PARAMETRIC AND NONPARAMETRIC APPROACHES FOR MULTISENSOR DATA FUSION

by

Bing Ma

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2001

Doctoral Committee:

Professor Alfred O. Hero III, Co-chair
Associate Professor Sridhar Lakshmanan, Co-chair
Associate Professor Kamal Sarabandi
Professor William J. Williams

ABSTRACT

PARAMETRIC AND NONPARAMETRIC APPROACHES FOR MULTISENSOR DATA FUSION

by
Bing Ma

Co-chairs: Alfred O. Hero III and Sridhar Lakshmanan

Multisensor data fusion technology combines data and information from multiple sensors to achieve improved accuracies and better inference about the environment than could be achieved by the use of a single sensor alone. In this dissertation, we propose parametric and nonparametric multisensor data fusion algorithms with a broad range of applications.

Image registration is a vital first step in fusing sensor data. Among the wide range of registration techniques that have been developed for various applications, mutual information based registration algorithms have been accepted as one of the most accurate and robust methods. Inspired by the mutual information based approaches, we propose to use the joint Rényi entropy as the dissimilarity metric between images. Since the Rényi entropy of an image can be estimated with the length of the minimum spanning tree over the corresponding graph, the proposed information-theoretic registration algorithm can be implemented by a novel nonparametric graph-representation method. The image matching is performed by minimizing the length of the minimum spanning tree (MST) which spans the graph generated from the overlapping images. Our method also takes advantage of the minimum k-point spanning tree (k-MST) approach to robustify the registration against outliers in the images. Since this algorithm does not require any parametric model, it can be directly applied to a variety of image types.

We also propose a parametric sensor fusion algorithm for simultaneous lane and pavement boundary detection in registered optical and radar images. The fusion problem is formulated in a Bayesian setting where the deformable templates play the role of *a priori* density and the imaging likelihoods play the role of likelihood functions. Under these formulations, the fusion problem is solved by a joint maximum *a posteriori* (MAP) estimate. We first employ existing prior and likelihood models in the fusion framework and experimental results have shown that the fusion method outperforms single sensor based boundary detection algorithms. However, there are some drawbacks in the existing models. To improve the fusion algorithm, we propose to utilize concentric circular shape models to represent the boundaries and to employ Gaussian and log-normal densities to describe the optical and radar imaging processes. This fusion algorithm leads to a well conditioned parameter estimation problem and the optical and radar observation data are combined effectively and efficiently.

I dedicate this thesis to my family for their love and support.

ACKNOWLEDGEMENTS

I am very grateful to many individuals that have helped and supported me through my doctoral study at University of Michigan. First of all, I would like to express my thanks to my advisors Professor Alfred Hero and Professor Sridhar Lakshmanan for their excellent guidance and encouragement, their patience in discussing my questions and their original suggestions for some solutions during my Ph.D. work. Their broad knowledge and deep insights have been an important resource to my accomplishment.

I would like to express my gratitude to Dr. Karl Kluge at Artificial Intelligence Laboratory for his helpful discussion of the previous work on the boundary detection project and providing me with the source codes of the LOIS algorithm and the general image display software package in unix environments. I also thank Professors Kamal Sarabandi and William Williams for serving in my committee and helping with their ideas.

I owe a lot to my friends Tingfang Ji, Robinson Piramuthu, Hua Xie, and Jia Li for their warm friendship and help in both professional and personal perspectives. I would like to thank my colleagues: Mahesh Godavarti, Robby Gupta, John Choi, Hyungsoo Kim, Tara Javidi, Saowapak Sotthivirat, Vitoria Yee, Olgica Milenkovic, Selin Aviyente, and Naveen Kashyap. They have made my staying at Michigan an enjoyable experience.

I would like to thank Ms. Susan Yale, Beth Olson, Beth Lawson and Ann Pace for their kind help in my most needing times.

Finally I would like to express my thank to my mother and father with my whole heart for their endless love and support. I thank my sister for making me smile all the time and taking good care of my parents while I am far away from home.

TABLE OF CONTENTS

DEDICATION . . . . **ii**

ACKNOWLEDGEMENTS . . . . **iii**

LIST OF FIGURES . . . . **viii**

LIST OF TABLES . . . . **xii**

CHAPTER

I. Introduction 1

1.1 Image Registration with Minimum Spanning Tree Method — A Nonparametric Fusion Approach 2
1.2 Simultaneous Detection of Lane and Pavement Boundaries — A Parametric Fusion Approach 5
1.3 Outline of The Dissertation 8

II. Overview of Multisensor Data Fusion 10

2.1 Motivation for Sensor Fusion 11
2.2 Classification of Sensor Fusion Techniques 13
  2.2.1 Classification by Types of Sensor Data 14
  2.2.2 Classification by Levels of Representation 14
  2.2.3 Classification by Mathematical Foundation 18

III. Minimum Spanning Tree and Rényi Entropy 23

3.1 Minimum Spanning Tree 23
3.2 Limit Theorem and Rate of Convergence for MST Length Functionals 25
  3.2.1 Limit Theorem for MST Length Functionals 25
  3.2.2 Rate of Convergence for MST Length Functionals on Uniform Samples 25
  3.2.3 Rate of Convergence for MST Length Functionals on Non-uniform Samples 27

3.3 Rényi Entropy and Rényi Information Divergence 29
3.4 Estimation of Rényi Entropy with MST 31
3.5 Robustification of Rényi Entropy Estimator by k-MST 35
3.6 Limit Theorem of k-MST greedy approximation 40

IV. Image Registration Via Graph Matching 42

4.1 Image Registration Concepts and Formulation 42
4.2 Image Registration Via Maximizing Mutual Information 47
4.3 Minimum Rényi Entropy Criterion 49
4.4 Image Registration Procedure 61
    4.4.1 Geo-registration 62
    4.4.2 Medical Image Registration 65
4.5 Determination of the Rényi Entropy Order 69
4.6 Feature Extraction 75
    4.6.1 Feature Extraction Via Uniform Spatial Sub-sampling 76
    4.6.2 Feature Extraction Via Vector Quantization 78
    4.6.3 Feature Extraction Via Stratified Sampling with Centroid Refinements 82
4.7 MST Construction 84
4.8 Noise Removal with k-MST Technique 84
4.9 Sliding Sampling Grid Approach 91
4.10 Conclusions and Future Work 94

V. Sensor Systems for Road Boundary Detection 98

5.1 Road Boundary Detection Problem 98
5.2 Sensor Systems 103

VI. Fusion Algorithm for Lane and Pavement Boundary Detection with Existing Prior and Likelihood Models 109

6.1 Parabolic Models of Lane and Pavement Boundaries 109
6.2 Imaging Likelihoods 113
    6.2.1 Radar Imaging Likelihood 113
    6.2.2 Optical Imaging Likelihood 119
6.3 Joint MAP Estimate for Lane and Pavement Boundaries 121
    6.3.1 Empirical MAP 122
    6.3.2 Computation of Empirical MAP Estimate 126
    6.3.3 Performance Comparison 128
6.4 Confidence Measure for Parameter Estimate 135
    6.4.1 Objective Confidence Measure 135
    6.4.2 Curvature Estimation 137
    6.4.3 Ill-conditioned Parameter Estimation Problem 140

VII. Fusion Algorithm for Lane and Pavement boundary Detection with Concentric Circular Models 143

7.1 Concentric Circular Models 144
7.2 Decoupling of The Parameterization 150
7.3 Imaging Likelihoods 156
    7.3.1 Radar Imaging Likelihood 156
    7.3.2 Optical Imaging Likelihood 157
7.4 Joint MAP Estimate for Lane and Pavement Boundaries 160
7.5 Experimental Results 163
7.6 Conclusions 172

VIII. Summary and Future Work for Boundary Detection 174

8.1 Summary 174
8.2 Future Work 176

APPENDIX **177**

BIBLIOGRAPHY **187**

LIST OF FIGURES

Figure

3.1.1 An MST example 24

3.1.2 Examples for k-point MST's 26

3.4.3 2D uniform and torus sample example 33

3.4.4 Mean MST length for uniform and torus samples 34

3.4.5 Comparison of the MST length of uniform and torus samples 34

3.5.6 Signal, contaminated signal, and their corresponding MST's 36

3.5.7 Use k-MST to remove outliers 37

3.5.8 k-MST length vs. number of rejected points 38

3.5.9 Use k-MST to remove outliers 39

4.1.1 A multidate image registration example 43

4.4.2 EO image generation block diagram 64

4.4.3 Misaligned EO and reference images 65

4.4.4 Aligned EO and reference images 66

4.4.5 MRI image generation process 67

4.4.6 Brain images for registration 69

4.5.7 Rényi information divergence with different orders 70

4.5.8 Resolution of Rényi information divergence with different orders 71

4.5.9 Relative resolution of Rényi information divergence with different orders 72

4.5.10 Test and reference DEM images 72

4.5.11 Rényi information divergence with difference orders 73

4.5.12 Resolution of Rényi information divergence with difference orders . . . 74

4.5.13 Relative resolution of Rényi information divergence with difference orders 74

4.6.14 Histograms of the EO and reference images 76

4.6.15 Rényi information divergence with respect to the threshold 77

4.6.16 Thresholded EO and reference images 78

4.6.17 Feature vectors extracted via uniform sampling 79

4.6.18 Feature vectors via vector quantization 82

4.6.19 Feature vectors via stratified sampling with centroid refinements 83

4.7.20 MST demonstration for misaligned images 85

4.7.21 MST demonstration for aligned images 85

4.7.22 MST length for different test-reference image pairs 86

4.7.23 Result for EO-terrain map registration 86

4.8.24 Sub-sampled brain images for registration 87

4.8.25 Illustration of k-MST for pre-operation brain image over the 3-dimensional domain of pixel position and grey scale 88

4.8.26 kMST length as a function of pruned number 89

4.8.27 Sub-sampled images after outlier removal 89

4.8.28 MST length as a function of rotation angle for brain image registration . 90

4.8.29 Registration result for brain images 90

4.9.30 MST length over subgraphs 92

4.9.31 Total MST lengths versus test image indexes 93

5.1.1 Clutter due to presence of complex shadows 99

5.1.2 Missing data due to other vehicles occluding the boundaries of interest . 100

5.1.3 Low signal-to-noise ratio radar data 100

5.1.4 Non-lane or -pavement boundaries due to entry/exit ramps 101

5.2.5 WOLVERINE I 104

5.2.6 Radar sensor setup 105

5.2.7 A typical road scenario in clear/day-light weather 106

5.2.8 A typical road scenario in foggy/low-light weather 106

5.2.9 Diagram of the optical and radar multisensor fusion 108

6.1.1 Boundary templates for the optical and radar images 112

6.2.2 Rationale for using log-normal pdf. (a) shows a radar image overlayed with the correct positions for the left and right pavement boundaries. (b), (c), and (d) show histograms of the actual radar returns for the three regions – the road, the left-side and right-side of the road. Also shown in (b), (c), and (d) are the maximum likelihood fits of the log-normal, Rayleigh, and Gaussian pdf's to the radar return histograms. 115

6.3.3 A pair of optical and radar images 124

6.3.4 Different dynamic range of the matching functions for optical and radar images 124

6.3.5 Wrong edge detection results with fusion method and no-weight scheme 125

6.3.6 Edge detection by fixed weight scheme 126

6.3.7 Performance comparison of the fusion and single sensor based methods 130

6.3.8 Performance comparison of the fusion and single sensor based methods 131

6.3.9 Performance comparison of the fusion and single sensor based methods 132

6.3.10 Advantage of the fusion method over single sensor based method 133

6.3.11 Advantage of the fixed weighting scheme over no weight scheme 134

6.4.12 Cross sections of the radar normalized matching function with respect to their pavement shape parameters 138

6.4.13 A cross section of the normalized matching function and its quadratic approximation 140

7.1.1 A typical road scenario in a radar image 145

7.1.2 A typical road scenario in a radar image 145

7.1.3 Cases where circles do not intersect the cone 149

7.1.4 Cases where circles intersect the cone in an infeasible way 150

7.2.5 Reparameterization for the case where the center is to the left of the road 152

7.2.6 Reparameterization for the case where the center is to the right of the road 153

7.2.7 Unified parameterization for both lane and pavement boundaries 155

7.3.8 An ideal lane scenario on the ground plane 159

7.4.9 Boundary detection results with fusion method 162

7.5.10 Performance comparison of the fusion and single sensor based methods 165

7.5.11 Performance comparison of the fusion and single sensor based methods 166

7.5.12 Performance comparison of the fusion and single sensor based methods 167

7.5.13 Performance comparison of fusion algorithms 170

7.5.14 Examples to show what confidence measure indicates 172

LIST OF TABLES

Table

2.1 Comparison of Fusion Levels 19

6.1 Curvature results for some radar images 141

6.2 Curvature results for some optical images 141

7.1 Curvature results for some radar images with circular shape models . . . 171

7.2 Curvature results for some optical images with circular shape models . . 171

CHAPTER I

Introduction

The recent development of new sensors has created a need for data processing techniques that can fuse observations from a variety of different sensors. *Multisensor data fusion* is an evolving technology concerned with the problem of how to combine data and information from multiple sensors in order to achieve improved accuracies and better inference about the environment than could be achieved by the use of a single sensor alone.

The concept of multisensor data fusion is hardly new. Humans and animals have evolved the capability to use multiple senses to improve their ability to survive. The human or animal brain is a good example of a data fusion system. The brain integrates sensory information namely sight, sound, smell, taste, and touch data to achieve more accurate assessment of the surrounding environment and identification of threats, thereby humans and animals improve their chances of survival.

In recent years, multisensor data fusion has been extensively investigated by researchers in a variety of science and engineering disciplines, such as automated target recognition [7, 8], automatic landing guidance [23, 98], remote sensing [32, 35, 78], monitoring of manufacturing processes [14, 106], robotics [1, 4], and medical applications [36, 37], to mention but a few.

In this work, we focus on developing parametric and nonparametric multisensor data fusion algorithms. A parametric algorithm is based on the assumption of a parametric model. The algorithm consists of fitting the model to the data, and estimating the parameters of the model. In contrast, a nonparametric algorithm is not based on any parametric model. Thus the nonparametric algorithm is applied when the problem parameterization is unknown or unavailable.

Our work consists of the following two multisensor fusion objectives:

① Develop a nonparametric approach for *image registration* from different sensors.

② Develop a parametric fusion approach for *detection of lane and pavement boundaries* from optical and radar sensors.

Both of the objectives are in the fields of multisensor data fusion and certainly they have some common characteristics. They share the basic data processing flow patterns — first put the measurements from sensors into a common coordinate system, then extract feature vectors from the measurements and finally combine the feature vectors together to achieve accurate detection, estimation, and classification. However, since the nonparametric approach is model independent while the parametric one is based on a model that is described with parameters, the two objectives are only loosely connected. For convenience and clarity of further explanation, we will separately pursue the two objectives.

1.1 Image Registration with Minimum Spanning Tree Method — A Nonparametric Fusion Approach

Image Registration refers to the process of aligning images so that their details overlap accurately. Images are usually registered for the purpose of combining or comparing them. Image registration is indispensable for such tasks as data fusion, navigation, clinic studies, and motion detection. A wide range of registration techniques has been developed for many different types of applications and data, such as mean squared alignment [42], correlation registration [43], moment invariant matching [18]. Basically, each method exploits different image features of the object such as boundaries, moments and texture, which are extracted from the image, and uses them to solve the matching problem.

In 1995, Viola and Wells [103] and Collignon et al. [17] independently proposed an information-theoretic approach for matching of images. In their work, mutual information is introduced as a measure for evaluating the similarity between images and image registration is achieved by maximizing the mutual information between the to-be-registered images.

Inspired by the mutual information based matching algorithm, we have developed a registration algorithm using Rényi entropy as a dissimilarity metric between two images. When two images are properly matched, corresponding anatomical areas should overlap and the resulting joint probability distribution contains high values for the intensity values in the anatomical areas, i.e., the joint probability distribution is relatively highly concentrated. Thus, the Rényi entropy of the overlapping images should be small. Since misregistration increases the dispersion of the joint probability distribution, i.e., increases the Rényi entropy, one should be able to obtain registration by finding the configuration corresponding to the minimum Rényi entropy. This algorithm does not require any parametric model, so it is a nonparametric approach and can be applied to a broad range of image types directly.

The minimum spanning tree (MST) is a graph-theoretic technique which determines the dominant skeletal pattern of a point set by mapping the shortest path of linear, nearest-neighbor connections. A spanning tree over n vertices is a connected acyclic graph which passes through all these vertices. The MST is the spanning tree with minimal length. In [96] the MST length was proved to provide a consistent estimator for Rényi entropy.

The k-point MST denotes the minimum spanning tree over a specified k vertices out of the n vertices in the graph. The minimum k-point spanning tree (k-MST) is defined as that k-point MST which has minimal length among k-point MST's spanning each of the $\left(\begin{array}{c} n \\ k \end{array}\right)$ possible subsets of k vertices. For a noise contaminated signal, the k-MST offers a robust estimate for the entropy of the signal by virtue of its ability to judiciously prune straggler points [40].

Thanks to the MST's capability of providing an estimator for the Rényi entropy, the proposed minimum Rényi entropy based image registration algorithm is boiled down to a novel graph-representation method. The image matching is performed by minimizing the length of the MST that spans the graph generated from the overlapping images. Our method also employs the k-MST approach to robustify the registration against outliers in the images.

As the complexities of the MST algorithm and k-MST greedy approximation algorithm are polynomial in the number of vertices, it is crucial to efficiently extract feature vectors from the original images in order to reduce the number of vertices in the graph and hence reduce the computational complexity of the registration algorithm. The feature vectors must be able to well represent the original image without losing too much information; on the other hand, the number of feature vectors have to be small enough so that the MST and k-MST techniques can handle them. Since these two objectives are in conflict, feature extraction is a challenging task. We propose a number of feature extraction techniques in this dissertation and the experimental results are very promising.

1.2 Simultaneous Detection of Lane and Pavement Boundaries — A Parametric Fusion Approach

Lane and pavement boundary detection is an enabling or enhancing technology which will have significant impact on the next generation of automotive systems such as road departure or lane excursion warning, intelligent cruise control, and, ultimately, autonomous driving.

Lane and pavement boundary detection problem is particularly difficult when no prior knowledge of the road geometry is available (such as from previous time instants [22, 49, 75]) and when the detection algorithms have to locate the boundaries even in situations where there may be a great deal of clutter in the images.

Deformable templates have been employed to model the lane and/or pavement boundaries. In [56, 68], the authors proposed a global shape model to describe the boundaries and thus the boundary detection problem becomes a problem of estimating the model parameters, i.e., a problem of finding the mode of an objective function with respect to the model parameters. These model-based methods operate well and have the advantage of having quite small search space for the optimization of the objective function. In [56], a vision-based algorithm was developed for locating lane boundaries, while in [68] a model-based method was proposed for detecting pavement boundaries in radar images.

Note that previously lane boundary detection in optical images [53, 54, 55, 56, 83], and pavement boundary detection in radar images [48, 59, 60, 68] have always been studied separately. However, a single sensor, either optical or radar sensor, has limited ability to sense and identify the relevant features in varying environments. For example, the optical sensor is not able to operate in a poorly illuminated environment, while the radar sensor can not distinguish the lane markers on the road.

To take advantage of the strengths (and overcome the weaknesses) of both the optical and radar sensors, we design parametric fusion algorithms to effectively combine the two different types of sensory data such that we can obtain a more precise and robust interpretation of the sensed environment. In the proposed fusion algorithms, the lane and pavement boundaries are represented by deformable templates and the distribution of the deformation parameters provides the *a priori* probability density funtion (pdf). Furthermore, the optical and radar imaging processes are described by imaging likelihood functions. With the *a priori* pdf and imaging likelihood functions, the goal of detecting boundaries via fusing optical and radar observations can be achieved with maximum *a posteriori* (MAP) estimation.

Polynomial shape models have been one of the prevailing deformable templates used for describing the lane and pavement boundaries in the past years. Specifically, in previous work that is closely related to our fusion mission [56, 68], parabolic shape models were employed in separate lane and pavement boundary detection. For the detection algorithms presented in [56] and [68], the radar imaging process is modeled with a log-normal pdf and the optical imaging process is described with an empirical matching function. To make most of existing boundary shape models and imaging likelihood functions, our first effort in detecting boundaries with fusion techniques is to jointly estimate the shape parameters with the MAP method where the parabolic shape models play the role of prior information and the log-normal pdf and the empirical matching function play the role of likelihood functions. Since this fusion technique combines information from both optical and radar images, the boundary detection results are shown to be more accurate and more reliable than single sensor based detection algorithms, especially in an adverse environment.

Although the detection results are promising with the fusion algorithm using existing prior shape models and likelihood functions, there are some drawbacks that prevent us from exploring the most of the fusion algorithm. First, since the parameters in the parabolic shape model have different units and are of different orders of magnitude, the MAP estimation is an inherent ill-conditioned problem. To get rid of this inherent pitfall of the parabolic model, we propose to use concentric circular shape models to describe the lane and pavement boundaries. Circular shape models lead to a much better conditioned estimation problem due to the compatibility of their parameters, namely, parameters share the same units and are of the same order of magnitude.

The optical likelihood function of the previous approach results in complications in the joint estimation problem. The empirical matching function inherited from single optical sensor lane detection algorithm is not a valid likelihood function since it is not normalized to a probability density function. In the radar and optical fusion algorithm, the empirical function has to be carefully weighted so that each sensor makes a fair contribution to the joint likelihood. In [71] we experimentally selected the weights according to the minimum mean square criterion which yield reasonably good results, but this empirical matching function make systematic and theoretically sound weight picking an impossible task. Inspired by the log-normal radar image likelihood function, we propose to model the optical imaging process as a Gaussian process which leads to a well defined likelihood function that can be easily manipulated with the likelihood from the radar sensor.

Then in our second effort in improving the fusion algorithm, we employ concentric circular shape models to represent the lane and pavement boundaries, and utilize the log-normal and Gaussian pdf’s to describe the radar and optical imaging processes. This new fusion algorithm is expected to yield a well conditioned estimation problem and combines the optical and radar modalities effectively and efficiently.

Although we develop this parametric fusion algorithm with a specific application in mind, we are aware that this fusion methodology can be applied to other model-based applications as well.

1.3 Outline of The Dissertation

In Chapter II, we give a brief historical overview of work related to multisensor data fusion.

Chapters III and IV deal with the nonparametric image registration methodology via graph matching. The main idea that separates this work from the large volume of work on nonparametric image registration algorithms is establishing the connection between minimum Rényi entropy and image registration and proposing the use of graph matching techniques to register a variety of images. In Chapter III, we discuss some critical properties, such as asymptote and convergence rate, of the MST length functional and its application in Rényi entropy estimation. In addition, k-MST is also introduced for robustification of entropy estimators. In Chapter IV, we present the criterion and procedure of image registration with graph matching techniques. First we describe a general image registration procedures, and detail the spatial transformation and feature extraction for MST construction. Then, we present the construction of MST for both noiseless and noisy images for EO-terrain map registration and MRI image registration.

In chapters V to VIII we present a parametric fusion approach for simultaneous detection of lane and pavement boundaries. In Chapter V we state the general settings of this boundary detection problem and pointed out that among the state-of-art techniques, deformable templates have been through promising progress in boundary detection society. We present a model-based fusion algorithm in Chapter VI. The fusion problem is formulated in a Bayesian setting where the deformable templates play the role of *a priori* density and the imaging likelihoods play the role of likelihood functions. In this chapter we make use of the existing boundary shape models and imaging likelihood functions in the fusion setting. We describe the joint maximum *a posteriori* (MAP) estimate for the boundary detection and show the experimental results in Section 6.3. And confidence measures for the lane and pavement boundary parameter estimates are studied in the end of Chapter VI. Chapter VII presents the boundary detection with the proposed concentric circle model, which makes the detection problem much better conditioned due to the compatibility of the model parameters. In this chapter, we also propose a Gaussian pdf to model the optical imaging process. And we show the advantage of this fusion algorithm over the algorithm describe in Chapter VI. Finally in Chapter VIII we provide the summary on the boundary detection algorithms and discuss the future work.

In the appendix, we give a preliminary result for the convergence rate of MST length of samples drawn from non-uniform densities.

CHAPTER II

Overview of Multisensor Data Fusion

Information acquisition consists of two fundamental processes — *sensor measuring* and *feature extraction*. The sensor measuring process is to obtain certain measurements of quantities (sensory data), which are dependent on the structures in the environment and their configuration. The feature extraction process is to derive from the sensory data specific items of information about meaningful structures in the environment. The sensor measuring process can be interpreted as a mapping of the state of the world into a set of images of much lower dimensionality. The usual operations involved in the sensor measuring process are sampling and projection. Both of the processes result in loss of information. The feature extraction process can be treated as the process of inverting the imaging map. However, since the sensor measuring process loses information in general, the imaging map is non-invertible. Therefore, the feature extraction task is fundamentally that of solving an ill-posed problem and must be approached with some reliable and robust mathematical tools.

Multisensor data fusion has emerged as the method of choice for resolving these problems. Data fusion techniques combine data from multiple sensors, and related information from associated databases to achieve improved accuracies and more specific inferences than could be achieved by the use of a single sensor alone. In recent years, multisensor data fusion has received significant attention among researchers in different disciplines of science and engineering, such as automated target recognition [7, 8, 10, 11, 24, 34], automatic landing guidance [23, 98], remote sensing [32, 35, 78], monitoring of manufacturing processes [14, 106], robotics [1, 4], and medical applications [36, 37, 92, 93].

Techniques to combine or fuse data are drawn from diverse set of more traditional disciplines including: digital signal processing [95, 102], statistical estimation [76, 91], control theory [25, 94], artificial intelligence [19, 73], and classic numerical methods [33, 105].

2.1 Motivation for Sensor Fusion

Before processing further, a worthwhile discussion would be about the motivation behind sensor fusion research and why the great effort that has been put into finding robust methods for fusing sensor data is natural.

Although the computer techniques have been developed so rapidly in the last few decades and people have witnessed an explosion in the use of electronic data processing with applications becoming ubiquitous, developing several important data processing applications, such as automatic navigation systems, is more difficult than foreseen. Systems developed with well-defined data interfaces have access to the information they require. Automatic navigation systems deal with the real world and unfortunately, the real world is not well-defined. Thus automation systems need to be able to react to unforeseen situations. These systems use sensors as their interface to a changing and possibly hostile environment.

Sensors are not reliable interfaces in their current form. Sensors are devices that collect data about the world around them. They range from inexpensive cameras to earth observation satellites costing millions of dollars. In spite of this variety, all sensors have a few characteristics in common. Every sensor device has a limited accuracy, which is subject to the effect of system noise, thermal noise, etc., and will function incorrectly under some adverse conditions.

The ability of automation systems to interact with their environment has been severely impeded by the state of current sensor technology. Sensors are currently not accurate nor reliable enough for effective use in many automation applications.

Sensor fusion seeks to overcome the drawbacks mentioned above by making the most of existing sensor technology. Using measurements from several independent sensors makes a system less vulnerable to the failure of a single component. Combining measurements from several different kinds of sensors can give a system more accurate information than otherwise possible. Combining several measurements from the same sensor makes a system less sensitive to noise because that in the measurements of the same environment at different times the signal components are highly correlated while the noise components are independent.

In addition to the advantages present in combing sensor measurements of the same type, note that decision making often depends on several different aspects of the same situation. Animals often rely on several different senses to make decision. Pit vipers locate their prey by using their eyes and heat sensors located at the sides of their head. Consider how many senses are used by a human being when eating. For automatic systems to react reasonably to their environment they will often need to combine input from many separate sources.

Another problem that sensor fusion attacks is information overload. A correct decision is almost always a well informed and timely one. The amount of time needed to reach a decision increases rapidly with the amount of information available. Sensor fusion is necessary to combine information in a way that removes inconsistencies and presents clearly the best interpretation of measurements input from many individual sources. Through sensor fusion, we can combine measurements from several different sensors in order to combine different aspects of the environment into one coherent structure.

When done properly, sensor fusion combines input from many independent sources of limited accuracy and reliability to give information of known accuracy and proven reliability. In summary, the advantages of sensor fusion over single sensor processing are due to the redundancy, diversity and complementarity among multiple sensors.

- Redundancy is caused by the use of multiple sensors to measure the same entity. It is well known that redundancy reduces uncertainty. This can be appreciated from the fact that for multiple sensors, the signal related to the measured quantity is often correlated, whereas the uncertainty associated with each individual sensor tends to be uncorrelated.

- If multiple sensors are of different nature, they measure the same scene with different laws of physics, and we obtain physical sensor diversity. Another diversity, spatial diversity, which offers different viewpoints of the sensed environment simply by having sensors in different locations, also plays a very important role in multisensor fusion.

- Multiple sensors observe a subset of the environment space, and the union of these subsets makes up broader environment observation. In this way, we achieve data complementarity.

2.2 Classification of Sensor Fusion Techniques

There are several criteria to categorize current sensor fusion techniques. These criteria include types of sensor data, levels of representation, mathematical fusion algorithms and so on.

2.2.1 Classification by Types of Sensor Data

Sensor fusion may be performed on time series, redundant, and/or complementary sensor data. Time series fusion, by far the most common, allows for filtering of noisy sensor data and is commonly used in target tracking applications [10]. Redundant sensors acquire data in parallel, and allow comparisons to be made among simultaneous measurements. An example of this is the use of multiple ultrasonic range finders on a mobile robot for obstacle detection and avoidance. Complementary sensor fusion incorporates information about different physical aspects of the environment. Complementary sensor data is often used in recognition and world modeling tasks. This complementary case is of special interest to the applications we deal with in this dissertation.

2.2.2 Classification by Levels of Representation

Applications of multisensor fusion may be characterized by the level of representation given to data during the fusion process. Observational data may be combined, or fused, at a variety of levels — signal, pixel, feature, and symbol levels. Note that these levels of fusion are only a rough classification of representation possibilities, and in no way can capture the subtlety of numerous applications. For instance, pixel-based images such as those used in medical imaging may be treated as spatially discrete two-dimensional non-causal signals. Despite the obvious pixel-based representation of these signals, the mathematical techniques used to process and fuse these data are more closely related to signal-based techniques [64].

2.2.2.1 Signal-level fusion

Signal-level fusion refers to the combination of the signals of a group of sensors in order to provide a signal that is usually of the same form as the original signals but of greater quality. The signals from sensors can be modeled as random variables corrupted by uncorrelated noise, and the fusion process can be considered as an estimation procedure. As compared to the other types of fusion, signal-level fusion requires the greatest degree of registration between the sensory information. The fusion requires both temporal and spatial registration. The most common techniques for signal-level fusion consist of weighted averaging and Kalman filtering.

Weighted averaging method takes a weighted average of the composite signals, where the weights are determined by the estimated variances of the signals [20]. Kalman filtering uses the statistical characteristics of a measurement model to recursively determine estimates for the fused data [57].

2.2.2.2 Pixel-level fusion

Pixel-level fusion can be used to increase the information content associated with each pixel in an image formed through a combination of multiple images, thus it may result in more reliable segmentation and more discriminating features for further processing. Pixel-level fusion may be performed on time series images obtained from a single sensor or images obtained from multiple sensors which are measuring the same physical phenomena such as multiple visual image sensors or acoustic sensors. The fused image can be produced either through pixel-by-pixel fusion or through the fusion of associated local neighborhood of pixels in each of the component image. Many of the general multisensor fusion methods can be applied to the pixel-level fusion. Here we particularly point out three useful method for fusion at the pixel level — logical filters, mathematical morphology, and simulated annealing.

Logical filtering is one of the most intuitive method of fusing the data from two pixels. The choice of the logical operators is dependent on the fusion applications [3]. Mathematical morphological method [31] in image analysis transforms each pixels of an image through the use of a set of morphological operators. These operators are derived from the basis operations of set union, intersection, difference, and their conditional combinations. Lee employed binary morphology to fuse images from a pair of millimeter-wave radars operating at different frequencies [63]. Simulated annealing is a relaxation-based optimization technique. In image fusion applications, simulated annealing considers pixel values and their neighborhood as states of atoms or molecules in a physical system. An energy function is assigned to the physical system and determines its Gibbs distribution. General temperature reductions in the energy function are utilized to relax or anneal the physical system toward a global minimum energy state which corresponds to the maximum *a posteriori* estimate of the true image given a corrupted observational image. Landa and Scheff have applied simulated annealing for the pixel-level fusion of the images from two cameras in order to estimate depth [61].

2.2.2.3 Feature-level fusion

If the sensors are measuring different physical phenomena, then the sensor data must be fused at feature/symbol level. Using features to represent the sensory information not only reduces the complex of the processing procedure but also increases the reliability of the processing results. Feature-level fusion involves the extraction of representation features from multiple sensor observations. These features are matched to corresponding features in a symbolic world model. Typical features extracted from an image and used for fusion include edges and regions of similar intensity. When multiple sensors have similar features at the same location, the likelihood that the features are actually present can be significantly increased and thus fusion improves the processing accuracy.

Feature-level sensor fusion requires less stringent registration than signal- and pixel-level fusion. Techniques proposed for feature-level fusion consist of tie statistic, and Gauss-Markov estimation with constraints. When information from multiple sensors is being used for classification and decision purposes, tie statistic provides a means of measure so that perceived features of the environment can be compared to known features [28]. The tie statistic allows an unknown sample probability density function to be quickly and effectively classified. Pollard *et al.* proposed Gauss-Markov estimation together with geometric constraints for the feature-level fusion of multiple stereo views of a wireframe model of an object [82]. A covariance matrix is used to store information concerning the constraints and is used as a database for elementary geometric reasoning.

2.2.2.4 Symbol-level fusion

Symbol-level fusion can effectively integrated the information from multiple sensors at the highest level of abstraction. Symbol-level fusion is commonly employed in the applications where multiple sensors are of different nature or refer to different regions of the environment. The symbols used for fusion can be derived from the processing of the individual sensory information, or through symbolic reasoning processes that may make use of prior knowledge from a world model or sources external to the system. In one of the applications which will be addressed in the dissertation, the simultaneous detection of lane and pavement boundaries, the prior knowledge of the lane and pavement boundaries is that they are parallel concentric arcs under the flat earth assumption, while the likelihoods indicate the likely presence of certain lane or pavement boundaries in the environment.

The most common type of symbol-level fusion application is pattern recognition. Feature information is extracted from sensor data, defining a point in the feature space. This point may be mapped to a symbolic interpretation of the world based on that symbol's neighborhood in the feature space. Such a neighborhood function may be defined by probability theory [101], Dempster-Shafer's theory of evidence [88], fuzzy logic [51], neural networks [108], or other means.

The prevailing techniques for symbol-level fusion include Bayesian (Maximum *A Posteriori*) estimation, Dempster-Shafer evidential reasoning, and fuzzy set theory. Bayesian estimation combines sensory information according to the rules of probability theory. This multisensor fusion approach is the core method used in our work and will be discussed in details in Chapter IX, I, and XI. Dempster-Shafer reasoning technique for sensor fusion allows each sensor to contribute information at its own level of detail. Dempster-Shafer evidential reasoning is an extension of the Bayesian approach. It makes explicit any lack of information concerning a proposition's probability by separating firm belief for the proposition from just its plausibility. Valin *et al.* applied this technique for fusion of imaging and non-imaging sensor information for airborne surveillance [100]. Fuzzy logic is a type of multiple-valued logic. It allows the uncertainty in multisensor fusion to be directly represented in the fusion process by allowing each position, as well as the actual implication operator, to be assigned a real number from 0.0 to 1.0 to indicate its degree of truth. Fuzzy logic technique has been used to fuse information for an on-line and real time vehicle detection system [46].

A comparison of fusion levels is given in Table 1, adapted from [67].

2.2.3 Classification by Mathematical Foundation

Independent of the level of representation used, a variety of popular mathematical techniques for sensor fusion appear in the literature. These methods generally perform

Characteristics	*Signal-level*	*Pixel-level*	*Feature-level*	*Symbol-level*
Type of sensory information	single or multi-dimensional signals	multiple images	features extracted from signals/images	symbol representing decision
Representation level of information	low	low to medium	medium	high
Model of Sensory Information	random variable corrupted by noise	stochastic process on image or pixels with multidimensional attributes	geometrical form, orientation, position, and temporal extent of features	symbol with associated uncertainty measure
Degree of registration				
spatial:	high	high	medium	low
temporal:	high	medium	medium	low
Means of registration				
spatial:	sensor coalignment	sensor coalignment, shared optics	geometrical transformations	spatial attributes of symbol, if necessary
temporal:	synchronization or estimation	synchronization	synchronization	temporal attributes of symbol, if necessary
Fusion method	signal detection and estimation	image estimation or pixel attribute combination	geometrical and temporal correspondence, feature attribute combination	logical and statistical inference
improvement due to fusion	reduction in expected variance,improved detection	increase in performance of image processing tasks	reduced processing, richer, more accurate feature data	increase in truth or probability values

Table 2.1: Comparison of Fusion Levels

a data-reduction mapping from multiple inputs to a smaller number of outputs. Inputs may be raw sensor measurements, pixel values, extracted features, or signal estimators; outputs may be estimated state, recognized objects or events, enhanced features, etc. An important aspect of each technique is the way in which it models uncertainty in sensor information. Since it is impossible to introduce all mathematical fusion techniques in this dissertation, we just present two mathematical techniques that are related to our work — Bayesian Inference and Dempster-Shafer Theory of Evidence.

2.2.3.1 Probability and Bayesian Inference Techniques

Probability theory, which measures the likelihood of an event, was first developed by Blaise Pascal in the seventeenth century as a means of solving gambling problems. It was later advanced by Thomas Bayes in the eighteenth century and by Pierre de Laplace in the early nineteenth century. Probability-based inference techniques have withstood mathematical scrutiny for hundreds of years, and are the foundation of most sensor fusion applications.

The basic assumption of probability-based sensor fusion is that the uncertainly in sensor information may be modeled by uncorrelated random noise. Decision and estimation rules based on probability include mean-square error and maximum *a posteriori* (MAP), which minimizes the probability of error, Bayes risk, which minimizes the probable cost of error, and maximum likelihood, which estimates a parameter without assuming a prior probability distribution for values of the parameter.

We will discuss this in Chapters VI and VII.

2.2.3.2 Dempster-Shafer Theory of Evidence

The theory of evidential combination was proposed by Dempster [21] and extended by Shafer [88]. The theory models uncertainty as belief in one or more propositions or ignorance. The set of all possible mutually exclusive hypothesis is called the frame of discernment, Θ. Let 2^{Θ} denote the set of all subsets of Θ. The *basic probability assignment* (bpa) of a set, A, is a function, usually denoted by m, which assigns an evidential weight to the set, such that

- $m(\phi) = 0$, where ϕ is the empty set.
- $\sum_{A \subseteq \Theta} m(A) = 1$.

Let $A \subseteq \Theta$. A is assigned a bpa representing a proposition's ignorance of which hypothesis is correct. This representation of evidential weight is what makes Dempster-Shafer reasoning different from probability. For instance, let $H_1, H_2, \cdots, H_n$ be n exclusive hypothesis. In probability theory, the probabilities of $H_1, H_2, \cdots, H_n$ sum to one, that is,

$$\sum_{i=1}^{n} Pr(H_i) = 1 \tag{2.1}$$

In the Dempster-Shafer Theory, the bpa belief mass for $H_1, H_2, \cdots, H_n$ may sum to less than one:

$$\sum_{i=1}^{n} m(H_i) \leq 1 \tag{2.2}$$

where, instead,

$$\sum_{i=1}^{n} m(H_i) + m(A) = 1 \tag{2.3}$$

where $A \subseteq \Theta$.

Subsets $A \subseteq \Theta$ with $m(A) > 0$ are called the *focal elements of m*. Two bpa's m_1 and m_2 can be combined using Dempster's rule [21] and evidence gathering accumulates belief in each hypothesis set A is obtained:

$$m(A) = \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{\sum_{A_1 \cap A_2 \neq \phi} m_1(A_1) m_2(A_2)} \tag{2.4}$$

The belief of a hypothesis set H_k is defined as

$$\text{bel}(H_k) = \sum_{A \subseteq H_k} m(A) \tag{2.5}$$

The plausibility of the hypothesis set H_k is defined as

$$\text{pls}(H_k) = 1 - \text{bel}(\sim H_k) \tag{2.6}$$

where $\sim H_k$ denotes the complementary hypothesis of H_k.

As evidence is gathered, the rule of combination increases the belief and plausibility of those hypotheses supported by the evidence, and decreases the bpa of the ignorance factor. (If no weight is assigned to ignorance, Dempster's rule of combination reduces to basic probability theory.) It is intuitively appealing to model ignorance in this way, especially when evidence is supplied from human experts.

An illustration of the Dempster-Shafer theory applied to target detection is given by Lee *et al.* [65]. Murphy discusses Dempster-Shafer theory in terms of its utility for sensor fusion for autonomous mobile robots [77]. In more quantitative applications, Lee and Leahy [64] compare the use of Dempster-Shafer, MAP, and maximum likelihood algorithms for multi-sensor image segmentation problems, showing a marginal improvement in the results of the Dempster-Shafer implementations over MAP. (ML ignores the prior information introduced by a Markov Random Field used in the other methods and, therefore, resulted in a much higher misclassification rate.)

However, Cheeseman [12] argues that these features can be provided equally well by Bayesian techniques. For instance, it is possible to measure the probability of error in a probability estimation; this value decreases with the addition of evidence just as the Dempster-Shafer "ignorance" value does. Given the strong mathematical history of probability theory, Dempster-Shafer reasoning remains controversial, but it has nevertheless enabled some significant applications of data fusion in uncertainty.

CHAPTER III

Minimum Spanning Tree and Rényi Entropy

3.1 Minimum Spanning Tree

Discrete points distributed in space can be characterized using a representational network called its minimum spanning tree (MST). This MST technique aims to quantify spatial dot patterns by revealing hidden nearest-neighbor correlations. The MST approach has recently been applied in several research areas, such as VLSI circuit layout and network provisioning [44, 84], two sample matching [29], pattern recognition [15], clustering [81, 107], nonparametric regression [5], and testing for randomness [41].

Given a set $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ of n points in $\mathbb{R}^d$, a spanning tree $\mathcal{T}$ is a connected acyclic graph which passes through all coordinates associated with the point set $\mathcal{X}_n$ (see Figure 3.1.1 for an illustration of a spanning tree). $\mathcal{T}$ is specified by an ordered list of edges e_{ij} connecting certain pairs (X_i, X_j), $i \neq j$, along with a list of edge adjacency relations. The edges e_{ij} connect all n points such that there are no paths in the graph that lead back to any given point.

For a given edge weight exponent $\gamma \in (0, d)$, the power weighted length $L_{\mathcal{T}}(\mathcal{X}_n)$ of a specified tree $\mathcal{T}$ is the sum of all edge lengths raised to power γ,

$$L_{\mathcal{T}}(\mathcal{X}_n) = \sum_{e_{ij} \in \mathcal{T}} |e_{ij}|^{\gamma} \,. \tag{3.1}$$

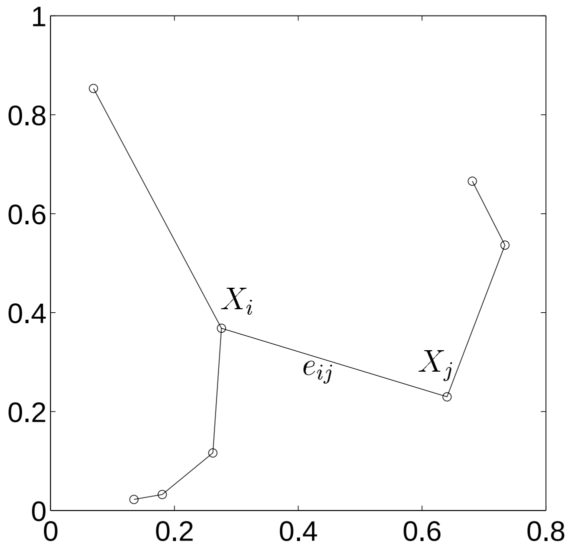


Figure 3.1.1: An MST example

The minimum spanning tree $\mathcal{T}^*$ is the spanning tree which minimizes the total edge weight $L_{\mathcal{T}}(\mathcal{X}_n)$ of the graph among all possible spanning trees over the given vertices, i.e., the power weighted length $L^*(\mathcal{X}_n)$ of the minimum spanning tree $\mathcal{T}^*$ is

$$L^*(\mathcal{X}_n) = \sum_{e_{ij} \in \mathcal{T}^*} |e_{ij}|^\gamma = \min_{\mathcal{T}} L_{\mathcal{T}}(\mathcal{X}_n) \tag{3.2}$$

The tree shown in Figure 3.1.1 is indeed a minimum spanning tree over the given vertices.

It is evident that the MST representation of a point process is naturally translation and rotation invariant. This invariance is important for image processing applications where patterns of interest may be articulated at arbitrary orientation and spatial positions, e.g., as occurs in automated pattern matching, radar target detection, and industrial inspection. The MST can be constructed in time polynomial in n, the number of vertices.

For any subset $\mathcal{X}_{n,k}$ of k ($0 < k \leq n$) points in $\mathcal{X}_n$ we define the k-point MST, denoted as $\mathcal{T}_{\mathcal{X}_{n,k}}$, which spans $\mathcal{X}_{n,k}$. For fixed k, the elements of the subset $\mathcal{X}_{n,k}$ are distinct and there are $\begin{pmatrix} n \\ k \end{pmatrix}$ possible k-point subsets of $\mathcal{X}_n$. In Figure 3.1.2 we give two examples of k-point MST's over different subsets of the original point set. Figure 3.1.2(a) shows the k-point MST on the subset $\mathcal{X}_{n,k} = \{X_3, X_4, X_5, X_6, X_7, X_8\}$, while Figure 3.1.2(b) shows the k-point MST on the subset $\mathcal{X}'_{n,k} = \{X_1, X_2, X_4, X_5, X_6, X_7\}$. These two examples show that for different subsets, the resultant k-point MST's have different lengths. The k-point MST in Figure 3.1.2(a) has the minimal length among all possible k-point MST's and the k-point MST in Figure 3.1.2(b) has the maximal length.

The minimum k-point spanning tree (k-MST) is defined as that k-point MST which has minimal length. Thus the k-MST spans a subset $\mathcal{X}^*_{n,k}$ defined by

$$L(\mathcal{X}^*_{n,k}) = \min_{\mathcal{X}_{n,k}} L^*(\mathcal{X}_{n,k}). \tag{3.3}$$

The k-point MST shown in Figure 3.1.2(a) is actually the k-MST for $n = 8$ and $k = 6$.

3.2 Limit Theorem and Rate of Convergence for MST Length Functionals

3.2.1 Limit Theorem for MST Length Functionals

Given a set $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ of n points in $\mathbb{R}^d$, Steele has proved that the length of the MST over $\mathcal{X}_n$ has the following asymptotic property [96]

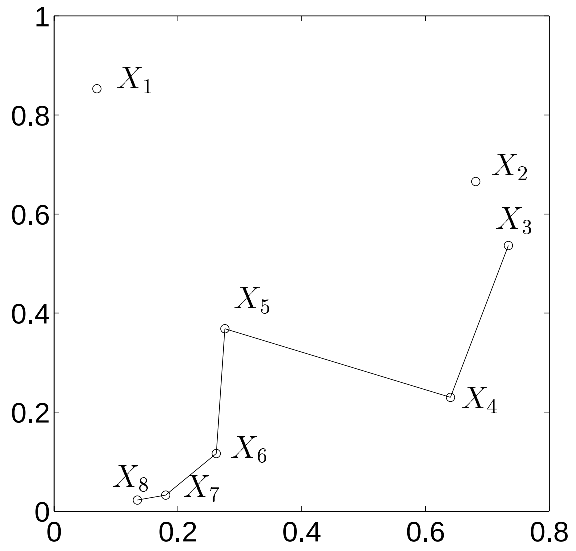
$$\lim_{n\to\infty} \frac{L(\mathcal{X}_n)}{n^{\frac{d-\gamma}{d}}} = \beta_\gamma \int (f(x))^{\frac{d-\gamma}{d}} dx \qquad (a.s.) \tag{3.4}$$

where β_γ is a constant which only depends on the edge weight exponent γ. In particular, β_γ is independent of the distribution of the X_i's.

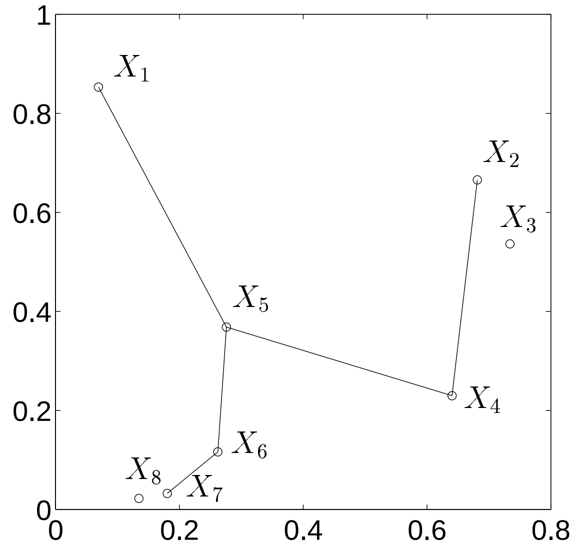
3.2.2 Rate of Convergence for MST Length Functionals on Uniform Samples

In [86] Redmond and Yukich derived the rate of convergence for MST length functionals L for independent identically distributed (i.i.d.) uniform random vectors $U_1, \ldots, U_n$ in $[0, 1]^d$ for $\gamma = 1$. Their result for the rate of convergence is given in the following theorem.

Theorem 3.2.1 *Given a set $\mathcal{U}_n = \{U_1, U_2, \ldots, U_n\}$ of n points in $[0, 1]^d$ with the uniform distribution, the convergence rate for the MST length functional L is as follows:*



(a) k-point MST spanning over a subset $\mathcal{X}_{n,k} = \{X_3, X_4, X_5, X_6, X_7, X_8\}$



(b) k-point MST spanning over a subset $\mathcal{X}'_{n,k} = \{X_1, X_2, X_4, X_5, X_6, X_7\}$

Figure 3.1.2: Examples for k-point MST's

(a) if $d \geq 3$,

$$\left| E[L(U_1, \ldots, U_n)] - \beta_1 n^{\frac{d-1}{d}} \right| = O(n^{\frac{d-2}{d}}) \tag{3.5}$$

(b) if $d = 2$,

$$\left| E[L(U_1, \ldots, U_n)] - \beta_1 n^{\frac{1}{2}} \right| = O(1) \tag{3.6}$$

3.2.3 Rate of Convergence for MST Length Functionals on Non-uniform Samples

In our work, we have derived a theorem establishing convergence rates for the MST length functional L for i.i.d. but non-uniform random vectors $X_1, X_2, \ldots, X_n$ in $[0,1]^d$. Our derivations closely follow the method that Redmond and Yukich have employed in [86]. Here we give the description of the results and some remarks. Please see Appendix for the proof of the following theorem.

Theorem 3.2.2 *Let $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ be a set of n i.i.d. realizations from the pdf $f(x)$ in $[0,1]^d$. Assume that for $\nu \in (0,1)$, f^ν is of bounded variation over $[0,1]^d$ and let v_1 be the total variation of $f^{\frac{d-1}{d}}$ and v_2 the total variation of $f^{\frac{d-2}{d}}$ over $[0,1]^d$. Also assume that $\int_{\{x: f(x)>0\}} f^{-\frac{1}{d}}(x)\, dx < \infty$ Let $\mathcal{Q}^m = \{Q_i\}_{i=1}^{m^d}$ be a uniform partition of $[0,1]^d$ into m^d subcubes Q_i with edges parallel to the axes and with edge length m^{-1}. Then for fixed resolution factor m, the convergence rate of the MST length functional over $\mathcal{X}_n$ is*

$$\begin{aligned} &\left| \frac{EL(X_1, \ldots, X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int f^{\frac{d-1}{d}}(x) dx \right| \\ &\quad \leq K_1 n^{-\frac{1}{d}} m \int f^{\frac{d-2}{d}}(x) dx + K_2 n^{-\frac{d-1}{d}} m^{d-1} + K_1 n^{-\frac{1}{d}} m^{-d+1} \sum_{i=1}^{m^d} v_2(Q_i) \\ &\qquad + \beta_1 m^{-d} \sum_{i=1}^{m^d} v_1(Q_i) + o\left(n^{-\frac{d-1}{d}}\right) \end{aligned} \tag{3.7}$$

where $K_1, K_2,$ and β_1 are constants. Namely, K_1 and K_2 are dependent on the subadditivity, superadditivity and continuity constants of the MST length functional (see Appendix *for the definition of these constants), and* β_1 *is the asymptote of the MST length functional for i.i.d. uniform random vectors* $\mathcal{U}_n = \{U_1, U_2, \ldots, U_n\}$ *in* $[0,1]^d$*,*

$$\lim_{n\to\infty} \frac{L(\mathcal{U}_n)}{n^{\frac{d-1}{d}}} = \beta_1 \qquad (a.s.) \tag{3.8}$$

Remarks:

The right hand side (RHS) of (3.9) can be rewritten in terms of the ratio $\frac{n}{m^d}$, equal to the number of points per cell, and (3.7) becomes

$$\begin{aligned} &\left| \frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int f^{\frac{d-1}{d}}(x)dx \right| \\ &\quad \le K_1 \left(\frac{1}{n/m^d}\right)^{\frac{1}{d}} \int f^{\frac{d-2}{d}}(x)dx + K_2 \left(\frac{1}{n/m^d}\right)^{\frac{d-1}{d}} \\ &\qquad + K_1 \left(\frac{1}{n/m^d}\right)^{\frac{1}{d}} \left(m^{-d}\sum_{i=1}^{m^d} v_2(Q_i)\right) + \beta_1 \left(m^{-d}\sum_{i=1}^{m^d} v_1(Q_i)\right) + o\left(\frac{1}{n/m^d}\right)^{\frac{d-1}{d}} \end{aligned} \tag{3.9}$$

Thus the approximation error on the left hand side (LHS) of (3.9) converges to zero as $n/m^d \to \infty$, and $n, m \to \infty$. For $d > 2$, the dominant term in the convergence rate is the first term on RHS of (3.7) which converges to zero at rate $n^{-\frac{1}{d}}$ and with rate constant equal to the Rényi entropy of order $\frac{d-2}{d}$ of f plus a term $K_1 \left(m^{-d}\sum_{i=1}^{m^d} v_2(Q_i)\right)$ which decreases in m. When m is fixed, there is an irreducible error (last term of RHS of (3.9)) which is proportional to the total variation $v_1(Q_i)$ of $f^{\frac{d-1}{d}}$ over the cells of volume m^{-d}. This error can be interpreted as the bias of the MST Rényi entropy estimator when a greedy approximation algorithm is used to construct the MST over each cell and the MST's are patched together. Thus the bound (3.9) illustrates the nature of the fundamental bias-variance tradeoff in the MST entropy estimator: for increasing number of cells the bias term on RHS in (3.9) decreases but the other terms in this expression increase unless n is also increased. Hence these other terms can be viewed as variance terms. These variance terms are the factors $\frac{1}{n/m^d}$ to various powers multiplied by bounded quantities.

- The convergence rate can be considered for two cases:

  ① When $d = 2$,

$$\left| \frac{EL(X_1, \ldots, X_n)}{n^{\frac{1}{2}}} - \beta_1 \int f^{\frac{1}{2}}(x)\, dx \right| \leq (K_1 + K_2) n^{-\frac{1}{2}} m$$
$$+\beta_1 m^{-2} \sum_{i=1}^{m^2} v_1(Q_i) + K_1 n^{-\frac{1}{2}} m^{-1} \sum_{i=1}^{m^2} v_2(Q_i) + o\left(\frac{1}{\sqrt{n}}\right) \quad (3.10)$$

  ② When $d > 2$,

$$\left| \frac{EL(X_1, \ldots, X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int f^{\frac{d-1}{d}}(x) dx \right|$$
$$\leq K_1 n^{-\frac{1}{d}} m \int f^{\frac{d-2}{d}}(x) dx + K_2 n^{-\frac{d-1}{d}} m^{d-1} + K_1 n^{-\frac{1}{d}} m^{-d+1} \sum_{i=1}^{m^d} v_2(Q_i)$$
$$+\beta_1 m^{-d} \sum_{i=1}^{m^d} v_1(Q_i) + o\left(n^{-\frac{d-1}{d}}\right) \quad (3.11)$$

- In both cases $d = 2$ and $d \geq 2$, if m is sufficiently large so that the underlying density f is close to piecewise constant, the total variations v_1 and v_2 of $f^{\frac{d-1}{d}}$ and $f^{\frac{d-2}{d}}$ will be small compared to the first terms in the right hand side of (3.10) and (3.11). That is, the error between the expectation of the MST length functional and its asymptote is reduced at the rates $n^{-\frac{1}{2}}$ for $d = 2$ and $n^{-\frac{d-1}{d}}$ for $d > 2$.

3.3 Rényi Entropy and Rényi Information Divergence

Rényi entropy [87] is a more general entropy than Shannon entropy. The Rényi entropy $H_\alpha(f)$, also called the α-entropy, for a continuous pdf f is defined as

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \int f^\alpha(x)\, dx, \quad (3.12)$$

for $0 < \alpha < \infty, \alpha \neq 1$. The parameter α is called information order.

The Shannon entropy function

$$H(f) = -\int f(x) \log f(x)\, dx \tag{3.13}$$

is the limiting case of Rényi entropy when the order α approaches 1, namely,

$$H(f) = \lim_{\alpha \to 1} H_\alpha(f). \tag{3.14}$$

If we take the limit of (3.12) as $\alpha \to 0$, we obtain the logarithm of the volume of the support set,

$$H_0(f) = \log(\mu\{x : f(x) > 0\}). \tag{3.15}$$

Thus the zeroth order Rényi entropy gives the measure of the support set of the density f.

Rényi information divergence is a distance measure between densities. Given a test density f and a reference density f_0, the order α Rényi information divergence of f and the reference density f_0 is defined as

$$I_\alpha(f, f_0) = -\frac{1}{1-\alpha} \log \int \left(\frac{f(x)}{f_0(x)} \right)^\alpha f_0(x) dx \tag{3.16}$$

For any order α, the information divergence takes on its minimum value (equals zero) if and only if $f = f_0$ (a.e.). $I_\alpha(f, f_0)$ reduces to the Rényi entropy $H_\alpha(f)$ when f_0 is equal to a uniform density over $[0,1]^d$. There are two special cases of interest for the order α. For $\alpha = \frac{1}{2}$ the Rényi information divergence becomes the log Hellinger distance squared where the Hellinger distance is defined by

$$d_H(f, f0) = \frac{1}{2} \int \left(\sqrt{f(x)} - \sqrt{f_0(x)} \right)^2 dx \tag{3.17}$$

For $\alpha \to 1$, the Rényi information divergence approaches the Kullback-Liebler divergence,

$$D(f \| f_0) = \int f(x) \log \frac{f(x)}{f_0(x)}\, dx \tag{3.18}$$

Since Rényi introduced this generalized entropy and information divergence in the early sixties, the Rényi entropies and information divergences have been successfully used in the information theory and statistics [6, 58].

There is one critical property for fractional order Rényi entropies and information divergences. For $\alpha \in (0,1)$, the order α Rényi information divergence always exists. This is a desirable property for a distance measure between two densities. In our image registration application, we shall take the advantage of Rényi entropies of fractional orders.

3.4 Estimation of Rényi Entropy with MST

Suppose $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ is a random sample from the continuous density f, and $L^*(\mathcal{X}_n)$ denotes the power weighted length of the minimum spanning tree over $\mathcal{X}_n$ (3.1). Let $\alpha = \frac{d-\gamma}{d}$. By the asymptotic property of the MST length functional (3.4), we obtain an estimator of Rényi entropy from the total edge weight of the MST,

$$\hat{H}_\alpha(f) = \frac{1}{1-\alpha}\left[\log \frac{L^*(\mathcal{X}_n)}{n^\alpha} - \log \beta_\gamma\right] \tag{3.19}$$

It follows directly from the result of [96] that the estimate $\hat{H}_\alpha$ using MST length is a strongly consistent estimator of H_α.

Figures 3.4.3 and 3.4.5 are an example which illustrates the use of minimum spanning trees as entropy discriminant between two different distributions. First let us define a torus density as

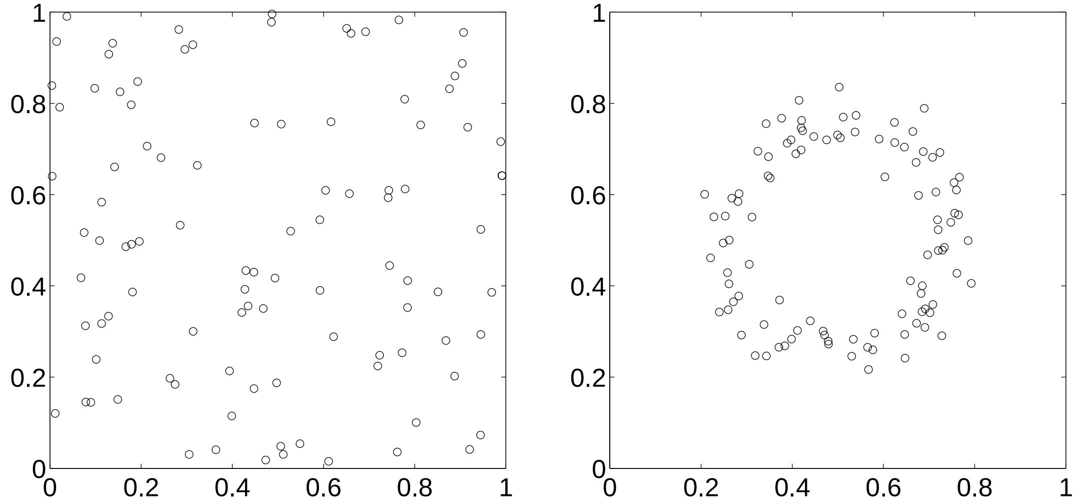
$$f(x) = c \exp\left\{-\frac{1}{2} 625 \ \left(\| \ x - [0.5 \quad 0.5] \ \| \ -0.25\right)^2\right\} \tag{3.20}$$

where c is a normalizing constant, $\|x\| = \sqrt{x_1^2 + x_2^2}$ for $x = (x_1, x_2)$. The constant contours of this density are circles for which the maximum contour is a circle of radius 0.25 and center $[0.5, 0.5]$. The other contours specify an annulus.

We study two different distributions on the unit square $[0, 1]^2$. The results are shown in Figure 3.4.3. The left column corresponds to a uniform density while the right column corresponds to a torus density specified as (3.20). In Figures 3.4.3(a) and (b) we plot single realizations of 100 random samples from the uniform and torus densities, respectively. Figures 3.4.3(c) and (d) are the corresponding MST's for these realizations. Note that for these realizations the overall length of MST for the uniform sample points is larger than that of the more concentrated torus sample points. According to our calculation, the total length of the MST spanning the uniform samples demonstrated in Figure 3.4.3(a) is 6.61, while the total length of the MST spanning the torus samples demonstrated in Figure 3.4.3(b) is 3.19.

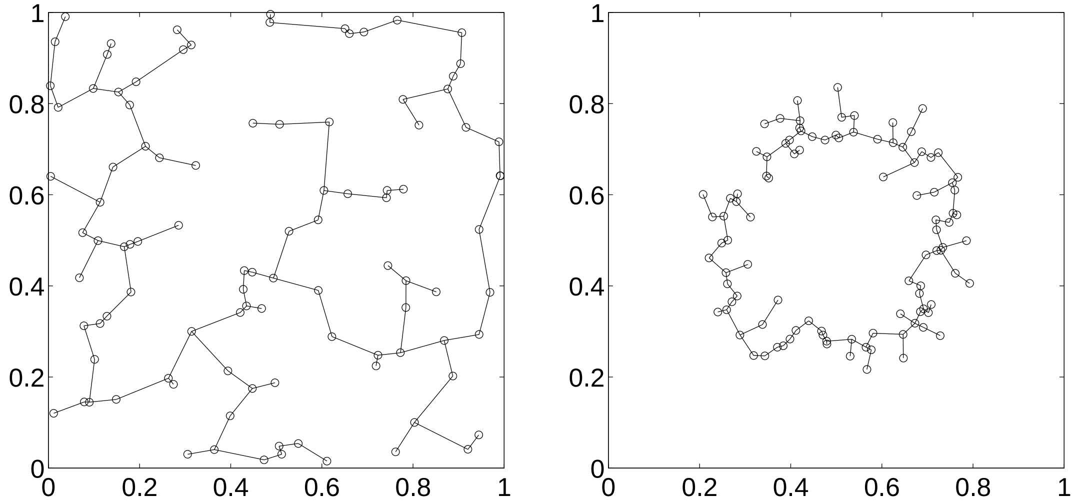
The mean length of the MST versus the number of sample points for each of the distributions is shown in Figures 3.4.4(a) and (b), computed on the basis of a large number of independent simulations of the two densities. The x-axis stands for the number of sample points generated from the uniform or torus density, and the y-axis stands for the mean value of the total length of the MST spanning the sample points.

Let n denote the number of sample points generated from a certain density. Note that for large n the mean length curves appear to increase with sub-linear rates and the rate constants depend on the underlying distribution of the random samples. Figure 3.4.5(a) shows the direct comparison of these two mean length curves as a function of n. Figure 3.4.5(b) shows the length curves normalized by $\sqrt{n}$ and transformed by $2\log(\cdot)$. It is observed that for both the uniform and the torus distributions the normalized and transformed length of the MST converges to two different constant levels. Furthermore, the asymptote for the uniform distribution is larger than that for the torus distribution. In fact, as was remarked in Hero and Michel [39], the difference between the asymptotes is equal to the difference between the Rényi entropies of order $1/2$ of the respective distributions.



(a) Random sample from uniform density

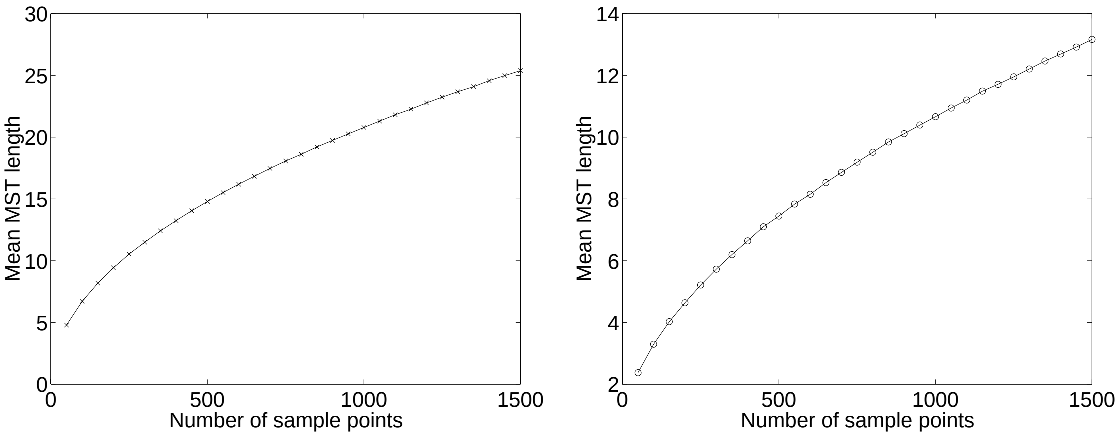
(b) Random sample from torus density



(c) MST over uniform sample

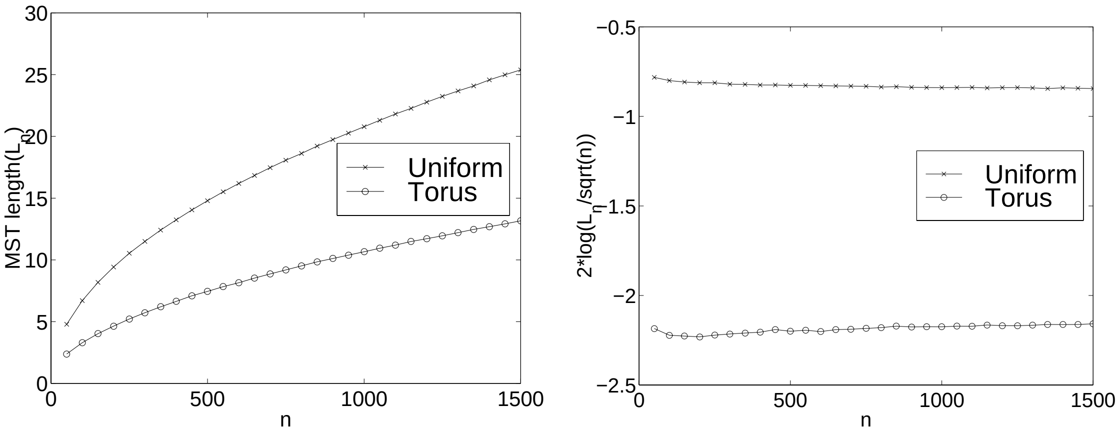
(d) MST over torus sample

Figure 3.4.3: 2D uniform and torus sample example



(a) Mean MST length for uniform samples (b) Mean MST length for torus samples

Figure 3.4.4: Mean MST length for uniform and torus samples



(a) (b)

Figure 3.4.5: Comparison of the MST length of uniform and torus samples

This example illustrates that the MST provides a consistent estimator of Rényi entropy (3.19).

3.5 Robustification of Rényi Entropy Estimator by k-MST

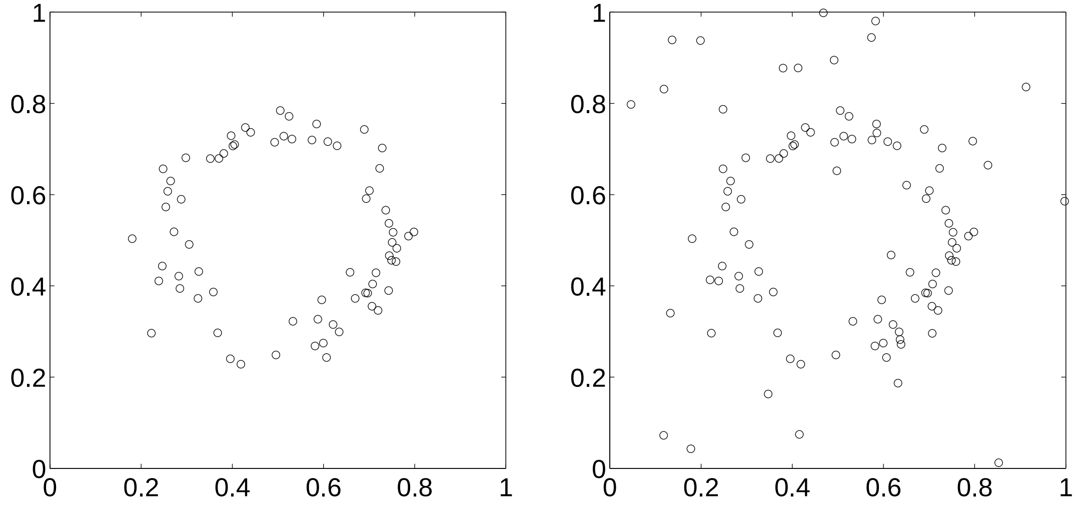
Now suppose that $\mathcal{X}_n$ is a random sample from a contaminated density

$$f = (1-\epsilon)f_s + \epsilon f_n,\ 0 \le \epsilon < 1 \tag{3.21}$$

where f_s is the density of interest (generating the signal) and f_n is the noise density (generating noise or outliers). Let f_s be the torus density specified by (3.20). In Figure 3.5.6 we show that the MST is sensitive to outlier contamination. Figure 3.5.6(a) shows 70 samples realized from the signal density f_s. The MST through the 70 samples is plotted in Figure 3.5.6(c). Figure 3.5.6(b) shows the sum of the 70 signal samples from f_s and 30 uniformly distributed noise samples over $[0,1]^2$. The MST through this contaminated signal is plotted in Figure 3.5.6(d). The MST in Figure 3.5.6(c) captures the shape of the uncontaminated torus density, and thus its length could be used to provide a reliable Rényi entropy estimator. But the MST in Figure 3.5.6(d) is severely influenced by the contamination of the noise and the MST length can only give erroneous entropy estimate. Thus the MST length function is not robust to noise.

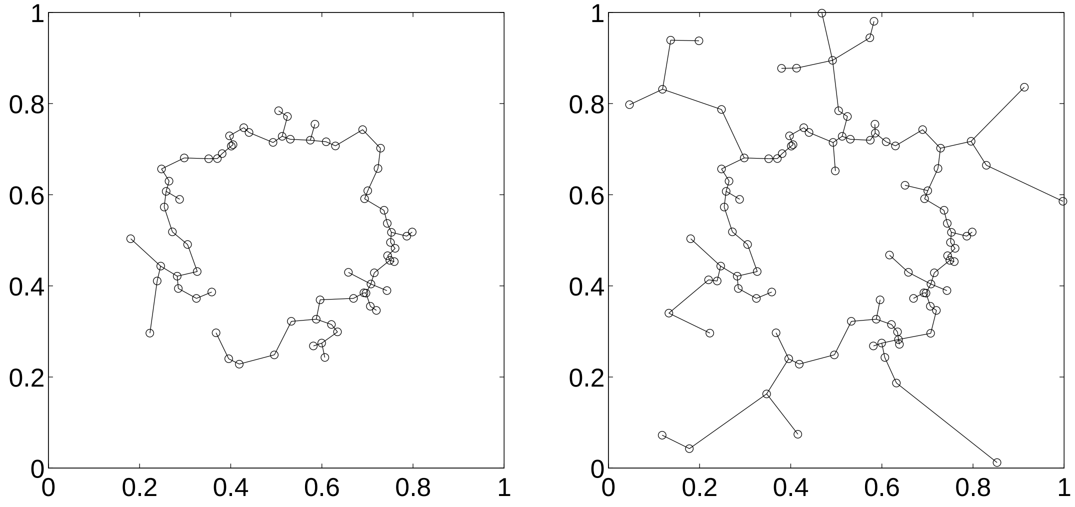
A solution to this lack of robustness is to utilize the k-MST technique to eliminate the outliers and then apply the MST entropy estimator.

For a uniform distributed noise, the noise tends to produce points that are further away from their “nearest neighbors” than points from f_s. Let $k = (1-\epsilon)n$. Then the k-MST will span most signal points and get rid of noise points (Figure 3.5.7). Figure 3.5.7(a) shows the k-MST over the contaminated samples with $k = 30$. The points which are not spanned by the k-MST are considered as outliers, and in this way we have successfully



(a) signal samples from f_s

(b) Contaminated by noise f_n

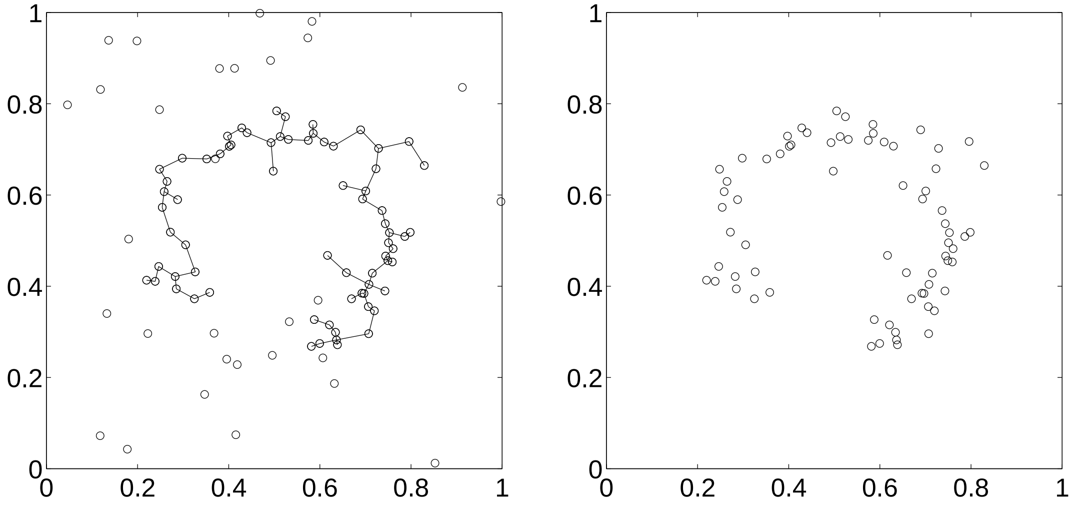


(c) MST over signal samples

(d) MST over contaminated samples

Figure 3.5.6: Signal, contaminated signal, and their corresponding MST's

removed the noise points (Figure 3.5.7(b)) and the k-MST achieves robustness to uniform noise [39, 40].



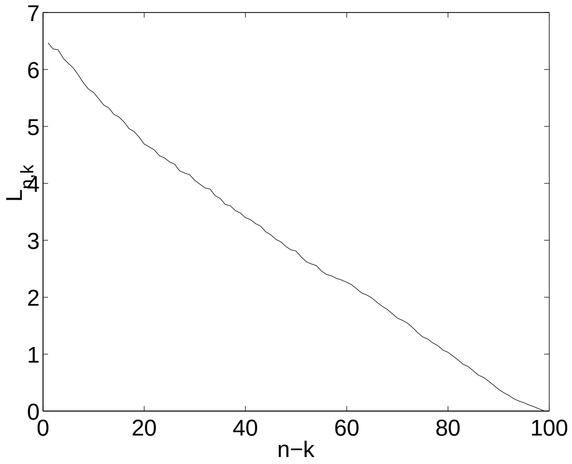
(a) k-MST over contaminated signal

(b) After removing outliers

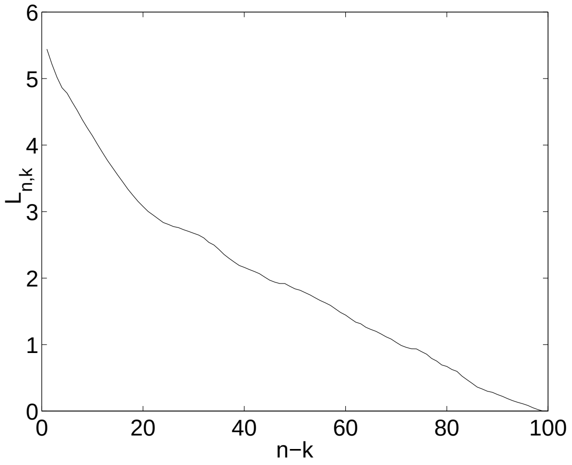
Figure 3.5.7: Use k-MST to remove outliers

In most cases, ϵ is unknown. If we remove less points than necessary, the outliers will still affect the construction of MST; however, if we remove too many more points than necessary, some signal points will be eliminated and thus the resulting MST will become vulnerable. The key to a practical k-MST robustification algorithm is to accurately estimate the appropriate number of points we want to reject.

Consider the k-MST length $L(\mathcal{X}^*_{n,k})$ with respect to the number of rejected points $n - k$. In Figure 3.5.8 the k-MST length is plotted for realizations of two distributions. Figure 3.5.8(a) corresponds to the uniform distribution over $[0, 1]^2$ and Figure 3.5.8(b) corresponds to signal samples generated from the torus distribution contaminated by the uniform noise. It is evident that the k-MST length curve for uniform sample (shown in Figure 3.5.8(a)) decreases linearly as $n - k$ increases, i.e., as k decreases. It has been proved that the mean k-MST length curve for uniform density is linear with respect to the number of rejected points $n - k$ [38]. On the other hand, the k-MST length curve for the contaminated sample from the torus density (Figure 3.5.8(b)) appears to be separated into two segments such that the left segment is close to be linear. The break point is approximately $n - k = 30$.



(a) k-MST length vs. number of rejected points for uniform samples



(b) k-MST length vs. number of rejected points for contaminated samples

Figure 3.5.8: k-MST length vs. number of rejected points

The estimate of the break point can be implemented by approximating the k-MST length curve by two piecewise linear curves. The least mean squared error measure is served as the stopping criterion. The intersection of the two linear curves is the estimate of the break point. With this method, we are able to identify the best estimate $\hat{\epsilon}$ of ϵ. In this example, $\hat{\epsilon} = 0.22$. Let $\hat{k} = (1 - \hat{\epsilon})n$. The corresponding $\hat{k}$-MST is shown in Figure 3.5.9. From the figure we notice that the remaining samples can characterize the signal density f_s. Therefore the length $L(\mathcal{X}^*_{n,\hat{k}})$ of the $\hat{k}$-MST gives a provably robust estimate of the Rényi entropy of f_s when used in (3.19) with n replaced by $\hat{k}$.

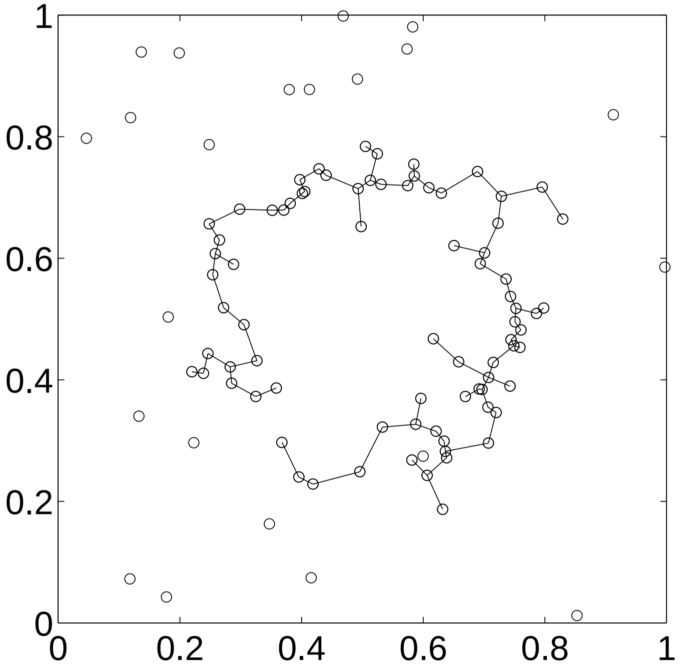


Figure 3.5.9: Use k-MST to remove outliers

Hero and Michel [39] proposed a greedy approximate k-MST construction method and proved that the k-MST length indeed provides a robust entropy estimator for noisy signals. We shall describe their greedy algorithms and the limit theorem in the next section.

3.6 Limit Theorem of k-MST greedy approximation

Since the computation of the exact minimal k-MST has a complexity which is exponential in the number of points n, several researchers have proposed polynomial-time approximations and have investigated the approximations' asymptotic properties [85, 39]. Hero and Michel give a generalized approximation algorithm constructing graphs in d dimensions, $d > 1$, in [39]. Suppose that the point samples are drawn from distributions with support $[0,1]^d$. Their generalized algorithm is described in the following.

The greedy approximation algorithm is divided into three steps:

① Uniformly partition $[0,1]^d$ into m^d cells Q_i of resolution $1/m$. Let $\mathcal{Q}^m = \{Q_i\}_{i=1}^{m^d}$.

② Find the smallest subset $B_k^m = \cup_i Q_i$ of partition elements containing at least k points.

③ Select the k points $\mathcal{X}_{n,k}$ out of the smallest subset such that $L(\mathcal{X}_{n,k})$ is minimized.

Step 3 requires finding a k-point minimal graph on a much reduced set of points, which is typically only slightly larger than k if m is suitably chosen. And hence this greedy approximation algorithm can be performed in polynomial time.

Define $\xi = k/n$. Let $\sigma(\mathcal{Q}^m)$ be the sigma algebra of $\mathcal{Q}^m$. If for any $C \in \sigma(\mathcal{Q}^m)$ satisfying $P(C) \geq \xi$ the set $A \in \sigma(\mathcal{Q}^m)$ has the following characteristic

$$P(C) \geq P(A) \geq \xi, \tag{3.22}$$

then A is called a *minimal resolution-*$1/m$ *set of probability at least* ξ. The class of all such sets is denoted $\mathcal{A}_\xi^m$. Hero and Michel have proved the following theorem in [39].

Theorem 3.6.1 *Let* $\mathcal{X}_n$ *be an i.i.d. sample from a distribution having density* $f(x)$. *Fix* $\xi \in [0,1], \gamma \in (0,d)$. *Let* $f^{\frac{d-\gamma}{d}}$ *be of bounded variation over* $[0,1]^d$ *and denote by* $v(A)$ *its* *total variation over a subset* $A \subseteq [0,1]^d$*. Then, the total edge weight* $L(\mathcal{X}_{n,k}^{Gm})$ *of a* k*-point graph constructed by the resolution-*$1/m$ *greedy algorithm satisfies*

$$\limsup_{n\to\infty} \left| \frac{L(\mathcal{X}_{n,k}^{Gm})}{n^{\frac{d-\gamma}{d}}} - \beta_\gamma \int_{A_\xi^m} f^{\frac{d-\gamma}{d}}(x)\,dx \right| < \delta, \tag{3.23}$$

where $A_\xi^m \in \mathcal{A}_\xi^m$ is any minimal resolution-$1/m$ set of probability at least ξ,

$$\begin{aligned} \delta &= 2m^{-d}\beta_\gamma \sum_{i=1}^{m^d} v(Q_i \cap \partial\mathcal{A}_\xi^m) + C_3(p_{\mathcal{A}_\xi^m} - \xi)^{\frac{d-\gamma}{d}} \\ &= O(m^{\gamma-d}) \end{aligned} \tag{3.24}$$

For any Borel set A in $[0,1]^d$ having $P(A) > 0$ define the conditional density

$$f(x|A) = \frac{f(x)}{P(A)} I_A(x)$$

where $I_A(x)$ is the indicator function of A. The Rényi entropy of $f(x|A)$ of order $\alpha \in (0,1)$ is defined as

$$H_\alpha(f|A) = \frac{1}{1-\alpha} \log \int f^\alpha(x|A)\,dx. \tag{3.25}$$

This is called the conditional Rényi entropy given A. Let A_0 be the probability-at-least-ξ Borel subset of $[0,1]^d$ which minimizes $H_\alpha(f|A)$,

$$H_\alpha(f|A_0) = \inf_{\{A\in\mathcal{B}:P(A)\geq\xi\}} H_\alpha(f|A) \tag{3.26}$$

Define the following function of $L(\mathcal{X}_{n,\lfloor\xi n\rfloor}^{Gm})$

$$\hat{H}_\alpha \triangleq \frac{1}{1-\alpha}\left(\log \frac{L(\mathcal{X}_{n,\lfloor\xi n\rfloor}^{Gm})}{(\lfloor\xi n\rfloor)^\alpha} - \log\beta_\gamma\right) \tag{3.27}$$

Then an immediate consequence of Theorem 3.6.1 is the following theorem.

Theorem 3.6.2 *Under the assumptions of Theorem 3.6.1* $\hat{H}_\alpha$ *is a strongly consistent estimator of the minimum conditional Rényi entropy* $H_\alpha(f|A_0)$ *of order* $\alpha \in (0,1)$ *as* $m,n \to \infty$.

CHAPTER IV

Image Registration Via Graph Matching

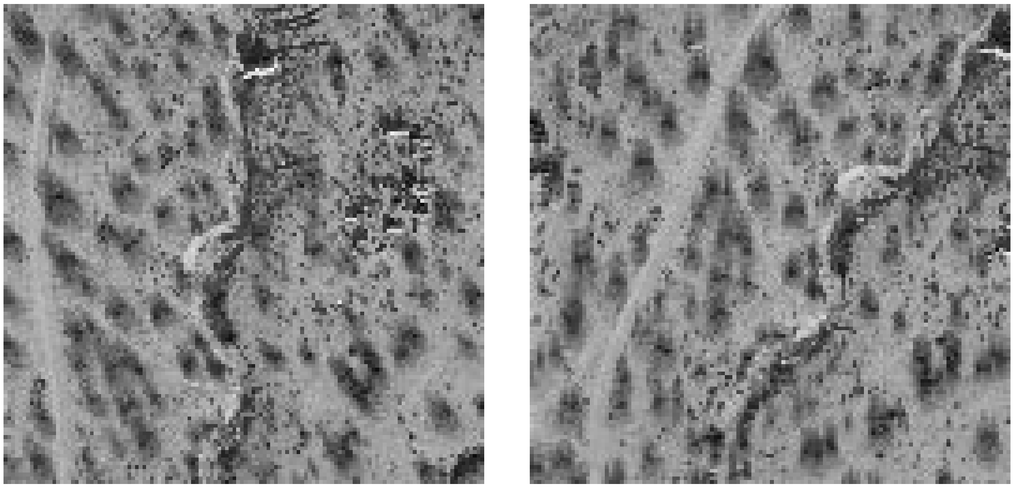
4.1 Image Registration Concepts and Formulation

Image registration refers to the process of aligning images so that their details overlap accurately. The images might be acquired at different times, from different viewing angles, or with different sensors. An example of multidate image registration is shown in Figure 4.1.1. Two images I_1 and I_0 (Figures 4.1.1(a) and (b)) are obtained in the Mojave desert at different times and with different viewing angles. Call I_1 as the test image and I_0 as the reference image. The registration result is achieved by rotating and translating the test image I_1 to align with the reference image I_0 (see Figure 4.1.1(c)).

Images are usually registered for the purpose of combining or comparing them. Therefore image registration is a vital first step in many image sequence analysis applications, e.g., fusing multiple sensor data. Actually image registration can be regarded as a special case of image fusion.

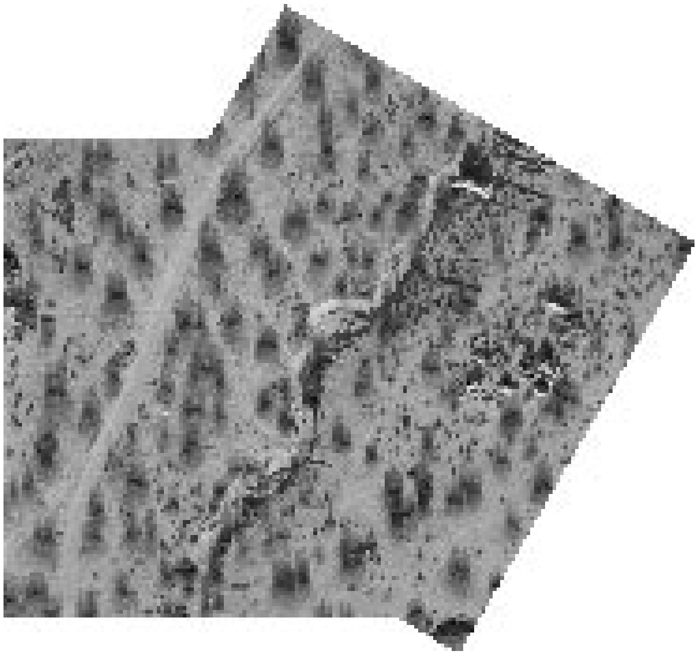
The need to register images has arisen in many practical problems in diverse fields:

- integrating information taken from different sensors,
- finding changes in images taken at different times or under different conditions,
- inferring three-dimensional information from images in which either the sensor or



(a) Image I_1

(b) Image I_0



(c) Registration result

Figure 4.1.1: A multidate image registration example

the objects in the scene have moved, and

- for model-based object recognition.

Registering two images always involves an operation that changes the locations and intensities of pixels in one image to line up with corresponding pixels in the other image. Image registration can be defined as a mapping between two images both spatially and with respect to intensity.

If we consider the images to be registered as two-dimensional arrays denoted by I_1 and I_0 where (x, y) are the coordinates of pixels and $I_1(x, y)$ and $I_0(x, y)$ each map to their respective intensity (or other measurement) values, then the mapping between images can be expressed as

$$I_0(x, y) = g(I_1(f(x, y))) \tag{4.1}$$

where f is a two-dimensional spatial-coordinate transformation, i.e., f is a transformation which maps two spatial coordinates, x and y, to new spatial coordinates x' and y',

$$(x', y') = f(x, y) \tag{4.2}$$

and g is a one-dimensional intensity transformation.

The intensity transformation is not always necessary, and often a simple lookup table determined by sensor calibration techniques is sufficient. For this reason we will ignore the intensity transformation in our work. Finding the parameters of the optimal spatial or geometric transformation is generally the key to any registration problem. The types of spatial transformation can range from a simple rigid transformation, containing only translations and rotations, to a fully elastic transformation, which locally deforms one image to fit the other.

Currently, the most common approach to registration is to extract a few outstanding characteristics of the data, which are called *control points, tie-points*, or *reference points*.

The control points in both images are matched in pairs and used to compute the parameters of a geometric transformation. Most available systems follow this registration approach, and because automated procedures do not always offer the needed reliability and accuracy, current systems assume some interactive choice of the control points. But such a point selection represents a repetitive, labor- and time-intensive task which becomes prohibitive for large amounts of data. Also, since the interactive choice of control points in some images, such as satellite images, is quite difficult, too few points, inaccurate points, or ill-distributed points might be chosen thus leading to large registration errors. And with the increase in the number of images collected every day from different sensors, automated registration of images has become a very important issue.

The aim of automatic registration is to estimate the geometric transformation in a manner robust to the local changes in the scene without human-machine interactive activities. In the past years, automatic image registration is broadly studied in computer vision [16, 79], remote sensing [74, 99], stereo vision [26, 45], and biomedical image analysis [9, 80].

Image registration methods vary depending on the choice of the following three components:

① **Feature space** $\mathcal{F}$

The feature space contains the representation information in the images that will be used for matching.

② **Search space** $\mathcal{T}$

The search space is the class of transformations that is capable of aligning the images.

③ **Dissimilarity metric** d

The dissimilarity metric describes the relative merit for each test. The minimal dissimilarity metric gives the optimal registration result.

Given the test image I_1 and the reference image I_0, F_1 and F_0 are feature vectors extracted from Images I_1 and I_0, respectively. For certain transformation $T \in \mathcal{T}$, we define the dissimilarity metric as the distance between the reference feature vectors F_0 and the transformed test feature vectors F_1,

$$d_T(I_1, I_0) = dist(T(F_1), F_0).$$

Under this framework, the problem of image registration becomes the problem of searching the optimal mapping T^* such that

$$T^* = \arg\min_{T \in \mathcal{T}} d_T(I_1, I_0)$$

Image registration may be broken into 3 distinct steps:

- extraction of alignment match features in both images,
- computation of the dissimilarity metric values between the transformed feature vectors and the reference feature vectors, and
- finding the optimal transformation which corresponds to the minimum dissimilarity metric value, and applying the mapping to move each image pixel from its current position to the corrected position.

A wide range of registration techniques has been developed for many different types of applications and data, such as mean squared alignment [42], Fourier descriptor matching [13], correlation registration [43], moment invariant matching [18], maximizing mutual information [17, 103] and others. Basically, each method exploits different image features of the object such as boundaries, moments or texture, which are extracted from the image, and uses them to solve the matching problem. Among these methods, the mutual information-based registration method has been broadly investigated in recent years and has yielded excellent experimental results. We will discuss this method in detail in the next section.

4.2 Image Registration Via Maximizing Mutual Information

In 1995, Viola et al. [103] and Collignon et al. [17] independently proposed to register images by maximizing the mutual information (MI) between the images. In this approach, mutual information serves as a similarity metric in the registration process. In order to be consistent with the registration formulation in the previous section, we define the dissimilarity metric to be the negative of the mutual information of the images.

Since 1995, several separate studies have established the effectiveness of mutual information based medical image registration [72, 80, 90, 104]. At present, mutual information is accepted by many as one of the most accurate and robust retrospective registration measures.

Mutual information is a basic concept from information theory, measuring the statistical dependence between two random variables or the amount of information that one variable contains about the other.

Given two random variables, X and Y, with marginal probability density functions, $f_X(x)$ and $f_Y(y)$, and joint probability density function, $f_{XY}(x,y)$. Mutual information, $I(X,Y)$, measures the degree of dependence of X and Y by measuring the distance between the joint distribution $f_{XY}(x,y)$ and the distribution associated to the case of complete independence $f_X(x) \cdot f_Y(y)$, by means of the Kullback-Leibler measure, i.e.,

$$I(X,Y) = \int\int f_{XY}(x,y) \log \frac{f_{XY}(x,y)}{f_X(x) f_Y(y)} \, dx \, dy \tag{4.3}$$

Mutual information is related to entropy by the equations

$$\begin{aligned} I(X,Y) &= H(X) + H(Y) - H(X,Y) && (4.4)\\ &= H(X) - H(X|Y) && (4.5)\\ &= H(Y) - H(Y|X) && (4.6)\end{aligned}$$

with $H(X)$ and $H(Y)$ being the entropy of X and Y, respectively, $H(X,Y)$ their joint entropy, and $H(X|Y)$ and $H(Y|X)$ the conditional entropy of X given Y and of Y given X, respectively

$$\begin{aligned} H(X) &= -\int\int f_X(x)\log f_X(x)\; dx && (4.7)\\ H(X,Y) &= -\int\int f_{XY}(x,y)\log f_{XY}(x,y)\; dx\; dy && (4.8)\\ H(X|Y) &= -\int\int f_{XY}(x,y)\log f_{X|Y}(x|y)\; dx\; dy. && (4.9)\end{aligned}$$

The entropy $H(X)$ is known to be a measure of the amount of uncertainty about random variable X, while $H(X|Y)$ is the amount of uncertainty left in X when knowing Y. Hence, from (4.5), $I(X,Y)$ is the reduction in the uncertainty of the random variable X by the knowledge of another random variable Y, or, equivalently, the amount of information that Y contains about X.

In order to apply the concept of mutual information to image registration, let us consider the image intensity values, x and y, of a pair of corresponding pixels in the two image that are to be registered to be random variables X and Y, respectively. Then estimates for the joint and marginal distributions $f_{XY}(x,y)$, $f_X(x)$, and $f_Y(y)$ can be obtained by applying any of a number of density estimation algorithms. Collignon et al. employed *histogram* methods to estimate the densities [17], that is, the density estimate is obtained by normalization of the joint and marginal histograms of the two grey scale images. Viola et al. applied *Parzen window* methods to estimate the densities from samples. Namely, they approximated the underlying density for X by a superposition of Gaussian densities (called a Gaussian mixture) centered on the elements of a sample A drawn from X:

$$f_X(x) \approx \frac{1}{N_A} \sum_{x_i \in A} G_\sigma(x - x_i), \tag{4.10}$$

where

$$G_\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{4.11}$$

is a Gaussian distribution with variance σ^2.

After the underlying joint and marginal densities are estimated, the mutual information between a pair of images can be easily calculated using numerical or analytical integration techniques. The mutual information registration criterion states that the mutual information of the image intensity values of corresponding pixel pairs is maximal if the images are geometrically aligned. Suppose we want to register two images, I_1 and I_0. The mutual information based registration algorithm declares that the images are geometrically aligned by the transformation T^* for which $I(T^* I_1, I_0)$ is maximal among all possible transformations.

Because no limiting constraints are imposed on the nature of the relation between the intensities in the images to be registered and no assumptions are required regarding image content, e.g., image parameterization, nor the imaging modalities involved, the mutual information criterion is very general and powerful. It allows for robust and completely automated registration of multi-modal images without prior segmentation, feature extraction or other preprocessing steps.

4.3 Minimum Rényi Entropy Criterion

Although the mutual information approach is very powerful, it has a drawback — in order to calculate the mutual information of the images to be registered, we have to estimate the underlying probability density functions of their intensities. As mentioned above, Viola et al. [103] estimated the underlying joint and marginal densities with the Parzen window method, while Collignon et al. [17] employed histogram method to obtain the estimated densities. Unfortunately, current probability density function estimation algorithms, including both Parzen window and histogram methods, are computationally intensive or unreliable, especially for few data samples. The accumulation of the errors in the density estimation process might lead to completely wrong registration.

Aware of this fact and motivated by this information-theoretic approach, we propose a novel image registration method based on the joint Rényi entropy of the images. If two images are perfectly aligned, which means that the two images are identical, the joint Rényi entropy of the overlapping images is the same as the Rényi entropy of any one of them. If the two images are not aligned, the union of the two images certainly has a higher entropy than any one of the two images. Therefore, if the two images are perfectly aligned, the joint Rényi entropy of the overlapping images is smaller than that of the overlapping misaligned images, i.e., the overlapping aligned images have the minimum Rényi entropy.

In the following we give mathematical justification of legitimacy of minimizing the joint Rényi entropy over transformations of the test images. The first result is concavity.

Theorem 4.3.1 *The Rényi entropy of fractional order,*

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \int f^\alpha(x)dx, \quad \alpha \in (0,1), \tag{4.12}$$

is a concave function, i.e., for any two densities f_0 *and* f_1*, and* $\forall\ \beta \in [0,1]$*, we have*

$$H_\alpha(\beta f_0 + (1-\beta)f_1) \geq \beta H_\alpha(f_0) + (1-\beta)H_\alpha(f_1) \tag{4.13}$$

and equality holds if and only if $f_0 = f_1$ *a.e..*

Proof: First we show that the function $g(y) = y^\alpha,\ y \geq 0$, is a concave function. Its second derivative is

$$\frac{d^2 g}{dy^2} = \alpha(\alpha - 1)y^{\alpha-2}. \tag{4.14}$$

Since $\alpha \in (0, 1)$, $\alpha - 1 < 0$. And $y \geq 0$, thus $\frac{d^2 g}{dy^2} \leq 0$. As $g(y)$ for $y \geq 0$ is concave, for two pdf's f_0 and f_1 the following inequality holds

$$[\beta f_0(x) + (1 - \beta) f_1(x)]^\alpha \geq \beta f_0^\alpha(x) + (1 - \beta) f_1^\alpha(x). \tag{4.15}$$

Integrating both sides of (4.15) over x, we have

$$\int [\beta f_0(x) + (1 - \beta) f_1(x)]^\alpha \, dx \geq \beta \int f_0^\alpha(x) \, dx + (1 - \beta) \int f_1^\alpha(x) \, dx. \tag{4.16}$$

Define function $h(y) = \frac{1}{1-\alpha} \log y$, for $y > 0$. It is a concave function, too. The concavity is immediately justified with the fact

$$\frac{d^2 h}{dy^2} = -\frac{1}{1 - \alpha} \frac{1}{y^2} < 0.$$

That is,

$$\frac{1}{1 - \alpha} \log[\beta y_0 + (1 - \beta) y_1] \geq \beta \frac{1}{1 - \alpha} \log y_0 + (1 - \beta) \frac{1}{1 - \alpha} \log y_1 \tag{4.17}$$

Utilizing (4.16) and (4.17), we have

$$\begin{aligned}
&H_\alpha(\beta f_0 + (1 - \beta) f_1) \\
&= \frac{1}{1 - \alpha} \log \int [\beta f_0(x) + (1 - \beta) f_1(x)]^\alpha \, dx \\
&\geq \frac{1}{1 - \alpha} \log \left[\beta \int f_0^\alpha(x) \, dx + (1 - \beta) \int f_1^\alpha(x) \, dx \right], && \text{by (4.16)} \\
&\geq \beta \frac{1}{1 - \alpha} \log \int f_0^\alpha(x) \, dx + (1 - \beta) \frac{1}{1 - \alpha} \log \int f_1^\alpha(x) \, dx, && \text{by (4.17)} \\
&= \beta H_\alpha(f_0) + (1 - \beta) H_\alpha(f_1) && (4.18)
\end{aligned}$$

■

Suppose that two noiseless images to be registered are I_0 and I_1 and Image I_0 is taken as the reference image. Feature vectors F_0 and F_1 are extracted from I_0 and I_1, respectively. Assume that the underlying densities for feature vectors F_0 and F_1 are f_0 and f_1, respectively, and also assume that the cardinalities of F_0 and F_1 satisfy

$$\frac{\text{Card}(F_0)}{\text{Card}(F_0) + \text{Card}(F_1)} = \beta, \quad 0 \leq \beta \leq 1. \tag{4.19}$$

Then the mixture density for overlapping feature vectors is $\beta f_0 + (1-\beta) f_1$. The Rényi entropies of the individual images, I_0 and I_1, are $H_\alpha(f_0)$ and $H_\alpha(f_1)$, respectively. The joint Rényi entropy of the overlapping images is $H_\alpha(\beta f_0 + (1-\beta) f_1)$.

Define the difference between the joint Rényi entropy and the linear combination of the individual entropies as

$$\triangle H_\alpha(\beta, f_0, f_1) \overset{\triangle}{=} H_\alpha(\beta f_0 + (1-\beta) f_1) - \left[\beta H_\alpha(f_0) + (1-\beta) H_\alpha(f_1)\right], \quad \alpha \in (0,1). \tag{4.20}$$

This difference is called *Jensen difference* and is a measure of dissimilarity between two densities f_0 and f_1 (often used in biology) [6].

From Theorem 4.3.1, we conclude that $\triangle H_\alpha(\beta, f_0, f_1) = 0$ iff $f_0 = f_1$ a.e. Therefore, if two images I_0 and I_1 are perfectly registered, then $\triangle H_\alpha(\beta, f_0, f_1) = 0$; in the other direction, if $\triangle H_\alpha(\beta, f_0, f_1) = 0$, then the two images I_0 and I_1 are perfectly registered. Hence, $\triangle H_\alpha(\beta, f_0, f_1) = 0$ can be regarded as a noiseless image registration criterion. If two noisy images are registered, the term $\triangle H_\alpha(\beta, f_0, f_1)$ will never be exactly zero. In this case, it is reasonable to register the images by searching for the minimal value of $H_\alpha(\beta f_0 + (1-\beta) f_1)$ over all affine transformations of the domain of f_1.

An alternative registration criterion for matching two images is minimization of the Rényi information divergence between the two images. As before, we assume that the two images to be registered are I_1 and I_0, respectively. Their extracted feature vectors are represented by F_1 and F_0, with underlying densities f_1 and f_0, respectively. Then the Rényi information divergence between the two images is

$$D_\alpha(f_0, f_1) = \frac{1}{\alpha - 1} \log \int f_1^\alpha(x) f_0^{1-\alpha}(x)\, dx, \quad \alpha \in (0,1). \tag{4.21}$$

We next compare the discrimination capabilities of the two proposed registration criteria,

- $\triangle H_\alpha(\beta, f_0, f_1)$, the Jensen difference between the two images.
- $D_\alpha(f_0, f_1)$, the Rényi information divergence between the two images.

For two densities f_0 and f_1 of a random vector X, define $f_0(x) - f_1(x) = \triangle_x$ and $\triangle = \max_x |\triangle_x|$. The comparison will be carried out by evaluating these two criteria for the case when $\triangle$ is very small, i.e., when the two densities f_0 and f_1 are very close.

Theorem 4.3.2 *Let* $f_{\frac{1}{2}} = \frac{1}{2}(f_0 + f_1)$*, the following asymptotic representation of the fractional Rényi entropy of a convex combination* $\beta f_0 + (1-\beta) f_1$ *holds for* $\beta \in [0,1]$*:*

$$\begin{aligned} & H_\alpha(\beta f_0 + (1-\beta) f_1) \\ = \;& H_\alpha(f_{\frac{1}{2}}) + \frac{\alpha}{1-\alpha}\left(\beta - \frac{1}{2}\right) \frac{\int f_{\frac{1}{2}}^\alpha(x) \left(\frac{f_0(x) - f_1(x)}{f_{\frac{1}{2}}(x)}\right) dx}{\int f_{\frac{1}{2}}^\alpha(x)\, dx} + o(\triangle_2) \end{aligned} \tag{4.22}$$

where

$$\triangle_2 = \frac{\int f_{\frac{1}{2}}^\alpha(x) \left(\frac{f_0(x) - f_1(x)}{f_{\frac{1}{2}}(x)}\right)^2 dx}{\int f_{\frac{1}{2}}^\alpha(x)\, dx} \tag{4.23}$$

Proof: Let $f_{1-\beta}(x) = \beta f_0(x) + (1-\beta) f_1(x)$. It can be written as

$$f_{1-\beta}(x) \;=\; \frac{1}{2}[f_0(x) + (1-\beta)(f_1(x) - f_0(x))] + \frac{1}{2}[f_0(x) + \beta(f_0(x) - f_1(x))]$$

$$\begin{aligned} &= f_{\frac{1}{2}}(x) + \frac{1}{2}(2\beta - 1)(f_0(x) - f_1(x)) \\ &= f_{\frac{1}{2}}(x)\left(1 + \frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right) \end{aligned} \tag{4.24}$$

A Taylor series expansion of $f^{\alpha}_{1-\beta}(x)$ yields

$$\begin{aligned} f^{\alpha}_{1-\beta}(x) &= f^{\alpha}_{\frac{1}{2}}(x)\left(1 + \frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)^{\alpha} \\ &= f^{\alpha}_{\frac{1}{2}}(x) + \alpha f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right) + \frac{\alpha(\alpha-1)}{2} f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)^{2} + o(\triangle^2) \end{aligned} \tag{4.25}$$

Taking logarithm on both sides of (4.25) and then dividing by $1-\alpha$, we have

$$\begin{aligned} &\frac{1}{1-\alpha}\log\int f^{\alpha}_{1-\beta}(x)dx \\ &= \frac{1}{1-\alpha}\log\int\left[f^{\alpha}_{\frac{1}{2}}(x) + \alpha f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)\right. \\ &\qquad\qquad \left. + \frac{\alpha(\alpha-1)}{2} f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)^{2} + o(\triangle^2)\right] dx \\ &= \frac{1}{1-\alpha}\log\left\{\int f^{\alpha}_{\frac{1}{2}}(x)dx\left[1 + \frac{\alpha\int f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)dx}{\int f^{\alpha}_{\frac{1}{2}}(x)dx}\right.\right. \\ &\qquad\qquad \left.\left. + \frac{\frac{\alpha(\alpha-1)}{2}\int f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)^{2}dx}{\int f^{\alpha}_{\frac{1}{2}}(x)dx} + o(\triangle^2)\right]\right\} \\ &= \frac{1}{1-\alpha}\log\int f^{\alpha}_{\frac{1}{2}}(x)dx + \frac{1}{1-\alpha}\log\left[1 + \frac{\alpha\int f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)dx}{\int f^{\alpha}_{\frac{1}{2}}(x)dx}\right. \\ &\qquad\qquad \left. + \frac{\frac{\alpha(\alpha-1)}{2}\int f^{\alpha}_{\frac{1}{2}}(x)\left(\frac{(2\beta-1)\triangle_x}{2f_{\frac{1}{2}}(x)}\right)^{2}dx}{\int f^{\alpha}_{\frac{1}{2}}(x)dx} + o(\triangle^2)\right] \end{aligned} \tag{4.26}$$

Since $\log(1+x) = x - \frac{x^2}{2} + o(x^2)$, we have

$$
\begin{aligned}
H_\alpha(\beta f_0 + (1-\beta)f_1) &= \frac{1}{1-\alpha}\log\int f^\alpha_{1-\beta}(x)dx \\
&= H_\alpha(f_{\frac{1}{2}}) + \frac{\alpha}{1-\alpha}\frac{2\beta-1}{2}\frac{\int \frac{\triangle_x}{f_{\frac{1}{2}}(x)} f^\alpha_{\frac{1}{2}}(x)dx}{\int f^\alpha_{\frac{1}{2}}(x)dx} + \frac{\alpha}{2}\left(\frac{2\beta-1}{2}\right)^2 \frac{\int \left(\frac{\triangle_x}{f_{\frac{1}{2}}(x)}\right)^2 f^\alpha_{\frac{1}{2}}(x)dx}{\int f^\alpha_{\frac{1}{2}}(x)dx} \\
&\quad -\frac{\alpha^2}{2(1-\alpha)}\left(\frac{2\beta-1}{2}\right)^2 \frac{\left(\int \frac{\triangle_x}{f_{\frac{1}{2}}(x)} f^\alpha_{\frac{1}{2}}(x)dx\right)^2}{\left(\int f^\alpha_{\frac{1}{2}}(x)dx\right)^2} + o(\triangle_3) \qquad (4.27) \\
&= H_\alpha(f_{\frac{1}{2}}) + \frac{\alpha}{1-\alpha}\left(\beta - \frac{1}{2}\right)\frac{\int f^\alpha_{\frac{1}{2}}(x)\left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)dx}{\int f^\alpha_{\frac{1}{2}}(x)\,dx} + o(\triangle_2) \qquad (4.28)
\end{aligned}
$$

■

Theorem 4.3.3 *The following asymptotic representation of the fractional Jensen difference of two densities f_0 and f_1 holds for $\beta \in [0,1]$:*

$$
\begin{aligned}
&\triangle H_\alpha(\beta, f_0, f_1) \\
&= \frac{\alpha\beta(1-\beta)}{2}\left(\frac{\int \left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)^2 f^\alpha_{\frac{1}{2}}(x)dx}{\int f^\alpha_{\frac{1}{2}}(x)dx} + \frac{\alpha}{1-\alpha}\left(\frac{\int \left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right) f^\alpha_{\frac{1}{2}}(x)dx}{\int f^\alpha_{\frac{1}{2}}(x)dx}\right)^2\right) \\
&\quad + o(\triangle_3) \qquad (4.29)
\end{aligned}
$$

where

$$f_{\frac{1}{2}}(x) = \frac{1}{2}(f_0(x) + f_1(x))$$

and

$$\triangle_3 = \frac{\int f^\alpha_{\frac{1}{2}}(x)\left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)^3 dx}{\int f^\alpha_{\frac{1}{2}}(x)\,dx} \qquad (4.30)$$

Proof: Letting $\beta = 0$ and $\beta = 1$ in (4.27), we have

$$\begin{aligned}
H_\alpha(f_0) &= H_\alpha(f_{\frac{1}{2}}) + \frac{\alpha}{1-\alpha}\frac{\int \frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)} f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx} + \frac{\alpha}{2}\frac{\int \left(\frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)}\right)^2 f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx} \\
&\quad - \frac{\alpha^2}{2(1-\alpha)}\frac{\left(\int \frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)} f_{\frac{1}{2}}^{\alpha}(x)dx\right)^2}{\left(\int f_{\frac{1}{2}}^{\alpha}(x)dx\right)^2} + o(\triangle_3) \\
H_\alpha(f_1) &= H_\alpha(f_{\frac{1}{2}}) - \frac{\alpha}{1-\alpha}\frac{\int \frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)} f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx} + \frac{\alpha}{2}\frac{\int \left(\frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)}\right)^2 f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx} \\
&\quad - \frac{\alpha^2}{2(1-\alpha)}\frac{\left(\int \frac{\frac{1}{2}\triangle_x}{f_{\frac{1}{2}}(x)} f_{\frac{1}{2}}^{\alpha}(x)dx\right)^2}{\left(\int f_{\frac{1}{2}}^{\alpha}(x)dx\right)^2} + o(\triangle_3)
\end{aligned} \tag{4.31}$$

Substituting (4.27) and (4.31) into (4.20), we obtain the Jensen difference

$$\begin{aligned}
&\triangle H_\alpha(\beta, f_0, f_1) = H_\alpha(\beta f_0 + (1-\beta)f_1) - [\beta H_\alpha(f_0) + (1-\beta)H_\alpha(f_1)] \\
&= \frac{\alpha\beta(1-\beta)}{2}\left(\frac{\int \left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)^2 f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx} + \frac{\alpha}{1-\alpha}\left(\frac{\int \left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right) f_{\frac{1}{2}}^{\alpha}(x)dx}{\int f_{\frac{1}{2}}^{\alpha}(x)dx}\right)^2\right) \\
&\qquad + o(\triangle_3)
\end{aligned} \tag{4.32}$$

■

Theorem 4.3.4 *The Rényi information divergence of fractional order $\alpha \in (0,1)$ between two densities f_0 and f_1 has the asymptotic representation*

$$D_\alpha(f_0, f_1) = \frac{\alpha}{4}\int f_{\frac{1}{2}}(x)\left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)^2 dx + o(\triangle_3) \tag{4.33}$$

where $f_{\frac{1}{2}}$ and $\triangle_3$ are as defined in Theorem 4.3.3.

Proof: We can rewrite the density f_0 as

$$f_0(x) \;=\; \frac{1}{2}(f_0(x)+f_1(x))+\frac{1}{2}(f_0(x)-f_1(x)) = f_{\frac{1}{2}}(x)+\frac{1}{2}\triangle_x \tag{4.34}$$

Similarly, we have

$$f_1(x) \;=\; f_{\frac{1}{2}}(x)-\frac{1}{2}\triangle_x \tag{4.35}$$

Using Taylor series expansion, we have

$$\begin{aligned} f_1^{\alpha}(x) &= f_{\frac{1}{2}}^{\alpha}(x)-\alpha f_{\frac{1}{2}}^{\alpha-1}(x)\left(\frac{\triangle_x}{2}\right)+\frac{\alpha(\alpha-1)}{2}f_{\frac{1}{2}}^{\alpha-2}(x)\left(\frac{\triangle_x}{2}\right)^2+o(\triangle_x^3) \\ f_0^{1-\alpha}(x) &= f_{\frac{1}{2}}^{1-\alpha}(x)+(1-\alpha)f_{\frac{1}{2}}^{-\alpha}(x)\left(\frac{\triangle_x}{2}\right)+\frac{\alpha(1-\alpha)}{2}f_{\frac{1}{2}}^{-\alpha-1}(x)\left(\frac{\triangle_x}{2}\right)^2+o(\triangle_x^3) \end{aligned} \tag{4.36}$$

Then

$$f_1^{\alpha}(x)f_0^{1-\alpha}(x) \;=\; f_{\frac{1}{2}}(x)-(2\alpha-1)\frac{\triangle_x}{2}-\alpha(1-\alpha)f_{\frac{1}{2}}^{-1}(x)\left(\frac{\triangle_x}{2}\right)^2+o(\triangle_x^3) \tag{4.37}$$

The Rényi information divergence becomes

$$\begin{aligned} &D_{\alpha}(f_0,f_1)=\frac{1}{\alpha-1}\log\int f_1^{\alpha}(x)f_0^{1-\alpha}(x)dx \\ &= \frac{1}{\alpha-1}\log\int\left(f_{\frac{1}{2}}(x)-(2\alpha-1)\frac{\triangle_x}{2}-\alpha(1-\alpha)f_{\frac{1}{2}}^{-1}(x)\left(\frac{\triangle_x}{2}\right)^2+o(\triangle_x^3)\right)dx \\ &= \frac{1}{\alpha-1}\log\left(1-\alpha(1-\alpha)\int_{\frac{1}{2}}^{-1}(x)\left(\frac{\triangle_x}{2}\right)^2dx\right) \\ &= \alpha\int f_{\frac{1}{2}}^{-1}(x)\left(\frac{\triangle_x}{2}\right)^2dx \\ &= \frac{\alpha}{4}\int f_{\frac{1}{2}}(x)\left(\frac{f_0(x)-f_1(x)}{f_{\frac{1}{2}}(x)}\right)^2dx \end{aligned} \tag{4.38}$$

■

If we define $E_f[g(x)] = \int f(x)g(x)\ dx$, for pdf f. Then (4.33) and (4.29) can be rewritten as

$$D_\alpha(f_0,f_1) = \frac{\alpha}{4}E_{f_{\frac{1}{2}}}\left[\frac{f_0-f_1}{f_{\frac{1}{2}}}\right]+o(\triangle_3) \tag{4.39}$$

$$\triangle H_\alpha(\beta,f_0,f_1) = \frac{\alpha\beta(1-\beta)}{2}\left\{E_{\tilde{f}^\alpha_{\frac{1}{2}}}\left[\left(\frac{f_0-f_1}{f_{\frac{1}{2}}}\right)^2\right]+\frac{\alpha}{1-\alpha}\left(E_{\tilde{f}^\alpha_{\frac{1}{2}}}\left[\frac{f_0-f_1}{f_{\frac{1}{2}}}\right]\right)^2\right\} +o(\triangle_3) \tag{4.40}$$

where $\tilde{f}^\alpha_{\frac{1}{2}}(x) \overset{\triangle}{=} \frac{f^\alpha_{\frac{1}{2}}(x)}{\int f^\alpha_{\frac{1}{2}}(x)\ dx}$ is a "tilted" pdf.

There are a number of interesting properties regarding $D_\alpha(f_0,f_1)$ and $\triangle H_\alpha(\beta,f_0,f_1)$:

- The divergence criterion $D_\alpha(f_0,f_1)$ depends on α, the information order, only through a scale factor, while the Jensen difference criterion $\triangle H_\alpha(\beta,f_0,f_1)$ is more strongly dependent on α.

  ① When α approaches 0, tail differences between the two densities f_0 and f_1 are much more influential on $\triangle H_\alpha(\beta,f_0,f_1)$ than on $D_\alpha(f_0,f_1)$.

  ② When α approaches 1, central differences between the two densities become highly pronounced in $\triangle H_\alpha(\beta,f_0,f_1)$. Therefore, if the images to be registered have concentrated regions of interest, we should choose α close to 1 to enlarge the discriminative capabilities of the registration criterion.

- The ratio of the numbers of feature vectors extracted from the two to-be-registered images, β, does not influence $D_\alpha(f_0,f_1)$, while this ratio does affect $\triangle H_\alpha(\beta,f_0,f_1)$. Furthermore, $\triangle H_\alpha(\beta,f_0,f_1)$ has the maximal discriminative capability for $\beta = \frac{1}{2}$, i.e., when two images yield the same number of feature vectors, the criterion $\triangle H_\alpha(\beta,f_0,f_1)$ has the best registration capability.

Note that all the three terms, $H_\alpha(\beta f_0+(1-\beta)f_1)$, $H_\alpha(f_0)$, and $H_\alpha(f_1)$, in the Jensen difference $\triangle H_\alpha(\beta,f_0,f_1)$ can be computed by employing the Rényi entropy estimator using MST techniques (3.19). In this way, this Jensen difference criterion does not require direction computation of probability density estimates, which is usually computationally expensive and may be unreliable. Therefore computing $H_\alpha(\beta f_0+(1-\beta)f_1)$ is simpler and more accurate than computing $D_\alpha(f_0, f_1)$. In this dissertation we will focus on the Jensen difference $\triangle H_\alpha(\beta, f_0, f_1)$ as the dissimilarity metric for image registration. Computing $\triangle H_\alpha(\beta, f_0, f_1)$ involves evaluating three terms, $H_\alpha(\beta f_0+(1-\beta)f_1)$, $H_\alpha(f_0)$, and $H_\alpha(f_1)$. In order to simplify the registration process, we want to have as few as possible terms in the objective function.

Theorem 4.3.5 *For fixed density f_0 and $\beta \in [0, 1]$, we have*

$$\arg_{f_1} \{\triangle H_\alpha(\beta, f_0, f_1) = 0\} = \arg\min_{f_1} H_\alpha(\beta f_0 + (1-\beta)f_1). \tag{4.41}$$

Proof: It is clear that $\triangle H_\alpha(\beta, f_0, f_1) = 0$ implies $f_0 = f_1$ a.e. (by Theorem 4.3.1), then

$$H_\alpha(\beta f_0 + (1-\beta)f_1) = H_\alpha(f_0). \tag{4.42}$$

The Rényi entropy of the mixture density $\beta f_0 + (1-\beta)f_1$ is always no less than the minimum Rényi entropy of any individual density, f_0 and f_1, i.e.,

$$H_\alpha(\beta f_0 + (1-\beta)f_1) \geq \min(H_\alpha(f_0), H_\alpha(f_1)). \tag{4.43}$$

Therefore from both (4.42) and (4.43) we conclude that $H_\alpha(\beta f_0 + (1-\beta)f_1)$ is minimized when $f_0 = f_1$ a.e., that is, $\triangle H_\alpha(\beta, f_0, f_1) = 0$ implies that $H_\alpha(\beta f_0 + (1-\beta)f_1)$ is minimized.

In the other direction, if $H_\alpha(\beta f_0 + (1-\beta)f_1)$ is minimized, i.e., $f_0 = f_1$ a.e., then we immediately conclude that $\triangle H_\alpha(\beta, f_0, f_1) = 0$. ■

Since the criteria $\min H_\alpha(\beta f_0 + (1-\beta)f_1)$ and $\triangle H_\alpha(\beta, f_0, f_1) = 0$ are equivalent, i.e., both are minimized for $f_1 = f_0$, and the former criterion has lower complexity, in the sequel, we will investigate the minimal joint Rényi entropy, $H_\alpha(\beta f_0 + (1-\beta)f_1)$, of the overlapping images as the registration criterion.

Recall that the Rényi entropy of a density $f(x)$ can be estimated by the normalized length of the minimum spanning tree over the corresponding graph connecting n i.i.d. realizations $\{x_1, x_2, \ldots, x_n\}$ of $f(x)$. Thus the objective of image registration can be stated as: find a transformation T on image I_1 which minimizes the length of the minimum spanning tree connecting the vertices generated from two images I_0 and I_1. With this property in mind, we successfully avoid the process of estimating underlying probability distribution and directly match two images based on the image intensities of the pixel pairs.

To apply the minimum joint Rényi entropy criterion to the image registration problem, we can choose to either work on the image intensity values or work on the feature vectors extracted from the images. It is important to keep in mind that we intend to employ minimum spanning tree methods in the registration process, and the computational complexity of constructing the minimum spanning tree is polynomial in the number of vertices. Reduction of the number of vertices is therefore critical for the algorithmic complexity to be manageable. Since feature extraction can greatly decrease the number of representation vectors, we will work on the feature vectors instead of working directly on the intensity values.

Suppose that we need to register a pair of images I_1 and I_0. Assume that there exists a mapping $T : I_1 \to I_0$ that relates the intensity values of corresponding pixel pairs. Let TI_1 denote the image transformed from I_1 with the transformation T. Let $\{TX_j\}_{j=1}^n$ and $\{Y_k\}_{k=1}^m$ be the point processes in R^d representing feature vectors extracted from TI_1 and

I_0, respectively. $\{TX_j\}_{j=1}^n$ are i.i.d. random variables with the underlying unknown densities f_{TX}, and $\{Y_k\}_{k=1}^m$ are i.i.d. random variables with the underlying unknown densities f_Y. Then the joint Rényi entropy of TX and Y is $H_\alpha(\beta f_Y + (1-\beta)f_{TX})$, $\beta = \frac{m}{m+n}$, and it satisfies:

$$H_\alpha(\beta f_Y + (1-\beta)f_{TX}) \geq H_\alpha(\beta f_Y + (1-\beta)f_{T^*X}) \geq H_\alpha(f_Y) \tag{4.44}$$

where T^* is the *best* transformation function such that the transformed image I_1 aligned with image I_0. Note that the second inequality may not be a strict equality since when images are acquired either by the same sensor at different times or by different sensors at the same or different times, a number of distortions prevent the two images from ever being "perfectly registered" to each other.

Under such framework, the image registration problem can be stated as a minimization problem:

$$T^* = \arg\min_{T \in \mathcal{T}} H_\alpha(\beta f_Y + (1-\beta)f_{TX}) \tag{4.45}$$

i.e., our objective is to find the optimal transformation T^* which minimizes the joint Rényi entropy. Utilizing (3.19), we turn to the problem of searching for the transformation T^* such that the graph corresponding to the union of $\{TX_j\}_{j=1}^n$ and $\{Y_k\}_{k=1}^m$ has the shortest minimum spanning tree length,

$$T^* = \arg\min_{T \in \mathcal{T}} L_{MST}(TX, Y) \tag{4.46}$$

4.4 Image Registration Procedure

Suppose that we are given two images I_1 and I_0. If the two images are obtained from different sensors, we pre-process one of the two images, say I_1, and convert I_1 to the image plane on which I_0 lies according to the physical characterization of each of the sensors. Alternatively, we convert both images to a reference image plane. Thus we get two images that lie on the same image plane. In order to focus on the registration algorithm of finding appropriate spatial transformations, from now on, we only consider the registration of two images I_1 and I_0 on the same image plane. Furthermore, we assume I_0 is the reference image and will transform I_1 to match I_0. We will call image I_1 the test image.

For each candidate spatial transformation T, the image registration process follows three steps:

① Apply spatial transformation T to I_1 and obtain the transformed image TI_1.

② Extract feature vectors $\{TX_j\}_{j=1}^n$ and $\{Y_k\}_{k=1}^m$ from both TI_1 and I_0, respectively.

③ Generate a graph G from the mixture of feature vectors $\{TX_j\}_{j=1}^n$ and $\{Y_k\}_{k=1}^m$, then construct the MST on G and calculate the length of the resulting MST, $L_{MST}(TX, Y)$.

The above three steps are repeated for all possible transformations $T \in \mathcal{T}$. Then we declare the transformation T^* satisfying (4.46) as the optimal transformation and the corresponding transformed image is matched to the reference image.

Since we will concentrate on developing the graph matching approach for image registration, we will not address various spatial-coordinate transformations in this dissertation. Instead, since the transformations are quite dependent on the application, we will give two examples to illustrate the different types of transformations.

4.4.1 Geo-registration

The first application is to register two images taken on different sensor planes by potentially different sensor modalities for geo-registration applications. Our objective is to register two types of images — a set of electro-optical(EO) images and a terrain height map. For this multisensor image registration problem, there usually exists distortions between the two types of images. The distortions are due to difference acquisition conditions of the images such as shadowing, diffraction, terrain changes over time, clouds blocking the illumination sources, seasonal variations, etc. Existence of such differences between the images to be registered requires that the registration algorithms to be robust to noise and other small perturbations in intensity values.

In our image registration problem, the set of EO images are generated from the *a priori* digital elevation model (DEM)[1] of a terrain patch (the terrain height map) at different look angles (determined by the sensor's location) and with different lighting positions. With different sensor and light locations, we can simulate the distractions mentioned above. For example, shadows are generated by taking into account both the sensor location and the lighting location as follows. The scene is first rendered using the lighting source as the viewing location. Depth values (distance from the light source) are generated for all pixels in the scene and stored in a depth buffer. Next, the scene is rendered using the sensor's location as the viewpoint. Before drawing each pixel, its depth value as measured from the sensor is compared to the transformed depth value as measured from the light source. This comparison determines if a particular pixel is illuminated by the source. Shadows are placed on those pixels that fail this comparison. The EO image generation flowchart is shown in Figure 4.4.2.

In this geo-registration problem, since we have two types of images to register, the first step is to project one type of the images to the other image plane. In our case, we project the terrain height map to the EO image plane and take the resulting EO projection as the reference image. Our objective is to find the appropriate EO image which registers with the reference image, i.e., to find the correct viewing angles such that the corresponding EO image is the best match to the reference image. In this application, the change of viewing angles corresponds to the spatial-coordinate transformation. Figure 4.4.3 shows the EO

---

[1] DEM stores the terrain height information in a three dimensional array where each element of the array consists of the locations (x and y coordinates) and the height of the terrain at that location.

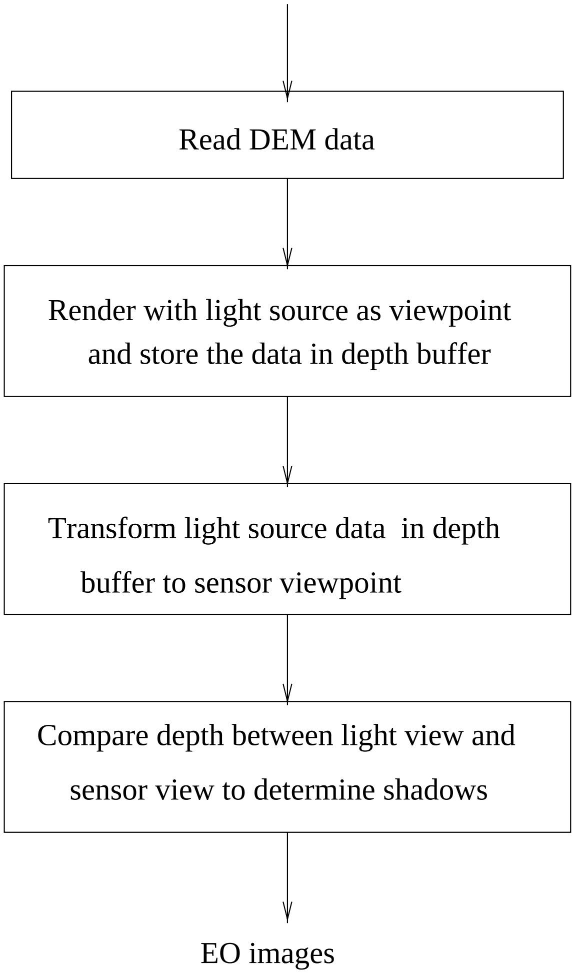
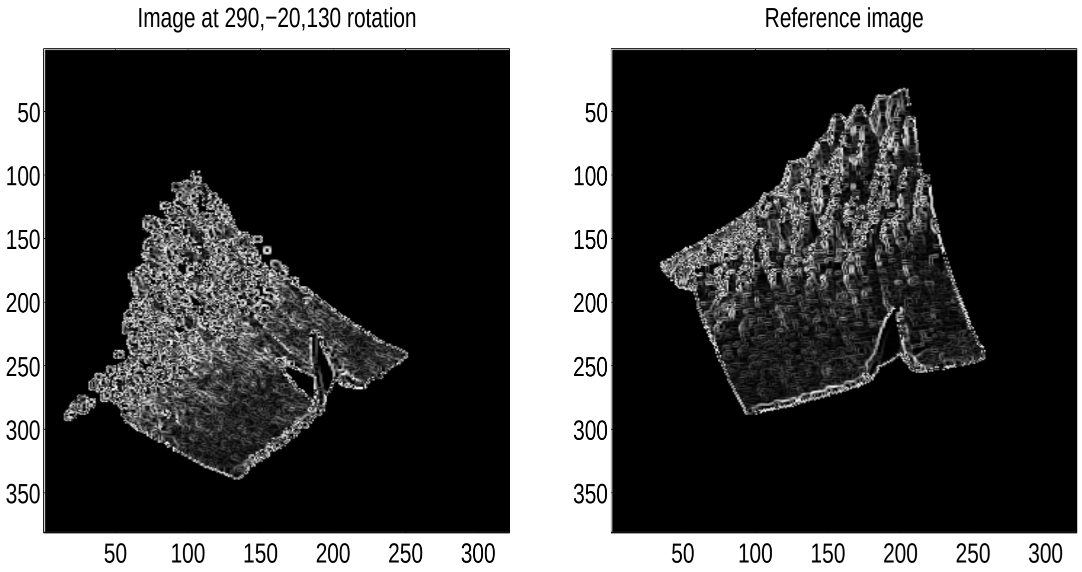


Figure 4.4.2: EO image generation block diagram

image with viewing angles (290, -20, 130) and the reference image. Clearly they are not aligned. Figure 4.4.4(a) shows the EO image with viewing angles (300, 0, 110), which is much better aligned with the reference image than the EO image with viewing angles (290, -20, 130).



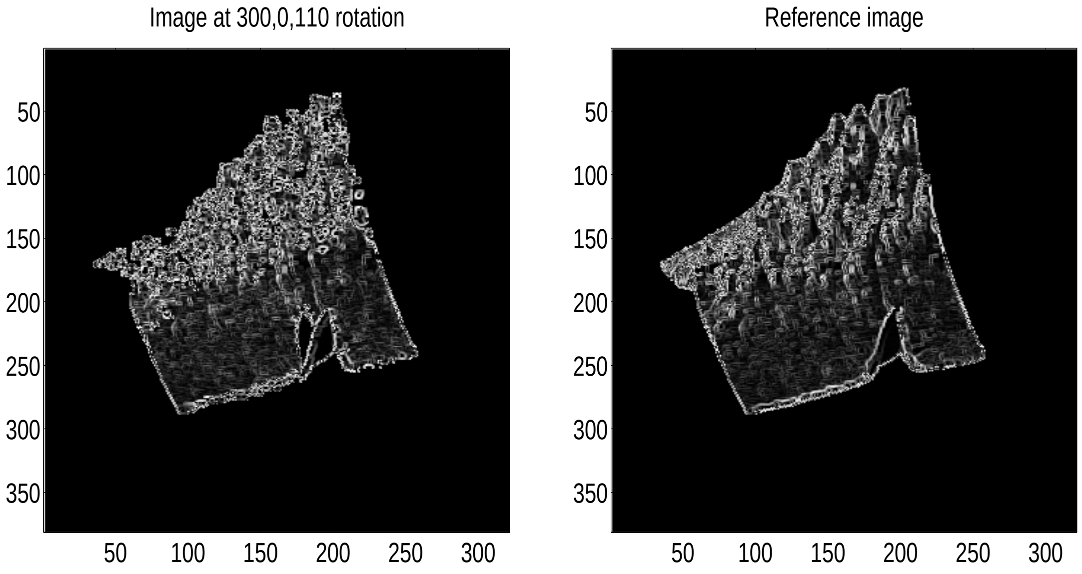
(a) (b)

Figure 4.4.3: Misaligned EO and reference images

4.4.2 Medical Image Registration

In some clinical applications, the images taken at different times are often utilized to analyze the health condition of a patient or to evaluate the operation performance. We will register magnetic resonance imaging (MRI) images via this graph matching method. Magnetic resonance imaging (MRI) is an imaging technique used primarily in medical settings to produce high quality images of the inside of the human body. MRI is based on the principles of nuclear magnetic resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules.

Magnetic resonance started out as a tomographic imaging modality for producing



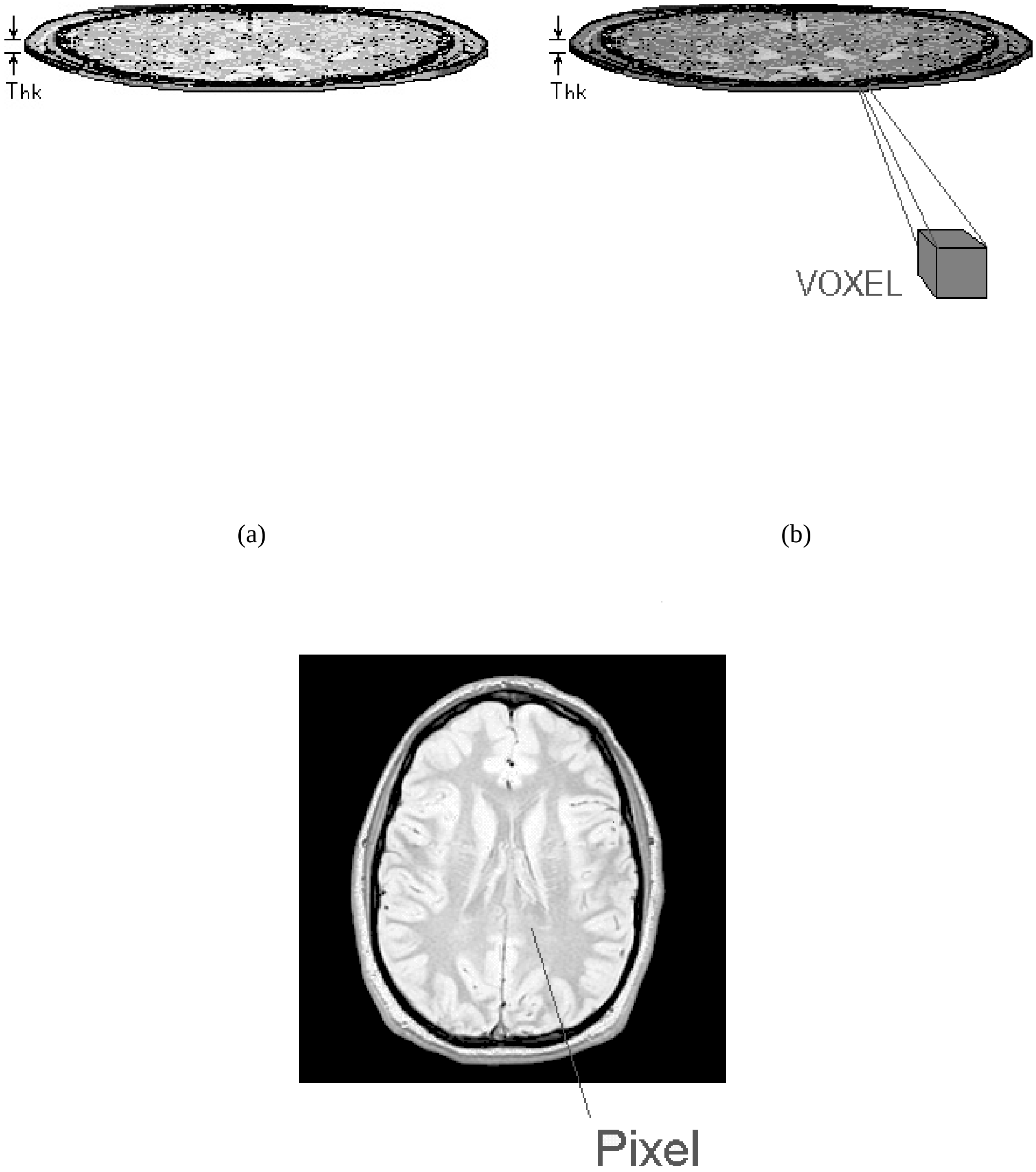
(a)

(b)

Figure 4.4.4: Aligned EO and reference images

NMR images of a slice though the human body. Each slice had a thickness (Thk)(4.4.5(a)). This form of imaging is in some respects equivalent to cutting off the anatomy above the slice and below the slice. The slice is said to be composed of several volume elements or voxels. The volume of a voxel is approximately 3mm^3(4.4.5(b)). The magnetic resonance image is composed of several picture elements called pixels. The intensity of a pixel is proportional to the NMR signal intensity of the contents of the corresponding volume element or voxel of the object being imaged.

MRI is based on the measurement of radio frequency electromagnetic waves as a spinning nucleus returns to its equilibrium state [47]. This MR phenomenon is due to the nuclear spin angular momentum which is possessed by atoms with an odd number of protons and/or neutrons. The principle of MR is based on these spins in the presence of the main (static) magnetic field B_0 (usually applied in the z-direction, which is the longitudinal axis), the radio frequency excitation field B_1 (usually applied in the xy plane, which is



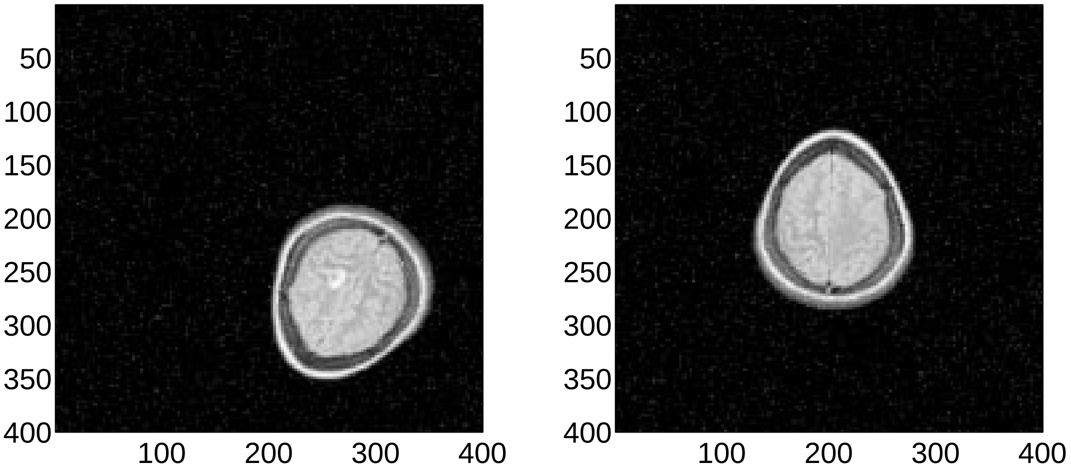
(a) (b)

(c)

Figure 4.4.5: MRI image generation process

the transverse plane) and the linear gradient fields G_s (usually applied in x and/or y directions for two dimensional imaging). When such an atom is placed in the static magnetic field B_0, the moment of the nucleus tends to line up with the field and all the spins possess the same frequency called *Larmor frequency*, which is proportional to B_0. If the atom is excited again by another magnetic field B_1, it emits a radio frequency signal as it returns to its equilibrium position. In this position, the magnetization vector precesses about the z-axis. If this excitation field is turned off, then the magnetization vector which was earlier aligned in the equilibrium position, now returns to its new equilibrium position along the z-direction producing an electromagnetic wave at Larmor frequency called free induction decay. This free induction decay signal can be detected using coils around the object, The time constant characterizing the return of this magnetization vector back to the z-direction is called T_1, while the time constant characterizing the decay of magnetization along the xy plane is called T_2. Spatial localization is achieved by applying linear gradient magnetic fields G_s in addition to the main field B_0. This gradient fields give spatial information through frequency and phase encoding of the received signal. One attractive feature of MRI is the ability to manipulate soft tissue contrast over a wide range of independent parameters in MRI. The main drawback of MRI is that MR imaging usually requires long scan times and expensive hardware. Nonetheless, it is preferred for its high resolution.

Figure 4.4.6 shows two magnetic resonance imaging (MRI) images. Each of the images are acquired from the same patient, taken from pre-operative and post-operative scans, respectively. Note that the left image, I_1, has an abnormal nodule (white spot) not present in the post-operation image, I_0. In this application, the spatial-coordinate transformation consists of translation and rotation.



(a) Pre-operative scan

(b) Post-operative scan

Figure 4.4.6: Brain images for registration

4.5 Determination of the Rényi Entropy Order

From the definition of Rényi entropy (3.12), we observe that there is a freedom α, the information order, that needs to be determined. Recall that we have restricted α to a fractional order in Section 3.3.

In order to determine the optimal order α in graph matching registration, we study the Rényi information divergence of the images to be registered when we vary the order α. In this study, we employ the histogram algorithm to estimate the underlying density functions of the image observations, and then calculate the information divergence using its definition (3.16).

For the geo-registration problem (Figure 4.4.3), the information divergence results with different order α are shown in Figure 4.5.7. In each subplot the x axis stands for the indexes of test images and the y axis stands for the Rényi information divergence. For any value of α, the information divergence achieves the minimum for the same image pair. And the divergence curves have the same envelope profile for all values of α. However, the difference between the minimum divergence and the second minimum divergence changes with the value of α. In order to achieve the maximum discrimination capability, it is better to have larger divergence difference.

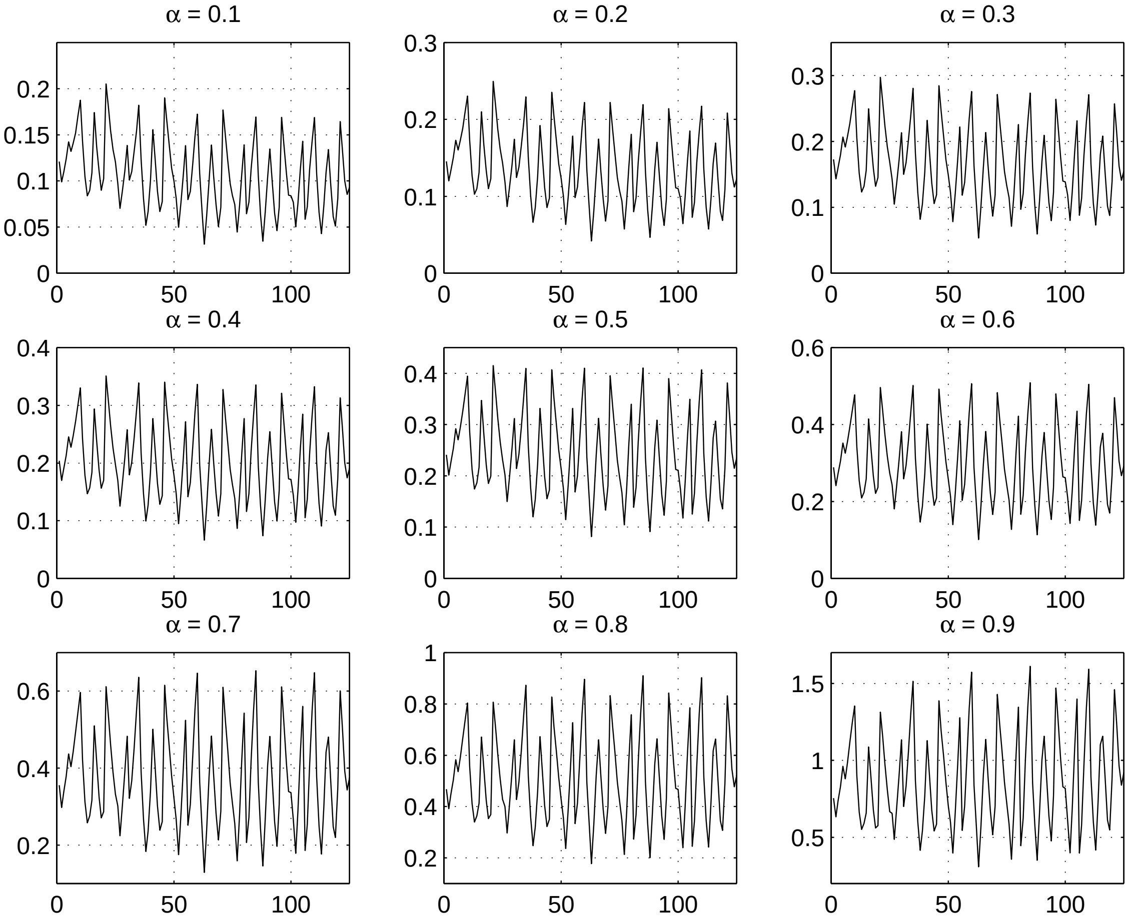


Figure 4.5.7: Rényi information divergence with different orders

For an order α, define the resolution of information divergence as the difference between the minimum value and the second minimum value of the Rényi information divergence. The resolution versus the order for the above geo-regsitration problem is plotted in Figure 4.5.8. Note that the difference increases as the order α increases. But at the same time, the dynamic range of the information divergence increases with the order α, too (Figure 4.5.7).

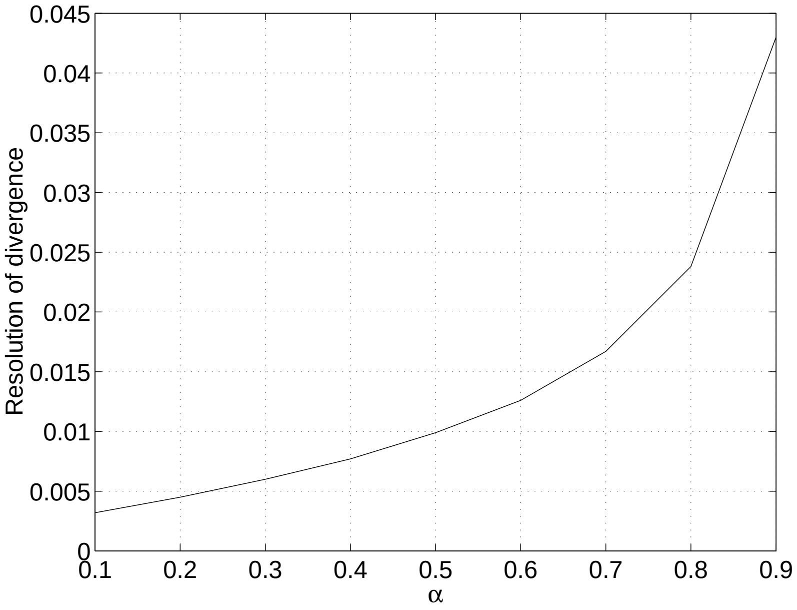


Figure 4.5.8: Resolution of Rényi information divergence with different orders

We are more interested in the relative resolution of the Rényi information divergence, which equals the resolution divided by the dynamic range of the information divergence (see Figure 4.5.9). The relative resolution for this application is also monotone increasing with the value of the order of Rényi information divergence. So in this registration application, the larger α, the better discrimnation capability.

We also investigate the resolution of Rényi information divergence for another georegistration problem. In this application, we are given a series of DEM images taken at different look angles and are asked to find the look angles of the registered image to the reference image. Figure 4.5.10 gives a test image and the reference image. The Rényi information divergence of the test and reference images with different orders is shown in Figure 4.5.11. In each subplot the x axis stands for the indexes of test images and the y axis stands for the Rényi information divergence. Similar to the previous example, for any values of α, the information divergence achieves the minimum for the same test-reference image pair. We plot the resolution and relative resolution of the information divergence

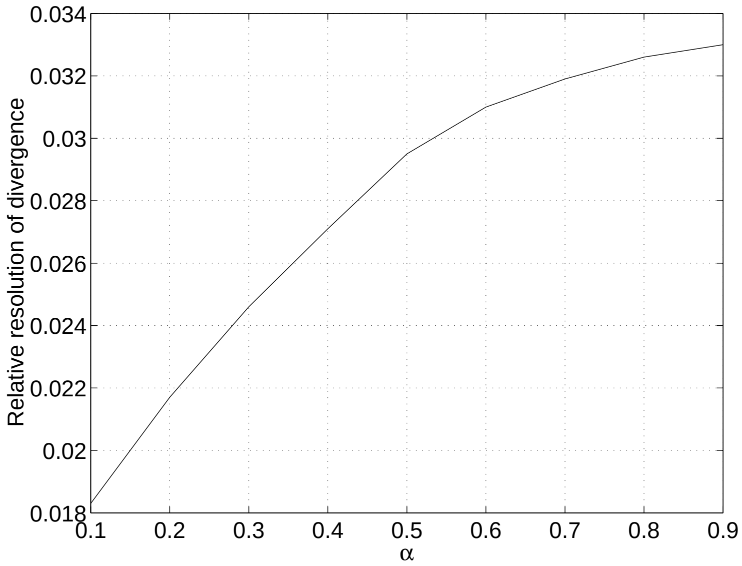
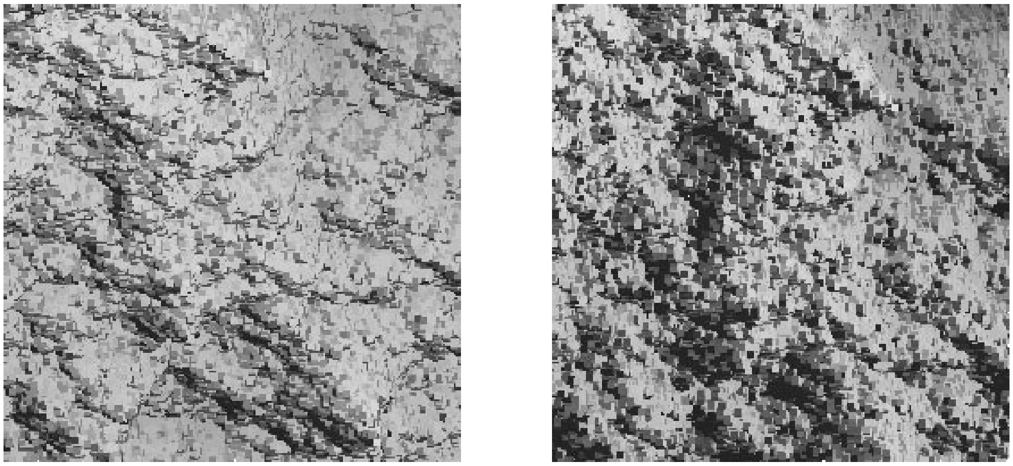


Figure 4.5.9: Relative resolution of Rényi information divergence with different orders

in Figures 4.5.12 and 4.5.13, respectively. From Figures 4.5.13 we see that the relative resolution achieves the maximum for $\alpha = 0.5$, which means the information divergence of order 0.5 has the the best discrimination capability among all possible fractional orders.



(a) Test image

(b) Reference image

Figure 4.5.10: Test and reference DEM images

Note that for the two appications, the best values of α, the order of information divergence, are not the same. In the first geo-registration application, the objects of interest are

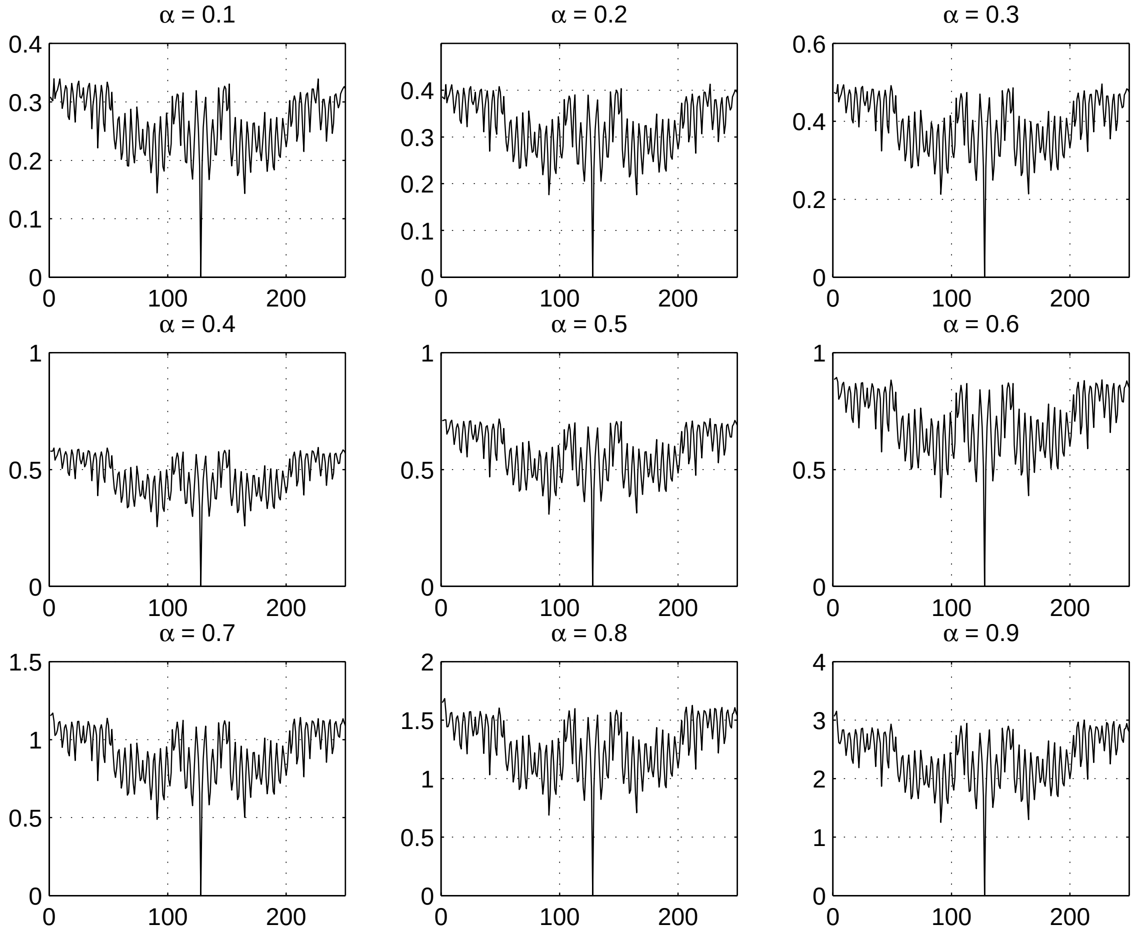


Figure 4.5.11: Rényi information divergence with difference orders

located close to the centers of images and hence the underlying densities of the images are highly centered. The Rényi information divergence achieves the maximum when the order approaches 1. On the contrary, in the second registration application, the objects of interest are scattered in the images and hence the underlying densities of the images have heavier tails than those in the first application. So we have to use lower value of α than unity to emphasize these tails.

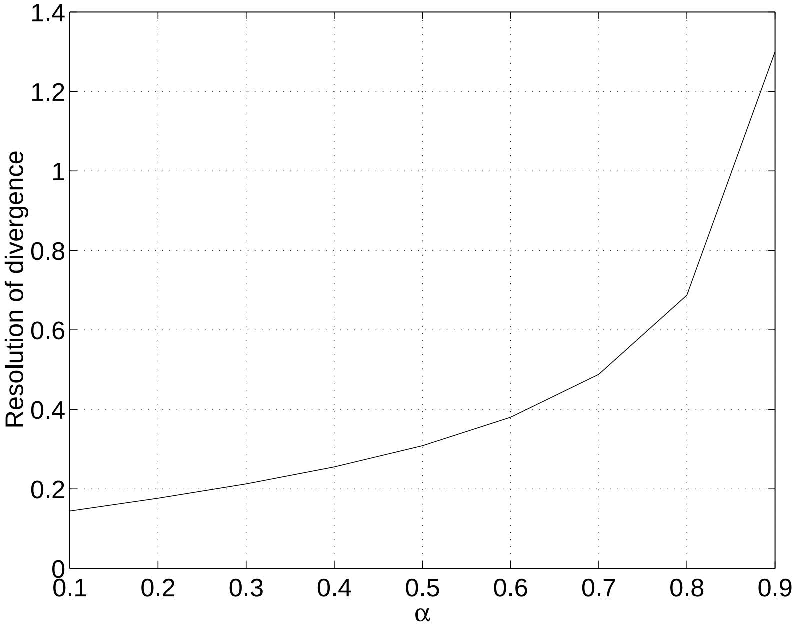


Figure 4.5.12: Resolution of Rényi information divergence with difference orders

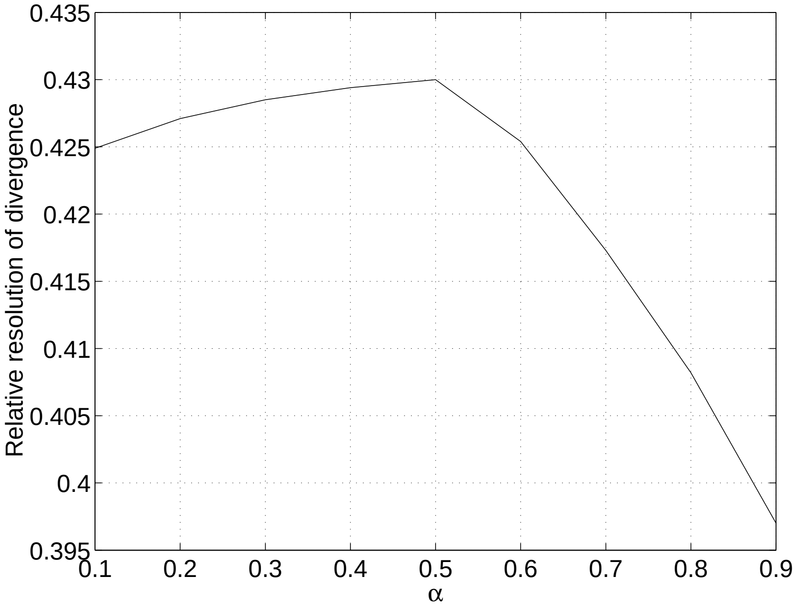


Figure 4.5.13: Relative resolution of Rényi information divergence with difference orders

4.6 Feature Extraction

For the registration problem, we want to register a test image I_1 to the reference image I_0. We first do spatial-coordinate transformation on the image I_1 to get a stack of images with different transformation parameters. Let us denote the stack of images as $\{I_1^{(n)}\}$. For the purpose of conciseness, we use I_1 to denote an image from the image set $\{I_1^{(n)}\}$.

A digital image is considered as a two-dimensional array. Each pixel is identified by its coordinates (x, y) and intensity $I(x, y)$. Alternatively, each pixel can be treated as a three-dimensional vector, $(x, y, I(x, y))$, and thus a pixel can be treated as a point in the three dimensional space. With this point of view, the pixels of the image can be considered as vertices in the three-dimensional space and then a minimum spanning tree can be constructed over these vertices.

Consider registering two 256×256 images, which are quite small images in real world applications. Each image contains $256 \times 256 = 65536$ pixels. If we try to find the MST over all image pixels, the MST should be constructed over as many as $65536 \times 2 = 131072$ vertices, which is computationally prohibitive since MST algorithm is implemented in polynomial time.

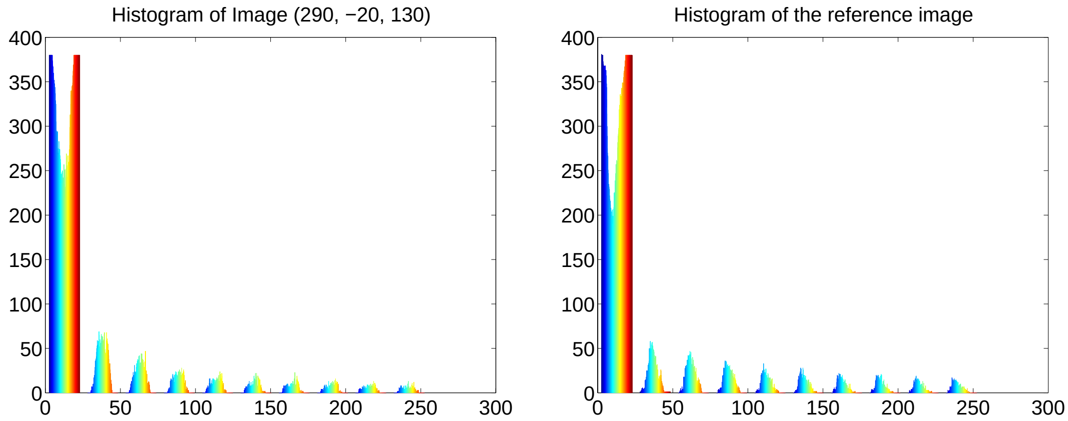
In order to make constructing MST feasible for the image registration problem, appropriate features must be extracted to compress the original great amount of data. The features should be able to well represent the original image and the number of feature vectors should be under the limit that MST is able to handle for computational complexity concern.

In our work, we investigate three approaches to extract feature vectors from the original image — uniform spatial sub-sampling, vector quantization, and stratified sampling with centroid refinements. We will illustrate the feature extraction process using the EO –

terrain map registration application (Figure 4.4.3).

4.6.1 Feature Extraction Via Uniform Spatial Sub-sampling

Before extracting features, we notice that the background (dark area) of the image does not contribute much to the registration. The histograms of both images are shown in Figure 4.6.14. From the histograms we empirically determine the threshold to be 30.



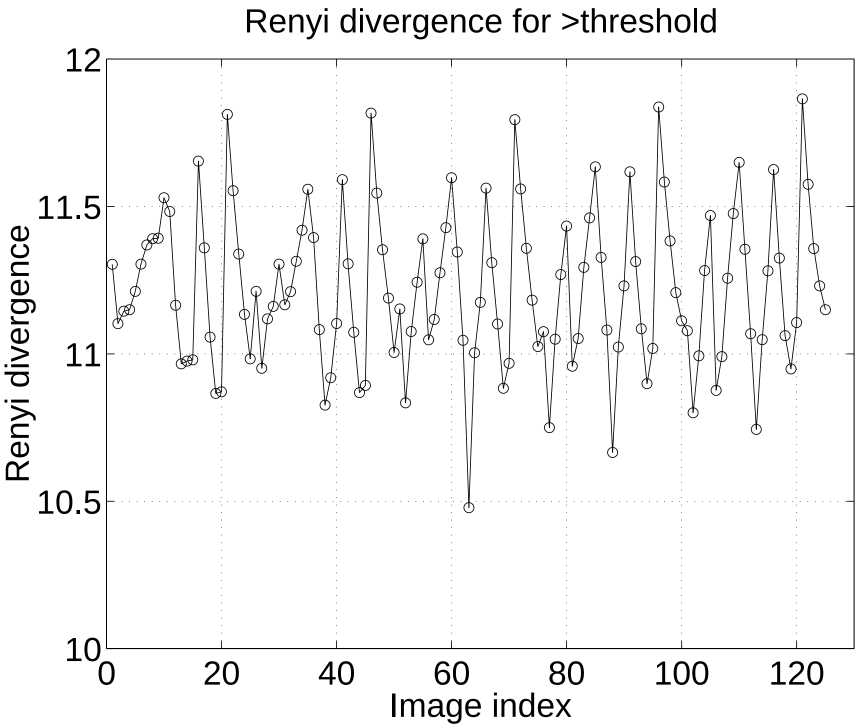
(a) (b)

Figure 4.6.14: Histograms of the EO and reference images

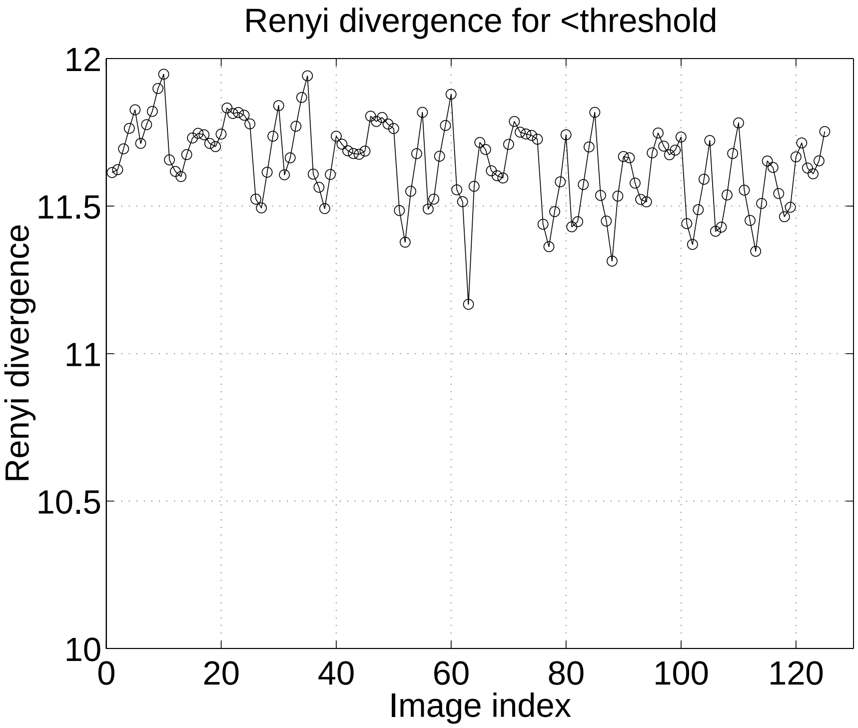
In order to justify that the background (dark area with intensity values less than 30) of the image does not much affect the registration result, we plot the Rényi information divergence for the pixels with intensity values greater (or less) than the threshold in Figure 4.6.1(a) (or Figure 4.6.1(b)). We can see that the Rényi information divergence changes in a small range for pixels with intensity values less than the threshold, but the number of those pixels are large. Thus removing those pixels can significantly reduce the number of vertices in the corresponding graph while not losing too much information.

The thresholded images are shown in Figure 4.6.16.

After thresholding the image, we sub-sample the thresholded image with a uniform grid. The pixels in the sub-sampled image will serve as the feature vectors to generate the

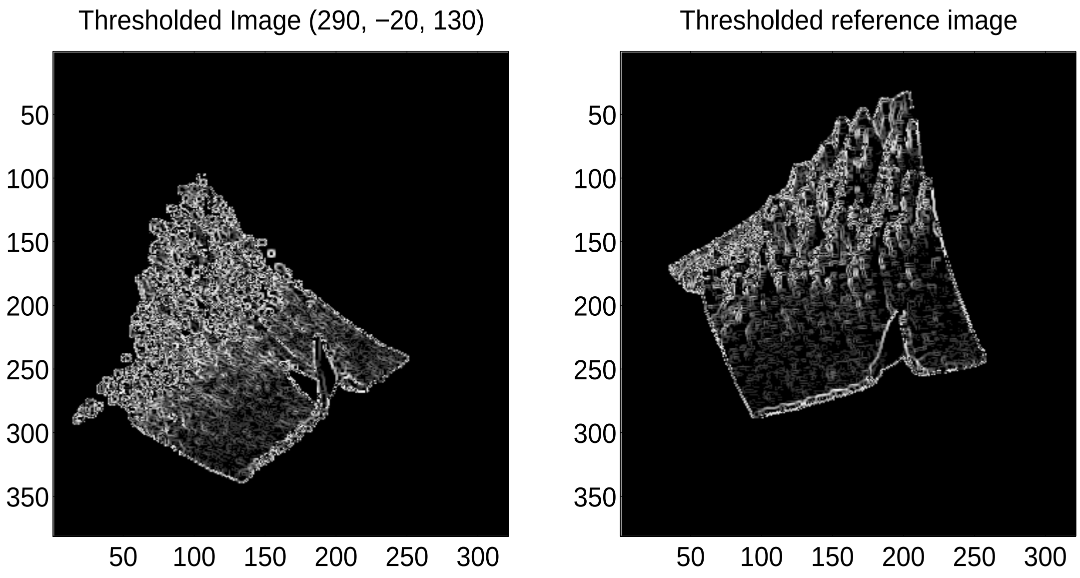


(a)



(b)

Figure 4.6.15: Rényi information divergence with respect to the threshold



(a) (b)

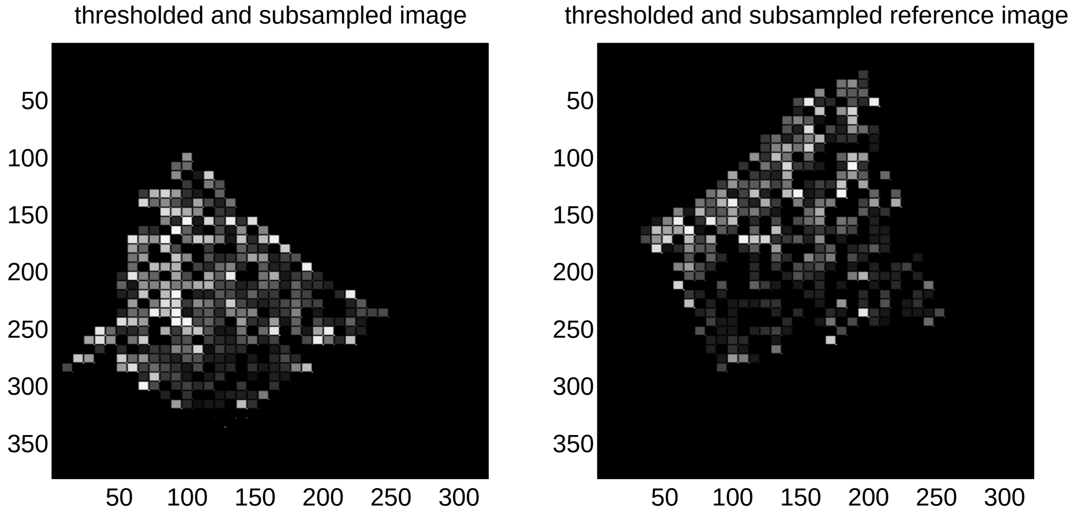
Figure 4.6.16: Thresholded EO and reference images

graph, i.e., pixel $I(x, y)$ in the sub-sampled image corresponds to the vertex $(x, y, I(x, y))$ in the generating graph. The sub-sampling rate is selected such that the number of vertices in the graph is computationally feasible for constructing MST while the remaining pixels contain enough information of the original image. Figure 4.6.17 shows the sub-sampling result with a sub-sampling rate of 8.

4.6.2 Feature Extraction Via Vector Quantization

Spatial sub-sampling greatly decreases the number of vertices in the graph. However, the number of pixels in the image cannot be decreased too much since spatially distinct patterns and texture may be lost. Therefore sub-sampling inevitably removes a great amount of useful information for registration. Indeed at very low sampling rate, Nyquist tells us that we will lose so much information that the resulting feature vectors are a very poor representation of the original image.

Some advanced techniques other than spatial sub-sampling can be applied to the fea-



(a) (b)

Figure 4.6.17: Feature vectors extracted via uniform sampling

ture extraction problems to reduce the number of vertices in the graph without losing much fidelity. In spatial sub-sampling algorithm, each coordinate of a pixel is uniformly quantized. Basically, the spatial sub-sampling algorithm partitions the image into uniform rectangular cells and places the vertices of the desired graph at the centers of each cell. Instead of sub-sampling the image on each coordinate separately, we propose to group the row and column coordinates together and quantize them as a single block. This idea is an extension of vector quantization (VQ) [30] for lossy data compression.

In vector quantization, we take L samples from the source and treat each sample as a component of a vector of dimension L. Both the encoder and decoder have a codebook that is comprised of L-dimensional code-vectors, which are selected to be the representation vectors for the source samples. For an encoder with Voronoi partition, each source sample vectors is compared to all the code-vectors in the codebook, and the closest code-vector is chosen to represent the input. The distance between two vectors are defined as

$$D = \sum_{l=1}^{L} (X_l - V_l)^2, \tag{4.47}$$

where X_l is the value of the lth element of the vector, and V_l is the value of the lth element of the code-vector. This algorithm partitions the sample space into Voronoi cells that corresponds to the code-vectors. Optimal codebooks for VQ can be derived mathematically for i.i.d. sources with certain distributions, such as, uniform, Gaussian, and Laplacian distributions.

In the case of VQ spatial sub-sampling, we treat the row and column coordinates of each pixel from the image as a two dimensional data vector. Given a certain number of vertices N, we design a codebook with N code-vectors, that minimize the average squared distance $\bar{D}$, which is given by

$$\bar{D} = \frac{1}{\sum_{i=1}^{N} M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} (r^{(i,j)} - V_1^{(i)})^2 + (c^{(i,j)} - V_2^{(i)})^2 \tag{4.48}$$

where M_i is the number of pixels in the ith cell, $r^{(i,j)}$ and $c^{(i,j)}$ are the row and column coordinates of the jth pixel in the ith cell, and $V_1^{(i)}$ and $V_2^{(i)}$ are the row and column coordinates of the vertex that represents the ith cell.

We propose to employ vector quantization technique to obtain the representation vectors for a given image. These representation vectors will serve as the feature vectors for constructing the MST. The feature extraction algorithm is as follows:

- Preset the number of representation vectors in each image (length of code book).
- Do vector quantization on the coordinates of the pixels.
- In each Voronoi region, the representation vector is the Voronoi centroid with the mean intensity of the region as its intensity.

We use the Linde-Buzo-Gray (LBG) algorithm [66] to implement the vector quantization. Let $\{X_n\}_{n=1}^{N}$ denote the input vector set, namely, X_n is a two dimensional vector consisting of the coordinates of the n-th pixel, i.e., $X_n = (i_n, j_n)$, where i_n and j_n are the row and column position for the n-th pixel . Let $\{Z_m\}_{m=1}^M$ be the representation vectors, where M is a pre-determined constant. The LBG algorithm proceeds as follows:

① Start with an initial set of representation vectors $\{Z_m^{(0)}\}_{m=1}^M$. Set $k = 0, \bar{D}^{(0)} = 0$. Select threshold ϵ.

② Find quantization regions

$$V_m^{(k)} = \{X_n | d(X_n, Z_m) < d(X_n, Z_l), \forall l \neq m\}, l = 1, 2, \ldots, M - 1$$

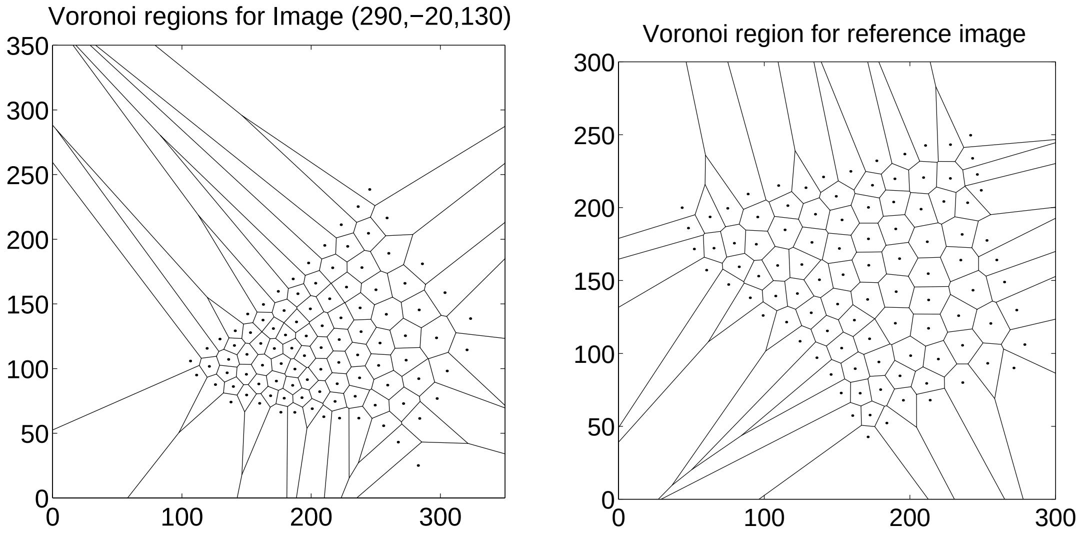
③ Compute the average distortion $\bar{D}^{(k)}$ between the input vectors and the representation vectors.

④ If $\frac{\bar{D}^{(k)} - \bar{D}^{(k-1)}}{\bar{D}^{(k)}} < \epsilon$, stop; otherwise, continue.

⑤ $k = k + 1$. Find new representation vectors $\{Z_m^{(k)}\}_{m=1}^M$ that are the average value of the elements of each of the quantization regions $V_m^{(k-1)}$. Go to Step 2.

In Figure 4.6.18 we show the vector quantization regions for the thresholded images (shown in Figure 4.6.16). Here we set the number of representation vectors to be 100 in each image.

Vector quantization technique significantly reduces the number of feature vectors without losing much useful information and thus the MST construction over these feature vectors are much faster. However, in order to have a satisfactory representation of the original image, the number of iterations in the LBG algorithm has to be sufficiently large. Therefore the vector quantization algorithm is computationally expensive, too. Thus we need to achieve a compromise between the computational load for vector quantization and the load for the MST algorithm.



(a) (b)

Figure 4.6.18: Feature vectors via vector quantization

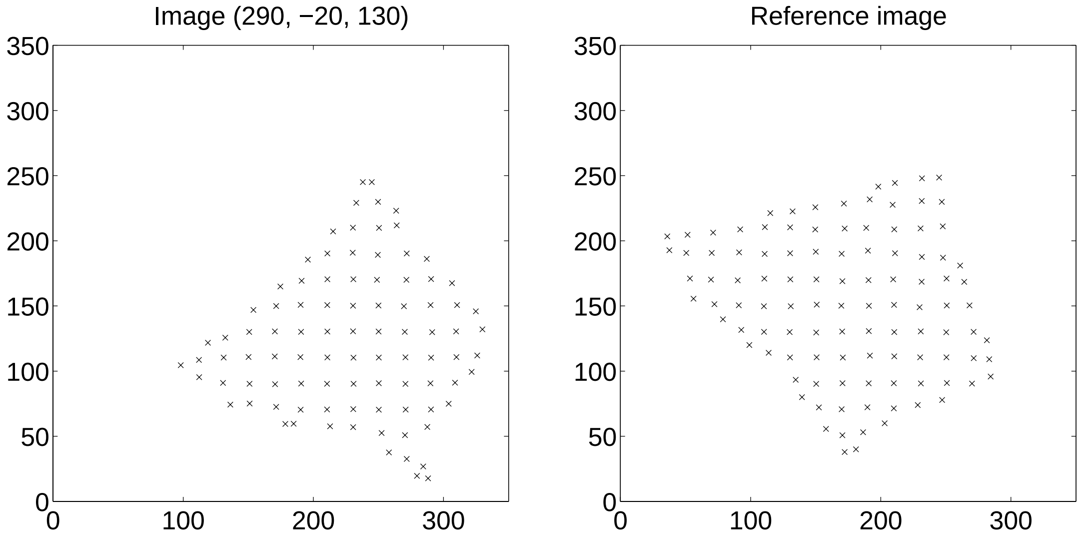
4.6.3 Feature Extraction Via Stratified Sampling with Centroid Refinements

The computation overhead of the vector quantization based feature extraction method results from the large number of iterations in the LBG algorithm. The reason for such large number of iterations is that the underlying distribution of the coordinates of the image pixels with intensities higher than a threshold is far from uniform, but the initial representation vectors of the LBG algorithm are selected to be uniform distributed. The significant difference between the initial and final representation vectors results in long computational time. One approach to speed up the LBG algorithm is to obtain a good initialization. However, a good initialization usually requires a thorough knowledge of the underlying distribution of the pixel coordinates, which is not easy to obtain.

From Figure 4.6.18, we notice that the representation vectors are concentrated in the region of interest and in this region the distribution of representation vectors is fairly close to the uniform distribution. By taking advantage of this fact, we propose a “quick” low complexity approximation of the Voronoi cells and centroids in the vector quantization method. This feature extraction technique is described as follows:

- First divide the image with a uniform grid.
- Throw away the cells whose number of pixels is less than a pre-determined threshold. The remaining cells are regarded as approximations to the Voronoi cells in the vector quantization method.
- Calculate the coordinate centroid of each remaining cell. Take the centroid as the representation vector with the mean intensity of the cell as its intensity.

Figure 4.6.19 illustrates the result of applying this feature extraction technique to the thresholded images (shown in Figure 4.6.16).



(a) (b)

Figure 4.6.19: Feature vectors via stratified sampling with centroid refinements

4.7 MST Construction

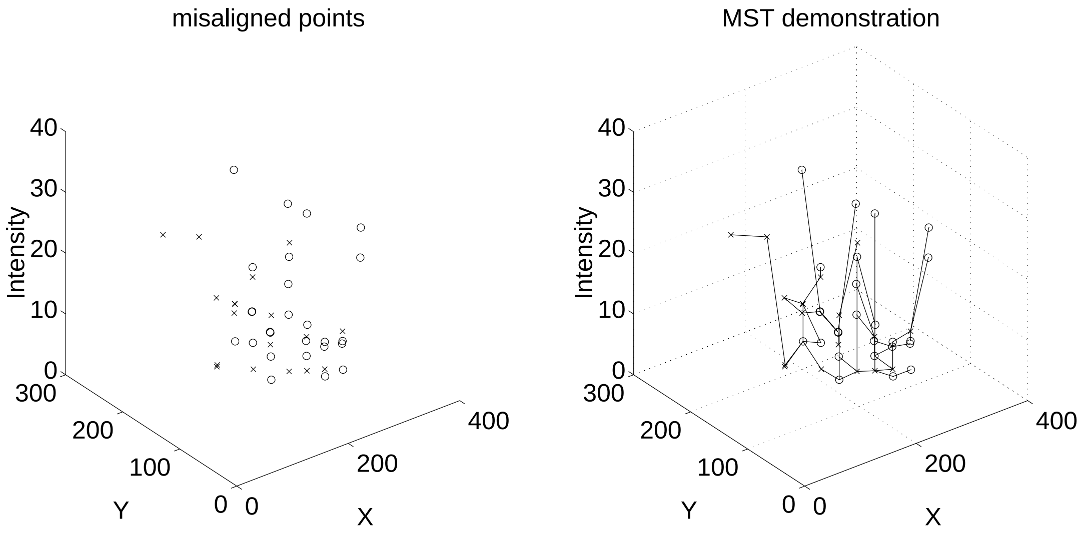
The feature vectors $(x, y, F(x, y))$ extracted from the original image are used to generate a graph in a three-dimensional space and each feature stands for a vertex described by a vector of length three, $(x, y, F(x, y))$. For image registration, we generate two graphs G_1 and G_0 from the image to be registered I_1 and the reference image I_0, respectively. To register the two images I_1 and I_0, first we overlap the two graphs G_1 and G_0 together and get a mixture graph G, and then we construct the MST over the graph G.

In order to clearly render the overlapping graphs and the corresponding MST, when we plot the figures, we use many fewer points than those actually required by image registration. Figure 4.7.20 demonstrates the MST over misaligned images, while Figure 4.7.21 shows the MST over aligned images. In both Figures 4.7.20(a) and 4.7.21(a), circle points denote the pixels from Image I_1 and cross points denote the pixels from Image I_0. From Figures 4.7.20(a) and 4.7.21(a) we see that for misaligned images, the representation points have larger distances than those for aligned images. Therefore the corresponding MST for the misaligned images has a longer length than that for the aligned images (Figures 4.7.20(b) and 4.7.21(b)).

We repeat the MST construction process over all the images in the image set $\{I_1^{(n)}\}$. The MST length is plotted in Figure 4.7.22. The x-axis stands for the image index, which corresponding to the viewing angles from the aircraft. The minimum of MST length indicates the best matching of the EO image and the reference image, which are shown in Figure 4.7.23.

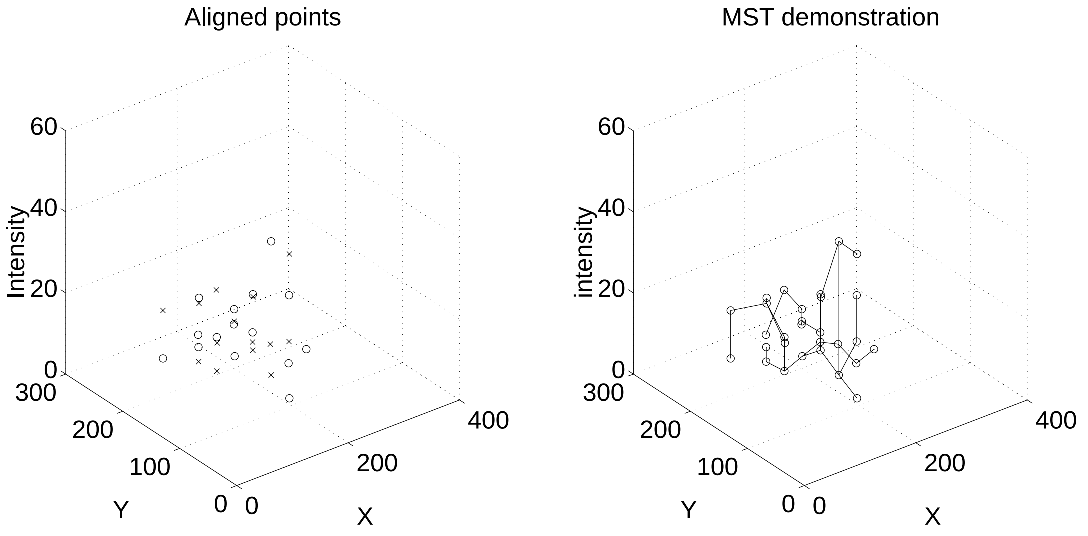
4.8 Noise Removal with k-MST Technique

So far we have not considered the noisy image registration yet. The existence of noise in the image will influence the construction of MST and hence affect the MST length.



(a) (b)

Figure 4.7.20: MST demonstration for misaligned images



(a) (b)

Figure 4.7.21: MST demonstration for aligned images

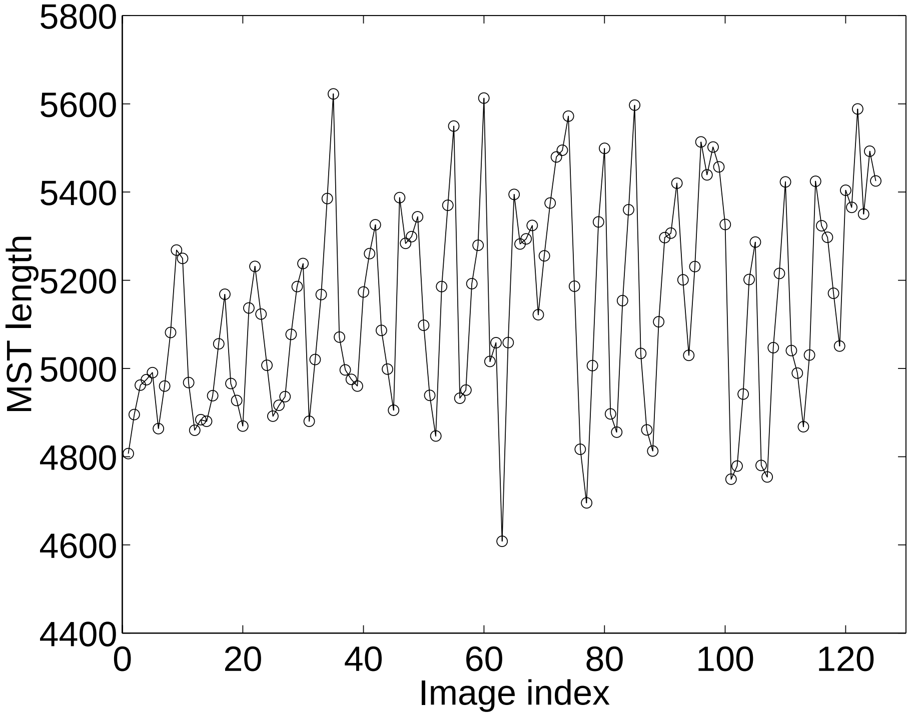
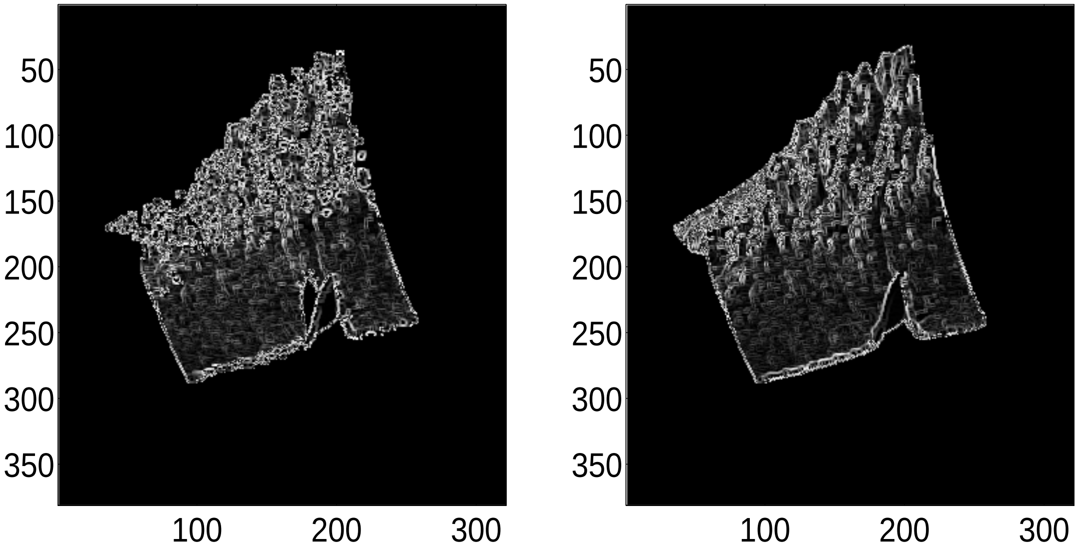


Figure 4.7.22: MST length for different test-reference image pairs



(a) Matching image

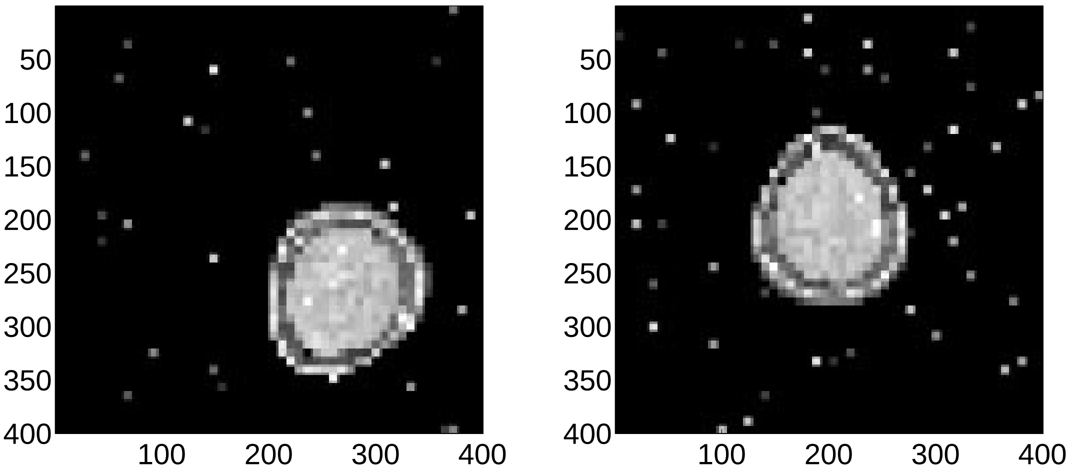
(b) Reference image

Figure 4.7.23: Result for EO-terrain map registration

In order to robustly register two noisy images, the first and crucial step is to remove the noise from the images. As we described in the Section 3.5, k-MST is a good candidate for noise removal. Since k-MST is time consuming, it is impractical to find the optimal sub-tree spanning a large number of nodes. Thus, noise removal will be implemented after the feature extraction. After applying k-MST algorithm on each feature image to remove outliers, we can employ the registration algorithm discussed in the preceding sections to register the noise-removed images.

To illustrate the registration of two noisy images, let us consider the registration of the images shown in Figure 4.4.6. The noise in the images is salt-and-pepper noise.

The feature vectors extracted via uniform sub-sampling method are shown in Figure 4.8.24.



(a) Pre-operation image

(b) Post-operation image

Figure 4.8.24: Sub-sampled brain images for registration

An illustration of the k-MST over the graph generated from pre-operation brain image I_1 is shown in Figure 4.8.25. Those unconnected vertices are further away from the center of the tree, which corresponds to the signal part of the mixture, and they are very unlikely observations from the underlying signal density. The k-MST total length as a function of

$n - k$, the number of pruned points, is plotted in Figure 4.8.26, from which we can easily determine the optimal knee, i.e., the number of pruned points, which is $n - \hat{k} = 23$ in this case. The pruned images are shown in Figure 4.8.27.

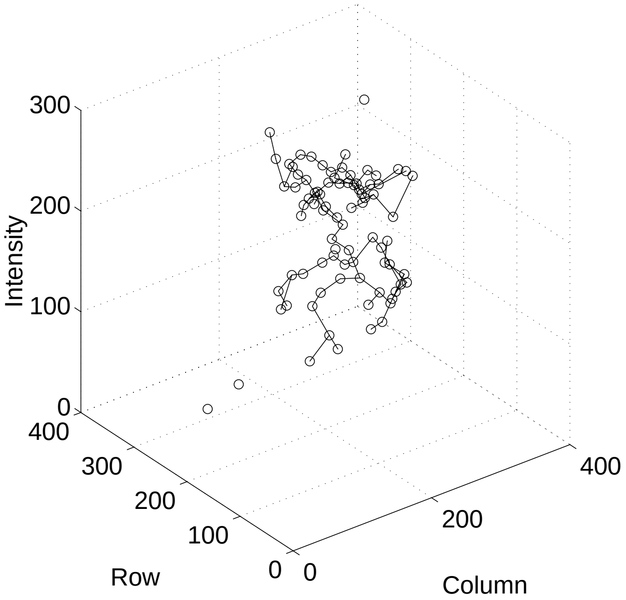


Figure 4.8.25: Illustration of k-MST for pre-operation brain image over the 3-dimensional domain of pixel position and grey scale

After pruning noisy points an initial alignment was performed by translating pre-operation brain image I_1 to make the centroids of I_1 and I_0 coincide. Then we rotate Image I_1, merge the two images and construct the MST on the resulting bitmaps. The MST length as a function of the rotation angle is plotted in Figure 4.8.28. The minimal MST length occurs at the rotation angle $214°$. The final result is shown in Figure 4.8.29. Figure 4.8.29(a) shows the matching result of I_1, while Figure 4.8.29(b) demonstrates the registration error defined as the difference between the matched image I_1 and the image I_0.

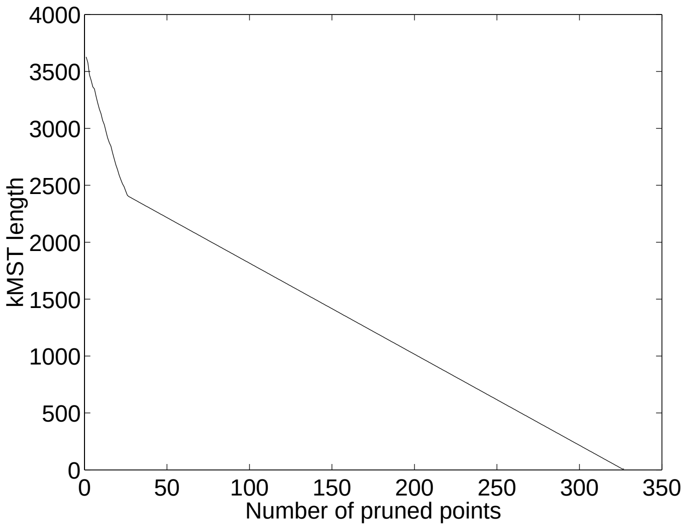
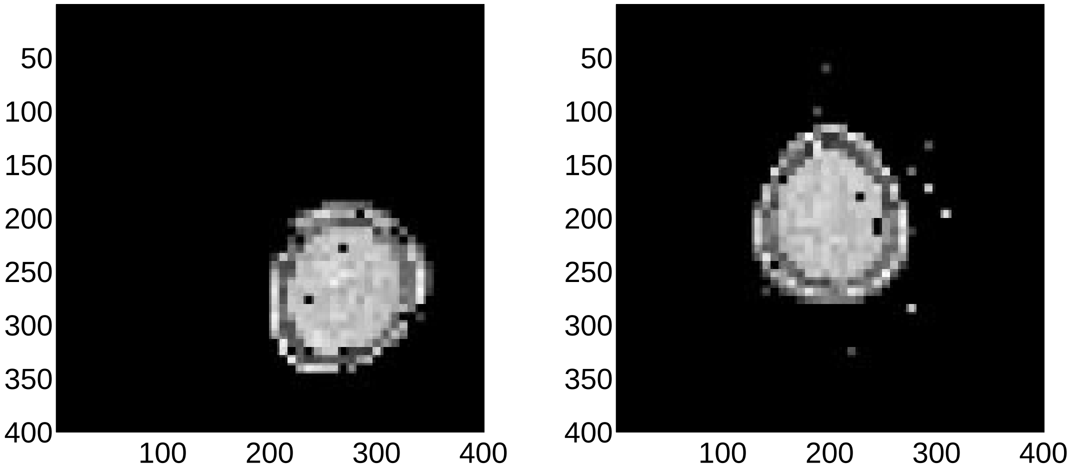


Figure 4.8.26: kMST length as a function of pruned number



(a) Pre-operation

(b) Post-operation

Figure 4.8.27: Sub-sampled images after outlier removal

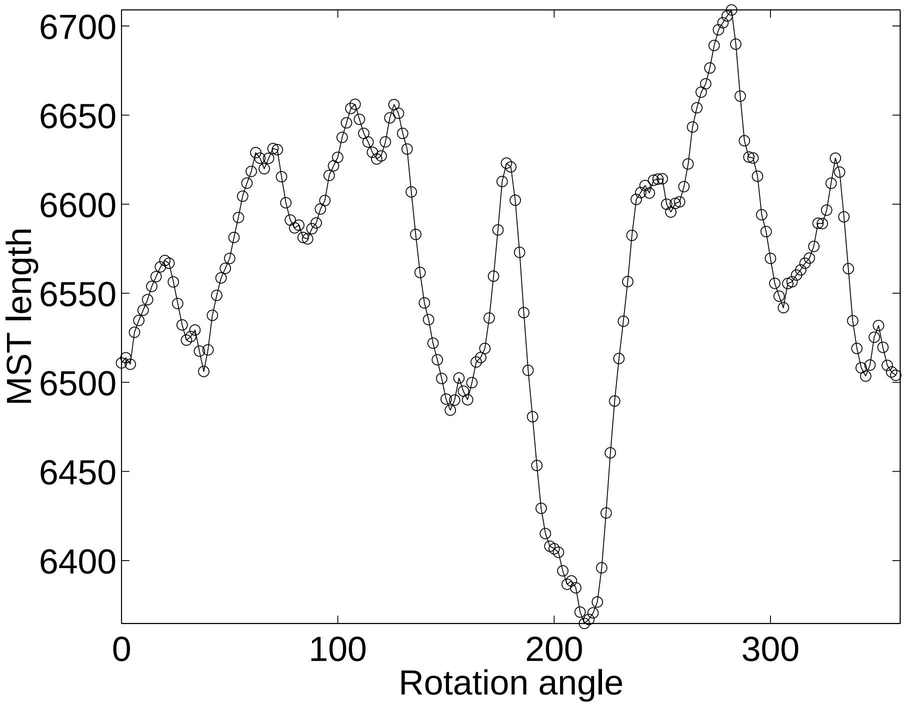
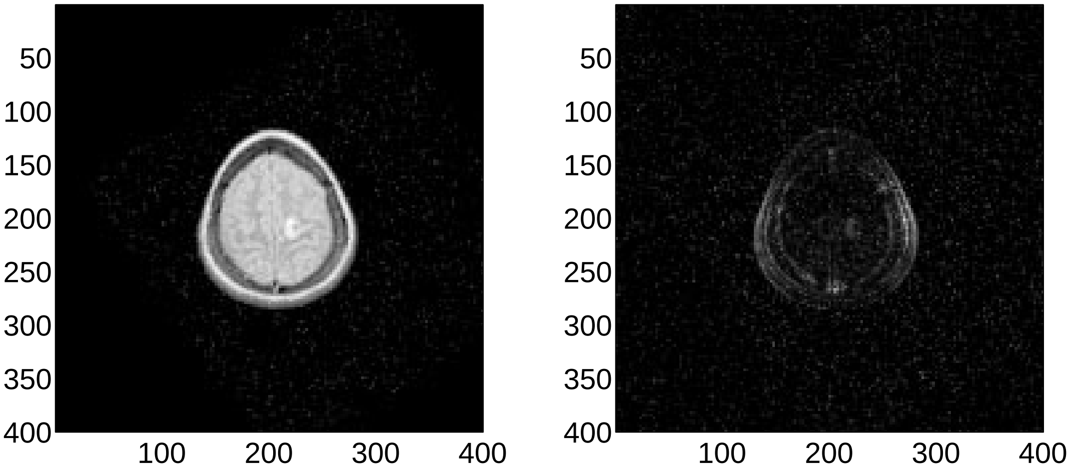


Figure 4.8.28: MST length as a function of rotation angle for brain image registration



(a) Registered pre-operation brain image

(b) Registration error

Figure 4.8.29: Registration result for brain images

4.9 Sliding Sampling Grid Approach

In order to improve the accuracy and robustness of the registration via MST matching, we proposed the following sliding sampling grid method:

① Coarsely sample the original images with high sampling rate to extract the feature vectors.

② Construct the MST over the graph generated from overlapped feature vectors.

③ Shift the sampling grid in a neighborhood and repeat Steps 1 and 2 for each shifting.

④ The summation of all MST lengths gives a dissimilarity metric between images.

We apply this registration technique to the geo-registration application. The sampling grid is shifted in a neighborhood of 3×3. The results are shown in Figures 4.9.30 and 4.9.31. Figure 4.9.30 shows the MST length functions for various shifted sampling grids where the numbers above each subplot are the shift positions in the row and column directions, respectively (We assume the first subplot is corresponding to the original sampling grid, which has (0,0) above the subplot)). Figure 4.9.31 shows the total MST length by adding the MST length over all subgraphs.

From the results, we observe that although in certain subgraphs, the minimum of the MST length does not occur for the registered image pair, the total MST length achieves its minimum for the registered image pair. The accuracy and robustness of the sliding sampling grid method are due to the fact that the individual MST length is correlated for the registered image pair, but not correlated for the misregistered image pair.

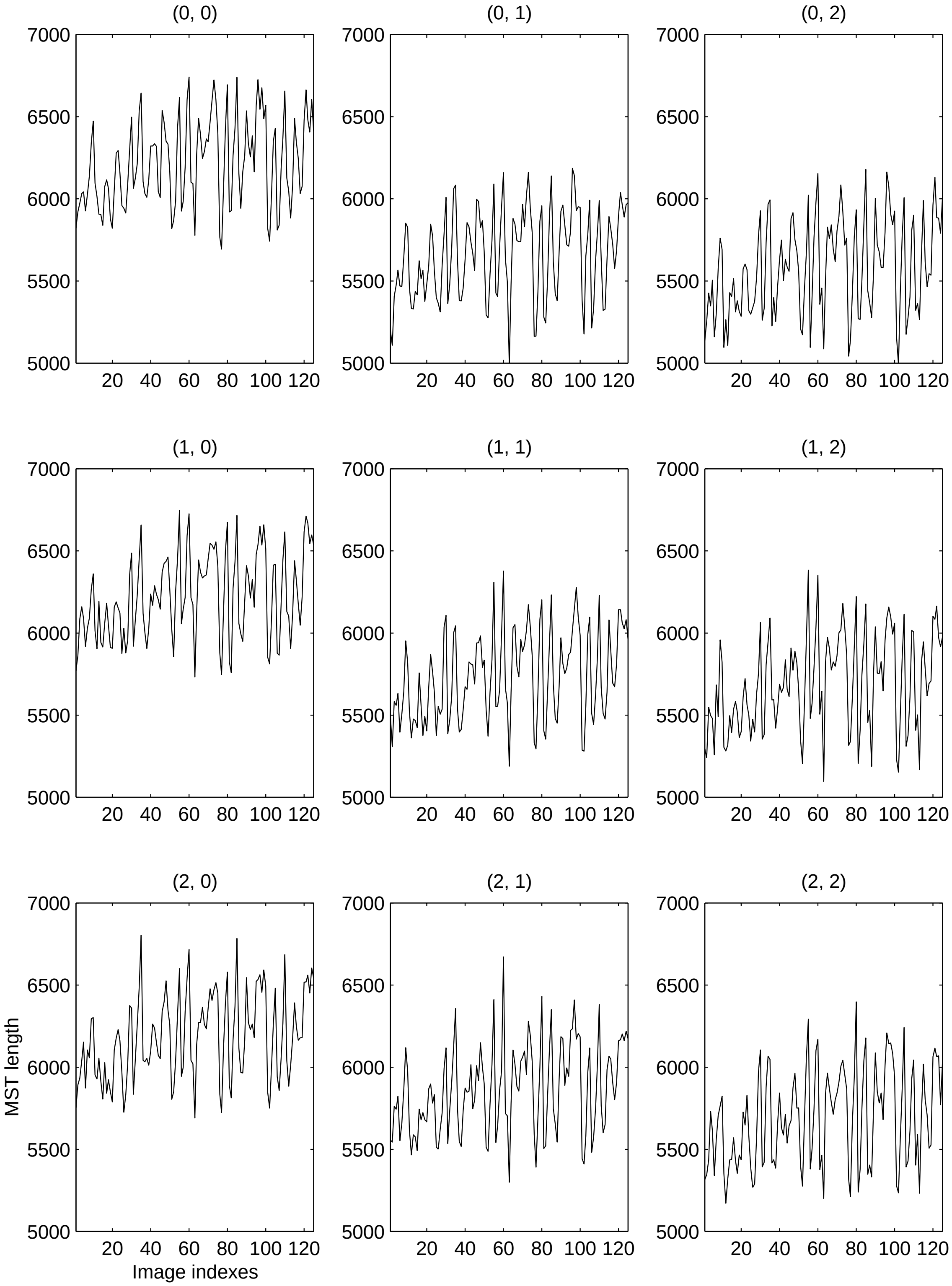


Figure 4.9.30: MST length over subgraphs

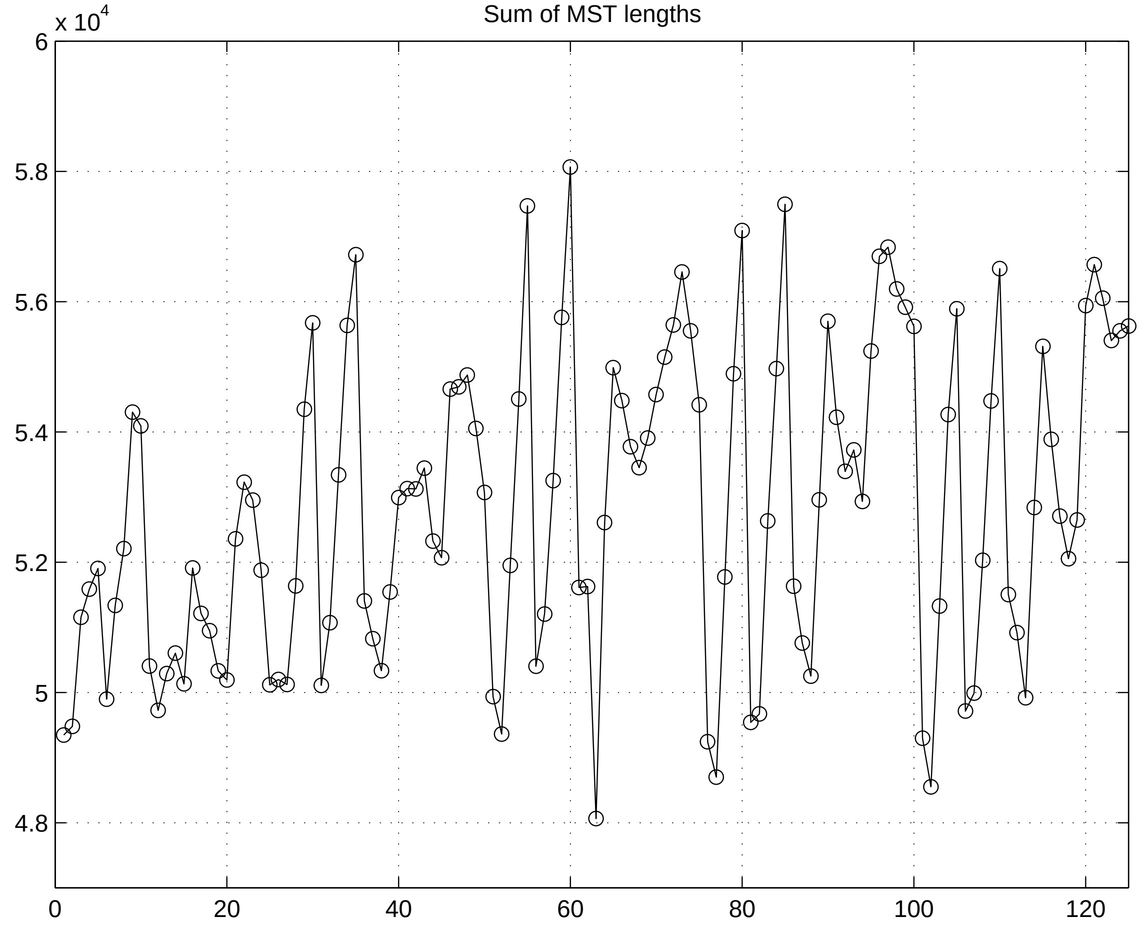


Figure 4.9.31: Total MST lengths versus test image indexes

4.10 Conclusions and Future Work

Registration is a fundamental task in image processing and quite a few registration techniques have been developed in various fields. In this chapter, we have proposed a graph-theoretic technique for image registration. The registration process is carried out on the feature vectors extracted from the original images. This registration algorithm uses the joint Rényi entropy of overlapping feature vectors as a dissimilarity metric between the two images. From the classical work of Steele [96], the Rényi entropy of feature vectors can be estimated with the power weighted length of the minimum spanning tree over the corresponding vertices in the generated graph. Thus we are able to make use of an equivalent dissimilarity metric, the MST length over the overlapping graphs generated from the original images, to align two images. Our method also takes advantage of the minimum k-point spanning tree approach to robustify the registration against outliers in the images. Since the k-MST length provides a robust and reliable distance measure between the images to be registered, this graph matching registration algorithm yields accurate registration results and is not sensitive to noise and small difference between images.

Comparison of Feature Extraction Strategies

Since the computational complexities of MST construction and approximate k-MST construction are both in polynomials of the number of vertices in the graph and the corresponding image size is usually large for practical use, feature extraction is a critical step for registering images with graph matching algorithms. In our work, we have proposed three strategies to extract feature vectors from the original to-be-registered images:

① extracting feature vectors with uniform spatial sub-sampling approach,

② extracting feature vectors with vector quantization algorithm, and

③ extracting feature vectors with stratified sampling with centroid refinements.

Extracting feature vectors with uniform spatial sub-sampling approach is straightforward and there is almost no additional computational load for the feature extraction process. The drawback of this approach is that the sub-sampling rate cannot be too high in order to keep enough information for registration process, and the number of vertices in the generated graph is still too large for effectively constructing the MST and approximate k-MST. Therefore, the feature vectors generated with the uniform sub-sampling approach are not satisfactorily effective. For the geo-registration application, the running time of the whole image registration process is about 24 minutes for satisfactory registration results. (All running times in this section are counted with Matlab codes.)

Vector quantization is a generalization of spatial sub-sampling. Since vector quantization takes into account the underlying densities of the feature vectors, it yields the most effective features among the three proposed strategies. Unfortunately, since a good initialization of representation vectors are not available in most applications, the LBG algorithm is time consuming. Thus the computational overhead caused by vector quantization procedure is not negligible compared to the complexity of constructing MST's and k-MST's. The image registration algorithm with this vector quantization approach consumes less time than the registration method with the uniform sub-sampling approach. For the geo-registration application, the running time of the whole image registration process is about 19 minutes for satisfactory registration results.

Extracting feature vectors with stratified sampling with centroid refinements is a low complexity approximation to the vector quantization based feature extraction approach. It takes advantage of the fact that most to-be-registered images have high information concentrated in the region of interest. Thus this method provides quite effective feature vectors and the registration process with this feature extraction method is much faster than the previous two approaches. For similar registration results, the registration process with stratified sampling with centroid refinements is implemented in about 11 minutes.

Future Work

We have seen that feature vectors extracted from the original images play a very important role in the graph matching image registration algorithm, both in registration accuracy and computational complexity. If the feature vectors well represents the information in the original images and their cardinality is not an obstacle for constructing MST and k-MST in a noisy environment, the graph matching registration method will achieve satisfactory results. Since “good” features are application dependent, it is difficult to find universally perfect features. However, it is known that some features are of great interest in image registration field. For example, boundaries of the objects, pixels with very high intensity, and rapidly changing regions. All these features contains great amount of information and should be helpful in registration process. In the future, we shall investigate the effectiveness and efficiency of more features and try to find a set of “good” features for some typical registration applications.

We should point out that for a given application, the extracted feature vectors used for matching images are not restricted to a single type. Then another direction of improving the image registration algorithms via graph matching is to employ multiple types of features. This multiple-type-feature registration method can be divided into two steps: first, a minimum spanning tree is constructed for each type of feature vectors; then a combining technique needs to be designed to appropriately incorporate the lengths of the MST’s over different features. The challenge of this method lies in the design of effective combining technique. It should take into account the dimensions and cardinalities of different types of feature vectors, and it should put more weights on the lengths of those MST’s that correspond to the more discriminative features. When we take advantage of more than one discriminative features in a single registration application, the number of the feature vectors in each type is usually small and it is easy and fast to construct the MST over those feature vectors. Thus, we can improve the accuracy of the graph-representation registration algorithm and reduce the computational complexity as well.

CHAPTER V

Sensor Systems for Road Boundary Detection

5.1 Road Boundary Detection Problem

Lane and pavement boundary detection is an enabling or enhancing technology which will have significant impact on the next generation of automotive systems such as road departure or lane excursion warning, intelligent cruise control, and ultimately autonomous driving. All of these applications have potential use in both military and civilian contexts. We shall give two examples to show the critical role of lane and pavement boundary detection. In the application of drowsy driver warning, knowledge of the pavement boundaries relative to the vehicle enables a driver assistance system to determine if the driver is running off the road. In the application of forward collision warning, pavement boundaries help disambiguate potential collision threats in terms of their relevance to the vehicle's intended path.

Lane and pavement boundary detection problem is particularly difficult when no prior knowledge of the road geometry is available (such as from previous time instants – see [22, 49, 75]) and when the detection algorithms have to locate the boundaries even in situations where there may be a great deal of clutter in the images.

Many gradient-based detection algorithms, which are applicable for structured edges including lane and pavement boundaries, apply a threshold to the image gradient magnitude to detect edges – see [50, 53, 54, 55] and the references therein. When a visual image has clearly visible lanes, and when the radar image has uniform regions with good separation between the regions, good performance can be obtained with these algorithms. However, real road scenes seldom give rise to such clean images –

① **Clutter**

Images may have structured noise, perhaps due to the presence of complex shadows (Figure 5.1.1).

Figure 5.1.1: Clutter due to presence of complex shadows

② **Missing data**

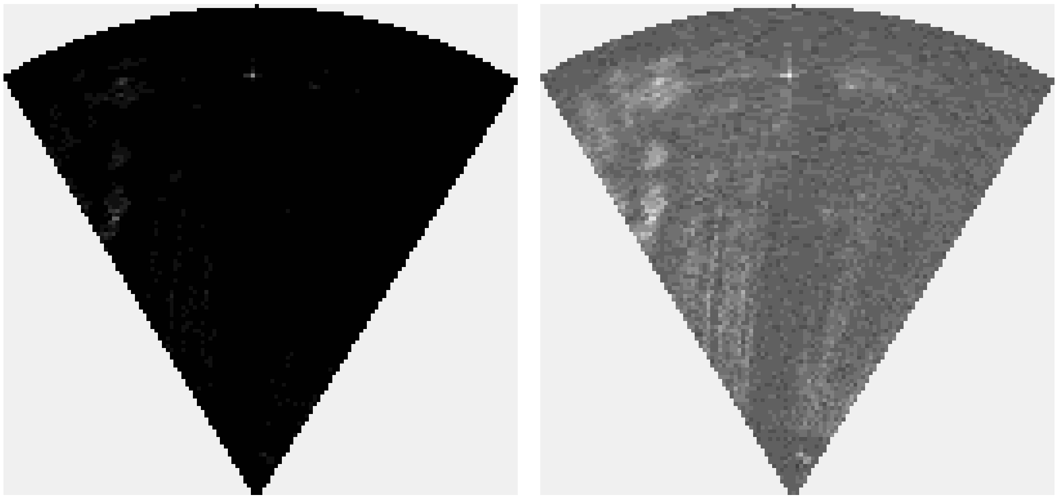
Valuable data might be missing, perhaps due to other vehicles occluding the boundaries of interest (Figure 5.1.2).

③ **Low SNR**

The signal-to-noise ratio might be inherently very poor perhaps due to the limitation of the (radar) imaging process (Figure 5.1.3). Figure 5.1.3(a) is the original raw radar data, while Figure 5.1.3(b) is the enhanced radar image obtained by nor-

Figure 5.1.2: Missing data due to other vehicles occluding the boundaries of interest

malizing the power values with respect to the maximum, taking one-fourth root of the pixel intensities in the raw radar image, and scaling the resultant values to a maximum of 255. Henceforth, all the radar images shown are enhanced versions of the corresponding raw radar images. However, all the processing is done on the raw(non-enhanced) radar images.



(a) Raw radar data

(b) Enhanced radar data

Figure 5.1.3: Low signal-to-noise ratio radar data

④ **Spurious boundaries**

Spatially significant non-lane or -pavement boundaries might be present, perhaps due to entry/exit ramps (Figure 5.1.4).

Figure 5.1.4: Non-lane or -pavement boundaries due to entry/exit ramps

Needless to say, it is difficult to select thresholds which eliminate noise edges without also eliminating many of the edge points of interest, and so the conventional edge detection algorithms described in [50, 53, 54, 55] are not suitable for our boundary detection problem under the above mitigating conditions.

A class of successful methods that overcome the thresholding problem are studied in [55, 56, 59, 68, 83]. These methods work directly with the image intensity array, as opposed to separately detected edge points, and use a global model of lane and pavement boundary shape. Two examples from this class are particularly relevant to our work:

① In Reference [56] the authors present a vision-based real-time algorithm called LOIS for locating lane and pavement boundaries using a deformable template global shape model. The global shape model adaptively adjusts and aligns a template so that it best matches the underlying features of the lane and pavement-boundary over the entire image. At the heart of LOIS is a matching function that encodes the knowledge that the edges of the lane should be near intensity gradients whose orientation are perpendicular to the lane edge. This allows strong magnitude gradients to be discounted if they are improperly oriented and weak magnitude gradients to be boosted if they are properly oriented. LOIS is shown to work well under a wide variety of conditions, including cases with strong mottled shadows and broken or interrupted lane markings, which pose a challenge for gradient-based lane detection schemes.

② In Reference [68] we present a method for detecting pavement boundaries in radar images. Like in LOIS, here too a deformable template model is used. The biggest difference though is in the matching function: [68] uses the log-normal probability model of the radar imaging process, which was proposed in [59]. This function encodes the knowledge that boundaries of the pavement should divide the image into three “relatively” homogeneous regions.

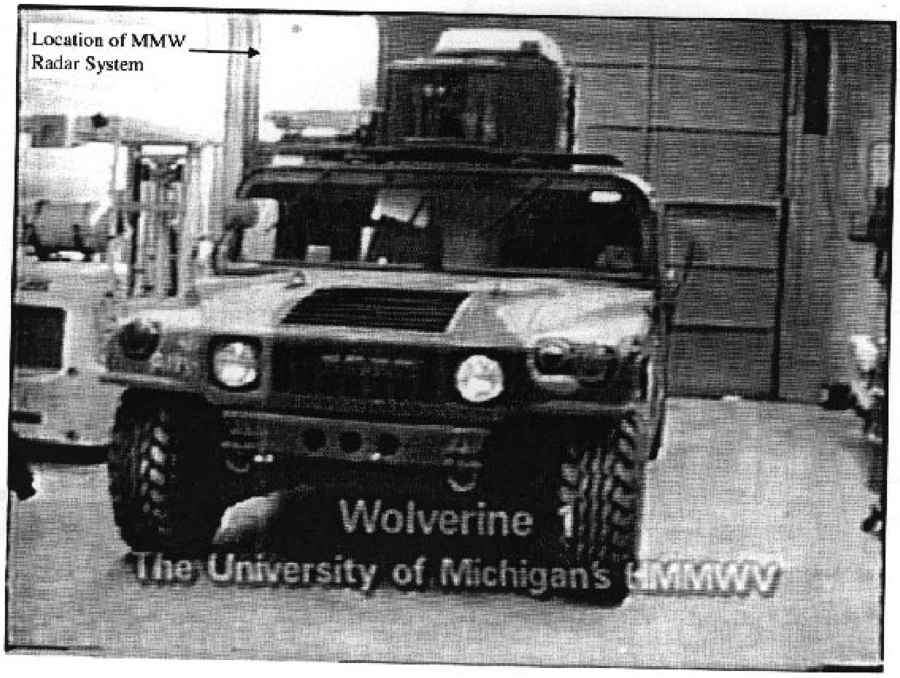
In both references [56] and [68] the boundary detection problem on hand is reformulated as a Bayesian estimation problem, where the deformable template model plays the role of a prior pdf and the matching function plays the role of a likelihood pdf, respectively.

Note that previously lane boundary detection in optical images [53, 54, 55, 56, 83], and pavement boundary detection in radar images [59, 48, 60, 68] have always been studied separately. However, a single sensor, either optical or radar sensor, limits itself in the ability to sense and identify the relevant features in varying environments. For example, the optical sensor is not able to operate in a poorly illuminated environment, while the radar sensor can not distinguish the lane markers on the road. To take advantage of the strengths (and overcome the weaknesses) of both the optical and radar sensors, it is natural to think of combining the two different types of sensed data together since multiple sensors will provide more information and hence a better and more precise interpretation of the sensed environment.

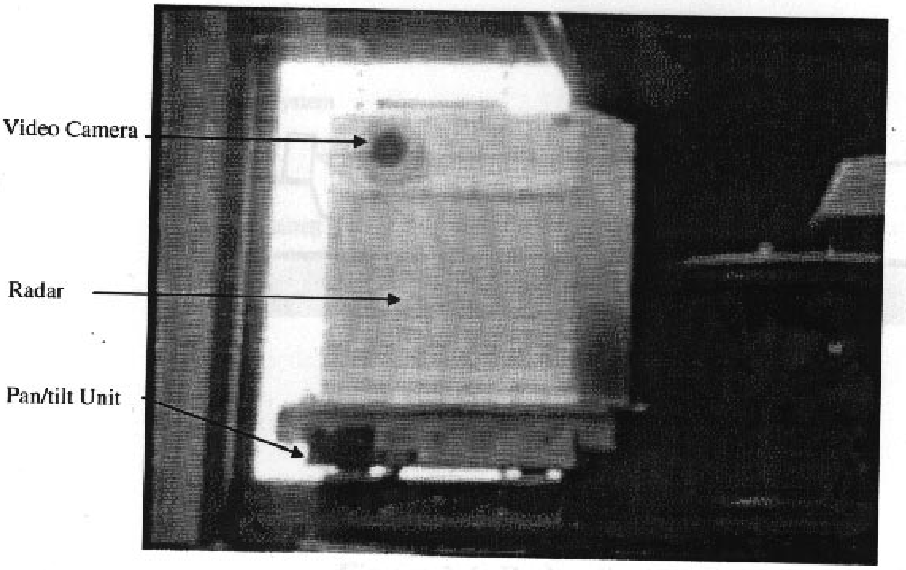
5.2 Sensor Systems

The radar data and visual images used in this work are all obtained from an imaging platform mounted on top of WOLVERINE I (*Wheeled Onroad Lab Vehicle Enabling Research Into New Environments*), a self-contained test-bed vehicle shown in Figure 5.2.5. The imaging sensors consist of a millimeter-wave radar sensor and a bore-sighted optical/vision sensor. The radar sensor is a 77GHz frequency modulated continuous wave (FMCW) radar sensor, with a maximum range of 128 meters (resolution 0.5 meters) and angular field of view of 64° (resolution 1°). The radar sensor setup is shown in Figure 5.2.6. The radar and optical sensors observe the same road scenario simultaneously in order to acquire a pair of co-registered images of that scene (Figures 5.2.7 and 5.2.8). The problem of interest is simultaneous detection of lane and pavement boundaries using observations from both the modalities. Consideration of both the modalities is meaningful because lane and pavement boundaries for the same road scene are highly correlated.

The information contained in the optical image depends on the reflectivity of the road illuminated by natural visible light. The lane boundaries, i.e., the white or yellow lane markers, constitute one of the two boundaries of interest to us and are clearly visible in the optical image if the image is obtained in a well-illuminated environment (see Figure 5.2.7(a)). The radar image is obtained by illuminating the road scene with electro-magnetic radiation in the millimeter-wave spectrum – see [48] and [59] for a detailed discussion of this image acquisition process. The relatively smooth road surface forward scatters much of this incident electro-magnetic power and hence returns very little power back to the radar; the side of the road, because it is made up of a coarser structure than the

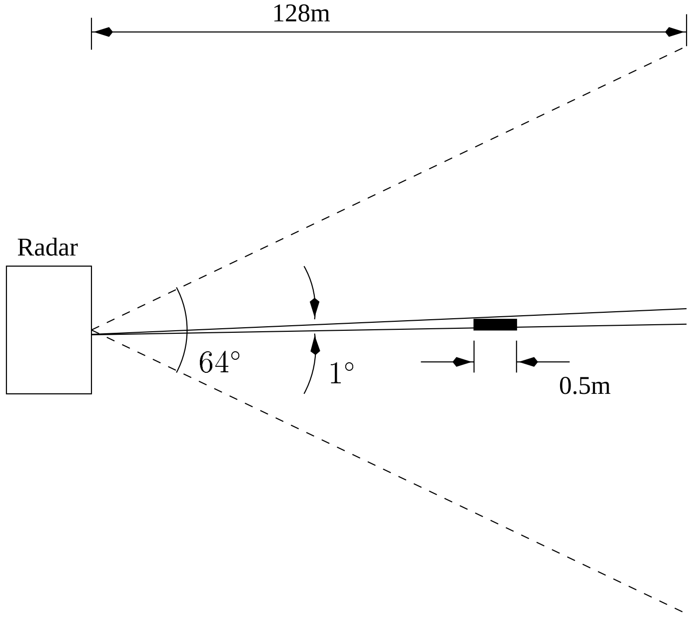


(a) WOLVERINE I

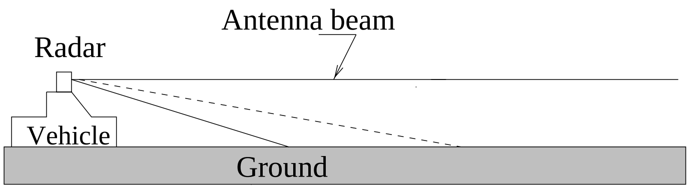


(b) imaging platform

Figure 5.2.5: WOLVERINE I

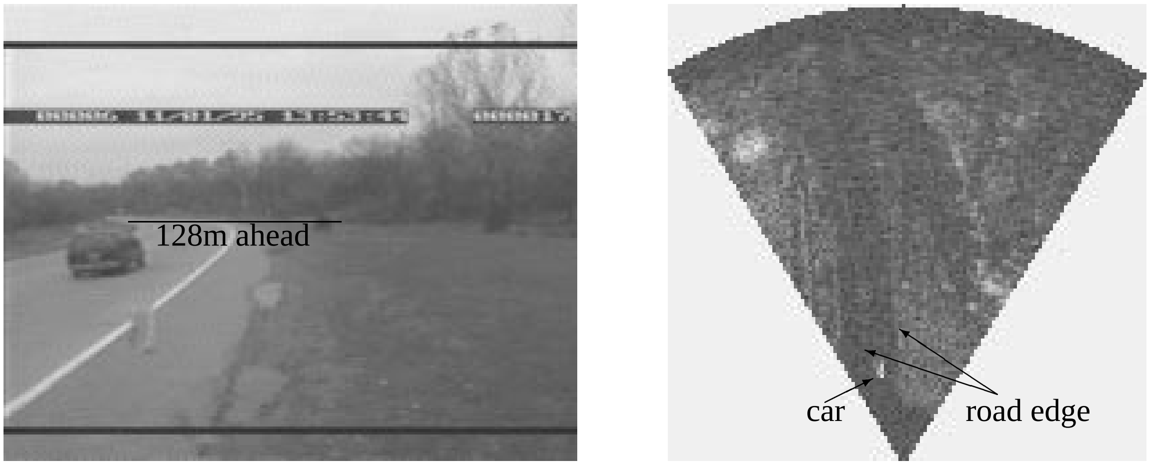


(a) Top view



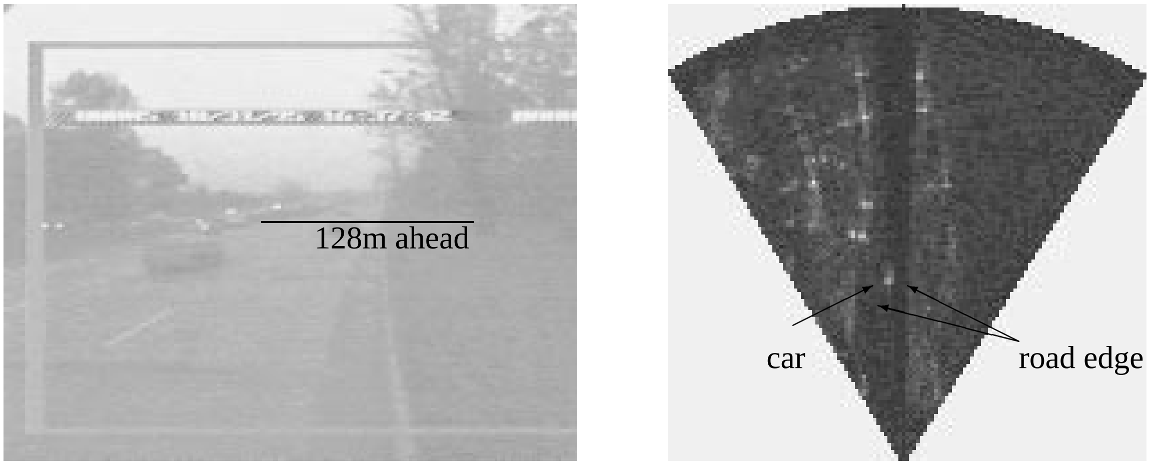
(b) Side view

Figure 5.2.6: Radar sensor setup



(a) Optical image (b) Radar image

Figure 5.2.7: A typical road scenario in clear/day-light weather



(a) Optical image (b) Radar image

Figure 5.2.8: A typical road scenario in foggy/low-light weather

road, returns a slightly higher amount of power. Thus, in the radar image, the road region is a little darker than the road sides, and the boundaries of the road delineate these three regions of different contrast (Figure 5.2.7(b)).

Evidently, either optical or radar sensor used alone has limited capabilities for resolving ambiguities and providing consistent descriptions of the road scenario due to the operating range and limitation which characterize the sensor. The optical sensor can provide high signal-to-noise ratio images in a well-illuminated environment such as a sunny day (Figure 5.2.7(a)). Such images, which clearly reflect the lane information, are sufficient for the lane boundary detection task. However, since it is a passive sensor and works at visible light wavelengths, in an ill-illuminated environment, e.g., at night or in foggy weather, the optical sensor will fail to provide sufficient information about the lane boundary (Figure 5.2.8(a). The radar sensor, on the contrary, being an active sensor and operating at millimeter wavelengths, has the ability to penetrate through rain, snow, fog, darkness, etc., i.e., it can operate under all weather conditions and provides an “alternate” image of the scenario in front of the vehicle (Figures 5.2.7(b) and 5.2.8(b)). Thus the radar image, regardless of the illumination situations, can give us the pavement boundary information, and the precise geometry of the pavement boundaries can be subsequently used in a number of driver warning and vehicle control tasks. The downside of the radar image though is its notoriously poor signal-to-noise ratio and low spatial resolution when compared to a visual image of the same road scene.

Since the optical and radar sensors provide different but complementary information about the road scene ahead of the vehicle, if the two types of information are combined appropriately and efficiently, the accuracy of the detected lane and pavement boundaries can be improved. The optical and radar fusion system shown in Figure 5.2.9 exploits this redundancy, diversity and complementarity between the two modalities.

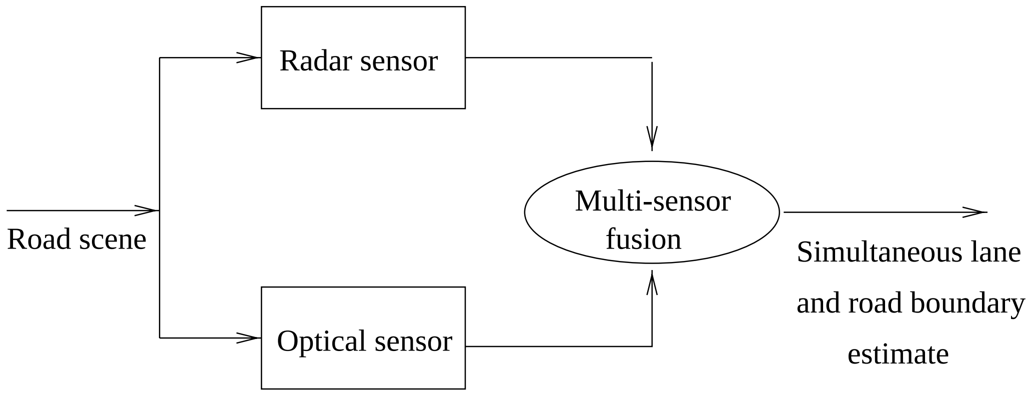


Figure 5.2.9: Diagram of the optical and radar multisensor fusion

Since two sensors, radar and optical sensors, are used to measure the same entity, the observational data are redundant. The redundancy of the data results from the fact that the lane and pavement boundaries in the visual and radar images, respectively, are highly correlated. Both lane and pavement boundaries are concentric arcs on the ground-plane and the lane boundaries are inside the road region, while the degradations of this boundary information introduced by the visual and radar imaging processes have completely different characteristics.

Since the radar and optical sensors measure the same scene with different laws of physics, we obtain physical sensor diversity. Since both sensors are placed at almost the same location, spatial diversity is not an issue in our work. However, it is because of the absence of spatial diversity that we can greatly simplify the registration process.

The optical sensor offers information about the lane boundaries, while the radar sensor provides information about the pavement boundaries. Thus each of them observes a subset of the environment space, and the union of these subsets makes up the whole road scenario. In this way, we achieve data complementarity.

CHAPTER VI

Fusion Algorithm for Lane and Pavement Boundary Detection with Existing Prior and Likelihood Models

We have seen that in previous work [56, 68], the authors studied separate lane or pavement boundary detection algorithms using only optical or radar images. In their boundary detection efforts, the lane and pavement boundaries are represented by parabolic curves, the radar imaging process is modeled with a log-normal pdf, and the optical imaging process is described with an empirical matching function. In this chapter, we propose a fusion algorithm for detecting lane and pavement boundaries using existing prior shape models and imaging likelihood functions.

6.1 Parabolic Models of Lane and Pavement Boundaries

In most cases, we can assume that *a priori* knowledge regarding the shape of the lane and pavement boundaries in the optical and radar images is available. A commonly used shape model for lane and pavement boundaries assumes that they can be approximated by concentric circular arcs on a flat ground plane. Such arcs, at least within a reasonable field of view, for small-to-moderate curvatures, are well approximated by parabolic curves on the ground plane,

$$x = \frac{1}{2}ky^2 + my + b \tag{6.1}$$

where the parameter k is the curvature of the arc, m is the tangential orientation, and b is the offset. While the radius of curvature and tangential orientation of left and right lane and pavement boundaries will differ slightly, constraining the left and right lane and pavement boundaries to have the same parameters k and m closely approximates the actual edge shapes for all but very small radii of curvature. So, it is assumed that concentric lane and pavement boundaries share the same parameters k and m. Then the only parameter that distinguishes the boundaries is the offset parameter b, that is, the left and right lane and pavement boundaries are characterized by different values of b.

The radar image is composed of reflections from the ground, and its domain is indeed the ground plane. So, (6.1) can be directly applied to model the shape of pavement boundaries in the radar image. The domain of the optical image, however, is a perspective projection of the ground plane, and therefore (6.1) needs to be rewritten in order to model the shape of lane boundaries in the image plane. Assuming a tilted pinhole camera perspective projection model, parabolic curves in the ground plane (6.1) transform into hyperbolic curves in the image plane:[1]

$$c = \frac{k'}{r - hz} + b'(r - hz) + vp' \tag{6.2}$$

where

$$\begin{aligned} k' &= \eta_k\, k, \\ vp' &= \eta_m\, m + \eta_{m,k}\, k + \eta, \text{ and} \\ b' &= \eta_b\, b + \eta_{b,m}\, m + \eta_{b,k}\, k. \end{aligned} \tag{6.3}$$

In other words, the k' parameter is linearly proportional to the curvature of the arc on the ground plane. The vp parameter is a function of the tangential orientation of the arc on the ground plane, with some coupling to the arc curvature as well. The b' parameter is

---

[1] See [54] for a derivation

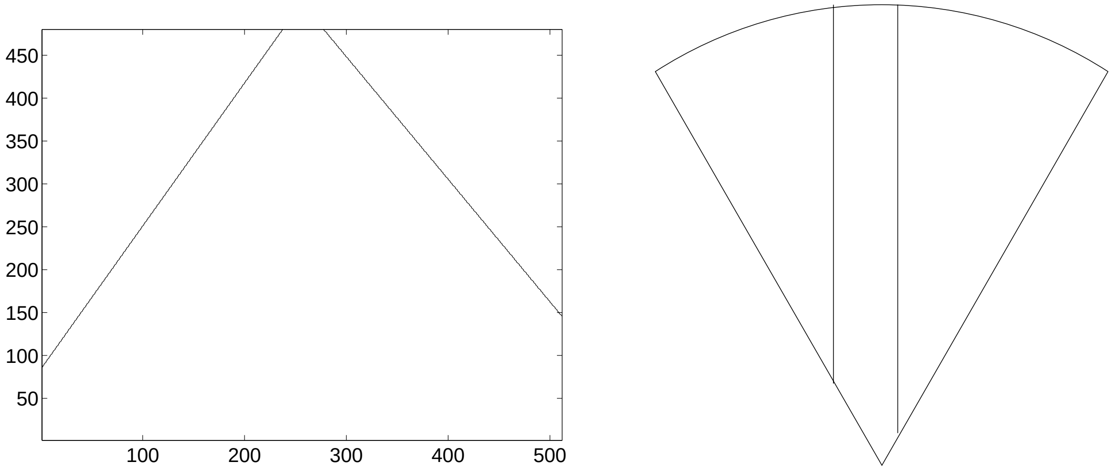
a function of the offset of the arc from the camera on the ground plane, with couplings to arc curvature and tangential orientation. The constants $\eta_k, \eta_m, \eta_{m,k}, \eta, \eta_b, \eta_{b,m},$ and $\eta_{b,k}$ depend on the camera geometry (resolution, focal length, height of the camera from the ground plane, and camera tilt).

Let $\underline{\theta}^o = \{k', vp', b'_L, b'_R\}$ and $\underline{\theta}^r = \{k, m, b_L, b_R\}$ denote the unknown lane and pavement boundaries' parameters, respectively. (In b_L, b_R, b'_L, b'_R the subscripts L and R indicates the offsets corresponding to the left and right boundaries.) Let $\underline{\theta} = \{\underline{\theta}^r, \underline{\theta}^o\}$ denote their adjoinment. By changing the values of $\underline{\theta}$, various lane and pavement boundary shapes can be realized – see Figure 6.1.1. The templates of the upper row in Figure 6.1.1 illustrate a straight road scenario with the deformation parameters $k' = 0,\ \ vp' = 256,\ \ b'_L = -0.6,\ \ b'_R = 0.7,\ \ hz = -30,\ \ b_L = -12.0,\ \ b_R = 2.4$. The templates of the lower row demonstrate a curved road scene with $k' = 600,\ \ vp' = 340,\ \ b'_L = -3.2,\ \ b'_R = -1.0,\ \ hz = 200,\ \ b_L = -14.2,\ \ b_R = 0.4.$

The problem of simultaneous detection of the lane and pavement boundaries is now equivalent to the problem of estimating $\underline{\theta}$. The elements of $\underline{\theta}$ have to satisfy some constraints, and for some elements of $\underline{\theta}$ the range of physically meaningful values they can possibly assume is known *a priori*. Given a hypothetical $\underline{\theta}$, its fidelity to the observed optical and radar images can also be assessed. In this report, we choose a probabilistic Bayesian framework to express the constraints, the *a priori* beliefs, and the assessment of fidelity to data.

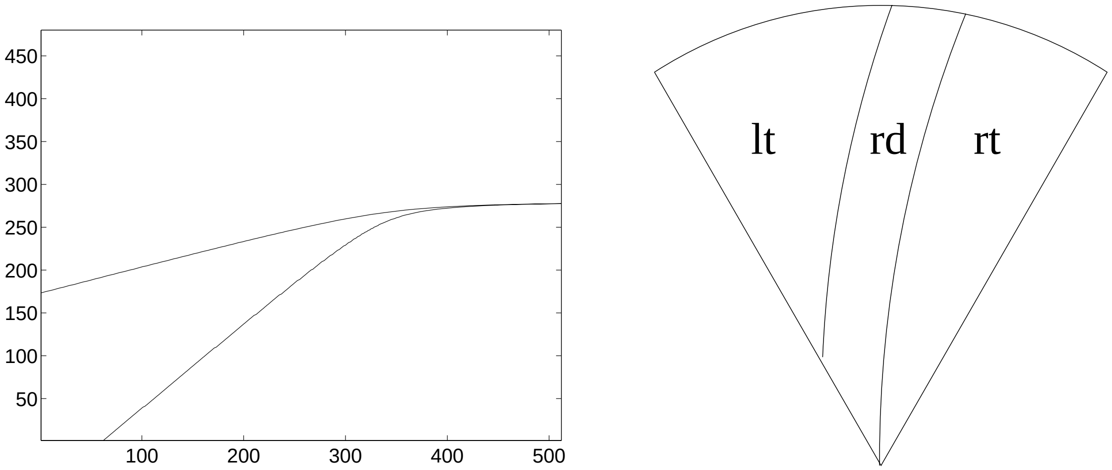
We present the so-called prior pdf here (the likelihood pdf's are in the next chapter):

$$\begin{aligned} p(\underline{\theta}) &= p(\underline{\theta}^o, \underline{\theta}^r) \\ &= I_{b'_L > b_L}(b'_L, b_L) \ \times\ I_{b'_R < b_R}(b'_R, b_R) \times\ \delta(k' - \eta_k\, k) \\ &\quad \times \delta(vp' - [\eta_m\, m + \eta_{m,k}\, k + \eta]) \end{aligned}$$



(a) optical domain

(b) radar domain



(c) optical domain

(d) radar domain

Figure 6.1.1: Boundary templates for the optical and radar images

$$\times \frac{2}{\pi}\,\text{atan}[\beta_{o,1} \times (b'_R - b'_L)] \;\times\; \frac{2}{\pi}\,\text{atan}\left[\frac{\beta_{o,2}}{b'_R - b'_L}\right]$$
$$\times \frac{2}{\pi}\,\text{atan}[\beta_{r,1} \times (b_R - b_L)] \;\times\; \frac{2}{\pi}\,\text{atan}\left[\frac{\beta_{r,2}}{b_R - b_L}\right] \tag{6.4}$$

where $I_A(x,y)$ is an indicator function,

$$I_A(x,y) = \begin{cases} 1, & \text{if } (x,y) \text{ satisfies relation } A \\ 0, & \text{otherwise} \end{cases} \tag{6.5}$$

and $\delta(x)$ is the Kronecker delta function,

$$\delta(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.6}$$

The terms on the first two lines of (6.4)'s RHS, correspond to the constraints that the elements of $\underline{\theta}$ have to satisfy. The first two terms impose the constraint that the lane markers be contained within the pavement region, the last two terms impose the constraint that the lane boundaries' curvature and orientation be precisely related to the pavement boundaries' curvature and offset via (6.3). The terms on the last two lines of (6.4)'s RHS expresses the *a priori* beliefs that lanes and pavements can be neither be too narrow nor too wide.

6.2 Imaging Likelihoods

6.2.1 Radar Imaging Likelihood

The true boundaries of the pavement separate the observed radar image into three homogeneous regions associated with the road surface, the left side of the road, and the right side of the road. So, given a parameter set $\underline{\theta}^r$, its fidelity to the observed radar image is assessed by how homogeneous the corresponding three (road, left, and right) regions are. A log normal probability law [2] is used to derive the homogeneity criteria.

We caution the reader that the radar returns over the the left, right and road regions are not truly homogeneous. Sources for non-homogeneity including point-like scatterers (cars, sign posts, retro-reflectors, etc.), changes in the off-road surface (grass, gravel, etc.) and presence of periodic off-road structures (bridges, fences, trees, etc.). Modeling all such variability is impossible. The log-normal pdf is meant to be a reasonable but low-complexity (two parameters per region) approximation to the actual variations in the data.

The rationale for using the log-normal law, as opposed to normal, exponential, or Rayleigh laws, is due to previous studies [27, 89, 62]. To appreciate the appropriateness of the log-normal pdf to describe radar returns, we refer the reader to Fig. 6.2.2, which demonstrates that log-normal is an excellent approximation to the radar return.

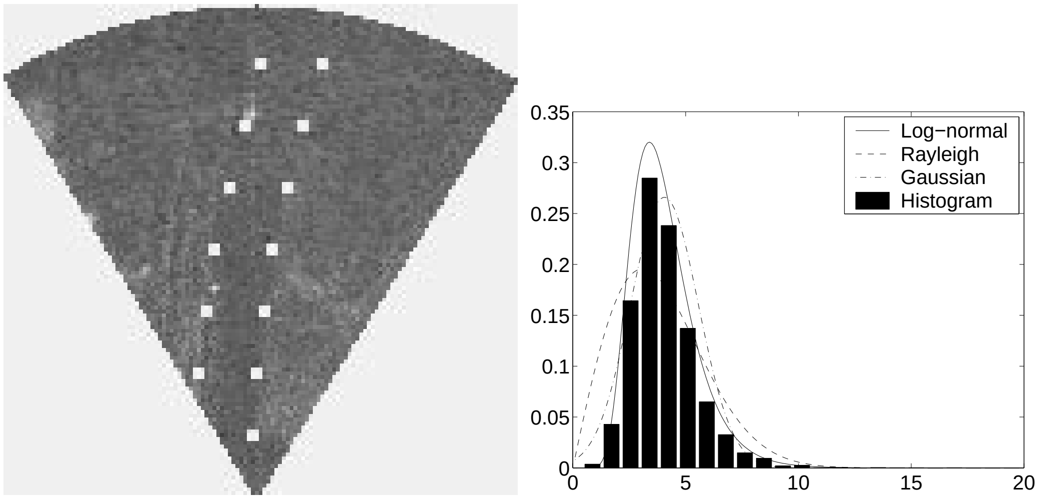
Let $\mathcal{L} = \{(r, \phi),\ 1 \le r \le r_{max},\ \phi_{min} \le \phi \le \phi_{max}\}$ denote the range and azimuth coordinates of the pixels in the millimeter-wave radar image Z^r. Given the parameters $\underline{\theta}^r$ for pavement boundaries, the conditional probability of Z^r taking on a realization z^r (corresponding to a single observation) is given by

$$p(z^r \mid \underline{\theta}^r) = \prod_{(r,\phi)\in\mathcal{L}} \frac{1}{z^r_{r\phi}\sqrt{2\pi\sigma^2_{r\phi}(\underline{\theta}^r)}} \exp\left\{-\frac{1}{2\sigma^2_{r\phi}(\underline{\theta}^r)}\left[\log z^r_{r\phi} - \mu_{r\phi}(\underline{\theta}^r)\right]^2\right\} \tag{6.7}$$

where $\mu_{r\phi}(\underline{\theta}^r),\ \sigma^2_{r\phi}(\underline{\theta}^r)$ denote the mean and variance of the region to which the pixel (r, ϕ) belongs. In (6.7) $\underline{\theta}^r$ is explicitly referred to emphasize the dependencies of means $\mu_{r\phi}$ and variances $\sigma^2_{r\phi}$ on the unknown parameters $\underline{\theta}^r$. However, henceforth, in order to make the representations concise, we will omit the explicit references to $\underline{\theta}^r$.

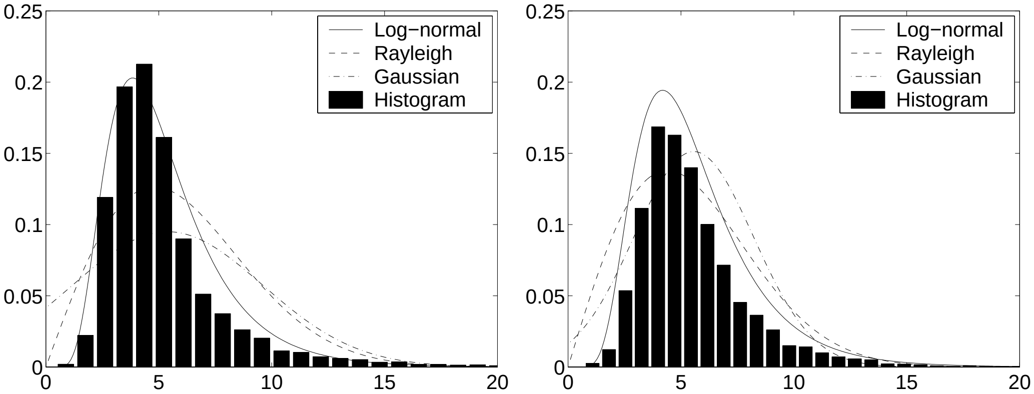
In its present form, (6.7) is not very useful for assessing the data fidelity of $\underline{\theta}^r$, due to the presence of nuisance parameters $\mu_{r\phi}$ and $\sigma^2_{r\phi}$. In the sequel, (6.7) is rewritten as a function of these nuisance parameters, and it reveals an intuitively appealing homogeneity criterion. Some additional notations are necessary for further derivation:

- Let N denote the number of pixels in $\mathcal{L}$.



(a) Radar image

(b) Road region



(c) Left-side region

(d) Right-side region

Figure 6.2.2: Rationale for using log-normal pdf. (a) shows a radar image overlayed with the correct positions for the left and right pavement boundaries. (b), (c), and (d) show histograms of the actual radar returns for the three regions – the road, the left-side and right-side of the road. Also shown in (b), (c), and (d) are the maximum likelihood fits of the log-normal, Rayleigh, and Gaussian pdf's to the radar return histograms.

- Also let $\mathcal{L}^{rd}$, $\mathcal{L}^{lt}$, $\mathcal{L}^{rt}$ denote the Cartesian coordinates of the pixels in the road, left-side and right-side regions, respectively, and $(\mu^{rd}, [\sigma^{rd}]^2, N^{rd})$, $(\mu^{lt}, [\sigma^{lt}]^2, N^{lt})$ and $(\mu^{rt}, [\sigma^{rt}]^2, N^{rt})$ denote the means, variances and the numbers of pixels of the corresponding regions.

- Define

$$\begin{aligned} \underline{\mu} &= \left[\mu^{rd} \quad \mu^{lt} \quad \mu^{rt}\right]^T, \\ \underline{\sigma}^2 &= \left[[\sigma^{rd}]^2 \quad [\sigma^{lt}]^2 \quad [\sigma^{rt}]^2\right]^T, \\ \log \underline{Z}^r &= \left[\log z^r_{1\phi_{min}}, \cdots, \log z^r_{1\phi_{max}}, \log z^r_{2\phi_{min}}, \cdots, \right. \\ &\quad \left. \log z^r_{2\phi_{max}}, \cdots, \log z^r_{N\phi_{min}}, \cdots, \log z^r_{N\phi_{max}}\right]^T, \end{aligned}$$

and

$$\bar{Z}^r = \prod_{(r,\phi)\in\mathcal{L}} z^r_{r\phi}$$

- Finally, let $\mathcal{I}$ denote an indicator matrix defined as follows

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}(1,1) & \mathcal{I}(1,2) & \mathcal{I}(1,3) \\ \mathcal{I}(2,1) & \mathcal{I}(2,2) & \mathcal{I}(2,3) \\ & \vdots & \\ \mathcal{I}(N,1) & \mathcal{I}(N,2) & \mathcal{I}(N,3) \end{bmatrix}$$

where for every p such that $1 \le p \le N$, if p belongs to j-th region,

$$\mathcal{I}(p,k) = \begin{cases} 1, & \text{if } k = j, \; k \in \{1,2,3\} \\ 0, & \text{if } k \neq j, \; k \in \{1,2,3\} \end{cases}$$

Now (6.7) can be expressed in terms of the above notations:

$$p(z^r \mid \underline{\theta}^r) = \left[\prod_{(r,\phi)\in\mathcal{L}} \frac{1}{z^r_{r\phi}}\right] \prod_{(r,\phi)\in\mathcal{L}} \frac{1}{\sqrt{2\pi\sigma^2_{r\phi}}} \exp\left\{-\frac{1}{2\sigma^2_{r\phi}}[\log z^r_{r\phi} - \mu_{r\phi}]^2\right\}$$

$$\begin{aligned}
&= \frac{1}{\bar{Z}^r} \prod_{(r,\phi)\in\mathcal{L}_{rd}} \frac{1}{\sqrt{2\pi[\sigma^{rd}]^2}} \;\exp\left\{ -\frac{1}{2[\sigma^{rd}]^2} \,[\log z^r_{r\phi} - \mu^{rd}]^2 \right\} \\
&\quad \times \prod_{(r,\phi)\in\mathcal{L}_{lt}} \frac{1}{\sqrt{2\pi[\sigma^{lt}]^2}} \;\exp\left\{ -\frac{1}{2[\sigma^{lt}]^2} \,[\log z^r_{r\phi} - \mu^{lt}]^2 \right\} \\
&\quad \times \prod_{(r,\phi)\in\mathcal{L}_{rt}} \frac{1}{\sqrt{2\pi[\sigma^{rt}]^2}} \;\exp\left\{ -\frac{1}{2[\sigma^{rt}]^2} \,[\log z^r_{r\phi} - \mu^{rt}]^2 \right\} \\
&= \frac{1}{\bar{Z}^r} \frac{1}{(2\pi)^{N/2} \|\mathrm{diag}[\mathcal{I}\,\underline{\sigma}^2]\|^{1/2}} \\
&\qquad \times \exp\left\{ -\frac{1}{2} \left[\log \underline{Z}^r - \mathcal{I}\,\underline{\mu}\right]^T \left[\mathrm{diag}(\mathcal{I}\,\underline{\sigma}^2)\right]^{-1} \left[\log \underline{Z}^r - \mathcal{I}\,\underline{\mu}\right] \right\}
\end{aligned} \tag{6.8}$$

Given a hypothetical shape of the pavement boundaries, $\underline{\theta}^r$, the nuisance parameters $\underline{\mu}$ and $\underline{\sigma}^2$, corresponding to the means and variances of the three regions, can be empirically estimated from the observed radar image z^r by a maximum likelihood method. Note that the likelihood is in a normal form, and so the maximum likelihood and least squares estimate of the nuisance parameters are equivalent,

$$\begin{aligned}
\hat{\underline{\mu}} &= \left[\mathcal{I}^T\mathcal{I}\right]^{-1} \mathcal{I}^T \log \underline{Z}^r \\
\hat{\underline{\sigma}}^2 &= \left[\mathcal{I}^T\mathcal{I}\right]^{-1} \mathcal{I}^T \left(\log \underline{Z}^r - \mathcal{I}\,\hat{\underline{\mu}}\right)^T \left(\log \underline{Z}^r - \mathcal{I}\,\hat{\underline{\mu}}\right)
\end{aligned} \tag{6.9}$$

Substituting these estimates of $\underline{\mu}$ and $\underline{\sigma}^2$ back into (6.8), and taking the logarithm results in

$$\begin{aligned}
\log p(z^r|\underline{\theta}^r) = &-N^{rd}\log\hat{\sigma}^{rd} - N^{lt}\log\hat{\sigma}^{lt} \\
&-N^{rt}\log\hat{\sigma}^{rt} - \log\bar{Z}^r - \frac{1}{2}N(1+\log 2\pi)
\end{aligned} \tag{6.10}$$

This result leads to the following interpretation:

> Given a hypothetical pavement boundary shape $\underline{\theta}^r$, the fidelity of this shape to the observed radar image z^r is assessed by the sum of the logarithm of the variances of the observed image over the corresponding three (road, left-side,

and right-side) regions. The smaller the sum of these variances, the better the fidelity.

Weighted Log-normal Likelihood

It is indicated in (6.10) that part of the objective is trying to maximize $\log p(z^r|\underline{\theta}^r)$, i.e., to minimize $N^{rd}\log\hat{\sigma}^{rd} + N^{lt}\log\hat{\sigma}^{lt} + N^{rt}\log\hat{\sigma}^{rt}$, i.e., trying to minimize the standard deviations, and hence, variances of the three regions. However, in the radar image, it is clear that the backscatter distribution of the pavement is virtually homogeneous while the backscatter distributions of the roadsides are much more complicated. Therefore, we might prefer having an even smaller variance in the road region at the price of having slightly larger variances in the roadside regions.

In the above method, the variances are weighted proportionately to the number of pixels in their respective regions. In order for the standard deviation of the road, $\hat{\sigma}^{rd}$, to weigh more heavily, in terms of its contribution to the likelihood, N^{rd} would have to be proportionately large. The same is true for the other two regions as well.

In order to re-enforce our *a priori* belief that road pixels tend to be homogeneous (at least compared to the pixels belonging to either side of the road), and to overcome the undue influence of bright point scatterers in the roadside regions, we propose a new radar imaging likelihood function that gives the region of the road a different weight w^{rd} from those given to the roadside regions.

$$\begin{aligned}\log p(z^r|\underline{\theta}^r) = &-w^{rd}N^{rd}\log\hat{\sigma}^{rd} - N^{lt}\log\hat{\sigma}^{lt} \\ &-N^{rt}\log\hat{\sigma}^{rt} - \log\bar{Z}^r - \frac{1}{2}N(1+\log 2\pi)\end{aligned} \tag{6.11}$$

In this report, we will utilize (6.11) as the log-likelihood of radar imaging process in a Bayesian estimation scheme. To make the representation of radar imaging log-likelihood function concise, we rewrite (6.11) as

$$\log p(z^r|\underline{\theta}^r) = L^r(z^r, \underline{\theta}^r) - c^r \tag{6.12}$$

where

$$L^r(z^r, \underline{\theta}^r) \;=\; -\left(w^{rd} N^{rd} \log \hat{\sigma}^{rd} + N^{lt} \log \hat{\sigma}^{lt} + N^{rt} \log \hat{\sigma}^{rt}\right) \tag{6.13}$$

$$c^r \;=\; \log \bar{Z}^r + \frac{1}{2} N(1 + \log 2\pi) \tag{6.14}$$

$L^r(z^r, \underline{\theta}^r)$ is the effective matching part of the log-likelihood function while c^r is a constant for a specified image and thus can be neglected in the estimation process. From now on, we call $L^r(z^r, \underline{\theta}^r)$ as radar matching function.

6.2.2 Optical Imaging Likelihood

In [56] the authors give a fairly robust optical imaging likelihood function. In our work we make a little modification of their proposed likelihood and describe the imaging likelihood based on an energy function, which directly interprets the observed optical image without regard to how it is formed. The energy function, and hence the likelihood, encodes the knowledge that the edges of the lane should be near intensity gradients whose orientation should be perpendicular to the lane edge. More specifically, given a hypothetical parameter set of underlying edges $\underline{\theta}^o = \{k', vp', b'_L, b'_R\}$, we assume that the likelihood of observing the optical image Z^o is given by

$$p(z^o|\underline{\theta}^o) = \gamma(\underline{\theta}^o)\; e^{-E^o(z^o, \underline{\theta}^o)} \tag{6.15}$$

where $E^o(z^o, \underline{\theta}^o)$ denotes an energy function and $\gamma(\underline{\theta}^o)$ is a normalizing constant for a given optical deformation parameter set $\underline{\theta}^o$,

$$\gamma(\underline{\theta}^o) = \frac{1}{\int e^{-E^o(z^o, \underline{\theta}^o)} dz^o} \tag{6.16}$$

The energy function $E^o(z^o, \underline{\theta}^o)$ can be described in the following steps:

- Define the Cauchy density function

$$f(\alpha, x) \triangleq \frac{\alpha}{\pi} \frac{1}{1 + \alpha^2 x^2} \tag{6.17}$$

- Let $g_m(r, c)$ be the gradient magnitude at pixel (r, c), with $g_d(r, c)$ being the corresponding gradient direction. Instead of the observed optical image, the gradient features are used in the representation of the energy function.

- Define the edges of the lane in the image by the curves:

$$\begin{aligned} S_L(r, c, \underline{\theta}^o) &\triangleq \frac{k'}{r - hz} + b'_L(r - hz) + vp' \\ S_R(r, c, \underline{\theta}^o) &\triangleq \frac{k'}{r - hz} + b'_R(r - hz) + vp' \end{aligned} \tag{6.18}$$

Given these definitions, the energy function of observing an image gradient field given a set of lane shape parameters $\underline{\theta}^o$ is

$$E^o(z^o, \underline{\theta}^o) = E^o(g_m, g_d, \underline{\theta}^o) \tag{6.19}$$

$$\begin{aligned} = \; & -\sum_{(r,c)} g_m(r, c) \times f(\alpha_m, c - S_L(r, c, \underline{\theta}^o)) \\ & \quad \times f(\alpha_d, \cos(g_d(r, c) - \text{atan}(\frac{d}{dr} S_L(r, c, \underline{\theta}^o)))) \\ & -\sum_{(r,c)} g_m(r, c) \times f(\alpha_m, c - S_R(r, c, \underline{\theta}^o)) \\ & \quad \times f(\alpha_d, \cos(g_d(r, c) - \text{atan}(\frac{d}{dr} S_R(r, c, \underline{\theta}^o)))) \end{aligned} \tag{6.20}$$

In other words, the contribution made by a pixel to the energy is the gradient magnitude at that pixel, multiplied by a function whose value decreases as the pixel column gets further from the lane edge and a function whose value decreases as the gradient direction at the pixel becomes less perpendicular to the tangent to the lane edge.

The logarithm of the optical imaging likelihood is

$$\log p(z^o | \underline{\theta}^o) = -E^o(z^o, \underline{\theta}^o) + \log \gamma(\underline{\theta}^o) \tag{6.21}$$

Define the optical matching function $L^o(z^o, \underline{\theta}^o)$ as the negative of the energy function $E^o(z^o, \underline{\theta}^o)$,

$$L^o(z^o, \underline{\theta}^o) = -E^o(z^o, \underline{\theta}^o) \tag{6.22}$$

6.3 Joint MAP Estimate for Lane and Pavement Boundaries

Since the prior distribution of the deformation parameters and the imaging likelihood functions are available, we shall pose the lane and pavement edge detection problem in a Bayesian framework. Let z^r be a realization of the radar random field Z^r and z^o be a realization of the optical random field Z^o. The optical and radar fusion detection problem can be solved by the joint MAP estimate

$$\begin{aligned} \hat{\underline{\theta}} &= \{\hat{\underline{\theta}}^r, \hat{\underline{\theta}}^o\} = \arg\max_{\underline{\theta}} p\left(\underline{\theta} | z^r, z^o\right) \\ &= \arg\max_{\{\underline{\theta}^r, \underline{\theta}^o\}} p\left(\underline{\theta}^r, \underline{\theta}^o | z^r, z^o\right) \end{aligned} \tag{6.23}$$

According to the Bayes' rule, we have

$$\hat{\underline{\theta}} = \arg\max_{\{\underline{\theta}^r, \underline{\theta}^o\}} \frac{p(z^r, z^o, \underline{\theta}^r, \underline{\theta}^o)}{p(z^r, z^o)} \tag{6.24}$$

Since the denominator $p(z^r, z^o)$ is fixed by the observation, the above formula can be rewritten as

$$\hat{\underline{\theta}} = \arg\max_{\{\underline{\theta}^r, \underline{\theta}^o\}} p(z^r, z^o, \underline{\theta}^r, \underline{\theta}^o) \tag{6.25}$$

By the chain rule of conditional probability, we have

$$p(z^r, z^o, \underline{\theta}^r, \underline{\theta}^o) = p(\underline{\theta}^r) p(z^r | \underline{\theta}^r) p(\underline{\theta}^o | \underline{\theta}^r, z^r) p(z^o | \underline{\theta}^o, z^r, \underline{\theta}^r) \tag{6.26}$$

Given the road shape parameter $\underline{\theta}^r$, the lane shape parameter $\underline{\theta}^o$ is independent of the radar observation z^r, i.e.,

$$p(\underline{\theta}^o | \underline{\theta}^r, z^r) = p(\underline{\theta}^o | \underline{\theta}^r), \tag{6.27}$$

and also given the lane shape parameter $\underline{\theta}^o$, the optical observation z^o is independent of the radar observation z^r and road shape parameter $\underline{\theta}^r$, i.e.,

$$p(z^o|\underline{\theta}^o, z^r, \underline{\theta}^r) = p(z^o|\underline{\theta}^o) \tag{6.28}$$

Substituting (6.27) and (6.28) into (6.26), and then substituting (6.26) into (6.25), we have

$$\begin{aligned} \hat{\underline{\theta}} &= \arg \max_{\{\underline{\theta}^r, \underline{\theta}^o\}} p(\underline{\theta}^r)\, p(z^r|\underline{\theta}^r)\, p(\underline{\theta}^o|\underline{\theta}^r)\, p(z^o|\underline{\theta}^o) \\ &= \arg \max_{\{\underline{\theta}^r, \underline{\theta}^o\}} p(\underline{\theta}^r, \underline{\theta}^o)\, p(z^r|\underline{\theta}^r)\, p(z^o|\underline{\theta}^o) \end{aligned} \tag{6.29}$$

Making use of the logarithm of the density functions, the joint MAP estimate turns to

$$\begin{aligned} \hat{\underline{\theta}} &= \arg \max_{\underline{\theta}}\ \log p(z^r, z^o, \underline{\theta}) \\ &= \arg \max_{\{\underline{\theta}^r, \underline{\theta}^o\}} \{\log p(\underline{\theta}^r, \underline{\theta}^o) + \log p(z^r|\underline{\theta}^r) + \log p(z^o|\underline{\theta}^o)\} \\ &= \arg \max_{\{\underline{\theta}^r, \underline{\theta}^o\}} \{\log p(\underline{\theta}^r, \underline{\theta}^o) + L^r(z^r, \underline{\theta}^r) + L^o(z^o, \underline{\theta}^o) + \gamma(\underline{\theta}^o)\} \end{aligned} \tag{6.30}$$

Calculating $\gamma(\underline{\theta}^o)$ is intractable, as it involves an integration over all the realizations of Z^o (see (6.16)). We will use a relative weighting to compensate for neglecting $\gamma(\underline{\theta}^o)$.

6.3.1 Empirical MAP

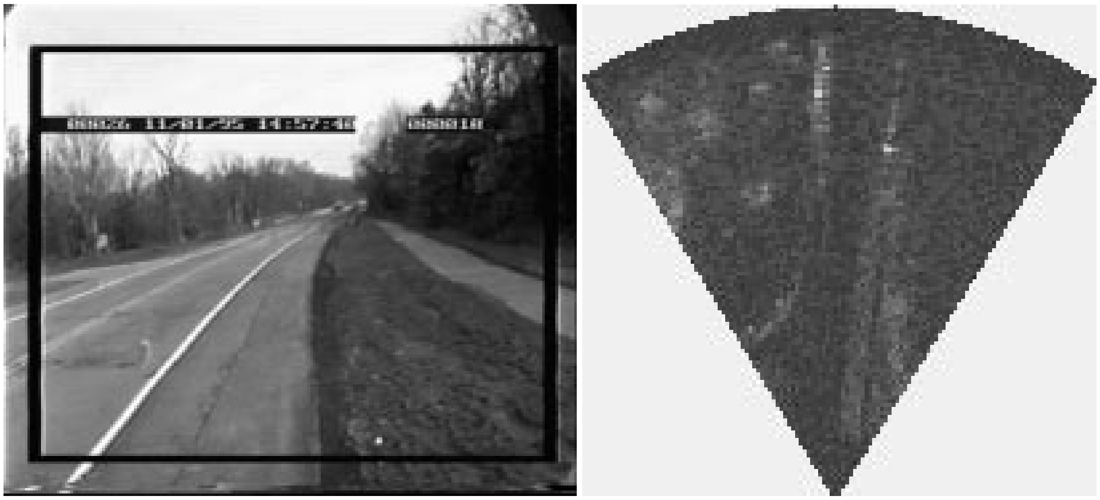
Since $\gamma(\underline{\theta}^o)$ in (6.30) is intractable, it is impossible to obtain a theoretical estimate for deformation parameters $\underline{\theta}$ based on (6.30). The primary difficulty we met in this edge detection problem is that for the two imaging likelihood functions, the radar imaging likelihood is normalized, while the optical imaging likelihood is not. Instead of computing the normalizing factor $\gamma(\underline{\theta}^o)$, we turn to the empirical MAP estimate,

$$\hat{\underline{\theta}} = \arg \max_{\{\underline{\theta}^r, \underline{\theta}^o\}} \{\log p(\underline{\theta}^r, \underline{\theta}^o) + L^r(z^r, \underline{\theta}^r) + \beta \cdot L^o(z^o, \underline{\theta}^o)\} \tag{6.31}$$

We expect the weighting factor β to play the same role as the normalizing constant $\gamma(\underline{\theta}^o)$.

The matching functions $L^r(z^r, \underline{\theta}^r)$ and $L^o(z^o, \underline{\theta}^o)$ are relative rather than absolute measure. This is the root cause of the problem — the matching functions are of different dynamic ranges within the parameter space of interest. Figure 6.3.4 shows the different dynamic ranges of the matching functions for an optical and radar image pair shown in Figure 6.3.3. The dynamic range of the radar matching function is 1.4×10^3, while the dynamic range of the optical matching function is 4.5×10^4. If the two matching functions are combined without weighting, i.e., let $\beta = 1$, then the instant result of the difference in dynamic ranges is that in most cases, the optical image dominates the joint estimate. Figure 6.3.5 gives an example of such dominance. Figure 6.3.5(a) shows the lane boundary detection result when only the optical observation data are used. The erroneously detected lane boundaries curve too much to the right because the optical data at the far range are not able to provide sufficient information for the curving of the road. The pavement boundary detection result shown in Figure 6.3.5(b) is quite correct. However, when we jointly detect the boundaries with no-weighting fusion algorithm, since the dynamic range of the optical matching function is much larger than the dynamic range of the radar matching function, the optical observation data dominate the joint estimate. Even though the the radar data yield correct curve direction by itself, it cannot correct the wrong curve direction with the fusion approach (see Figures 6.3.5(c) and (d)).

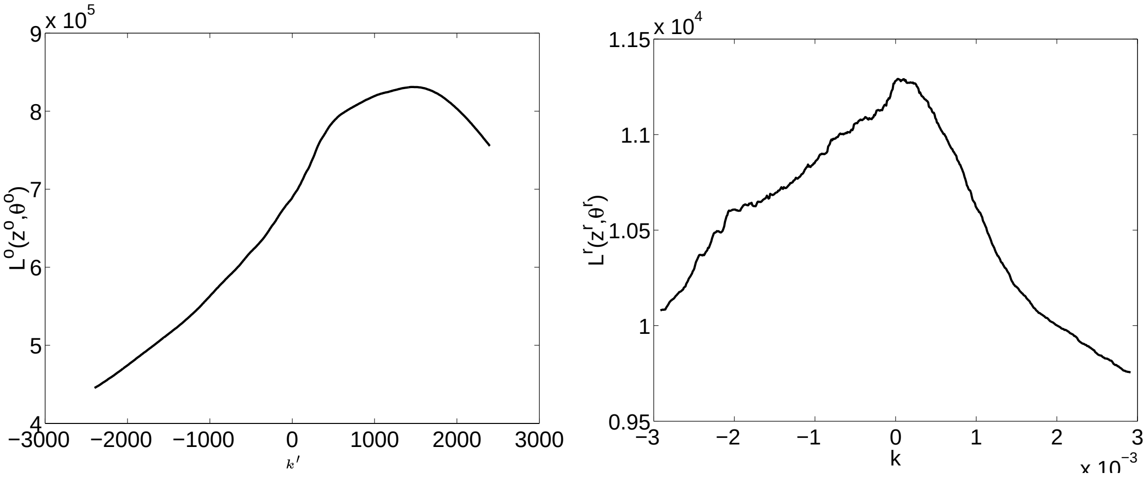
In order to overcome this inherent deficiency, we scale (using the weight β) one of the matching functions so that the ranges of variation for the weighted matching functions are approximately the same. This fixed weight β is empirically derived by examining the individual matching functions for a (training) set of optical and radar image pairs. Our extensive experiments seem to indicate that $\beta = 0.01$ gives good performance for most image pairs (Figure6.3.6).



(a) Optical image

(b) Radar image

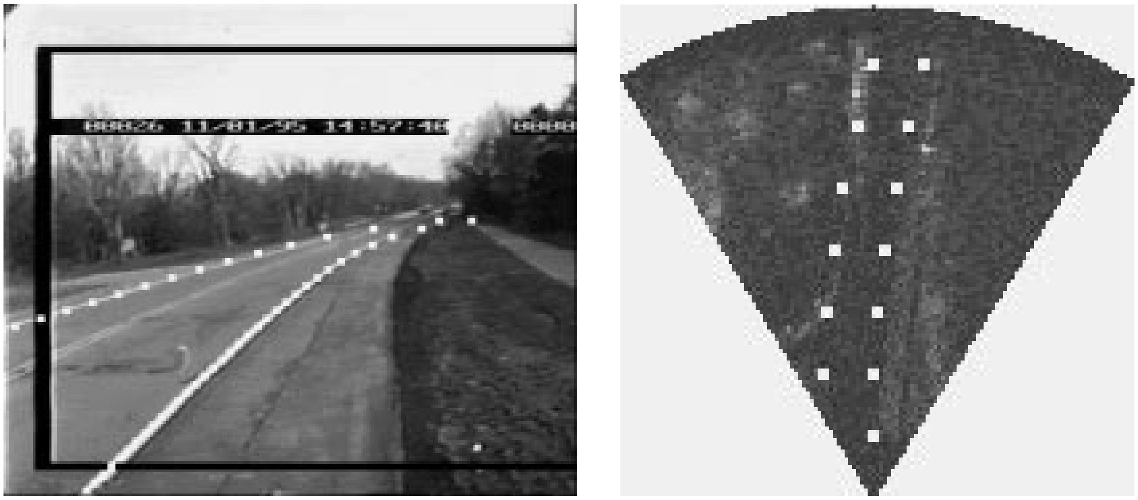
Figure 6.3.3: A pair of optical and radar images



(a) Optical matching function

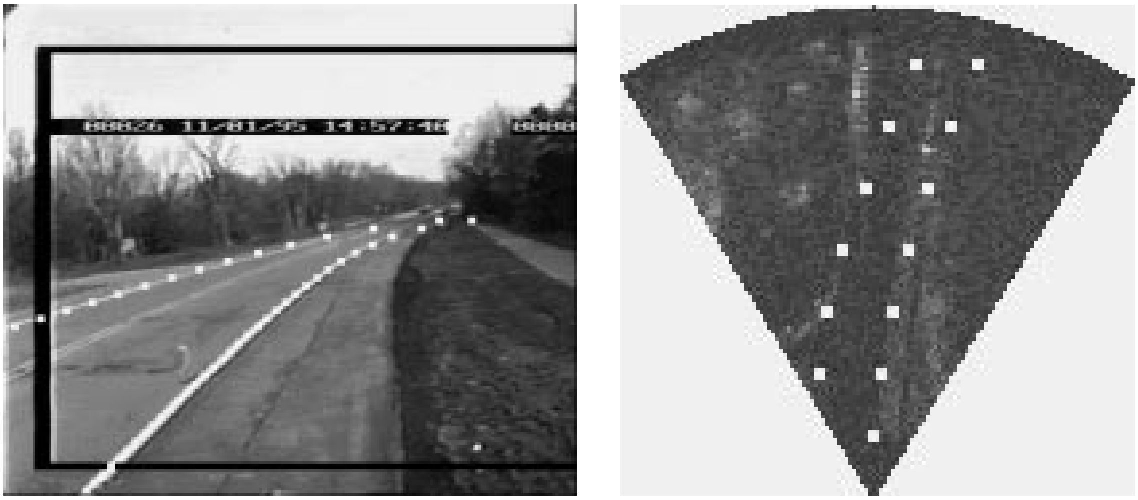
(b) Radar matching function

Figure 6.3.4: Different dynamic range of the matching functions for optical and radar images



(a) Single sensor based lane edge detection

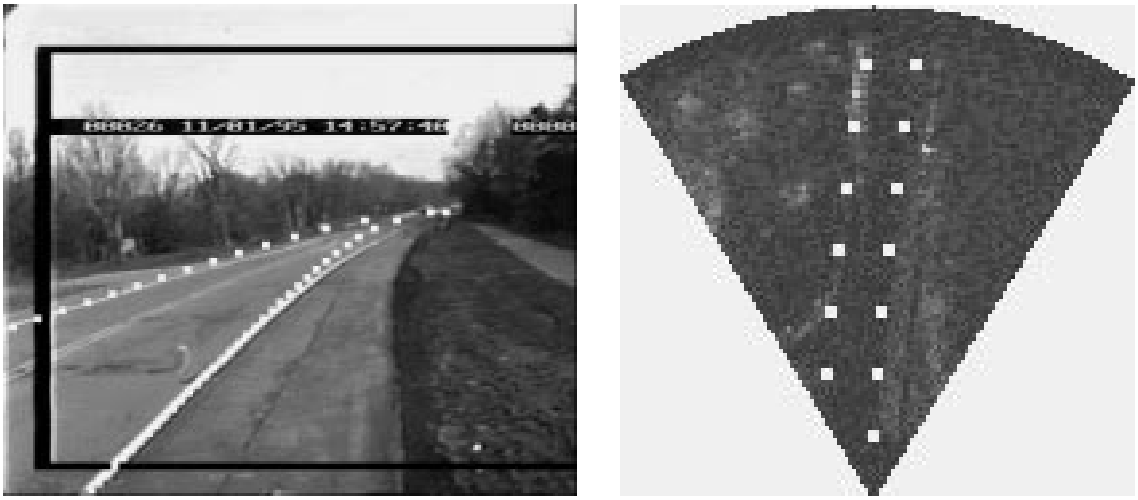
(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method while $\beta = 1$

(d) Pavement edge detection with fusion method while $\beta = 1$

Figure 6.3.5: Wrong edge detection results with fusion method and no-weight scheme



(a) Lane edge detection

(b) Pavement edge detection

Figure 6.3.6: Edge detection by fixed weight scheme

6.3.2 Computation of Empirical MAP Estimate

The problem is one of obtaining the maximum in (6.31). This problem is equivalent to finding the mode of a six-dimensional density surface. The surface is non-concave with many local maxima, hence we can not just apply the greedy search algorithms such as conjugate gradient methods. In our work, we suggest two techniques to find the global maximum.

6.3.2.1 Multi-resolution Pseudo-Exhaustive Search

Exhaustive search can find the optimal solutions at the cost of unacceptable computation resources in some optimization problems. For the problem we are considering, exhaustive search is not feasible due to the large searching space. Instead, a multi-resolution pseudo-exhaustive search method is studied, aiming for an acceleration of the matching process while maintaining the accuracy and robustness of the method. First, we constrain the parameters in appropriate ranges. Then we select a set of coarse step sizes (coarse grid) for the parameters and do the pseudo-exhaustive search to find the maximum of the joint MAP objective function (6.31). Once this coarse maximum is found, the corresponding estimated parameters are taken as the center of a finer search procedure with finer step sizes (finer grid) and smaller ranges of the parameters. Repeat the above step until the desired parameter grid size is reached.

6.3.2.2 Metropolis Algorithm with Geometric Annealing

Although the multi-resolution pseudo-exhaustive search gives us relatively accurate solutions, since we have six parameters to estimate, the search procedure is very time consuming. To accelerate the maximization procedure, we employ a sub-optimal approach, the Metropolis algorithm [56] with a geometric annealing schedule [97], to perform this maximization,

① Set $i = 0$, and initialize $\underline{\theta}^{(0)}$.

② Calculate $\log P(z^r, z^o, \underline{\theta}^{(i)})$.

③ Pick $\tilde{\underline{\theta}}$ at random among all the possible parameter values in the neighborhood of $\underline{\theta}^{(i)}$.

④ Calculate $\log P(z^r, z^o, \tilde{\underline{\theta}})$.

⑤ Calculate $\rho^{(i)} = \exp\left(\frac{\log P(z^r, z^o, \tilde{\underline{\theta}}) - \log P(z^r, z^o, \underline{\theta}^{(i)})}{T^{(i)}}\right)$, where $T^{(i)} = T_{init}\left(\frac{T_{final}}{T_{init}}\right)^{\frac{i+1}{max\_iter}}$.

⑥ Update the curve deformation parameters

$$\underline{\theta}^{(i+1)} = \begin{cases} \tilde{\underline{\theta}} & \text{if } \rho^{(i)} \geq 1 \\ \tilde{\underline{\theta}} & \text{w.p. } \rho^{(i)} \text{ if } \rho^{(i)} < 1 \\ \underline{\theta}^{(i)} & \text{otherwise} \end{cases}$$

⑦ Set i = i+1 and go to step 2.

6.3.3 Performance Comparison

We have applied the proposed multisensor fusion method to jointly detect the lane and pavement boundaries in registered radar and optical images. We have implemented the scheme described in previous sections on a data set containing 25 optical and radar image pairs. These image pairs were acquired under a variety of imaging (light, weather, etc.) conditions.

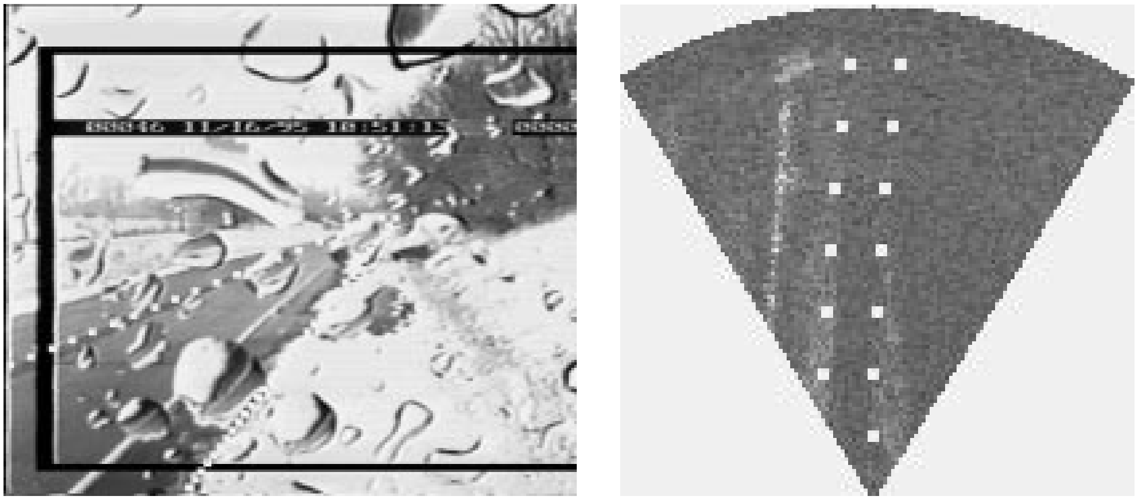
Since multiple (optical and radar) sensors provide more information and hence a more precise interpretation of the sensed environment, the performance of lane and pavement boundary detection is robust and accurate. To illustrate the advantage of the fusion detection algorithm over the single sensor based algorithm, we also implemented the MAP estimator described in [68] to detect the pavement boundaries in radar images alone and applied the LOIS algorithm presented in [56] to locate the lane boundaries using only optical images, and we plot the detection results with both single sensor based and fusion algorithms together in Figures 6.3.7, 6.3.8, and 6.3.9. In all the plots and for representative road scenes, the results obtained via independent optical and radar edge detection algorithms are illustrated in the upper rows, while the results using the fusion method are shown in the lower rows.

For the optical image shown in Figure 6.3.7(a), the snow line in the right side of the road dominates the lane boundary detection and thus the single optical image processing yields erroneous lane boundary detection result. However, the independent radar image processing succeeds in detecting the pavement boundaries (Figure 6.3.7(b)). In the fusion framework, since the prior constraints set on the shape parameters require that the lane boundaries be inside the pavement boundaries, the fusion algorithm corrects the estimation of the lane boundary offset parameters, and thus provides us the accurate lane and pavement boundary detection (Figures 6.3.7(c) and (d)). In this example, we have shown that the radar image improves the lane detection in the optical image.

For the radar image shown in Figure 6.3.8(b), the left side region to the road is also very homogeneous and serves as a spurious road leading to the wrong pavement detection using the radar information alone. We declare the pavement detection wrong by looking at the corresponding optical image — the road should curve to the right instead of curving to the left. However, the corresponding optical image provides enough information to yield correct lane detection (Figure 6.3.8(a)). The fusion algorithm, by combining both radar and optical sensory data, gives satisfactory lane and pavement boundary detection results (Figures 6.3.8(c) and (d)). This example has demonstrated that the information in the optical image improves the pavement edge detection in the radar image.

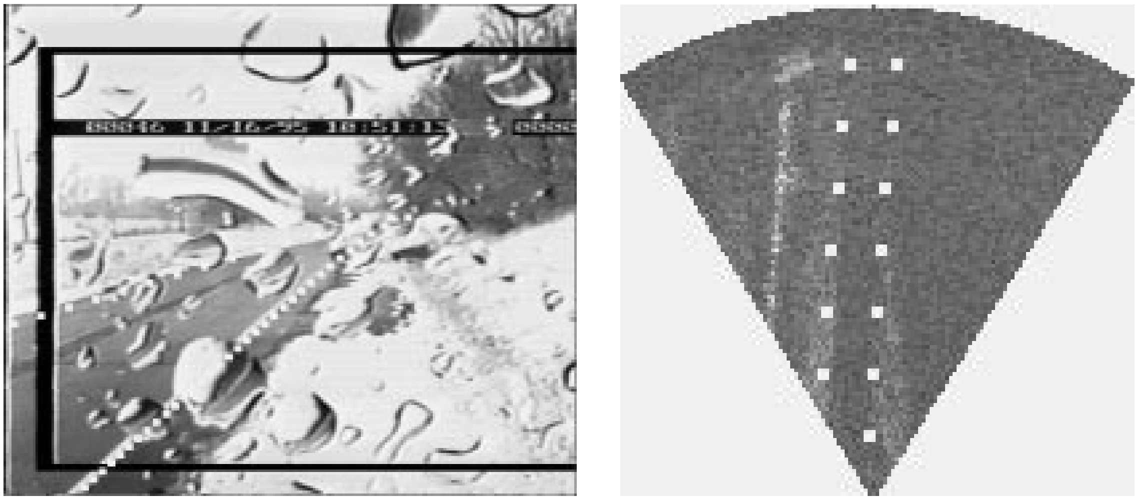
Both examples (Figures 6.3.7 and 6.3.8) indicate that the fusion method outperforms single sensor based method. We also noticed that fusion does not degrade the performance of the individual detection results when they are good by themselves (see Figure 6.3.9).

To compare the advantage of the fusion method over the single sensor detection methods and to appreciate the role of likelihood weighting, we undertake a large experiment. For the database of 25 optical and radar image pairs referred to earlier, we hand-picked ground truth and plot the detection errors compared to ground truth in Figures 6.3.10 and 6.3.11. Figure 6.3.10 shows the reduction in detection errors when the two data modalities are fused, compared to processing them individually. Figure 6.3.11 shows the reduction in detection errors when the likelihoods are relatively weighted using $\beta = 0.01$, compared to when no weighting is used.



(a) Single sensor based lane edge detection

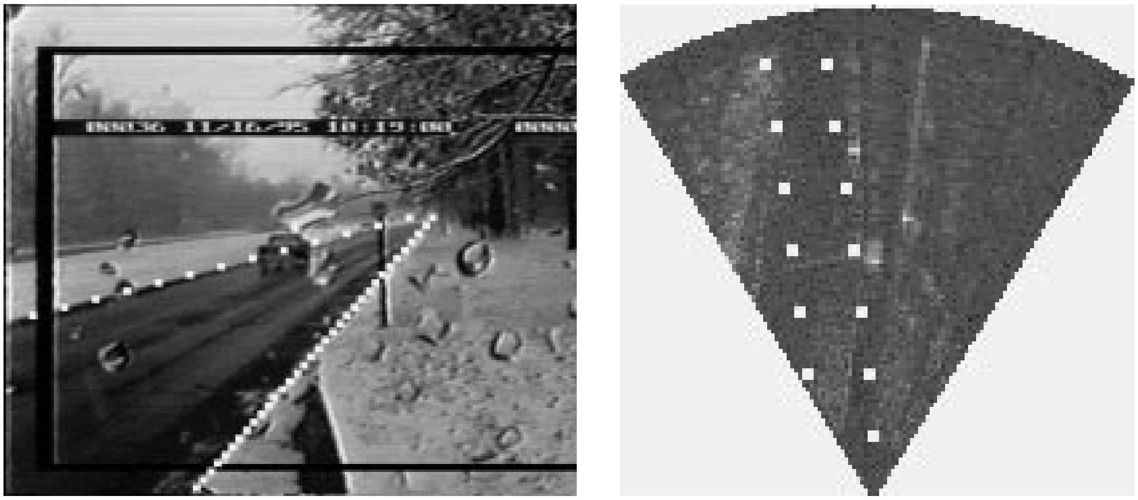
(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method

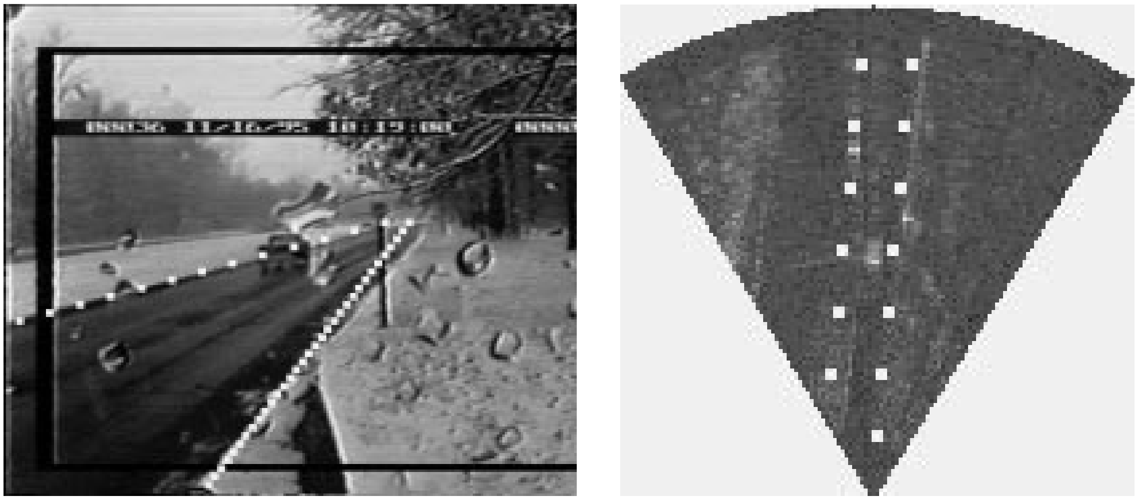
(d) Pavement edge detection with fusion method

Figure 6.3.7: Performance comparison of the fusion and single sensor based methods



(a) Single sensor based lane edge detection

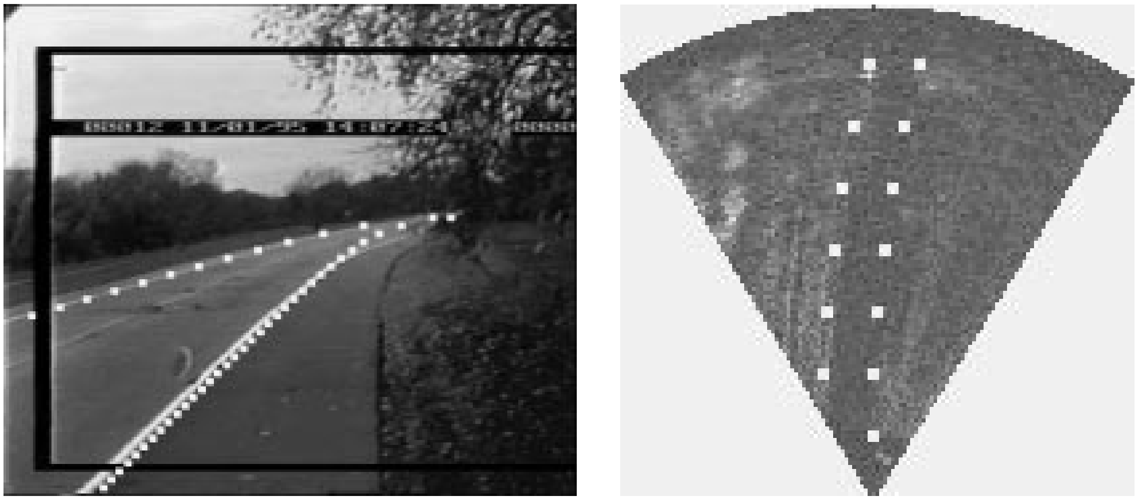
(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method

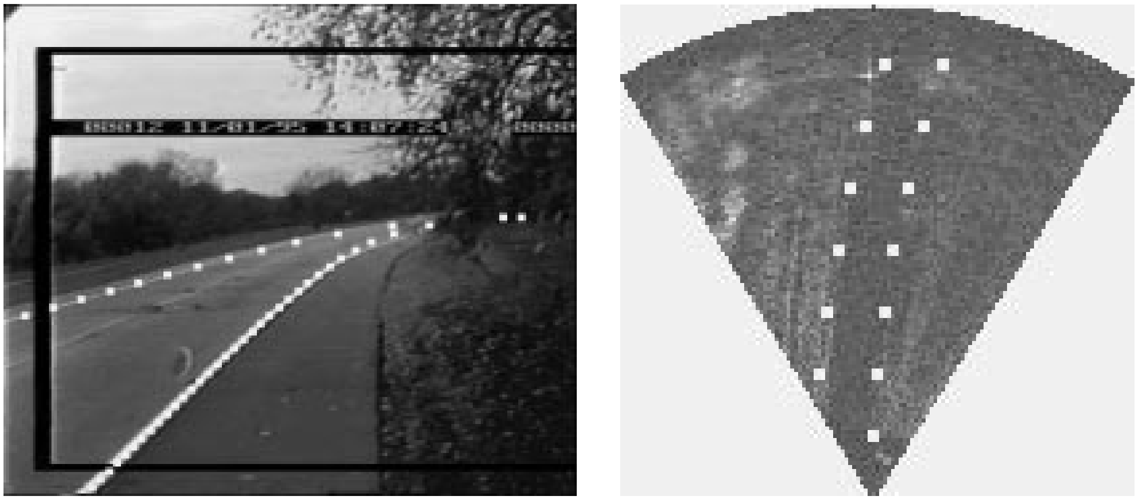
(d) Pavement edge detection with fusion method

Figure 6.3.8: Performance comparison of the fusion and single sensor based methods



(a) Single sensor based lane edge detection

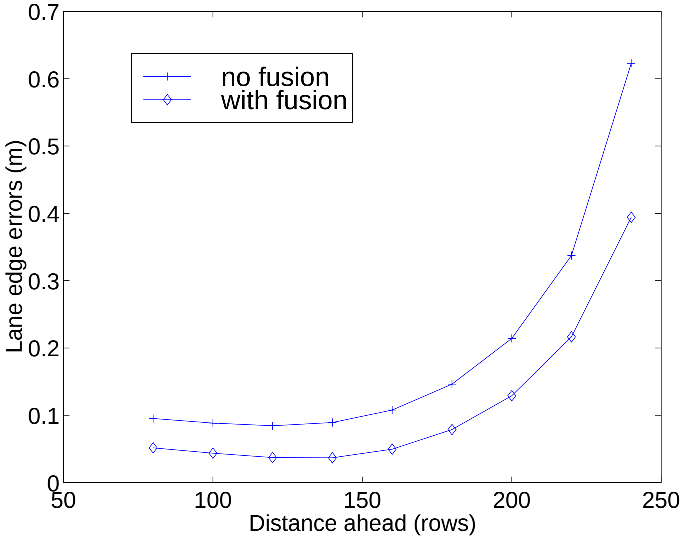
(b) Single sensor based pavement edge detection



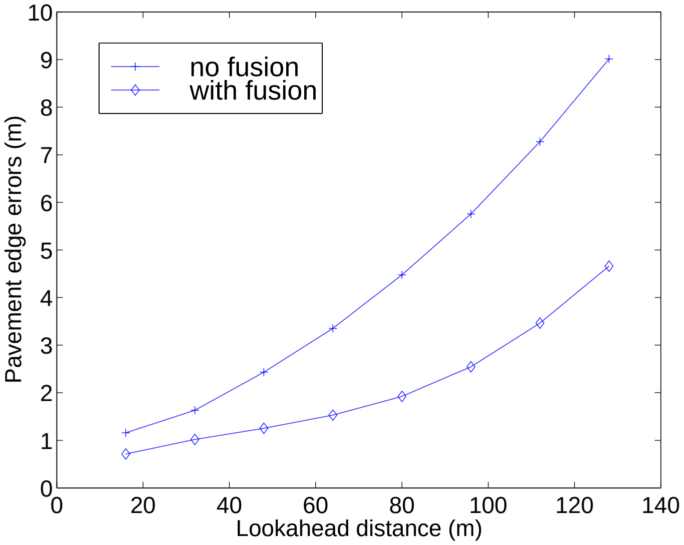
(c) Lane edge detection with fusion method

(d) Pavement edge detection with fusion method

Figure 6.3.9: Performance comparison of the fusion and single sensor based methods

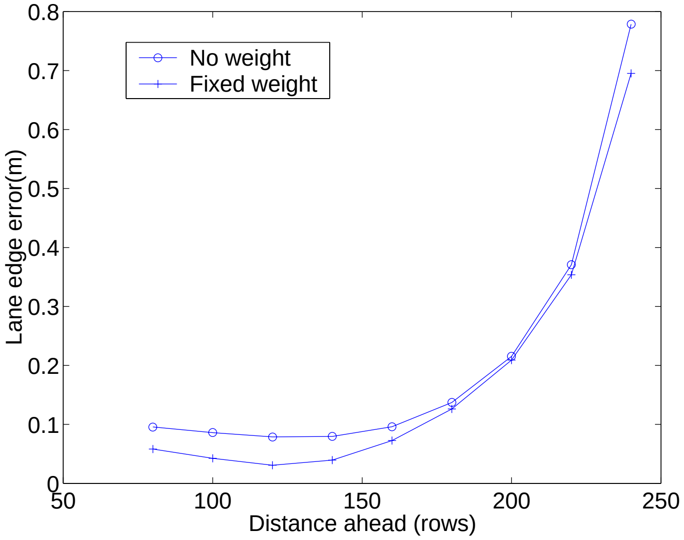


(a) Errors for lane edge detection

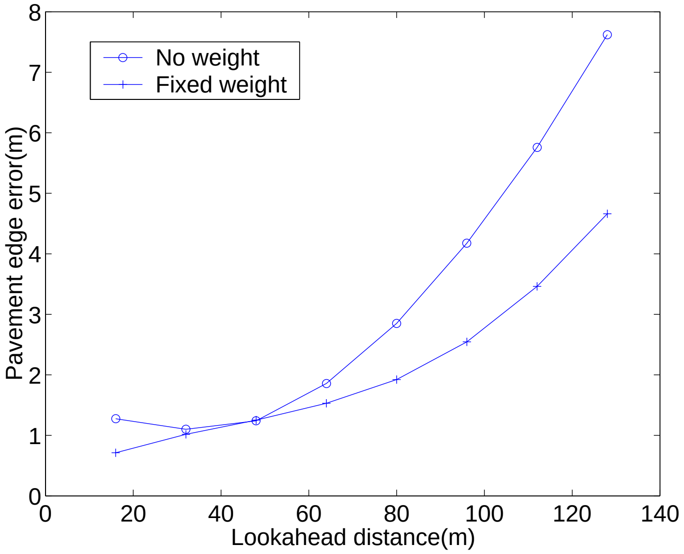


(b) Errors for pavement edge detection

Figure 6.3.10: Advantage of the fusion method over single sensor based method



(a) Errors for lane edge detection



(b) Errors for pavement edge detection

Figure 6.3.11: Advantage of the fixed weighting scheme over no weight scheme

6.4 Confidence Measure for Parameter Estimate

6.4.1 Objective Confidence Measure

Over the years, lane and pavement detection systems have enjoyed systematic improvements to their performance, including the techniques we just addressed in Chapters V and VI. There are currently studies underway to compare the performance of several systems side-by-side, and commercial availability of several such systems is imminent. While this is indeed a positive development, and a feather in the cap for the intelligent vehicles community, systematic assessment of the system's performance is sadly lacking. Compared to the volume of papers on various systems of lane and pavement detection and tracking the number of papers that provide frameworks for assessing their performance is miniscule.

In the preceding sections, we solve the simultaneous detection of lane and pavement boundaries with multisensor fusion approach. Central to the fusion methodology is an assessment of accuracy (i.e., confidence) of the lane and pavement boundary estimates. To be specific, let us consider the performance of single sensor based boundary detection approaches. For some pair of radar and optical images, let $\hat{\underline{\theta}}^r$ be the pavement boundary parameters estimated from the radar image only, and $\hat{\underline{\theta}}^o$ the lane boundary parameters estimated from the optical image only. $\hat{\underline{\theta}}^r$ and $\hat{\underline{\theta}}^o$ have different degrees of accuracy due to the different qualities of the original radar and optical images. In order to reflect the confidence difference between the lane and pavement boundary parameter estimates, different weights (the role β plays in (6.31)) should be imposed on the matching functions, $L^r(z^r, \underline{\theta}^r)$ and $L^o(z^r, \underline{\theta}^o)$, when they are combined in the joint MAP estimate.

Since $L^r(z^r, \underline{\theta}^r)$ and $L^o(z^r, \underline{\theta}^o)$ have different dynamic ranges when we vary the shape parameters $\underline{\theta}^r$ and $\underline{\theta}^o$, it is unfair to compare them directly. To remedy this dilemma, we introduce the normalized matching functions

$$L_n^r(z^r, \underline{\theta}^r) = \frac{L^r(z^r, \underline{\theta}^r) - \min_{\underline{\theta}^r} L^r(z^r, \underline{\theta}^r)}{\max_{\underline{\theta}^r} L^r(z^r, \underline{\theta}^r) - \min_{\underline{\theta}^r} L^r(z^r, \underline{\theta}^r)} \tag{6.32}$$

$$L_n^o(z^o, \underline{\theta}^o) = \frac{L^o(z^o, \underline{\theta}^o) - \min_{\underline{\theta}^o} L^o(z^o, \underline{\theta}^o)}{\max_{\underline{\theta}^o} L^o(z^o, \underline{\theta}^o) - \min_{\underline{\theta}^o} L^o(z^o, \underline{\theta}^o)} \tag{6.33}$$

For a given radar image, both the subtracting term in the numerator, $\min_{\underline{\theta}^r} L^r(z^r, \underline{\theta}^r)$, and the constant, and the denominator, $\max_{\underline{\theta}^r} L^r(z^r, \underline{\theta}^r) - \min_{\underline{\theta}^o} L^o(z^o, \underline{\theta}^o)$, are constants. So for certain parameters $\underline{\theta}^r$, the normalized matching function $L_n^r(z^r, \underline{\theta}^r)$ is a monotone function of the original matching function $L^r(z^r, \underline{\theta}^r)$, i.e., $L_n^r(z^r, \underline{\theta}^r)$ provides similar fidelity of the parameters to the image observation as $L^r(z^r, \underline{\theta}^r)$ does, only different in scales. Same with the optical case. In addition, the normalized matching functions have identical dynamic range, which is $[0, 1]$. For these two comparable normalized matching functions, more weight need to be imposed on the function whose parameter estimates are more reliable.

In this section, we will address an objective accuracy assessing technique for the parameter estimates of the lane and pavement boundaries. The idea is to assess the sensitivity of the normalized matching functions with respect to their shape parameters. To evaluate the sensitivity of a normalized matching function with respect to a parameter, a common used methodology would be using the curvature (second derivative) of the normalized matching function with respect to that parameter. The larger the curvature, the more confidence in that parameter estimate. Then we define the confidence metric for the normalized matching function as the average curvature among all the shape parameters.

Unfortunately, we do not have the exact likelihood function for the optical imaging process, instead, what we have is a empirical matching function. For the radar imaging process, we do have the exact likelihood function. However, since both radar imaging likelihood function and optical matching function are not explicit in the shape parameters, it is impossible to analytically calculate the derivatives. Therefore, we will be satisfied if we can approximate the curvatures.

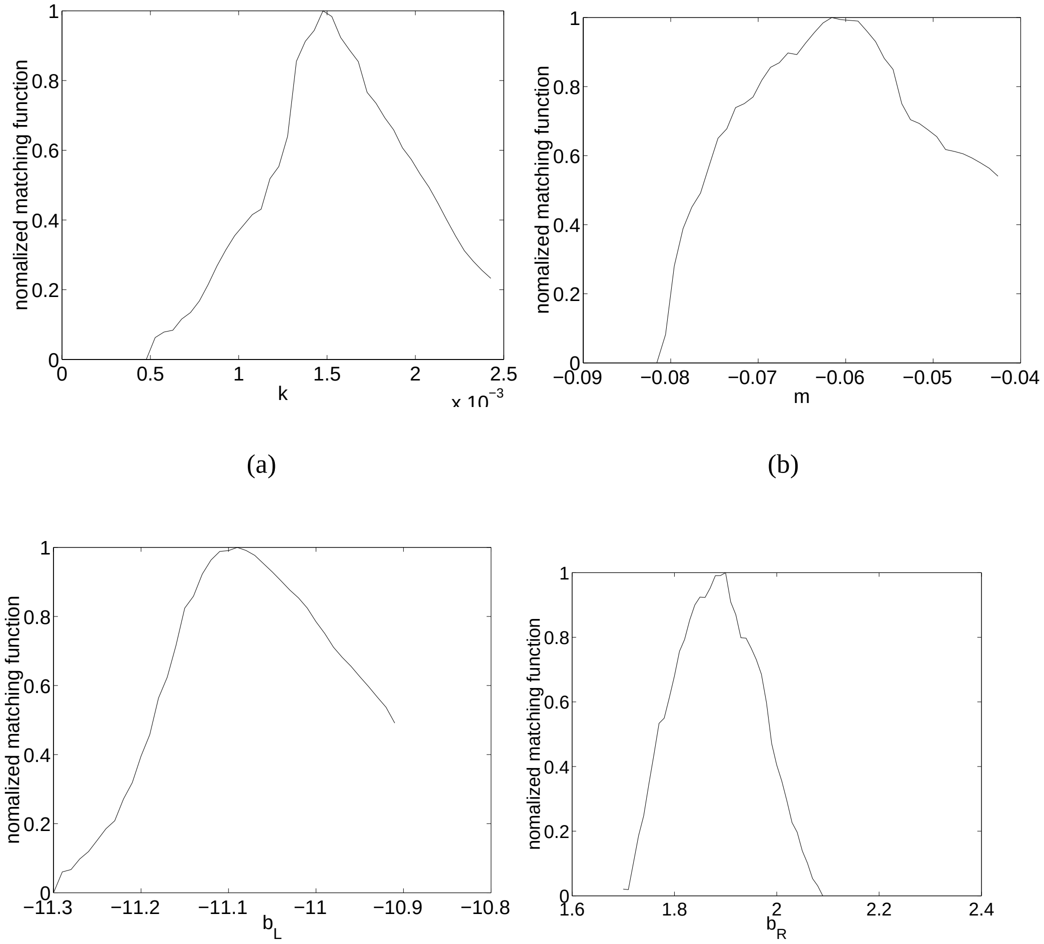
6.4.2 Curvature Estimation

Let us utilize an example to illustrate the variation of the normalized matching function values with respect to each shape parameter. We applied the MAP estimate algorithm proposed in [68] to the radar image shown in Figure 6.3.9 (a) and obtained the shape parameter estimates $\underline{\hat{\theta}} = \{\hat{k}, \hat{m}, \hat{b}_L, \hat{b}_R\}$. In Figure 6.4.12, we plot the cross sections of the normalized radar matching function surface ($L_n^r(z^r, \underline{\theta}^r)$) along one of the shape parameters while setting all other parameters as their best estimates. For example, Figure 6.4.12(a) shows the normalized matching function values as a function of the parameters k when setting the other parameters $m = \hat{m}, b_L = \hat{b}_L, b_R = \hat{b}_R$.

Let $L(x)$ denote the normalized matching values as a function of the parameter x while setting all other parameters to their optimal estimates. Let $\hat{x}$ denote the optimal estimate of x. Suppose $\triangle$ is the step size of x when we calculate the normalized matching function values L. Assume that there are $2M + 1$ normalized matching function values, $L_{\hat{x}} = \{L(\hat{x} + i\triangle), i = -M, \ldots, -2, -1, 0, 1, 2, \ldots, M\}$, available for the purpose of calculating the curvature. Then assessing the reliability of $\hat{x}$ is to estimate the second derivative of the normalized matching function $L(x)$ at the point $\hat{x}$ given the discrete data set $L_{\hat{x}}$.

One frequently used numerical approach to estimating the derivatives is using the difference between adjacent function values to approximate the derivatives. Let $L'(x)$ and $L''(x)$ denote the first and second derivatives, respectively.

$$\begin{aligned} L'(\hat{x}) &= \frac{L(\hat{x} + \triangle) - L(\hat{x})}{\triangle} \\ L'(\hat{x} + \triangle) &= \frac{L(\hat{x} + 2\triangle) - L(\hat{x} + \triangle)}{\triangle} \end{aligned}$$



(a)

(b)

(c)

(d)

Figure 6.4.12: Cross sections of the radar normalized matching function with respect to their pavement shape parameters

Then

$$\begin{aligned} L''(\hat{x}) &= \frac{L'(\hat{x}+\triangle) - L'(\hat{x})}{\triangle} \\ &= \frac{L(\hat{x}+2\triangle) - 2L(\hat{x}+\triangle) + L(\hat{x})}{\triangle^2} \end{aligned} \tag{6.34}$$

This estimate is quite straightforward and easy to implement, however, since it just uses three function values to make the estimation, the performance of this estimator is very sensitive to noise and the choice of the step size $\triangle$.

We propose using a quadratic curve to approximate the normalized matching function around the parameter estimate and then we use the curvature of the quadratic curve as the curvature of the likelihood function. Let $f_{a_1,a_2}(x) = a_1(x-\hat{x})^2 + a_2$ represent a quadratic curve. And for certain choices a_1 and a_2, we calculate the data sequence $f_{a_1,a_2,\hat{x}} = \{f_{a_1,a_2}(\hat{x}+i\triangle), i = -M, \ldots, -2, -1, 0, 1, 2, \ldots, M\}$. In order to approximate L with f_{a_1,a_2}, we need to minimize the total squared error between sequences $L_{\hat{x}}$ and $f_{a_1,a_2,\hat{x}}$ with respect to a_1 and a_2, while keeping $f_{a_1,a_2}(x)$ less than $L(x)$ for any x,

$$\begin{aligned} &\min_{a_1,a_2} \sum_{i=-M}^{M} \left[f_{a_1,a_2}(\hat{x}+i\triangle) - L(\hat{x}+i\triangle)\right]^2 \\ &s.t. \;\; f_{a_1,a_2}(x) \le L(x), \\ &\qquad \forall x \in \{\hat{x}+i\triangle, i = -M, \ldots, -2, -1, 0, 1, 2, \ldots, M\} \end{aligned} \tag{6.35}$$

The constraints are meant to guarantee $a_1 < 0$.

Solving this constrained optimization problem we get the optimal quadratic curve approximation with parameters $\hat{a}_1$ and $\hat{a}_2$. Then $-2\hat{a}_2$ is taken as the estimated curvature. An example of this curve approximation is shown in Figure 6.4.13 where the solid curve is the true normalized matching function while the dashed curve is the quadratic curve approximation.

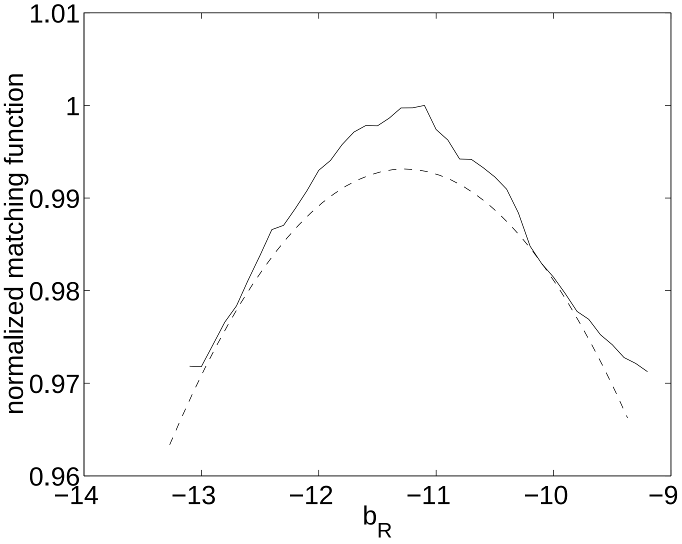


Figure 6.4.13: A cross section of the normalized matching function and its quadratic approximation

6.4.3 Ill-conditioned Parameter Estimation Problem

We have applied the proposed method to calculate the curvature of the normalized matching function with respect to each shape parameter. In Tables 6.1 and 6.2, we give the curvature results for some radar and optical images, respectively. With respective to a shape parameter, the curvature estimates are in the same order of magnitude for different images acquired under similar conditions, however, for different shape parameters, even in the same image, the curvature estimates vary dramatically. For example, in the radar image “anhinesb”, the curvature estimates with respect to k and b_L are 1.63×10^7 and 1.11, respectively. The curvatures’ dramatic difference in the order of magnitudes makes it very difficult to assess the confidence metric for the normalized matching function because the largest curvature completely dominates the average of the curvatures. Furthermore, it also implies that the surface of the objective function varies dramatically in one parameter while negligible in another parameter, which means that this estimation problem is very ill-conditioned.

Image name	Curv($\hat{k}$)	Curv($\hat{m}$)	Curv($\hat{b}_L$)	Curv($\hat{b}_R$)
anhinesb	1.63×10^7	2555.71	1.11	3.99
anhinesc	1.11×10^7	2358.94	5.48	15.32
anhinesg	1.99×10^7	2391.72	1.76	9.10
evhines	1.30×10^7	2418.19	3.34	3.06
hinesbev	1.48×10^7	1950.46	8.82	11.61

Table 6.1: Curvature results for some radar images

Image name	Curv($\hat{k}'$)	Curv($\hat{vp}'$)	Curv($\hat{b}'_L$)	Curv($\hat{b}'_R$)
anhinesb	1.07×10^{-5}	0.02	39.17	642.00
anhinesc	0.71×10^{-5}	0.02	30.78	455.71
anhinesg	0.58×10^{-5}	0.01	36.00	497.98
evhines	1.02×10^{-5}	0.01	9.78	398.15
hinesbev	0.86×10^{-5}	0.02	38.82	511.61

Table 6.2: Curvature results for some optical images

This ill-conditionedness is due to the inherent shortcomings of this parameterization scheme. Let us take the model parameters, $\underline{\theta}^r = \{k, m, b_L, b_R\}$, for the pavement boundaries as an example to illustrate why.

- **Model parameters are of different units —** the curvature parameter k is in units of inverse length, the orientation parameter m is dimensionless, and the offset parameters b_L and b_R are in units of length.

- **Model parameters are of different order of magnitudes —** the curvature parameter k is in the order of 10^{-3}, the orientation parameter m is in the order of 10^{-2}, and the offset parameters b_L and b_R are in the order of 10^1.

- **Variations in model parameters result non-uniform shape change —** a substantial change in the curvature parameter results in the pavement boundary shape changing only at distances far away from the host vehicle, while a small change in the offset parameter leads to a completely different scenario.

The lane boundary parameters, $\underline{\theta}^o = \{k', vp', b'_L, b'_R\}$, have similar curvature performance as the pavement boundary parameters, $\underline{\theta}^r$, and $\underline{\theta}^o$ have similar pitfalls: first of all, the model parameters are of different units; and secondly, the model parameters are of different order of magnitudes; and lastly, variations in the model parameters result non-uniform shape change.

However, we do notice that the curvature results for lane (pavement) boundary offset parameters b'_L and b'_R (b_L and b_R) are actually comparable (See Tables 6.2) because b'_L and b'_R have the same physical meaning — offsets, and they are of the same units and of the same order of magnitudes. The compatibility between the parameters enables us to utilize curvatures to evaluate the reliability of these parameter estimates. In most optical images (Figures 6.3.7(a), 6.3.8(a), and 6.3.9(a)), right lane boundaries are much clearer than their left counterparts. I.e., image observations provide more information on right boundaries than on the left boundaries. Therefore, the confidence of the matching function, L_n^o, in $\hat{b}'_R$, the right lane boundary offset, should be better than the confidence in $\hat{b}'_L$, the left lane boundary offset. In Table 6.2 we observe that the curvatures of right lane boundary offsets, Curv$(\hat{b}'_R)$, are usually larger than those of right lane boundary offsets, Curv$(\hat{b}'_L)$.

The above observation has demonstrated that the proposed confidence measure indeed assesses the reliability of the parameter's estimate. Then in order to get rid of the inherent ill conditionedness in parameter estimation for lane and pavement boundary detection problem, we have to find other shape models whose parameters are compatible. In the next chapter, we will propose a concentric circular shape model that can overcome the inherent shortcomings of the parabolic model.

CHAPTER VII

Fusion Algorithm for Lane and Pavement boundary Detection with Concentric Circular Models

Automated detection of lane and pavement boundaries has been broadly studied by researchers in intelligent transportation. Many state of art systems for detecting and tracking pavement boundaries use *a priori* shape models to mathematically describe the appearance of these boundaries. The use of prior shape models allows these systems to reject false boundaries (such as entry/exit ramps) and also overcome image clutter (shadows) and occlusion.

Several types of shape models have been employed, and the choice of model is usually made from the standpoint of accommodating all possible variations (in width, orientation, curvature, tilt, etc.) of the lane and pavement boundaries relative to the host vehicle. Polynomial (quadratic or cubic) shape models are the ones of choice. Polynomial shape models have improved the performance of lane and pavement boundary detection and tracking systems with higher accuracy and better reliability [22, 52, 69, 71] than the conventional gradient based methods. For example, we utilized quadratic shape models in our joint lane and pavement boundary detection application and we illustrated promising boundary detection results in Chapter VI. Unfortunately, as we have shown in the last chapter for a special case, parabolic shape models, polynomial shape models have inherent pitfalls for their parameters' estimation. Due to the parameters' different units and different dynamic ranges, the relating estimation problem is usually very ill conditioned.

It is common knowledge that in the United States public highway system lane and pavement boundaries are laid as concentric circles, at least over small segments. In this chapter we propose to directly use a circular model to describe the boundaries. Indeed the polynomial shape models are intended as an approximation to the circular model, but to our knowledge the circular model itself has never been used before.

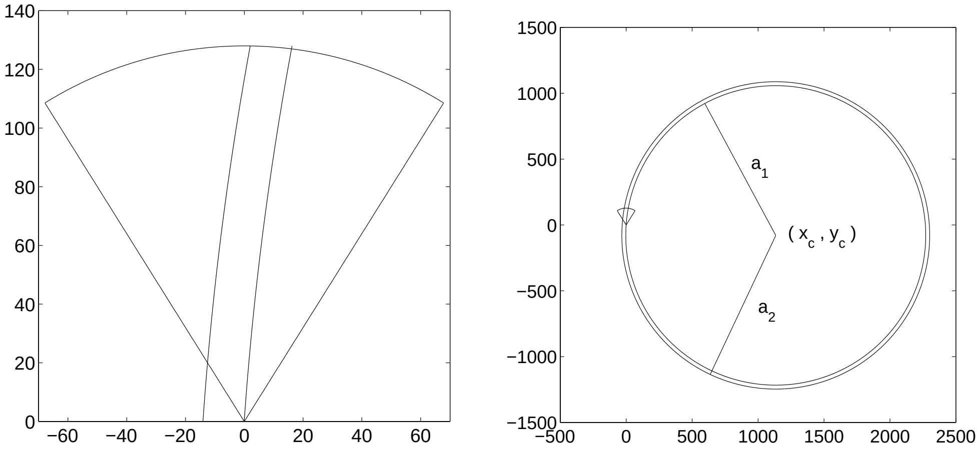
7.1 Concentric Circular Models

The field-of-view of a radar image on the ground plane is a cone shaped region with the apex at the position where the host vehicle stands. For an optical image, its field-of-view on the ground plane is also cone shaped, but perhaps with a different radius and azimuth from that of a radar image. Therefore, in our application domain, a typical road (or lane) scenario can be modeled by an intersection of a cone with two concentric circles (see Figures 7.1.1 and 7.1.2 for an illustration of road scenes). The cone represents the field-of-view of the radar (or optical) sensor, and the two circles represent the left and right pavement (or lane) boundaries.

Let us consider the representation of the pavement boundaries first. Assuming that the apex of the cone is at the origin $(0, 0)$, we represent the coordinates (x, y) of the pavement boundaries by circles centered at (x_c, y_c) with radii a_1 and a_2, respectively

$$(x - x_c)^2 + (y - y_c)^2 = a_{1,2}^2 \ , \tag{7.1}$$

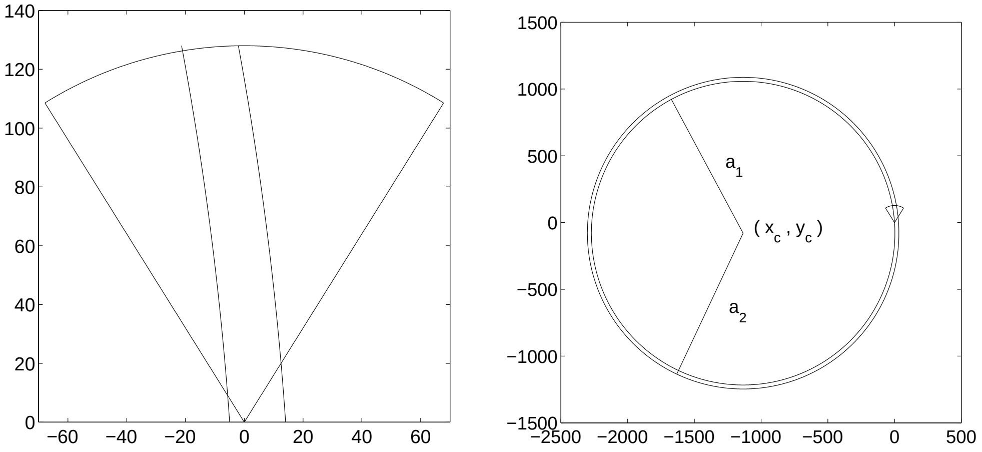
where $a_1 < a_2$. That is, a_1 is always the radius of the smaller circle on which either the left or right boundary is laid. So in this parameterization, the two boundaries are not distinguished as left and right. However, given the parameters $x_c, y_c, a_1,$ and a_2, it is easy to tell which radius corresponds to the left boundary — if $x_c > 0$, then the larger



(a) Field-of-view of the radar sensor

(b) Complete plot of the circular shape model

Figure 7.1.1: A typical road scenario in a radar image



(a) Field-of-view of the radar sensor

(b) Complete plot of the circular shape model

Figure 7.1.2: A typical road scenario in a radar image

radius, a_2, corresponds to the left boundary (Figure 7.1.1); otherwise, the smaller radius, a_1, corresponds to the left boundary (Figure 7.1.2). The reason for this parameterization is the ease of description of the constraints on the parameters. (This issue will be addressed in later this section.) The parameters $\underline{\theta}_c^r = \{x_c, y_c, a_1, a_2\}$ are capable of representing the pavement boundaries.

Figure 7.1.1(b) shows the complete plot of the cone and concentric circles, and Figure 7.1.1(a) gives the segment of the model shown in Figure 7.1.1(b) that corresponds to the field-of-view of the radar sensor. Similar with Figure 7.1.2.

Similarly, the lane boundaries on the ground plane can be represented as

$$(x - x_c)^2 + (y - y_c)^2 = (a'_{1,2})^2 \tag{7.2}$$

where a'_1 and a'_2 are the radii of the circles that corresponds to the lane boundaries. Then $\underline{\theta}_c^o = \{x_c, y_c, a'_1, a'_2\}$ are the shape parameters for lane boundaries on the ground plane. Note that the lane and pavement boundaries share the same parameters (x_c, y_c) since they are concentric circles.

The domain of the radar image is actually the ground plane and (7.1) can be directly applied to model the shape of pavement boundaries in the radar image.

The domain of the optical image, however, is a perspective projection of the road scenario on the ground plane, and therefore we need to transform the optical image data and the lane boundary model onto the same image plane.

There are two possible approaches to accomplish this goal. One is similar to that technique for the parabolic shape model — apply the perspective projection operator to the circular shape model (7.2) on the ground plane to obtain a lane boundary shape model on the image plane. An alternative method is to use (7.2) to describe the lane boundaries on the ground plane and project the optical image data onto the ground plane with the inverse perspective projection. For each pixel with row and column (r, c) in the optical image plane, its corresponding Cartesian coordinates (x, y) on the ground plane can be calculated by

$$\begin{aligned} y &= \frac{h\,[1 - r_f^2(r - c_r)(c_r - h_z)]}{r_f(r - h_z)} \\ x &= c_f(c - c_c)\sqrt{\frac{y^2 + h^2}{1 + r_f^2(r - c_r)^2}} \end{aligned} \tag{7.3}$$

where h_z is the horizon row, h is the height of the focal point above the ground plane, r_f is the height of a pixel on the image plane divided by the focal length, c_f is the width of a pixel on the image plane divided by the focal length, and c_c and c_r are one half of the numbers of columns and rows, respectively.

Note that in the second approach mentioned above, the parameters of the lane boundaries, $\{x_c, y_c, a_1', a_2'\}$, have the same physical units (namely, units of length) and properties as the parameters of the pavement boundaries, $\{x_c, y_c, a_1, a_2\}$. This is a desirable characteristic to ensure unit compatibility in the fusion algorithm. Therefore, we will take the second approach, i.e., we project the optical image onto the ground plane and employ (7.2) as the lane boundary representation. For the purpose of conciseness, during further algorithm derivation for lane boundary detection, the optical image is referred to the projected optical image on the ground plane.

We have shape parameters $\underline{\theta}_c^r = \{x_c, y_c, a_1, a_2\}$ for the pavement boundaries and $\underline{\theta}_c^o = \{x_c, y_c, a_1', a_2'\}$ for the lane boundaries. Let $\underline{\theta}_c = \{\underline{\theta}_c^r, \underline{\theta}_c^o\} = \{x_c, y_c, a_1, a_2, a_1', a_2'\}$ denote the complete parameter set for the circular shape models. Also for further study and comparison, let $\underline{\theta}_q^r = \{k, m, b_1, b_2\}$ and $\underline{\theta}_q^o = \{k', vp', b_1', b_2'\}$ denote the parameters for the parabolic models addressed in Chapter VI. Similarly, define $\underline{\theta}_q = \{\underline{\theta}_q^r, \underline{\theta}_q^o\} = \{k, m, b_1, b_2, b_1', b_2'\}$. Note that we exclude the parameters k' and vp' in this parameter set because they can be calculated from the parameters k and m (6.3).

There are a number of advantages of the new parameterization over the previous one. First of all, the new model better reflects the road shape of the real world. Circular shape models exactly describe the road boundaries, while parabolic shape models only approximate the road boundaries. For the circular shape model, the six parameters x_c, y_c, a_1, a_2, a'_1 and a'_2 have the same units – that of length. In addition, the ranges of feasible values for them have the same order of magnitude. On the contrary, as we mentioned at the end of the previous chapter, the parameters k, m, b_L, b_R, b'_L, and b'_R have different units and different dynamic ranges. We also observe that for both shape models, the number of parameters are the same, or in other words neither model is more complex to describe than the other.

What sets the models apart are the constraints on $\underline{\theta}_c$ and $\underline{\theta}_q$ which are imposed to satisfy in order to result in a "feasible" pair of left and right lane and pavement boundaries. In the quadratic case, the feasibility region is a hypercube with respect to the model parameters $\{k, m, b_L, b_R, b'_L, b'_R\}$. In the circular model, the feasibility region is not so simple. The feasibility region has the following restrictions:

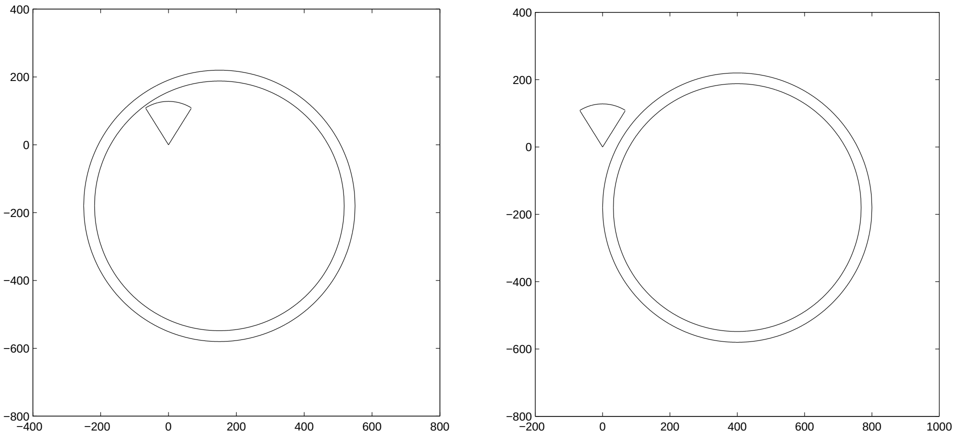
① The two circles corresponding to the pavement boundaries must intersect the cone. That is, the cone cannot be totally inside the inner circle, neither can it be outside the outer circle (see Figure 7.1.3[1]). The constraint on the model parameters is

$$a_1 - p_1 \leq x_c^2 + y_c^2 \leq a_2 + p_2 \tag{7.4}$$

where p_1 and p_2 are two appropriately selected positive numbers that will allow the corresponding pavement boundaries to both be offset to either the left or the right of the origin for extreme cases. Of course, the left boundary can be offset to the left of the origin, and the right boundary to its right.

---

[1] The plots in Figures 7.1.3 and 7.1.4 only demonstrate the relationship between the circles and the cone. In order to make the plots easily readable, we enlarged the size of the cone and the width of the road, while decreased the size of the circles.



(a)

(b)

Figure 7.1.3: Cases where circles do not intersect the cone

② The image data are acquired under the assumption that the host vehicle is still within the road (or at least the shoulder). As such, the cases shown in Figure 7.1.4 are not realistic as they correspond to scenarios when the host vehicle is entirely off the road. The corresponding constraint on the model parameters is

$$\alpha_{min} \leq \text{atan}\left(\frac{y_c}{x_c}\right) \leq \alpha_{\max} \tag{7.5}$$

③ The lane should be positioned inside the road region, i.e., the two circles on which the lane boundaries are laid are between the two circles on which the pavement boundaries are laid.

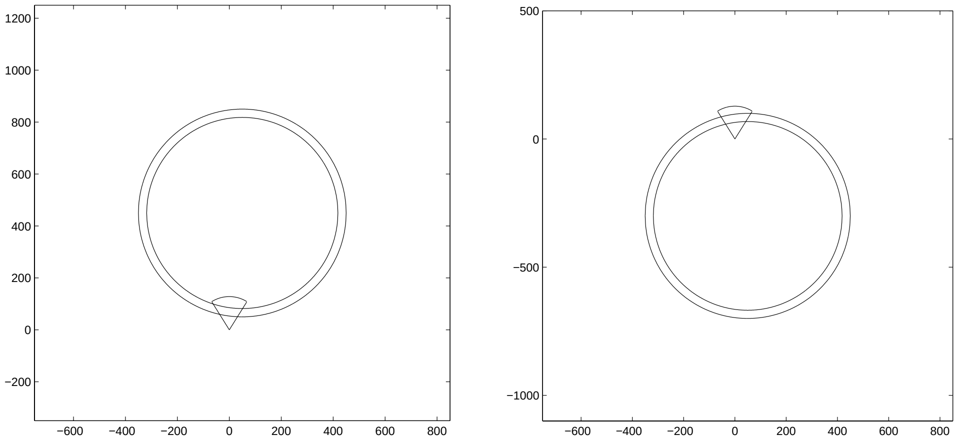
$$a_1 < a'_1 < a'_2 < a_2, \tag{7.6}$$

④ Finally, the road width has to be within minimum and maximum limits,

$$W_{min} \leq a_2 - a_1 \leq W_{max} \tag{7.7}$$

Similarly, there are upper and lower limits for the lane width,

$$W'_{min} \leq a'_2 - a'_1 \leq W'_{max} \tag{7.8}$$



(a) (b)

Figure 7.1.4: Cases where circles intersect the cone in an infeasible way

The shape model (7.1) and the constraints (7.4), (7.5), (7.6) (7.7), and (7.8) constitute the prior knowledge regarding pavement boundaries. This results in a so-called prior pdf:

$$\begin{aligned} P(\underline{\theta}_c) \quad &= \quad \frac{1}{\gamma_c} \cdot I_{[a_1-p_1, a_2+p_2]}(x_c^2+y_c^2) \;\cdot\; I_{[\alpha_{min},\alpha_{max}]}\left(\text{atan}\left(\frac{y_c}{x_c}\right)\right) \\ &\qquad \cdot I_{[W_{min},W_{max}]}(a_2-a_1) \;\cdot\; I_{[W'_{min},W'_{max}]}(a'_2-a'_1) \\ &\qquad \cdot I_{[a_1,a'_2]}(a'_1) \;\cdot\; I_{[a'_1,a_2]}(a'_2) \end{aligned} \qquad (7.9)$$

where γ_c is a normalizing constant and $I_A(x)$ is an indicator function,

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

This prior pdf is empirically derived, and it is simply a uniform distribution over the space of feasible model parameters, and zero elsewhere.

7.2 Decoupling of The Parameterization

It is obvious that the road width has both upper and lower bounds (see constraint (7.5)) and the lane should be located between the pavement boundaries(see constraint (7.6)).

Therefore, the four radius parameters, a_1, a_2, a_1' and a_2', are highly correlated. The apex of the cone, where the host vehicle is located, is inside the road, at least on the shoulder. Therefore, x_c and y_c are constrained by the parameters a_1 and a_2 (see constraint (7.4)). The dependence among the parameters makes the parameter constraints very complicated, as we have seen in (7.9). And more critically, the dependence among the parameters makes the estimation problem more difficult.

As a remedy to the difficulty caused by the parameterization, we propose an alternative parameter set to describe the boundaries. The objective of this new parameterization is to remove the high correlation among the shape parameters.

We remove the dependence among the pavement boundary parameters first. Instead of using correlated parameters a_1 and a_2, we propose to use parameters a_1 and w_2, the distance between the left and right pavement boundaries. And to eliminate the dependence between a_1 and x_c, we replace x_c with x_c', the horizontal coordinate of the intersection point of the circle with radius a_1 and the line segment passing the circle center and parallel to the x axis. Figure 7.2.5 illustrates the new parameter set with right curved pavement boundaries. In this case, the center is to the right of the circles. The right pavement boundary corresponds to the circle with radius a_1. C is the center of the circles. CA is a line segment parallel to the x axis and passing through the point C. A is the intersection of the line segment CA and the circle with radius a_1. Then the Cartesian coordinates of A are (x_c', y_c), and

$$x_c = x_c' + a_1. \tag{7.10}$$

The radius of the left boundary, a_2, can be represented by the sum of the radius of the right boundary, a_1, and the distance between the two boundaries, w_2.

$$a_2 = a_1 + w_2 \tag{7.11}$$

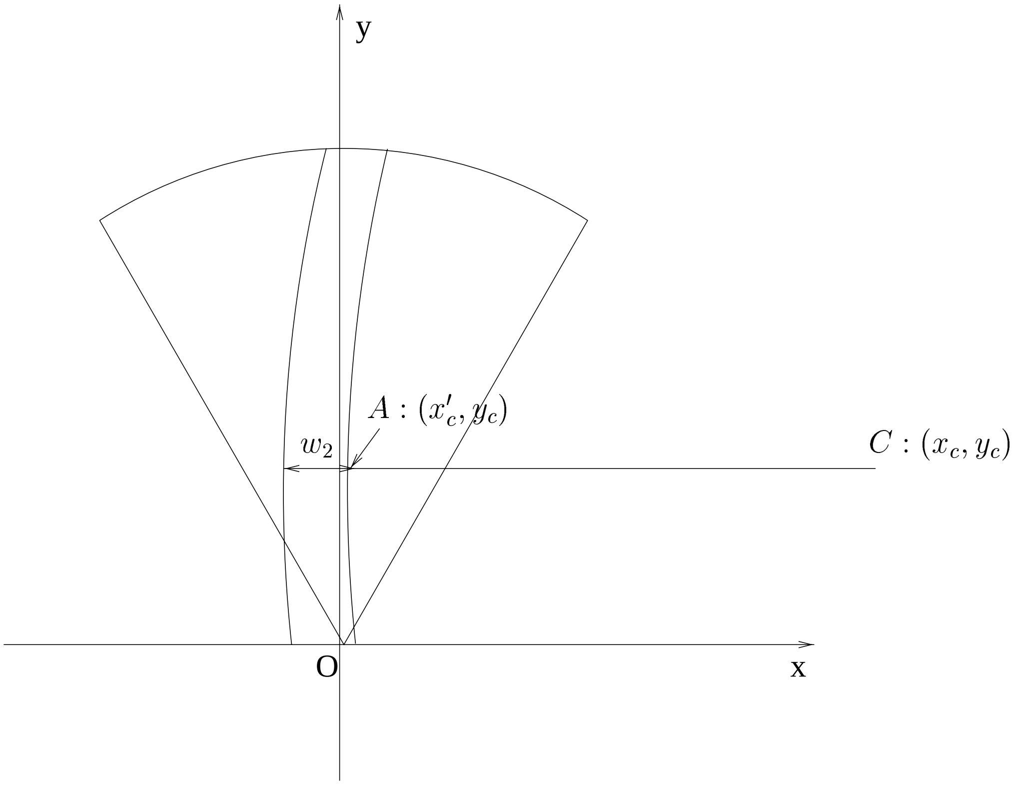


Figure 7.2.5: Reparameterization for the case where the center is to the left of the road

Figure 7.2.6 shows left curved boundaries. In this case, the center is to the left of the circles and the left pavement boundary corresponds to the circle with radius a_1. Let CA still be the line segment parallel to the x axis connecting two points, C and A. Then (7.11) holds for this case, too. And A's coordinates (x_c', y_c) satisfies

$$x_c = x_c' - a_1 \tag{7.12}$$

In order to unify (7.10) and (7.12) we define a new parameter a as

$$a = \begin{cases} a_1, & \text{if the center is to the right of the boundaries,} \\ -a_1, & \text{otherwise.} \end{cases} \tag{7.13}$$

In other words, the magnitude of a is the radius value of the smaller circle, and the sign of a depends on the relative position of the center and the boundaries. Then it is immediate to get a_1 given a,

$$a_1 = |a| \tag{7.14}$$

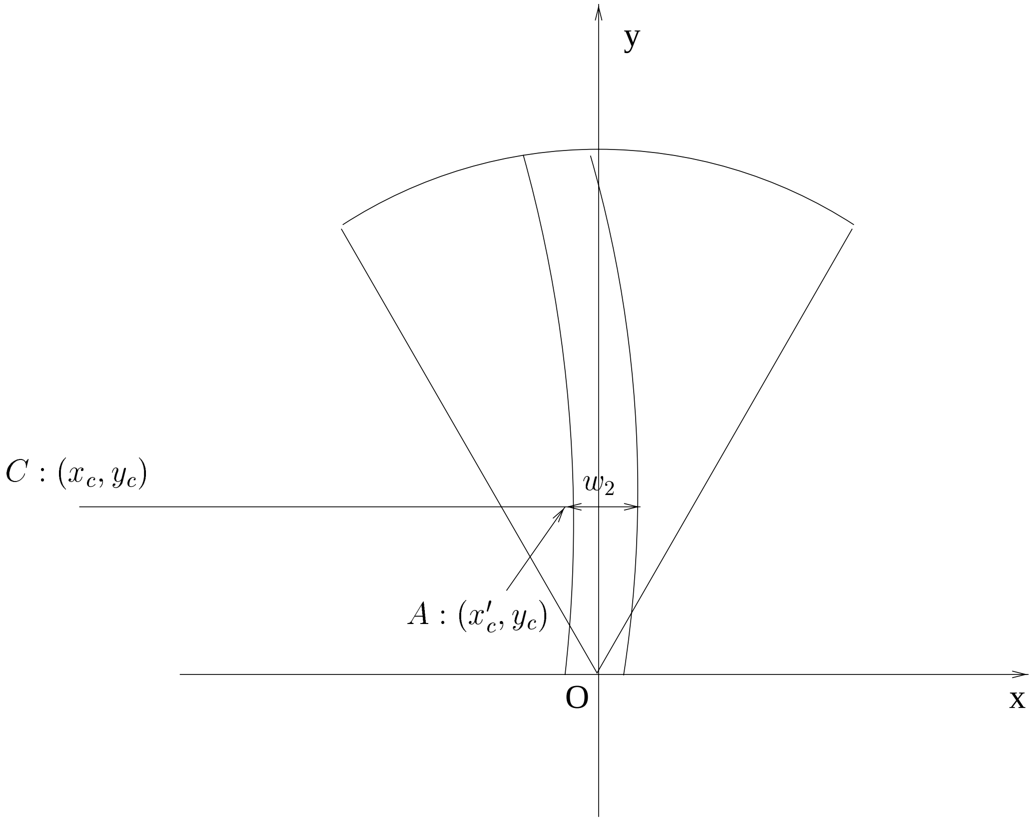


Figure 7.2.6: Reparameterization for the case where the center is to the right of the road

With this new parameter a, rewriting (7.10) and (7.12) in a united form,

$$x_c = x'_c + a. \tag{7.15}$$

And with (7.13), (7.11) is rewritten as

$$a_2 = w_2 + |a| \tag{7.16}$$

Define the new pavement shape parameters as $\underline{\theta}_c^r = \{a, w_2, x'_c, y_c\}$, the previous parameters $\{a_1, a_2, x_c, y_c\}$ can be easily calculated with the relationships (7.14), (7.16), and (7.15). That is, $\{a, w_2, x'_c, y_c\}$ are capable of representing the pavement boundaries and they are independent of each other.

To eliminate the dependence among the lane shape parameters $\{a'_1, a'_2, x_c, y_c\}$, a similar decoupling technique is applied. The new lane shape parameters $\underline{\theta}_c^o = \{a', w'_2, x'_c, y_c\}$ satisfying

$$a' = \begin{cases} a'_1, & \text{if the center is to the right of the boundaries,} \\ -a'_1, & \text{otherwise.} \end{cases} \tag{7.17}$$

$$\begin{aligned} w'_2 &= a'_2 - |a'| \quad (7.18) \\ x'_c &= x_c - a' \quad (7.19) \end{aligned}$$

In the fusion framework, we could easily define the joint shape parameters $\underline{\theta}_c = \{\underline{\theta}_c^r, \underline{\theta}_c^o\} = \{a, w_2, a', w'_2, x'_c, y_c\}$. But in this parameterization, the lane and pavement boundaries a and a' are coupled, which is not a desirable property. To remove this remaining dependence, we propose a unified parameter set for both lane and pavement boundaries.

In Figure 7.2.7, we illustrate the unified shape parameters $\{a, w_2, w'_1, w'_2, x'_c, y_c\}$ for both lane and pavement boundaries. In this plot, the solid curves represent the pavement boundaries and the dashed curves represent the lane boundaries. For this parameter set, a is defined by (7.13). The distance parameters w_2, w'_1, and w'_2 satisfying

$$a_2 = w_2 + |a| \tag{7.20}$$

$$a'_1 = w'_1 + |a| \tag{7.21}$$

$$a'_2 = w'_2 + |a| \tag{7.22}$$

And the x coordinate of the circle center is determined by (7.15).

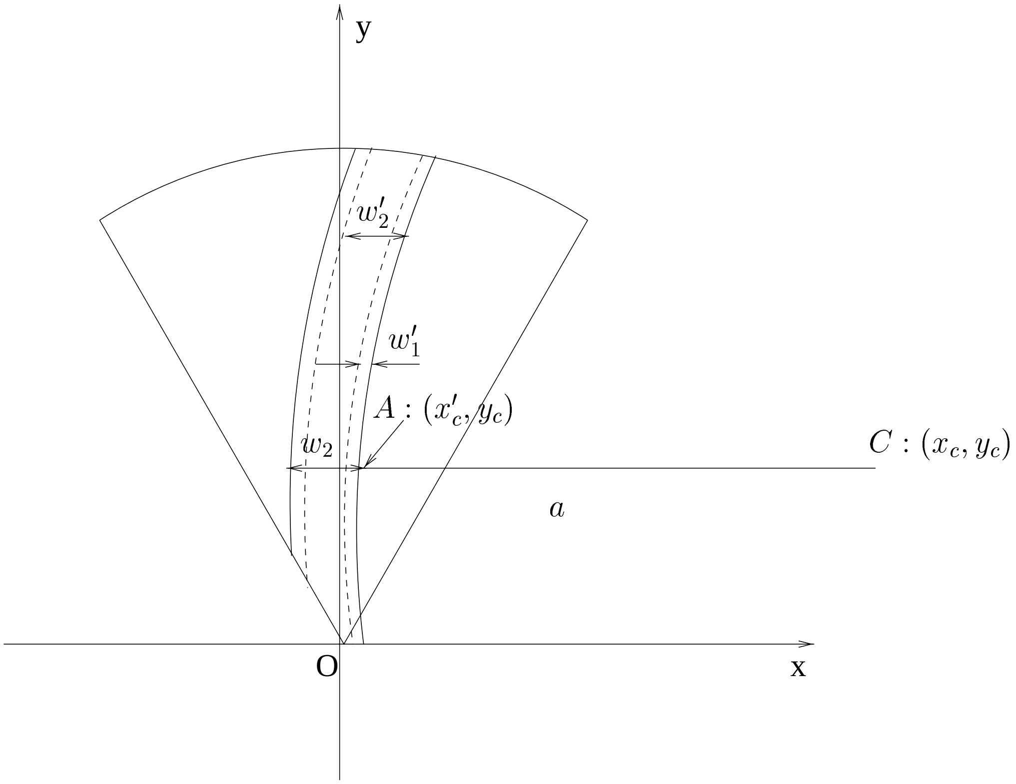


Figure 7.2.7: Unified parameterization for both lane and pavement boundaries

The parameters $\{a, w_2, w'_1, w'_2, x'_c, y_c\}$ are able to represent the lane and pavement boundaries and they are independent from each other. In our fusion algorithm, we propose to use this parameterization to describe the prior information on the boundaries. Let us define $\underline{\theta}_c = \{a, w_2, w'_1, w'_2, x'_c, y_c\}$. The constraints on $\underline{\theta}_c$ can be easily derived from the previous constraints on $\{a_1, a_2, a'_1, a'_2, x_c, y_c\}$ as follows

$$\begin{aligned} P(\underline{\theta}_c) \;=\; & \frac{1}{\gamma_c} \cdot I_{[|a|-p_1,|a|+w_2+p_2]}((a+x'_c)^2+y_c^2) \cdot I_{[\alpha_{min},\alpha_{max}]}\left(\text{atan}\left(\frac{y_c}{a+x'_c}\right)\right) \\ & \cdot I_{[W_{min},W_{max}]}(w_2) \cdot I_{[W'_{min},W'_{max}]}(w'_2-w'_1) \cdot I_{[0,w'_2]}(w'_1) \cdot I_{[w'_1,w_2]}(w'_2) \end{aligned} \tag{7.23}$$

where γ_c is a normalizing constant.

7.3 Imaging Likelihoods

We have seen that in the fusion algorithm presented in Chapter VI, the joint MAP estimation is approximated by an ad hoc empirical parameter estimation. The root cause for this approximation is that the optical and radar matching functions are not compatible and different weights have to be imposed on them when they are combined in the fusion framework. Although the radar imaging process is represented with a log-normal density, the optical imaging process is described with an empirical likelihood function, which makes it impossible to theoretically derive the weights.

To avoid the ad hoc approximation of the joint MAP estimation, we propose to use real probability densities to represent both the radar and optical imaging processes in this chapter, and hence no weighting scheme is needed in this newly proposed fusion algorithm.

7.3.1 Radar Imaging Likelihood

In Chapter VI we presented a log-normal pdf as a model for the radar imaging process. With the circular shape parameters, we will evaluate (6.7) as the radar likelihood function and (6.13) as the radar matching function. In order to be consistent with other symbols in this chapter, we rewrite the likelihood and matching functions as

$$p(z^r \mid \underline{\theta}_c^r) = \prod_{(r,\phi)\in\mathcal{L}} \frac{1}{z^r_{r\phi}\sqrt{2\pi\sigma^2_{r\phi}(\underline{\theta}_c^r)}} \exp\left\{-\frac{1}{2\sigma^2_{r\phi}(\underline{\theta}_c^r)}\left[\log z^r_{r\phi} - \mu_{r\phi}(\underline{\theta}_c^r)\right]^2\right\} \tag{7.24}$$

and

$$L_c^r(z^r, \underline{\theta}_c^r) \;=\; -\left(w^{rd} N^{rd} \log \hat{\sigma}^{rd} + N^{lt} \log \hat{\sigma}^{lt} + N^{rt} \log \hat{\sigma}^{rt}\right) \tag{7.25}$$

7.3.2 Optical Imaging Likelihood

In Section 6.2.2 we utilized an empirical likelihood function (6.15) to describe the optical imaging process. Since this likelihood function involves a computationally intractable normalizing constant, $\gamma(\underline{\theta}_q^o)$, it becomes a major obstacle in the fusion procedure (6.30). The value of $\gamma(\underline{\theta}_q^o)$ effectively determined different weightings of the two matching functions, $L^o(z^o, \underline{\theta}_q^o)$ and $L^r(z^r, \underline{\theta}_q^r)$. (In (6.31), β introduces a different weight on the optical matching function.) Since there is no analytical approach to derive the weight β, we experimentally determine the value of β by trying different weights on a training set of a large number of radar and optical image pairs. This trial-and-error approach of choosing the weight value is not only inaccurate but also time consuming. Since the root cause of the weighting scheme is that the optical imaging likelihood function is not normalized, to avoid this dilemma we propose propose a normalized Gaussian density to describe the optical imaging process.

Given boundary shape parameters $\underline{\theta}_c^o$, we assume that the optical image gradient magnitude G_m is the ideal gradient magnitude $S(\underline{\theta}_c^o)$ contaminated with additive white Gaussian noise W^o,

$$G_m = S(\underline{\theta}_c^o) + W^o, \tag{7.26}$$

where W^o are i.i.d. Gaussian random variables with mean 0 and unknown variance σ^2, and where $S(\underline{\theta}_c^o)$ is a tapered/spike function image with unknown height. Thus the optical imaging process can be modeled by the conditional probability of the optical random field

Z^o taking a realization z^o, which is a Gaussian pdf,

$$p(z^o|\underline{\theta}_c^o) = \prod_{(x,y)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[g_m(x,y) - s(\underline{\theta}_c^o, x, y)]^2\right\} \tag{7.27}$$

To better explain the tapered image, let us define a spike function

$$f(\alpha, d) \triangleq \frac{1}{1+\alpha d^2}, \tag{7.28}$$

where α is a smoothing constant which controls the effective width of the spike function. Then the intensity value of the tapered image $S(\underline{\theta}_c^o)$ at the pixel (x,y), $s(\underline{\theta}_c^o, x, y)$, can be written as

$$s(\underline{\theta}_c^o, x, y) = A\, f(\alpha, d_1(x,y)) + A\, f(\alpha, d_2(x,y)) \tag{7.29}$$

where A is the unknown height and d_1 and d_2 are the distances from the pixel (x,y) to the left and right lane boundaries, respectively.

Substituting (7.29) back to (7.27), we have the optical imaging likelihood function

$$\begin{aligned} p(z^o|\underline{\theta}_c^o) &= \prod_{(x,y)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[g_m(x,y) - Af(\alpha, d_1(x,y)) - Af(\alpha, d_2(x,y))]^2\right\} \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{(x,y)}[g_m(x,y) - Af(\alpha, d_1(x,y)) - Af(\alpha, d_2(x,y))]^2\right\} \end{aligned} \tag{7.30}$$

One thing worth mentioning is how to compute $d_1(x,y)$ and $d_2(x,y)$ for a given pixel (x,y). In Figure 7.3.8, C is the center of the concentric circles, and P is a point with Cartesian coordinates (x,y) in the field-of-view of the optical image. CP is a line segment connecting Points C and P. Point A is the intersection of the right lane and the extension of CP. Since CA is the radius of the circle corresponding to the right lane boundary, CA is perpendicular to the tangent of the lane boundary. Thus, the length of PA is the distance from the point P to the right lane boundary, i.e.,

$$\begin{aligned} d_1(x,y) &= \|PA\| = |\|CA\| - \|PA\|| = |a_1' - d_0| \\ &= |a_1' - \sqrt{(x-x_c)^2 + (y-y_c)^2}| \\ &= ||a'| + w_1' - \sqrt{(x - x_c' - a)^2 + (y-y_c)^2}| \end{aligned} \tag{7.31}$$

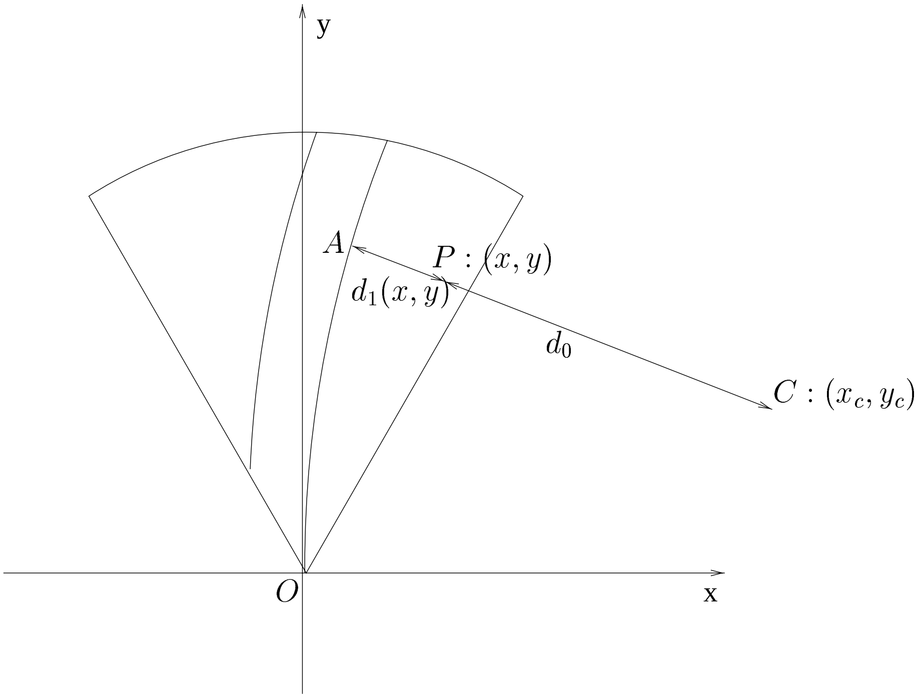


Figure 7.3.8: An ideal lane scenario on the ground plane

In (7.30), the nuisance parameters A and σ^2, corresponding to the maximum value of the gradient magnitude of the ideal lane boundaries and variance of the additive Gaussian noise, can be empirically estimated from the observed optical image z^o by maximum likelihood. Since the likelihood function is a normal distribution, the maximum likelihood and least squares estimate of the nuisance parameters are equivalent,

$$\hat{A} = \frac{\sum_{(x,y)} g_m(x,y)[f(\alpha, d_1(x,y)) + f(\alpha, d_2(x,y))]}{\sum_{(x,y)}[f(\alpha, d_1(x,y)) + f(\alpha, d_2(x,y)]^2}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{(x,y)}[g_m(x,y) - \hat{A}f(\alpha, d_1(x,y)) - \hat{A}f(\alpha, d_2(x,y))]^2 \tag{7.32}$$

Substituting these estimates of A and σ^2 back into (7.30), and taking the logarithm results in

$$\log p(z^o|\underline{\theta}_c^o) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\hat{\sigma}^2 + \frac{N}{2}\log N - \frac{N}{2}, \tag{7.33}$$

where N is the number of pixels in the optical image, and hence is a constant regardless of different lane boundary parameters. Therefore, the important term in this log-likelihood function is $-\frac{N}{2}\log\hat{\sigma}^2$, i.e., given a hypothetical lane boundary parameters $\underline{\theta}_c^o$, the fidelity of the hypothesized boundary to the observed optical image z^o is evaluated by the logarithm of the error residual between the model and the observed image. A smaller residual gives better fidelity in the sense of better (higher) likelihood of the model. Define the optical matching function as

$$L_c^o(z^o, \underline{\theta}_c^o) = -\frac{N}{2}\log\hat{\sigma}^2. \tag{7.34}$$

7.4 Joint MAP Estimate for Lane and Pavement Boundaries

Since the prior distribution of the deformation parameters and the imaging likelihood functions are available, we shall pose the lane and pavement edge detection problem in a Bayesian framework. Given z^r as a realization of the radar random field Z^r and z^o as a realization of the optical random field Z^o, the lane and pavement boundary detection with fusion technique can be solved by the joint MAP estimate

$$\begin{aligned}
\hat{\underline{\theta}}_c &= \{\hat{\underline{\theta}}_c^{\,r}, \hat{\underline{\theta}}_c^{\,o}\} = \arg\max_{\underline{\theta}_c} p\left(\underline{\theta}_c|z^r, z^o\right) \\
&= \arg\max_{\{\underline{\theta}_c^r, \underline{\theta}_c^o\}} p\left(\underline{\theta}_c^r, \underline{\theta}_c^o|z^r, z^o\right) \\
&= \arg\max_{\{\underline{\theta}_c^r, \underline{\theta}_c^o\}} p(\underline{\theta}_c^r, \underline{\theta}_c^o)\, p(z^r|\underline{\theta}_c^r)\, p(z^o|\underline{\theta}_c^o)
\end{aligned} \tag{7.35}$$

The derivation of (7.35) follows exactly the same steps as the derivation of (6.29).

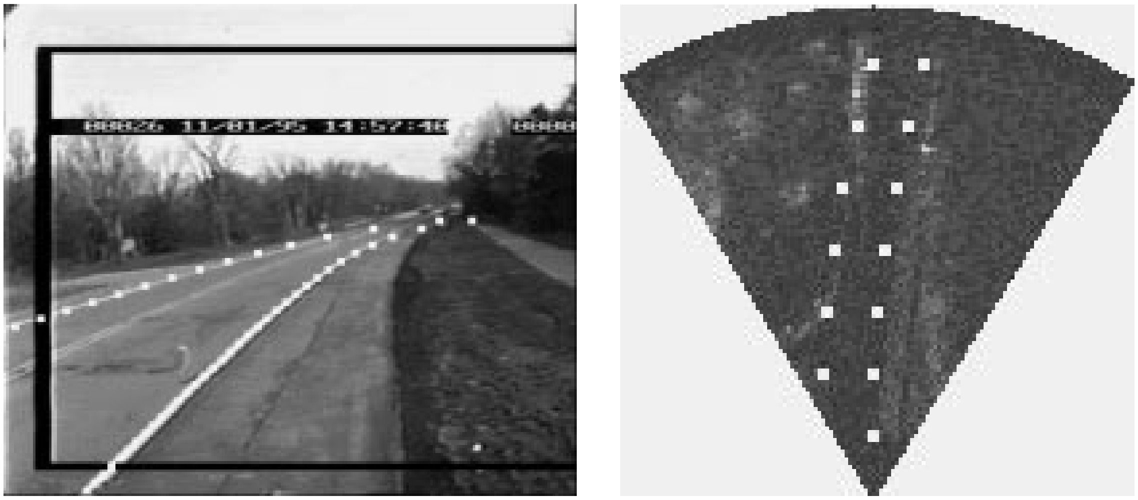
Taking logarithm of the objective function, the joint MAP estimate turns to

$$\begin{aligned}\underline{\hat{\theta}}_c &= \arg\max_{\{\underline{\theta}_c^r, \underline{\theta}_c^o\}} \{\log p(\underline{\theta}_c^r, \underline{\theta}_c^o) + \log p(z^r|\underline{\theta}_c^r) + \log p(z^o|\underline{\theta}_c^o)\} \\ &= \arg\max_{\{\underline{\theta}_c^r, \underline{\theta}_c^o\}} \{\log p(\underline{\theta}_c^r, \underline{\theta}_c^o) + L_c^r(z^r, \underline{\theta}_c^r) + L_c^o(z^o, \underline{\theta}_c^o)\} \qquad (7.36)\end{aligned}$$

We have seen that the main obstacle in the fusion algorithm proposed in Chapter VI is how to weighting the radar and optical matching functions. The root cause for this weighting scheme is due to the dramatically different dynamic ranges of two different matching functions. In Chapter VI, we only got an empirical solution by experimentally choosing the weighting constant β in (6.31).

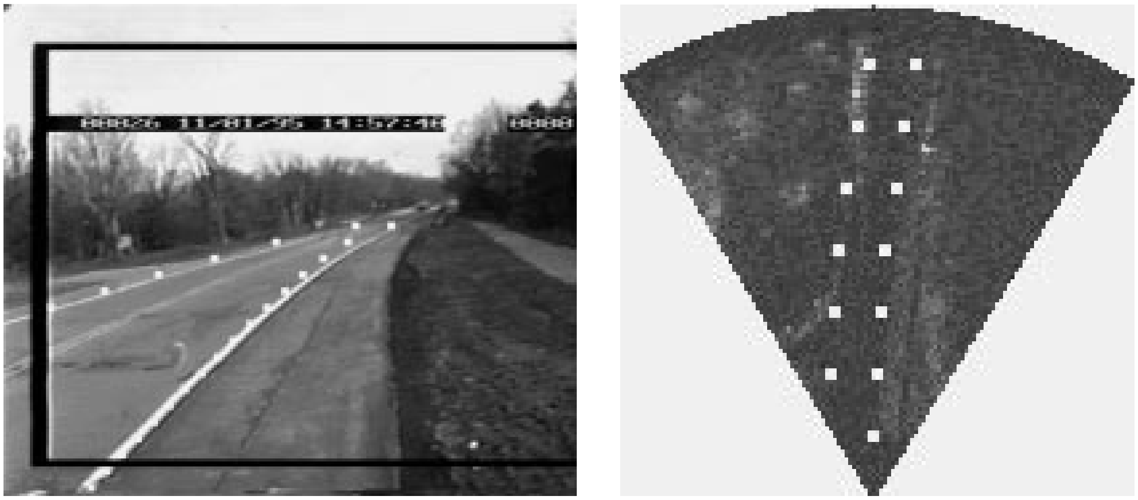
In the fusion algorithm presented above in Equation (7.36), the radar imaging likelihood is modeled as a log-normal pdf and optical imaging likelihood is described by a Gaussian density. Since both likelihood functions are normalized probability density functions, they are compatible. Then the effective parts of the logarithm of the imaging likelihood functions, i.e., the radar and optical matching functions should have well behaved dynamic ranges. For the optical and radar image pair shown in Figure 6.3.3, the dynamic ranges of the optical and radar matching functions are 1.38×10^4 and 3.20×10^3, respectively. Since the optical image has higher resolution than the radar image, the optical imaging log-likelihood tends to be more discriminative than the radar imaging log-likelihood in a well illuminated environment. Therefore, the two dynamic ranges are actually compatible. And hence in this fusion algorithm we do not have to worry about the different weightings on the matching functions, $L_c^r(z^r, \underline{\theta}_c^r)$ and $L_c^o(z^o, \underline{\theta}_c^o)$.

In Figure 7.4.9, we show the lane and pavement boundary detection results with single sensor based detection algorithms and the proposed joint MAP estimate method (7.36). We observe that the lane boundary detection result shown in Figure 7.4.9(a) is not right but the pavement boundary detection result shown in Figure 7.4.9(b) is correct, and that the fusion algorithm yield the correct joint boundary detection (Figures 7.4.9(c) and (d)). That is, the optical image does not dominate the parameter estimation in fusion process. In the parabolic shape model case, the optical image does dominate the parameter estimation in fusion process (see Figure 6.3.5).



(a) Single sensor based lane edge detection

(b) Single sensor based pavement edge detection



(c) Lane edge detection

(d) Pavement edge detection

Figure 7.4.9: Boundary detection results with fusion method

7.5 Experimental Results

We have implemented the proposed joint boundary detection algorithm to locate lane and pavement boundaries in registered optical and radar images. In the implementation, the boundary appearance is described by concentric circular shape models. We have carried out three categories of experiments in our boundary detection effort.

① Detect lane boundaries using the optical image alone. MAP estimation algorithm is employed to detect the lane boundaries in the optical image where the circular shape template with its parameters' distribution plays the role of the *a priori* information and the optical imaging process (7.30) plays the role of the likelihood function.

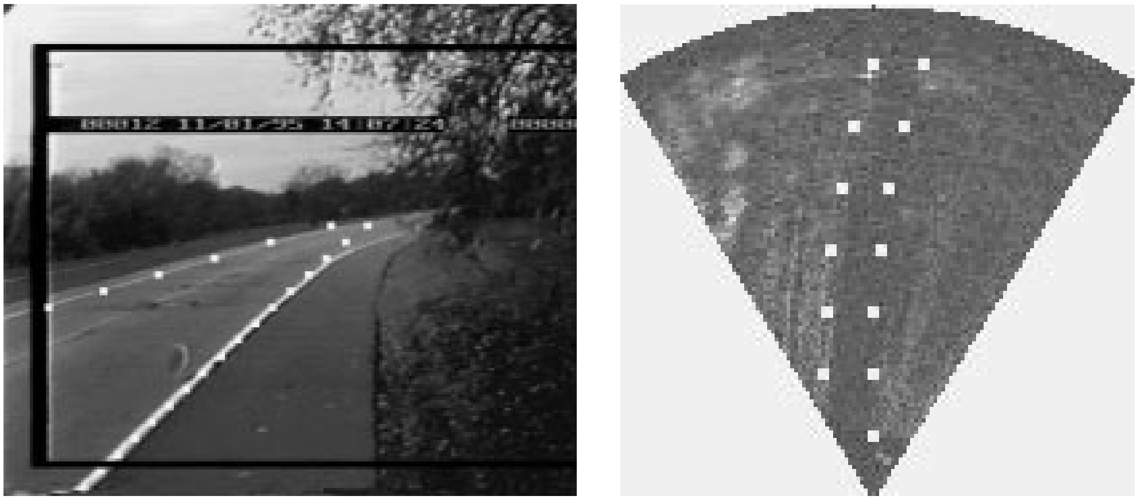
② Detect pavement boundaries using the radar image alone. MAP estimation algorithm [70] is employed to detect the pavement boundaries in the radar image. In the MAP estimator, the circular shape template and its parameters' distribution serve as the *a priori* information and the radar imaging process (7.24) serves as the likelihood function.

③ Jointly detect lane and pavement boundaries with fusion approach (7.36) using information from both optical and radar images. In this fusion approach, circular shape models are utilized to describe the lane and pavement boundaries, and the Gaussian and log-normal densities are employed to represent the optical and radar imaging likelihood functions.

In Figure 7.5.10 we show the detection results obtained with the above three methods for a pair of optical and radar images, both of good quality. Figure 7.5.10(a) shows the detected lane boundaries in the optical image and Figure 7.5.10(b) shows the detected pavement boundaries in the radar image. Both of them are quite satisfactory. This is due to the fact that both the optical and radar images are of high quality and each of them provides enough information for the boundary detection. Figures 7.5.10(c) and (d) show the boundary detection results using the proposed fusion approach. And we observe that the fusion algorithm does not degrade the boundary detection performance compared to that of single sensor based algorithms when both the optical and radar images are of good quality.

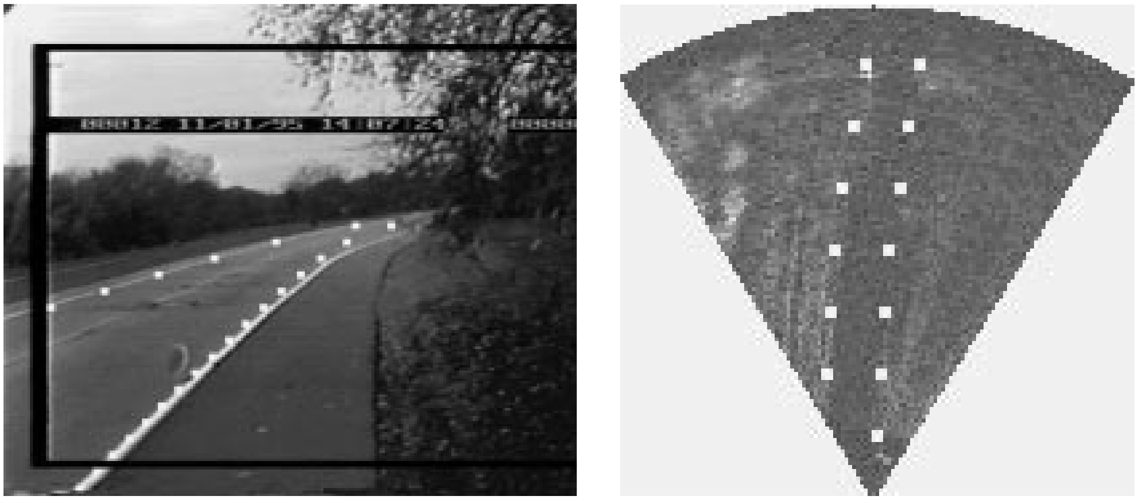
In Figure 7.5.11 we show the detection results for a pair of optical and radar images of different qualities. The optical image is degraded by the presence of snow. Wrong lane boundary detection result is obtained when only the optical image is used (Figure 7.5.11(a)). However, the radar image still offers sufficient information to correctly detect the pavement boundaries (Figure 7.5.11(b)). In the fusion approach, since we make use of information from both optical and radar sensors to jointly detect the lane and pavement boundaries, the radar image helps the lane detection in the optical images (Figures 7.5.11(c) and (d)).

In Figure 7.5.12 we show the detection results for a pair of fair-quality optical and bad-quality radar images. The single sensor based algorithms do not operate well in either lane or pavement boundary detection. Figure 7.5.12(a) gives the lane detection result in the optical image. The traffic sign to the right of the road misleads the detected boundaries curving to the left. In Figure 7.5.12(b), a homogeneous region to the left of the road results in wrong pavement boundary detection. Information from both optical and radar images is explored in the fusion approach and the redundancy, diversity and complementarity between the optical and radar sensors significantly improve the boundary detection performance. In Figures 7.5.12(c) and (d), we show that satisfactory results have been achieved with the joint boundary detection algorithm.



(a) Single sensor based lane edge detection

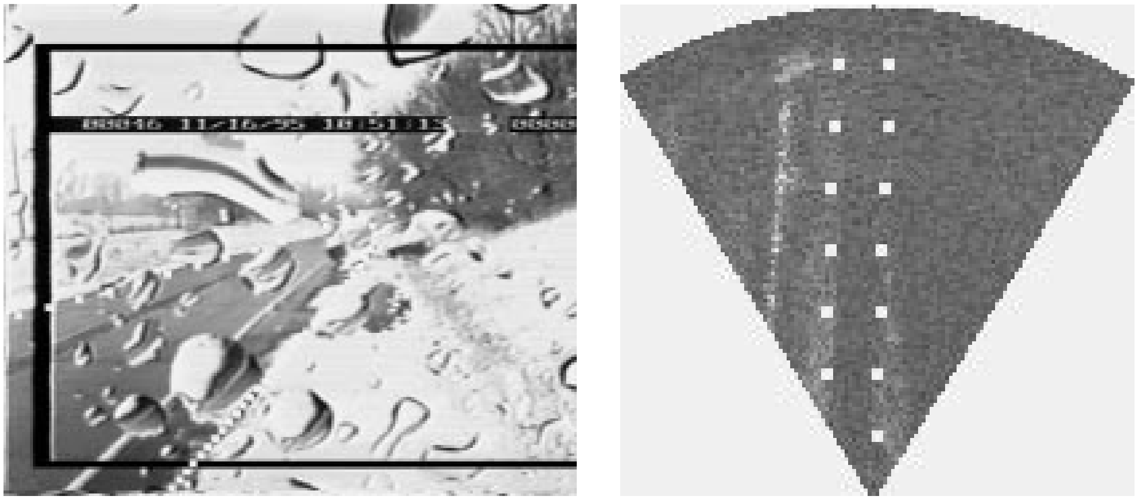
(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method

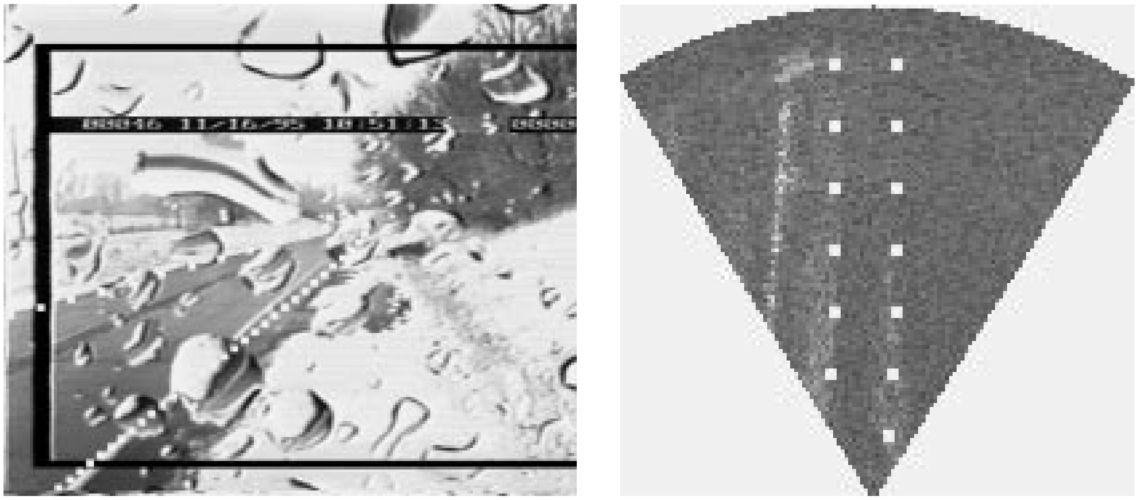
(d) Pavement edge detection with fusion method

Figure 7.5.10: Performance comparison of the fusion and single sensor based methods



(a) Single sensor based lane edge detection

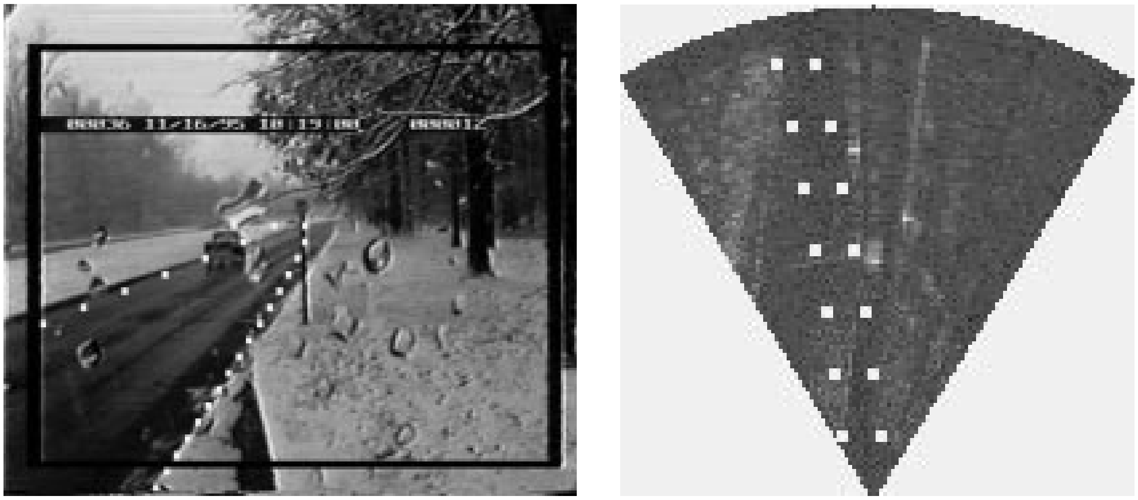
(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method

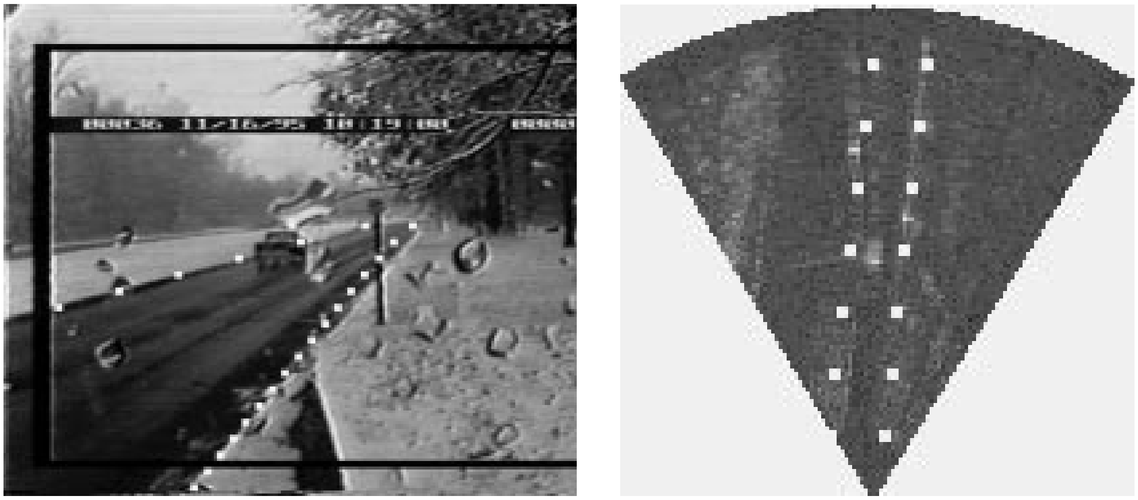
(d) Pavement edge detection with fusion method

Figure 7.5.11: Performance comparison of the fusion and single sensor based methods



(a) Single sensor based lane edge detection

(b) Single sensor based pavement edge detection



(c) Lane edge detection with fusion method

(d) Pavement edge detection with fusion method

Figure 7.5.12: Performance comparison of the fusion and single sensor based methods

All the examples have demonstrated that circular shape models and the newly formulated radar and optical likelihoods are indeed successful in detecting lane and pavement boundaries. And the proposed fusion algorithm improves the boundary detection performance when either the optical or the radar image is unable to provide sufficient information by itself. We note that the proposed fusion algorithm does not degrade the performance of the individual detection results when they are good by themselves.

In chapter VI, the fusion algorithm suffers from the non-compatibility of optical and radar matching functions, whose main cause is that the optical imaging process is modeled by a non-normalized pdf. To compensate the different dynamic ranges of the two matching functions, an empirically selected weight (β in (6.31)) is imposed on the optical matching function when it combines with the radar matching function in the fusion process. Since the weight is not analytically derived, it is not a perfect number that completely reflects the difference between the two matching functions.

On the contrary, the joint MAP estimator proposed in this chapter overcomes this difficulty by modeling optical imaging process with a Gaussian pdf. Then in the fusion setting we have two normalized pdf's and the sum of their logarithms is naturally derived from the basic MAP formulation. Therefore, no more fancy weighting scheme needs to be introduced into this fusion work. And most importantly, since no empirical numbers are involved in the parameters' estimation, this algorithm yields more accurate boundary detection.

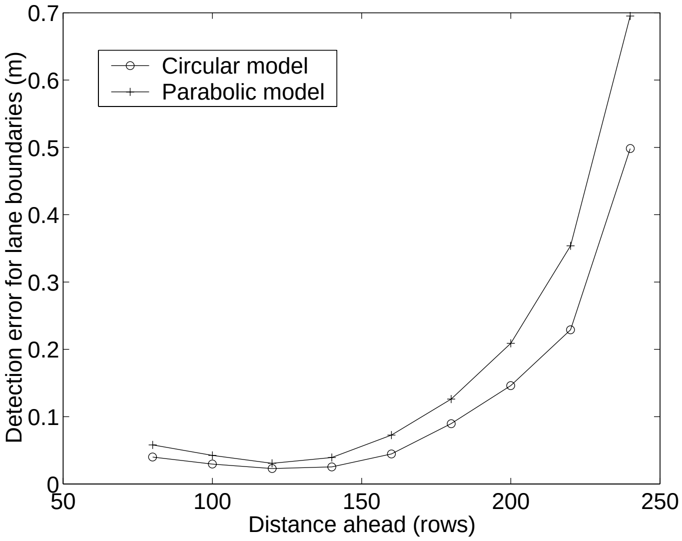
Another merit of the fusion algorithm proposed in this chapter is that it adopts circular shape models instead of parabolic shape models to better reflect the road scenes in the real world.

To compare the improvement of the fusion algorithm proposed in this chapter over the fusion algorithm described in chapter VI, we applied both algorithms to the database of 25 optical and radar image pairs referred earlier. The detection errors are plot in Figure 7.5.13. The fusion algorithm proposed in this chapter is labeled as “Circular model” and the fusion algorithm described in chapter VI is label as “Parabolic model” in this plot. Both Figures 7.5.13(a) and (b) demonstrate that the circular model fusion algorithm outperforms the parabolic model fusion algorithm in detecting the lane and pavement boundaries.

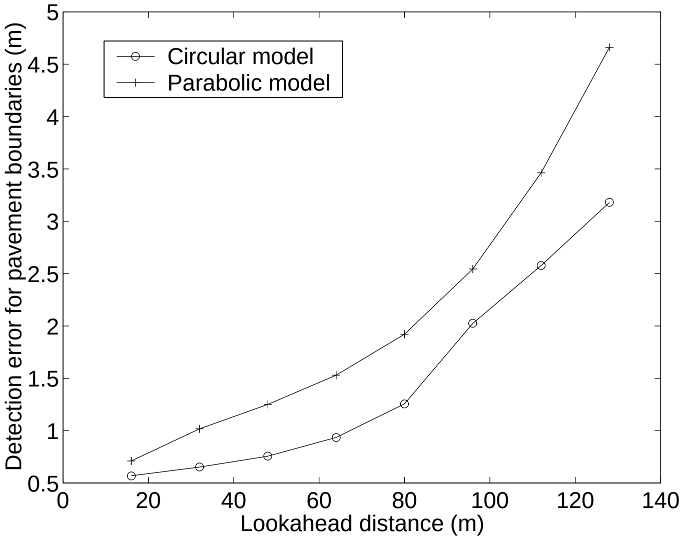
As we have stated at the beginning of this chapter, the circular shape parameters, $\{a_1, a_2, x_c, y_c\}$ for pavement boundaries and $\{a_1', a_2', x_c, y_c\}$ for lane boundaries, possess a number of advantages over the parabolic shape parameters. The circular shape parameters are of the same units and the same order of magnitude for the dynamic ranges. And the variation of any parameter uniformly affects the boundary appearance. An immediate implication from these merits is that the confidence measures defined in Chapter 6.4 shall have a desired property for parameter estimation — different parameters’ confidence measure should be compatible.

We take the pavement boundary model as an example to illustrate the compatible confidence measures. We know that the two parameter sets $\{a, w_2, x_c', y_c\}$ and $\{a_1, a_2, x_c, y_c\}$ are equivalent in representing the boundaries, and the former parameter set is a decoupled version of the latter one. So the sensitivity of the matching function with respect to the decoupled parameters is more of our concern in the context of parameter estimation. In Tables 7.1 and 7.2 we gave the curvature results for some radar and optical images, respectively. From Table 7.1 we observe that the curvatures for different shape parameters are in the same order of magnitude, which leads the parameter estimation a much better conditioned problem than the parameter estimation problem for the parabolic shape models.

In each radar or optical image, the curvature of y_c, the circle center’s vertical coordinate, is smaller than all other curvatures, which means that the radar or optical matching



(a) Errors for lane boundary detection



(b) Errors for pavement boundary detection

Figure 7.5.13: Performance comparison of fusion algorithms

Image name	Curv$(\hat{a})$	Curv$(\hat{w}_2)$	Curv$(\hat{x}_c')$	Curv$(\hat{y}_c)$
anhinesb	9.49×10^{-3}	9.43×10^{-3}	1.35×10^{-3}	0.44×10^{-3}
anhinesc	9.12×10^{-3}	8.53×10^{-3}	1.42×10^{-3}	0.56×10^{-3}
anhinesg	8.77×10^{-3}	7.71×10^{-3}	0.75×10^{-3}	0.55×10^{-3}
evhines	8.09×10^{-3}	7.35×10^{-3}	1.38×10^{-3}	0.47×10^{-3}
hinesbev	8.11×10^{-3}	8.14×10^{-3}	1.06×10^{-3}	0.47×10^{-3}

Table 7.1: Curvature results for some radar images with circular shape models

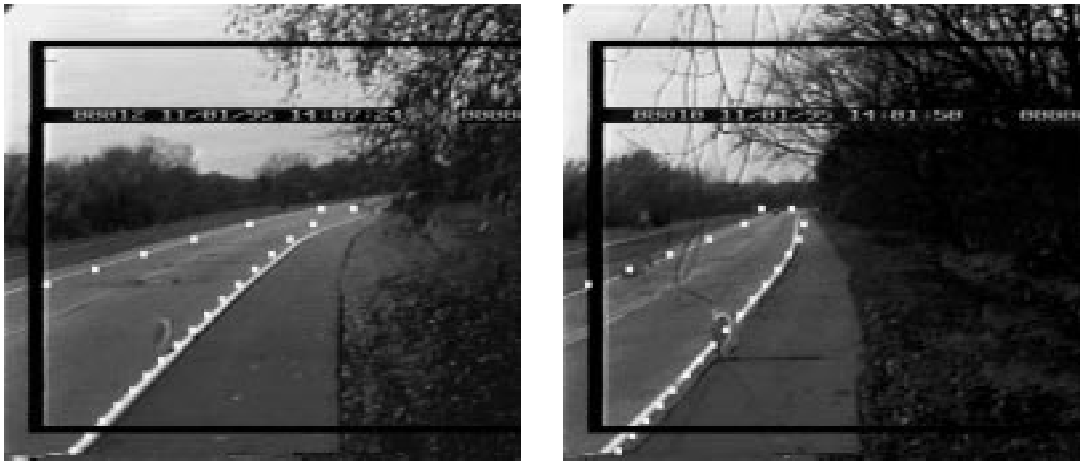
Image name	Curv$(\hat{a}')$	Curv$(\hat{w}_2')$	Curv$(\hat{x}_c')$	Curv$(\hat{y}_c)$
anhinesb	4.17×10^{-2}	1.34×10^{-2}	2.96×10^{-2}	0.03×10^{-2}
anhinesc	4.18×10^{-2}	1.16×10^{-2}	2.30×10^{-2}	0.01×10^{-2}
anhinesg	1.46×10^{-2}	11.46×10^{-2}	3.93×10^{-2}	0.34×10^{-2}
evhines	1.89×10^{-2}	3.73×10^{-2}	2.38×10^{-2}	0.04×10^{-2}
hinesbev	4.61×10^{-2}	1.40×10^{-2}	3.78×10^{-2}	0.01×10^{-2}

Table 7.2: Curvature results for some optical images with circular shape models

function is not so sensitive with the parameter y_c as with other parameters. This coincides with our knowledge that changing y_c a little bit will not change the whole road scenario in our field of view.

In most radar images, observation data near the left pavement boundary behave similarly as those near the right boundary. Thus the sensitivities of the matching function to both left and right pavement boundaries are very close. That is, the confidence measure values for parameters a and w_2 shall be very close. In Table 7.1, the curvature results for a and w_2 have successfully justified that this is the case.

On the contrary, in optical images, the left and right lane boundaries are of different qualities due to their different distances from the camera. In Figure 7.5.14 we show two different optical images with their boundary detection results. The two images are called “anhinesb”(Figure 7.5.14(a)) and “anhinesg”(Figure 7.5.14(b)), respectively. In both images, the right lane boundaries are closer to the camera and hence they have higher resolution than the left ones. In Image “anhinesb” the right lane is laid on the circle with the smaller radius, and thus the parameter a corresponds to the right boundary and the parameter w_2 corresponds to the left boundary. Therefore, the confidence measure for a is considerably larger than that for w_2 (the first row in Table 7.2). In Image “anhinesg” the left lane is laid on the circle with the smaller radius. Thus a corresponds to the left boundary and w_2 corresponds to the right boundary. So the confidence measure for a is considerably smaller than that for w_2 (the third row in Table 7.2).



(a) anhinesb (b) anhinesg

Figure 7.5.14: Examples to show what confidence measure indicates

With experiments on real data we have shown the advantages of circular shape models over parabolic models. In circular shape models, the parameters are compatible with each other and their confidence measures indeed indicate the confidence of the parameters’ estimates.

7.6 Conclusions

In this chapter we have proposed a new type of deformable templates to describe the lane and pavement boundaries — concentric circular shape model. Since the US highway systems actually laid the lane and pavement boundaries on concentric circles, circular shape models are better choices than their polynomial approximations. With experiments we have shown the advantages of this new parameterization over polynomial models. Possessing the same unit, the same order of magnitude, similar confidence measures of their estimates, the circular shape parameters result in a much better conditioned parameter estimation problem.

In this new fusion algorithm we also have adopted a Gaussian pdf to model the optical imaging process. Since only normalized pdf's are involved in the joint MAP estimator, no weighting scheme is necessary to compensate the difference between matching functions as occurred in the boundary detection algorithm presented in Chapter VI. Without any experimentally selected weight, our experiments have shown that this fusion algorithm yields more accurate and robust lane and pavement boundary detection results than the algorithm which uses the empirical imaging likelihood function (Chapter VI).

CHAPTER VIII

Summary and Future Work for Boundary Detection

8.1 Summary

Part of this dissertation work concentrates on simultaneous lane and pavement boundary detection using optical and radar sensors mounted on the host vehicle. In our work, the lane and pavement boundaries are represented by deformable templates and thus the boundary detection becomes a problem of estimating the shape parameters. We model the optical and radar imaging processes with likelihood functions. Then the boundary detection problem is set in a Bayesian framework and joint maximum *a posteriori* method is employed to estimate the shape parameters for both lane and pavement boundaries. Since the fusion method makes most of the information from both optical and radar images, it yields more accurate and reliable boundary detection results than the separate lane/pavement boundary detection algorithms.

We investigated two boundary shape models — parabolic and circular shape models. Parabolic shape models have been used in previous separate lane/pavement boundary detection algorithms and they can capture the basic characteristic of the boundary shapes within the field-of-view of the sensors. However, due to the incompatibility of the parabolic shape parameters, the boundary parameters' estimation is inevitably an ill conditioned problem. on the contrary, circular shape models better represent the lane and pavement boundaries in the US highway system since the highways are constructed based on the concentric circular models. We have proposed this never used circular shape models for our boundary detection algorithm and obtained much better conditioned estimation problems than parabolic shape models due to the same unit and same order of magnitude of the circular shape parameters.

In our previous work on pavement boundary detection algorithm using radar images alone, we proposed to employ a log-normal pdf to describe the radar imaging likelihood function [68]. In their work on lane boundary detection using optical images alone, Kluge and Lakshmanan presented an empirical matching function to describe the fidelity of the observation data to certain shape parameters. To take advantage of the previous work, we employ the log-normal pdf and the empirical matching function to describe the radar and optical imaging likelihood functions in our first effort in fusing information from both radar and optical images to simultaneously detect the lane and pavement boundaries. But since the optical imaging matching function is not a normalized pdf, the radar and optical likelihood functions are of different dynamic ranges and different weights must be applied to them. We experimentally choose the weights and in most cases the fusion algorithm yields satisfactory detection results.

Since experimentally choosing the weights on two imaging likelihoods is not only inaccurate but also time consuming, we have proposed a Gaussian pdf to describe the optical imaging process. With the log-normal pdf describing the radar imaging process, we have two normalized pdf's for the imaging likelihood functions and no different weights are necessary in the process of fusion. We have proposed an alternative fusion algorithm where circular shape models play the role of *a priori* information on the lane and pavement boundaries and the Gaussian and log-normal pdf's play the role of imaging likelihood functions. Joint MAP method is also applied to estimate the shape parameters and better detection results are achieved than the first fusion algorithm.

8.2 Future Work

Although circular shape models better describe the boundaries and give more accurate detection results, it has problems in representing straight lane and pavement boundaries. The radii of the concentric circles are infinity for straight boundaries and theoretically, our detection algorithms are unable to handle such situations. In our implementation of the proposed detection algorithms, we use very large radii to approximate the infinite radii and the results are quite satisfactory in our experiments. To improve the detection results for straight boundaries, we can adopt a detection algorithm using two shape models to describe the boundaries. The two shape models are linear and circular models. The boundary detection algorithm operates as follows: First, detect the lane and pavement boundaries with the linear model; Second, with the estimated parameters of the linear model as the initialization, detect the boundaries with the circular shape model. Finally, compare the residual errors of the detection results with linear and circular models, the detection result with the smaller residual error is the optimal detection results that we are looking for.

Another direction for this boundary detection problem is to take into account the different resolutions of the pixels in the optical (radar) image at near and far ranges. It is obvious that the observation data at the near range have higher resolutions than the observation data at the far range. If we apply different and appropriate weights on the data at different distances from the host vehicle so that it can take advantage of the resolution difference, then the detection results should not be misled by some noise points at the far range and yield erroneous detection results.

APPENDIX

APPENDIX

Proof of Convergence Rates of MST Length Functionals for Non-uniform Samples

Before we start deriving the rate of convergence of MST length functionals L of i.i.d. non-uniform random vectors $X_1, X_2, \ldots, X_n$ in $[0,1]^d$, we shall give some properties that are critical in the derivation. Let F be a finite subset of $[0,1]^d$, $d \geq 2$, i.e., F is a set of points. And let W be a real-valued function defined on F.

- *Null condition*: $W(\phi) = 0$, where ϕ is the null set.
- *Subadditivity*: There exists a constant C_1 with the following property: If $\mathcal{Q}^m = \{Q_i\}_{i=1}^{m^d}$ is a uniform partition of $[0,1]^d$ into m^d subcubes Q_i with edge parallel to the axes and with edge length m^{-1} and volume m^{-d} and if $\{q_i\}_{i=1}^{m^d}$ is the set of points in $[0,1]^d$ that translate each Q_i back to the origin such that $Q_i - q_i$ has the form $m^{-1}[0,1]^d$, then for every finite subset F of $[0,1]^d$,
$$W(F) \leq m^{-1} \sum_{i=1}^{m^d} W\left(m[F \cap Q_i - q_i]\right) + C_1 m^{d-1} \tag{A.1}$$
- *Superadditivity*: For the same conditions as above on Q_i, m, and q_i, there exists a constant C_2 with the following property:
$$W(F) \geq m^{-1} \sum_{i=1}^{m^d} W\left(m[F \cap Q_i - q_i]\right) - C_2 m^{d-1} \tag{A.2}$$

- *Continuity*: There exists a constant C_3 such that for all finite subsets F and G of $[0,1]^d$,

$$|W(F \cup G) - W(F)| \leq C_3(\text{card}(G))^{(d-1)/d} \tag{A.3}$$

where card(G) is the cardinality of the subset G.

The functional W is said to be a *continuous subadditive functional* if it satisfies the null condition, sudadditivity and continuity. W is said to be a *continuous superadditive functional* if it satisfies the null condition, superadditivity and continuity.

Many continuous subadditive functionals W on $[0,1]^d$ are naturally related to a *dual* superadditive functional W^*, where $W(F) + 1 \geq W^*(F)$ for every finite subset F and where for i.i.d. uniform random vectors $U_1, \ldots, U_n$ in $[0,1]^d$,

$$|\ E[W(U_1, \ldots, U_n)] - E[W^*(U_1, \ldots, U_n)]\ | \leq C_4 n^{(d-2)/d} \tag{A.4}$$

with C_4 a constant. We point out that the dual W^* is not uniquely defined and is any superadditive Euclidean functional satisfying (A.4). It has been shown that the boundary-rooted version of W, namely, one where points may be connected to the boundary of the unit cube, usually has the requisite property (A.4) of the dual.

A continuous subadditive functional W (and its dual W^*) is said to be a *quasi-additive continuous functional* if W and W^* satisfy the approximation property (A.4)

Redmond and Yukich have proved that MST length functionals are quasi-additive [86].

We shall derive the rate of convergence of the MST length functionals on non-uniform samples with two steps — first consider blockwise constant distributions, then proceed to arbitrary distributions.

A.1 Convergence Rates of MST Length Functionals for Blockwise Constant Distributed Samples

Assume that the underlying density of the sample points $X_1, \ldots, X_n$ is blockwise constant

$$f(x) = \sum_{i=1}^{m^d} \phi_i 1_{Q_i}(x) \tag{A.5}$$

where $\{Q_i\}_{i=1}^{m^d}$ is a uniform partition of the unit cube $[0,1]^d$ with edge length m^{-1}. Furthermore, assume that $\phi_i > 0$ for all i.

Since the MST length functionals enjoy the property of subadditivity, we have

$$\begin{aligned} L(X_1, \ldots, X_n) \quad & \leq m^{-1} \sum_{i=1}^{m^d} L\left(m[\{X_1, \ldots, X_n\} \cap Q_i - q_i]\right) + C_1 m^{d-1} \\ & = m^{-1} \sum_{i=1}^{m^d} L(U_1, \ldots, U_{n_i}) + C_1 m^{d-1} \end{aligned} \tag{A.6}$$

because in each Q_i the samples are drawn from a uniform distribution. Note that n_i is a random variable with the Binomial distribution $B(n, \phi_i m^{-d})$.

Taking expectations on both sides of above inequality,

$$\begin{aligned} EL(X_1, \ldots, X_n) \quad & \leq m^{-1} \sum_{i=1}^{m^d} EL(U_1, \ldots, U_{n_i}) + C_1 m^{d-1} \\ & = m^{-1} \sum_{i=1}^{m^d} E_{n_i}\left[EL(U_1, \ldots, U_{n_i} | n_i)\right] + C_1 m^{d-1} \end{aligned} \tag{A.7}$$

For uniform samples $U_1, \ldots, U_n$ in $[0,1]^d$, Redmond and Yukich proved the following rates of convergence for MST length functionals L,

$$|EL(U_1, \ldots, U_n) - \beta_1 n^{\frac{d-1}{d}}| \leq K_1 n^{\frac{d-2}{d}} \tag{A.8}$$

where $K_1 = C_3 + C_1 \left(\frac{n}{m^d}\right)^{-\frac{d-1}{2d}}$.

Using the result (A.8) and continuity of the MST length functionals, we have

$$\begin{aligned}
&EL(X_1,\ldots,X_n)\\
&\quad\le m^{-1}\sum_{i=1}^{m^d} E_{n_i}\left[\beta_1 n_i^{\frac{d-1}{d}}+K_1 n_i^{\frac{d-2}{d}}\right]+C_1 m^{d-1}\\
&\quad= m^{-1}\sum_{i=1}^{m^d}\left[\beta_1 E\left(n_i^{\frac{d-1}{d}}\right)+K_1 E\left(n_i^{\frac{d-2}{d}}\right)\right]+C_1 m^{d-1}\\
&\quad= m^{-1}\beta_1 n^{\frac{d-1}{d}}\sum_{i=1}^{m^d} E\left(\left(\frac{n_i}{n}\right)^{\frac{d-1}{d}}\right)+\ m^{-1}K_1 n^{\frac{d-2}{d}}\sum_{i=1}^{m^d} E\left(\left(\frac{n_i}{n}\right)^{\frac{d-2}{d}}\right)+C_1 m^{d-1}
\end{aligned}\tag{A.9}$$

For the Binomial distributed random variable n_i, we employ the Taylor's series expansion and obtain

$$\begin{aligned}
\left(\frac{n_i}{n}\right)^{\frac{d-1}{d}} &= \left[\phi_i m^{-d}+\left(\frac{n_i}{n}-\phi_i m^{-d}\right)\right]^{\frac{d-1}{d}}\\
&= \left(\phi_i m^{-d}\right)^{\frac{d-1}{d}}+\frac{d-1}{d}\left(\phi_i m^{-d}\right)^{-\frac{1}{d}}\left(\frac{n_i}{n}-\phi_i m^{-d}\right)\\
&\quad+\frac{1}{2}\frac{d-1}{d}\left(-\frac{1}{d}\right)\left(\phi_i m^{-d}\right)^{-\frac{d+1}{d}}\left(\frac{n_i}{n}-\phi_i m^{-d}\right)^2+o\left(\frac{n_i}{n}-\phi_i m^{-d}\right)^2
\end{aligned}\tag{A.10}$$

and

$$\begin{aligned}
\left(\frac{n_i}{n}\right)^{\frac{d-2}{d}} &= \left[\phi_i m^{-d}+\left(\frac{n_i}{n}-\phi_i m^{-d}\right)\right]^{\frac{d-2}{d}}\\
&= \left(\phi_i m^{-d}\right)^{\frac{d-2}{d}}+\frac{d-2}{d}\left(\phi_i m^{-d}\right)^{-\frac{2}{d}}\left(\frac{n_i}{n}-\phi_i m^{-d}\right)\\
&\quad+\frac{1}{2}\frac{d-2}{d}\left(-\frac{2}{d}\right)\left(\phi_i m^{-d}\right)^{-\frac{d+2}{d}}\left(\frac{n_i}{n}-\phi_i m^{-d}\right)^2+o\left(\frac{n_i}{n}-\phi_i m^{-d}\right)^2
\end{aligned}\tag{A.11}$$

Taking the expectations on both sides of the above two inequalities, we have

$$\begin{aligned}
E\left[\left(\frac{n_i}{n}\right)^{\frac{d-1}{d}}\right] &= \left(\phi_i m^{-d}\right)^{\frac{d-1}{d}}-\frac{d-1}{2d^2}\left(\phi_i m^{-d}\right)^{-\frac{d+1}{d}}E\left[\left(\frac{n_i}{n}-\phi_i m^{-d}\right)^2\right]+o\left(n^{-2}\right)\\
&= \left(\phi_i m^{-d}\right)^{\frac{d-1}{d}}-\frac{d-1}{2nd^2}\left(\phi_i m^{-d}\right)^{-\frac{d+1}{d}}(\phi_i m^{-d})(1-\phi_i m^{-d})+o\left(n^{-2}\right)
\end{aligned}\tag{A.12}$$

and

$$\begin{aligned} E\left[\left(\frac{n_i}{n}\right)^{\frac{d-2}{d}}\right] &= \left(\phi_i m^{-d}\right)^{\frac{d-2}{d}} - \frac{d-2}{d^2}\left(\phi_i m^{-d}\right)^{-\frac{d+2}{d}} E\left[\left(\frac{n_i}{n} - \phi_i m^{-d}\right)^2\right] + o\left(n^{-2}\right) \\ &= \left(\phi_i m^{-d}\right)^{\frac{d-2}{d}} - \frac{d-2}{nd^2}\left(\phi_i m^{-d}\right)^{-\frac{d+2}{d}} (\phi_i m^{-d})(1 - \phi_i m^{-d}) + o\left(n^{-2}\right) \end{aligned} \tag{A.13}$$

Then substituting (A.12) and (A.13) back into (A.9), we have

$$\begin{aligned} &EL(X_1, \ldots, X_n) \\ &\quad \leq m^{-1}\beta_1 n^{\frac{d-1}{d}} \sum_{i=1}^{m^d} E\left(\left(\frac{n_i}{n}\right)^{\frac{d-1}{d}}\right) + m^{-1}K_1 n^{\frac{d-2}{d}} \sum_{i=1}^{m^d} E\left(\left(\frac{n_i}{n}\right)^{\frac{d-2}{d}}\right) + C_1 m^{d-1} \\ &\quad = \beta_1 n^{\frac{d-1}{d}} \sum_{i=1}^{m^d} \phi_i^{\frac{d-1}{d}} m^{-d} - \beta_1 n^{-\frac{1}{d}} \frac{d-1}{2d^2} \sum_{i=1}^{m^d} \phi_i^{-\frac{1}{d}} + \beta_1 n^{-\frac{1}{d}} \frac{d-1}{2d^2} \sum_{i=1}^{m^d} \phi_i^{\frac{d-1}{d}} m^{-d} \\ &\qquad + K_1 n^{\frac{d-2}{d}} \sum_{i=1}^{m^d} \phi_i^{\frac{d-2}{d}} m^{-(d-1)} - K_1 n^{-\frac{2}{d}} \frac{d-2}{d} \sum_{i=1}^{m^d} \phi_i^{-\frac{2}{d}} m \\ &\qquad - K_1 n^{-\frac{2}{d}} \frac{d-2}{d} \sum_{i=1}^{m^d} \phi_i^{\frac{d-2}{d}} m^{-(d-1)} + m^{-1}\beta_1 n^{\frac{d-1}{d}} o\left(n^{-2}\right) + C_1 m^{d-1} \\ &\quad = \beta_1 n^{\frac{d-1}{d}} \int \phi_i^{\frac{d-1}{d}} dx - \beta_1 n^{-\frac{1}{d}} m^d \frac{d-1}{2d^2} \int \phi_i^{-\frac{1}{d}} dx + \beta_1 n^{-\frac{1}{d}} \frac{d-1}{2d^2} \int \phi_i^{\frac{d-1}{d}} dx \\ &\qquad + K_1 n^{\frac{d-2}{d}} m \int \phi_i^{\frac{d-2}{d}} dx - K_1 n^{-\frac{2}{d}} m^{(d-1)} \frac{d-2}{d} \int \phi_i^{-\frac{2}{d}} dx dx \\ &\qquad - K_1 n^{-\frac{2}{d}} m \frac{d-2}{d} \int \phi_i^{\frac{d-2}{d}} + m^{-1}\beta_1 n^{\frac{d-1}{d}} o\left(n^{-2}\right) + C_1 m^{d-1} \end{aligned} \tag{A.14}$$

Changing $n \to nm^d$ and dividing both sides of (A.14) with $(nm^d)^{\frac{d-1}{d}}$, we rewrite (A.14) as

$$\begin{aligned} &\frac{EL(X_1, \ldots, X_{nm^d})}{(nm^d)^{\frac{d-1}{d}}} \\ &\quad \leq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx + \beta_1 n^{-1} m^{-d} \frac{d-1}{2d^2} \int \phi^{\frac{d-1}{d}} dx \\ &\qquad + K_1 n^{-\frac{1}{d}} \int \phi^{\frac{d-2}{d}} dx - K_1 n^{-\frac{d+1}{d}} m^{-2} \frac{d-2}{d} \int \phi^{-\frac{2}{d}} dx \\ &\qquad - K_1 n^{-\frac{d+1}{d}} m^{-d} \frac{d-2}{d} \int \phi^{\frac{d-2}{d}} dx + m^{-1}\beta_1 o\left(n^{-2}m^{-2d}\right) + C_1 n^{-\frac{d-1}{d}} \end{aligned} \tag{A.15}$$

When m is sufficiently large,

$$\frac{EL(X_1,\ldots,X_{nm^d})}{(nm^d)^{\frac{d-1}{d}}} \\ \leq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx + K_1 n^{-\frac{1}{d}} \int \phi^{\frac{d-2}{d}} dx + C_1 n^{-\frac{d-1}{d}} \tag{A.16}$$

For fixed m and changing back $nm^d \to n$,

$$\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} \\ \leq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} m^d \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx + K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + C_1 n^{-\frac{d-1}{d}} m^{d-1} \tag{A.17}$$

Similarly we have the inequality for the dual functional L^*,

$$\frac{EL^*(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} \\ \geq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} m^d \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx + K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + C_2 n^{-\frac{d-1}{d}} m^{d-1} \tag{A.18}$$

By the inequality between the MST length functional and its dual,

$$L(X_1,\ldots,X_n) + 1 \geq L^*(X_1,\ldots,X_n), \tag{A.19}$$

we have

$$\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} \geq \frac{EL^*(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - n^{-\frac{d-1}{d}} \tag{A.20}$$

Therefore, with (A.18) and (A.20) we can obtain

$$\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} \geq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} m^d \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx \\ + K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + C_2 n^{-\frac{d-1}{d}} m^{d-1} - n^{-\frac{d-1}{d}} \tag{A.21}$$

For fixed m, we can define $C_2' = C_2 - m^{-d-1}$ and (A.21) becomes

$$\begin{aligned}\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} \geq \beta_1 \int \phi^{\frac{d-1}{d}} dx - \beta_1 n^{-1} m^d \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx \\ + K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + C_2' n^{-\frac{d-1}{d}} m^{d-1}\end{aligned} \tag{A.22}$$

Define $K_2 = \max(C_1, C_2')$ and from (A.17) and (A.22), we have

$$\begin{aligned}&\left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int \phi^{\frac{d-1}{d}} dx\right| \\ &\quad\leq -\beta_1 n^{-1} m^d \frac{d-1}{2d^2} \int \phi^{-\frac{1}{d}} dx + K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + K_2 n^{-\frac{d-1}{d}} m^{d-1}\end{aligned} \tag{A.23}$$

Define $K_3 = -\beta_1 \frac{d-1}{2d^2}$ and we have

$$\begin{aligned}&\left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int \phi^{\frac{d-1}{d}} dx\right| \\ &\quad\leq K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}} dx + K_2 n^{-\frac{d-1}{d}} m^{d-1} + K_3 n^{-1} m^d \int \phi^{-\frac{1}{d}} dx\end{aligned} \tag{A.24}$$

(A.24) can be rewritten as a function of the number of pixels in each cell,

$$\begin{aligned}&\left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int \phi^{\frac{d-1}{d}} dx\right| \\ &\quad\leq K_1 \left(\frac{1}{n/m^d}\right)^{\frac{1}{d}} \int \phi^{\frac{d-2}{d}} dx + K_2 \left(\frac{1}{n/m^d}\right)^{\frac{d-1}{d}} m^{d-1} + K_3 \left(\frac{1}{n/m^d}\right) \int \phi^{-\frac{1}{d}} dx\end{aligned} \tag{A.25}$$

A.2 Convergence Rates of MST Length Functionals for Arbitrary Continuous Distributed Samples

Hero and Michel have proved the following lemma in [39]. This lemma gives the upper and lower bounds on the difference of the integrals of two density functions.

Lemma 1: For $\nu \in [0,1]$ let f^ν be of bounded variation over $[0,1]^d$ and denote by v its total variation over $[0,1]^d$. Define the resolution $1/m$ block density approximation

$\phi(x) = \sum_{i=1}^{m^d} \phi_i \, 1_{Q_i}(x)$ where $\phi_i = m^d \int_{Q_i} f(x)dx$. Then

$$0 \leq \int_A |\phi^\nu(x) - f^\nu(x)|dx \leq m^{-d} \sum_{i=1}^{m^d} v(Q_i). \tag{A.26}$$

For $\nu = \frac{d-1}{d}$ let us denote the corresponding total variation as v_1 and for $\nu = \frac{d-2}{d}$ let us denote the corresponding total variation as v_2. Also denote by v_0 the total variation of $f^{-\frac{1}{d}}$. Suppose that the underlying density f satisfies the assumptions in Lemma 1. By using Schwartz inequality with (A.24) and (A.26), we have the convergence rate of MST length functionals for arbitrary distributed samples,

$$\begin{aligned}
&\left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int f^{\frac{d-1}{d}}(x)dx\right| \\
&\quad= \left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int \phi^{\frac{d-1}{d}}(x)dx + \beta_1 \int \phi^{\frac{d-1}{d}}(x)dx - \beta_1 \int f^{\frac{d-1}{d}}(x)dx\right| \\
&\quad\leq \left|\frac{EL(X_1,\ldots,X_n)}{n^{\frac{d-1}{d}}} - \beta_1 \int \phi^{\frac{d-1}{d}}(x)dx\right| + \beta_1 \left|\int \phi^{\frac{d-1}{d}}(x)dx - \int f^{\frac{d-1}{d}}(x)dx\right| \\
&\quad\leq K_1 n^{-\frac{1}{d}} m \int \phi^{\frac{d-2}{d}}(x)dx + K_2 n^{-\frac{d-1}{d}} m^{d-1} \\
&\qquad\qquad + K_3 n^{-1} m^d \int \phi^{-\frac{1}{d}}dx + \beta_1 m^{-d} \sum_{i=1}^{m^d} v_1(Q_i) \\
&\quad\leq K_1 n^{-\frac{1}{d}} m \left[\int f^{\frac{d-2}{d}}(x)dx + m^{-d} \sum_{i=1}^{m^d} v_2(Q_i)\right] + K_2 n^{-\frac{d-1}{d}} m^{d-1} \\
&\qquad\qquad + K_3 n^{-1} m^d \left[\int f^{-\frac{1}{d}}(x)dx + m^{-d} \sum_{i=1}^{m^d} v_0(Q_i)\right] + \beta_1 m^{-d} \sum_{i=1}^{m^d} v_1(Q_i) \\
&\quad\leq K_1 n^{-\frac{1}{d}} m \int f^{\frac{d-2}{d}}(x)dx + K_2 n^{-\frac{d-1}{d}} m^{d-1} + K_3 n^{-1} m^d \int f^{-\frac{1}{d}}(x)dx \\
&\qquad\qquad + \beta_1 \left(m^{-d} \sum_{i=1}^{m^d} v_1(Q_i)\right) + K_1 n^{-\frac{1}{d}} m \left(m^{-d} \sum_{i=1}^{m^d} v_2(Q_i)\right) \\
&\qquad\qquad + K_3 n^{-1} m^d \left(m^{-d} \sum_{i=1}^{m^d} v_0(Q_i)\right)
\end{aligned} \tag{A.27}$$

■

In the above expression, K_1 is related to subadditivity and continuity of the MST length functional, K_2 is related to subadditivity of the MST length functional and superadditivity of the dual of the MST length functional, and K_3 is related to the asymptotic constant of the MST length functional.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. A. Abidi and R. C. Gonzalez, *Data Fusion in Robotics and Machine Intelligence*, Academic Press, 1992.

[2] J. Aitchison and J. A. C. Brown, *The Log-normal Distribution*, Cambridge University Press, 1957.

[3] P. Ajjimarangsee, “Neural network model for fusion of visible and infrared sensor outputs,” in *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 153–160, 1988.

[4] P. K. Allen, *Robotic object recognition using vision and touch*, Kluwer Academic Publishers, 1987.

[5] D. Banks, M. lavibe, and H. J. Newton, “The minimal spanning tree for nonparametric regression and structure discovery,” in *Proceedings of the 24th Symposium on the Interface*, pp. 370–374, 1992.

[6] M. Basseville, “Distance measures for signal processing and pattern recognition,” *Signal Processing*, vol. 18, pp. 349–369, 1989.

[7] R. E. Bethel and G. J. Paras, “PDF multisensor multitarget tracker,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 153–168, 1998.

[8] B. Bhanu, “Automatic target recognition:state of the art survey,” *IEEE Trans. Aerosp. Elect. Syst.*, vol. 22, no. 4, pp. 364–376, 1986.

[9] M. Bhattacharya and D. Majumder, “Registration of CT and MR images of alzheimer’s patient: A shape theoretic approach,” *Pattern Recognition Letters*, vol. 21, no. 6, pp. 531–548, 2000.

[10] D. C. Borghys, P. S. Verlinde, C. Perneel, and M. Acheroy, “Long-range target detection in a cluttered environment using multisensor image sequences,” in *Proc. of SPIE*, volume 3068, pp. 569–578, 1997.

[11] W. M. Brown and C. W. Swonger, “A prospectus for automatic target recognition,” *IEEE Trans. Aerosp. Elect. Syst.*, vol. 25, no. 3, pp. 401–410, 1989.

[12] P. Cheeseman, “In defense of probability,” in *Proceedings of the International Conference on Artificial Intelligence*, pp. 1002–1009, 1985.

[13] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156–1168, Dec 1994.

[14] S. Chen and Y. W. Jen, "Data fusion neural network for tool condition monitoring in CNC milling machining," *International Journal of Machine Tools and Manufacture*, vol. 40, no. 3, pp. 381–400, 2000.

[15] N. Chowdhury and C. A. Murthy, "Minimal spanning tree based clustering technique: relationship with Bayes classifier," *Pattern Recognition*, vol. 30, no. 11, pp. 1919–1929, Nov 1997.

[16] E. Coiras, J. Santamaria, and C. Miravet, "Segment-based registration technique for visual-infrared images," *Optical Engineering*, vol. 39, no. 1, pp. 282–289, 2000.

[17] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multimodality medical image registration using information theory," in *Proceedings of the XIVth International Conference on Information Processing in Medical Imaging*, volume 3, pp. 263–274, Ile de Derder, France, June 1995, Kluwer Academic Publishers.

[18] X. Dai and S. Khorram, "Feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2351–2362, 1999.

[19] B. V. Dasarathy, "Sensor fusion potential exploitation - innovative architectures and illustrative applications," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 24–38, 1997.

[20] A. R. de Saint Vincent, "A 3-D perception system for the mobile robot hilare," in *IEEE Int. Conf. Robotics and Automation*, pp. 1105–1111, San Francisco, 1986.

[21] A. Dempster, "Upper and lower probabilities induced by a multi-valued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325–339, 1967.

[22] E. D. Dickmanns and B. D. Mysliwetz, "Recursive 3-D road and relative ego-state recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 199–213, Feb. 1992.

[23] E. D. Dickmanns, S. Werner, S. Kraus, and R. Schell, "Experimental results in autonomous landing approaches by dynamic machine vision," in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 2220, pp. 304–313, Orlando, FL, USA, 1994.

[24] O. E. Drummond, "Methodologies for performance evaluation of multitarget multisensor tracking," in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3809, pp. 355–369, Denver, CO, USA, 1999.

[25] A. El-Fallah, R. Mahler, B. Ravichandran, and R. Mehra, “Adaptive data fusion using finite-set statistics,” in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3720, pp. 80–91, 1999.

[26] M. D. Elstrom and P. W. Smith, “Stereo-based registration of multi-sensor imagery for enhanced visualization of remote environments,” in *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, volume 3, pp. 1948–1953, Detroit, MI, USA, 1999.

[27] A. Farina, A. Russo, and F. A. Studer, “Coherent radar detection in log-normal clutter,” *IEE Proc., Part F: Communications, Radar and Signal Processing*, vol. 133, no. 1, pp. 39–54, Feb 1986.

[28] G. M. Flachs, J. B. Jordan, C. L. Beer, D. R. Scott, and J. J. Carlson, “Feature space mapping for sensor fusion,” *Journal of Robotic Systems*, vol. 7, no. 3, pp. 373–393, 1990.

[29] J. H. Friedman and L. C. Rafsky, “Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests,” *The Annals of Statistics*, pp. 697–717, 1979.

[30] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1991.

[31] C. R. Giardina and E. R. Dougherty, *Morphological methods in image and signal processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

[32] B. Haack and M. Bechdol, “Multisensor remote sensing data for land use/cover mapping,” *Computers, Environment and Urban Systems*, vol. 23, no. 1, pp. 53–69, 1999.

[33] D. L. Hall and R. J. Linn, “Survey of commercial software for multisensor data fusion,” in *Proc. SPIE Conf. Sensor Fusion and Aerospace Applications*, Orlando, Apr. 1991.

[34] R. C. Harney, “Sensor fusion for target recognition: a review of fundamentals and a potential approach to multisensor requirement allocation,” in *Proc. of SPIE*, volume 2269, pp. 316–335, 1994.

[35] D. Haverkamp and C. Tsatsoulis, “Information fusion for estimation of summer MIZ ice concentration from SAR imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1278–1291, 1999.

[36] A. Hernandez, O. Basset, I. Magnin, A. Bremond, and G. Gimenez, “Fusion of ultrasonic and radiographic images of the breast,” in *Proc. IEEE Ultrasonics Symposium*, pp. 1437–1440, San Antonio, TX, USA, 1996.

[37] A. I. Hernandez, G. Carrault, F. Mora, L. Thoraval, G. Passariello, and J. M. Schleich, "Multisensor fusion for atrial and ventricular activity detection in coronary care monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 10, pp. 1186–1190, 1999.

[38] A. Hero and O. Michel, "Robust entropy estimation stratigies based on edge weighted random graphs," in *1998 SPIE International Symposium on Optical Science, Engineering and Instrumentation*, San Diego, July 1998.

[39] A. Hero and O. Michel, "Asymptotic theory of greedy approximations to minimal k-point random graphs," *IEEE Trans. on Information Theory*, vol. IT-45, pp. 1921–1939, Sept 1999.

[40] A. Hero and O. Michel, "Estimation of Rényi information divergence via pruned minimal spanning trees," in *1999 IEEE Workshop on Higher Order Statistics*, Caesaria, Israel, 1999.

[41] R. Hoffman and A. K. Jain, "A test of randomness based on the minimal spanning tree," *Pattern Recognition Letters*, pp. 175–180, 1983.

[42] D. P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment with an image," *Int. J. Comput. Vis.*, vol. 5, pp. 195–212, 1990.

[43] J. Igbokwe, "Geometrical processing of multi-sensoral multi-temporal satellite images for change detection studies," *International Journal of Remote Sensing*, vol. 20, no. 6, pp. 1141–1148, 1999.

[44] A. Jain and J. W. Mamer, "Approximations for the random minimal spanning tree with application to network provisioning," *Operations Research*, vol. 36, no. 4, pp. 575–584, 1988.

[45] A. E. Johnson and S. B. Kang, "Registration and integration of textured 3D data," *Image and Vision Computing*, vol. 17, no. 2, pp. 135–147, 1999.

[46] E. Jouseau and B. Dorizzi, "Neural networks and fuzzy data fusion – application to an on-line and real time vehicle detection system," *Pattern Recognition Letters*, vol. 20, no. 1, pp. 97–107, 1999.

[47] A. C. Kak, *Principles of computerized tomographic imaging*, New York : IEEE Press, 1988.

[48] K. Kaliyaperumal, S. Lakshmanan, and K. Kluge, "Detecting roads and obstacles in radar images," *to appear in IEEE Transactions on Vehicular Technology*, 2000.

[49] W. Kasprzak and H. Niemann, "Adaptive road recognition and ego-state tracking in the presence of obstacles," *International Journal of Computer Vision*, vol. 28, no. 1, pp. 5–26, June 1998.

[50] S. K. Kenue, "LANELOK: Detection of lane boundaries and vehicle tracking using image-processing techniques, parts I and II," in *SPIE Mobile Robots IV*, 1989.

[51] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[52] K. Kluge and C. Thorpe, "The YARF system for vision-based road following," *Mathematical and Computer Modelling*, vol. 22, no. 4-7, pp. 213–234, Aug - Oct 1995.

[53] K. C. Kluge, *YARF: An open-ended framework for robot road following*, PhD thesis, Carnegie Mellon University, 1993.

[54] K. C. Kluge, "Extracting road curvature and orientation from image edge points without perceptual grouping into features," in *Proc. Intell. Vehic. '94 Symp.*, pp. 109–114, Sept. 1994.

[55] K. C. Kluge, "Performance evaluation of vision-based lane sensing: some preliminary tools, metrics and results," in *IEEE Conference on Intelligent Transportation Systems*, 1997.

[56] K. C. Kluge and S. Lakshmanan, "A deformable-template approach to lane detection," in *Proc. Intell. Vehic. '95 Symp.*, Sept. 1995.

[57] D. J. Kriegman, T. E., and T. O. Binford, "Stereo vision and navigation in buildings for mobile robots," *IEEE Trans. Robot. Automat.*, vol. 5, no. 6, pp. 792–803, 1989.

[58] S. Kullback, *Information theory and statistics*, New York, Dover Publications, 1968.

[59] S. Lakshmanan and D. Grimmer, "Deformable template approach to detecting straight edges in radar images," *IEEE Trans. on PAMI*, vol. 18, no. 4, pp. 438–443, Apr. 1996.

[60] S. Lakshmanan, A. K. Jain, and Y. Zhong, "Detecting straight edges in millimeter-wave images," in *IEEE Inter. Conf. Image Processing*, volume 1, 1996.

[61] J. Landa and K. Scheff, "Binocular fusion using simulated annealing," in *Proc. IEEE 1st Int. Conf. Neural Networks*, pp. 327–334, San Diego, CA, USA, 1987.

[62] J. S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images.," *Comput Graphics Image Process*, vol. 17, no. 1, pp. 24–32, Sept 1981.

[63] J. S. J. Lee, "Multiple sensor fusion based on morphological processing," in *Proceedings of SPIE - The International Society for Optical Engineering*, pp. 94–100, 1988.

[64] R. H. Lee and R. Leahy, "Multi-spectral tissue classification of MR images using sensor fusion approaches," in *Proceedings of the SPIE: Medical Imaging IV: Image Processing*, volume 1233, pp. 149–157, 1990.

[65] C. Lin, S. Amphay, and B. M. Sundstrom, “Sensor fusion with passive millimeter wave and laser radar for target detection,” in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3703, pp. 57–67, 1999.

[66] Y. Linde, A. Buzo, and R. M. Gray, “An algorithm for vector quantization design,” *IEEE Transactions on Communications*, vol. COM-28, pp. 84–95, Jan. 1980.

[67] R. C. Luo and M. G. Kay, “Data fusion and sensor integration: State-of-the-art 1990’s,” in *Data Fusion in Robotics and Machine Intelligence*, M. A. Abidi and R. C. Gonzalez, editors, pp. 7–135, Academic Press, 1992.

[68] B. Ma, S. Lakshmanan, and A. Hero, “Detection of curved road edges in radar images via deformable templates,” in *Proc. IEEE. Intl. Conf. Image Proc.*, 1997.

[69] B. Ma, S. Lakshmanan, and A. Hero, “Road and lane edge detection with multisensor fusion methods,” in *Proc. IEEE. Intl. Conf. Image Proc.*, pp. 686–690, 1999.

[70] B. Ma, S. Lakshmanan, and A. Hero, “Pavement boundary detection via circular shape models,” in *Proc. IEEE. Intelligent Vehicles Symposium*, 2000.

[71] B. Ma, S. Lakshmanan, and A. Hero, “Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion,” *IEEE Trans. Intelligent Transportation Systems*, vol. 1, no. 3, pp. 135–147, Sept. 2000.

[72] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, Apr 1997.

[73] F. Matia and A. Jimenez, “Multisensor fusion: an autonomous mobile robot,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 22, no. 2, pp. 129–141, 1998.

[74] M. McGuire and H. S. Stone, “Techniques for multiresolution image registration in the presence of occlusions,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1476–1479, 2000.

[75] E. Mizutani, T. Kozek, and L. O. Chua, “Road lane marker extraction by motion-detector CNNs,” in *Proc. IEEE Intl Joint Conf. on Neural Networks*, pp. 503–508, 1998.

[76] M. Moghaddam, J. Dungan, and J. Coughlan, “Fusion of AIRSAR and TM data for parameter classification and estimation in dense and hilly forests,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pp. 305–307, Hamburg, Germany, 1999.

[77] R. R. Murphy, “Dempster-Shafer theory for sensor fusion in autonomous mobile robots,” *EEE Transactions on Robotics and Automation*, vol. 14, no. 2, pp. 197–206, Apr 1999.

[78] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, “Multiresolution-based image fusion with additive wavelet decomposition,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1204–1211, 1999.

[79] T. O’Donnell, S. Aharon, S. S. Halliburton, A. Gupta, G. Funka-Lea, and R. D. White, “Multi-modality model-based registration in the cardiac domain,” in *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 790–791, Hilton Head Island, SC, USA, 2000.

[80] N. Pagoulatos, D. Haynor, and Y. Kim, “Image-based registration of ultrasound and magnetic resonance images: A preliminary study,” in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3976, pp. 156–164, San Diego, CA, USA, 2000.

[81] N. R. Pal and J. Biswas, “Cluster validation using graph theoretic concepts,” *Pattern Recognition*, vol. 30, no. 6, pp. 847–857, Jun 1997.

[82] S. B. Pollard, T. P. Pridmore, J. Porrill, J. E. W. Mayhew, and J. P. Frisby, “Geometrical modeling from multiple stereo views,” *International Journal of Robotics Research*, vol. 8, no. 4, pp. 3–32, 1989.

[83] D. Pomerleau and T. Jochem, “Rapidly adapting machine vision for automated vehicle steering,” *IEEE Expert*, vol. 11, no. 2, pp. 19–27, 1996.

[84] R. C. Prim, “Shortest connection networks and some generalizations,” *Bell Syst. Tech. Journ.*, pp. 1389–1401, 1957.

[85] R. Ravi, R. Sundaram, M. V. Marathe, D. J. Rosenkrantz, and S. S. Ravi, “Spanning trees – short or small,” *SIAM Journal on Discrete Mathematics*, vol. 9, no. 2, , 1996.

[86] J. E. Redmond, C.and Yukich, “Limit theorems and rates of convergence for Euclidean functionals,” *The Annals of Applied Probability*, vol. 4, no. 4, pp. 1057–1073, 1994.

[87] A. Renyi, “On measures of entropy and information,” in *Proc. 4th Berkeley Symp.Math. Stat. and Prob.*, volume 1, pp. 547–561, 1961.

[88] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, 1976.

[89] M. L. Slolnik, *Introduction to radar systems*, McGraw-Hill, 1980.

[90] P. J. Slomka, J. Mandel, A. Fenster, and D. Downey, “Automated 3D registration of magnetic resonance angiography, 3D power Doppler, and 3D B-mode ultrasound images of carotid bifurcation,” in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3979, pp. 332–341, San Diego, CA, USA, 2000.

[91] T. M. Smith and P. A. Kelly, “Random sets technique for information fusion applied to estimation of brain functional images,” in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3661, pp. 1158–1169, 1999.

[92] D. A. Socolinsky and L. B. Wolff, "Image fusion for enhanced visualization of brain imaging," in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3658, pp. 352–362, 1999.

[93] B. Solaiman, R. Debon, F. Pipelier, J.-M. Cauvin, and C. Roux, "Information fusion: application to data and model fusion for ultrasound image segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 10, pp. 1171–1175, 1999.

[94] C. Sossai, P. Bison, G. Chemello, and G. Trainito, "Sensor fusion for localization using possibility theory," *Control Engineering Practice*, vol. 7, no. 6, pp. 773–782, 1999.

[95] M. Soumekh, "Signal subspace fusion of uncalibrated sensors with application in SAR and diagnostic medicine," *IEEE Transactions on Image Processing*, vol. 8, no. 1, pp. 127–137, 1999.

[96] J. M. Steele, "Growth rate of Euclidean minimal spanning trees with power weighted edges," *Ann. Probab.*, pp. 1767–1787, 1988.

[97] D. N. Strenki and S. Kirkpatrick, "Analysis of finite-length annealing schedules," *Algorithmica*, pp. 346–366, 1991.

[98] B. T. Sweet and C. L. Tiana, "Image processing and fusion for landing guidance," in *Proc. of SPIE*, volume 2736, pp. 84–95, 1996.

[99] O. Thepaut, K. Kpalma, and J. Ronsin, "Automatic registration of ERS and SPOT multisensor images in a data fusion context," *Forest Ecology and Management*, vol. 128, no. 1, pp. 93–100, 2000.

[100] P. Valin, A. Jouan, and E. Bosse, "Fusion of imaging and non-imaging sensor information for airborne surveillance," in *Proc. Sensor Fusion: Architectures, Algorithms, and Applications*, volume 3719, pp. 126–138, Orlando, FL, USA, 1999.

[101] H. Van Trees, *Detection, estimation, and modulation theory*, Wiley, New York, 1968-71.

[102] P. Verlinde, P. Druyts, G. Chollet, and M. Acheroy, "Multi-level data fusion approach for gradually upgrading the performances of identity verification systems," in *Proceedings of SPIE - The International Society for Optical Engineering*, volume 3719, pp. 14–25, 1999.

[103] P. Viola and W. Wells, "Alignment by maximization of mutual information," in *Proceedings of the 5th International Conference on Computer Vision*, pp. 16–23, Cambridge, MA, USA, June 1995.

[104] P. Viola and W. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, Sep 1997.

[105] F. Wright, “The fusion of multi-source data,” *Signal*, pp. 39–43, Oct. 1980.

[106] M. Yelverton, B. Cusson, T. Timmons, and K. Stoddard, “Using automatic fault detection to improve diffusion furnace performance and reduce wafer scrap,” *MICRO*, vol. 17, no. 4, pp. 27–33, 1999.

[107] C. Zahn, “Graph-theoretical method for detecting and describing gestalt clusters,” *IEEE Trans. on Computers*, pp. 68–86, 1971.

[108] J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul, MN, 1992.