which were good-road, up-probable-road, down-probable-road, up-bad-road and down-probable-road. By applying these membership functions to the RGB space, Mohammedzadeh et al.(2008) obtained 125 different classes. There are several classifiers used in literature. Artificial neural networks, fuzzy-classifiers and statistical classifiers are popularly used classifiers.

1.5.4.1 Spectral Classification

In spectral classification, pixels are classified according to their spectral features. In most of the satellite or aerial images there are some non-road regions that spectrally similar to the roads so these regions are also classified as road when only spectral features are used. For this reason spectral classification is generally used as an initial step in classification. There are a lot of spectral classifiers used in literature. Hauptfleisch(2010) stated that classical statistics, artificial neural networks and fuzzy classifiers are the most popular methods.

In some of the studies that are based on classical statistics, probability theory is used in order to obtain information. Oddo et al.(2000) used a maximum likelihood classifier with the assumption that roads have uniform spectral features. They defined mean and variance as the parameters and calculated the distance of the patterns to the probability curve. Then using a threshold they classified the pixels as road or non-road. Chen et al.(2004) used a histogram Bayesian classification approach. Firstly, they choose a set of road pixels and a set of non-road pixels. Then by using the spectral distribution of these sets, they made a decision on a pixel if it is belong to a road or not.

Artificial neural networks(ANN) are also used in spectral classification step of the algorithms. In (Fiset et al., 1998), it is aimed to update a road map database. For this purpose they try to match the existing roads on the map with the satellite image precisely. They realized this matching by using a multi-layer neural network which is trained to detect roads on the satellite image. In (Li et al., 2003), a neural network based spectral classifier is used with pixel-based knowledge post-processing method, for the purpose of extraction of road and water informations. In (Mokhtarzade and Zoej, 2007), it is examined the possible usage of ANNs in road detection algorithms. They compared several neural network structure and they concluded that instead of using spectral features of a single pixel, using spectral features of a pixel with its neighbors as input to the neurons can increase the performance.

In (Shackelford and Davis, 2003), a hierarchical pixel-based fuzzy classification method is used to spectrally classify the image. Outputs of this classification step are road, building, water, grass, tree, bare, soil and shadow. After this classification step, they applied object-based analysis for further processing. In (Mohammadzadeh et al., 2004), a fuzzy process developed in which a matrix of membership degrees is obtained for each pixel and then a rule is used to form fuzzy outputs. Finally, features are extracted by using a defuzzification step.

1.5.4.2 Structural Classification

In structural classification, generally input image is firstly pre-processed for the purpose of extraction of structural features. Then these structural features of the objects in the image are used to classify the objects. In (Jin and Davis, 2003), firstly, segmentation is applied to the image and then by using structural features, road centerlines are extracted. On the other hand, a multi-scale curvilinear structure detector is used as a second structural classifier. Then outputs of these blocks are combined using optimum path search algorithm. In (Hu and Tao, 2005), firstly, in order to identify road candidates, a binary template matching is used on perpendicular profiles along the road direction. Then by using the results of this analysis, the width, centerline and lateral sides of the roads are extracted. In (Zhang and Couloigner, 2006), firstly, k-means clustering method is used to convert input image to a segmented image. Then several shape descriptors are defined and used in classification to truly and precisely identify roads. Doucette et al.(2004) detected centerline of parallel edges and then using The Self-

Organized Mapping algorithm extracted a low-level road map. Also Bacher and Mayer(2005) used the regions between the parallel edges where spectral distribution is uniform on gray values. These regions are added to the training set.

1.5.4.3 Textural Classification

In textural classification, textural features of the image are used. Different regions in the image have different textural features. Definitions of 14 textural features are given by Haralick(1979). In (Ohanian and Dubes, 1992) it is stated that only three of Haralick features are enough for the purpose of textural classification (cited by Hauptfleisch(2010)). These features are contrast, entropy and angular second-moment. In (Mao and Jain, 1992), a multi-resolution simultaneous autoregressive (MR-SAR) model used for the purpose of textural classification and segmentation. It is stated by Hauptfleisch(2010) that this method is also used in Zhang and Couloigner(2006)and (Hui et al., 2006).

1.5.4.4 Contextual classification

Contextual classification is also a method used in classification step in literature. Two different contextual feature classes which are global features and local features are defined by Hinz et al.(1999). According to the features of region of interest, different road detection routines are used. This is determined with global features such as spectral and structural features. There might be some objects around the roads whose shape or position is an indicator of a road. Road signs, cars or tree lines are some examples to the local contextual information.

Baumgartner et al.(1997) used five different local contextual objects which are parallel features, shadows, vehicles, rural driveways and driveways up to buildings. Also they used three global contextual objects which are forested region, urban region or rural region. After the decision of global contextual feature, they used local contextual features for the purpose of classification.

In (Yang and Wang, 2007), four global contextual information defined which are rural, urban, montane and hybrids of suburban and rural regions. By using an edge density histogram model, sub-parts of the image are classified.

1.5.5 Post-Processing Methods

In post-processing operations features of input image are not generally used. Pixels or regions are determined as road or non-road without looking at their features. This determination process is independent from the features of input image, it is based on the binary final image. For this reason, these operations are called post-processing operations rather than classification. For example, assuming that there is a tunnel on a roadway. Spectral, textural or structural features at the location of tunnel will not be similar to the features of roads so the part of roadway with the tunnel will be classified as non-road. However in post-processing operations, by applying edge linking method, this discontinuity can be removed.

Generally, there are some regions or pixels on the classified image which are mis-classified by the classifier, namely, false positives and false negatives. These regions or pixels which were not classified truly by the classifiers, can be corrected by using post-processing operations. In addition to this, post-processing operations are also used for the purpose of obtaining more good and meaningful final images.

1.5.5.1 Morphological Operations

On the other hand, there exists a lot of cavities, zig-zaged edges and defects on the binary classified image. These kind of factors can be removed by applying morphological operations on the binary classified image. For example, in (Long and Zhao, 2005), most of the cavities are removed by applying convex hull algorithm and morphological opening. Also zig-zaged edges are smoothed by using contour tracing algorithm. In (Xu et al., 2009) after the classification step, by using morphological operations, holes and non-road regions are removed from the classified image.

Shi and Zhu(2002) converted RGB image to gray level then used a gray-level constant threshold filter to obtain a binary image. Then in next step, they detected lines from this binary image. As a result of pixel-by-pixel classification of the gray-level threshold filter, a lot of lines appears on the image. By using binary opening, closing, thinning and removing splinter lines, they obtain the road network.

1.5.5.2 Edge Linking

In (Zhao et al., 2002), after the spectral classification step, discontinuities on the road map are removed by using edge linking method. Although the performance of the classifier on these regions was poor, effect of this problem was removed by using post-processing operations on the binary classified image.

CHAPTER 2

BACKGROUND

In this chapter, background information about the methods used in this thesis are provided and examples are given for a better understanding. Color spaces used in this thesis are defined briefly in section 2.1. Then commonly used morphological operations for the purpose of post-processing such as opening and closing are described in section 2.2. In section 2.3, two different enhancement filters are described in which the input image is smoothed but the edges of the image are preserved. These filters are bilateral filtering and multi-scale gray-level morphological cleaning and strengthening(MMCSA). Then mean-shift segmentation procedure is explained in section 2.4. In section 2.5 Hough transform is explained. Finally, in section 2.6 edge linking method is explained.

2.1 Color Spaces

It is stated in (Singlia and Hemacllandran, 2011) that a color space is a model for representing color in terms of intensity values. There are a lot of color space representations. In this section, color spaces that are used in this thesis, which are RGB, XYZ, L*u*v* and L*a*b, are explained. Color spaces other than RGB space, are derived by applying linear or non-linear transformations to RGB space as stated in (Cheng et al., 2001).

2.1.1 RGB Color Space

Red, green, and blue are three primary colors that used in RGB representation of images. A digital image is two dimensional in spatial domain and every pixel of it has an intensity value. In RGB images, every pixel has three intensity values which are corresponding to the intensity of red, green and blue color in that pixel. According to Singlia and Hemacllandran(2011), RGB color space is the most popular color representation and is used in televisions and cameras but Comaniciu and Meer(2002) stated that it is non-linear and is not suitable for segmentation procedures.

2.1.2 CIE Color Spaces: XYZ, L*u*v and L*a*b*

CIE (Commission International de l'Eclairage) color system has three parameters which are X, Y and Z and it was developed to represent perceptual uniformity. XYZ values are linearly obtained from RGB values as given below.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.1}$$

Comaniciu and Meer(2002) mentions that CIE L*u*v* and CIE L*a*b spaces are obtained by applying non-linear transformation to XYZ. L^* is same in both while last two terms are different through chromatic coordinates. Definition of these spaces are given below.

$$\begin{aligned} L^* &= 116 f\left(\frac{Y}{Y_n}\right) - 16 \\ u^* &= 13L^*(u' - u_n') \\ v^* &= 13L^*(v' - v_n') \\ a^* &= 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* &= 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \end{aligned} \tag{2.2}$$

where u_n', v_n', X_n, Y_n and Z_n are equal to the value of reference white point where meaning of the subscript n is normalized. $f(t)$, u' and v' used in equation 2.2 are given below.

$$\begin{aligned} f(t) &= \begin{cases} t^{\frac{1}{3}}, & t \geq \frac{1}{113} \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & t < \frac{1}{113} \end{cases} \\ u' &= \frac{4X}{X+15Y+3Z} \\ v' &= \frac{9Y}{X+15Y+3Z} \end{aligned} \tag{2.3}$$

2.2 Morphological Operations

Morphological operations are crucial techniques used in image processing. Morphological operators are used to diminish effects of noisy features on binary and gray-level images. For example, they are used to fill small-black holes in a big white region, or vice versa. Also they can be used to smooth the zig-zags at the edges of binary images. There are several studies that focused on morphological operations. In this section, four fundamental morphological operators are explained. These are erosion, dilation, opening and closing. Generally a structuring element(SE) is used to apply these morphological operations. Shape and size of the structuring element effect the output image. One of the popular binary structuring element is a disk which is showed in Figure 2.1.

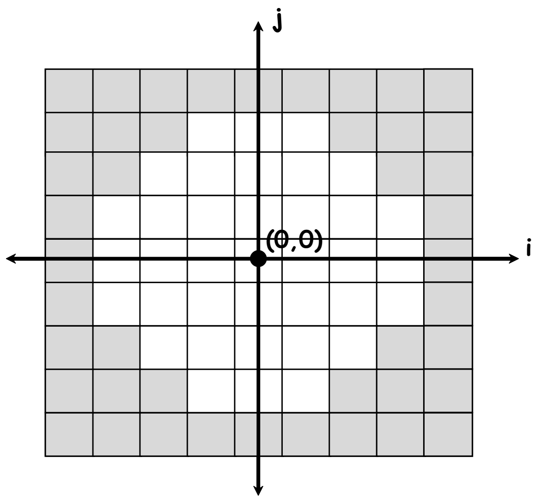


Figure 2.1: A disk-shaped binary structuring element with radius 3

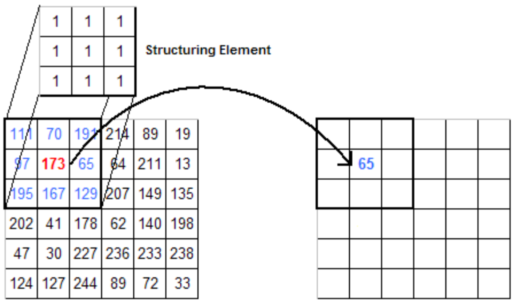
2.2.1 Erosion

Erosion which is represented with $\ominus$ term, is defined as given in equation 2.4.

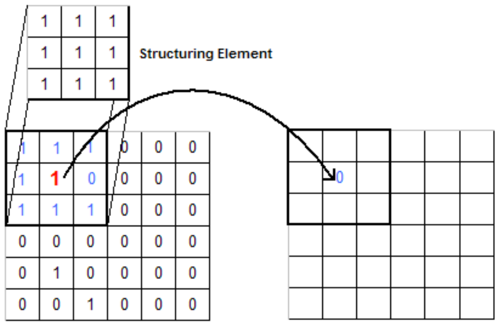
$$(x \ominus SE)(i, j) = \min_{(m,n)\in SE}\{x(i + m, j + n)\} \tag{2.4}$$

Erosion can be applied to both gray-level and binary images. In gray-level case, new value of a pixel is chosen as the minimum pixel value within the neighborhood of the pixel of interest where neighborhood is chosen according to the shape and size of the structuring element(SE).

Binary case can be thought as a sub-space of gray-level case where there are only two gray-levels which are one(1) and zero(0). In binary case, if there is a zero within the neighborhood then value of the pixel will be equal to zero after erosion. In Figure 2.2 both binary and gray-level erosion operation are illustrated where a square-shaped structuring element is used.



(a) Gray-level



(b) Binary

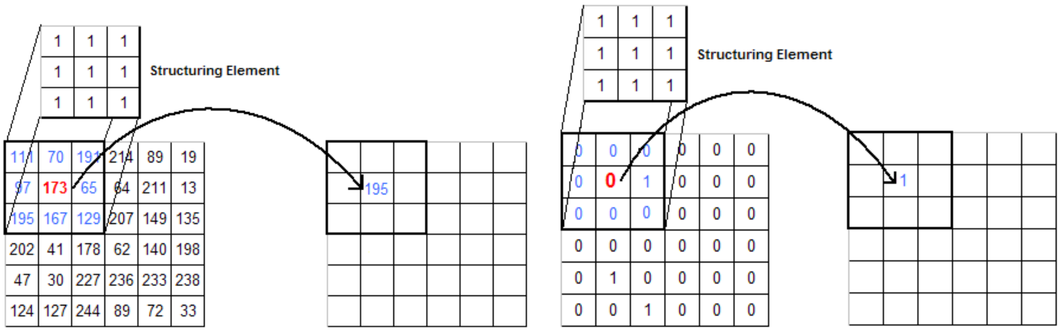
Figure 2.2: Erosion Operation

2.2.2 Dilation

Dilation which is represented with $\oplus$ term, is defined as given in equation 2.5.

$$(x \oplus SE)(i, j) = \max_{(m,n)\in SE} \{x(i - m, j - n)\} \tag{2.5}$$

In gray-level case, new value of a pixel is chosen as the maximum pixel value within the neighborhood of the pixel of interest in which neighborhood is chosen according to the shape and size of the structuring element(SE). Binary case can be thought as a sub-space of gray-level case in which there are only ones and zeros. In binary case, if there is a 1 within the neighborhood of pixel of interest, then new value of the pixel will be equal to 1. In Figure 2.3 both gray-level and binary case erosion operations are illustrated where a square-shaped structuring element is used.



(a) Gray-level (b) Binary

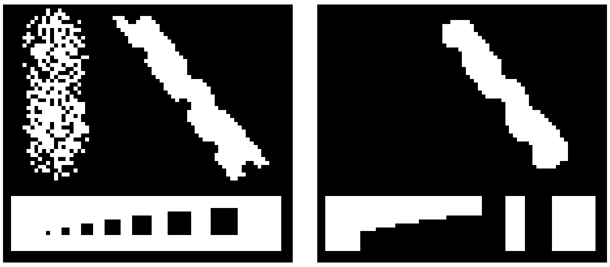
Figure 2.3: Dilation Operation

2.2.3 Opening

Opening is defined in two steps. Firstly erosion applied to the input image and then dilation applied to the eroded image. Mathematically opening represented with the $\circ$ symbol and is defined as given below.

$$(x \circ SE)(i, j) = ((x \ominus SE) \oplus SE)(i, j) \tag{2.6}$$

Erosion operation removes noisy features/pixels whose size is smaller than structuring element from the input image. In binary case erosion removes 1's. In gray-level case erosion removes brighter pixels. In erosion, features that are bigger than structuring elements size do not disappear as small features do, but they shrink. So when dilation applied to the eroded image, the features that are narrowed by erosion operation are expanded again. Hence a feature is removed if its size is smaller than the structuring element. In Figure 2.4 below, a binary input image and corresponding output image are given for the opening operation with disk-shaped structuring element whose radius is equal to 3. As it can be seen from the Figure 2.4, the upper right object is removed as a result of opening operation, because there were noise in that object.



(a) Input Image (b) Output Image

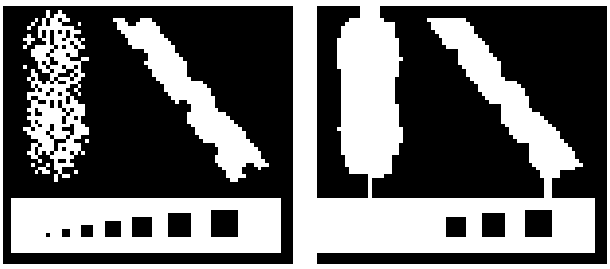
Figure 2.4: Opening Operation

2.2.4 Closing

Closing is defined in two steps. Firstly dilation applied to the input image and then erosion applied to the dilated image. Mathematically closing represented with the $\bullet$ symbol and is defined as given below.

$$(x \bullet SE)(i, j) = ((x \oplus SE) \ominus SE)(i, j) \tag{2.7}$$

Dilation operation causes to removal of black holes that are smaller than the structuring element. Black features that are bigger than structuring element shrink but do not disappear. Second step of closing is erosion. When erosion applied, big black/darker features expands but removed small features do not exist anymore. As a result, black holes or darker pixels that are smaller than structuring element are removed from the image. An example image and its corresponding output is given in Figure 2.5 below. An example image and its corresponding output is given in Figure 2.5 below.



(a) Input Image (b) Output Image

Figure 2.5: Closing Operation

As it can be seen from the upper left object of the image that black holes in the object are filled when closing operation is applied. And also it can be seen from the upper right object of the image that black cavities which are smaller than structuring element are also removed, therefore some zig-zags of input image are removed.

2.3 Edge Preserving Smoothing Filters

Spectral distribution of pixels of a road can vary locally due to the several factors such as different ages, road materials, angle of incidence of sun-light etc. As a result of these effects, a road is being divided into sub-parts, so classification performance becomes worse. In order to overcome this problem, smoothing filters can be used to decrease spectral variability that exist on a road. However as implied by Long and Zhao(2005), common smoothing filters such as Gauss Filter or Wavelet Filter generally smooths every color feature of an input image. Because of smoothing operation, edges or sharp transitions on the image are also being smoothed or blurred. Nevertheless, there are also edge-preserving filters which are developed in the literature. Two of these filters that are used in this thesis are explained in the following sub-sections.

2.3.1 Bilateral Filtering

Bilateral filter is proposed by Tomasi and Manduchi(1998) and it preserves edges while removing small-noisy color variations in the image. This is achieved by determining coefficients of an NxN filter by two features which are spatial closeness and range(spectral) closeness. Traditional smoothing filters only use spatial closeness as the decreasing factor of coefficients while going through outside from the center of the filter. As a result edges which are distinguished by their spectral differences are being softened without any preserving.

Bilateral filtering can be applied to both gray-level and color images. In color images, firstly RGB to L*a*b* transformation is used and then bilateral filter is applied to the transformed image. The bilateral filter which is given as an example of bilateral filtering in (Tomasi and Manduchi, 1998), is used in this thesis study. In this filter, two 2D Gaussian-shaped filters are used together in order to filter the input. The form of the Gaussian-shaped functions is given below.

$$c(i, j) = e^{\frac{-1}{2}(\frac{d(i,j)}{\sigma^2})} \tag{2.8}$$

where $-N < i < N$ and $-N < j < N$. σ is the bandwidth for the related feature space and $d(i, j)$ is the distance of the interested point from the center point of the filter whose value will be determined according to the filter output. One of the Gaussian-shaped filters is applied in spatial domain. Namely, the distance is considered as the distance in space. The other Gaussian-shaped filter is applied in the spectral domain. Namely, the distance is considered as the difference of the gray-level values of the compared pixels. An NxN filter will have N^2 coefficients. Every coefficient of the bilateral filter has two multipliers which are in the form as given below.

$$c_{space}(i, j) = e^{\frac{-1}{2}(\frac{d_S(i,j)}{\sigma_S^2})} \tag{2.9}$$

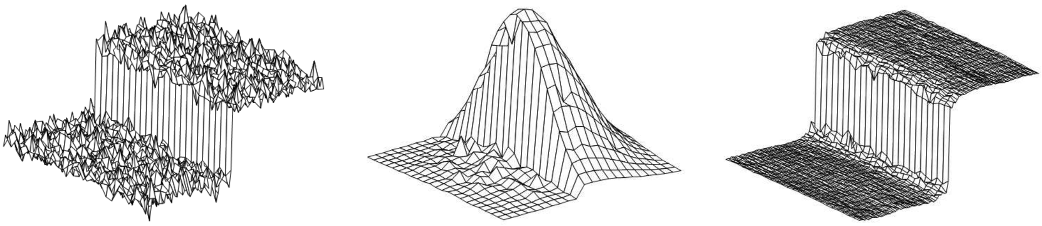
$$c_{spect}(i, j) = e^{\frac{-1}{2}(\frac{d_C(i,j)}{\sigma_C^2})} \tag{2.10}$$

$$f(i, j) = c_{space}(i, j) \times c_{spect}(i, j) \tag{2.11}$$

If $x(i, j)$ is sub-image part related to the point of interest, new value(y) of the filtered pixel will be in the form as given below.

$$y = \frac{1}{N^2} \sum_i \sum_j f(i, j) \times x(i, j) \tag{2.12}$$

It can be seen from the equation 2.12 that the output value of the point-of-interest is equal to weighted average of to the pixels that are within the neighborhood of it. Weights $f(i, j)$ are determined by both spatial and spectral distances. If one of the distance values of a neighbor point is big, its weight coefficient will be small. As a result, new value of the pixel-of-interest will be less effected by this pixel. Since the spectral distance at the edges is big, neighbor pixel that fall in to the other side of an edge will be less effective on the new value of the point-of-interest when compared with the neighbor pixels that are in the same side with the point-of-interest. In Figure 2.6, result of the bilateral filtering can be seen when a noisy step-function(a) is used as the input to the bilateral filter. Filter coefficients are shown at (b) and output of the filter is shown at (c).



(a) Input Image (b) Bilateral Filter (c) Output Image

Figure 2.6: Bilateral Filtering (Tomasi and Manduchi, 1998)

As it can be seen from the Figure 2.6-(b) that at the edge of the input image, filter coefficients are being decreased considerably because, spectral distance of the neighbor pixels is too big. Hence coefficient of the filter at those pixels become too small. By this way, edges are being preserved.

2.3.2 Multi-scale Gray-Level Morphological Cleaning and Strengthening

Multi-scale gray-level morphological cleaning and strengthening algorithm(MMCSA) is based on gray-level opening and closing operations, hence it is more complex than bilateral filtering.

Definition of gray-level opening and closing operations are given by Maragos(1989) and they were explained in section 2.2. Opening and closing operations include both erosion and dilation operations. Multi-scale opening and closing operations are described in (Chanda and Mukhopadhyay, 2002) as given below:

$$(x \circ nSE)(i, j) = ((x \ominus nSE) \oplus nSE)(i, j) \tag{2.13}$$

$$(x \bullet nSE)(i, j) = ((x \oplus nSE) \ominus nSE)(i, j) \tag{2.14}$$

Here, n is the scaling factor and nSE is obtained by dilating SE recursively $n-1$ times with itself. Mathematically nSE can be written as below.

$$nSE = \underbrace{SE \oplus SE \oplus ... \oplus SE}_{n-1\, times}$$

In MMCSA, multi-scale opened and closed images are being used several times. At each scale for $k = 2, 3..n$ ”tophat” and ”bothat” images are obtained. At n^{th} scale SE is dilated with itself n times so its radius is increased n times with its initial radius. Value of n can be determined by investigating the maximum size of the noise that will be removed. For example, in the process of road smoothing, maximum noise size will be equal to the maximum road width. ”Tophat” and ”bothat” images are defined as residual image in (Long and Zhao, 2005) and these images contain both noise and signal which are smaller than structuring element at scale k, kSE. Tophat(T_k) and bothat(B_k) images are defined as given below:

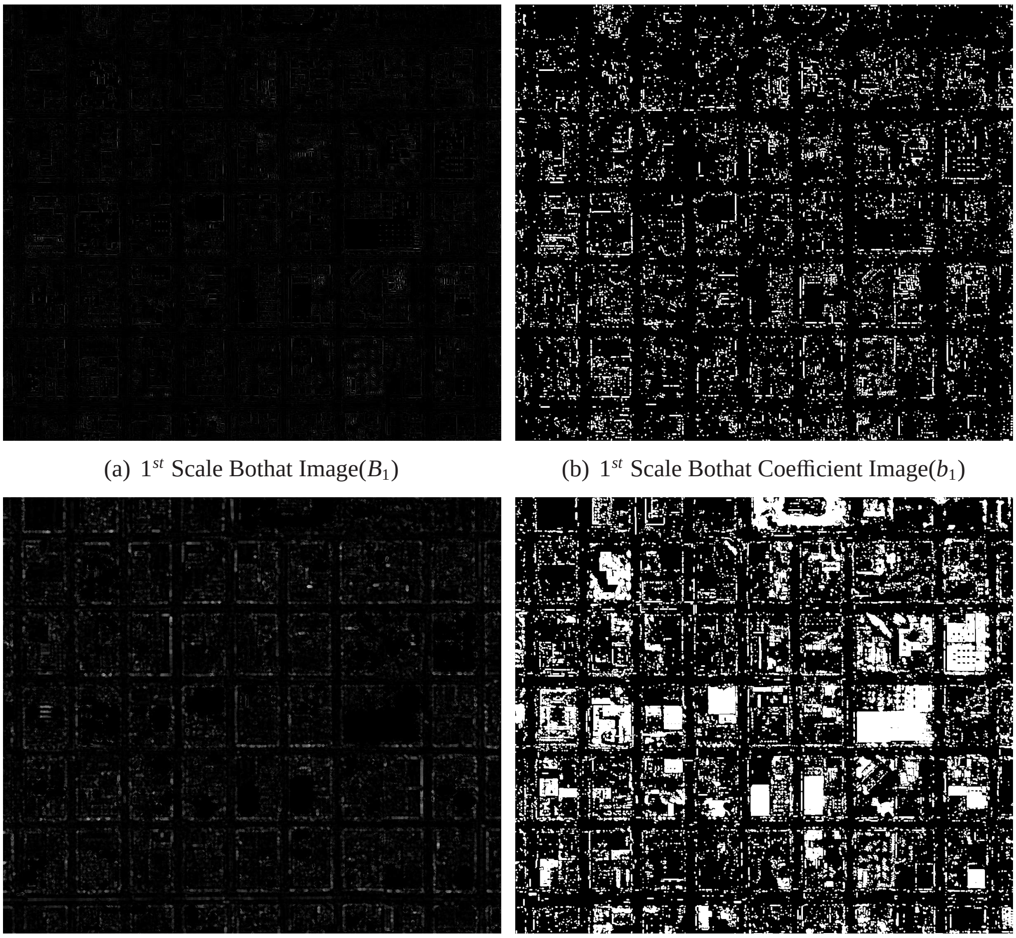
$$T_k = C_k - C_{k-1} \tag{2.15}$$

$$B_k = O_{k-1} - O_k \tag{2.16}$$

Here C_k is k-scale closed image and O_k is k-scale opened image, $C_0 = I$ and $O_0 = I$ where I is the input image. Tophat images corresponds to bright pixels and bothat images corresponds to dark pixels of input image. Output image Y of the algorithm is being computed as given below:

$$Y = \frac{1}{2}(C_n + O_n) + \sum_{k=1}^{n} t_k \times T_k - \sum_{k=1}^{n} b_k \times B_k \tag{2.17}$$

Here, b_k and t_k are coefficient images whose size are equal to the input image’s size where $k = 1, 2, ..., n$ and their value vary from pixel to pixel. These coefficient images are calculated from tophat and bothat images. Calculation of these coefficients is different in MMCSA-LZ and MMCSA-C. Multiplication in the equation 2.17 is performed as element by element. At every scale of the MMCSA, there will be tophat and bothat images. These images will have both noise and signal as mentioned above. All of the tophat and bothat images are used at the end of the algorithm as given in equation 2.17. Images in the figures below are intermediate images of MMCSA-LZ method.



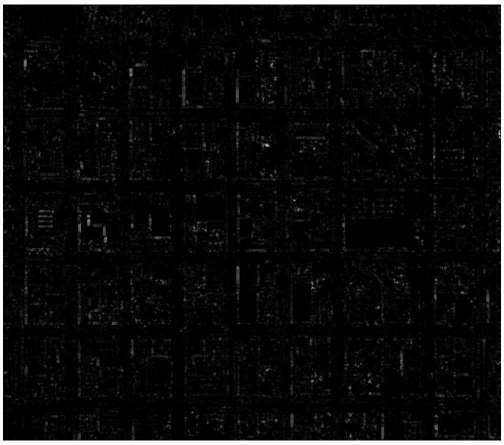
(a) 1^{st} Scale Bothat Image(B_1)

(b) 1^{st} Scale Bothat Coefficient Image(b_1)

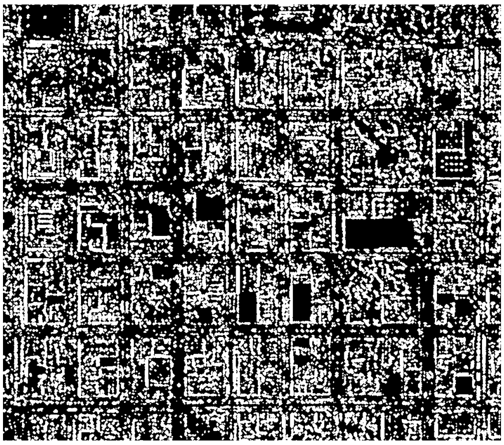
(c) 1^{st} Scale Tophat Image(T_1)

(d) 1^{st} Scale Tophat Coefficient Image(t_1)

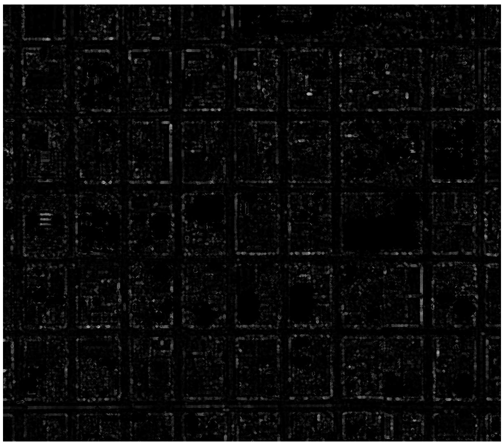
Figure 2.7: MMCSA 1^{st} Scale Images



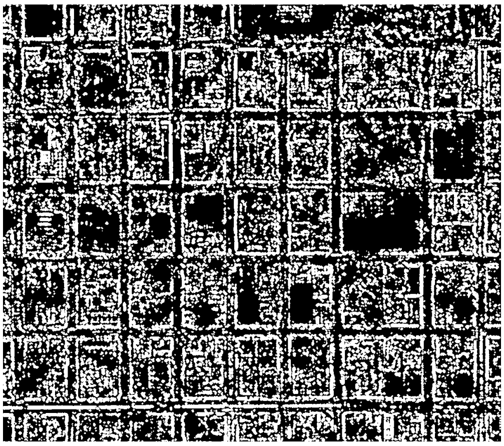
(a) 4^{th} Scale Bothat Image(B_4)



(b) 4^{th} Scale Bothat Coefficient Image(b_4)



(c) 4^{th} Scale Tophat Image(T_4)



(d) 4^{th} Scale Tophat Coefficient Image(t_4)

(e) Original Image(I) (Long and Zhao, 2005)

(f) Filtered Image(Y)

Figure 2.8: MMCSA 4^{th} Scale Images

2.3.2.1 MMCSA-C

Chanda and Mukhopadhyay(2002) states that according to their experimental results the best coefficient images are the ones in which weight of the coefficients decreases as the scale factor decreases. And they describe coefficient images as the exponential function of 2. Namely, $t_k = \frac{1}{2}^{n-k}$ and $b_k = \frac{1}{2}^{n-k}$ for $k = 1, 2, ..., n$. As a result their coefficient images are formed by the exponential terms of 2. At the end, residual image of last scale will be the most effective image on the output image and residual image of first scale will be the less effective image on the output image. The abbreviation MMCSA-C is used throughout this thesis, for the algorithm that is proposed by Chanda and Mukhopadhyay(2002).

2.3.2.2 MMCSA-LZ

Long and Zhao(2005) implements an automatic threshold selection method to filter tophat/bothat image. Briefly, they compute probability distribution of residual image and determine a threshold value, to filter tophat/bothat image. Threshold value is calculated from the square root of the second moment of probability distribution function. After the calculation of threshold value tophat/bothat image is filtered. Value of every pixel of tophat/bothat image is compared with the threshold value. The pixels that are smaller than the threshold are considered as noise and the coefficient related to that pixel in the coefficient image t_k/b_k are set to zero. If value of a pixel is greater than the threshold than related coefficient image pixel is set to one. Hence the coefficient images b_k and t_k used in the equation 2.17 are just formed by zeros and ones for MMCSA-LZ method. The abbreviation MMCSA-LZ is used throughout this thesis, for the algorithm that is proposed by Long and Zhao(2005).

After giving the fundamental principles of MMCSA now we can explain how it is smoothing noisy features while preserving edges. Normally opening and closing operations diminishes the sharp features in the images, i.e. the edges. However when they are averaged, output image will preserve its edge features. On the other hand, filtering tophat and bothat images will filter noise elements which are smaller than structuring element nSE. Since residual images are added to the final image by decreased effect, smoothing will be done for noisy features.

2.4 Mean-shift Segmentation

Theoretical background of mean-shift segmentation procedure is constructed by Fukunaga and Hostetler(1975). Then Cheng(1995) has analyzed and generalized mean-shift procedure. Mean-shift segmentation which is used in this thesis and explained in this section, is the procedure which is described by Comaniciu and Meer(2002). By using the theoretical results extracted in (Fukunaga and Hostetler, 1975), they published several articles about mean-shift procedure and related applications. (Comaniciu and Meer, 1997), (Comaniciu and Meer, 1999) and (Comaniciu and Meer, 2001) are some of their study about mean-shift procedure.

Mean-shift segmentation that is explained by Comaniciu and Meer(2002) is realized in four steps. First of all, image is filtered by using mean-shift procedure. As a result of filtering operation, convergence points in a d-dimensional space which are called modes are extracted. Then the mode points which are closer than spatial bandwidth h_s in spatial domain and closer than range bandwidth h_r in range domain are concatenated. Namely, their basin of attraction points are concatenated. Then every cluster is assigned with a label. As a last step, the regions that are smaller than M pixels are eliminated.

2.4.1 Mean-shift Filtering Procedure

First step is mean-shift filtering procedure, which is a non-parametric density gradient estimation method using kernels. Before the definition of the procedure, firstly it is suitable to define feature space. Dimensions of feature space can be any number, it is not restricted by mean-shift procedure. It is stated by Comaniciu and Meer(2002) that in segmentation based on mean-shift procedure, feature space is constructed by adding spatial coordinates of the pixels to the range(color) space of image and this is called joint spatial-range domain representation. There are two dimensions for an image in spatial domain and p dimensions in color space. In gray level case $p = 1$, while in RGB or L*u*v* case $p = 3$. In order to obtain a good segmentation, range differences on color space should correspond to the Euclidean distances in color space which represent the pixels. Hence L*u*v* space representation of image is being used as color space in feature space. This feature space is assumed as probability density function in mean-shift filtering procedure.

Fukunaga and Hostetler(1975) investigated a non-parametric density gradient estimation method by using a generalized kernel approach. They stated that the value of a probability density function or its gradient at a point, can be estimated using **sample observations** which are taken within a small region that surrounds the point-of-interest. According to them gradient density estimate at a point is essentially a weighted measure of the **observations** about the point-of-interest. The difference between each surrounding point and the point-of-interest is calculated and multiplied by a weighting factor. Then the sample mean of these weighted shifts is taken as the gradient density estimate. Mentioned procedure is called as mean-shift procedure in (Cheng, 1995) and (Comaniciu and Meer, 2002).

When there are n data points in d-dimensional space R^d, multivariate kernel density estimator at $\mathbf{x}$ point will be as below:

$$\widehat{f}(\mathbf{x}) = \tfrac{1}{n}\sum_{i=1}^{n} |\mathbf{H}|^{-1/2} K\left(\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right) \tag{2.18}$$

where $\mathbf{H}$ is a $d \times d$ dimensional symmetric positive definite bandwidth matrix. For the purpose of simplicity, the bandwidth matrix $\mathbf{H}$ is chosen as proportional to the identity matrix ($\mathbf{H} = h^2\mathbf{I}$). As a result, the kernel density estimator can be rewritten as below:

$$\widehat{f}(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) \tag{2.19}$$

$K(\mathbf{x})$ is the d-variate kernel that is satisfying following conditions.

$$\begin{aligned}
&\int_{R_d} K(\mathbf{x})d\mathbf{x} = 1\\
&\int_{R_d} \mathbf{x}K(\mathbf{x})d\mathbf{x} = 0\\
&\int_{R_d} \mathbf{x}\mathbf{x}^T K(\mathbf{x})d\mathbf{x} = c_K\mathbf{I}\\
&\lim_{\|\mathbf{x}\|\to\infty} \|\mathbf{x}\|^d K(\mathbf{x}) = 0
\end{aligned} \tag{2.20}$$

Here c_k is constant. K($\mathbf{x}$) is in the form as given below:

$$K(\mathbf{x}) = c_{k,d}k\left(\|\mathbf{x}\|^2\right) \tag{2.21}$$

and k(x) is the profile of the kernel and $c_{k,d}$ is the normalization constant. Using the profile notation, kernel density estimator can be rewritten as given below:

$$\widehat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \tag{2.22}$$

and the gradient of $\widehat{f}(\mathbf{x})$ is $\nabla\widehat{f}(\mathbf{x})$

$$\widehat{\nabla} f_{h,K}(\mathbf{x}) = \nabla\widehat{f}_{h,K}(\mathbf{x}) = \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(\mathbf{x}-\mathbf{x}_i)\,k'\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \tag{2.23}$$

If we define $g(x) = -k'(x)$ for all $x \in [0, \infty]$ except for a finite set of points then kernel $G(x)$ that uses $g(x)$ profile can be found as below.

$$G(x) = c_{g,d}\, g\left(\|\mathbf{x}\|^2\right) \tag{2.24}$$

where $c_{g,d}$ is normalization constant. If we rewrite the equation 2.23 by substituting $g(x)$

$$\begin{aligned}
\widehat{\nabla} f_{h,K}(\mathbf{x}) &= \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{x})\,g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\\
&= \frac{2c_{k,d}}{nh^{d+2}}\left(\sum_{i=1}^{n}\mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) - \mathbf{x}\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\right)\\
&= \frac{2c_{k,d}}{nh^{d+2}}\left(\frac{\sum_{i=1}^{n}\mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}\right)\left(\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)\right)
\end{aligned} \tag{2.25}$$

First term in parenthesis in equation 2.25 is the mean shift and can be abbreviated as $\mathbf{m}_{h,G}(\mathbf{x})$ and the remaining term is **proportional** to the density estimate at x computed with kernel G and can be defined as $\widehat{f}_{h,G}(\mathbf{x}) = \frac{c_{g,d}}{nh^d}\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)$. As a result equation 2.25 becomes as following.

$$\widehat{\nabla} f_{h,K}(\mathbf{x}) = \frac{2c_{k,d}}{h^2 c_{g,d}}\mathbf{m}_{h,G}(\mathbf{x})\widehat{f}_{h,G}(\mathbf{x}) \tag{2.26}$$

and yielding

$$\mathbf{m}_{h,G}(\mathbf{x}) = \frac{1}{2}h^2 c \frac{\widehat{\nabla} f_{h,K}(\mathbf{x})}{\widehat{f}_{h,G}(\mathbf{x})} \tag{2.27}$$

From this equation it is seen that the mean-shift vector is computed with kernel G at point $\mathbf{x}$ is proportional to the normalized density gradient estimate obtained with kernel K. As a result, mean-shift vector always points toward the direction of the maximum increase in density. From equation 2.27, it can be understood that mean is shifted to the region that contains majority of points. Mean-shift vector is aligned with local gradient estimate and if mean-shift vector is iteratively calculated, it will have a path to a stationary point which is the mode point.

Multivariate kernel used for mean-shift procedure in (Comaniciu and Meer, 2002) is given below:

$$K_{h_s,h_r}(\mathbf{x}) = \frac{C}{{h_s}^2 {h_r}^p} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right) \tag{2.28}$$

where $\mathbf{x}^s$ is the spatial part, $\mathbf{x}^r$ is the range part of a feature vector, h_s is spatial kernel bandwidth, h_r is range kernel bandwidth, C is the corresponding normalization constant and $k(x)$ is the common profile of the kernel used in both domains that is given below:

$$k(\mathbf{x}) = e^{\frac{-1}{2}\mathbf{x}} \tag{2.29}$$

2.4.2 Mode Detection

The second step of mean-shift segmentation is mode detection. Actually mode points are the points where $\widehat{\nabla} f_{h,K}(\mathbf{x}) = 0$. However because of bigger step sizes, it is impossible to converge to a mode point in finite number of iteration. Iterations have to be stopped at a level. This is determined by looking the magnitude of the mean-shift vector. Iterations are stopped if the magnitude of the mean-shift vector is smaller than a threshold. All of the points that converge to the same mode point are defined as basin of attraction of that mode point and basin of attraction points can be defined as a cluster.

2.4.3 Clustering

In third step of mean-shift segmentation, mode points that are closer as explained above are concatenated. In last step, smaller clusters or segments than M pixels are eliminated.

2.5 Hough Transform

Concept of Hough Transform was first introduced by Hough(1962) for a patent application and the popular form of the Hough transform, namely transformation to the radius-angle($\rho-\theta$) space is proposed by Duda and Hart(1972).

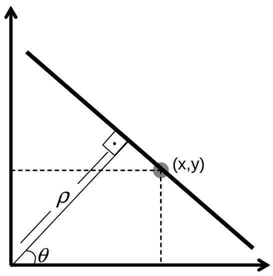


Figure 2.9: ρ and θ parameters of a line

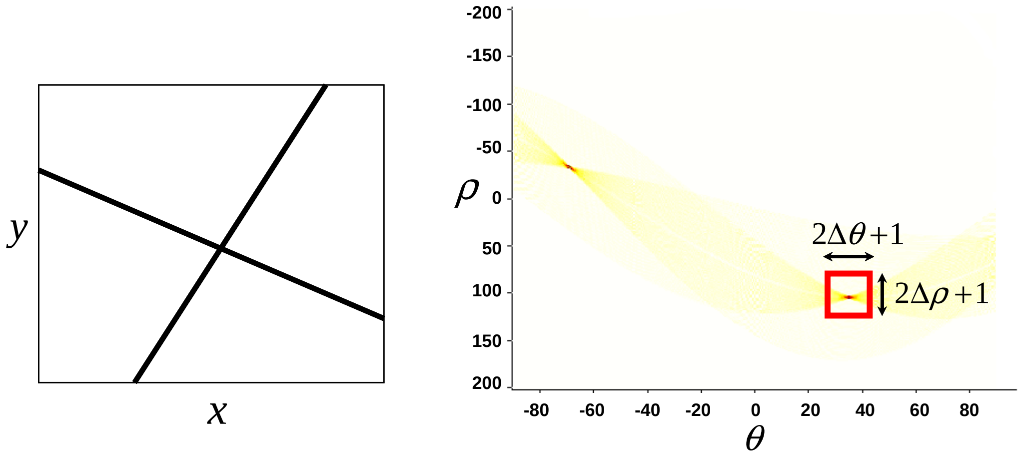
θ is the angle between the x-axis and the normal vector of a line which overpasses from the origin. Range of θ is $[0, \pi]$. ρ is the distance of the line to the origin and its range is $[0, \sqrt{M^2 + N^2}]$ where M and N are dimensions of input image. For the digital images, ρ and θ space is quantized and there are counters for each ρ and θ pair. Value of counter is the magnitude of that point in ρ and θ space. Defined parameters are shown in Figure 2.9. A line can be represented as (x,y) pairs satisfying the equation 2.30 given below:

$$\rho = x\cos(\theta) + y\sin(\theta) \tag{2.30}$$

where x and y are the spatial coordinates of an image pixel. If the equation 2.30 following conclusions can be extracted.

- A point in ρ-θ space, corresponds to a line in x-y space.
- A point in x-y space, corresponds to a sinusoidal in ρ-θ space.

In Figure 2.10 Hough domain representation of a synthetic image is given. It can be seen from the output image that some points are having value higher than others. These points mean that there is a line whose parameters are ρ_0 and θ_0. Actually there are a lot of points whose magnitude is greater than 1 since whatever two point will result to a point of magnitude 2.



(a) Spatial Domain (b) Hough Domain

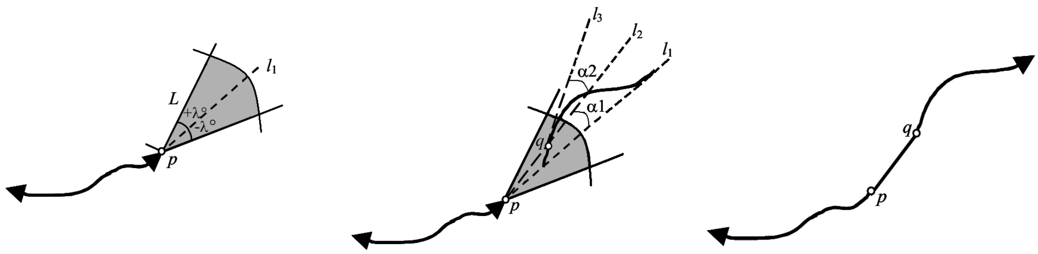
Figure 2.10: Hough Transformation

2.5.1 Line Detection

After the Hough transformation, peaks of the Hough domain representation are detected with a high-pass threshold filter and the ρ_k and θ_k are extracted for $k = 1, 2, ..., n$ if n-lines detected. Detected lines are drawn onto a null image whose size is $M \times N$, and this image is pixel-by-pixel multiplied with input image. Resulting points will show the lines in that image. However in most of the applications, generally there will be no ideal lines on the input image, on the contrary, there will be broken-off lines. After the line detection, points that are belonging to the same line are connected if the distance between them is smaller than a parameter, i.e, 'Fill Gap' parameter. And also smaller lines that only have a few pixels are removed. As an optional step, if $|\rho_a - \rho_b| < \delta\rho$ and $|\theta_a - \theta_b| < \delta\theta$, where $\delta\rho$ and $\delta\theta$ are the threshold values, for two lines a and b then these lines are combined. This operation can also be executed in Hough domain by using a rectangular window whose size is $2\delta\rho + 1$ and $2\delta\theta + 1$ around a detected peak point as illustrated in Figure 2.10-b.

2.6 Edge Linking Method

Edge linking method is explained in (Zhao et al., 2002) and (Christophe and Inglada, 2007). In this method, nearby edges are linked to each other if linking conditions exist. An example edge linking operation is illustrated in the Figure 2.11. The edge in the Figure 2.11-a whose end-point is p, is the analyzed edge and the other edge in the Figure 2.11-b,c whose start-point is q, is the nearby edge.



(a) Search Region (b) Measurements (c) Linking

Figure 2.11: Edge Linking (Zhao et al., 2002)

First of all, a radial search region is constructed whose centerline is in the direction of the end-point of the analyzed edge. The angular width of the radial search region is 2λ and the radius is L. Then it is checked if there is an end-point of a nearby edge in the radial search region. For each nearby edge end-point q within the radial search region, the cost function which is given in equation 2.31 is calculated and the edges that result in minimal cost are linked.

$$E(q) = |\alpha_1| + |\alpha_2| \tag{2.31}$$

Here α_1 is the angle between the line connecting the points p-q and the tangential line of the analyzed edge at point p. α_2 is the angle between the line connecting the points p-q and the tangential line of the nearby edge at point q.

2.7 Convex Hull Method

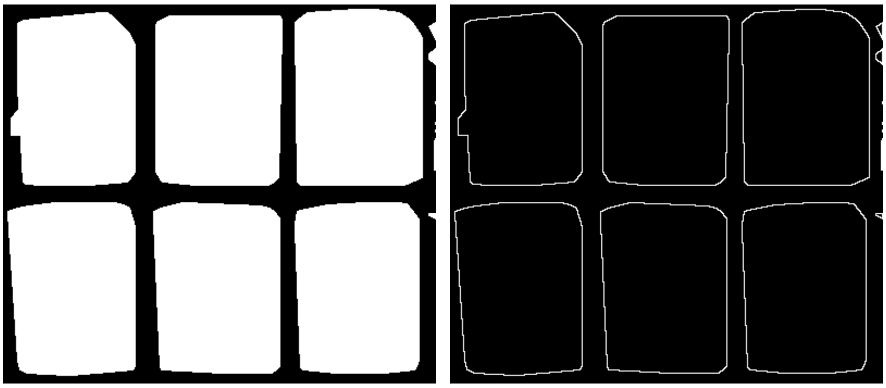
In Convex Hull Method, separated regions are converted to convex objects if they are not in concave shape. By applying this method, zig-zags on the edges of the regions are being removed. The method illustrated in the Figure 2.12 below.

(a) Input Image (b) Convex-Hull Method Applied

Figure 2.12: Convex-Hull Method

2.8 Contour Tracing Method

In Contour Tracing Method, edges of the separated regions are extracted. The method is illustrated in the Figure 2.13 below.



(a) Input Image (b) Contour Tracing Method Applied

Figure 2.13: Contour Tracing Method

2.9 Structural Features

In this thesis structural features will be analyzed in order to detect road segments. Barzohar and Cooper(1996) stated that the variance of the road width is small, the width of the roads changes slowly and the length of the roads is generally big (cited by Zhao et al.(2002)). Hence structural features of segments can be used to determine that if a segment can be a road or not.

There are several structural features that can be extracted from roads. Song and Civco(2004) had defined **perimeter/area ratio** as smoothness similarity and used it in the detection of road segments. As they stated perimeter/area ratio of a road will be high so a thresholding can be applied based on this feature. This ratio is useful to detect long-shaped segments, however, it is dependent on the width of segment. For a narrow segment this ratio is higher than the ratio of a wide segment where both segments equal in length. Generally pavements or openings between the double-direction ways are equal in length and narrow in width when compared with their neighbor road segments so instead of road segments these narrow non-road segments can be chosen as seed segments. Furthermore segments which are not similar in shape with road segments but having too much zig-zags on their edges can be considered as seed segments because this kind of segments generally have a big ratio of perimeter to the area.

The ratio of major-axis length to the minor-axis length of the segment is useful to detect long-shaped segments but dependence on the width of the segment is also a disadvantage for this property. On the other hand segments that have zig-zags on their edges are not problematic. One disadvantage of the ratio of major-axis length to the minor-axis length comes out in the case of a curve-shaped road segment. Since axis lengths are calculated according to the convex shape which is covering the segment, minor-axis length will be big, so the ratio will be small.

Solidity feature is defined as the ratio of the area of the segment to the area of the convex area that covers the segment. It can be used in conjunction with the ratio of major-axis length to the minor-axis length.

Area of discrete regions on a image is also used as a structural feature. Generally, the regions whose area is smaller than a threshold are removed in post-processing operations as performed in (Long and Zhao, 2005).

2.10 Gaussian Mixture Model

In Gaussian Mixture Model(GMM) an input space is modeled with multiple Gaussian distribution functions. Generally sample data points from a space which is being observed are used to obtain the Gaussian Mixture Model of observed space. Input space can be one-dimensional or more. For example, when RGB data is used as the input space each Gaussian mixture will be defined in three dimensions. One dimensional illustration of a Gaussian Mixture Model is given in Figure below. In this Figure, the dashed line is the input space and the solid lines are the modeled Gaussian functions.

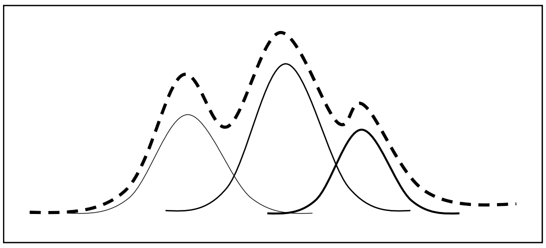


Figure 2.14: Illustration of Gaussian Mixture Model

Number of the mixtures can be defined parametrically. The sample data points that are used in the formation of a mixture called as cluster. Clusters are defined by using two parameters which are mean and covariance matrix. These parameters are estimated form the sample data points because it is not known which data point will contribute to the which Gaussian Model. In this thesis, Gaussian Mixture Model is trained (parameters are estimated) with maximum likelihood criterion. In this estimation **Expectation Maximization algorithm** is used. Maximum likelihood estimation of the mixture model is defined as given below.

$$P(\mathbf{x}|\mu, \Sigma) = \sum_{n=1}^{N} w_n p(x|\mu_n, \Sigma_n) \tag{2.32}$$

where N is the number of the Gaussian mixtures, w_n is the weight of n^{th} cluster and $\sum_{n=1}^{N} w_n = 1$, μ_n is the mean of the n^{th} cluster and Σ_n is the covariance matrix of the n^{th} cluster and d is the dimension of the sample data points which are represented with x. $p(\mathbf{x}|\mu_n, \Sigma_n)$ is the representation of n^{th} Gaussian cluster and defined as given below.

$$p\left(\mathbf{x}|\mu_n,\Sigma_n\right) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_n|^{\frac{1}{2}}} e^{\frac{1}{2}(\overline{x}-\overline{\mu_n})^t \Sigma_n^{-1}(\overline{x}-\overline{\mu_n})} \tag{2.33}$$

Expectation Maximization(EM) algorithm is an iterative method in which the objective is to maximize the likelihood ($P\left(\mathbf{x}|\mu,\Sigma\right)$) of sample data points, $\mathbf{x}$. In GMM case, there is a hidden variable in EM which is describing the membership of the sample points to the Gaussian mixtures. By using this hidden variable, a joint auxiliary function which is maximum when the likelihood of the data is maximum, is defined for all Gaussian mixtures. Maximum likelihood is obtained by searching the case in which the derivative of this function is equal to zero. When it is analyzed, at the end, update of the estimated values of mean(μ_n), variance(${\sigma_n}^2$) and weight(w_n) of Gaussian mixtures will be as given below.

$$\begin{aligned}
\mu_{nj} &= \frac{\sum_{i=1}^{n} x_i P(j|x_i,\mu_n\Sigma_n)}{\sum_{i=1}^{n} P(j|x_i,\mu_n\Sigma_n)} \\
\sigma_{nj} &= \frac{\sum_{i=1}^{n} \left(x_i-\mu_{nj}\right)^2 P(i+1|x_i,\mu_n\Sigma_n)}{\sum_{i=1}^{n} P(j|x_i,\mu_n\Sigma_n)} \\
w_{nj} &= \frac{1}{n}\sum_{i=1}^{n} P(j|x_i,\mu_n\Sigma_n)
\end{aligned} \tag{2.34}$$

These iterative calculations are performed until convergence occurs.

CHAPTER 3

PROPOSED ALTERNATIVE METHODS

In this chapter, the baseline algorithm, the proposed alternative methods and the algorithm proposed as a result of alternatives are explained. The algorithm suggested by Long and Zhao(2005) is selected as the baseline algorithm and implemented with MATLAB software. Some of the steps of the baseline algorithm are replaced with alternative methods and the results are compared. Firstly, the algorithm proposed by Long and Zhao(2005) explained in section 3.1. Then in section 3.2, alternative methods map is explained. In section 3.3, a multi-scale template matching filter which filters the input image according to the its structural features, is explained. Then in section 3.4, two different segment merging methods are explained. Proposed automatic seed selection and spectral classification method is given in section 3.5. Proposed structural verification method is given in section 3.6. Finally, in section 3.7 proposed post-processing operations are explained.

3.1 Baseline Algorithm

The algorithm proposed by Long and Zhao(2005) in which mean-shift segmentation is used, is selected as the baseline algorithm. Flow chart of the baseline algorithm is given in Figure 3.1 and it is explained below.

It is implied by Long and Zhao(2005) that the spectral variation of road pixels of satellite images is large, as a result of material types and aging of roads, shadow of buildings, road signs, all of which can be considered as noisy features. A filter that is diminishing effect of noisy features, increases the quality of the image. Hence firstly, input image is filtered using multi-scale gray-level morphological cleaning and strengthening algorithm (MMCSA-

LZ). The details of the MMCSA-LZ filter were given previously in section 2.3.2.2. The filter is applied to improve the quality of the image, so further steps of the algorithm work more efficiently. This operation can be thought as a step of pre-processing.

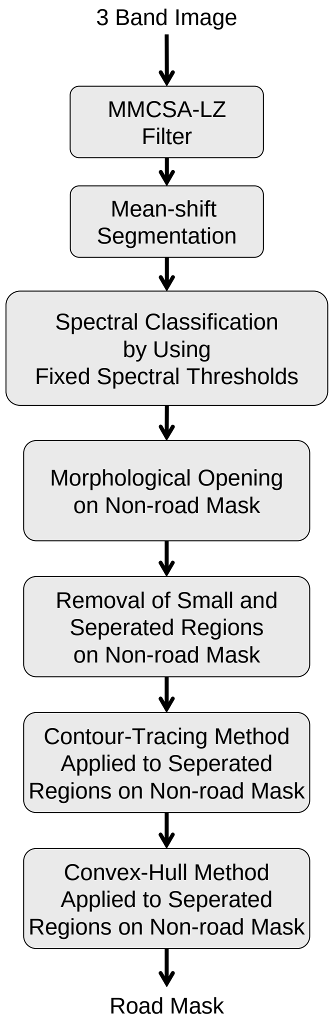


Figure 3.1: Chosen baseline algorithm proposed by Long and Zhao(2005)

Secondly, mean-shift segmentation procedure is applied to the filtered image. As explained in section 2.4, mean-shift segmentation procedure is applied in the joint spatial-range domain and its output is a segmented image, in which segments formed by the pixels that are within a neighborhood and whose spectral properties are similar.

After the segmentation step, a spectral filter is applied to the segmented image by Long and Zhao(2005). They chosen the lower and upper limits of the filter according to their experimental results. This step corresponds to the classification of road and non-road pixels. By applying this filter, the image is converted to a binary image showing road masks, which is formed by road pixels (zeros) and non-road pixels (ones). It is important to mention that roads are represented with zero (0) instead of one (1).

Since the binary image obtained at the classification step contains noisy small white regions and black holes a post-processing step is needed. In their study, region-based post-processing operations are performed. Firstly, binary opening operation is applied to the classified image in order to remove small connections that exist between different regions. Then using connected components analysis, all of the discrete regions on the image are labeled with different numbers. Then white regions that are smaller than an area threshold are removed. Similar to the white regions, there are black holes on the classified image. These holes are removed by using contour-tracing algorithm. Finally, convex hull algorithm is applied to the image, in order to remove zig-zag shaped edges. At the end, a road-edge network is extracted by Long and Zhao(2005).

3.2 Proposed Alternatives to the Baseline Algorithm

As mentioned in section 3.1, the algorithm that is proposed by Long and Zhao(2005) is chosen as the baseline algorithm. There are mainly three main steps in this algorithm which are pre-processing(1), classification(2) and post-processing(3). Pre-processing operations include an edge-preserving smoothing filter and mean-shift segmentation procedure. Classification is performed with a spectral filter using constant thresholds. After the classification step, several post-processing operations are performed. Main idea of the post-processing operations is to process regions in the classified image. All of the sub-steps of post-processing operations of the baseline algorithm are organized to process regions. In the Figure 3.2, the baseline algorithm is the one which goes through solid arrows. Steps of the resulting proposed algorithm which is obtained at the end of the experimental work flow in 4.4, are the alternative steps that have thick frames in the Figure 3.2.

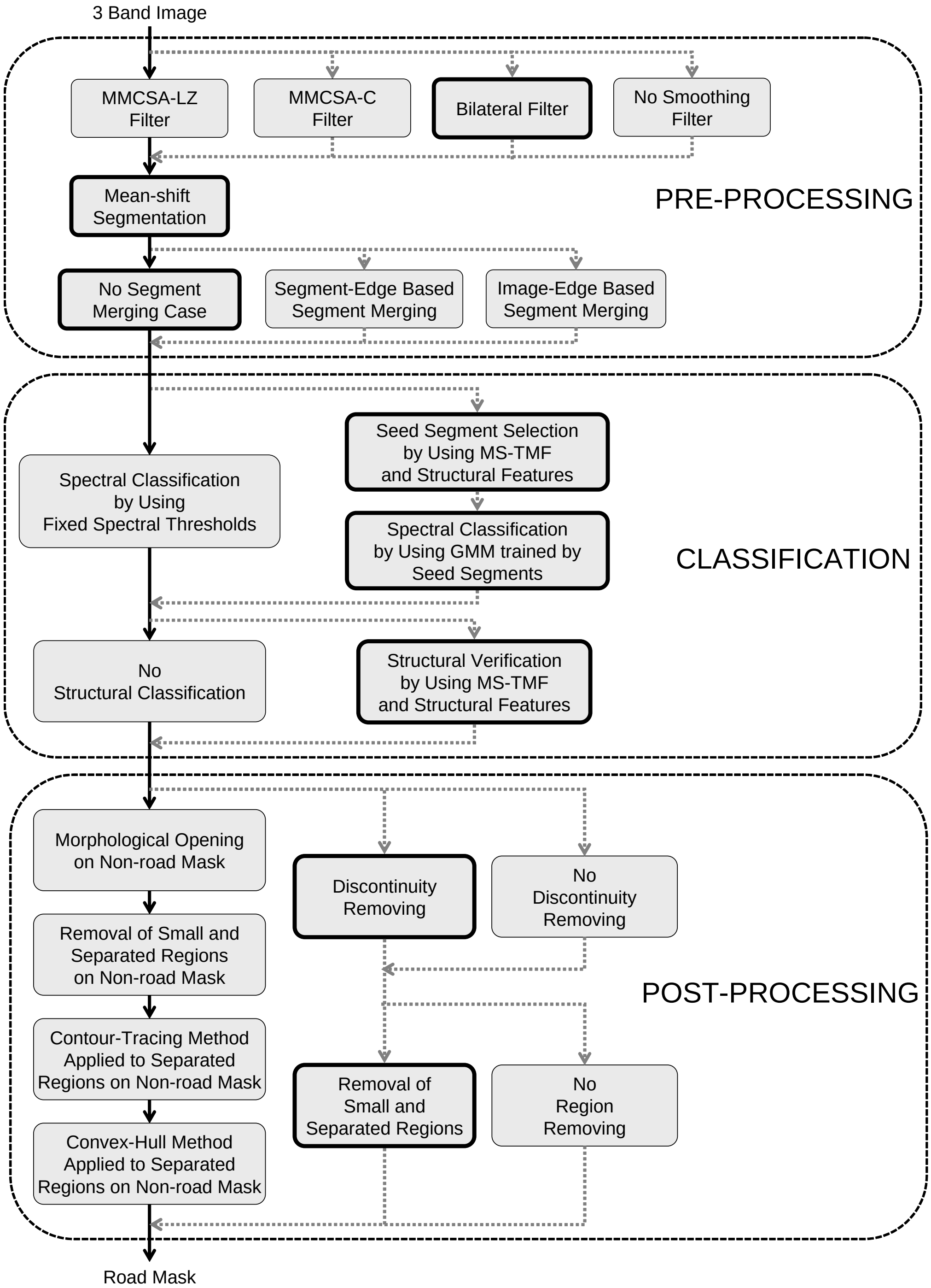


Figure 3.2: Alternative Methods Map

In order to increase the performance of the baseline algorithm, several alternative steps are proposed and performance comparisons are carried out. Alternative steps to the baseline algorithm are the blocks that are in the right-side of the Figure 3.2 which are going through dashed arrows.

In the pre-processing, different edge-preserving smoothing filters are compared. In some cases, applying segmentation procedure results in over-segmentation on roads. Since the proposed automatic spectral classification method is based on the analysis of the structural features of the segments, a proper segmentation is needed. In order to diminish effect of over-segmentation, two different segment-merging methods are proposed. These methods can be thought as post-operation of the mean-shift segmentation procedure.

In the classification, an automatic spectral classification method in which spectral threshold values are determined for each image separately by analyzing spectral features of seed segments, is proposed instead of the spectral filter whose thresholds are fixed. Seed segments are selected among all segments by analyzing structural features of them. A multi-scale template matching filter is used in seed segment selection and structural verification steps as it can be seen in the Figure 3.2. Seed segments are used to train a Gaussian Mixture Model (GMM) in which Expectation Maximization is used. Then weights of resultant clusters of GMM are compared with a threshold and eliminated if they are smaller than it. At the end, by using the mean and standard deviation parameters of passed-clusters, input image is spectrally classified.

In the post-processing, there are four sub-steps in baseline algorithm where all sub-steps are aimed to process regions on the classified image. Instead of region-based post-processing operations, alternative post-processing operations are proposed. There are multiple steps in region-based post-processing operations however all sub-steps of this post-processing method are related to each other so breaking the flow from an intermediate step is not preferred. Steps of the post-processing operations are explained in the related sections below.

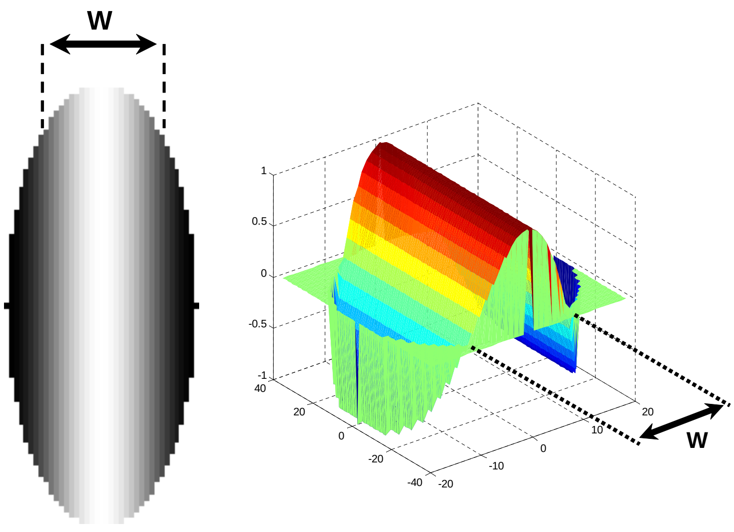
3.3 Multi-scale Template Matching Filter

In this section, the structure of the proposed multi-scale template matching filter(MS-TMF) is explained. Properties of the used template and multi-scale property are explained. MS-TMF can be applied to binary or gray-level images. Normally output of the filter is a gray-level image. However, in this thesis, value of each pixel of the output image is compared with **similarity rate** as it can be seen in the Figure 3.4 and a binary image is obtained. MS-TMF is used for the purpose of structural feature analysis of segments. Binary mask of a segment is filtered with MS-TMF and the binary output image of the filter is compared with the binary mask of input segment in order to decide on the segment if it is road or not. On the other hand, MS-TMF can be used for other purposes, for example, it can be used to post-process a road-mask image in order to obtain more straight edges.

In this thesis, MS-TMF is used to filter segments. Hence, in this section, properties of MS-TMF are explained for the case in which input is a segment rather than a road-mask or another image. The output is a binary image since each gray-level pixel is converted to 1 or 0 by comparing the value of that pixel with **similarity threshold**.

3.3.1 Constitution of the Template

In order to detect particular structural features in an image, template matching filters can be used. Aim of the usage of template matching filters is to remove the structures which are not similar to the template, from the input image. In road extraction algorithms, different road templates can be used to detect roads. In this thesis a 2-dimensional gray-level template is used.



(a) Gray 2-D View (b) Colored 3-D View

Figure 3.3: Cosine-disk template with a target width of 15 pixels

Three parameters are used in the definition of the template. These are target road width of the template, length coefficient and the template type. **Length coefficient** is the ratio of the length of the template to the its width. A 2-D cosine-disk template is used in the template matching filter structure as given in the Figure 3.3. There are **positive coefficients** (bright pixels in the Figure 3.3-a) at the center of the template and **negative coefficients** (dark pixels in the Figure 3.3-a) at the sides of the template. Number of the positive coefficients of the template in horizontal axis, is set to the target road width, **W**, of the template.

3.3.2 Template Matching Filter Structure

Since the orientation of a road (angle between the road and the x-axis) can have any value, the template matching filter has to be rotation invariant. For this purpose, rotated versions of a single template is used simultaneously. Mentioned filter structure is illustrated in Figure 3.4. In this figure, the term $\delta\theta$ represent the angular resolution that is used in rotation of template filter. Rotated versions of the filter are produced with $\delta\theta$ steps till to the 180°.

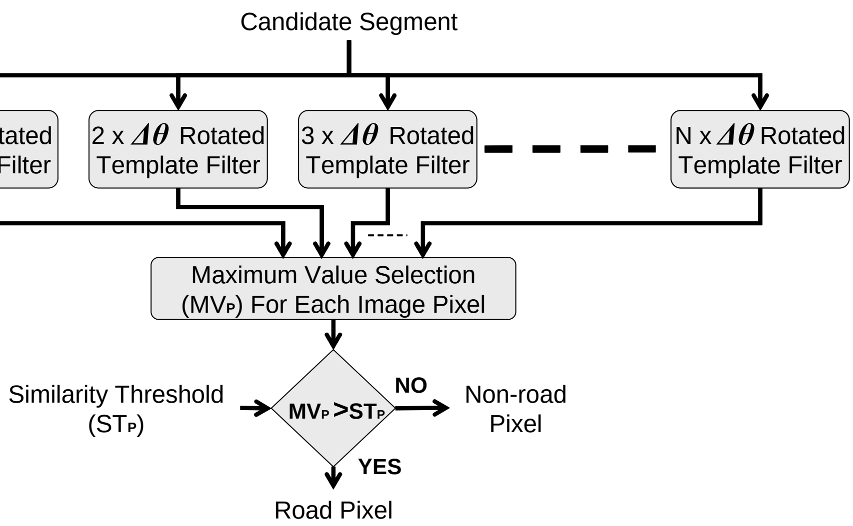


Figure 3.4: Template Matching Filter Structure

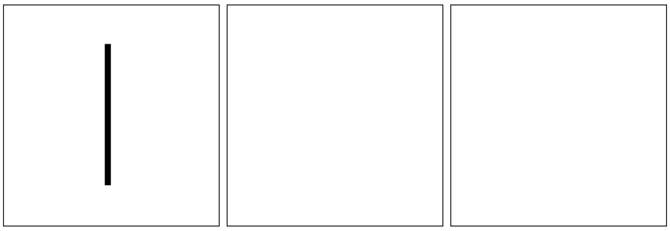
All rotated template filters will produce a value for each pixel. The output value of the rotated template for which the maximum value was obtained is chosen as the new value of that pixel. In other words, the output value of the rotated template which results in the maximum output is chosen. Then each pixel value is compared with a threshold which determined by using **similarity rate**. The pixels whose value is greater than this threshold are labeled with one and the others labeled with zero.

As it is known, in filtering, a window is defined around a pixel and that image window is pixel-by-pixel multiplied with the filter coefficients and the sum of the all multiplications is assigned to be the new value of that pixel. If the input is binary then the value of the output pixel becomes equal to the sum of some of filter coefficients. In this thesis, binary mask of segments is filtered with MS-TMF, hence, coefficients of the filter are used in the decision of the threshold value which is calculated by using **similarity rate**.

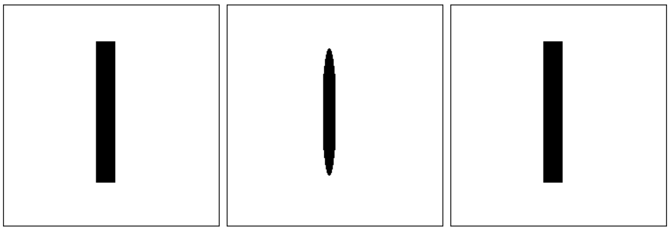
Similarity rate is defined as a ratio which can be between **0** and **1**. **0** corresponds to the sum of all coefficients in the template and **1** corresponds to the sum of all positive coefficients in the template. For example, assuming that sum of all coefficients of the template is -25, sum of all positive coefficients of the template is equal to 75 and the similarity rate is chosen as 0.45. Then the threshold will be equal to $(-25) + 75 - (-25)x0.45 = 20$. A pixel will be labeled with one if its value greater than 20.

3.3.3 Template Matching Filter - Examples

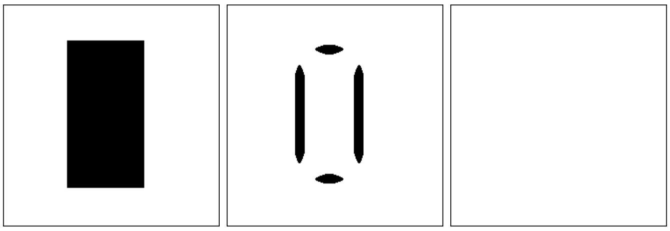
In this section, outputs of the template matching filter for different input cases, are given and operation of the filter is explained. In the Figure 3.5 some possible input segments to the proposed template matching filter are given in first column, the output images obtained as the output of the proposed template-matching filter with $ST_p = 0.25$ are given in second column and final decision for **road likelihood ratio** set to 0.30 is given in third column.



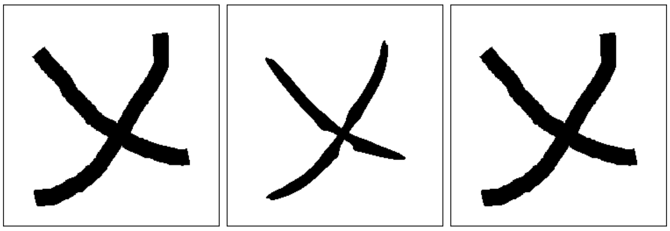
(a) Narrow Road Segment



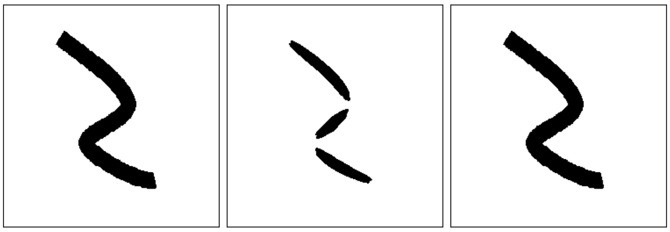
(b) Optimal Width Road Segment



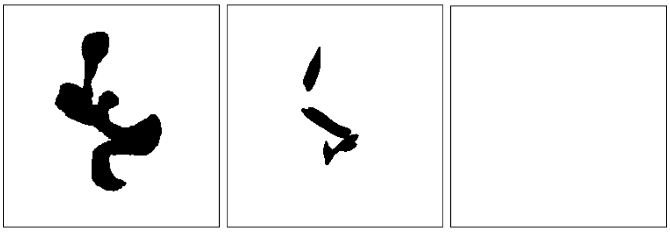
(c) A Wide Non-road Segment



(d) Road Segment with Cross-point



(e) Curve-shaped Road Segment



(f) A Non-road Segment whose Edges have Zig-zags

Figure 3.5: Possible Inputs, Outputs and Final Decision of the Template Matching Filter

Narrow Road Segment: When the template is fitted to the centerline of the segment in Figure 3.5-a, some positive coefficients at both side of the template will be multiplied with zero. Therefore the output value will be smaller than the maximum output value of the filter. If the template is shifted towards left or right, output value will be smaller than the output of the centerline fitted case.

Optimal Width Road Segment: When the template **fitted to** the centerline of the segment in Figure 3.5-b whose width is equal to the target width, positive coefficients of the template are multiplied with one and negative coefficients of the template are multiplied with zero. Therefore value of the output will be equal to the sum of the positive coefficients. As a result, **maximum output value** of the filter will be obtained since all of the positive coefficients are summed. When the template **did not fit to** the centerline of the segment in Figure 3.5-b, namely, if the template is shifted to the left side of the segment, positive coefficients of the template on the left side will be multiplied with zero and negative coefficients on the left side will be multiplied with one. Therefore output value will be **less than the maximum output value** of the filter.

A Wide Non-road Segment: When the template is fitted to center of the segment in Figure 3.5-c then the output value of the filter will be equal to the sum of all the template coefficients. Since there are negative coefficients, the output value will be very smaller.

Road Segment with Cross-point: When the input is the segment in Figure 3.5-d, the filter will not remove that segment as a non-road segment. This is the advantage of the template-matching filter when compared with easily calculated structural features such as the ratio of major-axis length to the minor-axis length, the ratio of perimeter to the area etc. These structural features are being used to make decision on a segment if it is a road or not. Since the segment have a cross-point it can be classified as non-road when structural features used. However when template matching filter is used, it will be classified as road segment if its width is equal to the target road width of the template.

Curve-shaped Road Segment: As in the case of the segment which has cross-point as in Figure 3.5-e, curve-shaped road segments can also be classified as road if template matching filter is used. On the other hand, as a result of rotation invariant property of the template matching filter, curvature that exists on the segment will not effect the decision.

A Non-road Segment whose Edges have Zig-zags: Perimeter to area ratio of road segments is usually big. However, as a result of their zig-zag shaped edges, some segments may have big perimeter to area ratio although they are not belonging to the roads. The segment in the Figure 3.5-f is an example to this kind of segments. If template matching filter is used then this segment is being classified as non-road.

In the second column of the Figure 3.5, corresponding outputs of the segments in column 1 of the Figure 3.5 are given. These outputs are obtained by choosing the **similarity rate** as 0.25. It can be seen that when the segments at Figure 3.5-b,d and e are input to the filter, output is similar to the input. However when the segments at Figure 3.5-a,c and f are input to the filter, output is not similar to the input. In the seed selection and structural verification steps that are explained in following sections, the area of output is divided to the area of the input. This ratio is named as **road likelihood ratio**. If this ratio is smaller than a threshold then corresponding segment is classified as non-road. For the example segments given in the Figure 3.5-a,b,c,d,e,f, road likelihood ratios are 0.0, 0.49, 0.18, 0.58, 0.46 and 0.27, respectively. So if the road likelihood threshold is chosen as 0.30 then only the segments b, d and e will be classified as road.

It is important to mention that examples given above were filtered with only a single target road width. Namely, **Multi-scale** property of MS-TMF is not used.

3.3.4 Multi-scale Property

In multi-scale filter structure several template matching filters having different width are used simultaneously and outputs of them are combined. By applying different target road widths, roads whose widths are different can be detected. Mentioned structure is given in Figure 3.6. The term P represent a pixel and TMF is the abbreviation of template matching filter. These rectangular blocks are containing the structure that is given in Figure 3.4. In addition to the similarity rate, road resolution parameter is used in multi-scale template matching filter structure.

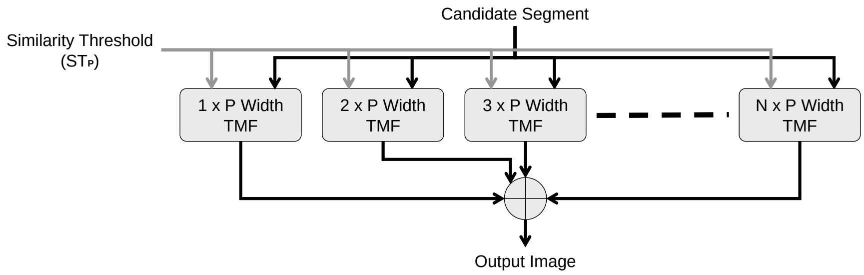


Figure 3.6: Multi-scale Template Matching Filter Structure

As explained in the section 3.3.2, similarity comparison is realized with a threshold which is calculated by using **similarity rate**. Since a rate is chosen as parameter rather than a threshold, it is possible to use only a **single similarity rate** value for all TMFs in MS-TMF structure. There is not any need to calculate a threshold value for every input image.

3.4 Segment Merging

In earlier studies, it has been observed that mean-shift segmentation procedure generally results in over-segmentation, that is dividing road segments into sub-segments. In addition to this drawback, long structures such as buildings which are similar in shape to the roads can lead to incorrect seed detection. In order to overcome these issues, two different segment-merging methods which are segment-edge based segment merging and image-edge based segment merging are proposed.

3.4.1 Segment-Edge Based Segment Merging

In the segment-edge based segment merging method which is shown in the Figure 3.7, firstly, lines on the edges of each segment are extracted by using line detection in Hough space which was explained in section 2.5. After the line detection step, the side information which is the position of a line, with respect to the selected segment is determined. Then segments that are forming a continuous structure with almost identical spectral features are merged by considering the angle, the start-point coordinate, the end-point coordinate and the side information of the lines detected on their edges. Mentioned steps are shown in Figure 3.7 below.

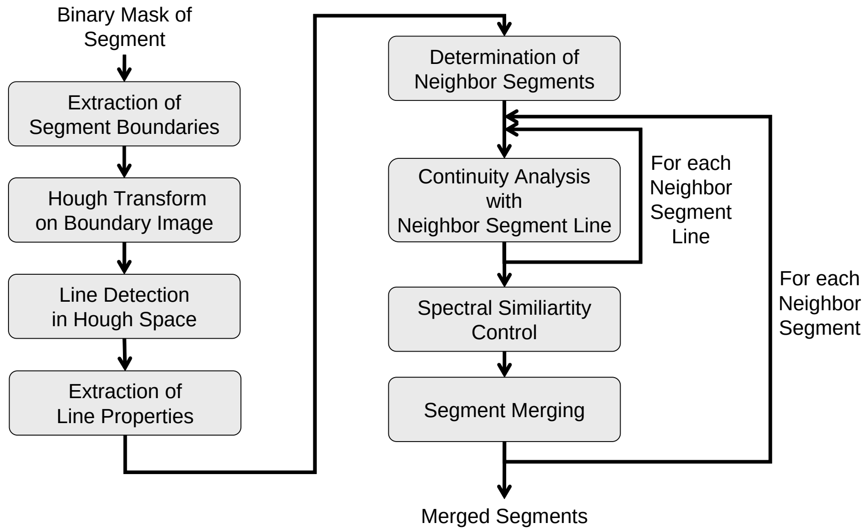


Figure 3.7: Segment-Edge Based Merging Operations of a Segment

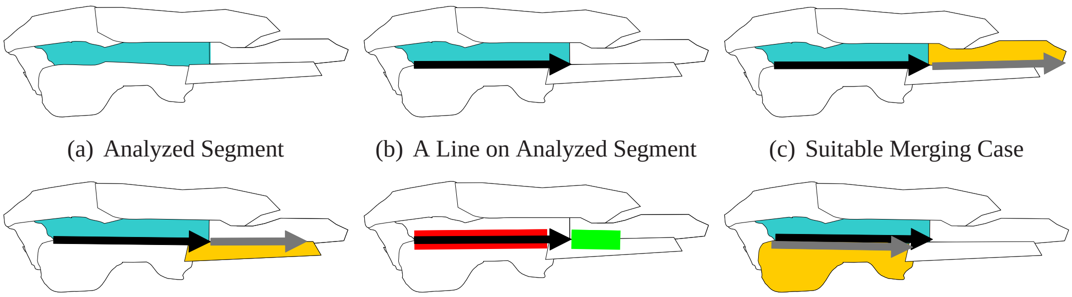
3.4.1.1 Determination of Lines on a Segment

First step of segment merging operation is to determine the lines on the boundaries of a segment. For this purpose, boundaries of the segment are extracted by applying edge detection on binary mask of the segment. Then Hough transform is applied to the edge mask of the segment and straight lines on the edges are detected with the method explained in section 2.5.1.

3.4.1.2 Segment Merging

In this section, segment merging operations are explained. For a better understanding, merging operations are explained for the blue segment given in Figure 3.8-a. Line of analyzed segment is shown with black arrow. Interested neighbor segments are filled with yellow and their lines are shown with gray arrows in Figure 3.8.

After the lines of the analyzed segment are determined, the neighbor segment list which is the list of the segments that are adjacent to the analyzed segment is extracted. Then for each line of analyzed segment, properties of each line of each neighbor segment is compared with the properties of the line of the analyzed segment. As mentioned previously, properties of a line are the start-point and end-point coordinates, the angle and the side information.



(a) Analyzed Segment

(b) A Line on Analyzed Segment

(c) Suitable Merging Case

(d) Incorrect Segment Merging Case prevented by side information analysis

(e) Continuity Analysis Regions

(f) Incorrect Segment Merging Case prevented by continuity analysis

Figure 3.8: Segment-Edge Based Segment Merging

Then four different analysis performed on the features of lines(black and gray arrows in Figure 3.8-c,d,f) in order to decide on segment merging. These are angular similarity analysis, continuity analysis, side information analysis and spectral similarity analysis. These analysis's are performed for each line of each neighbor segment.

Angular Similarity Analysis: In this analysis it is checked whether the difference of the angles of two lines is smaller than the $\delta\theta_{as}$ threshold or not. If it is not, segment merging is not applied.

Side Information Analysis: In this analysis, it is checked whether compared lines are in the same side of their segments or not. Side information of the lines is used in order to prevent incorrect segment merging case which is illustrated in Figure 3.8-d. In this case, segments are not forming a continuous shape however angles of lines are equal. As it can be seen from the Figure 3.8-d, the black line is located on the bottom side of the blue segment while the gray line is located on the top side of the yellow segment. Their side informations are different, hence segment-merging is not performed for the case Figure 3.8-d. Side information is obtained by calculating the outer product between the line of interest and its normal line where lines are treated as vectors.

Continuity Analysis: In this analysis, a region in the direction of the line-of-interest is extracted and it is checked whether the start-point or the end-point of the line of the neighbor segment, is in this region or not. This region is the green region in Figure 3.8-e. If this condition is satisfied then it is checked if the start-point or the end-point of the line of neighbor segment is not included in the forbidden region which is illustrated with red in Figure 3.8-

e. If this condition is also satisfied then it is decided that these two segments are forming a continuous shape. Here, forbidden region information is used in order to prevent the incorrect segment merging case which is illustrated in Figure 3.8-f. By using continuity analysis, incorrect segment merging case that is illustrated in the Figure 3.7-f is prevented.

Spectral Similarity Analysis: In this analysis, it is checked whether segments are spectrally similar or not. If they are not, segment-merging is canceled.

If the result of four analysis's are affirmative than segments are merged.

3.4.2 Image-Edge Based Segment Merging

In the image-edge based segment merging, firstly edge detection is performed on the gray-level input image. Then extracted edge map is processed with a series of operations.

3.4.2.1 Edge Processing

A post processing after edge detection is required before segment merging operation because, edge detection result contains a lot of small edge pieces, curve-shaped edges, intersection points on the edges and broken smooth road edges etc. which are useless or have some drawbacks. In order to remove these drawbacks, a series of edge processing operations are performed. Similar edge processing operations are performed in (Zhao et al., 2002) for the purpose of road extraction where long and straight edge lines are chosen to decide on road seeds. There are small differences and additional steps in the edge processing method that is proposed in this thesis than the method used in (Zhao et al., 2002). Steps of the proposed edge-processing method are given in Figure 3.9.

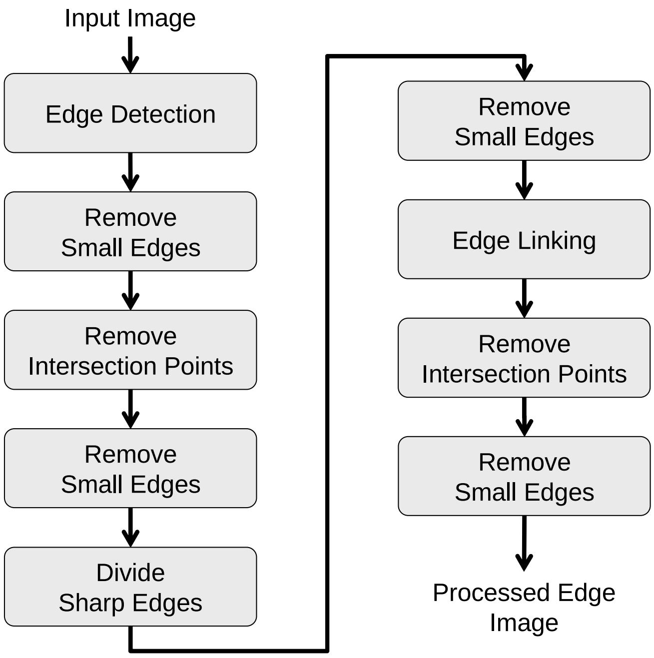
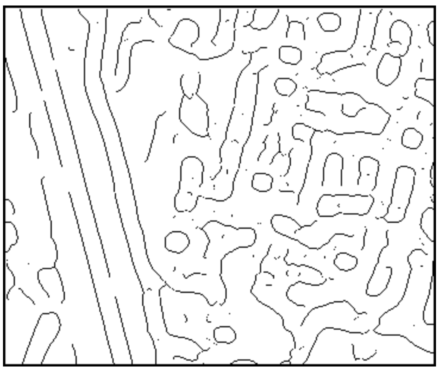


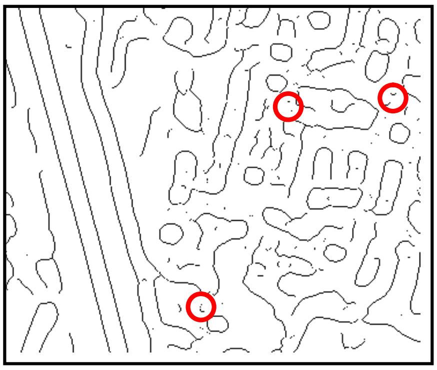
Figure 3.9: Edge Processing Operations

Removal of Small Edges

There is not any need to detect segments around small edges and there is not valuable information for this kind of edges. Hence, differently than the method used in (Zhao et al., 2002), the edges that are smaller than a **length threshold value** are removed in **several steps** of edge processing. In (Zhao et al., 2002), only noisy edge-points are removed at the initial steps. In Figure 3.10 an edge map and the edges that are remained after the removal of small edges are shown.



(a) Edge Map Image

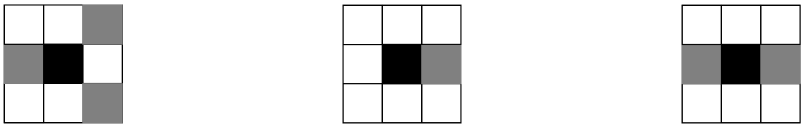


(b) Small Edges Removed

Figure 3.10: Removal of Small Edges

Removal of Intersection Points

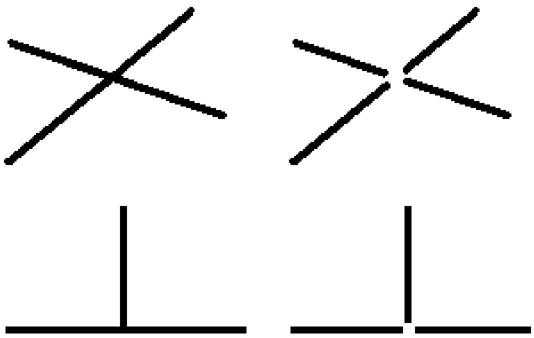
Intersection points are problematic because these points are complicating the determination of list of the neighbor segments and extraction of ordered pixel list of the edges. Therefore intersection points on the edges are removed differently than the method used in (Zhao et al., 2002). This is performed by analyzing neighborhood of each edge pixel. There can be three types of pixels on an edge which are intersection pixel, end-point pixel and regular pixel. A pixel is an intersection pixel if there are more than two neighbor edge pixels around it (Figure 3.11-a). If there is only one neighbor edge pixel around a pixel (Figure 3.11-b) then it is an end-point pixel. If there are two neighbor edge pixels around a pixel (Figure 3.11-c) then it is a regular pixel.



(a) Intersection Pixel (b) End-point Pixel (c) Regular Pixel

Figure 3.11: Edge Pixel Types

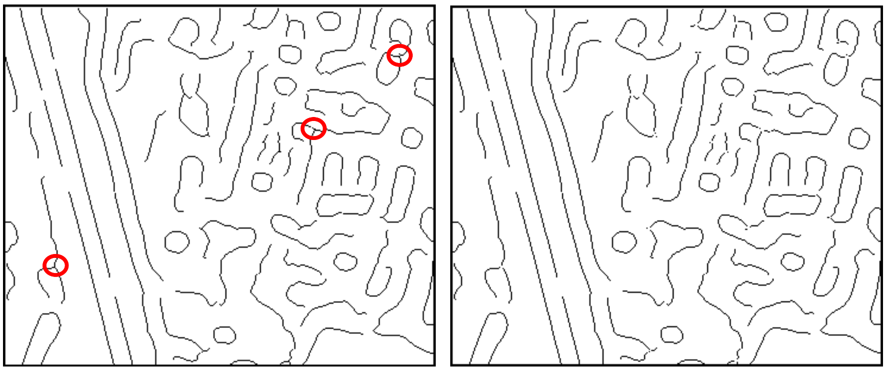
By analyzing the neighbor pixels of each pixel of an edge, intersection pixels on the edges are detected and removed. In Figure 3.12, removal of the intersection points operation is illustrated. After the removal of the intersection points, small edges that came out as a result of this operation, are also removed from the edge map.



(a) Example Edges (b) Intersections Removed

Figure 3.12: Removal of the Intersection Points - Example

In the Figure 3.13, removal of intersection points is shown for the test image.



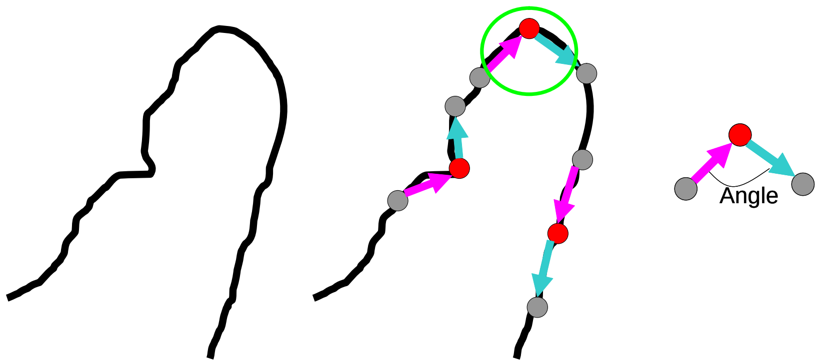
(a) Before Intersection Removal (b) Intersection Points Removed

Figure 3.13: Removal of the Intersection Points

Partitioning Sharp Edges

After the removal of intersection points together with the following small edge removal, all of the discrete edges are labeled with different numbers. For each edge, an ordered pixel list is extracted by starting from one of end-point pixel of the edge and inserting each pixel on the line to the list one by one. Then as applied in (Zhao et al., 2002), edges that are containing sharp transition are broken from the pixel where sharp transition occurs.

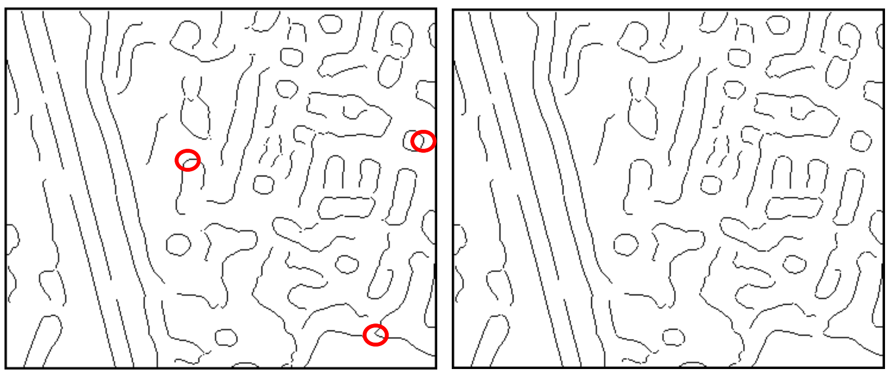
Forward-Backward Angle Analysis: Sharpness detection is performed by analyzing edge-points. For each analyzed pixel of the edge, two vectors defined. Analyzed pixel is the end-point of the first vector (red arrows in Figure 3.14-b) and the start-point of the second vector (blue arrows in Figure 3.14-b). Other end-points of the vectors are 5-pixel away from the analyzed pixel. In this analysis, the angle between these two vector (Figure 3.14-c) is calculated and if there is an angular difference greater than an angle threshold value then analyzed pixel is assumed to be the point where a sharp transition occurs. This is the same analysis with one of the two sharpness detection rules were used in (Zhao et al., 2002) which was called as **global case**. The other sharpness detection rule (called as local) which was used in (Zhao et al., 2002), is not used in this thesis. After the partitioning of the edges, small edges are discarded that came out as a result of the partitioning.



(a) An Example Edge (b) Forward-Backward Angle Analysis (c) Angle Between Points

Figure 3.14: Sharpness Detection

In the Figure 3.15, partitioning of sharp points is shown for the test image.



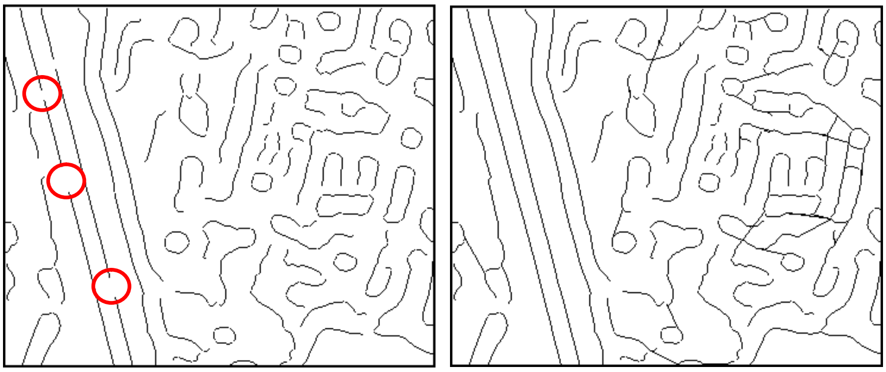
(a) Before Division of Sharp Edges (b) Sharp Edges Divided

Figure 3.15: Division of Sharp Edges

Edge Linking

As a result of trees, shadows, crossroads and etc. smooth road edges are generally broken into sub-pieces in unprocessed edge maps. In order to overcome this issue, edges that are forming a continuous edge are connected to each other. This method is used in (Zhao et al., 2002) and (Christophe and Inglada, 2007). In edge linking, Then a similar method to the edge linking method that is explained in section 2.6 is used to connect the edges that are forming a continuous smooth edge. In next step, the intersection points that appeared because of the edge linking are removed. Finally, the small edges that are came out as a result of the removal of the intersection points are discarded from the edge map. Consequently, processed edge map of the image is obtained.

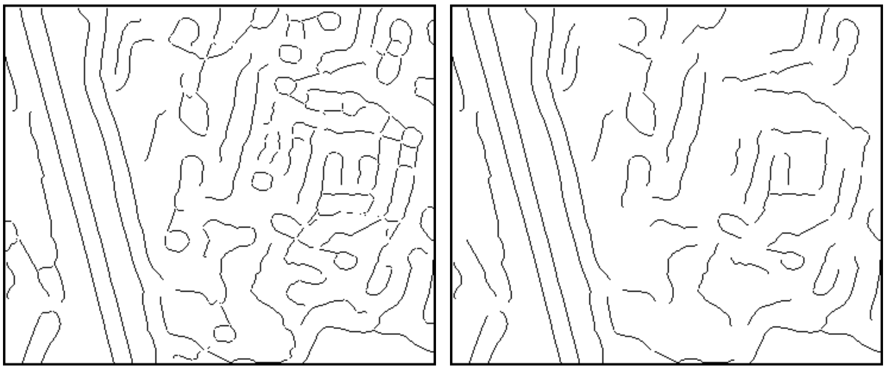
In the Figure 3.16, edge linking operation is shown for the test image.



(a) Before Edge Linking (b) Edges Linked

Figure 3.16: Edge Linking

Finally, in the Figure 3.17, final removal of intersection points and small edges are shown for the test image.

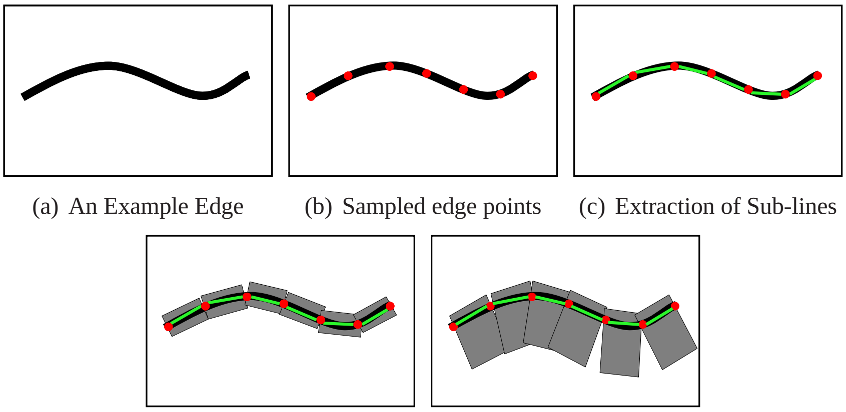


(a) Intersection Points Removed (b) Small Edges Removed

Figure 3.17: Final Operations

3.4.2.2 Determination of Neighbor Segments

For the purpose of the segment merging, segments that are located within the neighborhood of each edge are determined. Neighborhood analysis region is determined by using the shifted versions of the edge. As mentioned before edges are not in the form of a straight line so it is not possible to find the mathematical function of an edge with a short analysis. In order to overcome this problem, curve-shaped edges (Figure 3.18-a) are linearized. Pixels of the edge are sampled with a sample interval and lines that are connecting these sample points are used to represent curve-shaped edges (Figure 3.18-b,c). Then an analysis region of k-pixel width (gray regions in Figure 3.18-d) is formed around these small lines. Because there is an ambiguity between the position of the edge and positions of the segments, experimentally k=7 chosen for 1m resolution images. These analysis regions are connected together and segments which are overlapped by this region are assumed to be the neighbors of that edge. Then this analysis region is shifted towards one side of the edge and is widened to K-pixels and a segment is approved to be a neighbor segment of that side of the edge, if more than half of its area is covered by neighborhood analysis region(gray regions in Figure 3.18-e). Width of a single lane of a road can be thought as 3-meters and if the resolution of the image is 1-meter, a lane of a road will occupy 3-pixels. By widening the width of the neighborhood analysis region to the 20-pixels, roads whose widths up to 20-pixels(meters) are being covered. In other words, it is assumed that roads can not have more than 6-7 lanes. Neighborhood analysis is performed for both sides of the edge and the segments which are not included by neighbor segment list of the edge are removed from side neighbor segment lists.

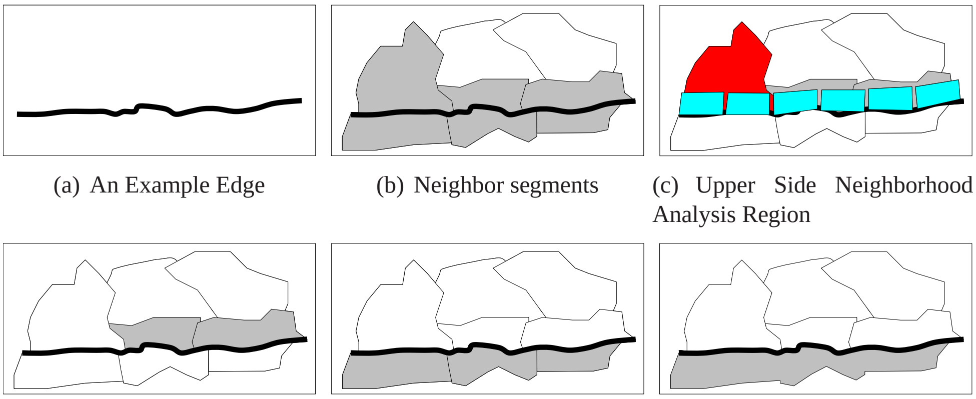


(a) An Example Edge (b) Sampled edge points (c) Extraction of Sub-lines

(d) Neighbor segment detection region(k=7) (e) Side-neighbor segment analysis region(K=20)

Figure 3.18: Extraction of a Neighborhood Analysis Regions

After the determination of the neighbor segment lists for each side of the edge, hierarchical top-down clustering method is applied separately on each side segment lists. In the decision of clustering, spectral features of the segments are used. Then the side with greater number of merged segments is chosen as the merging side for that edge and merging is applied only for that side. Mentioned merging decision is illustrated in Figure 3.19 below. For example, in this figure lower-side segments are chosen as to be merged since the number of merged segments is bigger than the upper-sides. The red segment in the Figure 3.19-c is not selected as a neighbor segment since more than half of its area is not covered by the neighborhood analysis region.



(a) An Example Edge (b) Neighbor segments (c) Upper Side Neighborhood Analysis Region

(d) Upper Side Segments (e) Lower Side Segments (f) Merged Segments

Figure 3.19: Image-Edge Based Segment Merging Decision

3.5 Automatic Spectral Classification

Instead of using constant spectral threshold values to classify the segmented image, an automatic spectral threshold selection method is proposed. This method is applied in two steps. Firstly, seed road segments are determined by analyzing structural features of all segments and using multi-scale template matching filter.

At the end of the structural feature analysis, a number of segments are chosen as seed segments. Spectral characteristic of these seed segments is extracted and then mean and standard deviation of spectral variation are calculated by using Gaussian Mixture Model. A window whose width is proportional to the standard deviation is centered at the mean and lower and upper limit of spectral filter are calculated.

3.5.1 Determination of Seed Segments

In road detection algorithms, classification is realized by comparing input space features with pre-defined features. These pre-defined features can be different for classification methods and types. For example, if the classifier is a neural network than a training procedure is performed. In training, sample data from input space is given to the neural network to learn the features of input space. In automatic classification, input space parameters have to be recognized automatically. If there exist some features that roads generally have then by using this features seed segments can be determined. Seed data is used to extract features of roads when the aim is road detection. In literature, generally spectral features of the seeds are analyzed and the image is classified by performing comparison with these spectral features. Briefly, seeds are the most critical part of classification step, so seed selection is crucial in automatic classification. By using general features of roads, seed segments can be determined. Then by using spectral features of seed segments classification can be performed. When it is analyzed it can be seen roads have recognizable structural features. Since segmentation is used in this thesis, structural features of segments can be used in seed selection. There are several structural features for a segment such as area, perimeter/area ratio, major-axis length/minor-axis length ratio, solidity, similarity rate and etc. In order to select road segments as seed segments, one or more of these structural features can be used.

3.5.1.1 Structural Features for Discriminating Road Segments

In order to determine seed segments, a priority based analysis is performed by analyzing structural features of all segments. The area of the segments, the ratio of the major-axis length to the minor-axis length of the segments and multi-scale template matching filter are used in priority-based analysis which is performed in three-steps.

In first step, segments are sorted according to their **ratio of the major-axis length to the minor-axis length**. This ratio ensures that segments that are not similar to the roads, are not listed at the top of the sorted list.

Then in second step, from top of this sorted list, the first 100 segments that have the biggest area are chosen. **Number of chosen segments** is defined as an input parameter to the method. Pavements, shadow of buildings and opening between the double-direction ways are similar in shape with narrow roads so there can be erroneous seed detection. By using the **area feature**, choosing narrow road-shaped non-road segments as seed is prevented. Additionally, choosing large-scaled segments as seed will give more realistic spectral characteristics which will be useful in classification step.

In the last step of the determination of seeds, binary mask of chosen segments are filtered with the **multi-scale template matching filter**(see section 3.3 for details) and the filtered segments whose road likelihood feature bigger than a threshold value are determined as seed segments of the roads. There are two parameters used with MS-TMF which are **similarity rate** and **road likelihood ratio**.

The analysis that explained above is used in selection of seed segments. Chosen seed segments are used in the decision of thresholds of automatic spectral classifier. Hence the road segments will be classified as road in classification step, if their spectral features similar to the seed segments even though their shape is not similar to the roads.

3.5.2 Classification by Using Spectral Properties of Seed Segments

All of the pixel values of seed segments are assumed as road pixels and Gaussian Mixture Model is trained by using Expectation Maximization algorithm by considering spectral features of these pixels. Then spectral threshold values are determined by using mean and standard deviation of the spectral distribution of the clusters. If number of the pixels in a cluster is a small piece of all seed pixels then that cluster is ignored. One-dimensional case of mentioned spectral filter for a single cluster is illustrated in the Figure 3.20. Mean and standard deviation of each cluster are calculated and a window whose width is equal to twice the standard deviation(σ_1) multiplied with a **spectral window size coefficient parameter** which is illustrated as k_1 in the Figure 3.20, is centered at the mean-point in spectral range. The start-point and end-point of window are chosen as lower and upper thresholds of spectral filter.

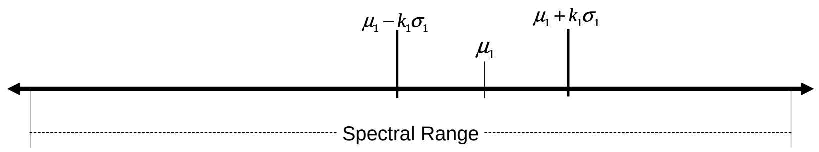


Figure 3.20: Spectral Filter for One Cluster

After the determination of all of the spectral windows of all clusters, image is filtered and a binary image classifying road (white) and non-road (black) regions is obtained. When obtaining binary image result of each cluster is analyzed separately. In this analysis, **spatial coverage limit parameter** which is a threshold is defined. If the ratio of the number of resultant road pixels is greater than the number of all image pixels then **spectral window size**

coefficient is made small 10% of its value. This operation consecutively performed until the spatial coverage limit is not exceeded. In order to not loose any spectral feature that exists in the original image, filtering is performed on the smoothed image rather than segmented image. After the classification step, segment information is reused to further process the classified image. Pixels which are classified as road are counted for each segment and this value is divided to the area of the segment. If this ratio is greater than a threshold which is defined as **segment pass ratio** then all of the pixels of that segment are considered as road pixel.

It is important to say that, if there are more than one clusters, different clusters can be related to different road types such as earth road, asphalt road etc or subtypes of the same road types. So there will can be two or more different filter window on spectral range. Briefly increasing the cluster number can aid to classify different road types simultaneously.

3.6 Structural Verification

Generally, there are non-road regions that are spectrally similar to the roads so when spectral classification is performed, these non-road regions are also classified as road. In order to remove these non-road regions, structural verification can be used. In road extraction literature, it is an agreeable assumption that roads generally have recognizable structural features. As it is explained in section 2.9, structural features can be used in road detection. In the proposed automatic spectral classification step (section 3.5), a binary road-mask image is obtained but segment informations are also preserved by investigating the coverage rate of segments. Hence structural feature analysis can be performed on these segments.

In baseline algorithm, there is not any structural verification after the spectral classification step. Instead post-processing operations are performed on classified binary image without the usage of segments that were obtained in segmentation step. The post-processing operations used in the baseline algorithm are morphological opening operation (for preparation of post-processing), removal of small regions (to make smaller false-positive space), contour-tracing algorithm (to fill black holes) and convex hull algorithm(to remove zig-zags from the edges of roads). Before applying post-processing operations, binary image is labeled by using connected component analysis in (Long and Zhao, 2005) and then post-processing operations are applied to the labeled regions. It means that although segment properties were available, they are not being used.

In structural verification step that is proposed in this thesis, segment features are being used. Firstly, the rate of the major axis length to the minor-axis length is calculated for all of the segments. As mentioned previously that the major/minor ratio is small if segment is in a form that is not similar to straight line. In order to take into account this drawback of the feature, solidity feature is also used in calculations. If solidity feature is big enough and major/minor ratio feature is small for a segment then it is removed. In next step, multi-scale template matching filter(MS-TMF)(see section 3.3 for details) is applied to the segments. MS-TMF is not applied to all segments because of zig-zags on the edges of the road-segments. It is revealed that MS-TMF is removing some of the road segments that are structurally similar to the roads because of zig-zags. In order to prevent this drawback, MS-TMF is applied only to the segments whose major/minor ratio is smaller than a threshold value. Namely, it is assumed that if a segment have a big major/minor ratio than it is considered as road. After the usage of MS-TMF, road likelihood feature is calculated by comparing areas of the segments and their outputs. Finally, the segments with small road likelihood feature are removed. Road likelihood feature is the measure that if a segment structurally similar to roads or not. It is defined as given in below equation.

$$Road\ Likelihood = \frac{Area\ of\ MS - TMF\ Output}{Area\ of\ Segment} \tag{3.1}$$

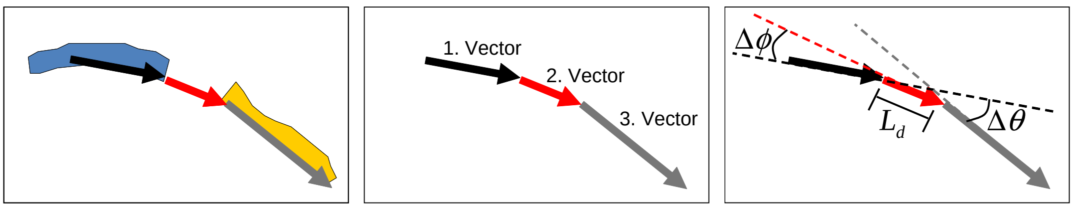
3.7 Proposed Post-Processing Operations

Discontinuity removing and small and separated region removing methods are used to post-process the classified image. Discontinuity removing method proposed in this thesis is similar to the edge linking method that is used in (Zhao et al., 2002) but not the same. There are different definitions and decision rules than the method used in (Zhao et al., 2002). Small and separated region removing is not a proposed method. It is being described here, since it is used in proposed post-processing operations. Here, area removing is used after the removal of discontinuities so it is aimed to prevent removal of small road parts that are separate from the road mask.

3.7.1 Discontinuity Removing

There may spectral or structural reasons that result in discontinuities that exist on classified road-masks. Shadows, cars and trees are spectrally different than roads so if these objects are on a road, there will be a discontinuity on the classified image. On the other hand, some roads are similar spectrally to their environment so in segmentation procedure, that part of the road is grouped with its environment pixels. As a result, discontinuity occurs in structural verification step at that region.

In order to remove discontinuities, segment informations are used. Similar to the segment-edge based segment merging method, for each segment, the list of nearby(but not adjacent) segments is extracted. Then edges of binary mask of the analyzed segment are extracted and Hough transform is applied to this edge map image. For each line on the boundary of the analyzed segment, continuity analysis is performed with each line of each nearby segment. In this analysis, three vectors used to determine if there is a continuity or not. First vector is the line of analyzed segment, second vector is the line of nearby segment and third is the vector which links the first vector with second vector. In order to make decision on a continuity, three comparison realized for three measurements. **First measure** is the length of discontinuity(L_d). This measure is the length of second vector and gives the length of interval that will be removed for the case of continuous segments. **Second measure** is the angular difference of first and second vector($\Delta\theta$) and **third measure** is the angular difference of first and third vector($\Delta\phi$). These measures are illustrated in the Figure 3.21. Comparisons for these measurements is as below.



(a) Nearby Segments and Their Lines

(b) Defined Vectors

(c) Defined Measurements

Figure 3.21: Discontinuity Removing

$\Delta\phi$: It is checked whether $\Delta\phi$ is smaller than 90° or not.

$\Delta\theta$: It is checked whether $\Delta\theta$ is smaller than an angular threshold or not.

L_d : It is checked if L_d is smaller than the length threshold which is defined as below:

$$L_d = L_{max} \times (cos\Delta\phi)^{k_{ad}} \tag{3.2}$$

Here L_{max} is the maximum discontinuity that can be removed and k_{ad} is a coefficient which is preventing the sharp connections.

If these three measurements are smaller than the defined thresholds then two segments are linked to each other. The width of the link is assigned as the average of average widths of analyzed segment and neighbor segment.

3.7.2 Small and Separated Region Removing

In this step, the image in which discontinuities are removed, is labeled by using connected component analysis. It is assumed that roads are connected to each other. But there are also some regions that are spectrally and structurally similar to the roads, so these regions are not removed in classification step. In this step, area feature of all separated regions(not segments) is calculated and the regions which are smaller than an area threshold are removed.

3.8 Resulting Proposed Algorithm

In the next chapter, results of experimental analysis will be presented. A work flow is performed in experimental analysis. At the end of the work flow, an improved version of the baseline algorithm is obtained. Since the details of the proposed algorithm are becoming clear after the experiments, the proposed algorithm will be defined at the end of the chapter 4.

CHAPTER 4

EXPERIMENTAL ANALYSIS

In this chapter, experimental results that were revealed by implementing the algorithms and methods in this thesis are presented. In order to measure the performance of the algorithms, a data set of satellite images is used where images are 1m resolution 3-band images. Image sizes vary between 1024x1024 pixels and 2048x2048 pixels. Pixel-based ground truth data of all test images is extracted manually.

In section 4.1, performance criteria that are common in performance measurement are presented and the ones used in this study are mentioned. In section 4.2, usage of alternative color spaces is presented. In section 4.4, experimental work flow that is performed during experimental analysis is explained. In the following sections, performance measurements and parameter tuning results for the steps of the experimental work flow are presented. Finally, in section 4.17, alternative algorithm that is proposed as an improvement to the baseline algorithm is given.

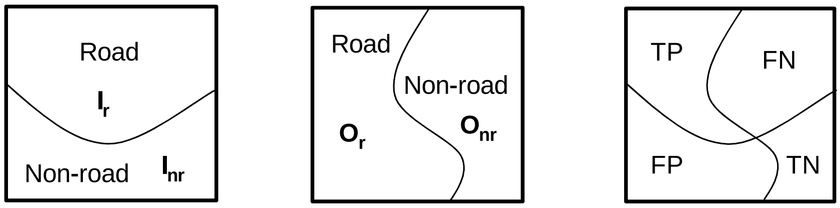
4.1 The Performance Criteria

There are several measures to evaluate the performance of proposed algorithms in literature. All of these measures are calculated with the usage of well-known comparison terms: true positive, true negative, false positive and false negative. In this section, these terms and the performance criteria that are used in the performance measurements will be explained.

An input image can be thought as an input space (I_s) which is formed by the road sub-space (I_r) and non-road sub-space (I_{nr}), where $I_r \cup I_{nr} = I_s$ and $I_r \cap I_{nr} = \oslash$. Similarly, an algorithm that tries to extract roads from an input image, generally gives a binary output image (O_s) which is formed by output road sub-space (O_r) and output non-road sub-space (O_{nr}) where $O_r \cup O_{nr} = O_s$ and $O_r \cap O_{nr} = \emptyset$.

Elements of these spaces can be pixels, segments, lines etc. Performance measurement is realized by comparing output spaces with input spaces where input space is represented with ground truth image. When the elements of spaces are chosen as **pixels**, ground truth image will be a binary image in which road pixels are filled with 1(one) and non-road pixels are filled with 0(zero). When the elements of spaces are chosen as **segment**, input image has to be segmented and roads have to be classified as road segments and non-road segments in order to form a ground truth. When the elements of spaces are chosen as **lines**, input and output road-spaces can be defined but definition of non-road space for both input and output can be complex because it is difficult to assign lines to non-road pixels. There can be infinite number of non-road lines. Hence, performance measurements that are using true negative space can not be evaluated. As mentioned above in the pixel case, all of the spaces can be defined easily by just marking the pixels as road or non-road. Disadvantage of the pixel based space definition is the extraction of ground truth data because it is a time-consuming process. In this thesis, pixel based road and non-road spaces are used to evaluate the performance. Ground truth images of test images are extracted manually for the purpose of performance evaluation.

Number of elements in both input space and output space are equal to each other i.e, equal to the number of pixels of input image, $nI_s = nO_s$. In a fully successful road extraction algorithm, $O_{nr} = I_{nr}$ and $O_r = I_r$. Generally, there will be misclassified pixels therefore these equalities will not be true for all algorithms. If input space and output space are merged pixel-by-pixel then there will be four sub-spaces in non-ideal extraction case. This is illustrated in Figure 4.1.



(a) Input Space (b) Output Space (c) Comparison Space

Figure 4.1: Illustration of Spaces

If a pixel $p_{i,j}$ is extracted as a road pixel by the proposed algorithm ($p_{i,j} \in O_r$) and it is in the ground truth ($p_{i,j} \in I_r$) then it is called **True Positive(TP)**. If that pixel (i.e $p_{i,j} \in O_r$) is not included by the ground truth ($p_{i,j} \in I_{nr}$) then it is called **False Positive(FP)**. Similarly if a pixel is not extracted as a road pixel by the proposed algorithm ($p_{i,j} \in O_{nr}$) and it is not in the ground truth ($p_{i,j} \in I_{nr}$) then it is called **True Negative(TN)**. If that pixel (i.e $p_{i,j} \in O_{nr}$) is included by the ground truth ($p_{i,j} \in I_r$) then it is called **False Negative(FN)**. Number of elements in output spaces TP, TN, FP and FN are counted and performance evaluations are realized according to these numbers. Mathematically, number of elements in each space are given for an $M \times N$ image in Equation 4.1.

$$
\begin{aligned}
nTP &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{i,j} \qquad \forall (p_{i,j} \in I_r) \quad and \quad (p_{i,j} \in O_r) \\
nTN &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{i,j} \qquad \forall (p_{i,j} \in I_{nr}) \quad and \quad (p_{i,j} \in O_{nr}) \\
nFP &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{i,j} \qquad \forall (p_{i,j} \in I_{nr}) \quad and \quad (p_{i,j} \in O_r) \\
nFN &= \sum_{i=1}^{M}\sum_{j=1}^{N} p_{i,j} \qquad \forall (p_{i,j} \in I_r) \quad and \quad (p_{i,j} \in O_{nr})
\end{aligned}
\tag{4.1}
$$

As mentioned above, there are several performance measurement methods in the literature. Popular performance measurement methods are explained below.

Precision which is also called as correctness, is the measure of how much the road-space which is extracted by an algorithm is correct. Its mathematical definition is given below:

$$
Precision = \frac{nTP}{nTP + nFP} \tag{4.2}
$$

Recall which is also called as **Completeness**, is the measure of how much the road-space which is extracted by an algorithm is matching with the ground truth. Its mathematical definition is given below:

$$Recall = \frac{nTP}{nTP + nFN} \tag{4.3}$$

Specificity which is also called as **True Negative Rate** is the measure of how much the non-road space which is extracted by an algorithm is really non-road space. Its mathematical definition is given below:

$$Specificity = \frac{nTN}{nTN + nFP} \tag{4.4}$$

Precision and recall evaluations individually can not represent the real performance of an algorithm. They are related to each other. For example, if every pixel on an image are classified as road ($O_r = O_s$) then recall will be equal to 1 since false positive space is not used in recall evaluation. On the other hand, precision will be equal to the ratio of the number of road pixels to the number of image pixels ($\frac{I_r}{I_s}$). Similarly, if False Positive space is empty, namely, if all the extracted road pixels are included by the ground truth but if they do not cover all of the ground truth ($O_r \subset I_r$ and $O_r \neq I_r$), then precision will be equal to 1. However in this case, recall will be equal to the ratio of the number of pixels that are found as road to the number of pixels of ground truth ($\frac{O_r}{I_r}$). Briefly, by considering only one of the precision-recall evaluations, the real performance of an algorithm can not be shown. For this reason, more complex measurement criteria are also defined for performance evaluation.

Accuracy is the measure of how the pixels in an image are classified truly. In other words, it measures correctness of both output road space O_r and non-road space O_{nr}. Its mathematical definition is given below:

$$Accuracy = \frac{nTP + nTN}{nTP + nTN + nFN + nFP} \tag{4.5}$$

Quality factor measures how much the classified road-space O_r and non-road space O_{nr} are matching to the ground truth in which false positive space is taken into account as an addition to the recall criteria. Its mathematical definition is given below:

$$Quality = \frac{nTP}{nTP + nFN + nFP} \tag{4.6}$$

F-measure is the harmonic mean of precision and recall. Its mathematical definition is given below:

$$F_{measure} = 2\frac{Precision \times Recall}{Precision + Recall}; \tag{4.7}$$

Note that generally, these measures are given as percentage ratios where calculated measure is multiplied by 100. In this thesis, all of the explained performance criteria are calculated for all test images. However precision, recall and quality factor measurements will be given in results.

4.2 Color Space Selection Alternatives

In this section, color space selection for different steps of the algorithms will be explained. Alternative map of color spaces used for different steps are given in Figure 4.2.

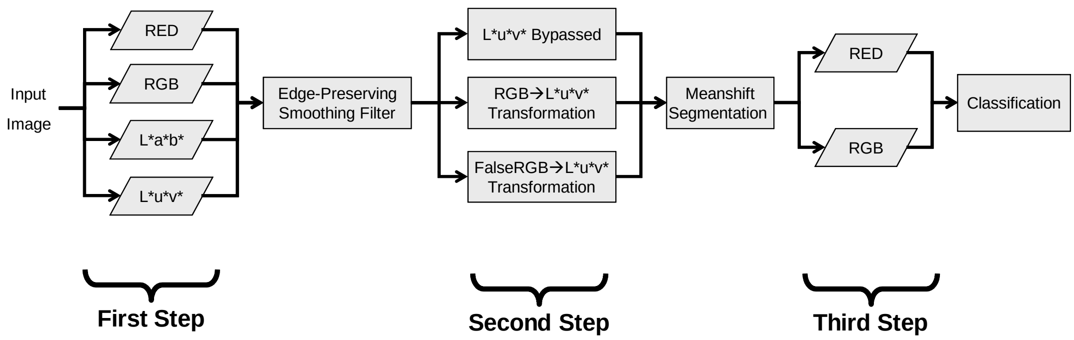


Figure 4.2: Alternative Map for Color Space Selection During Processing

4.2.1 Color Space Selection for Pre-processing Step

For the first color space selection, there are four different color space types as given in Figure 4.2 that are chosen as candidate for applying edge-preserving smoothing filter. A specific color information is not mentioned in baseline algorithm but it is mentioned that the test image used in it, is the red band of an IKONOS multi-spectral image. For this reason red band is chosen as an alternative color space for first selection. On the other hand, RGB data is available for the used test images so second color space alternative for first selection is selected as RGB.

In order to use multi-band images in filtering operations conversion to CIE spaces are used in literature. In the articles that describing bilateral filter (Tomasi and Manduchi, 1998) and mean-shift filter (Comaniciu and Meer, 2002), transformation to CIELAB or CIELUV color spaces is proposed. If R, G and B bands are processed separately, than blurring occurs on the edges of the image as stated by Tomasi and Manduchi(1998). They presented results for bilateral filtering using both RGB and L*a*b* and remarked the blurring effects on the result of RGB processing, which does not exist on the result of L*a*b* processing. Beside this, Comaniciu and Meer(2002) stated that CIE spaces are more suitable to represent Euclidean distance within their three dimensional space. Hence, two CIE space alternatives L*u*v and L*a*b* are added to to first step of alternative color space map. Note that L*u*v and L*a*b* spaces are obtained by applying mathematical transformations to the RGB space.

4.2.2 Color Space Selection for Segmentation Step

As mentioned in 2.4, mean-shift segmentation procedure that is used in this thesis is the one which is developed by Comaniciu and Meer(2002). They proposed to use L*u*v color space as input to the mean-shift segmentation. Hence all of the pre-processed images are converted to the L*u*v* color space in the second step of color space alternative map. In addition to this, it is mentioned by Long and Zhao(2005) that L*u*v* transformation is used in segmentation step of baseline algorithm, however they presented results of a red band image. As mentioned above, L*u*v* space is obtained from RGB space by using mathematical transformations. Hence it is understood that RGB data is used in the baseline algorithm for the mean-shift step.

4.2.3 Color Space Selection for Classification Step

Another color space selection is performed after the segmentation procedure. Classification can be performed on different color spaces. Three-band color space RGB is used in classification step. In addition to it, in order to represent the baseline algorithm, red band is presented as an another alternative.

4.3 Analysis of Baseline Algorithm

The first step of the baseline algorithm is edge-preserving smoothing filter which has three input parameters. (See section 2.3.2.2 for details.) Long and Zhao(2005) stated that only disk (circle)-shaped structuring element (SE) was used in MMCSA-LZ. They did not give any information about the final scale parameter (n) of the SE. In the experiments that are performed in this thesis, final scale is initially chosen as 4 before the parameter tuning steps. It means that SE is dilated with itself 4 times at the fourth scale. Second parameter is the threshold values that are used to remove small color differences from the residual tophat and bothat images. According to their experimental results, Long and Zhao(2005) stated that the best results are obtained with the threshold values that are in the range of $[1, 2]$. They used red band of an multi-spectral satellite image as the input to the algorithm. In second step of the baseline algorithm, mean-shift segmentation is used with the parameters spatial resolution(hS), range resolution(hR) and minimum region(MR). Spatial and range resolution parameters are given in (Long and Zhao, 2005). Minimum region parameter is chosen as 150 initially. After the segmentation step, Long and Zhao(2005) applied spectral classification with the constant spectral threshold values that are [21-54]. They stated that these spectral threshold values are determined as a result of their experiments. Finally, in post-processing step, they use binary opening operation with disk radius 3. Then at the end, regions that are smaller than a threshold are removed. They did not give the value of the area threshold hence it is chosen 1100 after parameter tuning.

4.4 Experimental Work Flow

Several experiments and implementations were done during studies. At the end, an experimental work flow is established as given in the Figure 4.3.

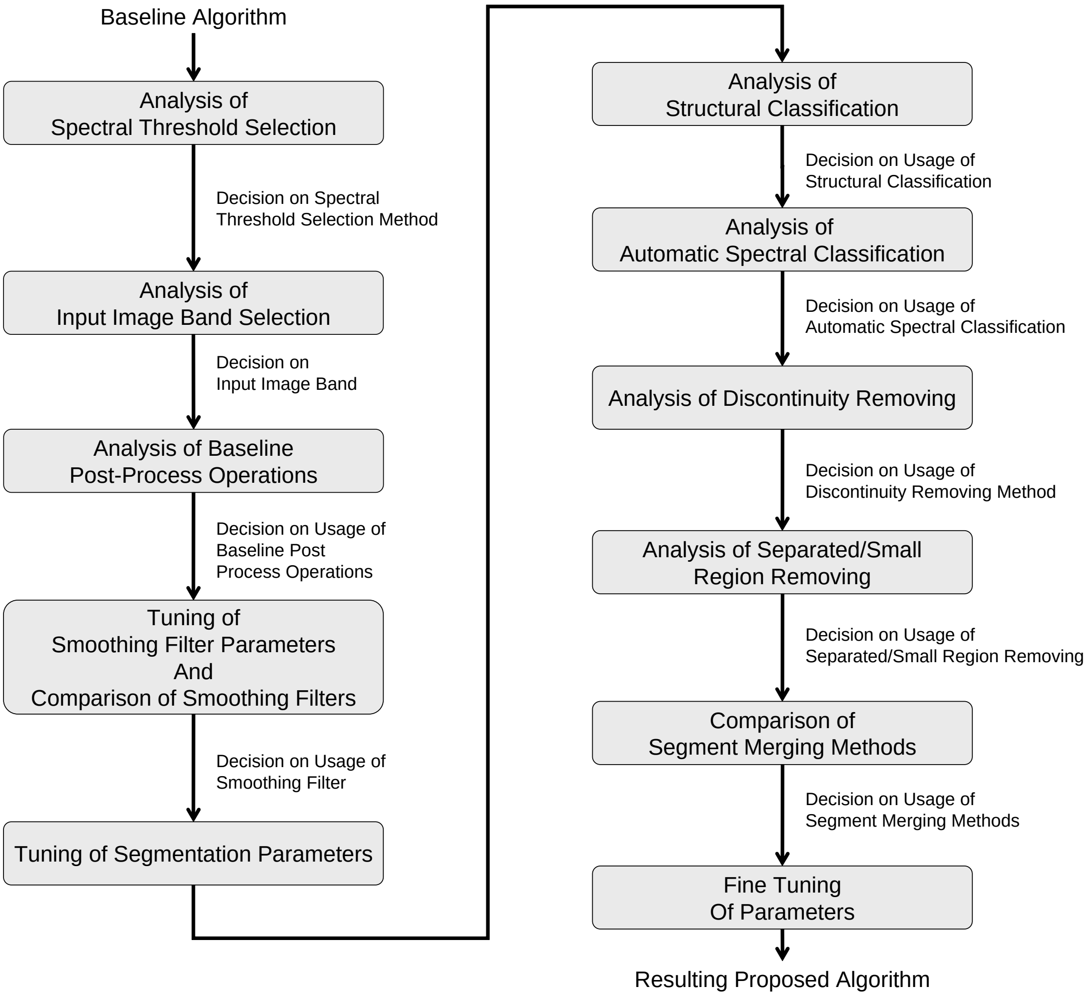
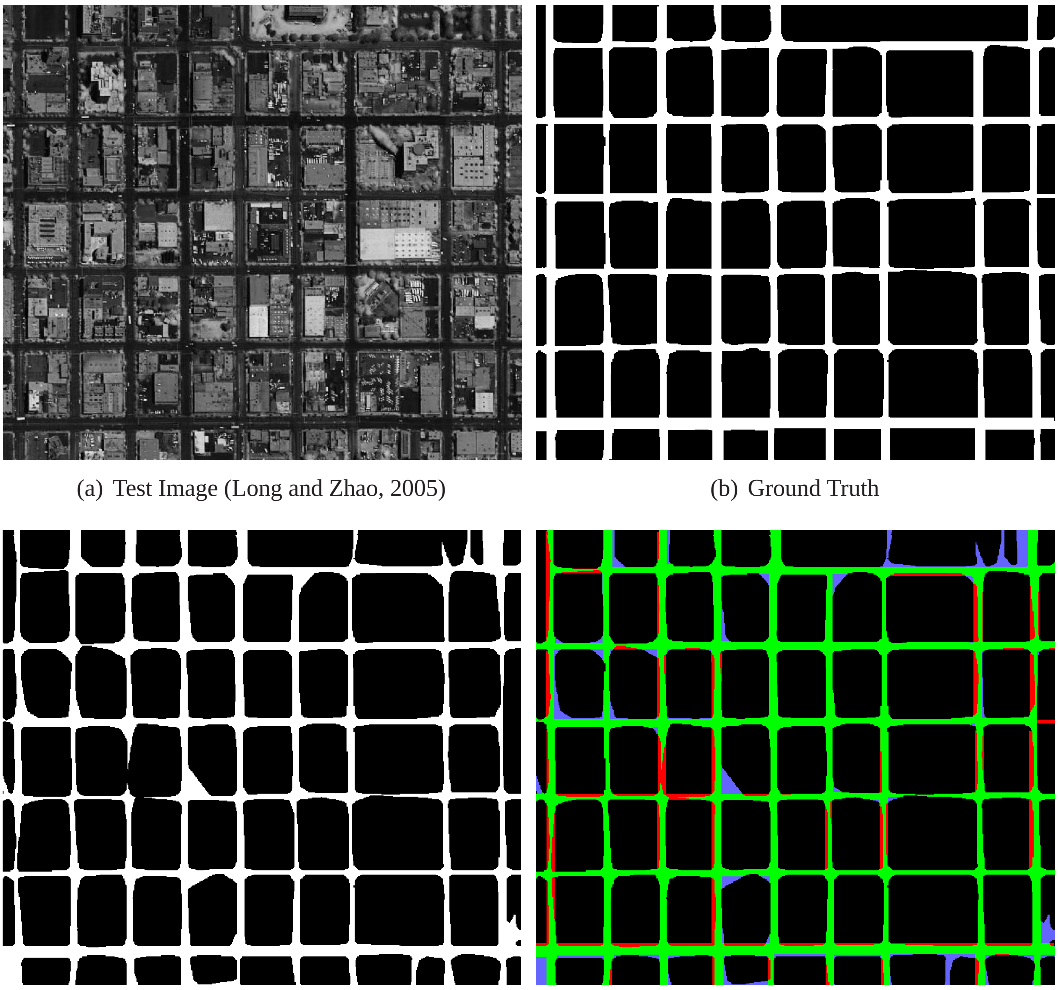


Figure 4.3: Experimental Work Flow

The baseline algorithm is applied as described in (Long and Zhao, 2005). First of all, by investigating the intermediate images and performance results, selection of spectral thresholds of classification step is changed. Then usage of 3-band(RGB) images is compared with usage of 1-band(red) images. In the next step, it is seen that region-based post-processing operations are decreasing the overall performance of the algorithm, so performance result comparison is done between region-based post-processing operations and no post processing operation case. Then parameter tuning is performed for 4 different smoothing filter and a comparison is done between those filters. The filter whose performance results are the best, is chosen as the new smoothing filter. Then parameter tuning is performed for the parameters of segmentation operation. After these parameter tuning operations, automatic spectral classification operation is used instead of classification with manually extracted spectral thresholds. Then effect of different segment-merging operations are investigated. Then an analysis is performed by investigating the effects of structural verification step. Parameter tuning is also performed for the parameters of this step. And at the end, effects of removal of discontinuities and small-separated regions is investigated.

4.5 Performance of Baseline Algorithm

Ground truth of the test image that is used in baseline algorithm is extracted manually and the resulting road mask is extracted from the road edge map that is presented in (Long and Zhao, 2005). Performance of the baseline algorithm on their test image is given in Figure 4.4 below.



(a) Test Image (Long and Zhao, 2005)

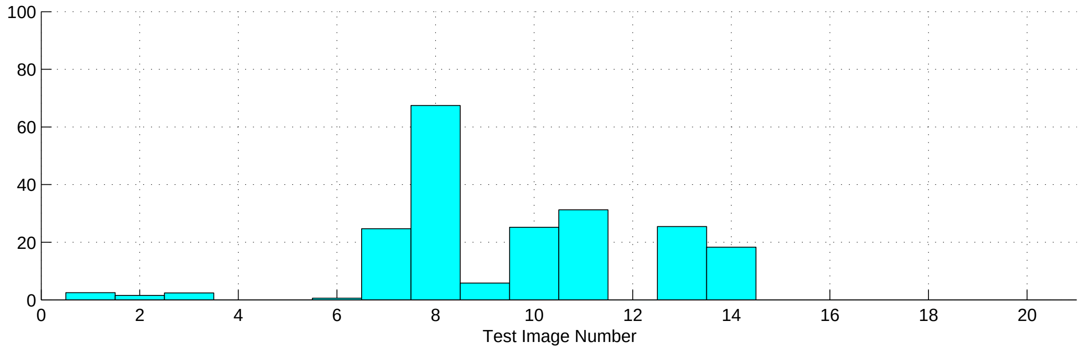
(b) Ground Truth

(c) Final Road Mask

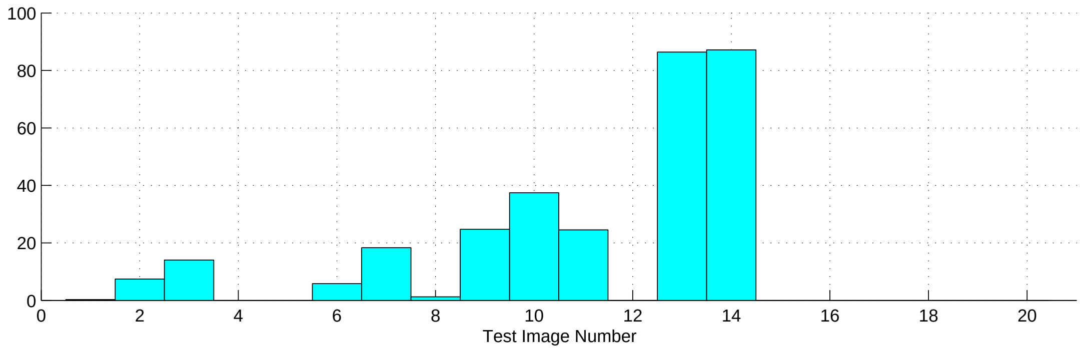
(d) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP) - Precision = %90.6, Recall = %84.8, Quality Factor = %77.9

Figure 4.4: Performance of Baseline Algorithm on its Test Image

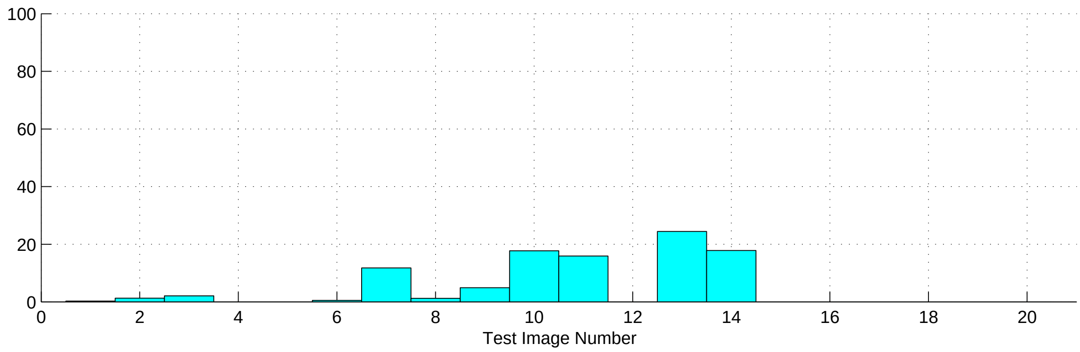
As it can be seen from the Figure 4.4, performance of the baseline algorithm on the test image that Long and Zhao(2005) used, is very good. Baseline algorithm is also applied to the test images by using the parameters that are presented in (Long and Zhao, 2005). Performance measurements are given in Figure 4.5 and mean and standard deviation of these graphics are given in Table 4.1.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.5: Performance Results for the Baseline Algorithm

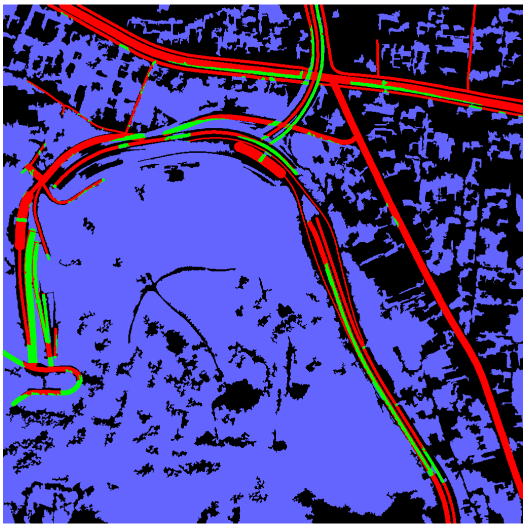
Table 4.1: Performance Results for the Baseline Algorithm

Applied Algorithm	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Baseline	10.25	17.32	15.37	26.73	4.90	7.86

It is seen that performance of the baseline algorithm was poor for most of the test images. Major reason of poor performance is the constant spectral thresholds used in classification step. The thresholds that proposed by Long and Zhao(2005) are not suitable for the data set used in this thesis. In Figure 4.6, output of the classification step of the baseline algorithm for the test image-2 and ground truth image(a) are given. It can be seen that the thresholds proposed by Long and Zhao(2005) are not suitable for the test image-2. This is the case for most of the test images.

(a) Ground Truth

(b) Classified Image

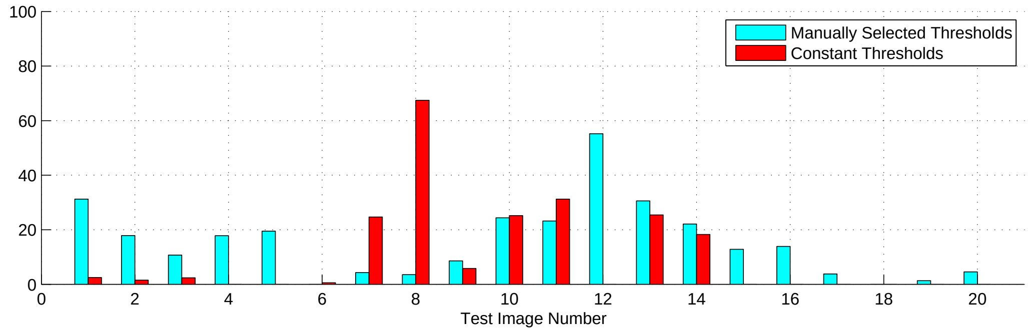


(c) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)

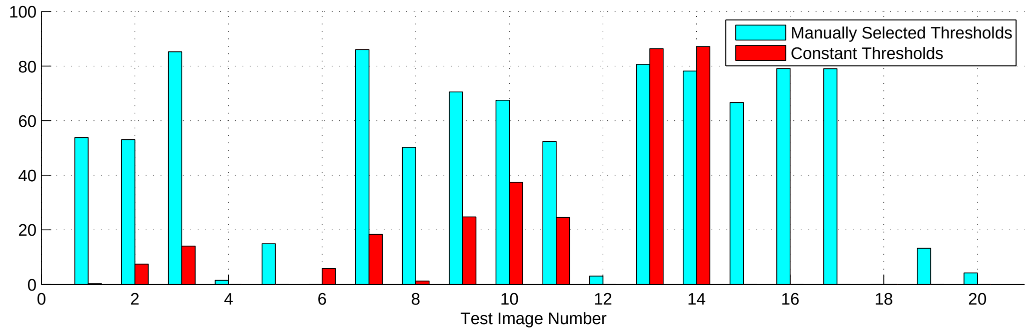
Figure 4.6: Effect of Constant Spectral Threshold Selection

4.6 Analysis of Spectral Threshold Selection

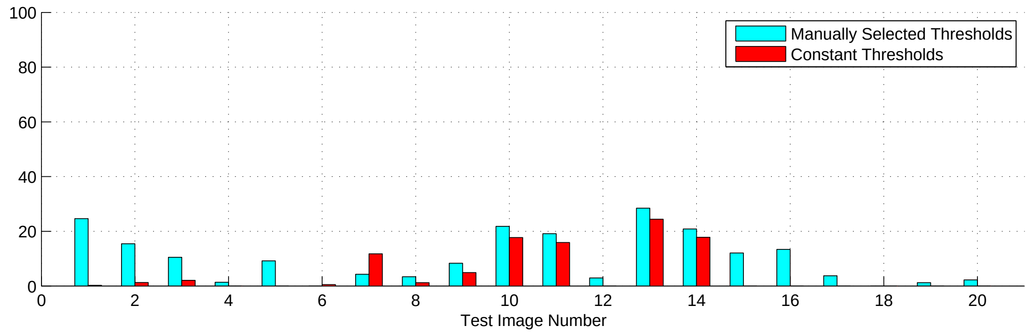
As it is explained in previous section, usage of constant spectral threshold values in classification step is not suitable for the data set. For this reason, instead of using constant spectral threshold values, spectral threshold values of roads in all test images are extracted manually and the baseline algorithm is executed using these new coefficients. In Figure 4.7, comparison of the performance results of two cases are shown. In Table 4.2 mean and standard deviation of the graphics in the Figure 4.7 are given. It can be seen that using manually selected spectral coefficient values is increased the overall performance of the algorithm.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.7: Comparison of Spectral Threshold Selection

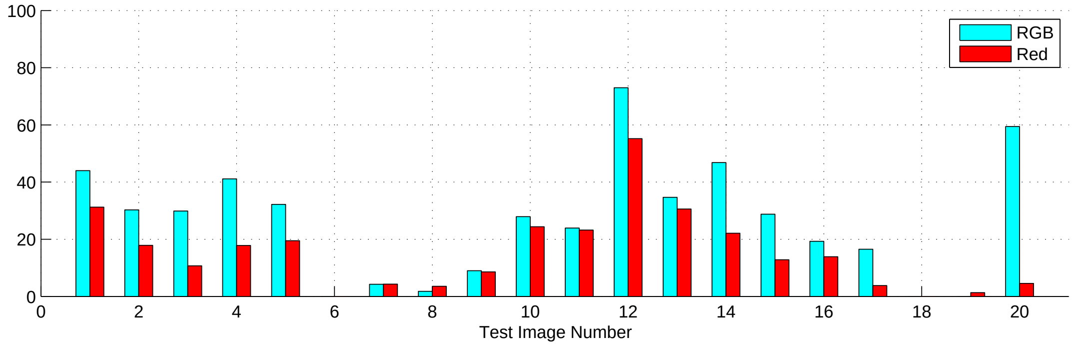
Table 4.2: Performance Comparison of Spectral Threshold Selection

Applied Thresholds	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Manual	15.27	13.65	46.97	33.29	10.16	8.95
Constant	10.25	17.32	15.37	26.73	4.90	7.86

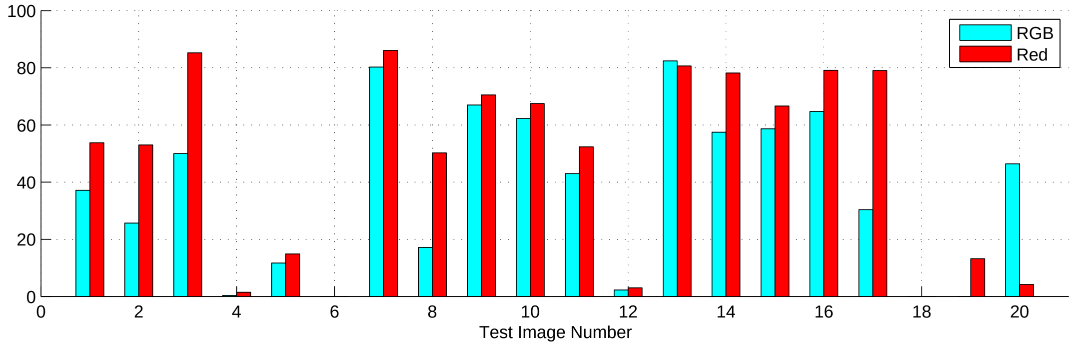
Manual extraction of spectral threshold values is not an appropriate and efficient method for an automated algorithm. However in the following steps of the experimental work flow different methods will be compared with each other in different steps. Constant spectral threshold values suggested in baseline algorithm, are not suitable with the data set used in this thesis and if these threshold values are used, then it will not be efficient to compare different methods in following operations since bad effect of these thresholds will be dominant in the final image. For this reason, instead of constant spectral thresholds manually selected spectral thresholds will be used in the following steps. Then in the step of automatic spectral classification, spectral threshold values will be determined automatically. Briefly, manually extracted spectral thresholds will be used temporarily in order to be able to continue on experimental work flow.

4.7 Analysis of Input Band Selection

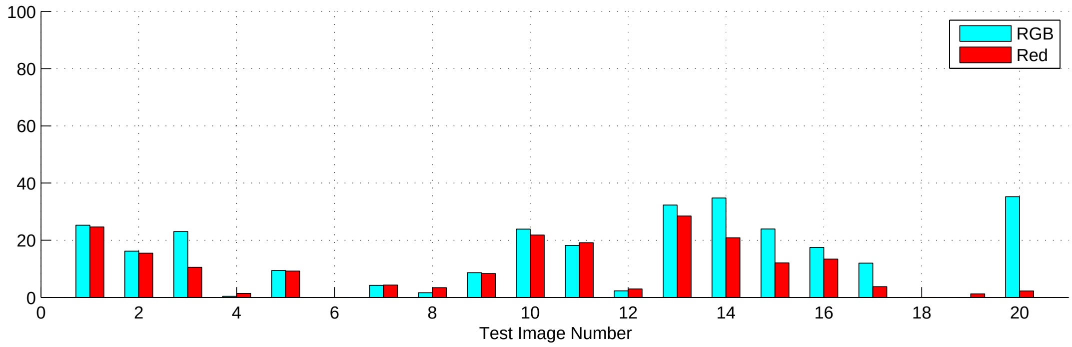
Recent satellite images are generally multi-spectral images. For this reason, instead of using red band of test images as in the baseline algorithm, RGB data can be used in the algorithm. In Figure 4.8, comparison of the performance results of two cases are shown. In Table 4.3 mean and standard deviation of the graphics in the Figure 4.8 are given. It can be seen that using three-band data is increased the overall performance of the algorithm. Hence in the following steps of experimental work flow RGB data will be used as input image.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.8: Comparison of Input Band Selection

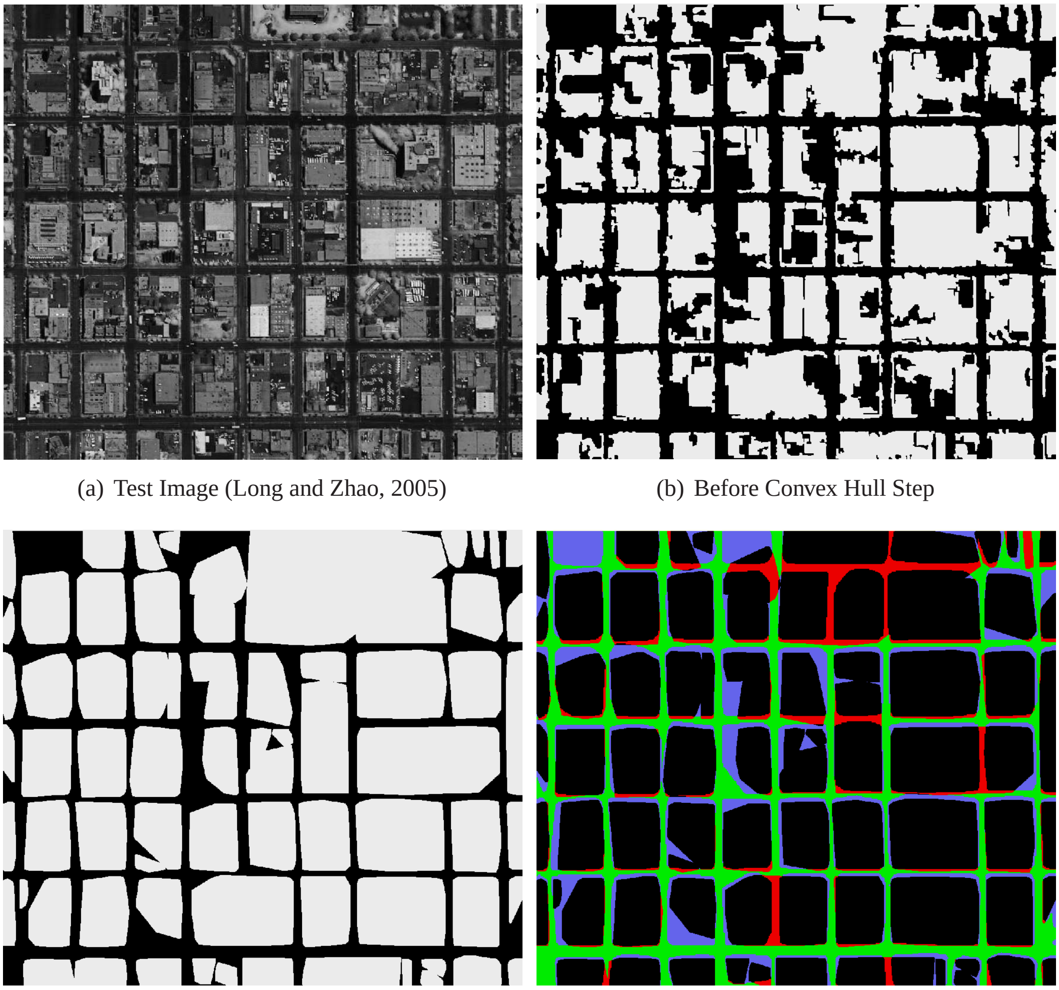
Table 4.3: Performance Comparison of Input Band Selection

Applied Bands	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
RGB	26.14	20.50	36.86	28.36	14.43	12.28
Red	15.27	13.65	46.97	33.29	10.16	8.95

4.8 Analysis of Baseline Post-Process Operations

There are four steps in the post-process operations that are applied in the baseline algorithm. These are binary opening, removal of small regions, contour-tracing and convex hull step. In this thesis study, these four steps are called as region-based post-processing since all of the operations are carried on the separated regions that exist on the classified image. In convex hull step, separated regions are labeled and concave-shaped regions are converted to convex shapes.

As it can be seen in the Figure 4.9, non-road regions of test image used in (Long and Zhao, 2005), are small regions which are not connected to each other as a result of grid type structure of roads in that particular city. In the Figure 4.9, test image, input and output of the convex hull step of the baseline algorithm and comparison of the result with ground truth are given.



(a) Test Image (Long and Zhao, 2005)

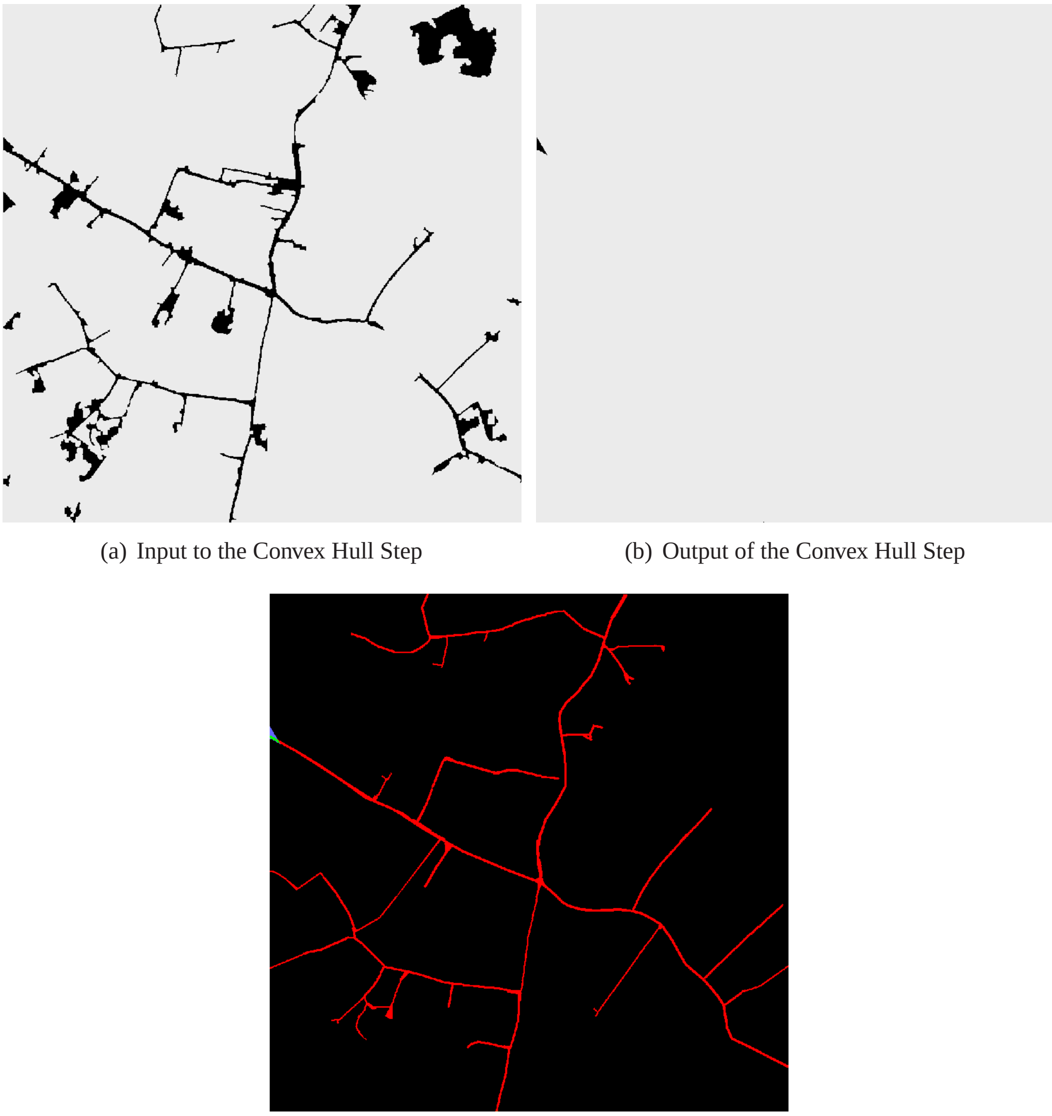
(b) Before Convex Hull Step

(c) After Convex Hull Step

(d) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)

Figure 4.9: Some Steps of Baseline Algorithm

In that particular city, non-road regions are in rectangular shape which can be thought as convex, and applying convex hull method removes only zig-zags on the edges of these regions. However this is not the case for rural areas and even for every urban area. Although the improvements that explained in above two sections, performance results for the test images 4, 6, 18 and 19, are still too low as it can be seen from the Figure 4.8. It is seen that applying region-based convex-hull step to the classified images is the reason of low performance results. In the Figure 4.10 input and output of the convex hull step of the baseline algorithm and comparison with the ground truth for the 4^{th} test image are shown.



(a) Input to the Convex Hull Step

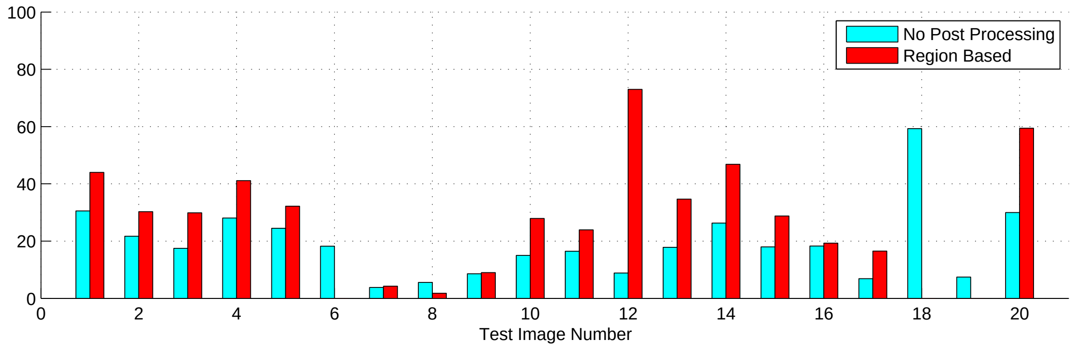
(b) Output of the Convex Hull Step

(c) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)

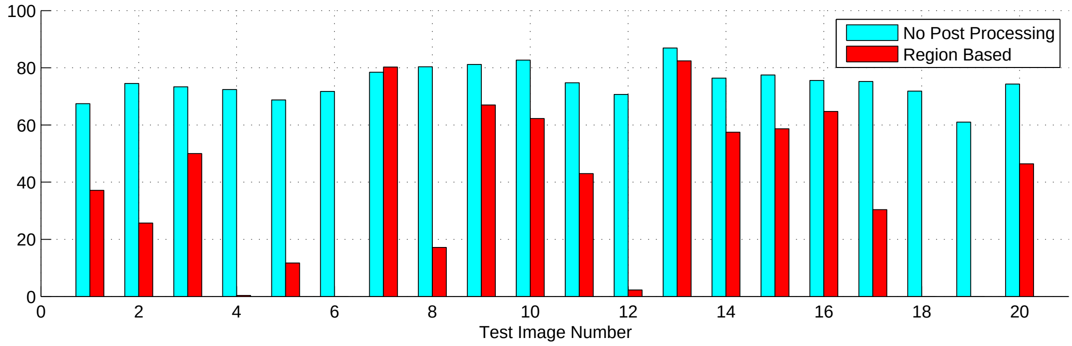
Figure 4.10: Effect of the Convex Hull Step

Generally roads have curve-shaped structures in some urban areas, in other words, non-road regions are concave regions rather than convex so applying convex hull method leads to overflow of non-road regions on to the road regions or vice versa. As it can be seen in the Figure 4.10, non-road regions are connected to each other and are covering all of the spatial domain so applying convex hull step is removing all of the road pixels that are existing between non-road regions.

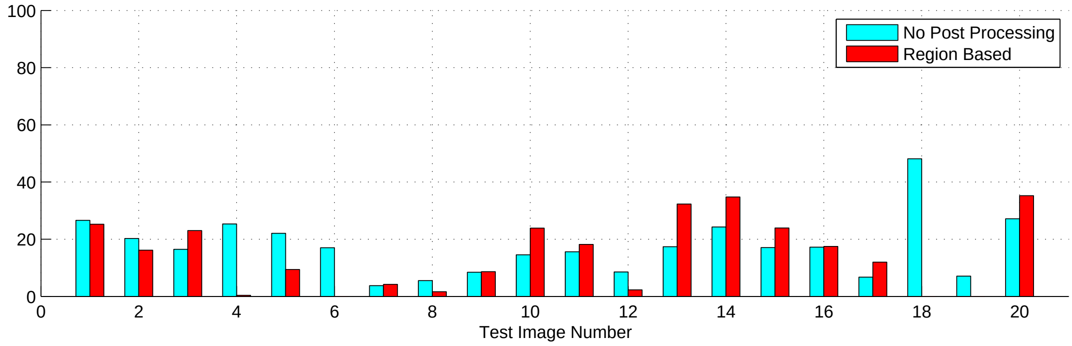
In Figure 4.11, comparison of the performance results of two cases are shown. In Table 4.4 mean and standard deviation of the graphics in the Figure 4.11 are given.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.11: Effect of Region-based Post Process Operations

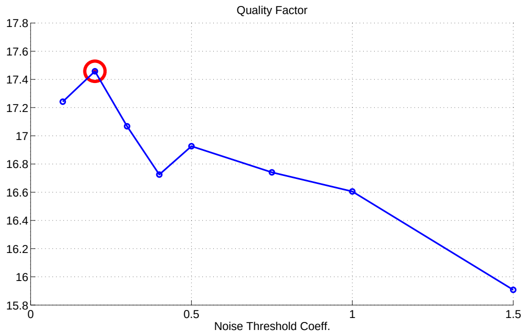
Table 4.4: Performance Comparison of Post Process Operations

Used Image Bands	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
No Post Process	19.14	12.52	74.75	5.75	17.46	10.21
Region-based	26.14	20.5	36.86	28.36	14.43	12.28

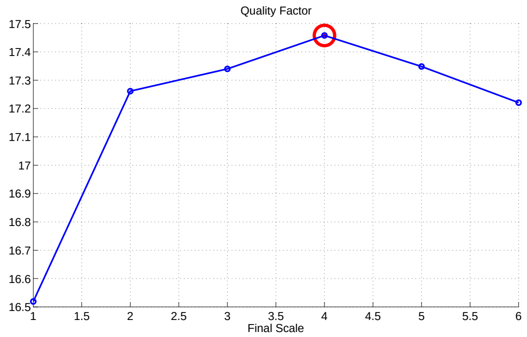
It is revealed from the results that region-based post processing operations that are used in (Long and Zhao, 2005), are decreasing the overall performance. The aim of the post-processing operations that are used in baseline algorithm, is to process separate regions on the binary classified image and four consequent operations that is containing convex hull step are all considered together. As a result of the comparison given above, it is decided to do not use the region-based post processing operations. It means that **No Post-Processing Operations** alternative will be used in the following steps of experimental work-flow. Effects of other post-processing alternatives will be analyzed after the analysis of smoothing filters.

4.9 Analysis of Smoothing Filters

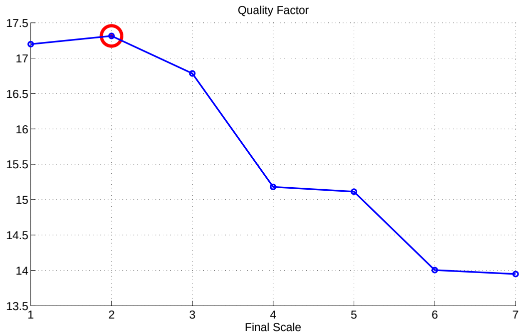
Up to this step, image enhancement filter MMCSA-LZ is used with the parameters that are suggested by Long and Zhao(2005). In this step, tuning of parameters of MMCSA-LZ, MMCSA-C and bilateral filtering is performed. Then performance comparison is realized with tuned parameters. Quality factor is proportional to both precision and recall and performance of a step is optimum at the point where quality factor is maximum so only quality factor results are given here. In the figure below tuning of parameters of MMCSA-LZ, MMCSA-C and bilateral filtering are shown.



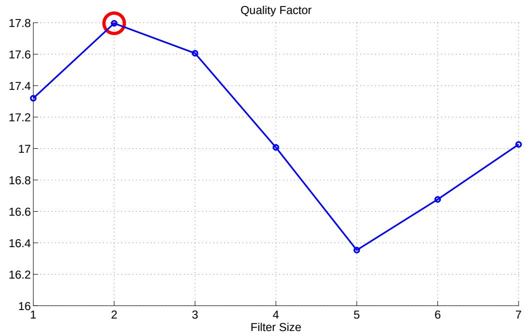
(a) MMCSA-LZ - Noise Threshold Coeff.



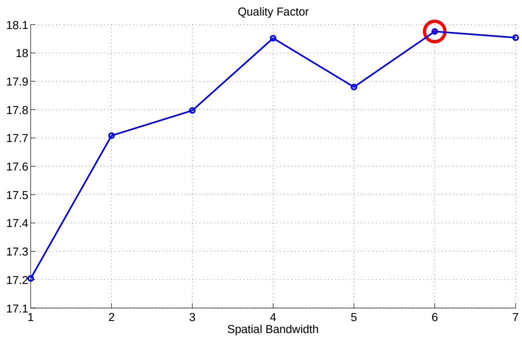
(b) MMCSA-LZ - Final Scale



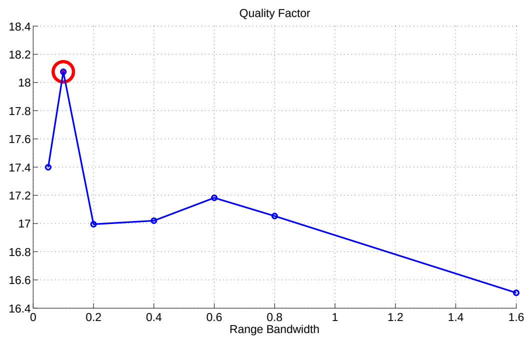
(c) MMCSA-C - Final Scale



(d) Bilateral Filter - Filter Size



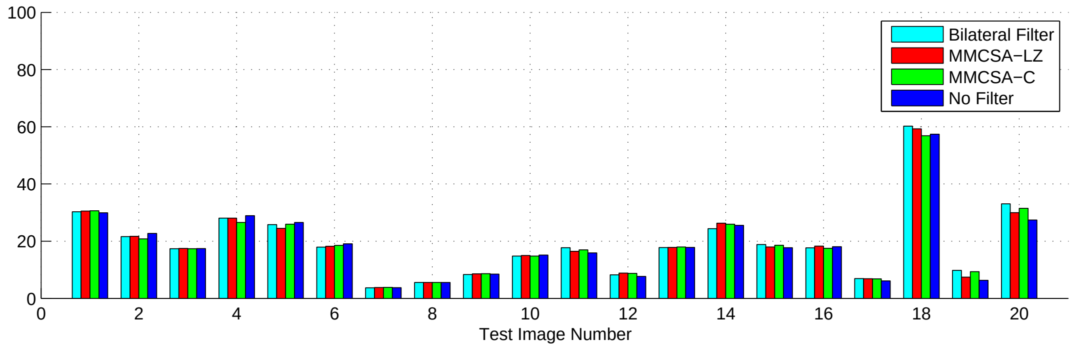
(e) Bilateral Filter - Spatial Bandwidth



(f) Bilateral Filter - Range Bandwidth

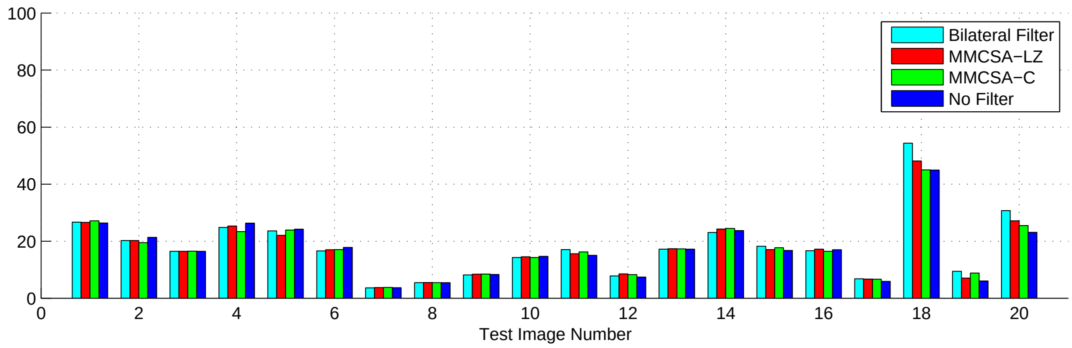
Figure 4.12: Parameter Tuning of Edge Preserving Filters

All possible combinations are used for the parameters in order to decide on optimum values. The values that are resulted in maximum performance are chosen as optimum value of related parameter. These points are marked with a big circle in Figure 4.12. After the decision of optimum parameters for edge preserving filters effect of four different case on overall performance compared. In Figure 4.13, comparison of the performance results of four cases are shown. In Table 4.5 mean and standard deviation of the graphics in the Figure 4.13 are given.



(a) Precision

(b) Recall



(c) Quality Factor

Figure 4.13: Comparison of Edge Preserving Filters

Table 4.5: Performance Comparison of Edge Preserving Filters

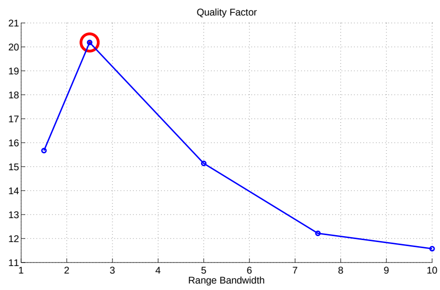
Used Image Bands	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Bilateral	19.41	12.74	76.83	6.58	18.08	11.36
MMCSA-LZ	19.14	12.52	74.75	5.75	17.46	10.21
MMCSA-C	19.15	12.09	74.26	7.57	17.31	9.58
No Filter	18.88	12.33	73.36	6.60	17.10	9.84

It can be seen from the Table 4.5 that bilateral filtering is the best among the others. So instead of MMCSA-LZ; bilateral filtering will be used as enhancement filter in the following steps of the experimental work flow.

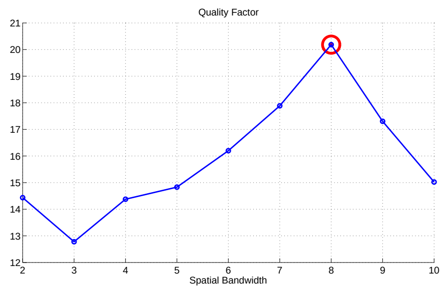
The algorithm at the second row of the Table 4.5 is modified version of the baseline algorithm. These modifications are explained in previous sections. This version of the baseline algorithm can be named as **Baseline-fitted**, since modifications are applied in order to fit the baseline algorithm to the data set used in this thesis.

4.10 Tuning of Segmentation Parameters

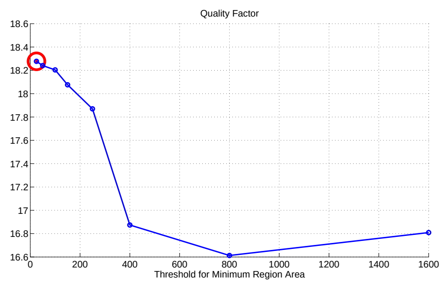
In this step, optimum values for the parameters of mean-shift segmentation step are determined by realizing parameter tuning. In the Figure 4.14, parameter tuning of optimum values of three parameters of mean-shift segmentation are shown. All possible combinations are used for the parameters in order to decide on optimum values. The values that are resulted in maximum performance are chosen as optimum value of related parameter. These points are marked with a big circle in Figure 4.14.



(a) Range Bandwidth



(b) Spatial Bandwith

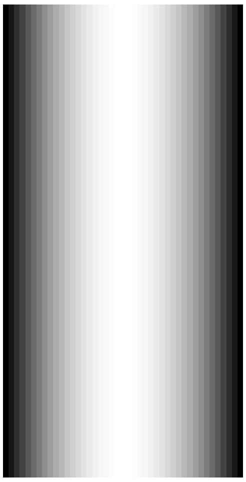


(c) Minimum Region

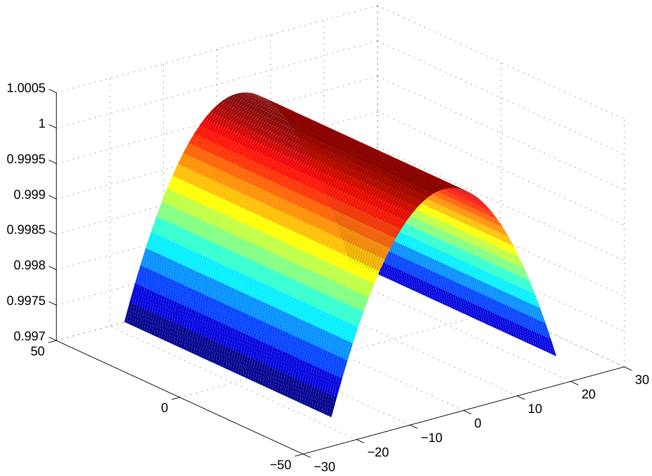
Figure 4.14: Tuning on Segmentation Parameters

4.11 Analysis of Structural Verification

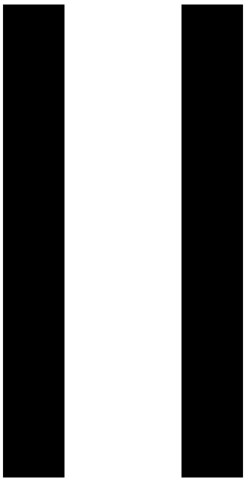
In some of the test images, although recall measurements were good enough, it is seen that precision measurements were too low. Especially, precision measurement of test images 7, 8, 9, 12, 17 and 19 were below 20%. This poor performance is due to the spectral similarity of road pixels with non-road pixels. As a result, most of the pixels of the image are classified as road when a spectral classification is applied. In order to decrease false positives, structural verification is performed on the spectrally classified image. As it is explained in section 3.6, multi-scale template matching filter(MS-TMF) is used as one of the structural feature extraction tool in structural verification step. In MS-TMF, a cosine-disk shaped template is used in this thesis. However different templates can also be used in template matching filter. For this purpose, four different templates are compared with each other. 2D and 3D view of the cosine-disk template is presented in section 3.3.1. Below 2-D and 3-D views of other three alternative templates which are gauss template, rectangle template and cosine template are given.



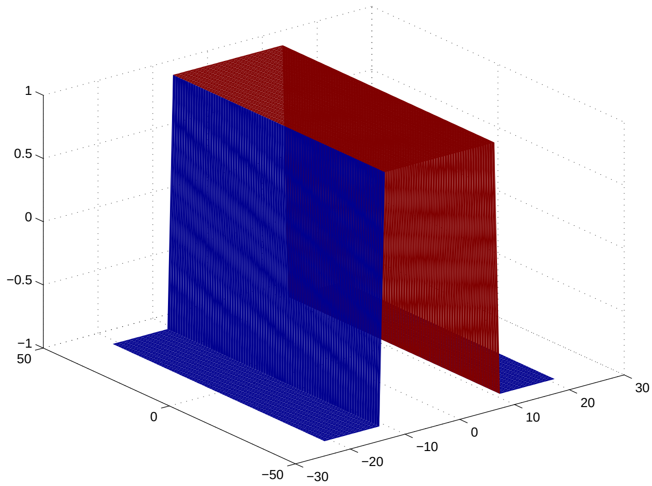
(a) 2D-View of Gauss Template



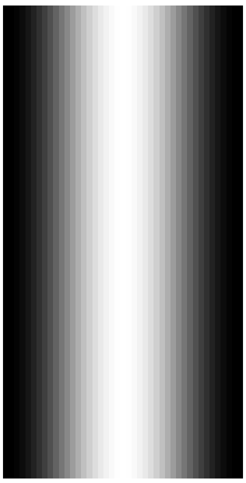
(b) 3D-View of Gauss Template



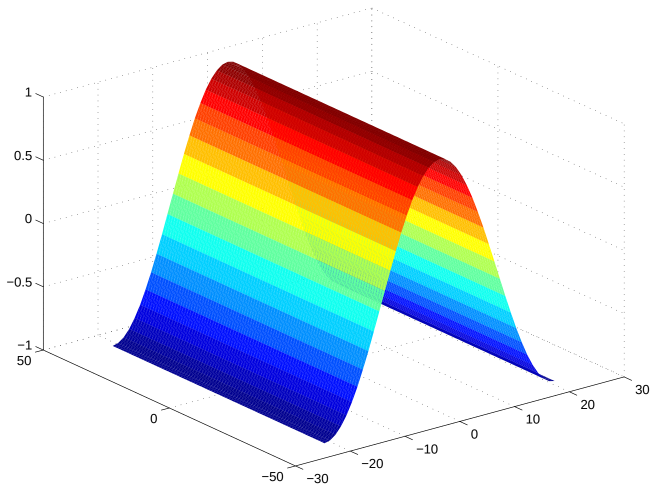
(c) 2D-View of Rectangle Template



(d) 3D-View of Rectangle Template



(e) 2D-View of Cosine Template

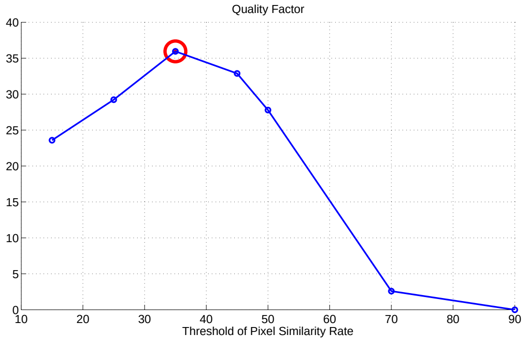


(f) 3D-View of Cosine Template

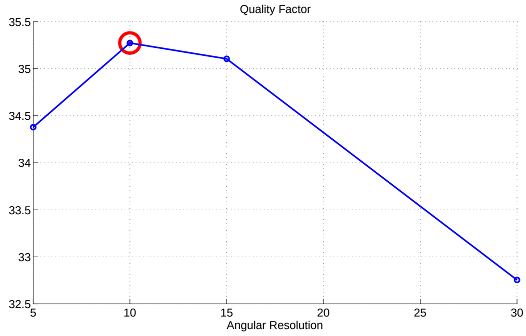
Figure 4.15: Alternative Templates for MS-TMF

4.11.1 Parameter Tuning

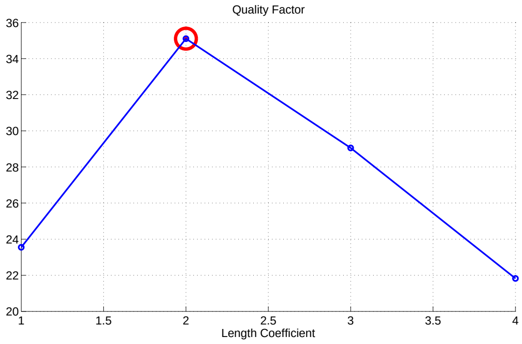
In the figure below tuning of the parameters of structural verification method, including MS-TMF parameters are given.



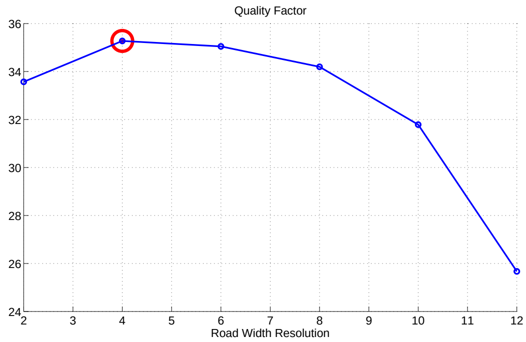
(a) TMF - Threshold for Pixel Similarity Rate



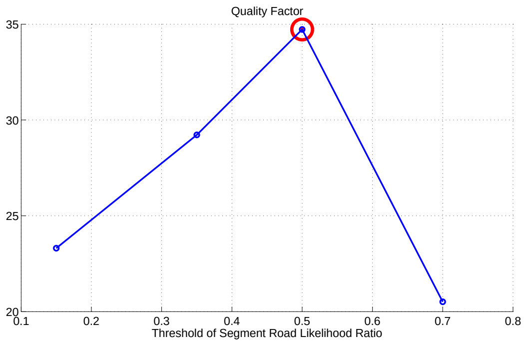
(b) TMF - Angular Resolution



(c) TMF - Length Coefficient



(d) TMF - Road Width Resolution



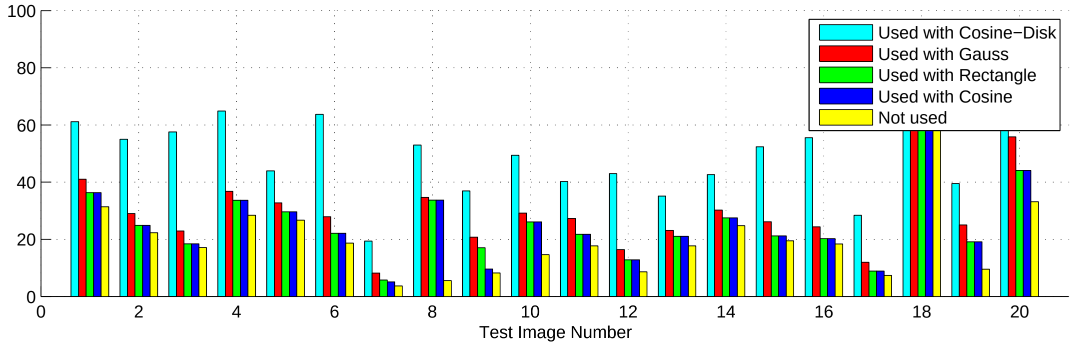
(e) Threshold for Segment Road Likelihood Ratio

(f) Threshold for Major/Minor Ratio

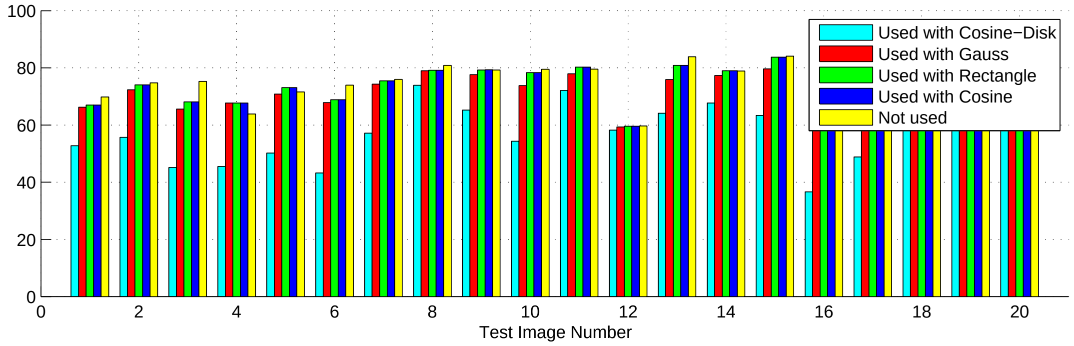
Figure 4.16: Parameter Tuning of Structural Verification Step

4.11.2 Performance Evaluation

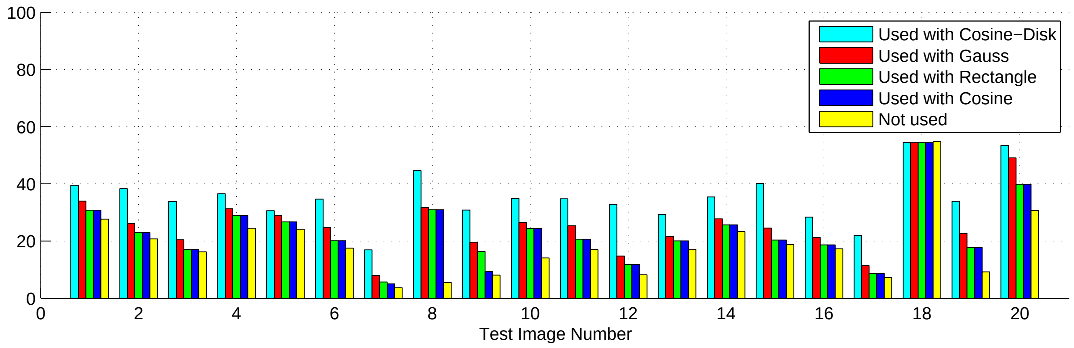
In Figure 4.17, effect of tuned structural verification method with different templates can be observed. In Table 4.6 mean and standard deviation of the graphics in the Figure 4.17 are given.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.17: Effect of Structural Verification

Table 4.6: Effect of Structural Verification

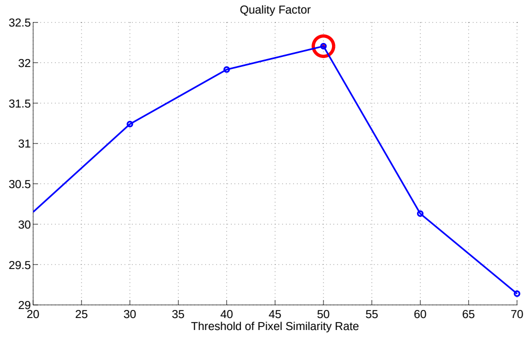
Structural Verification	Precision		Recall		Quality Factor	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Used with Cosine-Disk	48.50	12.89	58.97	12.00	35.27	8.86
Used with Gauss	29.20	12.61	72.71	6.58	26.19	10.93
Used with Rectangle	25.23	12.36	74.85	6.52	23.08	10.86
Used with Cosine	24.82	12.78	74.86	6.52	22.70	11.24
Not used	19.69	12.81	75.92	6.60	18.28	11.45

It can be seen from the Table 4.6 that structural verification is doubling the performance measurements. Hence structural verification method will be used in the following steps of the experimental work flow. Also it can be seen from the table 4.6 that the cosine-disk template is the best template among the alternative templates. Hence cosine-disk template will be used as the template of MS-TMF in the following steps of the experimental work flow.

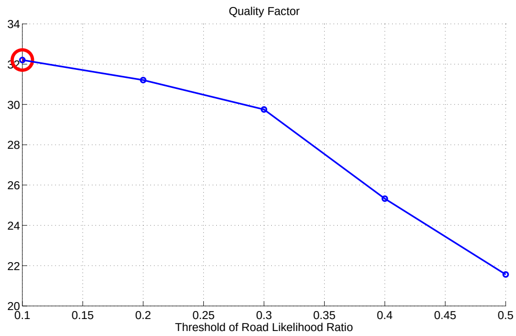
4.12 Analysis of Automatic Spectral Classification

In the first step of experimental work flow, due to the bad performance of the constant spectral thresholds used in baseline algorithm, it was decided to continue with manually extracted spectral threshold values temporarily. As it is mentioned previously, manual selection of threshold values is not a practical method, Hence, in this step, automatic spectral threshold selection method is applied in the spectral classification step of the algorithm. Parameter tuning is also performed for the parameters of proposed classification method. In the Figure 4.18, parameter tuning results can be observed.

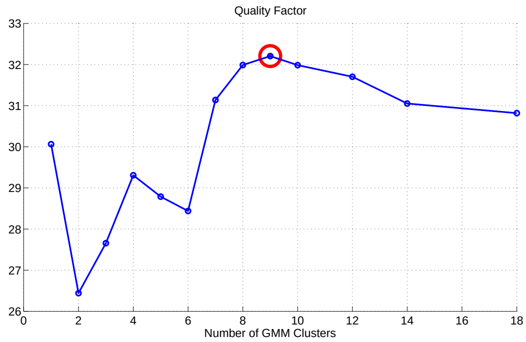
(a) Number of Seed Segments



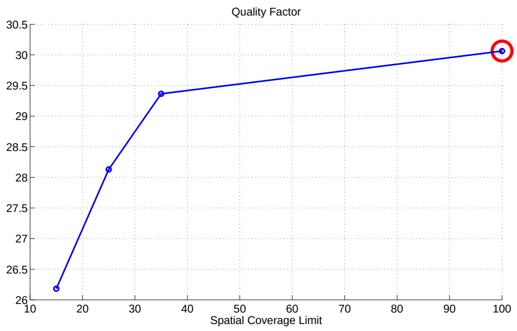
(b) Threshold for Pixel Similarity Rate



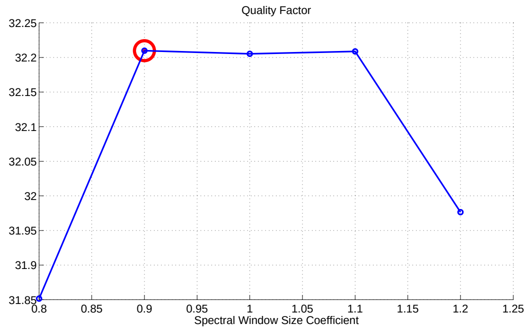
(c) Threshold for Segment Road Likelihood Ratio



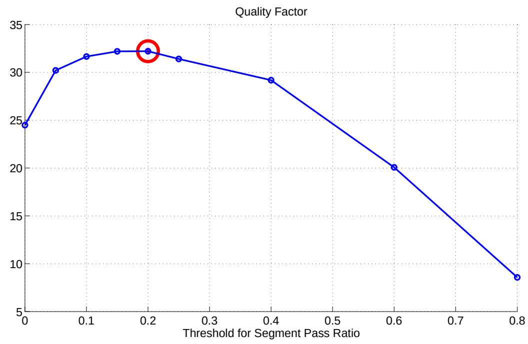
(d) Number of GMM Clusters



(e) Spatial Coverage Limit



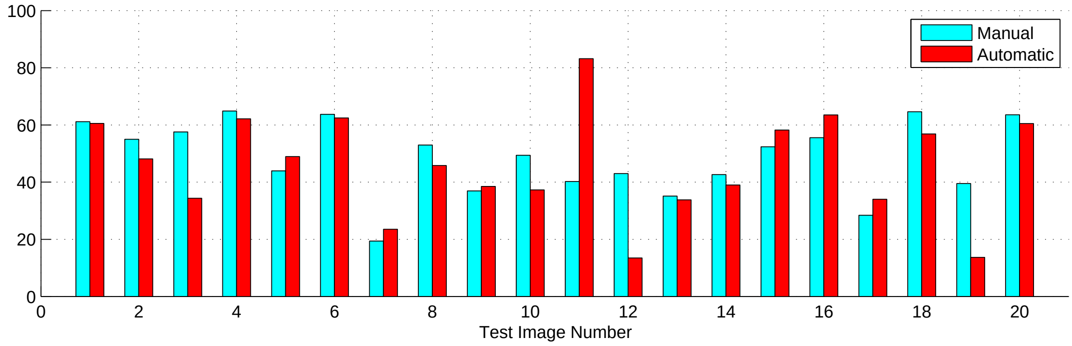
(f) Spectral Window Size Coefficient



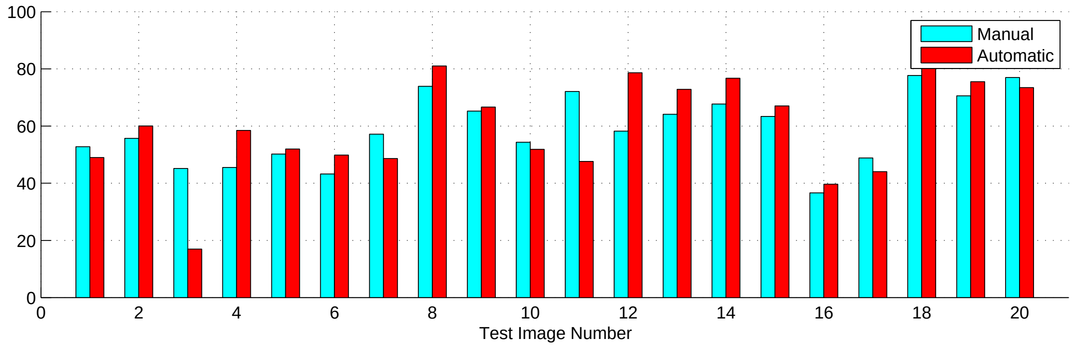
(g) Threshold for Segment Pass Ratio

Figure 4.18: Parameter Tuning of Automatic Classification Step

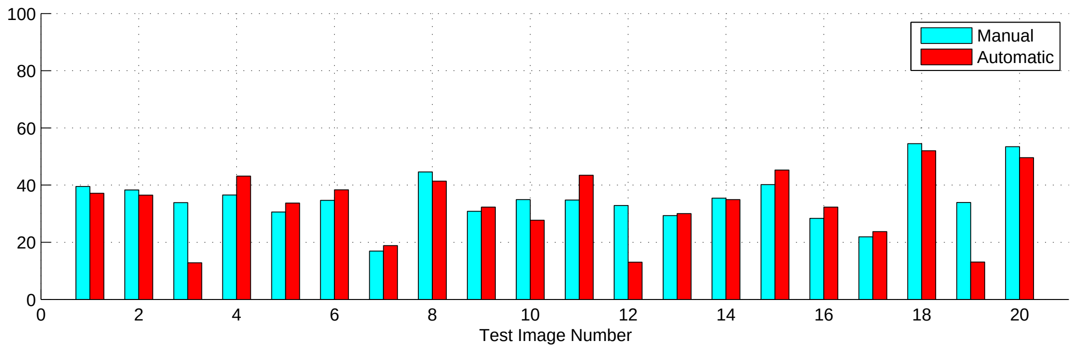
All possible combinations are used for the parameters in order to decide on optimum values. The values that are resulted in maximum performance are chosen as optimum value of related parameter. These points are marked with a big circle in Figure 4.18. In Figure 4.19, effect of automatic classification after parameter tuning can be observed. In Table 4.7 mean and standard deviation of the graphics in the Figure 4.19 are given.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.19: Performance Comparison Classification Methods

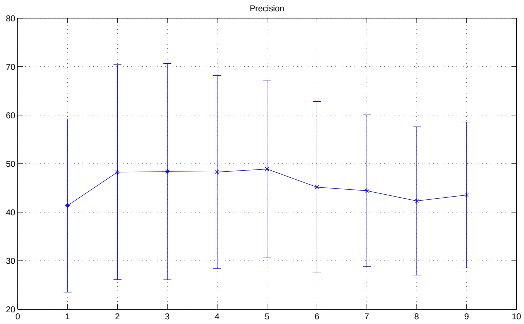
Table 4.7: Performance Comparison Classification Methods

Classification Type	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Manual	48.50	12.89	58.97	12.00	35.27	8.86
Automatic	45.90	17.97	59.80	17.13	32.96	11.83

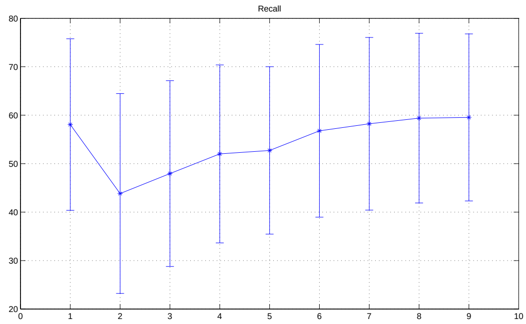
It can be seen from the Figure 4.19 and the Table 4.7 that automatic classification is not increasing overall performance but it's performance is around the performance of the manual classification. As it is mentioned previously, manual extraction of spectral threshold values is not an appropriate and efficient method for an automated algorithm. Hence automatic classification will be used in the following steps of the experimental work flow although it's performance is not as good as manual classification.

4.13 Analysis of Gaussian Mixture Model Parameters

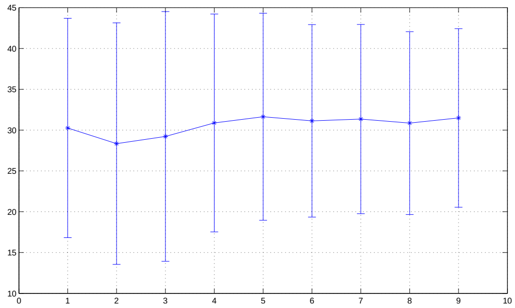
In Gaussian Mixture Model expectation maximization algorithm is used. Initial points of the algorithm are selected randomly. According to the selection of initial points different mixture of clusters can be obtained. For this purpose, GMM is trained several times and classification is performed for each case. Then it is checked whether overall performance of the road detection algorithm is being effected or not. In the Figure 4.20, horizontal axis represents the training count. For example when training count is 5, it means that GMM is trained 5 times. Resulting clusters of each training are used for classification separately. Classification is performed for all of the test images. Then mean of performance result of these 5 different case are averaged. In Figure 4.20 these average values are used. Vertical lines on the graphics are representing the standard deviation of calculated mean parameter. It can be seen from the figure that, as the training count increases standard-deviation decreases. There is a small difference between the case when training count equal to 1 and 9. Hence it can be said that training the Gaussian Mixture Model only for one time is acceptable.



(a) Precision



(b) Recall

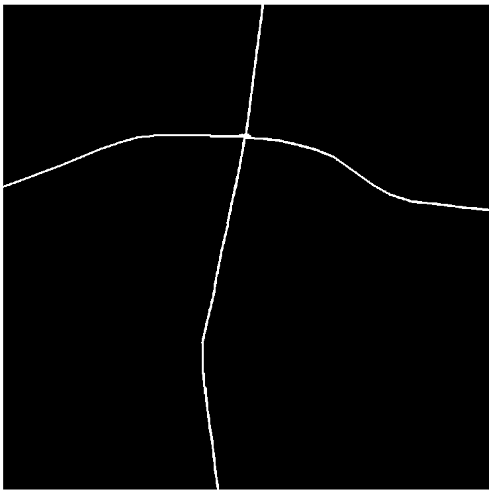


(c) Quality Factor

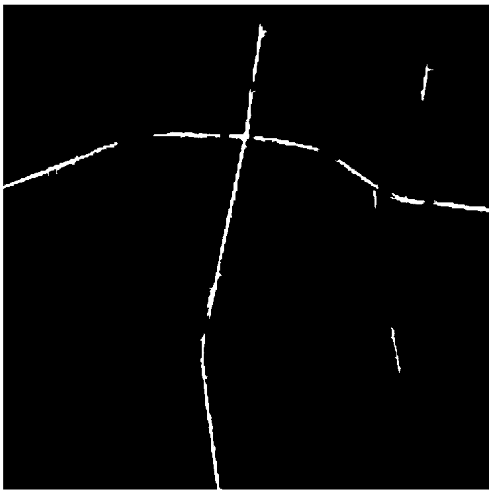
Figure 4.20: Gaussian Mixture Model Training Performance

4.14 Analysis of Discontinuity Removing Method

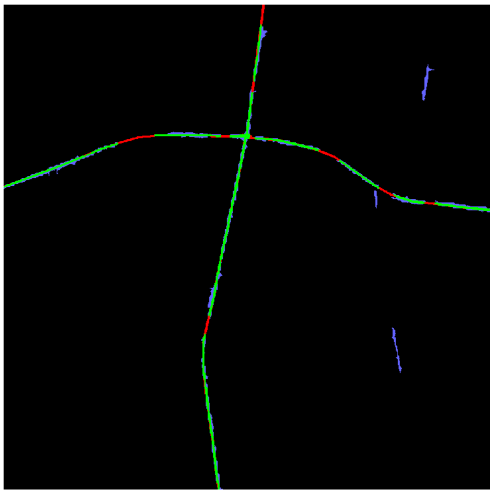
In most of the classified images it is seen that there are discontinuities on the roads. In order to remove these discontinuities, discontinuity removing method that is proposed in section 3.7.1, is used. In Figure 4.21, effect of discontinuity removing on a test image is shown.



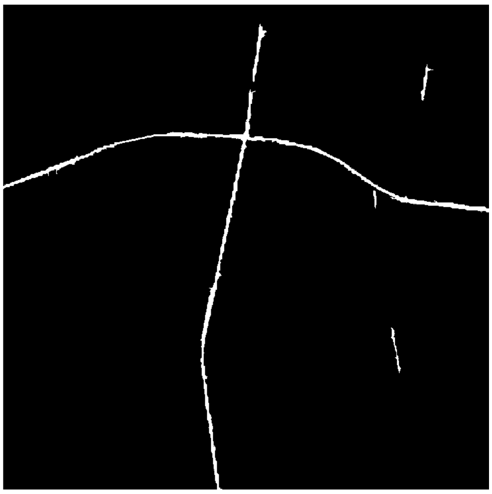
(a) Ground Truth



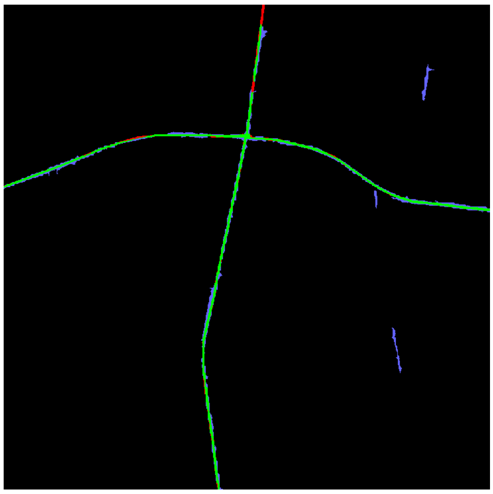
(b) Classified Image



(c) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)



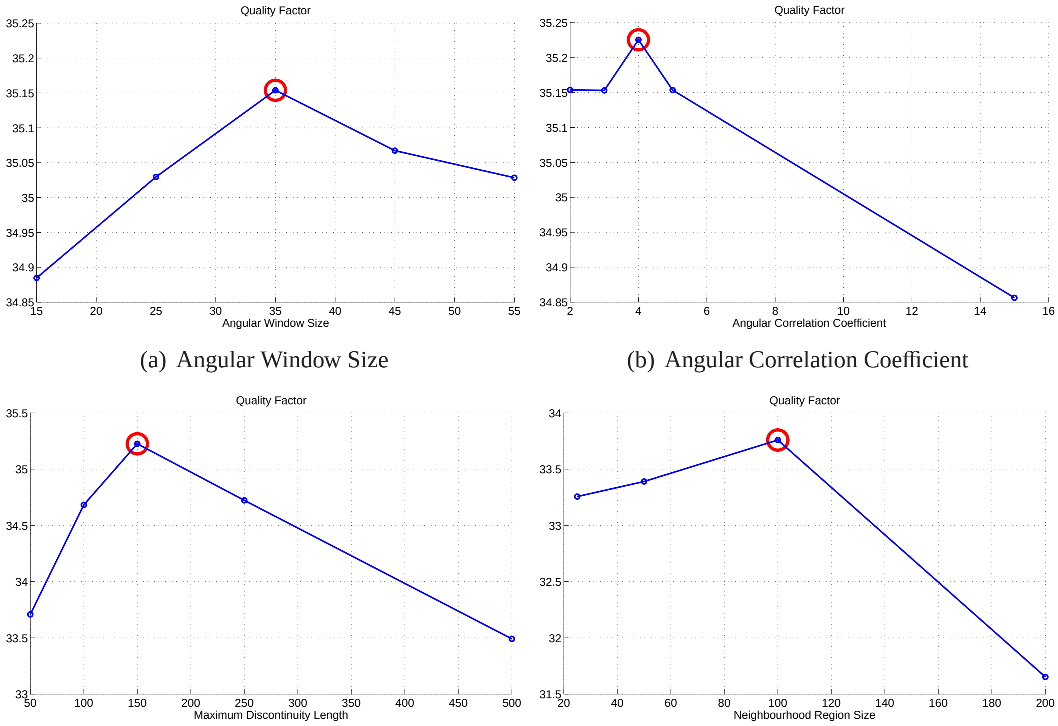
(d) Discontinuities Removed



(e) Comparison with Ground Truth after Discontinuity Removal(Red=FN, Green=TP, Blue=FP)

Figure 4.21: Effect of Discontinuity Removing

In Figure 4.22, tuning of the parameters of discontinuity removing method are given.

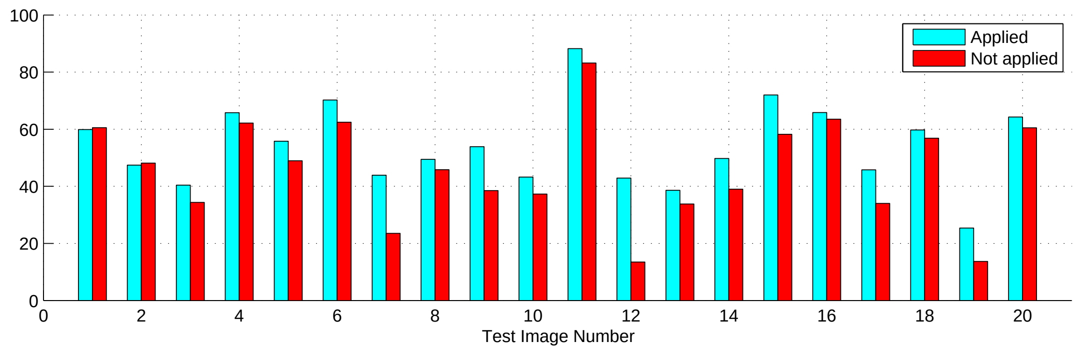


(a) Angular Window Size (b) Angular Correlation Coefficient

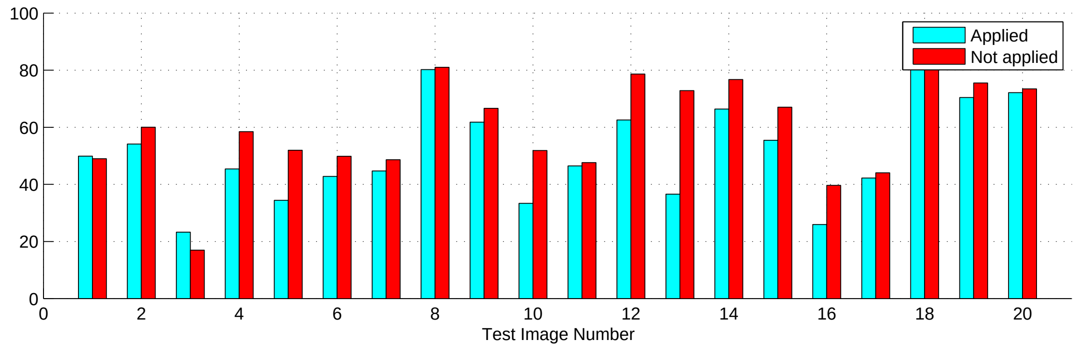
(c) Maximum Discontinuity Length (d) Neighborhood Search Region Size

Figure 4.22: Parameter Tuning of Discontinuity Removing Step

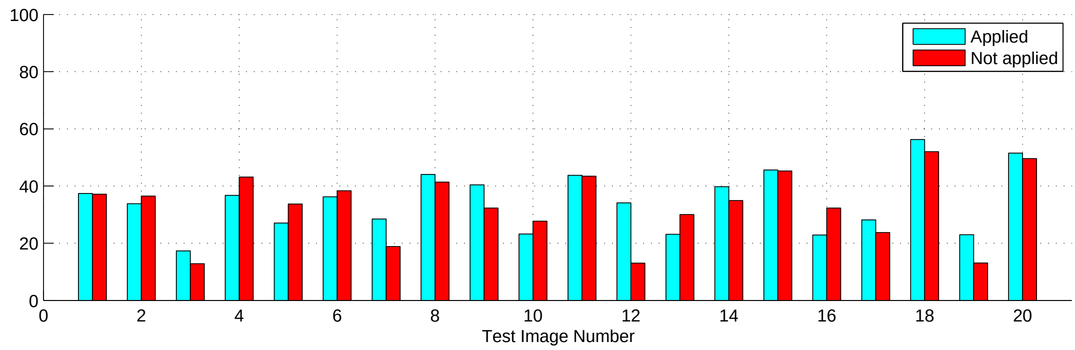
All possible combinations are used for the parameters in order to decide on optimum values. The values that are resulted in maximum performance are chosen as optimum value of related parameter. These points are marked with a big circle in Figure 4.22. In Figure 4.23, effect of discontinuity removing after parameter tuning can be observed. In Table 4.8 mean and standard deviation of the graphics in the Figure 4.23 are given.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.23: Effect of Discontinuity Removing on Overall Performance

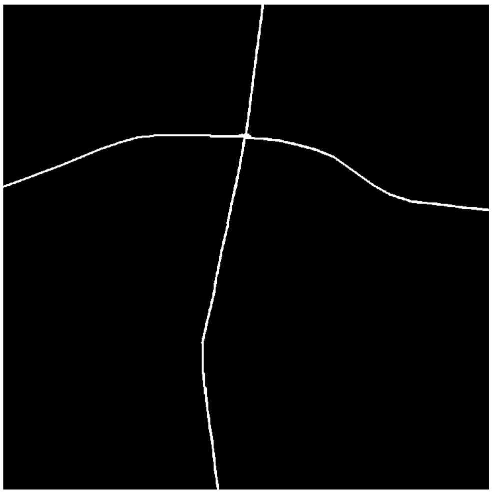
Table 4.8: Performance Comparison of Discontinuity Removing

Discontinuity Removing	Precision		Recall		Quality Factor	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Applied	54.12	14.40	51.94	18.02	34.64	10.50
Not applied	45.90	17.97	59.80	17.13	32.96	11.83

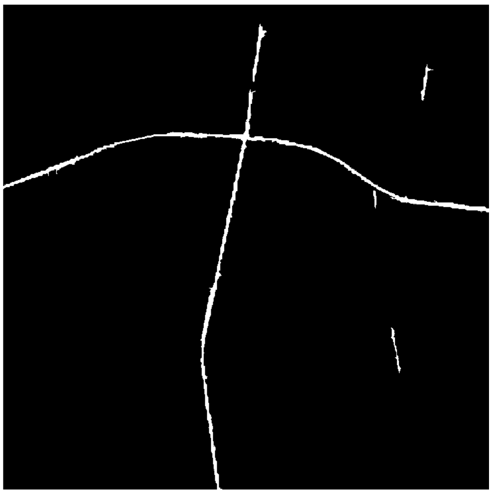
It can be seen from the Figure 4.23 and the Table 4.8 that discontinuity removing is increasing overall performance. Hence it will be used in the following steps of experimental work flow.

4.15 Analysis of Separated/Small Region Removing

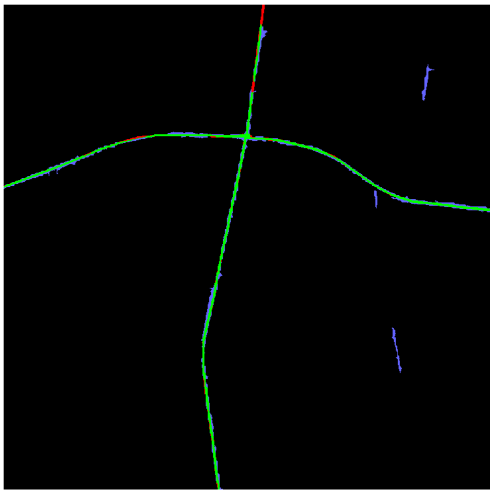
In most of the classified images it is seen that there are separated and small regions which are not road but classified as road since they are spectrally and structurally similar to the roads. In order to remove these regions, small and separated region removing step is applied. Effect of small and separated region removing step can be seen in the Figure 4.24.



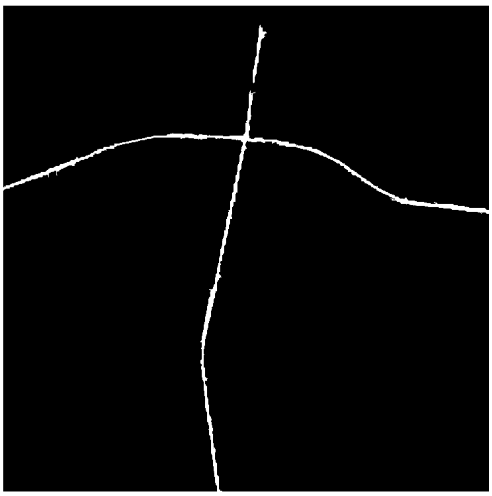
(a) Ground Truth



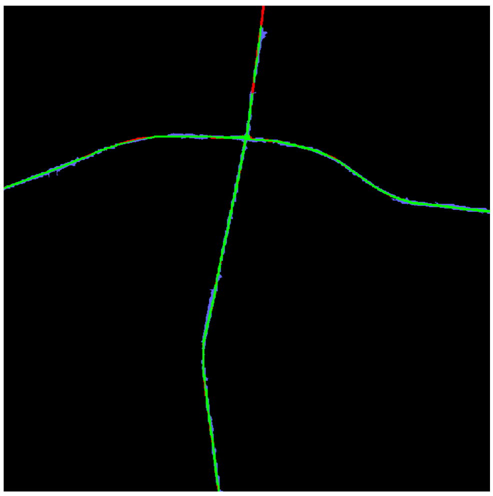
(b) Input Image



(c) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)



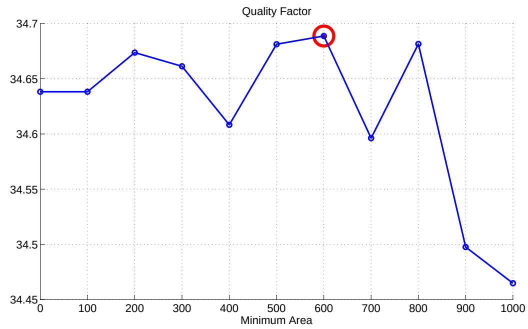
(d) Separated/Small Regions Removed



(e) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)

Figure 4.24: Effect of Separated/Small Region Removing

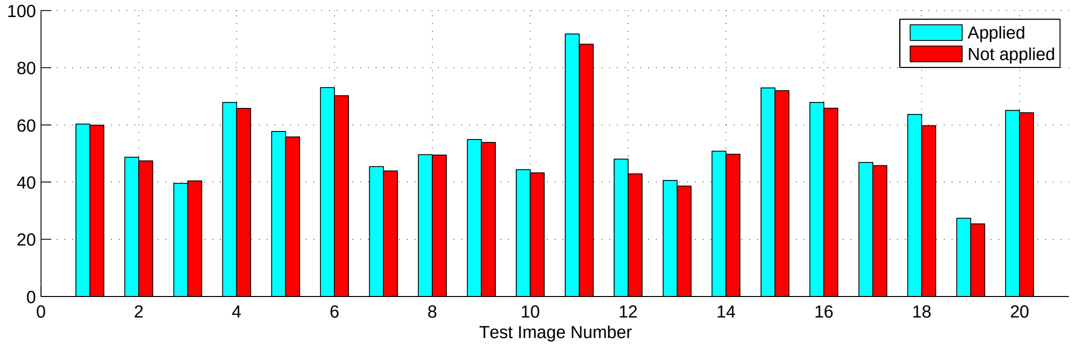
In Figure 4.22, tuning of the threshold for minimum region area of separated/small area removing method is given. The value that is resulted in maximum performance is chosen as optimum value of the parameter. These point is marked with a big circle in Figure 4.25.



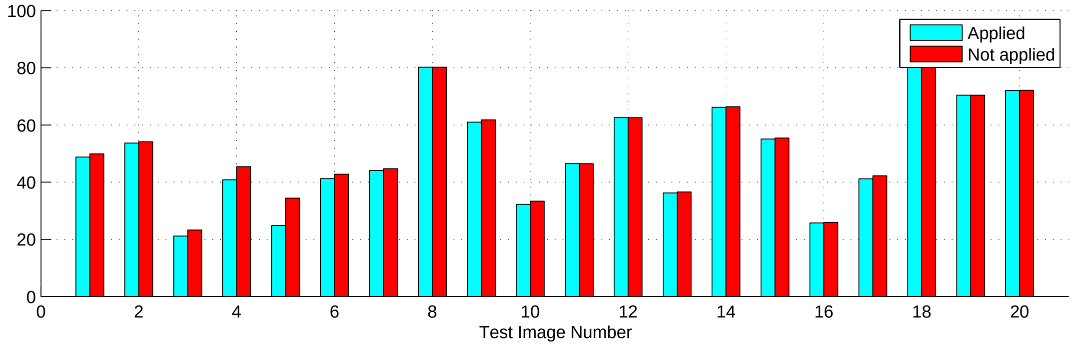
(a) Threshold for Minimum Region Area

Figure 4.25: Parameter Tuning of Separated/Small Area Removing Step

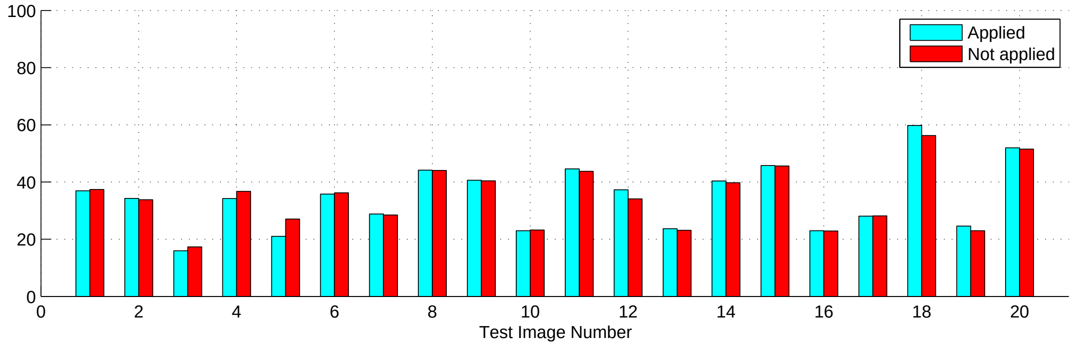
In Figure 4.26, effect of separated/small area removing after parameter tuning can be observed. In Table 4.9 mean and standard deviation of the graphics in the Figure 4.26 are given.



(a) Precision



(b) Recall



(c) Quality Factor

Figure 4.26: Effect of Separated/Small Area Removing on Overall Performance

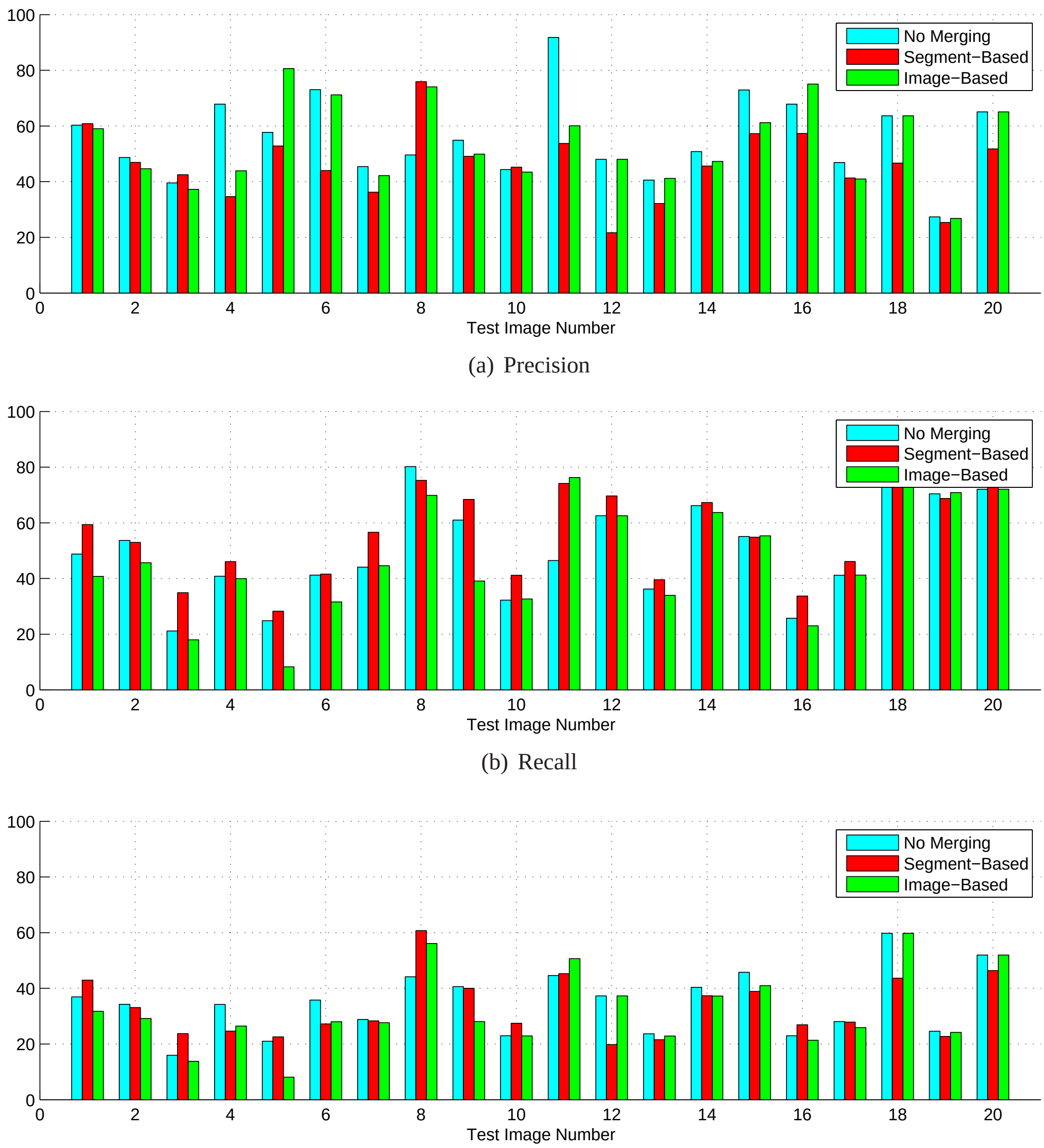
Table 4.9: Performance Comparison of Separated/Small Area Removing

Area Removing	**Precision**		**Recall**		**Quality Factor**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Applied	55.82	14.77	50.74	19.04	34.69	11.32
Not applied	54.12	14.40	51.94	18.02	34.64	10.50

It can be seen from the Figure 4.26 and 4.9 that separated/small area removing is increasing overall performance. Hence it will be used in the following steps of the experimental work flow.

4.16 Analysis of Segment Merging Methods

In automatic spectral classification step, seed segments are chosen by examining **structural features** of all segments. By using segment merging techniques before seed segment selection, it is aimed to increase road likelihood ratio of road segments. In Figure 4.27, effect of segment merging methods can be observed. In Table 4.10 mean and standard deviation of the graphics in the Figure 4.27 are given. Although segment merging operations were increasing the performance in earlier studies, at the end, it is observed that these operations are not increasing overall performance. There are two reasons for this result. Firstly, in earlier studies, instead of Gaussian Mixture Model(GMM), a different spectral classification method was being used. On the other hand, parameters of all steps were not tuned in earlier studies. In the earlier classification method, a few segments were being chosen to be as seeds and mean and standard deviation parameters of **each seed segment** were being used to classify the image where selection of seed segments were critical. However in GMM, instead of using the mean and standard deviation parameters of **each seed segment**, all of the pixel values are considered together and Expectation Maximization algorithm is applied in order to obtain **clusters** which are different than **seed segments**. Moreover the clusters whose weight is smaller than a threshold are not used in classification. Briefly, by the use of GMM and tuning of parameters, seed selection problem is resolved. As a result, segment merging operations are not applied since they are not increasing the overall performance.



(a) Precision

(b) Recall

(c) Quality Factor

Figure 4.27: Effect of Segment Merging Methods on Overall Performance

Table 4.10: Performance Comparison of Segment Merging Methods

Merging Method	Precision		Recall		Quality Factor	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
No Merge	55.82	14.77	50.74	19.04	34.69	11.32
Segment-Edge Based	46.05	12.56	56.36	17.12	33.06	10.86
Image-Edge Based	53.78	14.63	48.01	21.50	32.21	13.77

4.17 Resulting Proposed Algorithm

In this section, improved version of the baseline algorithm is proposed. The proposed algorithm is obtained by a series of experimental analysis as given in previous sections. At the end of experimental work flow proposed algorithm is obtained. The alternative method that is increased the performance is chosen as the method used for related step. Resulting proposed algorithm is given in Figure 4.28 below.

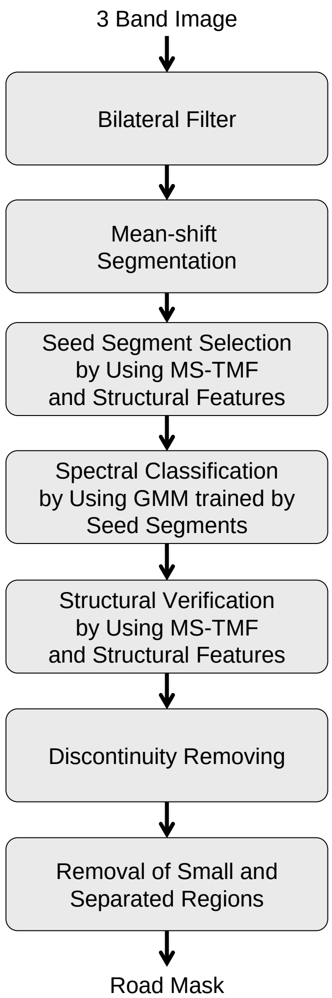
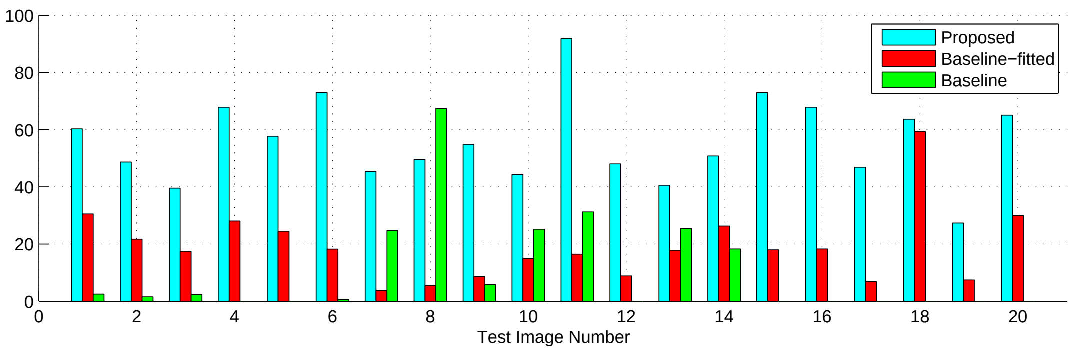


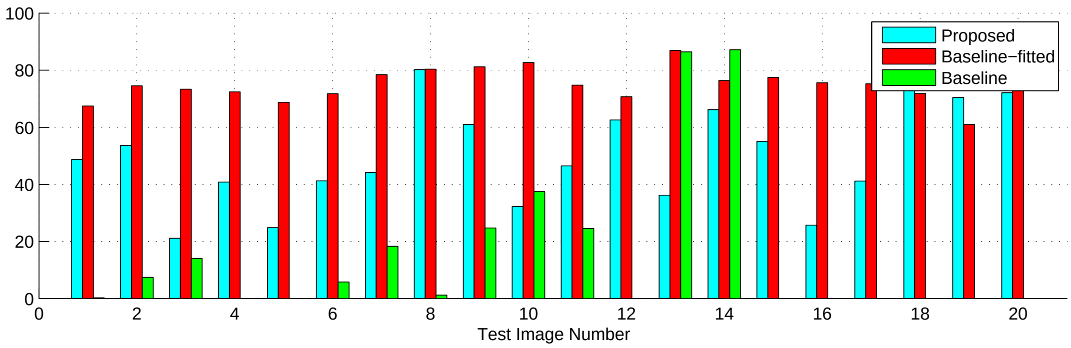
Figure 4.28: Resulting Proposed Algorithm

In this algorithm, first of all bilateral filtering is used for the purpose of removing small spectral variations on the roads. Then mean-shift segmentation is used to group spectrally similar pixels and to be able to analysis of structural features. Then seed segments chosen by investigating all of the segments according to their structural features. In seed selection, multi-scale template matching filter and some structural feature measurements are used to distinguish the most road likelihood segments from the others. Then by using spectral characteristics of seed segments, spectral classification is performed by determining spectral threshold values with Gaussian Mixture Model. As a result of classification a binary raw map of roads is obtained. Then using MS-TMF and structural feature measurements, structural verification is applied. After this step, discontinuity removing method is applied in order to connect broken-off roads. At the end, small and separated regions are removed from the image.

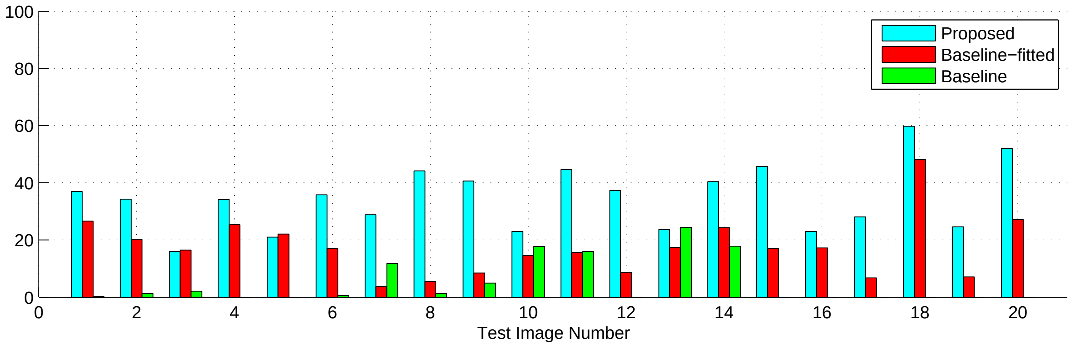
In the Figure 4.29, performance comparison of the proposed algorithm and the baseline algorithm is given and in Table 4.11, mean and standard deviation of performance measurements of the proposed algorithm and the baseline algorithm are given. There are two different results for the baseline algorithm in the Figure 4.29 and Table 4.11. In section 4.9, modified version of the baseline algorithm is defined as **Baseline-fitted** in which spectral classification parameters are fitted to the data set that is used in this thesis and post-processing operations are not used since the data set is not appropriate for these operations. These modifications are performed because, the image that was used by Long and Zhao(2005) has different spectral and structural features than the data set used in this thesis. Despite these modifications, it can be observed from the results that the performance of the resulting proposed algorithm is better than both versions of baseline algorithm.



(a) Precision



(b) Recall



(c) Quality Factor

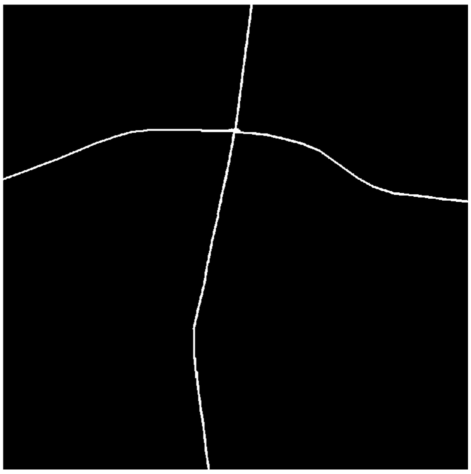
Figure 4.29: Performance Comparison of Alternative Algorithms

Table 4.11: Performance Comparison of Alternative Algorithms

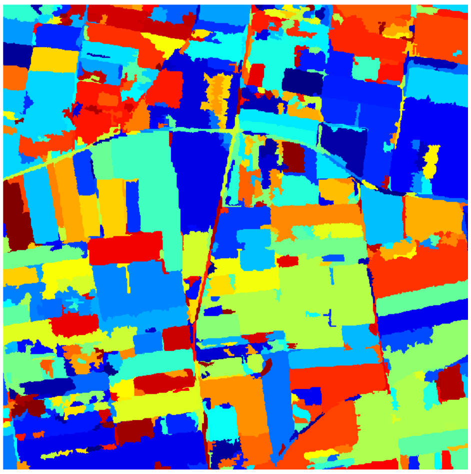
Applied Algorithm	Precision		Recall		Quality Factor	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Proposed	55.82	14.77	50.74	19.04	34.69	11.32
Baseline-fitted	19.14	12.52	74.75	5.75	17.46	10.21
Baseline	10.25	17.32	15.37	26.73	4.898	7.858

4.17.1 Step Details of the Proposed Algorithm

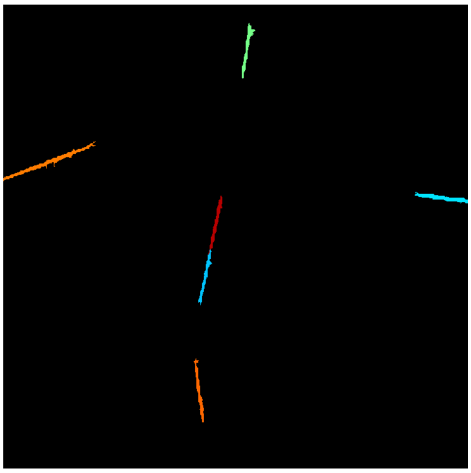
In this section, images that are obtained at several steps of the proposed algorithm are given in figures below. In classification step, 9 clusters are formed with the GMM. However only



(a) Ground Truth



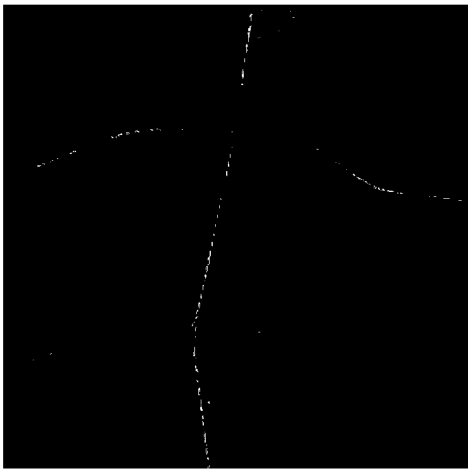
(b) Segmented Image



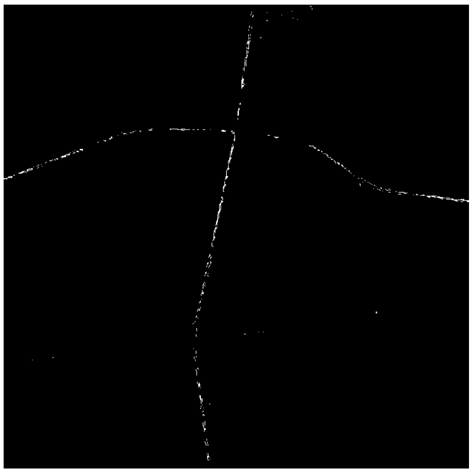
(c) Chosen Seed Segments

Figure 4.30: Proposed Algorithm - Pre-Processing and Seed Selection

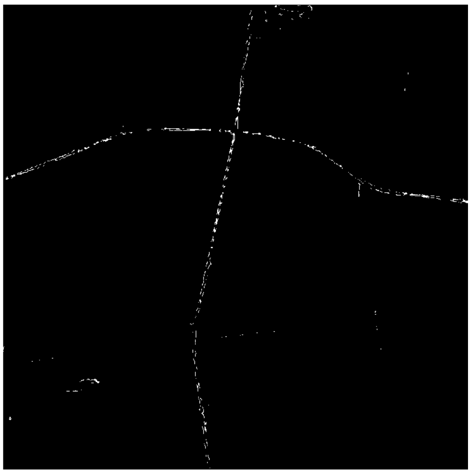
6 clusters are used to classify image as it can be seen from the following figures. Remaining 3 clusters are not used since their weights are smaller than a threshold. Also it can be aseen that, each cluster is adding additional information to the classified image.



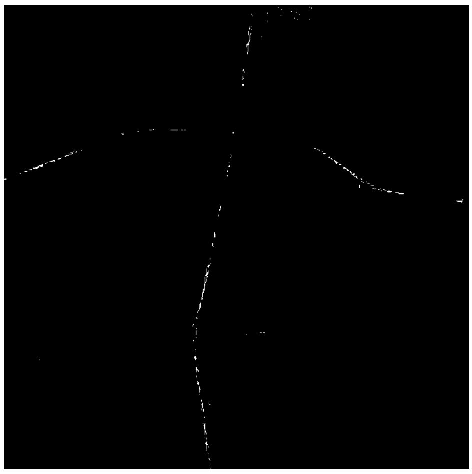
(a) Classification Result - Cluster 1



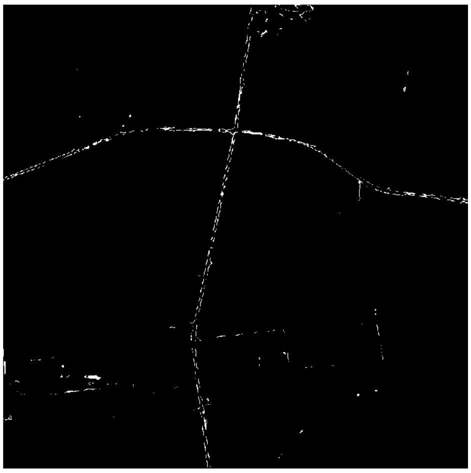
(b) Classification Result - Cluster 2



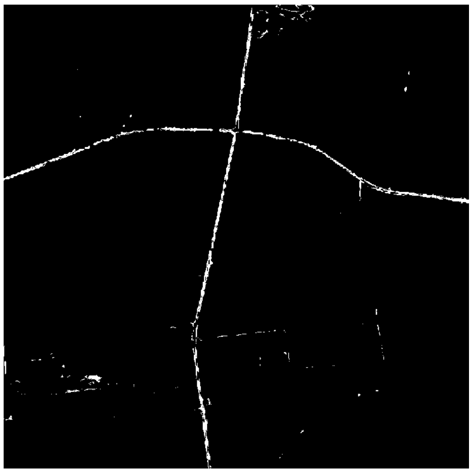
(c) Classification Result - Cluster 3



(d) Classification Result - Cluster 4

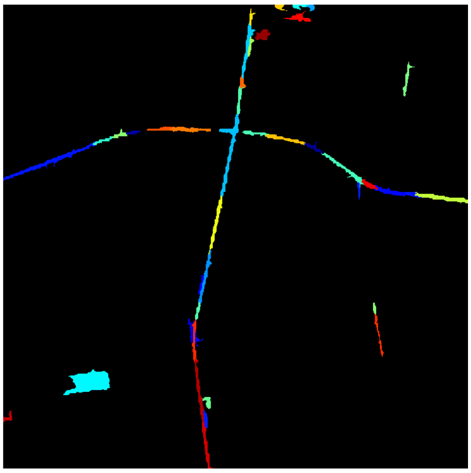


(e) Classification Result - Cluster 5

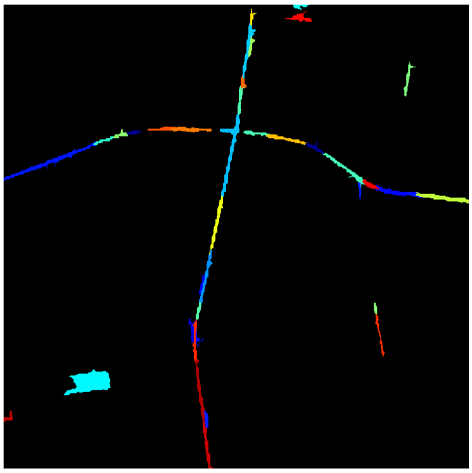


(f) Classification Result - All Clusters Together

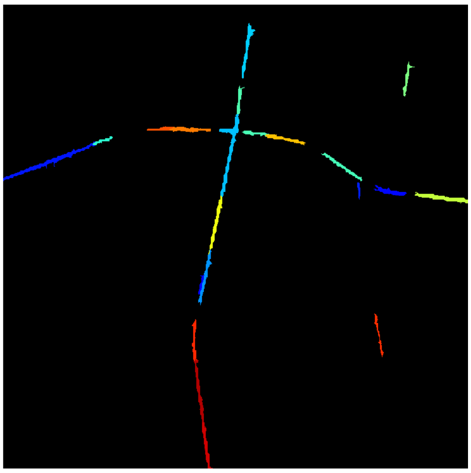
Figure 4.31: Proposed Algorithm - Spectral Classification



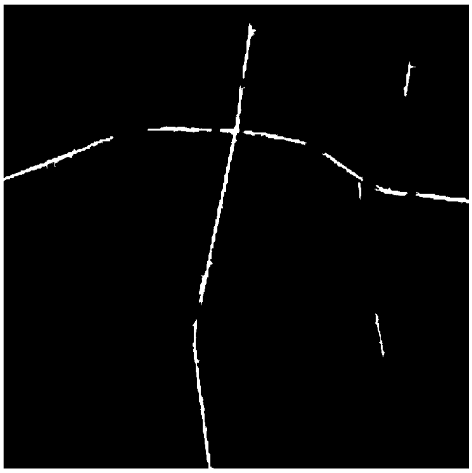
(a) Classification Result Converted to Segments



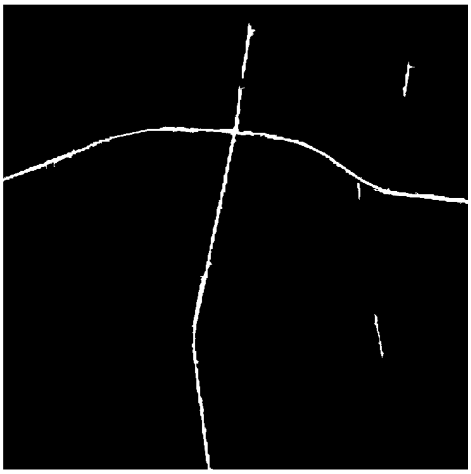
(b) Segments that have small Major/Minor Ratio Removed



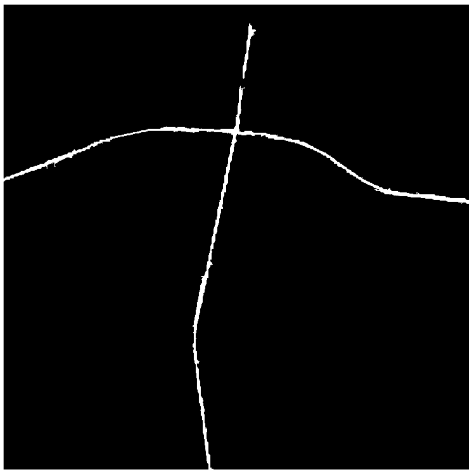
(c) Multi-scale Template Matching Filter Output



(d) Output of Structural Verification Step

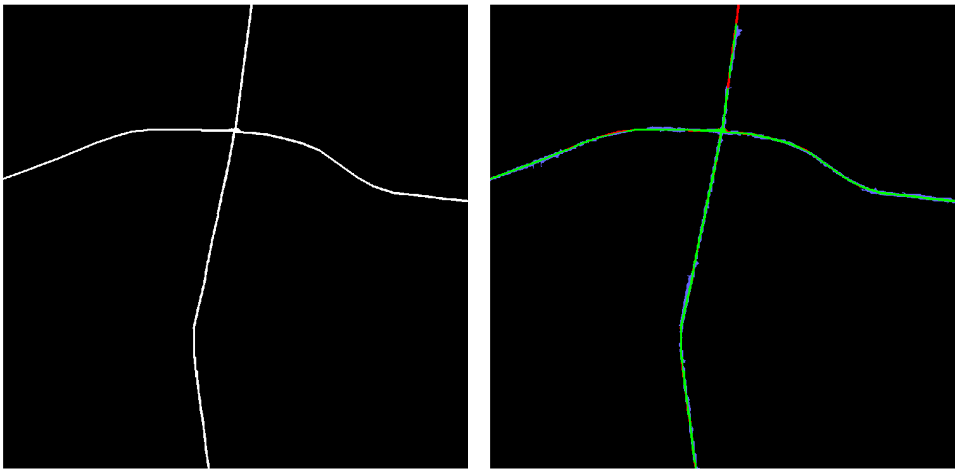


(e) Output of Discontinuity Removing Step



(f) Small and Separated Regions Removed

Figure 4.32: Proposed Algorithm - Post-Processing



(a) Ground Truth

(b) Comparison with Ground Truth(Red=FN, Green=TP, Blue=FP)

Figure 4.33: Proposed Algorithm - Final Result

The performance results and comparison of 20 test images are given in the Figure 4.29. It can be seen that performance of the proposed algorithm is varying among the test images. For example, performance on the test image 3 is very bad because seed selection for that image was not successful. As a result, performance became worse. On the other hand, performance of the algorithm is good for the test images 8, 11, 15 and 18.

CHAPTER 5

CONCLUSION

5.1 Summary

In this thesis, a previously defined road extraction algorithm using mean-shift segmentation procedure is chosen as the baseline algorithm. An experimental work flow is constructed and improvements on the baseline algorithm are presented by applying alternative processing steps. Using the alternative steps which gave the best performance, a new road extraction algorithm is proposed. Finally, experimental results that are obtained during the implementation and analysis of the proposed alternative methods and resulting proposed algorithm are given.

5.2 Discussion

In this section the results obtained by the analysis of the baseline algorithm and alternative processing steps are discussed.

5.2.1 Edge Preserving Smoothing Filters

It is observed from the comparison results that using edge-preserving smoothing filters increasing the overall performance of the algorithm. Although performance results are close to each other for the filters used which are MMCSA-LZ filter, MMCSA-C filter and bilateral filter, bilateral filter is the best among these filters.

5.2.2 Segment Merging Operations

Two different segment merging methods which are segment-edge based segment merging method(SEBSMM) and image-edge based segment merging method(IEBSMM) are proposed and compared in this thesis. They have some benefits and drawbacks. In SEBSMM, detected lines on the edges can become shorter than the expected length. And also line detection performance on curve-shaped edges is poor. As a result, mentioned method is not effective on curve-shaped segments.

Disadvantage of IEBSMM is, dependence of it on edge detection. Despite edge processing is applied after the edge detection, it is seen that road edges that are broken into sub-edges could not be corrected. Hence road segments around these edges are not merged. In addition to the broken edges, merging side selection is another disadvantage. Because, merging is applied on the side in which there are more segments to be merged and it is common that number of non-road segments within the neighborhood of an edge is more than road segments. As a result road segments are not merged.

SEBSMM is a local method since only neighbor segments are analyzed and merged. However IEBSMM is not as local as SEBSMM. Because, all of the segments that are neighbor to the analyzed edge are included in analysis. Hence, to make a decision of merging will be more complicated and will be more global than the decision of SEBSMM. Briefly it can be said that SEBSMM is more useful than IEBSMM.

It is observed from the experimental results that using segment-merging methods is not increasing the overall performance of the algorithm. So there is no need to merge segments. However segment merging operations can be useful for other application areas. For example, segment merging methods can reduce the effects of over-segmentation case.

5.2.3 Spectral Classification

It is seen for some of the test images that most of the image pixels were classified as road when predefined fixed spectral thresholds are used. Especially when there is a wide-range of test images, spectral features of roads can differ from image to image. For example, as it can be seen from the images in appendix, there are several type of images and spectral characteristic of images are differing form image to image. Hence, it is not possible to determine fixed threshold values that will give best performance for all images. Briefly, it is observed from the experimental results that automatic calculation of spectral threshold values for each image separately must be performed in order to obtain better performance results.

5.2.4 Structural Verification

It is observed from the experimental results that structural verification is a very important step. When it is used, the performance of the algorithm is approximately doubled. There are some non-road regions such as buildings, pavements, parking lots etc. which are spectrally similar to the roads. When only spectral classification is performed, these non-road regions will be also classified as road. Hence extra classification steps must be performed to remove these non-road regions. It is seen that performing structural verification is removing most of these non-road regions, because generally their shapes are not similar to the roads.

5.2.5 Post-processing operations

It is observed from the experimental results that using discontinuity removing and separated area removing methods increasing the overall performance.

5.2.6 Resulting Proposed Algorithm

It is observed that resulting proposed algorithm is more convenient than the baseline algorithm especially when the data set is including wide range of images. On the other hand, as it can be seen from the results that are presented in appendix, performance of the resulting proposed algorithm is better for the rural or semi-urban regions. Especially, main roads of most of the test images are detected with the resulting proposed algorithm.

5.3 Future Work

5.3.1 Segmentation Procedure

Mean-shift segmentation procedure is defined in **joint spatial-range domain** of input image and segments are formed with the pixels that are in the basin of attraction of a local mode in joint spatial-range domain. Mean-shift segmentation procedure is generally used in object tracking systems, raw segmentation of regular images etc. In these applications, segmented objects are compact but this is not the case for the roads because roads generally spread over the entire image so they are not compact objects. However since **spatial domain informations** are used in mean-shift segmentation procedure, roads are being divided into sub-segments (over-segmentation). If spatial domain informations are not used, roads may be detected as a single segment or as a few segments since road pixels are similar in range domain. Effects of mentioned modification of mean-shift segmentation procedure on resulting segments can be analyzed in future.

5.3.2 Segment Merging Methods

Instead of using SEBSMM or IEBSMM, a mixed alternative of them can be developed. In this alternative method, segments to be merged are selected as it is performed in SEBSMM however edge processing operations that described in IEBSMM can be used in line detection instead of Hough transform. In this alternative method, long sides of the road segments can be used in merging decision. By this way, merging of curve-shaped road segments can be easier and there can be no dependence to the Hough transform. This method may be implemented and analyzed in the future.

5.3.3 Structural Feature Enhancement

It is observed from the experimental results that some of the structural features, such as major/minor ratio, are not useful when they applied alone. Because there are a lot of concave-shaped segments. As it is explained in section 3.5.1.1, major/minor ratio of concave-shaped road segments is small. This issue may be solved by dividing segments to the sub-segments from their edge points at which sharp transitions occur. By this way, major/minor ratio may be used alone for the purpose of structural verification or classification.

5.3.4 Classification Features

There are different features on an image that can be used in classification which are spectral, structural, textural and contextual features. In this thesis, only spectral and structural features are used. If textural or contextual features are used in some steps of the classification then performance of the resulting proposed algorithm may be increased.

REFERENCES

Bacher U. and Mayer H. Automatic road extraction from multispectral high resolution satellite images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36:29–34, 2005.

Baltsavias E., OSullivan L., and Zhang C. Automated road extraction and updating using the atomi system-performance comparison between aerial lm, ADS40, IKONOS and quickbird orthoimagery. In *Proceedings of the 20th ISPRS Congress*, July 2004.

Barzohar M. and Cooper D.B. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. *IEEE Trans. PAMI*, 18(7):707–721, July 1996.

Baumgartner A., Steger C., Mayer H., and Eckstein W. Multi-resolution, semantic objects, and context for road extraction. In *Proceedings of Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, 1997.

Baumgartner A., Hinz S., and Wiedemann C. Efficient methods and interfaces for road tracking. *ISPRS*, 2002.

Bovik A. *Handbook of image and video processing*. Elsevier Academic Press, 2005.

Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–714, 1986.

Chanda B. and Mukhopadhyay S. An edge preserving noise smoothing technique using multiscale morphology. *ELSEVIER*, 82:527–544, 2002.

Chen C.C, Shahabi C., and Knoblock C.A. Utilizing road network data for automatic identification of road intersections from high resolution color orthoimagery. In *In Proceedings of the Second Workshop on Spatio-Temporal Database Management(STDBM'04), colocated with VLDB*, pages 1477–1480, 2004.

Chen T., Wang J., and Zhang K. A wavelet transform based method for road centerline extraction. *Photogrammetric Engineering and Remote Sensing*, 70:1423–1431, 2004.

Cheng H.D., Jiang X.H., Sun Y., and Wang J. Color image segmentation: advances and prospects. *The Journal of The Pattern Recognition Society*, 34:2259–2281, 2001.

Cheng Y. Mean shift, mode seeking, and clustering. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 17(8), Aug. 1995.

Chiang Y.Y. Automatic and accurate extraction of road intersections from raster maps. *Geoinformatica*, 13:121–157, Feb. 2008.

Christophe E. and Inglada J. Robust road extraction for high resolution satellite images. *ICIP 2007(International Conference on Image Processing)*, 24:437–440, March 2007.

Comaniciu D. and Meer P. Robust anaysis of feature spaces - color image segmentation. *IEEE*, pages 750–755, 1997.

Comaniciu D. and Meer P. Mean shift analysis and applications. In *7th International Conference on Computer Vision*, pages 1197–1203. IEEE, Sep. 1999.

Comaniciu D. and Meer P. The variable bandwidth mean shift and data-driven scale selection. In *8th International Conference on Computer Vision*, volume 1, pages 438–445. IEEE, 2001.

Comaniciu D. and Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 24(5), May 2002.

Doucette P., Agouris P., and Stefanidis A. Automated road extraction from high resolution multispectral imagery. *Photogrammetric Engineering and Remote Sensing*, 70:12:1405–1416, Dec. 2004.

Duda R. and Hart P. *Pattern classification and scene analysis*. John Wiley and Sons, 1973.

Duda R.O. and Hart P.E. Use of the hough transformation to detect lines and curves in pictures. Technical Report 1, Communications of ACM, Jan. 1972.

Fiset R., Cavayas F., Mouchot M., Solaiman B., and Desjardins R. Map-image matching using a multi-layer perceptron: the case of the road network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53:76–84, 1998.

Fukunaga K. and Hostetler L.D. The estimation of the gradient of a density function. *IEEE Transactions On Information Theory*, 21(1), Jan. 1975.

Geraud T. and Mouret J.B. Fast road network extraction in satellite images using mathematical morphology and markov random fields. *EURASIP Journal on Applied Signal Processing*, 16:2503–2514, 2004.

Grote A., Heipke C., Rottensteiner F., and Meyer H. Road extraction in suburban areas by region-based road subgraph extraction and evaluation. *Urban Remote Sensing Joint Event*, 2009.

Gruen A. and Li H. Road extraction from aerial and satellite images by dynamic programming. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):11–20, Aug. 1995.

Hall D. and Ball G. Isodata: A novel method of data analysis and pattern classification. Technical report, Stanford Research Institute, 1965.

Haralick R.M. Statistical and structural approaches to texture. In *Proceedings of the IEEE*, volume 67:5, pages 786–804, May 1979.

Hauptfleisch A.C. Automatic road network extraction from high resolution satellite imagery using spectral classiffication methods. Master's thesis, University of Pretoria, Feb. 2010.

Hinz S. and Baumgartner A. Automatic extraction of urban road networks from multi-view aerial imagery. *Journal of Photogrammetry and Remote Sensing*, 58:83–98, 2003.

Hinz S., Baumgartner A., Steger C., Mayer H., Eckstein W., Ebner H., and Radig B. Road extraction in rural an urban areas, 1999.

Hough P.V.C. Method and means for recognizing complex patterns. Technical Report 3069654, United States Patent Office, December 1962.

Hu X. and Tao C.V. A reliable and fast ribbon road detector using profile analysis and model-based verification. *International Journal of Remote Sensing*, 26(5):887–902, March 2005.

Hui G., Jilin L., and Yaya L. Road extracting based on texture analysis. In *Proceedings of the 16th International Conference on Artificial Reality and Telexistence-Workshops*, 2006.

Jin X. and Davis C.H. An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion. *Information Fusion*, 6: 257–273, Sep. 2003.

Li C., Xia D., Wang Y., and Zou H. Satellite sensing image analysis by integrating neural network with knowledge reasoning technique. In *Proceedings of SPIE Image Processing and Pattern Recognition in Remote Sensing*, 2003.

Long H. and Zhao Z. Urban road extraction from high-resolution optical satellite images. *International Journal of Remote Sensing*, 26(22):4907–4921, Nov. 2005. doi: 10.1080/01431160500258966.

Mao J. and Jain A.K. Texture classification and segmentation using multi-resolution simultaneous autoregressive models. *Pattern Recognition*, 25:2:173–188, Feb. 1992.

Maragos P. Pattern spectrum and multiscale shape representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11:701–716, July 1989.

Marr D. and Hildreth E. Theory of edge detection. In *The Proceedings of the Royal Society of London*, 1980.

Mei T., Li F., Qin Q., and Li D. Road extraction from remote sensing image using support vector machine. In *Proceedings of SPIE - Third International Symposium on Multispectral Image Processing and Pattern Recognition*, 2003.

Mena J.B. State of the art on automatic road extraction for GIS update - a novel classification. *Pattern Recognition Letters*, 24:3037–3058, March 2003.

Mohammadzadeh A., Tavakoli A., and Zoej M.V. Automatic linear feature extraction of iranian roads from high resolution multi-spectral satellite imagery. In *Proceedings of the 10th International Society for Photogrammetry and Remote Sensing Congress*, 2004.

Mohammedzadeh A., Zoej M.J. Valadan, and Tavakoli A. Automatic main road extraction form high resolution satellite imageries by means of self-learning fuzzy-GA algorithm. *Journal of Appiled Sciences*, 19(8):3431–3438, 2008.

Mokhtarzade M. and Zoej M.V. Road detection from high-resolution satellite images using artificial neural networks. In *Proceedings of SPIE Image Processing and Pattern Recognition in Remote Sensing*, volume 9, pages 32–40, 2007.

Nalwa V.S. and Binford T.O. On detecting edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:699–714, 1986.

Oddo L.A., Doucette P., and Agouris P. Automated road extraction via the hybridization of self-organization and model based techniques. In *Proceedings of the 29th Applied Imagery Pattern Recognition Workshop*, pages 32–38, Oct. 2000.

Ohanian P.P. and Dubes R.C. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25:819–833, Aug. 1992.

Ruskone R. and Airault S. Toward an automatic extraction of the road network by local interpretation of the scene. *In 46th Photogrammetric Week*, 1997.

Sarkar S. and Boyer K. Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:382–399, 1993.

Senthilnath J., Rajeshwari M., and Omkar S.N. Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method. *Journal of Indian Society of Remote Sensing*, 37:351–361, 2009.

Shackelford A.K. and Davis C.H. Urban road network extraction from high-resolution multispectral data. In *Proceedings of the 2nd GRSS and ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pages 142–146, May 2003.

Shi W. and Zhu C. The line segment match method for extracting road network from high-resolution satellite images. *IEEE Transactions on Geoscience and Remoete Sensing*, 40(2): 511–514, Feb. 2002.

Singlia M. and Hemacllandran K. Performance analysis of color spaces in image retrieval. *Assam University Journal of Science and Technology: Physical Sciences and Technology*, 7 (11):94–104, 2011.

Song M. and Civco D. Road extraction using SVM and image segmentation. *Photogrammetric Engineering Remote Sensing*, 70(12):1365–1371, Dec. 2004.

Tomasi C. and Manduchi R., editors. *Bilateral Filtering for Gray and Color Images*, 1998.

Xu G., Zhang D., and Liu X. Road extraction in high resolution images from google earth. *ICICS*, 2009.

Yang J. and Wang R.S. Classified road detection from satellite images based on perceptual organization. *International Journal of Remote Sensing*, 28:4653–4669, 2007.

Yuan J., Wang D., Yan L., and Li R. Automatic road extraction from satellite imagery using legion. In *Proceedings of International Joint Conference on Neural Networks*, pages 3471–3476. IEEE, June 2009.

Zhang C. Towards an operational system for automated updating of road databases by integration of imagery and geodata. *ISPRS Journal of Photogram-metry and Remote Sensing*, 58:166–186, 2004.

Zhang Q. and Couloigner I. Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery. *Pattern Recognition Letters*, 27:937–946, 2006.

Zhao H., Kumagai J., Nakagawa M., and Shibasaki R. Semi-automatic road extractıon from high-resolution satellite image. In *Photogrammetric Computer Vision*, number 45 in 3. ISPRS, Sep. 2002.