RECOGNITION OF OCCLUDED OBJECT USING WAVELETS

TIE HUA DU

NATIONAL UNIVERSITY OF SINGAPORE

2006

RECOGNITION OF OCCLUDED OBJECT USING WAVELETS

TIE HUA DU
(B.Eng., M. Sc.)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2006

Acknowledgements

This thesis and the research presented in this thesis were made possible by the support and guidance of many people. Without them, the completion of this work would not have been possible.

First and foremost, I would like to thank my supervisors, A/Prof. Kah Bin Lim and A/Prof. Geok Soon Hong who have provided me with a comprehensive vision of research, strong technical guidance, and valuable feedback on my research. They have given me confidence in my abilities and have also provided me with the freedom to pursue those areas in pattern recognition of particular interest to me during my Ph.D period.

I take this opportunity to express my sincere appreciation to Prof. ZuoWei Shen from the Mathematics Department, National University of Singapore, who has guided me to wavelet world. He has a very sharp mind in wavelet theory and its applications. My appreciation also goes to Dr. SuQi Pan who has helped me a lot to clear my doubts in wavelet and other problems in mathematics.

I would like to thank several colleagues who have provided me with both helpful comments and great friendship during the past three years. Particularly I would like to thank Mr. YingHe Chen, Mr. WeiMiao Yu and Mr. Hao Zheng.

I would also like to thank the members of the doctoral thesis committee and oral defense committee.

I wish also to thank National University of Singapore for awarding me the research scholarship and the Department of Mechanical Engineering for the use of facilities.

Last but not least, I wish to express my deep appreciation to my dear parents and parents in laws for their continuous support and affection all along my life. I feel indebted to their encouragement and moral support during the past years, and I owe them a lot of gratitude. I am especially indebted to my loving wife Yong Liu, for her care and understanding, patience, encouragement and everything she gives to me. And finally I would like to dedicate this thesis to my lovely son Chuang Du and Yi Du.

TABLE OF CONTENTS

Summary

Object recognition has extensive applications in many areas, such as visual inspection, part assembly, artificial intelligence, etc. It is a major and also a challenging task in computer vision. Although humans perform object recognition effortlessly and instantaneously, implementation of this task on machines is very difficult. The problem is even more complicated when there is partial occlusion situation. Many researchers have dedicated themselves into this area and made great contributions in the past few decades. However, existing algorithms have various shortcomings and limitations, such as their limited applicability to the polygonal shapes, and the necessary prior knowledge of the scale.

This research is aimed at developing a novel 2-D object recognition algorithm applicable for both stand-alone and partial occluded objects using wavelet techniques. Wavelet is a more recent mathematical tool in comparison with Fourier transform, and it has several exciting properties which can be well used in this research, e.g. multiresolution analysis, singularity detection and local analysis. A wavelet-based object recognition algorithm is presented in this thesis. The feature to represent the object is the wavelet representation of curve segments of the object boundary. To achieve the consistent boundary partitioning, a wavelet-based corner detection algorithm is proposed and verified. After partitioning, each curve segment is normalized, which makes it invariant to similarity transformation. An adaptive fast wavelets decomposition using bi-orthonormal wavelet is then applied on each segment to extract multiresolution representation, which facilitates hierarchical

matching. After thresholding to eliminate the noise and quantization error, the resultant scaling coefficients and wavelet coefficients are the features for recognition. In matching process, firstly, we match the features of segments between object in the scene and the model in an object database to find out segment-pair candidates with similar geometric shape. Hierarchical matching strategy is adopted to accelerate the matching speed. If valid segment-pairs between object in scene and model are found, relative orientation and scale information are then applied for further verification to eliminate false matching. Experiment results show that our proposed recognition algorithm is invariant to similarity transform, robust to partial occlusion, and that it is computationally efficient.

List of Tables

List of Illustrations

Chapter 1

Introduction

1.1 Background

An object recognition system finds objects in the real world from an image of the world, using object models which are known *a priori*. Object recognition has extensive applications in many areas, such as visual inspection, part assembly, artificial intelligence, etc. Although humans perform object recognition effortlessly and instantaneously, implementation of this task on machines is very difficult. It is a major and also a challenging task in computer vision. Many researchers have dedicated themselves into this area and made great contributions in the past few decades.

The object recognition problem can be defined as a labeling problem based on models of known objects. Stated formally, given an image containing one or more objects of interest and a set of labels corresponding to a set of models known to system, the system should assign correct labels to the regions, or a set of regions, in the image.

In this research project, we restrict ourselves to two-dimensional object recognition. It is assumed that all the real world objects are viewed by a camera directly located on top of them, so that the height variation can be neglected for an arbitrary orientation and position of the objects. This simplification is reasonable and the 2-D recognition is indeed important in many image analysis applications, and is widely applied to many fields.

An object is defined by its photometric and geometric features. Those methods which solely depend on photometric features may fail to identify object properly, since photometric features vary with circumstances such as illumination and environmental condition. In comparison, geometric features tend to be much more useful then photometric features in pattern recognition. The boundary of an object is one of the most important geometric features. Contour-based approaches are more popular than region-based approaches in literature. This is because human beings are thought to discriminate shapes mainly by their contour features. Another reason is because in many applications where recognition is based on shape, the contour is the only interest, whilst the content of the interior of the shape is not important. Moreover, contour-based approaches generally need less computational effort than region-based approaches. In this research project, the feature we used is also contour-based.

1.2 Recognition Process

Given an image containing several objects, the pattern recognition process consists of three major phases as shown in Fig.1.1 The first phase is called *image isolation*, in which each object is found and its image is isolated from the rest of the scene. The second phase is called *feature extraction*. This is where the objects are measured. A measurement is the value of some quantifiable property of an object. A feature is a function of one or more measurements, computed so that it quantifies some significant characteristic of the object. The feature extraction process produces a set of features that, taken together comprise the feature vector. This drastically reduces the amount of information necessary to represent all the knowledge upon which the subsequent classification decisions must be based. It is productive to conceptualize an *n*-dimensional space in which all possible n-element feature vectors reside. Thus, any particular object corresponds to a point in feature space. Feature extraction is the crucial phase for pattern recognition, the features extracted should be effective and the feature extraction process should be efficient. The third phase of pattern recognition is *classification*. Its output is merely a decision regarding the class to which each object belongs.

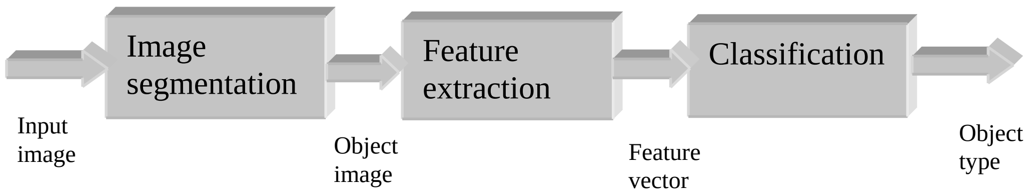


Fig. 1.1 The three phases of pattern recognition

Object recognition is not a single process, but a close combination of many image processing techniques, such as low level process (e.g. denoising, image enhancement and etc.), mid level process (e.g. segmentation and feature extraction) and high level process (e.g. feature mapping). In order to develop a successful object recognition system, each process needs to be specially designed to co-operate with the preceding process and subsequent process without flaw.

1.3 Problem Statement and Research Objective

Most recognition systems expect precise and complete information, which restrict their scope to simple application. In practice, one has to allow flexibility in the form of noisy scenes and partially occluded objects in different scales and in randomly oriented positions.

The object being recognized may be different from the model object in database in size, position and orientation (as shown in Fig. 1.2). We call these variations (scaling, translation and rotation) similarity transformation. Recognition of two dimensional objects regardless of these transformations is an important problem in pattern recognition. Therefore, the invariance of object representation to similarity transformation is an essential requirement.

Fig. 1.2 Object under similarity transformation
(a) A pliers (b) a pliers with similarity transformation

The recognition of individual objects with complete shapes regardless of similarity transformation has been studied for a long time, and can be handled without much difficulty with many existing techniques. Problems arise when the object is occluded. The occlusion takes place when an object is either overlapped or touched by another object (as shown in Fig. 1.3 (a)). This problem has significant importance in an industrial environment. Supposing that parts are moving on a conveyor belt for visual inspection, when parts touch or overlap each other, the vision system should be able to recognize correctly each of the occluded objects rather than to reject them as a single unidentifiable part. A similar situation arises when a robot tries to pick up a particular part from a bin in which different part types are jumbled together. Besides overlapping, when an object is not fully covered in an image or some portion of the object can not be seen due to some major defects of the image (as shown in Fig. 1.3 (b)), we categorize these situations as partial occlusion. The complexity and difficulty of object recognition induced by partial occlusion increase tremendously. The problem of recognizing partially occluded objects is considered as one of the most difficult problems in machine vision. Researchers have developed some algorithms using local features to deal with this problem, some progresses have been made and reported (as reviewed in Chapter 2), however, these works have their limitations and drawbacks in one way or another. The problem of recognizing partially occluded objects is still an open issue till date.

(a) (b)

Fig. 1.3 Object with partial occlusion

(a) A pliers is overlapped with a screwdriver (b) A pliers which two handles can not be seen

1.4 Object Representation- Criteria of Shape Descriptor

Object representation is the key issue of pattern recognition. A robust and effective object representation algorithm generally leads to a successful object recognition system. Object representation generally looks for effective and perceptually important shape features based on either object boundary information or from the object region. A thorough literature review of 2-D object representation techniques has been done by Tsang (2001), the pros and cons of each technique have also been discussed. Based on the extensive literature survey on object representation techniques done by many researchers and us, we shall conclude that: For general recognition purpose, a good shape descriptor should meet the following criteria:

a) Invariance under similarity transformations

A recognition system should be able to effectively find perceptually similar shapes from a database. A perceptually similar shape usually means rotated, translated and scaled shapes. Therefore, the shape descriptor must be essentially invariant under translation, rotation and scaling, which collectively are called Similarity Transform.

b) Stability

The shape descriptor should also be able to find noise corrupted shapes, distorted shapes and defective shapes, which are tolerated by human being when comparing shapes. This is also known as the robustness requirement.

c) Compactness

As shown by Karp (1972), the time used to match the shape descriptor of a scene object to a model may increase significantly with the number of features. Therefore, the size of shape descriptor must be as few as possible in order to make matching process easy and fast. Compact shape descriptors are highly desirable for indexing and online retrieval.

d) Completeness

The shape descriptor must contain characteristic information of the object shape as complete as possible. Only when the shape descriptor can describe adequately the object shape completely, can we then eliminate the ambiguity which may be encountered when we try to match the object in the scene to the model.

e) Hierarchical Representation

If a shape descriptor has a hierarchical coarse to fine representation characteristics, it can achieve a high level of matching efficiency. This is because shapes can be matched at coarse level to first eliminate large amount of dissimilar shapes, and at finer level, shape can be matched in details.

f) Generalization

A desirable shape descriptor should be application independent rather than only performing well for certain type of objects.

g) Efficiency

Low computational complexity is an important characteristic of a desirable shape descriptor. For a shape descriptor, low computational complexity means minimizing any uncertain or ad hoc factors that are involved in the derivation processes. The fewer the uncertain factors involved in the computation processes, the more robust the shape descriptor becomes. In essence, low computation complexity means clarity and stability.

h) Uniqueness

For two objects with different shapes, they should have distinctive different representation.

We set the above criteria as the benchmark to evaluate the object representation algorithms reviewed in the next chapter. We will also use it to examine the object representation presented in this thesis.

1.5 Local Features Vs Global Features

According to whether the object representation is based on the whole object or based on a small section/region, object representations can be largely classified into two types, global feature based and local feature based.

Global features are usually some characteristics of regions in images such as area, perimeter, moments, Fourier descriptors, Hough transformation, etc. They can be obtained either for a region by considering all points within a region or only for those points on the boundary of a region. The advantages of global-feature-based approaches are: the features are easier to determine and the number of features used for recognition is usually small, and the matching process is fast. However, one major setback of this approach is that it requires the objects being recognized to be wholly visible, non-overlapping, and not touching each other. Most pattern recognition algorithms developed for standalone object recognition do not work when partial occlusion takes place. The reason is that these algorithms are designed based on global features, which become completely useless when partial occlusion takes place.

On the other hand, local features are usually on the boundary of an object or represent a distinguishable small area of a region. Some commonly used local features are curvatures, boundary segments, and corners. Recognition approach using local features offers the advantage that if some of the descriptions are corrupted due to noise or occlusion, the remaining information may still be adequate for concluding the object identity, because the characteristics of the visible parts or intact portions of the object can also be obtained and used in the matching process.

Therefore, for this research project, in order to recognize partial occluded objects, the object representation must not only meet the criteria mentioned in the preceding section (Section 1.4), but must also be based on local features.

1.6 Motivation

Recognition of shapes which are incomplete or distorted is important in many image analysis applications. This is especially true in situations where ideal imaging conditions cannot be maintained. This problem has been studied by many researchers for two decades, but have not been resolved entirely yet. Existing techniques also have their limitations in many aspects. A thorough literature survey of related works is shown in Chapter Two.

Most of the existing 2-D object recognition systems use object representations in spatial domain. Generally, object representations in spatial domain suffer from two main drawbacks: sensitivity to noise and high dimensionality (Tsang, 2001). Therefore, object recognition algorithms based on spatial domain features have limited success in recognition performance. The problems can be solved in the following ways: histogram, moments, scale space, spectral transforms etc. Although histogram and scale space methods increase robustness to noise and compactness, matching using these methods can be very computationally expensive. Moment is robust and compact, however, higher order moments are either difficult to obtain or without physical meaning. Among the four solutions, spectral transform is the most promising.

In spectral transform, Fourier transform is the most dominant frequency analysis tool in the past two centuries used to extract object features (Gorman and Mitchell, 1988). Shape representation using Fourier descriptor is simple to compute, robust and compact. Wavelet transform is another spectral transform. It is a relatively recent development in applied mathematics in 1980s. But unlike Fourier transform that uses global sinusoids as the basis function, the wavelet transform is more efficient in representing and detecting local features of a curve due to the spatial and frequency localization property of wavelet bases. Moreover, wavelet transform can readily represent signal in multiple resolution compactly and efficiently. These properties possessed by wavelet motivated us to use wavelet transform technique to tackle partial occluded object recognition problem. The wavelet theory has reached a mature stage over the past few decades. It is a versatile tool with very rich mathematical content and wide applications. It has been employed in many fields and applications with great success, such as signal processing, data compression, image analysis, communication systems, biomedical imaging, radar, air acoustics, theoretical mathematics, control system. We therefore observe that wavelet has several promising properties that make it suitable to solve this occlusion problem, such as: singularity detection, multiresolution representation, noise insensitivity and computational efficiency. Many researchers have tried to solve object recognition problems using wavelet technique, and many contributions have been reported (Chuang et al. 1996, Tieng et al. 1997, Antoine et al. 1997, Yoon et al. 1998, Bui et al. 1999, Khalil et al. 2000, Yu et al. 2001, Tsang 2001, Khalil et al. 2002), and showed they outperformed traditional methods. These works have shown that wavelet is a promising tool for object recognition. However, research on applying the wavelets to the recognition of occluded objects is still lacking, and hence very few publications on partial occluded object recognition problem using wavelet can be found in the existing literature. Among the reported methods, many of them either make assumptions to simplify the problem or have limitations and drawbacks in some other aspects.

Nevertheless, the promising nature of the wavelet technique inspired us to employ it to solve problem on two dimensional partial-occluded object recognition.

1.7 Objectives

The objective of our research is to develop an object recognition system addressing the partial occlusion issue. The system should recognize standalone single object under similarity transformation, and also partial occluded object successfully and efficiently, by using wavelet technique. Our object recognition algorithm is designed with the following objectives; it should:

1) be able to handle standalone object with similarity transform;
2) be able to recognize object with moderate partial occlusion;
3) be computationally efficient;
4) be able to tolerate noise contamination; and
5) should outperform existing algorithms.

1.8 Our Scheme and Contributions

Our recognition algorithm consists of the following processes as illustrated in Fig. 1.4. The recognition system developed in this thesis is a model-based system. Therefore, the recognition process consists of two blocks: database construction and unknown object recognition.

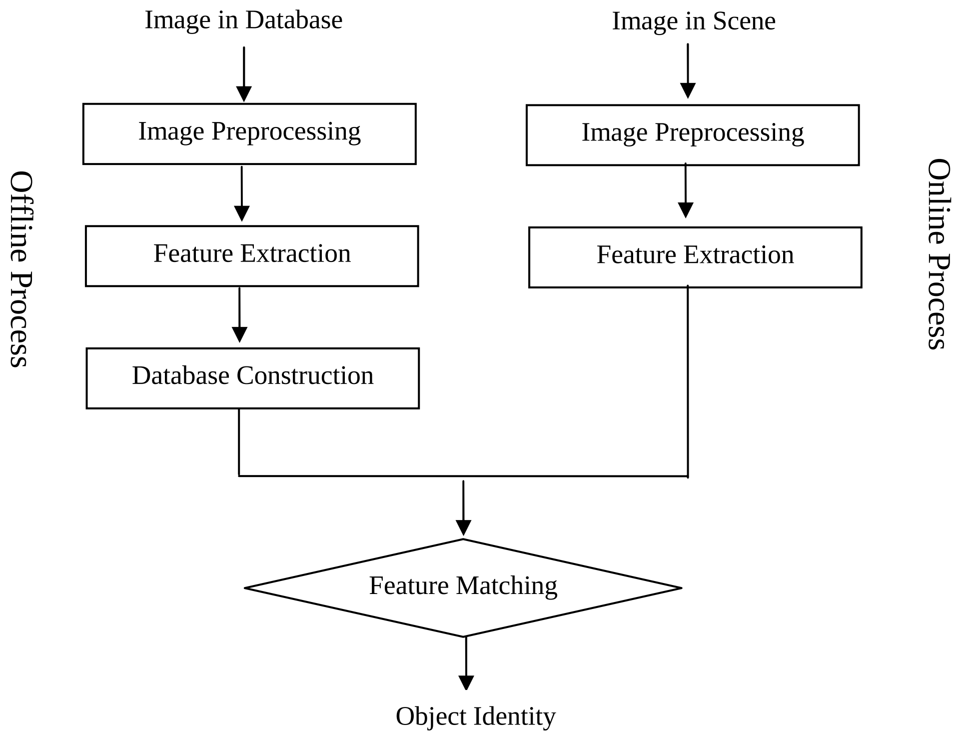


Fig. 1.4 Recognition process flow chart

A brief summary of a general recognition process is the following:

I. Database construction

We choose a set of good quality images as the candidates to construct the model database. Database construction is done offline to shorten the time required for recognition. For database construction, these images need to go through the following steps:

1) Image pre-processing and segmentation

   The images first undergo image enhancement, noise removal to enhance the quality of the image. After that, the edges of the objects are extracted and followed by boundary tracking.

2) Boundary partitioning

   The corner points on the object boundary are extracted using proposed wavelet-based scale-invariant corner detection methodology. Then, the object boundary is partitioned into curve segments in the way that each segment consists of two consecutive corners. We then shift the partition points away from the corners by a length which is proportional to the distance between these two consecutive corners.

3) Feature extraction

   For each curve segment, we normalize it so that it is translation, scaling and rotation invariant. After that, the normalized segment is represented by proposed wavelet representation. The representations of all the segments of the object form the feature matrix of the object.

4) Feature storing

We store the feature matrices of images containing objects with known identities together with their respective identities. Such that, if the feature matrix of an unknown object matches with any feature matrix in the database, the identity of the unknown object then can be retrieved from the database. The collection of feature matrices is called model database.

II. Unknown object recognition

After the completion of database construction, the recognition system is ready to recognize objects with unknown identity. Given an image of the scene which contains object(s) with unknown identity, the recognition system will enhance the image, detect the edges, track the object boundary, partition the object boundary and extract the features of the object(s). The pre-processing and feature extraction process are exactly the same for both model object and unknown object. Therefore, the algorithm discussed in chapter 4 & 5 for feature extraction is applicable for both model object and unknown object.

To recognize the unknown object in the scene, the feature matrix of unknown object needs to be matched with the feature matrices of the model objects one by one iteratively until a satisfactory match is found. If the number of models in the database is large, the iterative matching will be time consuming. Therefore, we designed a hierarchical matching algorithm which not only increases the efficiency of matching but also increases the matching accuracy.

This research project mainly addresses the three following issues which are crucial for the overall recognition system:

1. The proposed recognition system partitions the boundary of the object into a series of curve segments. To ensure the performance of proposed recognition system under the conditions of similarity transforms and partial occlusion, a curve partition algorithm to ensure invariance should be specially developed.
2. A compact, computational efficient, multiresolution and local object representation methodology must be devised to meet the stringent requirements of our particular recognition task.
3. A computational efficient matching algorithm wound be necessary to achieve efficient matching and high accuracy.

1.9 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2 gives a literature review on related research works in the recognition of 2-D standalone objects, as well as partial occluded objects which employ traditional techniques and novel wavelet techniques.

Chapter 3 introduces the mathematical background of wavelet which is essential in our object recognition process. It also highlights the superior properties of wavelet transforms over others, which facilitate our work.

Chapter 4 introduces the image preprocessing process adopted and our proposed boundary partition process using wavelet techniques. A specially designed wavelet-based corner detection method which is invariant to similarity transformation and partial occlusion is proposed first, followed by boundary partitioning.

Chapter 5 presents our wavelet-based object presentation algorithm. The invariance, stability, compactness, completeness, generalization and efficiency of proposed object representation are examined.

Chapter 6 describes the matching process with hierarchical matching strategy and decision making rule. The effectiveness and efficiency of this hierarchical matching process are discussed.

Chapter 7 demonstrates the efficiency and robustness of our proposed recognition system by extensive experiments in three aspects: standalone objects recognition with similarity transformation, partial occluded objects recognition with and without similarity transformation, and recognition of objects with boundary noise.

Chapter 8 concludes and summarizes the contributions from the research presented in this thesis. Some limitations of our proposed recognition approach are discussed. Potential future works are also presented.

Chapter 2

Literature Review

2.1 Introduction

The problem of recognizing two-dimensional object shapes has drawn much research attention. Many significant contributions have been made and published. In this chapter, we make a thorough literature review of related works.

Based on the natures of the features used, object recognition approaches can be categorized into *global* and *local* feature based approaches. *Global* features are usually some characteristics of the entire region or boundary. Those methods using *global* features such as area moments (Hu 1962, Teh et al. 1980, Khotanzad et al. 1990), curve moments (Chen 1993, Zhao et al. 1997) and Fourier descriptors (Persoon 1977, Richard et al. 1974, Etesami et al. 1985) have been well reported. The advantages of global-feature-based approaches are: the features are easily calculated and the number of features used for recognition is usually small, and the matching process is fast. However, one major setback of this approach is that it requires the objects being recognized to be wholly visible, non-overlapping, and not touching each other. Most pattern recognition algorithms developed for standalone object recognition fail to work when partial occlusion takes place. The reason is that these algorithms are designed based on global features, which are completely contaminated when partial occlusion takes place.

On the other hand, local features are usually on the boundary of an object or represent a distinguishable small area of a region. In order to develop a more efficient recognition system to handle not only the problem of isolated object recognition, but also the problem of partially occluded object recognition, many researchers have tried many approaches using various local features, such as boundary dominant points, curve segments, wavelet descriptors, etc. We can further categorize these local features into features in spatial domain and features in spectral domain. The former are usually the geometric primitives such as: corners, holes, curve segments (eg. Line and arc), etc. The features in spectral domain consist of Fourier descriptors and Wavelet descriptors. They are less sensitive to noise compared to the features in spatial domain. Among them, wavelet descriptor is the most promising one due to the possession of localization property in both spatial and frequency domains and multiresolution representation capability. In this chapter, we review several related works regarding partial occluded object recognition using various techniques.

2.2 Dominant-Point Based Approaches

Dominant points are rich in information, they are usually used as features for recognition. Some researchers (Ansari et al., 1990, Han et al., 1990, Lamdan et al. 1990, Tsang et al. 1994, Kim et al. 1996, Zhang et al. 2003) used dominant-point based recognition methods to recognize partially occluded objects.

Ansari and Delp (1990) used a set of landmarks, i.e., local extreme curvature points to represent each object. They introduced a local shape measure named “sphericity”, which was derived from the mapping of a set of three model landmarks to a set of three scene landmarks along the boundaries of the objects. A table of compatibility was constructed in order to store all the sphericity values. In the matching process, a technique named “hopping dynamic programming” guided the landmark matching through the compatibility table, in order to find a sequence of high-valued diagonal entries. This sequence of entries corresponded to a set of landmark represents a match between the scene and the model. A least-square-error technique was used to estimate the location of the model object in the scene, and to verify the hypothesis. The main problem addressed in this paper is the landmark matching for object recognition, and did not discuss method used in landmark extraction. It required the landmarks to have a consistent ordering. However, it is not a trivial task when both noise and occlusion are involved. Noise may generate spurious landmarks, and partial occlusion may break the boundary into discontinuous segments. This method works well when object being recognized have adequate landmarks and more than half of its landmarks can be detected in the correct sequential order. If only a few landmarks matched between object in scene and model, the final decision on recognition is ambiguous.

Han and Jang (1990) used the local maximum curvature points from curved boundary to represent object shape. The relative feature measure they used was the relative distance values between local maximum curvature points. An association graph method is adopted to identify objects, in which the nodes correspond to the local maximum curvature points of the occluded image. The presence of edge between two nodes indicates a high likelihood that the nodes belong to the same object. From the graph, a maximal clique was extracted. Using the minimum weight matching algorithm they proposed, a one-to-one correspondence was established between the nodes in the cliques and the local maximum curvature points in the object image. After estimating the location of the model object in the scene, the boundary consistency of the object was checked to verify the hypothesis. In order to increase the matching speed, a heuristic method has also been developed. However, by only considering the relative distance between corner points, it is insufficient to accurately determine the correspondence between objects in the scene and the model.

Zhang et al. (2003) recently proposed a shape space based approach for invariant object representation and recognition. They also make this approach capable of recognizing partial occluded objects by using partial Procrustean distance as the measure. In this article, the representation of the object in shape space is based on the landmarks, e.g. local curvature maxima. The shape space concept was introduced by Kendall (1977, 1984) and Bookstein (1984). By using shape space, the object can be represented as a point in a high-dimensional space, called shape space. In a shape space of 2D objects, all possible views of an object caused by translation, scaling and rotation are represented by a single point. Object recognition can be achieved by computing the Euclidean distance between the object and a model in the shape space. If the object is related to the model by similarity transform, the distance is approaching zero. The aim of using partial Procrustean distance is to ensure that we use only the “true” landmarks, i.e., those shared by the occluded and the model object. However, in practice, one may not know which landmarks are “true”, therefore a search would be needed. Random searching can be very computational intensive, therefore, they set a constraint which is all “true” marks are contiguous, i.e. an occlusion always cuts off a continuous curve from the object contour. This constraint drastically reduces the number of searches. However, it limits the application scope of this approach.

As reviewed above, object recognition approaches wholly rely on dominant points alone suffer from two drawbacks:

a) Dominant points alone are insufficient to form a complete integrated representation of an object. Therefore, the recognition result is uncertain.

b) Dominant point extraction is sensitive to noise contamination. Severe noise on boundary may generate spurious corners. Smoothing operation e.g. Gaussian smoothing can reduce the effect of noise, however, there is no proper guiding principle for choosing the proper width of a smoothing mask.

Thus, representation based on dominant points does not meet criteria (b) and (c) stated in Section 1.4. Therefore, dominant points based recognition algorithms are not optimal especially in situations involving occlusion.

2.3 Polygonal Approximation Approaches

Polygonal approximations represent the object boundary as a string of line segments. They are computed by using various criteria to determine “breakpoints” that yield the best polygons. The polygonal approximation (Bhanu et al. 1984, Price 1984, Ayache et al. 1986, Bhanu et al. 1987, Eric et al. 1989, Liu et al. 1990) has been widely used as a representation for recognizing occluded object or object with unknown scale.

Ayache et al. (1986) used polygons to represent objects and regarded polygon line segments as local features. Their matching process was a recursive hypothesis prediction and evaluation procedure. A prediction is made by matching a segment in the model with a segment in the scene by comparing local intrinsic feature measures. To evaluate the hypothesis, they matched additional segments of the model with a segment in the scene, updated the hypothesized position, and computed a quality score of the match. After a sufficient number of hypotheses had been evaluated or a very high quality measure was reached, they stopped the matching process. The hypothesis with the highest score was examined before being validated or rejected. Because they use the transformation information calculated from local feature measure of some portions to restrict the matching on other portions, they could guarantee match consistency not only in local portions, but also in global regions.

Liu and Srinath (1990) presented a polygonal approach to recognize and locate partially distorted two-dimensional shapes without regard to their orientation, location and size. They first calculate the curvature function from digitized image of an object. The points of local maxima and minima extracted from the smooth curvature function are used as control points to segment the boundary and to guide the boundary matching procedure. The boundary matching procedure considers two shapes at a time, one shape from the template data bank, and the other being the object to be classified. The procedure tries to match the control points in the unknown shape to those of a shape from the template data bank, and estimates the translation, rotation, and scaling factors to be used to normalize the boundary of the unknown shape. The chamfer 3/4 distance transformation and a partial distance measurement scheme are used as the final step to measure the similarity between these two shapes. The unknown shape is assigned to the class corresponding to the minimum distance. Experimental results showed that this algorithm works reasonably well even with moderate amount of noise. As they mentioned, proper selection of the value of standard deviation of Gaussian function was important to the success of this algorithm. They chose the standard deviation tentatively so that the boundary of the boundary was broken into 40 segments. However, the choice of 40 segments is shape and occlusion dependent. An automatic method for the selection of the standard deviation has not been presented. Therefore, the performance of this algorithm is limited.

A new method called supersegment has been proposed by Fridtjof et al. (1992) to increase the reliability of the polygonal approximation approach by performing segmentation of a boundary at varying thresholds. This method can be applied to scale-invariant recognition by using the angle between the neighboring segments and the arc length ratio as features. It improves upon the results obtained with the conventional polygonal approximation technique. However, it is still unstable with regard to break points, especially for curved objects.

Recognition methods rely on polygonal presentation only work well for polygonal objects. For non-polygonal objects, these methods have the drawback that they are unstable in finding break points. Therefore, polygonal approximation does not fulfill the generalization criteria (f) mentioned earlier in Section 1.4.

2.4 Curve Segment Approaches

To recognize object which is not polyhedral objects, researchers tried to describe object by circular arcs (Turney et al. 1985, Knoll et al. 1986, Kalvin 1986, Ettinger 1988, Grimson 1989). In order to make the representation even more complete and precise, some researchers used the combination of some basic geometric features, such as line, arc, corner and end to describe a contour (Tsang et al. 1992, Wei 1998, Sarkar 2003).

Tsang et al. (1992) proposed a technique for the recognition of occluded object which use corners and circular arcs as the features. The set of primitive features, together with their respective physical size, form a representation which contributes to an identity of the object concerned. The object boundary is extracted and transformed into the $\theta - S$ domain. A zero- and a first-order discontinuity detector are then employed to detect the corners and arc segments, respectively, on the object boundary. The terminals of a complete circular arc are localized with the use of regression analysis, and the total angular change is determined directly from the internal angle covered by the segment on the θ axis. The position and angular spans of a corner are reflected from the location and the size of the corresponding zero-order discontinuity. Classification of the features of an unknown object shape is performed by a multilayer artificial neural network which is capable of identifying distorted and incomplete input patterns. From the illustrations in this article, we can see corners and circular arcs can not cover the entire object boundary. Therefore, representation by corners and circular arcs is incomplete.

Lim et al. (1995) and Xin et al. (1995) proposed a scale-space based algorithm to detect line, arc, corner and end on a curve. This scale-space based geometric primitives detection algorithm is insensitive to noise and robust to similarity transform. Based on this, Wei (1998) proposed an object recognition system which can recognize non-occluded and partially occluded two-dimensional objects. In his work, methods of calculating the local feature measures and relative feature measures of these geometric primitives are introduced. The integration of these features and feature measures is applied to efficiently represent the object shape. An association graph method is used to match object in the scene with the objects in the model. The local feature measures is compared to find possible match pairs between scene and model. Then, relative feature measures are employed to find mutual consistency among the nodes. Boundary of the model object is superimposed on to the boundary of the scene object according to the translation, rotation and scaling parameters indicated by the maximal cliques for hypotheses verification.

Sarkar et al. (2004) proposed a method for approximating digital planar curves with line segments and circular arcs. They formulate the approximation task as an optimization problem, in the way that seeks the desired set of optimal breakpoints such that when line segments and circular arcs are appropriately fitted between all pairs of adjacent breakpoints, the fitting error is minimized. By using Genetic Algorithm (GA), the optimal (or near-optimal) solution can be reached fairly quickly. However, using only lines and arcs to represent object with arbitrary shapes can not achieve high accuracy. Therefore, the recognition algorithm based on this representation hardly differentiates different objects with similar shapes.

Although the combination of these geometric primitives represents most geometric features of the object boundary over dominant points method and polygonal representation method, most man-made objects can be fully represented by these geometric primitives. However, some objects with irregular shape may not be fully represented by the above mentioned geometric primitives; some portion may not belong to any of the categories, such as involutes. Therefore, this representation is incomplete. Another drawback of this approach is its inefficiency to determine the end points of arcs and lines precisely due to noise and scaling, therefore the local measures of arcs and lines are not accurate either. On the whole, representation using the combination of the geometric primitives does not meet criteria (a) and (d) mentioned in Section 1.4. Thus, the performance of object recognition approaches based on this representation is limited.

2.5 Other Approaches

Lim et al. (2004) used curve moment invariants to confront partial occluded object recognition problem. Curve moment (Chen, 1993) is an improved moment invariants which are computed based only on the shape boundary, and hence they are even more computational efficient over traditional moment invariants (Hu, 1962). They first partition object boundary into a sequence of curve segments. Then the curve moments at different order of each curve segment are calculated. The collection of the curve moment invariants of all the segments on the object boundary form the representation of the object. A hierarchical matching strategy is adopted to speed up the matching process. Curve moment representation proposed in this article is invariant to similarity transform and compact in size. However, similar to the traditional moment invariant, the curve moment is a many to one correspondence. Therefore, the representation using curve moment is not unique. Moreover the physical meaning of curve moments at higher order is unknown. Therefore, this representation does not possess hierarchical representation capability.

Salari and Blaji (1991) proposed an object recognition method using a B-spline representation of the boundary. Curve segments on the boundary are represented using B-splines which are piecewise polynomial curves guided by a sequence of points. The B-spline control points found from the boundary points are then used to extract local features of the curve. Then, a Hough transform like method is applied to detect a consistent set of scale, rotation and translation parameters. After normalization using the transformation parameters extracted, the match measures are computed to validate the match. Despite the success of this algorithm, control points alone can not form a complete representation of the shape of the object.

2.6 Fourier Descriptors Approaches

Gorman and Mitchell (1988) represent an object contour by Fourier coefficients of contour segments. The breaking points of the contour are the vertices which result from a polygonal approximation of the contour. Each contour segment is a portion of the object contour and consists of three consecutive vertices. It begins from a vertex which is considered as the first vertex and then ends at the third vertex along the object contour. The feature values of each contour segment are the Fourier coefficients derived from tracing along the segment from the beginning to the end and then back to the beginning. The shape measure between a model and a scene contour segment is the norm squared distance between the Fourier coefficients of the two segments. An inter-segment distance table measuring the norm squared distances between the model and the scene contour segments is constructed. The table is augmented by repeating the rows. A backward dynamic programming procedure is then used to determine the minimum distance path starting from the first column to the last column of the augmented table. An entry along the minimum distance path that results from a diagonal transition corresponds to a match between the model and the scene segment, indicated by the row and the column indices of the entry. However, a simple, automatic method for selecting a threshold value for contour splitting was not proposed in this paper. During the experiments, it was assumed that the scale factor between the sizes of the known and unknown contours was approximately known. This assumption is not always valid in practical applications. Therefore, the performance will degrade significantly when the scale information is not available.

2.7 Wavelet Approaches

As mentioned by Zhang et al. (2004), object representation in spectral domain is the most promising one. Besides Fourier transform, wavelet transform is another recently developed technique which can describe object in both spectral domain and spatial domain. It preserves the high efficiency in computation, compactness and robustness as Fourier transform, and also possesses the properties of multi-resolution approximation, spatial description and singularity detection which are especially useful for this research work. Since the mathematical theory of wavelet has reached a mature stage, some researchers have tried to solve object recognition problem using wavelet techniques. Some pattern recognition approaches using wavelets have been reported (Chuang et al. 1996, Tieng et al. 1997, Antoine et al. 1997, Yoon et al. 1998, Bui et al. 1999, Khalil et al. 2000, Yu et al. 2001, Tsang 2001, Khalil et al. 2002). Since our algorithm also used wavelet technique, and we were enlightened by these works from several aspects, it is worth while to review them in detail. We do not restrict the review into object recognition of partially occluded object only.

Chuang et al. (1996) proposed a hierarchical planar curve descriptor that decomposes a curve into components of different scales using wavelet transform so that the coarsest scale components carry the global approximation information while the finer scale components contain the local detailed information. The effect of scaling, translation, and rotation of a planar curve on its wavelet descriptor was derived. Features extracted from the wavelet approach can be normalized so that we can handle the effect of rotation, translation, and scaling. The performance of a class of wavelet bases with different vanishing moment and symmetry properties was studied. A deformable wavelet descriptor is also proposed by interpreting the wavelet coefficients as random variables. In contrast to the scale-space filtering approach, which serves primarily as an analytical tool, the wavelet descriptor provides an effective synthesis tool as well. And compared with Fourier descriptor that uses global sinusoids as the basis functions, the wavelet descriptor is more efficient in representing and detecting local features of a curve due to the spatial and frequency localization property of wavelet bases. However, the features are normalized using averaged magnitude and phase over the entire boundary. Thus, the features obtained can be easily distorted by partial occlusion. Therefore, recognition system using this feature can not handle recognition of partial occluded object.

Tieng and Boles (1997) proposed an algorithm which could recognize a two-dimensional object of arbitrary shape using the wavelet transform zero-crossing representation. The object boundary is first represented by a normalized Modified Radial Function (MRF). The MRF is a representation of r which is the distance between the boundary points to the centroid as a function of l which is the arc length of the boundary. After that, in order to match objects with different sizes, the MRF is sampled into standard size which is a number of power-of-two to facilitate dyadic wavelet transform. A dyadic wavelet transform is applied to the normalized MRF using the first derivative of a cubic spline as the wavelet base. The position and magnitude of zero crossings of the wavelet coefficients of a few low resolution levels are extracted as the representation of the shape of the object. Four alternative dissimilarity functions which compare the unknown object and candidate model are also proposed. Since these dissimilarity functions require the compared representations to have the same number of zero-crossings, a false zero-crossing elimination algorithm is also presented. Experimental results showed that the performance of the proposed algorithm is much better than the use of Fourier descriptors-particularly in the presence of noise. Because construction of the MRF needs to shift the origin of the coordinate system to the centroid of the object, therefore the centroid of the object will be distorted by partial occlusion. Moreover, zero-crossings alone are insufficient to represent the shape of object completely.

Yoon et al. (1998) proposed a notion of object representations using curvature zero-crossings of the approximation of object contour at various scales. Three different representations are proposed to solve different recognition problems. Among them, the “scale-invariant” representation can be used to handle object matching in the presence of noise, occlusion, and scale variation. The scaling effect of an object is modeled by using the continuous wavelet transform. Then the scaled low-resolution boundaries are decomposed using discrete wavelet transform. The zero-crossings on the curvature functions over the desired scales form the “scale-invariant” representation of the object. However, match of the zero-crossings of the object contours between target and model is a necessary but insufficient condition to confirm that the object in the scene matches with the particular model. Further more, wavelet transform is used as low pass filter in this work; other than the computation efficiency, the authors did not show the remarkable analytical advantages of wavelet transform over other low-pass filter, such as Gaussian.

Bui and Chen (1999) proposed a novel set of descriptors using wavelet and Fourier transforms. They first transform the pattern (image) into polar coordinate (r,θ) using the centre of mass of the pattern as origin. Then, they apply the Fourier transform along the axis of the polar angle θ and the wavelet transform along the axis of radius r . This feature combined the advantages of Fourier transform (e.g. translation invariant) and Wavelet transform (e.g. multiresolution analysis). The features thus obtained are invariant under translation, rotation and scaling. Experimental results show that the proposed Fourier-Wavelet descriptor is an efficient representation which can provide a reliable recognition. However, this algorithm is only applicable for the recognition of objects which are wholly visible because of the transformation to polar coordinate system and the use of Fourier transform.

Khalil and Bayoumi (2000) proposed a technique to recognize 2D objects under similarity transform. They first make a continuous wavelet transform of the slope representation of the boundary of an object. Then, they extract the position and regularity of each singularity of the slope representation. The singularities of the slope representation correspond to the corners or reflection points of the object boundary. The feature vector is composed by the distance between adjacent singularities and the regularities (represented by Lipschitz exponent) of each singularity. The classifier used is the multilayer feedforward neural network. Experimental results show that it outperforms some traditional methods, such as the Fourier descriptors method and the moment invariants method especially in the presence of noise. However, the authors did not pay attention to partial occlusion problem here. Although each individual element of the feature vector adopted in this method itself is a local feature, this method can not recognize partially occluded object with similarity transformation. The reason is that the number of boundary points is normalized into 256 points prior to the extraction of object features, thus the boundary size is affected by both scaling and partial occlusion. Therefore, the feature obtained is distorted when partial occlusion occurs. In 2002, they (Khalil et al. 2002) also derived several affine invariant functions using different dyadic levels. It has been shown that these invariant functions outperform some traditional invariant functions. These affine invariant functions have the potential to be used as features in object recognition system.

Yu et al. (2001) developed a feature extraction algorithm using wavelet and Fractal. In this paper, the notion of feature extraction with wavelet and fractal theories is presented as a powerful technique in pattern recognition. This novel method of feature extraction includes utilizing the central projection transformation (CPT) to describe the shape, with the wavelet transformation to help in the boundary identification, and the fractal dimension to enhance the discrimination power. Its essential advantage is that it can be used to recognize more complex patterns than the traditional Fourier descriptors. Although the fractal dimensions reduce the dimensions of the feature tremendously, the discrimination capability is also reduced since the fractal dimension is a many-to-one correspondence. In another word, this representation is not complete and not unique. Moreover, this algorithm is not suitable for partially occluded object recognition because of the use of CPT and fractal dimension over the entire pattern.

Tsang (2001) proposed a planar curve descriptor which is invariant to translation, size, rotation and starting point using wavelet transform. They first extract the boundary of a two-dimensional object and express its coordinates as a complex number. After that, the boundary data is normalized so that it is invariant to the size and position. Then, a wavelet transform is applied using a second-order Daubechies wavelet. The feature vector is composed of the number of time that the lines at the predefined levels cut across the magnitude response of the wavelet transform and the total width of the cutting stripes at different levels. Experimental results showed that by using the proposed feature, one can recognize standalone objects efficiently regardless of similarity transformation. They also extend it to the recognition of occluded objects by incorporating local features which are the distance from the current stripe to the next stripe at the specified levels into the feature vectors. However, they simplified the problem by restricting the size of the objects to be classified fixed. The feature does not work when the size of object varies. This assumption makes the recognition problem much easier in comparison with ours.

From the literature review we presented above, we could conclude that wavelet descriptors are more promising over those spatial descriptors because they possess multiresolution representation capability and are less affected by noise contamination. Wavelet descriptors also outperforms Fourier descriptors because of the localization property in both frequency domain and spatial domain. However, since wavelet is a relative new technique, the research of object recognition using wavelet is still weak. Most existing wavelet based object recognition algorithms do not pay attention to occluded object recognition problem. Although Yoon et al. (1998) and Tsang (2001) have tried to solve partial occluded object recognition problem, however they did not fully utilize the strength of wavelet technique. The representations of object proposed by them are not complete and not unique. In conlcusion, the recognition of 2-D partial occluded objects using wavelets is a novel idea which has good potential, but it has not been fully explored.

Chapter 3

Introduction of Wavelet

3.1 Introduction

Wavelet technique plays a key row in this research project. In this thesis, a wavelet-based corner detection algorithm is proposed for boundary partition, followed by a feature extraction algorithm using wavelet multiresolution decomposition. Therefore, a brief introduction of wavelet transform's mathematical background and its attractive properties is essential prior to introducing our proposed recognition algorithm. For more detailed information about wavelet transform, readers can refer to (Daubechies, 1992, Mallat, 1998, Chui, 1992).

Although the Fourier transform has been the mainstay of transform-based image processing since the late 1950s, a more recent transformation, called the wavelet transform, is now making it even easier to compress, transmit, and analyze images. Unlike the Fourier transform, whose basis function are sinusoids, wavelet transforms are based on small waves, called wavelets, of varying frequency and limited duration. This allows wavelet transform to provide not only frequency (scale) information, but also the temporal information.

Grossmann and Morlet (1984) studied the wavelet transform in its continuous form and initially applied it to analyze geological data. However, at that time, the development of wavelet theory and its application were limited due to the lack of "good" wavelet basis which are smooth, compactly supported and orthonormal. A few years later, Daubechies (Daubechies, 1988), a female mathematician, constructed a class of wavelet bases which are smooth, compactly supported and orthonormal. In the same year, Mallat (Mallat, 1988) proposed a general method to construct wavelet bases. It is termed Multi-Resolution Analysis (MRA) and is intrinsically consistent with sub-band coding in signal analysis. The above achievements played an important role in the development of the wavelet theory. They made the wavelet a mature theory.

3.2 Multiresolution Analysis (MRA)

It is always desired to analyze a signal or a function in multiple resolutions, because the objects in an image are observed to occur at different scales, and the related features such as edges, for example, can be either a sharp transition from black to white or one that occurs gradually over a considerable distance. In general, a multiresolution approach to image representation or analysis seeks to exploit this idea. Beside *pyramid algorithm* (Burt and Adelson, 1983) and *subband coding* (Woods and O'neil, 1986), *wavelet transform* played an important role in the development of a unique mathematical theory called multiresolution analysis. By using wavelet transform, a function or signal can be viewed as composed of a smooth background with fluctuations or details on top of it, where the smooth background reveals the approximation of the function at coarse scale, while the fluctuations at finer scales represent the abruption details at higher resolutions. In most literature, MRA refers to the general method to construct wavelet bases proposed by Mallat (1988). However, the major concern of the research presented in this thesis is on how to represent the object into multiple resolutions instead of on how to construct a wavelet basis. Therefore, in this section, the MRA is introduced in the way to demonstrate how a function can be represented at multiresolution using wavelet. To avoid being bogged down in mathematical rigor, some complicated portions of MRA theory have been omitted. Readers may refer to Mallat's paper (Mallat, 1989b) for a complete understanding of MRA.

Given a scaling function $\varphi(x)$ that meets the MRA requirements, scaling function $\varphi(x)$ form a multiresolution analysis that must satisfy the following refinable property:

$$\varphi(x) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_0(n)\varphi(2x-n) \tag{3.1}$$

where $h_0(n)$ is called scaling vector.

Its counterpart wavelet function $\psi(x)$ can be expressed as a weighted sum of shifted, double-resolution scaling functions, and can be written as:

$$\psi(x) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_1(n)\psi(2x-n) \tag{3.2}$$

where $h_1(n)$ is the wavelet vector, it correlates with $h_0(n)$ as

$$h_1(n) = (-1)^n h_0(1-n) \tag{3.3}$$

The dilation and shift of scaling and wavelet function are defined as:

$$\varphi_{j,k}(x) = 2^{j/2}\varphi\left(2^j x - k\right) \quad j,k \in \mathbb{Z} \tag{3.4}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi\left(2^j x - k\right) \quad j,k \in \mathbb{Z} \tag{3.5}$$

Here, k determines the position of $\varphi_{j,k}\left(x\right)$ and $\psi_{j,k}$ along $x-$axis , j determines the width, and $2^{j/2}$ is used to normalize the amplitude.

The subspace or the closed span formed by the scaling function $\varphi_{j,k}(x)$ is denoted as

$$V_j = \overline{\underset{k}{\text{Span}}\{\varphi_{j,k}(x)\}} \tag{3.6}$$

The subspaces spanned by the scaling function at low scales are nested within those spanned at higher scales as shown in Fig. 3.1. That is,

$$V_{-\infty} \subset \cdots \subset V_0 \subset V_1 \subset V_2 \cdots \subset V_\infty \tag{3.7}$$

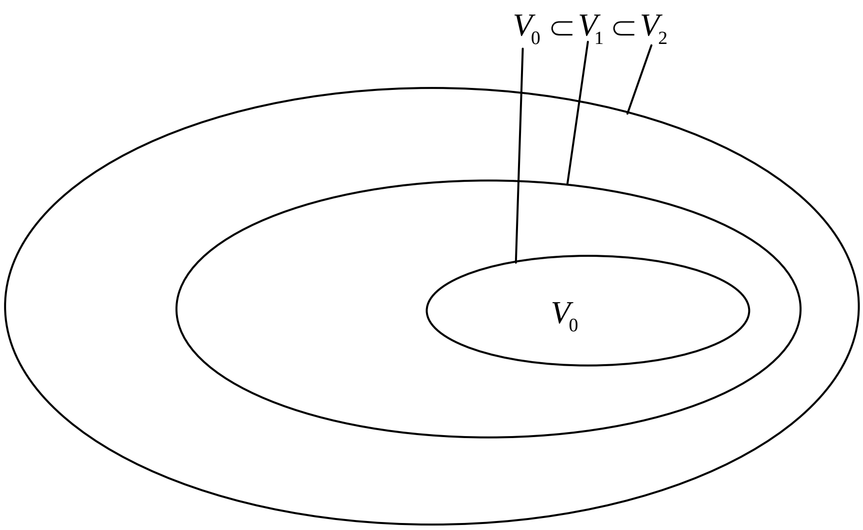


Fig. 3.1 The nested function spaces spanned by a scaling function

The subspace spanned by the wavelet function $\psi_{j,k}(x)$ is

$$W_j = \overline{\underset{k}{\text{Span}}\{\psi_{j,k}(x)\}} \tag{3.8}$$

The scaling and wavelet function subspaces are related by

$$V_{j+1} = V_j \oplus W_j \tag{3.9}$$

Where $\oplus$ denotes the union of space. The orthogonal complement of V_j in V_{j+1} is W_j, and all members of V_j are orthogonal to the member of W_j. Thus

$$\langle \varphi_{j,k}(x), \psi_{j,l}(x) \rangle = 0 \tag{3.10}$$

The relation is illustrated in Fig. 3.2.

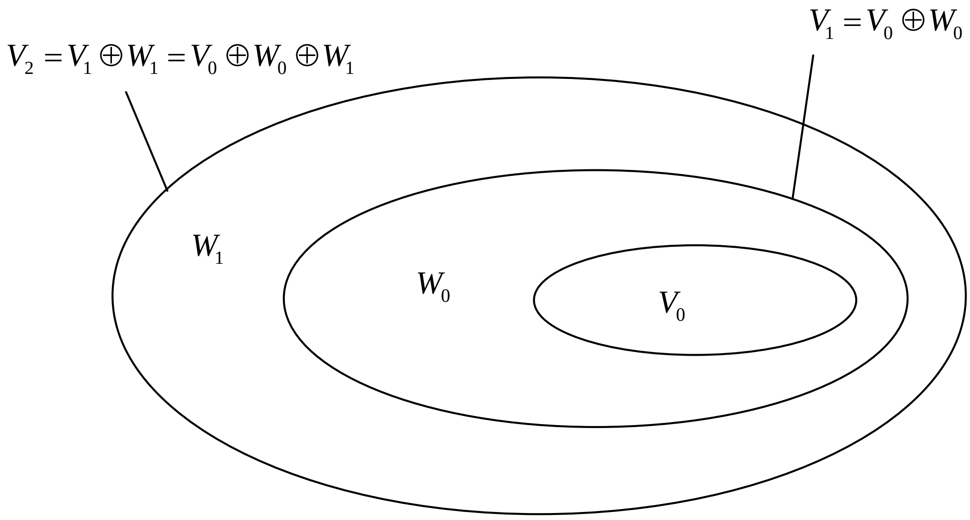


Fig. 3.2 The relationship between scaling and wavelet function spaces

Now the space of all measurable, square-integrable functions can be expressed as

$$L^2(R) = V_0 \oplus W_0 \oplus W_1 \oplus \cdots \tag{3.11}$$

The wavelet series expansion of a function $f(x) \in L^2(R)$ can be expressed as

$$f(x)=\sum_{k} c_{j_0}(k)\varphi_{j_0,k}(x)+\sum_{j=j_0}^{\infty}\sum_{k} d_j(k)\psi_{j,k}(x) \tag{3.12}$$

where j_0 is an arbitrary starting scale, and $c_{j_0}(k)$ and $d_j(k)$ are referred to scaling coefficients and wavelet coefficients, respectively. If the expansion function forms an orthonormal basis or tight frame, the scaling coefficients $c_{j_0}(k)$ and wavelet coefficients $d_j(k)$ can be obtained by

$$c_{j_0}(k)=\left\langle f(x),\varphi_{j_0,k}(x)\right\rangle=\int f(x)\varphi_{j_0,k}(x)dx \tag{3.13}$$

$$d_j(k)=\left\langle f(x),\psi_{j,k}(x)\right\rangle=\int f(x)\psi_{j,k}(x)dx \tag{3.14}$$

If the expansion functions are part of a biorthogonal basis, the φ and ψ terms in the equations must be replaced by their dual functions $\tilde{\varphi}$ and $\tilde{\psi}$ respectively.

3.3 Discrete wavelet transform

If the function $f(t)$ being expanded is a sequence of numbers, e.g. $f(t)$ is a function of discrete variables $t=0,1,2,\cdots,n-1$, the resulting coefficients are called the discrete wavelet transform (DWT) of $f(t)$. For this case, the series expansion defined in Equations, (3.12) through (3.13) become the DWT transform pair

$$c_{j_0}(k)=\left\langle f(t),\varphi_{j_0,k}(t)\right\rangle=\sum_{t} f(t)\varphi_{j_0,k}(t-k) \qquad t\in\mathbb{Z} \tag{3.15}$$

$$d_j(k)=\left\langle f(t),\psi_{j,k}(t)\right\rangle=\sum_{t} f(t)\psi_{j,k}(t-k) \qquad t\in\mathbb{Z} \tag{3.16}$$

Again, the two above equations are applicable for biorthogonal wavelet basis, in this case, the φ and ψ terms in the equations must be replaced by their dual functions $\tilde{\varphi}$ and $\tilde{\psi}$ respectively.

$f(t)$ can be reconstructed by inverse discrete wavelet transform.

$$f(t) = \sum_k c_{j_0}(k)\varphi_{j_0,k}(t) + \sum_{j=j_0}^{j_m}\sum_k d_j(k)\psi_{j,k}(t) \qquad t \in \mathbb{Z} \tag{3.17}$$

where j_m is the maximum resolution that can be decomposed. Due to the fact that $f(t)$ has limited number of data, its resolution is also limited.

3.4 Fast wavelet transform

Mallat (1989a, 1989b) defined a discrete wavelet transform algorithm that is more efficient than computing a full set of inner product. It applies two-band subband coding in an iterative fashion and builds the wavelet transform from the bottom up. That is, computing begins with small scale coefficients and ends at the coarsest scale. It is also called Mallat's herringbone algorithm [Mallat 1998].

Recalling the multiresolution refinement equation (3.1), and substituting it into equation (3.4), we can obtain

$$\begin{aligned}
\varphi_{j-1,k}(x) &= 2^{(j-1)/2}\varphi\left(2^{j-1}x-k\right) \\
&= 2^{(j-1)/2}\sum_{n\in z}\sqrt{2}h_0(n)\varphi\left(2^j x-2k-n\right) \\
&= \sum_{n\in z}2^{j/2}h_0(n)\varphi\left(2^j x-2k-n\right) \\
&= \sum_{n\in z}h_0(n)\varphi_{j,2k+n}(x) \\
&= \sum_{p\in z}h_0(p-2k)\varphi_{j,p}(x)
\end{aligned} \tag{3.18}$$

Similarly, substitute equation (3.2) into (3.5), we obtain

$$\psi_{j-1,k}(x) = \sum_{p \in z} h_1(p-2k)\varphi_{j,p}(x) \tag{3.19}$$

By substituting equation (3.18) into equation (3.15)

$$\begin{aligned} c_{j-1}(k) &= \sum_{p \in z} h_0(p-2k)c_j(p) \\ &= h_0(-n) * c_j(n)\big|_{n=2k} \end{aligned} \tag{3.20}$$

Similarly, by substituting equation (3.19) into equation (3.16), we can obtain

$$\begin{aligned} d_{j-1}(k) &= \sum_{p \in z} h_1(p-2k)c_j(p) \\ &= h_1(-n) * c_j(n)\big|_{n=2k} \end{aligned} \tag{3.21}$$

where $*$ is the convolution operation. Equations (3.20) and (3.21) indicate that $c_{j-1}(k)$ and $d_{j-1}(k)$ can be easily calculated by simply convolving $c_j(k)$ with the low-pass filter $h_0(-n)$ and high-pass filter $h_0(-n)$ respectively, and down-sampling it by two. This decomposition can be carried out hierarchically by further decomposing $c_{j-1}(k)$ using the methodology shown in Fig.3.3. Note that, here the highest scale coefficients are the original sampled function itself, e.g. $f(k) = c_j(k)$.

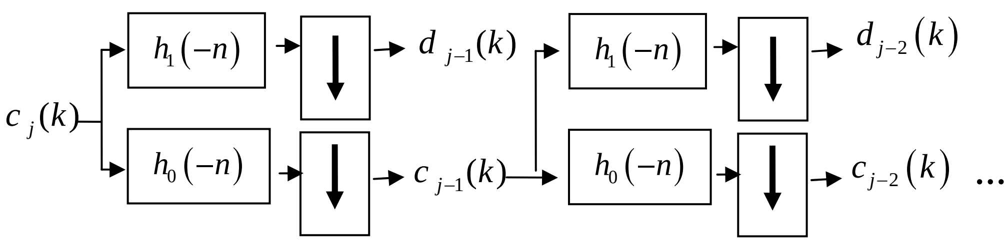


Fig. 3.3 Fast wavelet transform

The number of mathematical operations involved in the computation of the FWT of a length $M = 2^J$ sequence is in the order of $O(M)$. That is to say the number of floating-point multiplication and additions (using filter banks) is linear with respect to the length of the sequence. This compares favorably with FFT algorithm, which requires $O(M \log M)$ operations.

The inverse transform is obtained by reversing the process as shown in Fig. 3.4.

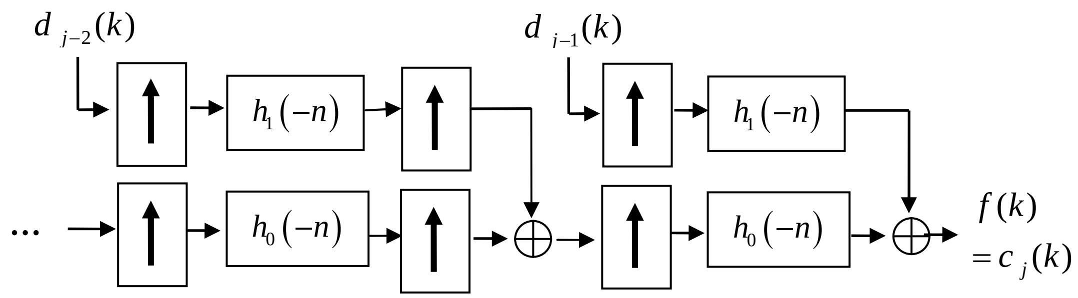


Fig. 3.4 Inverse discrete wavelet transform

3.5 Wavelet bases selection

Wavelet bases construction plays an important role in wavelet development. In the early stage of wavelet development, the roughness of wavelet made the mathematicians doubt about the existence of a good wavelet basis, until Daubechies (1988) constructed a family of orthonormal wavelets having compact support, and then Mallat (1989b) proposed a general method called MRA to construct wavelet bases. The purpose of this section is to introduce the properties of different wavelet bases commonly used in the areas of signal processing, image processing and pattern recognition, while the details on the construction of their mathematical formula is avoided here. The details can be found in Mallat (1998).

Notice that Continuous Wavelet Transform (CWT) is overcomplete. For overcomplete transforms, the restrictions on the basis functions are relatively mild. If $\psi(x)$ is a real-valued function whose Fourier spectrum $\Psi(s)$ satisfies the admissibility criterion (Grossman et al. 1984, Chui 1992), then $\psi(x)$ is called a basic wavelet. However, to extract object compact features, such over-complete transforms are inapplicable. Moreover, the fast wavelet transform algorithm is infeasible for overcomplete transforms using wavelet base with redundancy.

Daubechies (1988) constructed a family of orthonormal wavelets having compact support. For each wavelet base of order r, the set of wavelets

$$\{{}_r\psi_{j,k}(x)\} = \{2^{j/2}\,{}_r\psi(2^j x - k)\} \tag{3.22}$$

where j and k are integers, forms an orthonormal basis. Furthermore ${}_r\psi(x)$ is zero outside the interval $[0, 2r-1]$. The fast wavelet transform can be applied when using orthonormal wavelet bases, and the scaling vectors and wavelet vectors are the same for both forward wavelet transform and inverse wavelet transform. However, the orthonormal wavelet bases lack symmetric properties which are desirable for most cases.

The biorthogonal DWT requires two scaling vectors and two wavelet vectors rather than one each, but this does not increase the computational burden of the process. The biorthogonal transform, however, affords a much wider choice of wavelet shape than orthonormal transform, so it is preferable in many applications.

The choice of a basic wavelet is usually governed by the application. For example, for function representation, an orthonormal or biorthogonal basis is desirable or required, since the objective is to represent the function faithfully and compactly. An overcomplete transform increases the amount of data required to completely represent the function. For some specific feature detection, then it is more important to select a wavelet that is similar to the components of interest. In this thesis, the reason why we use a specified wavelet base for this particular task will be explained separately where necessary.

3.6 Properties of wavelet that are useful in this research project

- Localization in both spatial and frequency domain

The most important property of wavelet is that wavelet bases are localized in space domain, while the Fourier sinusoidal functions are not. This localization feature of wavelet, along with wavelets localization property in frequency domain, makes many functions sparse in wavelet domain. This sparseness results into a number of uses in this research project, such as corner detection and efficient object representation by eliminating insignificant coefficients.

- Multiresolution representation

The wavelet representation also provides a coarse-to-fine matching strategy in pattern recognition, called multi-resolution matching. The matching starts from the coarsest scale and moves on to the finer scales. The costs for different scales are quite different. Since the coarsest scale has only a small number of coefficients, the cost at this scale is much less than that of finer scales. In practice, the majority of the patterns can be filtered out during the coarse scale matching, while only few patterns will need information at finer scales to be identified. Therefore, the process of multi-resolution matching will be faster compared to the conventional matching techniques.

- Fast computation

Due to the existence of Fast Wavelet Transform (FWT) algorithm, the object representation using wavelet transform can be extracted with much lesser computation load than without using FWT. Thus, the overall object recognition algorithm can be more efficient, and online recognition can become possible.

- Representation of function at appropriate resolution

Wavelet representation can adjust the resolution according to the resolution of the input function, while representation using Fourier transform (Gorman et al. 1988) must be represented in designated resolution. Therefore, the wavelet representation is more precise and efficient.

- Wide choice of wavelet bases

Unlike Fourier transform, which utilizes only the sinusoidal function, wavelet transform have an infinite set of possible bases functions. By choosing the right wavelet base, we can detect and extract desired features. In addition, we are also able to represent a function efficiently and effectively.

Chapter 4

Preprocessing and Boundary Partitioning

4.1 Introduction

Feature extraction is the core of an object recognition algorithm. Given an input image, either from template database or scene, our proposed feature extraction algorithm yields a compact, effective and robust feature suitable for the intended partial occluded object recognition task. Our proposed feature extraction algorithm consists of the following steps as shown in Figure. 4.1. The given input image is first de-noised and enhanced to improve the image quality in order to ease the later processes. Then the boundary of an object in the image is obtained by edge detection, followed by boundary tracking to serialize the object boundary pixels into a sequence of coordinates of points. After that, the object boundary is partitioned into curve segments using corner points extracted by our proposed wavelet-based corner detection algorithm. Subsequently, the coordinates of each segment are normalized so that it is invariant under translation and rotation, followed by re-sampling it into the nearest upper number of power of two to facilitate dyadic discrete wavelet transform. The normalized coordinates of curve segment are decomposed using wavelet transform. Only significant wavelet coefficients are kept as the object features, while those insignificant wavelet coefficients are discarded to make the object representation compact and robust to noise contamination.

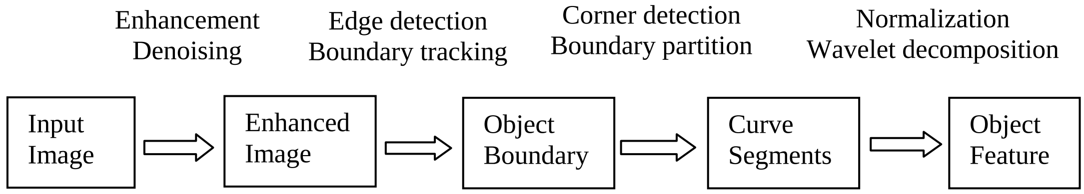


Figure 4.1 Feature extraction process

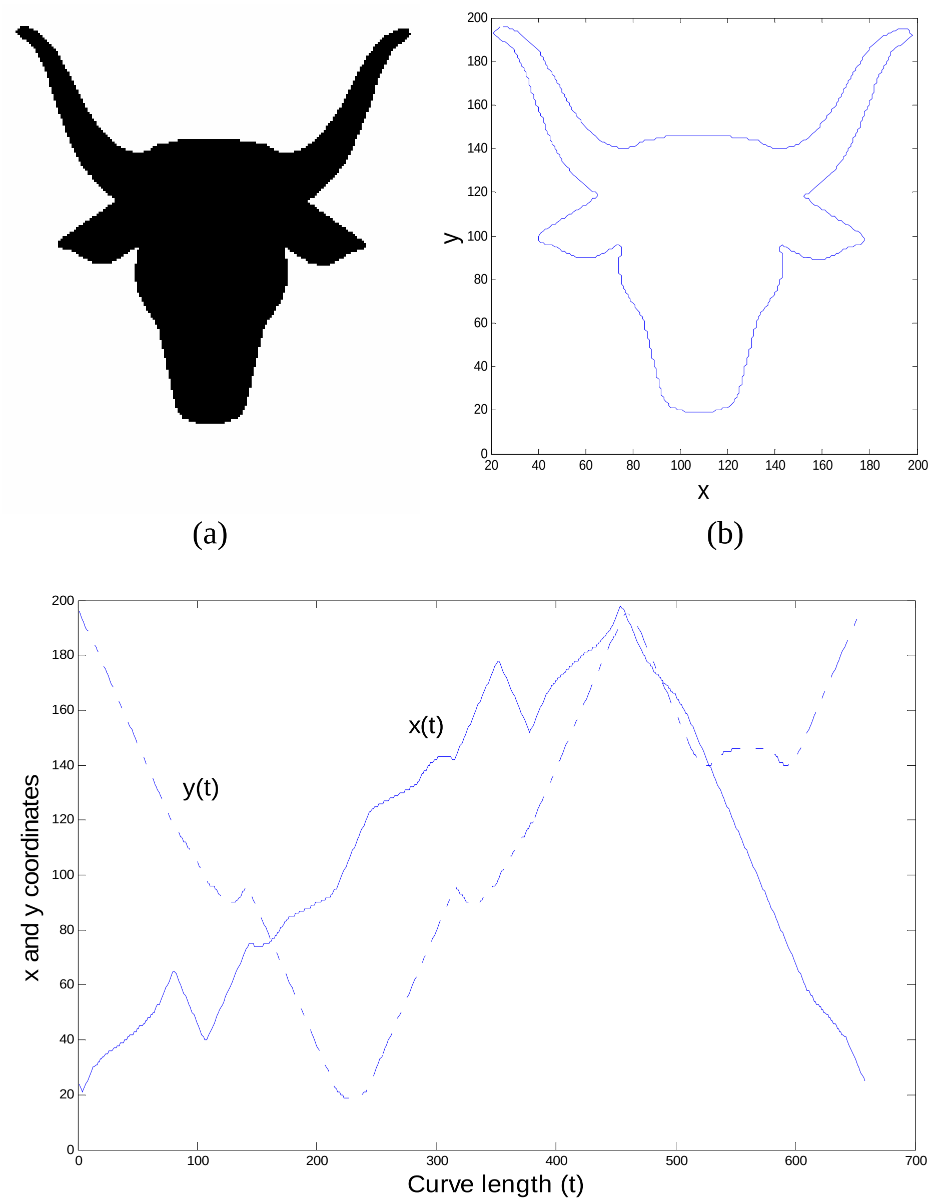
The feature extraction algorithm developed in our work is introduced in details in two chapters (Chapter 4 & 5). In this chapter, the preprocessing process procedure is briefly introduced first, followed by our proposed boundary partition algorithm including a novel wavelet based scale-invariant corner detection algorithm and boundary segmentation method which are the emphasis of this chapter.

4.2 Preprocessing

Although preprocessing tasks which include image enhancement, denoising, edge detection and boundary tracking are fundamental problems in machine vision, they are important to object recognition. They are regarded as separate problems in our work. Many researchers are dedicating themselves into these fundamental problems. Due to time constrain, the preprocessing techniques used in this research project are from existing methods, and they are briefly introduced here.

Given a digitized image $f(x, y)$ $x, y \in$ □ , we assume that input images are de-noised (Goudail, et al. 2004) and enhanced to increase the chance of success for further processing. Then the edges are detected using any edge detection method such as Canny edge detector (Canny, 1986). After edge detection, the edge pixels are obtained but in an unordered form. To serialize them, a tracing algorithm (Haig et al.

1992) is adopted to obtain an ordered list of the coordinates of points on the contour represented as $\{x(t), y(t)\}$ $t = 1, 2 \ldots n$ as shown in Figure 4.2, where n is the total number of points on the boundary.



(a) (b)

(c)

Figure 4.2 preprocessing process (a) original image $f(x, y)$ (b) the boundary of object (c) plot of $\{x(t), y(t)\}$ $t = 1, 2 \ldots n$ as a function of curve length t

4.3 Boundary partitioning

Splitting the object's boundary into segments is one of the common approaches to handle recognition of object with partial occlusion (Gorman et al. 1988). The purpose of boundary partitioning is to segment object boundary into independent segments, so that the intact segments are not affected by the partial occlusion on other portions. Therefore, 2-D shape boundary segmentation is required as a fundamental and important step in the recognition of partially occlude objects (Katzir et al., 1994).

Gorman et al. (1988) used a polygonal approximation approach to detect the vertices on the boundary, then partition it into curve segments such that each of them consists of three consecutive vertices. This approach assumes that the scale factor between the sizes of the known and unknown contours is approximately known. However, in this thesis, the scale information is assumed absolutely unknown. Moreover, polygonal approximation approach is unstable in finding the break points for non-polygonal shapes. Katzir et al. (1994) proposed a novel curve segmentation algorithm by transforming the curve into another one which intersects itself, then regard the points on the original curve corresponding to intersection points of the new curve as endpoints of segments. This segmentation algorithm is ideal for objects with smooth boundaries. However, it results in having segments with very small length for objects with sharp corners, which is undesired for this research project.

Attneave (1954) pointed out that points at which the curve bends most sharply are good partition points, because corner points are the most dominant points and are readily detected. The boundary partition algorithm proposed in this thesis resorts to the same approach. Since our recognition work involves similarity transform and partial occlusion, a corner detection algorithm which is invariant to scaling and partial occlusion is essential in order to have a consistent boundary partitioning even if when similarity transform and partial occlusion are present. However, the combination of scaling and partial occlusion makes consistent corner detection extremely difficult. The reason is that a corner is a local feature; it largely depends on the size of the support region. Unfortunately, when both scaling and partial occlusion occur, the optimal scales for corner detection can not be determined successfully because occlusion disfeatures the global image.

4.4 Literature survey of existing corner detection algorithm

An extensive literature survey has been carried out to search for a suitable corner detection algorithm from existing curve corner detection techniques. The main shortcoming of the single-scale algorithms (Pavlidis 1974, Freeman 1977, Rosenfeld 1973, Wuescher 1991, Fishler 1986, Rosenfeld 1975, Beus 1987, Sanka 1978, Anderson 1984, Cheng 1988) is that they work well only when the features of the object are of similar size and the size information is known *a priori*. However, since the object in question is assumed unknown, some researchers tried to obtain the size of the object by its area or perimeter. However, in this thesis, the object may be partially occluded, and therefore, we are unable to obtain the size information. In addition, corners on planar curves have their corresponding regions of support that may be quite different in size. Hence, it is not possible to define an optimal resolution in advance for the detection of corners. Consequently, traditional single scale corner detection algorithms tend to detect unwanted corners or miss the obvious corners in some situations. In addition, the corners detected for the same object in different scales (in case we cannot determine the size of object in scene due to partial occlusion) may not be consistent. Researchers (Asada 1986, Mokhtarian 1986, Pei 1992, Rattarangsi 1992) resorted to detect corners based on the integrated information at multiple scales instead of the information at a single scale. Asada (1986) proposed the curvature primal sketch to represent the change in boundary curvature. The primal sketch consists of various symbolic descriptions at multiple scales, such as corners, knots and ends, and is generated by convolving different size Gaussian windows with the curvature and tracking the locations of curvature discontinuities. Boundary symbolic features are detected at a different scale individually based on the *a priori* knowledge for extracting features at the corresponding best scales. The resulting representation is a multiple-scale interpretation of the boundary. However, the detection of an individual symbolic feature is still based on the signal at a single scale. As to Rattarangsi et al.(1992) proposed a new algorithm that detects corners by integrating information obtained at multiple scales. In their algorithm, the scale spaces of isolated single and double corners are first analyzed to investigate the behavior of the scale space. The scale space consists of the local modulus maxima of the differentiation of the boundary function presented at all scales. Then, the scale space is transformed into a tree. Finally, the corners are detected using a coarse-to-fine tree parsing technique. However, since many scales are required in the algorithm, this is time consuming. In addition, some false corners are still detected, e.g. arcs with small radius. The reason is that the algorithms detects corner only based on the locations of the local modulus maxima while the magnitude of the maxima is neglected.

Wavelet technique is a novel technique for corner detection. It has been an active research area since Mallat (1992) built a theoretical foundation for singularity detection. Quddus (1999) proposed a fast wavelet-based technique for corner detection. This boundary based techniques exploits the wavelet transform modulus maxima to detect corners. It is simple to implement and computationally efficient in comparison with the scale-space methods because the wavelet transform modulus maxima are only extracted on a few dyadic levels, and it makes use of the fast discrete wavelet transform which is much more computational efficient than Gaussian convolution. However, this method still detects false corners. Quddus (2002) presented yet another novel technique to detect corners using a concept of determine the natural scale using Singular Value Decomposition (SVD) of the wavelet coefficients. The SVD facilitates the selection of global natural scale in discrete wavelet transform. They define natural scale at the level associated with most dominant eigenvalue. Eigenvector corresponding to dominant eigenvalue is considered as the optimal scale. The corners are detected at the locations corresponding to modulus maxima at the optimal scale. However, the eigenvalue to determine the natural scale is calculated based on the wavelet coefficients of the entire boundary. Obviously, partial occlusion will affect the result of natural scale selection. Hence, this algorithm will under perform when occlusion occurs. Lee (1993) proposed a non-parametric algorithm for detecting and locating corners of planar curves. The algorithm is based on the multi-scale wavelet transform of the orientation of the curve which can effectively utilize both the information of local modulus maxima positions and magnitudes of the transform results. The false corner elimination is based on the intrinsic ratios of true corners derived by them. In 1995, they (Lee 1995) proposed an improved multi-scale corner detection algorithm using wavelet transform. The ramp-width of contour orientation profile, which can be computed using the transformed modulus of two scales, reveal the difference between corner and arc and is utilized in the detection of true corner points. They assume that the ramp length of the orientation function with smoothing factor of 3 of an ideal corner is 7. However, in real applications dealing with discrete image, such assumption is very hard to be fulfilled due to rotation, quantization and noise. Moreover, for our application we can not use only those perfect corners for segmentation, because the number of ideal corners are either too little or do not exist at all for some objects. We have written a computer program to simulate their algorithm, and the experimental result was not satisfied. Hua (2000) proposed a multiscale corner detection algorithm using the wavelet transform of parametric coordinates instead of the traditional contour orientation profile. A wavelet with 2 vanishing moments was used to detect the local modulus maxima of the object boundary that corresponds to corners. However, the cooperation of the local modulus maxima of the wavelet transform of the $x-$ and $y-$ coordinates is an ambiguity problem which has not been addressed by the authors. Moreover, this algorithm may not be rotation invariant, since no mathematical proof or evidence has been provided.

4.5 Proposed wavelet-based corner detection algorithm

Based on the literature review made above, we conclude that existing corner detection algorithms surveyed during the course of this thesis are not readily suitable to partition the object boundary into a consistent sequence of segments which is essential for this research project. Nevertheless, those wavelet based corner detection algorithms outperform the traditional algorithms providing better accuracy and computational efficiency. Therefore, we dedicated ourselves to develop a novel wavelet based corner detection algorithm which is capable to detection true corners even when similarity transform and partial occlusion occur.

In this section, a specially designed wavelet based corner technique which is robust to scale and partial occlusion in favor of this research project is proposed. The corner candidates detection method in the proposed wavelet based corner detection algorithm is adopted from Lee (1993, 1995) and Quddus (1999), but we proposed a novel false corners elimination method using Lipschitz exponent as a quantitative evaluation of the sharpness of corner candidates which is crucial for the overall performance of the corner detection algorithm.

The proposed corner detection algorithm consists of the 3 steps as shown in Figure. 4.3. we use the similar methodologies presented in Lee (1993, 1995) and Quddus (1999) to calculate the orientation profile and detect the corner candidates. Subsequently, we use our own novel method to eliminate false corners. The false corner elimination plays an important role in corner detection, and makes a big difference on the overall performance. In the following sections, we will briefly describe the orientation profile calculation and corner candidate detection, and then present our false corner elimination method in detail.

Figure 4.3 Corner detection process flow chart

4.5.1 Orientation profile calculation

Let $\{x(t), y(t)\}\ \ t = 1, 2...n$ represent the boundary of a planar object, where t is the arc length. The orientation is defined as:

$$\phi(t) = \tan^{-1} \frac{dy/dt}{dx/dt} \tag{4.1}$$

If the orientation profile of a boundary in discrete form is represented as the Freeman chain code, the orientation resolution is only $\pi/4$. To improve the orientation resolution, the orientation at point t is defined as

$$\phi(t) = \tan^{-1}\frac{y_{t+q} - y_{t-q}}{x_{t+q} - x_{t-q}} \tag{4.2}$$

Where q is the smoothing parameter whose value is generally chosen to be 3 (Lee 1993, Lee 1995, Mahmoud 2000).

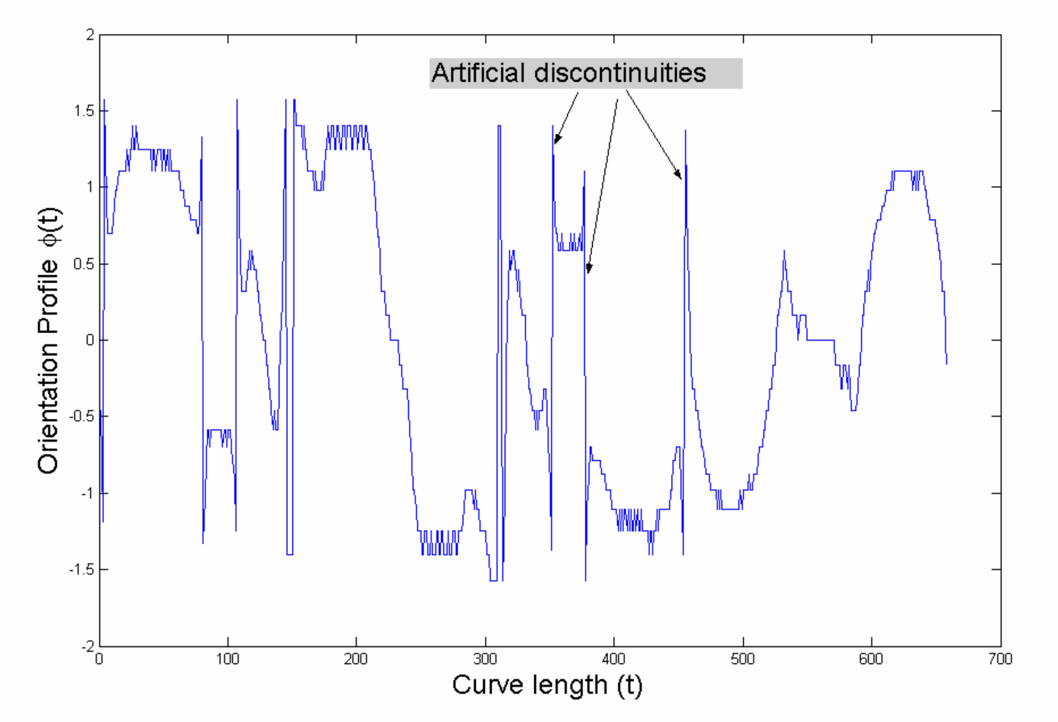


Figure 4.4 Orientation profile containing wrap-around error

The range of equation (4.2) is bounded between $-\pi/2$ and $\pi/2$. As a result, the wrap-around error occurs when $|\phi|$ exceeds $\pi/2$ and this error creates additive discontinuities as shown in Figure 4.4. To eliminate this error, an offset is added to the ϕ values to obtain the compensated true orientation profile $\phi_c(t)$

$$\phi_c(t) = \phi(t) + offset(k) \tag{4.3}$$

The $offset$ is initialized to zero, e.g. $offset(0) = 0$. It is then updated as:

$$offset(k+1) = \begin{cases} offset(k) + \pi & \text{if } \phi(t)\text{-}\phi(t-1) \geq \pi \times factor \\ offset(k) - \pi & \text{if } \phi(t)\text{-}\phi(t-1) \leq -\pi \times factor \end{cases} \tag{4.4}$$

Theoretically, the $offset$ value is changed when $\phi(t)\text{-}\phi(t-1) \geq \pi$. Practically, the offset value has to be changed when $\phi(t)\text{-}\phi(t-1)$ approaches π. Thus, the $factor$ has to be chosen to be slightly less than one. Based on our experimental simulations, a value of the $factor$ at 0.85 is a good choice. The orientation profile without wrap-around error of the bull head (Figure 4.2 a) is shown in Figure 4.5.

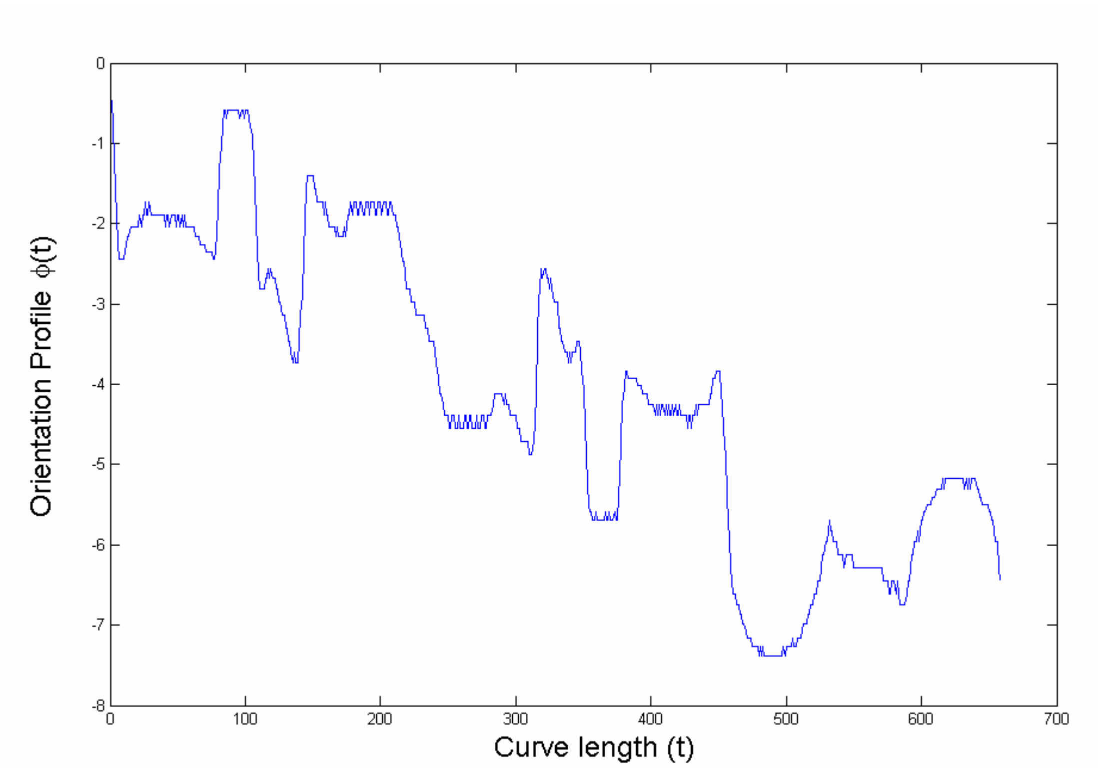


Figure 4.5 Orientation profile after offset

4.5.2 Corner candidates detection.

As illustrated in Figure 4.5, there are several 'steps' with sudden changes in amplitude which correspond to corners on the object boundary. The 'step' is a kind of singularity in mathematics. In (Mallat et al., 1992(a)), it shows that the wavelets have the capability to detect such singularities, it can also locate the position of the singularities precisely. They also proposed a quadratic wavelet as shown in Figure 4.6, which has compact support and such that the wavelet transform can be computed with a fast algorithm.

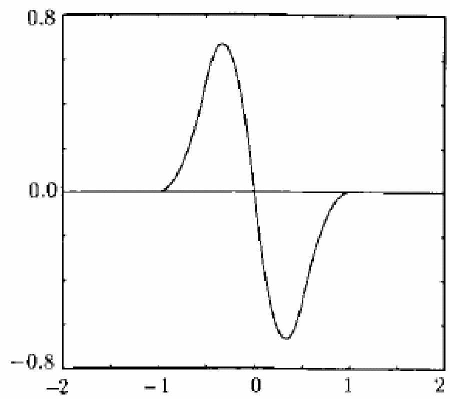


Figure 4.6 Quadratic spline wavelet

The orientation profile of the object boundary is transformed using the Quadratic wavelet into three dyadic scales. The local extrema on each scale are detected, both the location and amplitude are recorded. A point is set as local extremum if its modulus is greater than its nearest neighborhood. In practice, its modulus must be greater than a properly chosen threshold as well to eliminate some insignificant local modulus maxima caused by noise. Since the wavelet is the first derivative of a smoothing function, the wavelet transform modulus maxima are located where the signal has sharp transitions. Figure 4.7 is the dyadic wavelet transform of the orientation profile in Figure 4.5, c omputed with the wavelet shown in Figure 4.6(a), where the local modulus maxima are indicated by stars '*'.

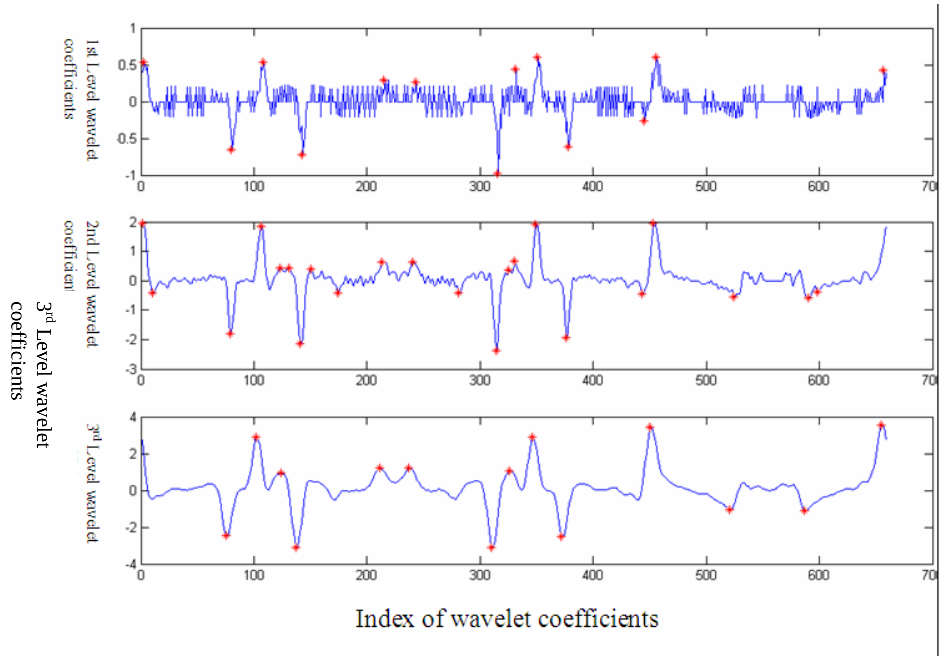


Figure 4.7 Wavelet transform of the function shown in figure 4.4, where the local modulus maxima is indicated by '*'

After the local modulus maxima of the wavelet transform of three dyadic scales are detected. A matching process is required to determine cross-scale correspondences of the extrema. A coarse-to-fine tracking process similar to the one in reference (Witkin, 1983) is adopted. Sine the scale is discrete, the correspondences can not be determined precisely. However, by using the three following criteria, the local modulus maxima can be linked with negligible probability of mis-match.

i. The distance of the local modulus maxima at two adjacent scales is less than a proper chosen threshold.

ii. The local modulus maxima at two adjacent scales have the same sign, e.g. both should be positive or negative.

iii. Only local modulus maxima that manage to propagate through the three scales are considered as eligible candidates.

The linking of the extrema (indicated by “*”) is shown in Figure 4.8. The location of the local modulus maxima at the finest scale stands for the real locations of corner candidates, because the support of the finest scale wavelet function is small compared with the corner distance. The detected corner candidates of the “bull head” is shown in Figure 4.9.

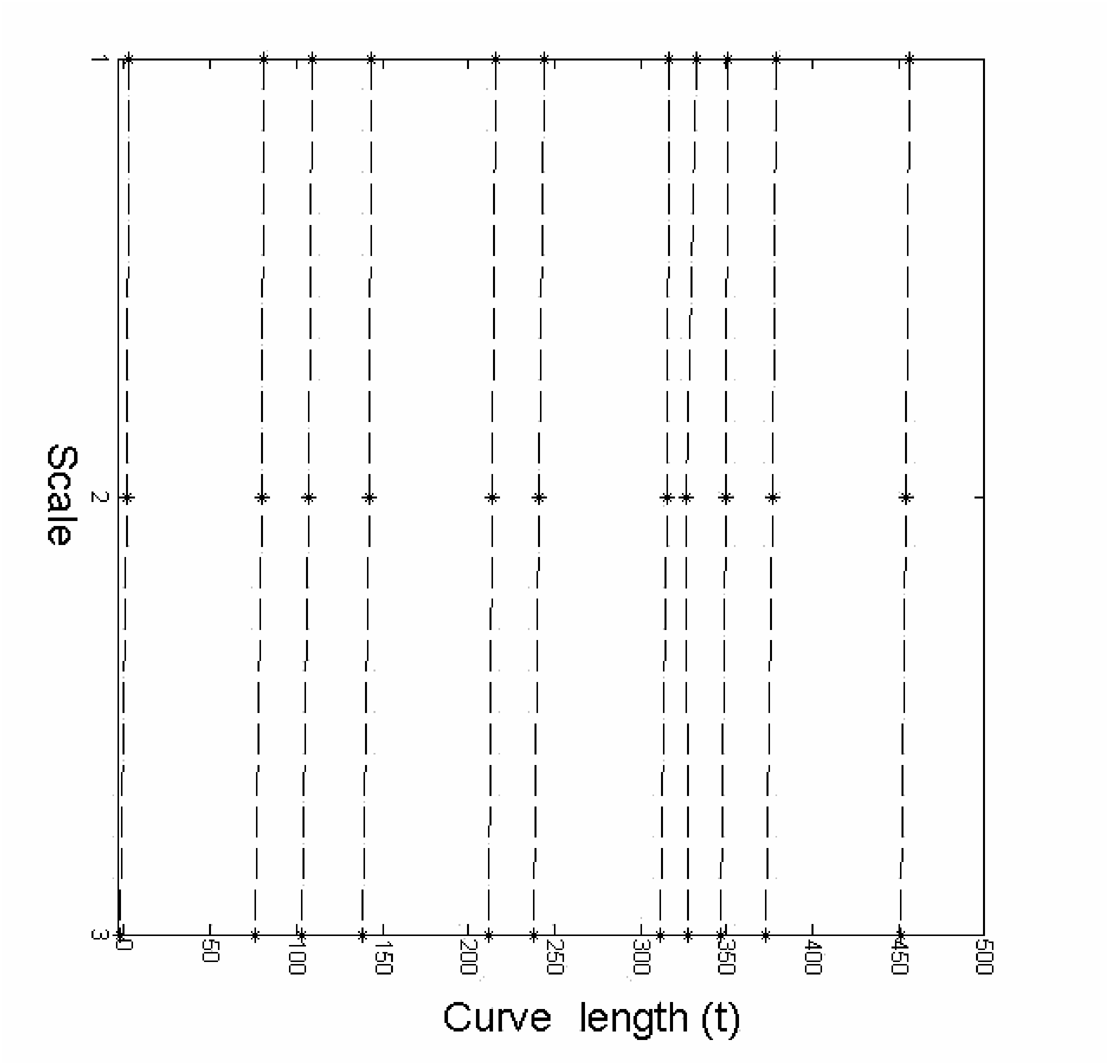


Figure 4.8 The linking of local extrema

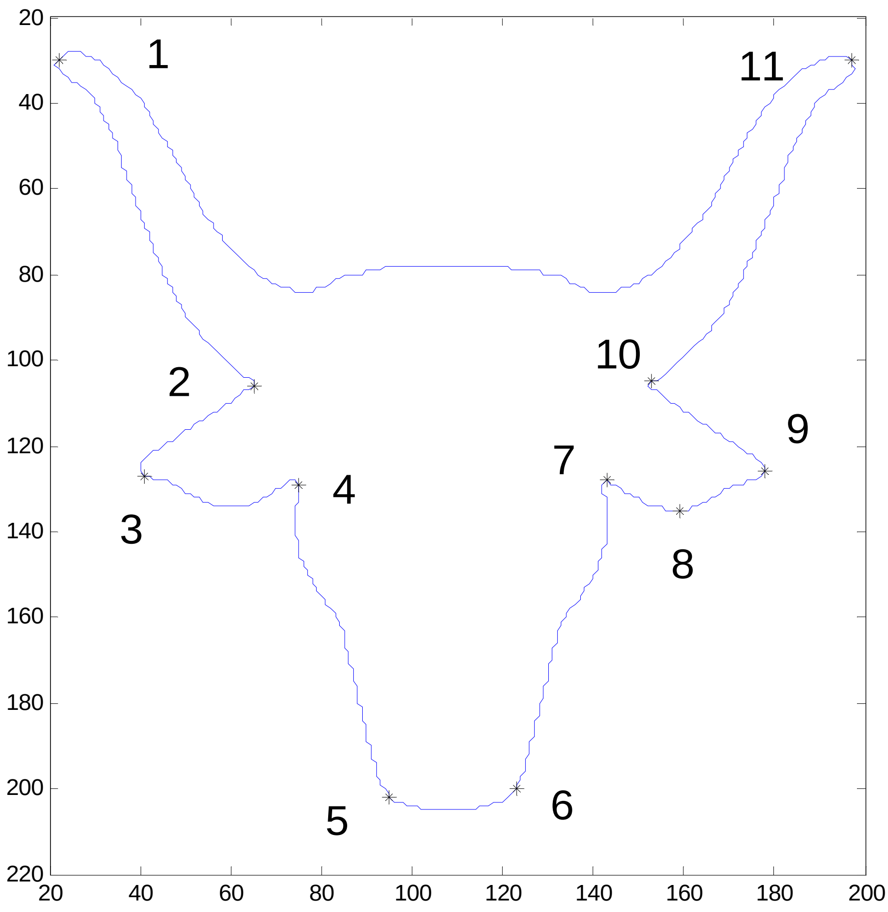


Figure 4.9 Corner candidates

4.5.3 False corner elimination using Lipschitz exponent.

Although only significant local modulus maxima are selected, and only which manage to propagate through the three scales are considered as eligible candidates, false corners still cannot be totally avoided. The reason is that it is impossible to set a hard-threshold to separate the local modulus maxima created by real corners from those caused by noise and arcs for objects with unknown scale and arbitrary nature. Quddus (1999) addressed this problem by normalizing the wavelet coefficients with respect to the maximum peak at the level. However, this normalization does not work well when there is partial occlusion, because it may affect the maximum peaks. Quddus (2002) tried to solve this problem in an alternate way by detecting corners using a concept of determining the natural scale using Singular Value Decomposition (SVD) of the wavelet coefficients. Again, the natural scale determination is affected by the partial occlusion. Lee et al. (1993, 1995) used the ratio of the inter-scale local modulus maxima to eliminate the false corners. However, their algorithm intends to detect only ideal corners formed by two straight lines. In this thesis, a mathematical term called ‘Lipschitz exponent’ is used to serve as a quantity measure of the corner’s sharpness to eliminate false corners. In mathematics, local regularity is often measured with Lipschitz exponents. If a function is smooth (nonsingular) at a particular point, the Lipschitz exponent of this point can describe the smoothness of the function at this point.

Mallat (1992(b)) provided an efficient algorithm to measure that local Lipschitz regularity from the wavelet transform modulus maxima over scales. A function is Lipschitz α at x_0, if and only if there exists a constant A such that at each wavelet coefficient modulus maxima $Wf(s,x)$ is

$$\left|Wf(s,x)\right| \le As^{\alpha} \tag{4.5}$$

where x is within the cone of x_0 defined by $\left|x - x_0\right| \le Cs$, and C is a constant, s is the scale.

Taking a logarithm on the both side of Equation (4.5), we obtain

$$\log_2 \left|Wf(s,x)\right| \le \log_2(A) + \alpha \log_2(s) \tag{4.6}$$

The Lipschitz exponent α therefore can be evaluated by the slope of $\log_2|Wf(s,x)|$ as a function of $\log_2(s)$, if we have the wavelet transform local maxima at two scales s_1 and s_2, the Lipschitz exponent can be examined by equation (4.7). In practice, we have the local maxima at more than two scales, the Lipschitz exponent is evaluated by approximating the slope of the $\log_2|Wf(s,x)|$ as a function of $\log_2(s)$ by minimum square error (MSE) method. The plot of the $\log_2|W\Phi_c(s,k)|$ as a function of $\log_2(s)$ of corner candidates 1 and 5 (Figure 4.9) are shown in Figure 4.10.

$$\alpha = \frac{\log_2|Wf(s_2,x)| - \log_2|Wf(s_1,x)|}{\log_2(s_2) - \log_2(s_1)} \tag{4.7}$$

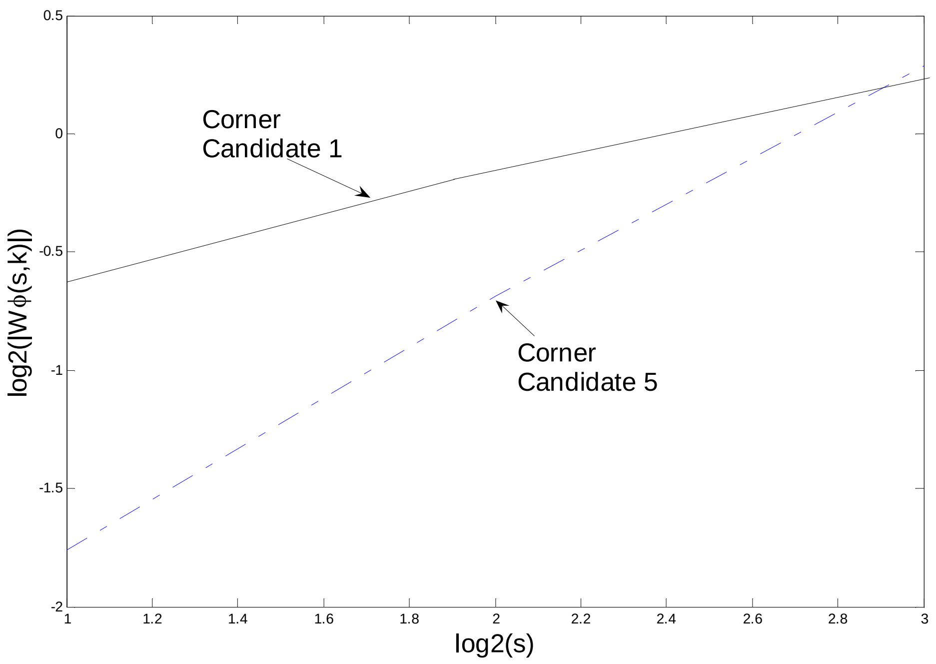


Figure 4.10 The decay of the $\log_2|W\Phi_c(s,k)|$ as a function of $\log_2(s)$ of corner candidates 1 and 5 as shown in Figure 4.9

The local Lipschitz exponents of the orientation profile are smaller than 1 as the orientation profile is continuous, therefore, it is sufficient to use a wavelet with only one vanishing moment to evaluate the local Lipschitz exponent. The Quadratic wavelet proposed by Mallat (1992b) as shown in Figure 4.6 is a good choice, because it has compact support and fast computational algorithm.

The relevance of Lipschitz exponent to evaluate the corners is proven in two points: Firstly, the Lipschitz exponent evaluation is correspond to the human perception of saliency of corners; Secondly, the similarity transform invariance of the Lipschitz exponent is mathematically proven.

The perceptual saliency of a corner is closely related to two geometrical terms: the acuteness and the sharpness. The more acute the angle is, the more salient the corner will be. The sharper the angle's tip is, the more salient the corner will be. We evaluate the relationship of Lipschitz exponent with this two terms by performing a computer simulation. Supposing a sharp corner $f(t) = \{x(t), y(t)\}$ with angle θ, is convoluted by a Gaussian function $g(t)$, where the width of the Gaussian function is controlled by σ.

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2 / 2\sigma^2} \tag{4.8}$$

The Gaussian functions with $\sigma = 2, 4, 8$ are shown in Figure 4.11. Figure 4.12 illustrates the effect of two corners with angle at 40 and 140 degrees convoluted with Gaussian function with $\sigma = 2, 4, 8$ respectively. It visually demonstrates the statement made above: The smaller the angle is and the smaller the σ is, the more salient the corner will be. Figure 4.13 shows the simulation result of how the Lipschitz exponent is correlated with the angle θ and the smooth factor σ of a corner. The simulation result illustrates that the Lipschitz exponent is a monotonic increasing function of the angle of corner θ and the smooth factor σ. Therefore, Lipschitz exponent is a unified measure of the perceptual saliency of the corner.

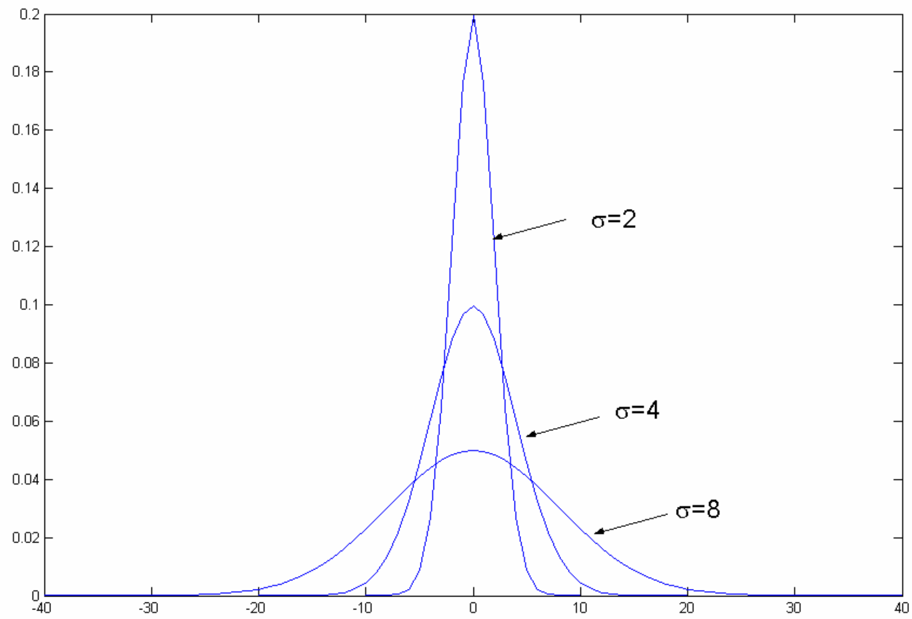
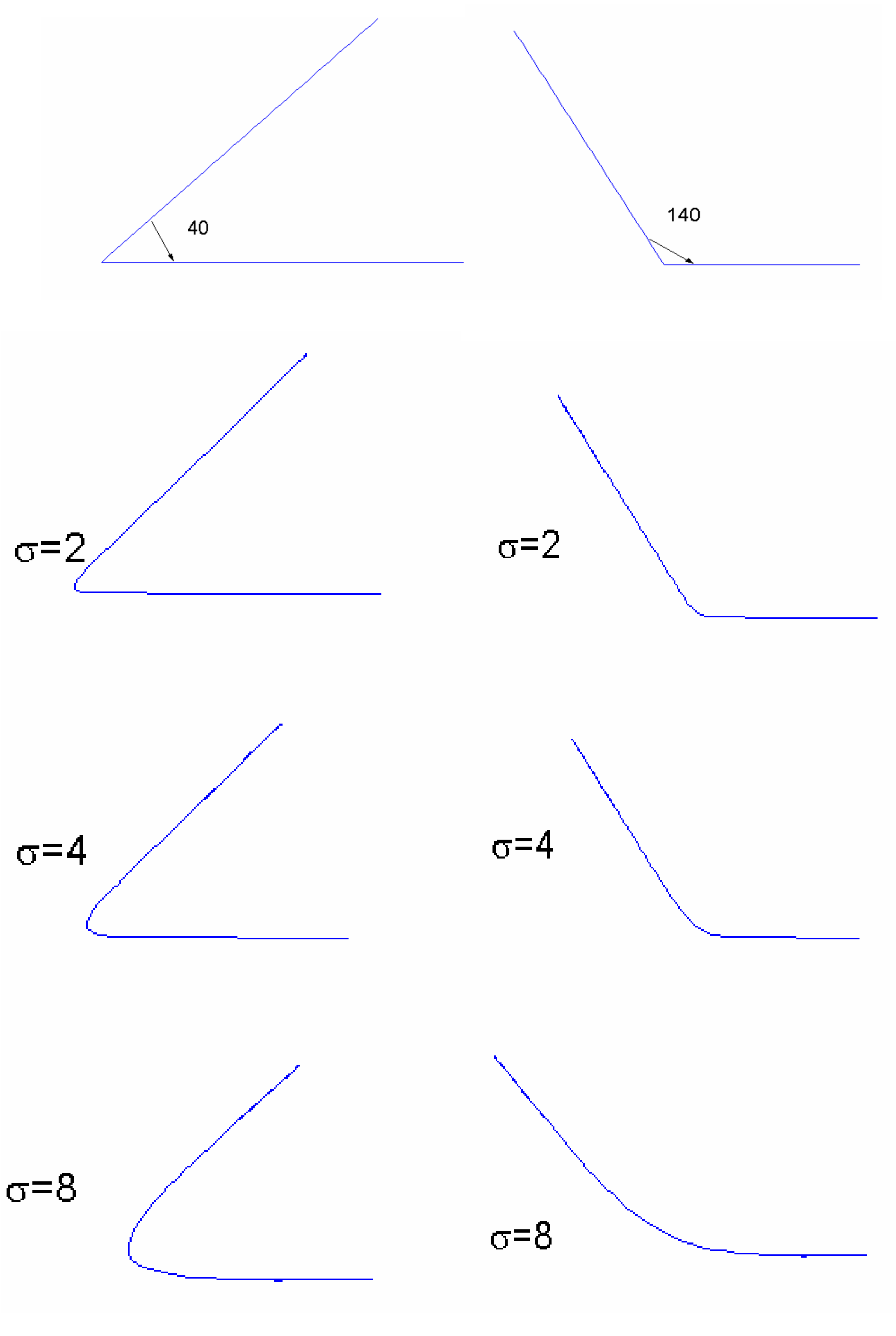


Figure 4.11 Gaussian Functions with $\sigma = 2,4,8$



(a) (b)

Figure 4.12 (a) Corner of angle 40 degree convoluted by Gaussian Functions with $\sigma = 2,4,8$
(b) Corner of angle 140 degree convoluted by Gaussian Functions with $\sigma = 2,4,8$

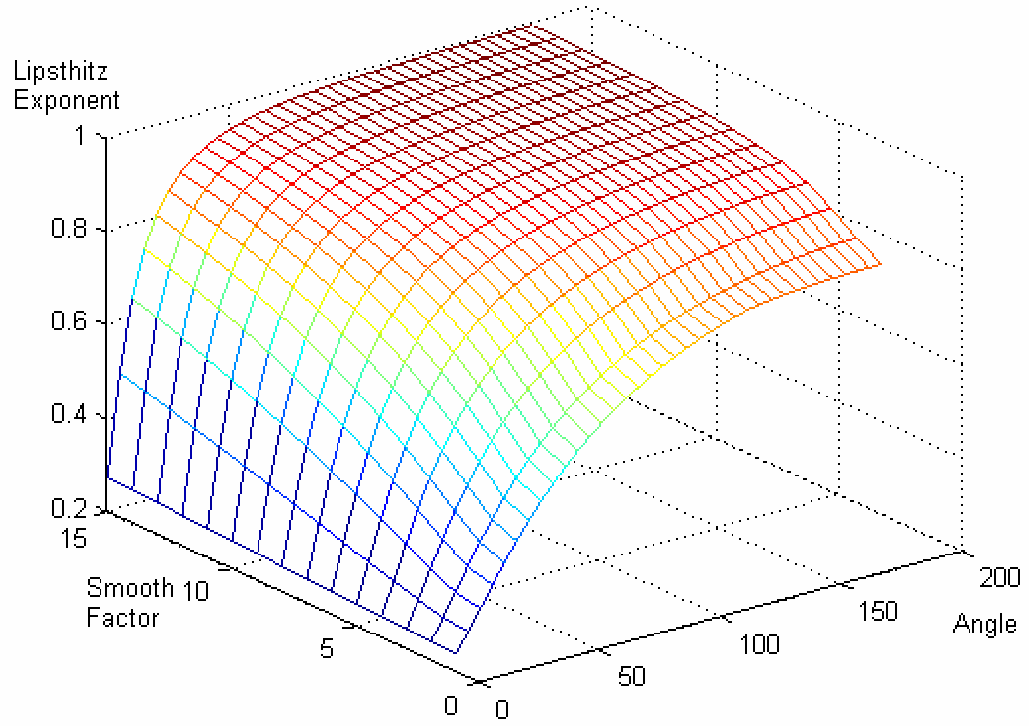


Figure 4.13 Relationship of Lipschitz Exponent with the angle of corners and the width of Gaussian kernel for smoothing

Since the proposed corner detection algorithm is supposed to be similarity transform invariant, to evaluate the corner, the Lipschitz exponent must be invariant to similarity transform as well. Obviously, the orientation profile $\phi_c(t)$ of an object boundary is translation and rotation invariant. For an object boundary scaled by ratio r , the coordinates of the scaled object is represented by $\{X(k),Y(k)\}=\{r*x(k/r),r*y(k/r)\}\quad k=1,2...N$, where $\{x(t),y(t)\}\quad t=1,2...n$ is the boundary of the original object, and $N=rn$. Then the orientation profile of the scaled object can be approximated as $\Phi_c(k)\approx\phi_c(k/r)$ (Note that it is not exactly equal to $\phi_c(t)$ since a fixed step three is used to obtain the orientation profile). The wavelet transform of $\Phi_c(k)$ in our work is derived as follows:

$$\begin{aligned} W\Phi_c(s,k) &= \int \Phi_c(k)\psi_s(k)dk \\ &= \int \phi_c(k/r)\psi_s(k)dk \\ &= r\int \phi_c(u)\psi_s(ru)du \\ &= \sqrt{r}\int \phi_c(u)\sqrt{\frac{r}{s}}\psi(\frac{ru}{s})du \\ &= \sqrt{r}\int \phi_c(u)\psi_{s/r}(u)du \\ &= \sqrt{r}W\phi_c(s/r,u) \end{aligned} \tag{4.9}$$

By substituting equation (4.9) into (4.7), the local Lipschitz exponent of the Φ_c can be derived as shown in Equation (4.10). This implies that the scaling of object would not change the local Lipschitz exponent.

$$\begin{aligned} \alpha(\Phi_c) &= \frac{\log_2 \left|W\Phi_c(s_2,k)\right| - \log_2 \left|W\Phi_c(s_1,k)\right|}{\log_2(s_2) - \log_2(s_1)} \\ &= \frac{\log_2 \sqrt{r}\left|W\phi_c(s_2/r,u)\right| - \log_2 \sqrt{r}\left|W\phi_c(s_1/r,u)\right|}{\log_2(s_2) - \log_2(s_1)} \\ &= \frac{\log_2 \left|W\phi_c(s_2/r,u)\right| - \log_2 \left|W\phi_c(s_1/r,u)\right|}{\log_2(s_2/r) - \log_2(s_1/r)} \\ &= \alpha(\phi_c) \end{aligned} \tag{4.10}$$

From the above computer simulation and mathematical proof, the Lipschitz exponent has been shown to be an appropriate measure of the saliency of the corner. Therefore, by thresholding the corner candidates' Lipschitz exponent with a proper chosen threshold, the false corners can be eliminated regardless of similarity transformation. The Lipschitz exponents of the corner candidates of the "bull head" shown in Figure 4.9 are listed in Table 4.1. We choose 0.6 as the threshold, hence, the false corners 5,6 and 8 are eliminated. The true corners are shown in Figure 4.14.

Table.4.1 The Lipschitz exponent of corner candidates and the evaluation result

Corner Candidate	Lipschitz Exponent	Result
1	0.2885	True corner
2	0.4056	True corner
3	0.3455	True corner
4	0.4423	True corner
5	*0.7014*	*False corner*
6	*0.7293*	*False corner*
7	0.4432	True corner
8	*0.8023*	*False corner*
9	0.3582	True corner
10	0.4234	True corner
11	0.2754	True corner

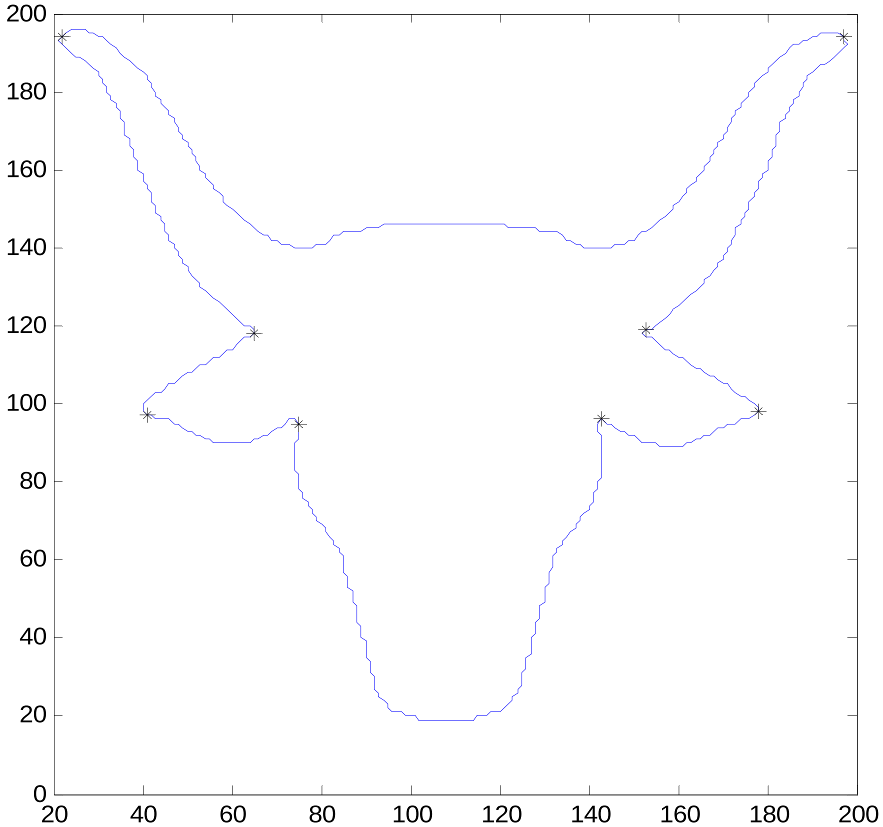
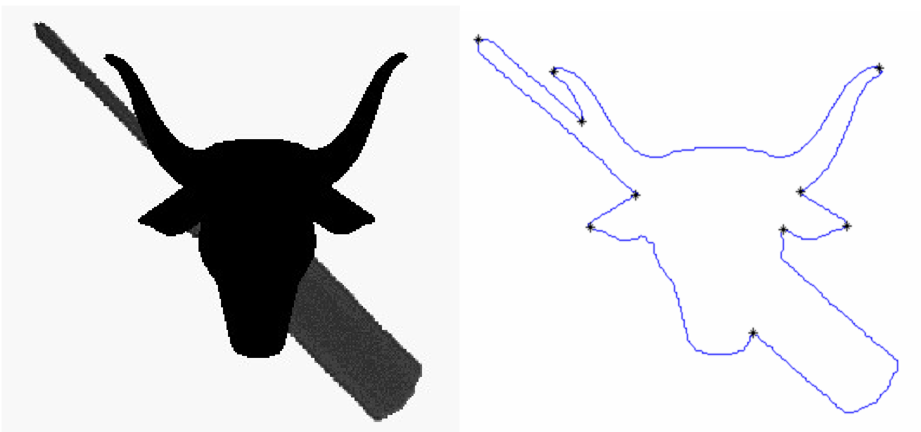


Figure 4.14 True corners after false corner elimination

Note that the Lipschitz exponent is measured by the decay of wavelet transform local maxima. Each wavelet transform local maximum depends only on its support region, therefore the occlusion on other region would not affect this wavelet transform. Hence, the Lipschitz exponent is robust to partial occlusion. The corner detection result of a bull head scaled by 1.5 times and occluded by a screwdriver is shown in Fig. 4.15. Extensive experiments have been done, and the results (which will be shown in Chapter 7) show that our proposed corner detection algorithm is invariant to similarity transform and robust to partial occlusion.



(a) (b)

Fig. 4.15. (a) Bull head scaled by 1.5 times occluded by screwdriver (b) Corner detection result

4.6 Boundary partitioning using detected corners

Having the corners detected by the proposed corner detection algorithm in hand, we have several options to partition the object boundary. Such as:

- The most intuitive way is to consider the portion between two consecutive corners as a segment. However, the segments segmented in this way do not contain corner information which is very important to object recognition, as the corners reflect important features in wavelet domain. Hence, segment contour in this manner will make the representation incomplete.
- Gorman et al.(1988) partition the boundary in the way that each segment consists of three consecutive corners. However, a successful match of segment requires three consecutive intact corners. To recognize a partially-occluded object, such requirement is too stringent to be fulfilled. Hence, partitioning in this way may lead to erroneous results in recognition. For instance, the wrench in Figure 4.16 does not have 3 consecutive corners when overlapped by the pliers. It wound not, therefore, be recognized as wrench.

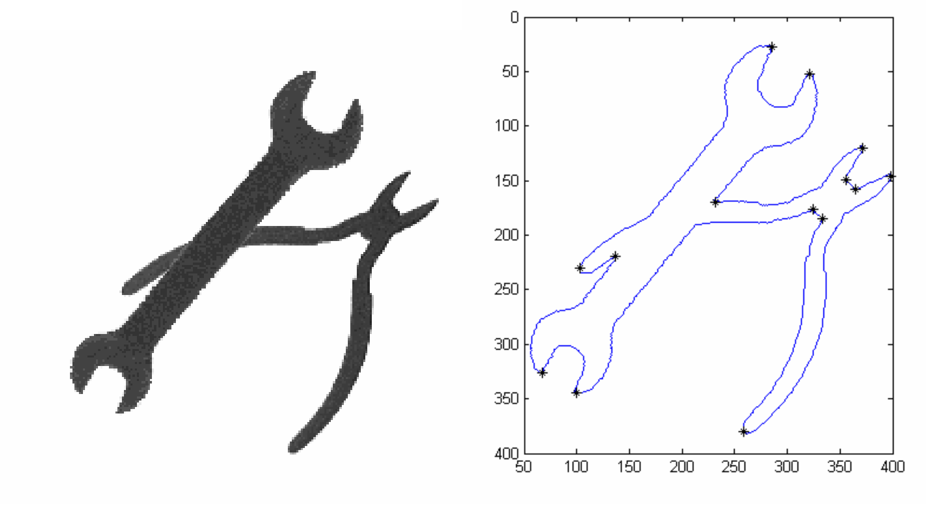


Figure 4.16 wrench overlapped by pliers

In order to overcome the drawbacks of the above two methods, we propose a new segmentation algorithm that can partition the object boundary into segments containing corner information, and each of them needing only two consecutive corners. We simply lengthen the curve segment between two consecutive corners by a length which is proportional (which we use 1/8 in the program) to the length between these two consecutive corners. Assume that we have two consecutive corners on the object boundary, represented by (x_{t_1}, y_{t_1}) and (x_{t_2}, y_{t_2}), where t_1 and t_2 are the indices of the two corners. The length (number of points on the boundary) between these two corners is $t_2 - t_1$. The segment formed using these two corners is defined by the following start and end point:

- The beginning point of the segment is $(x_{round(t_1-(t_2-t_1)/8)}, y_{round(t_1-(t_2-t_1)/8)})$, where *round* is to round $t_1 - (t_2 - t_1)/8$ to the nearest integers.
- And the end point is $(x_{round(t_2+(t_2-t_1)/8)}, y_{round(t_2+(t_2-t_1)/8)})$.

The boundary partitioning result of the bull head shown in Figure 4.2(b) is shown in Figure 4.17. The figure demonstrates that each segment contains two corners on both sides. Moreover, the proportional extension retains the size information of each segment so that make the later normalization process feasible.

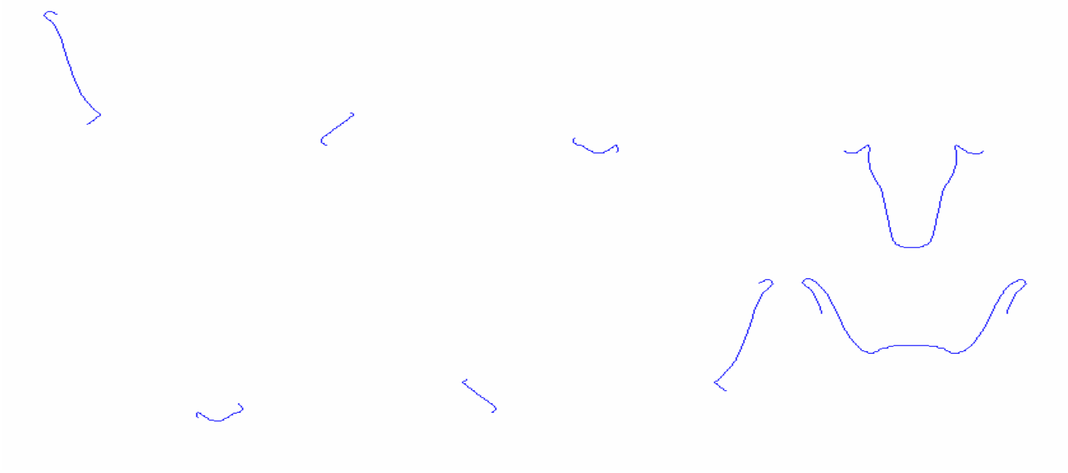


Figure 4.17 Extracted partitioned segments of Figure 4.2(b)

Chapter 5

Object Feature Extraction

5.1 Introduction

Feature extraction is the core of pattern recognition; robust and efficient features are essential to a successful pattern recognition algorithm. Existing object recognition algorithms addressing the partial occlusion problem use various kinds of object representation including dominant points, polygonal approximation, curve segment representation, local curve moment, Fourier descriptors, wavelet descriptors, etc. Despite the success of these existing representation methods, they all have drawbacks and limitations in some applications. For example, dominant points are insufficient to form a complete integrated representation of the object. It is not possible to approximate objects which have non-polygonal shape invariant to similarity transform using Polygonal approximation. Fourier descriptors are not localized in spatial domain, a local variation of the shape can affect all Fourier coefficients. A thorough literature survey of existing object representation techniques has been given in Chapter 2, the merits and drawbacks have been discussed.

In this chapter, the proposed novel wavelet-based object representation algorithm is presented. The proposed object representation consists of the wavelet representation of every individual segments and the relative position between them. To extract the wavelet representation of each segment, it is first normalized to eliminate the similarity transformation effect. Then, it is decomposed into multiple scales using wavelet transform. Finally, the insignificant coefficients are removed to make the representation more robust to noise and more compact. The relative scale and position between segments are also recorded. The proposed object representation algorithm is examined according to the evaluation criteria set in Section 1.4.

5.2 Curve Segment Normalization

The object in the scene may have different orientation, scale, and position. In order to make the object representation invariant to similarity transformation, each elemental representation of curve segment must be similarity transform invariant as well. Therefore, normalization is performed on each segment prior to feature extraction. Each segment undergoes the following normalization procedures:

i. To make the segment translation invariant, the curve segment represented by $[X, Y]$ $X = \{x_0, x_1, x_i ... x_m\}, Y = \{y_0, y_1, y_i ... y_m\}$ are shifted so that the start point (x_0, y_0) of the segment is positioned on the origin, the coordinates of curve segment after translation is changed according to equation (5.1):

$$[X_{trans}, Y_{trans}] = [X - x_0, Y - y_0] \tag{5.1}$$

Figure 5.2 shows the plot of the translation version of a curve of the bull head shown in Figure 5.1.

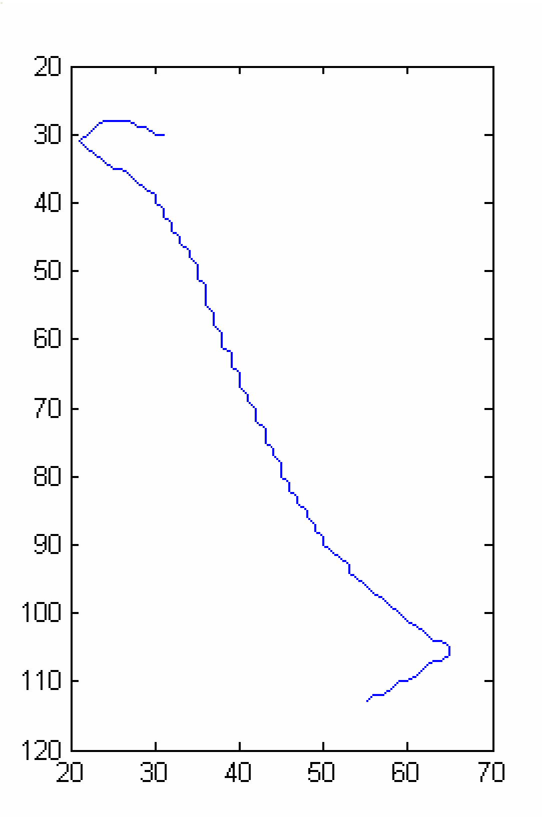


Fig. 5.1 Plot of a curve segment of the bull head

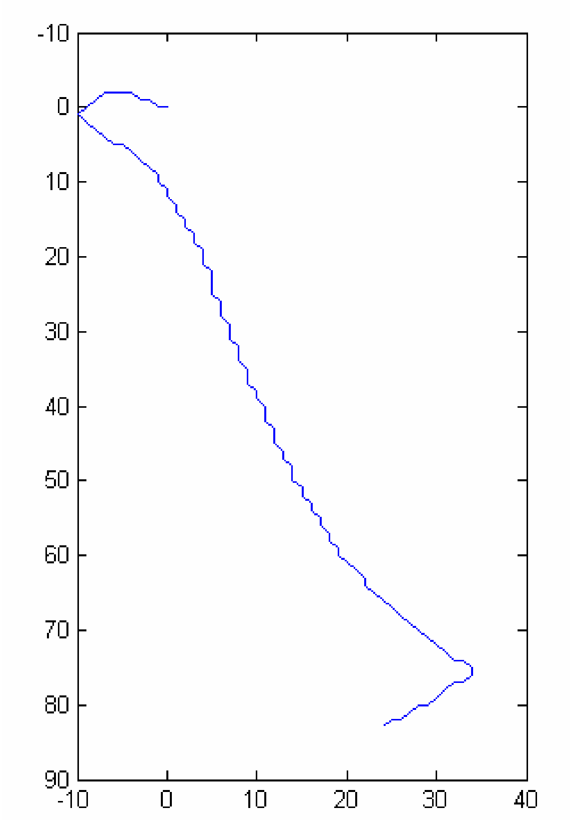


Fig. 5.2 Plot of the translated curve segment

ii. To make the segment rotation invariant, the curve segment is rotated by an angle θ so that the line joining the start and end points is along x axis as shown in equation (5.2), where θ is the angle between the said line and x axis given by equation (5.3).

$$[X_{rotate}, Y_{rotate}] = [X_{trans}\cos\theta + Y_{trans}\sin\theta, Y_{trans}\cos\theta - X_{trans}\sin\theta] \tag{5.2}$$

$$\theta = \tan^{-1}\left(\frac{y_m - y_0}{x_m - x_0}\right) \tag{5.3}$$

The plot of the rotated version of the curve segment after translation is shown in Figure 5.3.

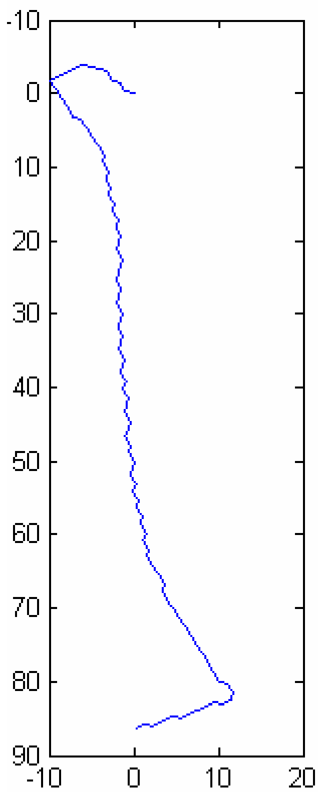


Fig. 5.3 Plot of the rotated curve segment after translation

iii. To make the segment scale invariant, the curve segment is normalized so that the distance between the start point to end point of the segment is 1 by scaling it by equation(5.4).

$$[X_{norm}, Y_{norm}] = [X_{rotate}, Y_{rotate}]/L \tag{5.4}$$

where $L = \sqrt{(x_m - x_0)^2 + (y_m - y_0)^2}$ is the distance between the start point to the end point.

The curve segment after size normalization is shown in Figure 5.4.

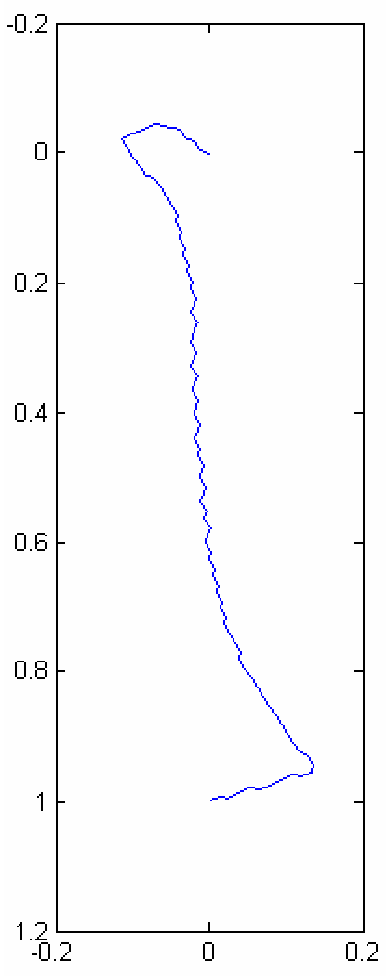


Fig. 5.4 Plot of the scaled curve segment after rotation and translation

iv. As described in Chapter 3, the fast wavelet transform of a discrete signal is done by convolution followed by down-sampling. Therefore, to facilitate the dyadic wavelet decomposition in the later process, the curve segment is resampled into number of power-of-two integer 2^n points by linear interpolation, where $2^{n-1} \leq m < 2^n$, and m is the original number of points of this segment. The curve segment after resampling is represented as $[X_{resa}, Y_{resa}]$.

The (x_0, y_0), θ and L of each segment are also recorded as part of the object representation to retain the spatial information of the curve segment.

5.3 Wavelet Decomposition

Multiresolution analysis strategy has been found to be very useful in pattern recognition areas, it can represent signal (object) by multiple resolutions so that the object can be matched in hierarchical order. Scale-space approach is one such strategy which represents signal at multiple resolutions by convolving the signal by Gaussian functions with different standard deviations (Witkin 1993, Lim et al. 1995(a), Lim et al. 1995(b)). However, since the Gaussian function is not an orthonormal basis, the multi-resolution representation generated by scale-space filtering has heavy redundancy. Researchers try to make the representation more compact by extracting only local maxima and (or) zero-crossings as the representation (Mokhtarian et al. 1986, 1992, 1995), however, such representation is still incomplete.

Wavelets lead to a powerful new approach to signal processing and analysis called Multiresolution theory (Mallat, 1989(a)). Multiresolution theory incorporates and unifies techniques from a variety of disciplines, including subband coding from signal processing, quadrature mirror filtering from digital speech recognition, and pyramidal image processing. A signal can be represented compactly and completely by wavelet transform using an orthogonal or bi-orthogonal wavelet basis. Moreover, the transient can be detected and represented readily using a well-chosen wavelet basis, the amplitude and location information can also be retained.

The proposed object representation algorithm represents each curve segment by the decomposed scaling and wavelet coefficients of the normalized x and y coordinates of the segment. The scaling and wavelet coefficients can represent the original signal uniquely, compactly and faithfully. In the following section, the wavelet decomposition is explained in detail, and the properties of the proposed representation are also discussed.

5.3.1 Level of decomposition

The curve segments of an object boundary after normalization may have different number of points depending on their original curve length. In order to represent the segment at a desired resolution and facilitate feature matching, the normalized x and y coordinates of the segment are decomposed according to the number of points on the curve segment. Suppose that the number of points on the normalized curve segment is 2^n, the normalized x and y functions of each segment are decomposed by $n-4$ level discrete wavelet transform, so that the number of the coarse level scaling and wavelet coefficients are fixed at 16 regardless of the length of original data as shown in Figure 5.5, where $c_{n-i}(X,Y)$ and $d_{n-i}(X,Y)$ are the scaling and wavelet coefficients of the normalized x and y coordinates on i th level, respectively, and 2^{n-i} is the number of coefficients at i th level. In such a way, the wavelet transform provides a natural Multiresolution representation, where small curve segment is represented with fewer levels while larger curve segment is represented with more levels. This feature is highly desired in object representation.

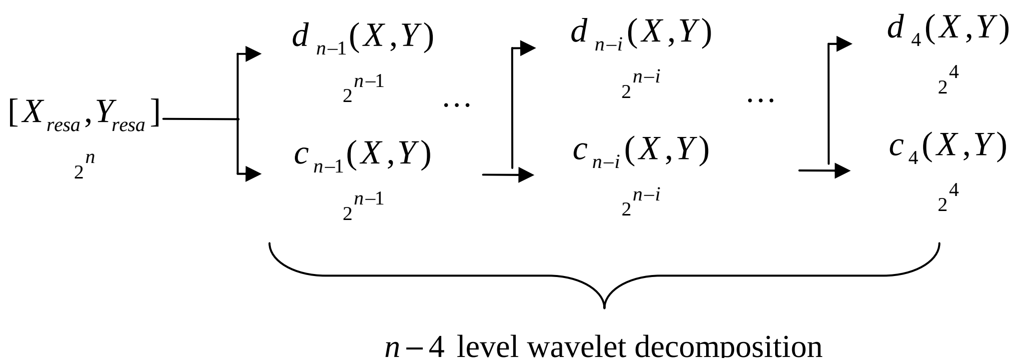


Fig. 5.5 Wavelet decomposition of the coordinates of the curve segment

5.3.2 Wavelet basis selection

As mentioned in Chapter 3, an overcomplete wavelet transform increases the amount of data required to represent the function. Therefore, an orthogonal or biorthogonal wavelet basis is desired for compact representation. Compared with orthogonal wavelet basis, biorthogonal wavelet basis are symmetric, hence possesses the linear phase property. Note that orthogonal wavelets are not symmetric except Haar wavelet. However, Haar wavelet lacks smoothness. The biorthogonal DWT requires two scaling vectors and two wavelet vectors rather than one each, but this does not increase the computational burden of the process.

Wavelet, like the quadratic spline wavelet, which is the first derivative of cubic spline, is suitable for detecting and representing abrupt changes as we used it for corner detection in chapter 4. But here the wavelet transform is applied on the x and y coordinates, and the change of the coordinates is continuous. Therefore, there would not be any sharp variation on the wavelet coefficients if we use a wavelet with only one vanishing moment. In order to represent the position and amplitude of the corners, a wavelet with vanishing moment of 2 is used. Generally, the larger is the support of the wavelet, the better will be the approximation power. However, more computational effort will be needed and the localization property of the wavelet will be poorer. The converse is true for a wavelet with a smaller support. Therefore, we need to select a wavelet base to compromise the approximation power by the localization property (computational complexity). After trying many wavelet bases, we chose bior2.4 (Daubechies, 1992) wavelet as the analysis wavelet for feature detection as it is a good compromise for our application. The decomposition together with reconstruction scaling together with wavelet functions and their corresponded filters are shown in Figure 5.6. In fact, during discrete wavelet decomposition, what we need are only the decomposition low-pass and high-pass filters. The low-pass filter $h_0(n)$ and high-pass filter $h_1(n)$ of bior2.4 are

$$\begin{aligned} h_0(n) &= (\frac{3}{128}, -\frac{3}{64}, -\frac{1}{8}, \frac{19}{64}, \frac{45}{64}, \frac{19}{64}, -\frac{1}{8}, -\frac{3}{64}, \frac{3}{128}) \\ h_1(n) &= (\frac{1}{4}, -\frac{1}{2}, \frac{1}{4}) \end{aligned} \tag{5.5}$$

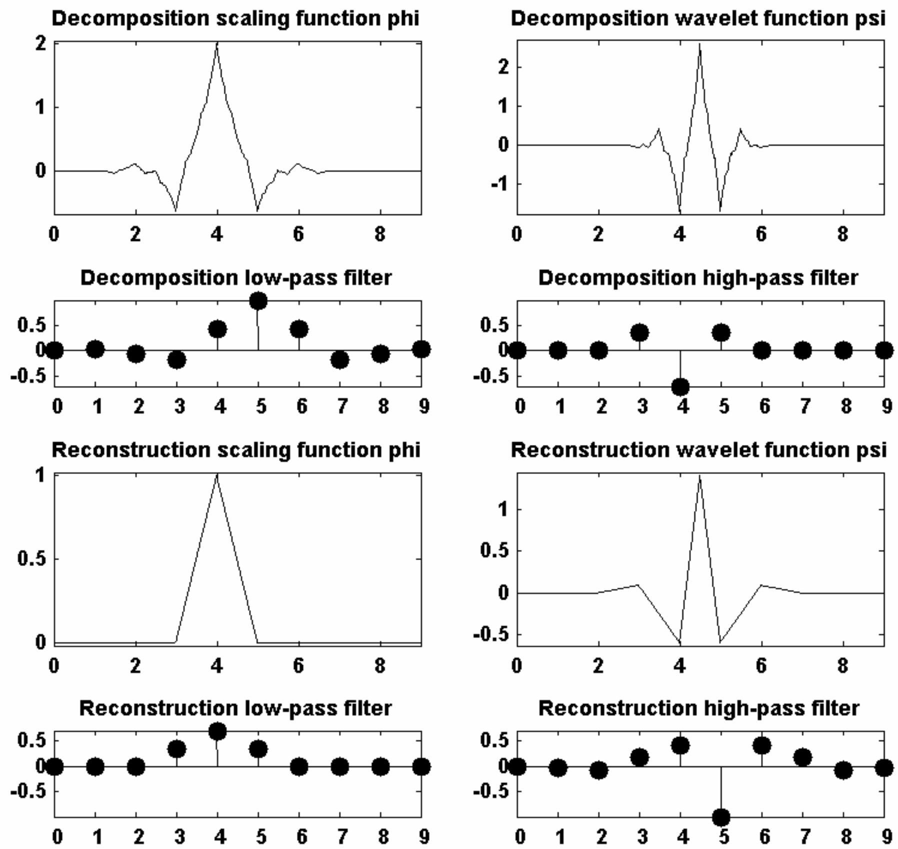
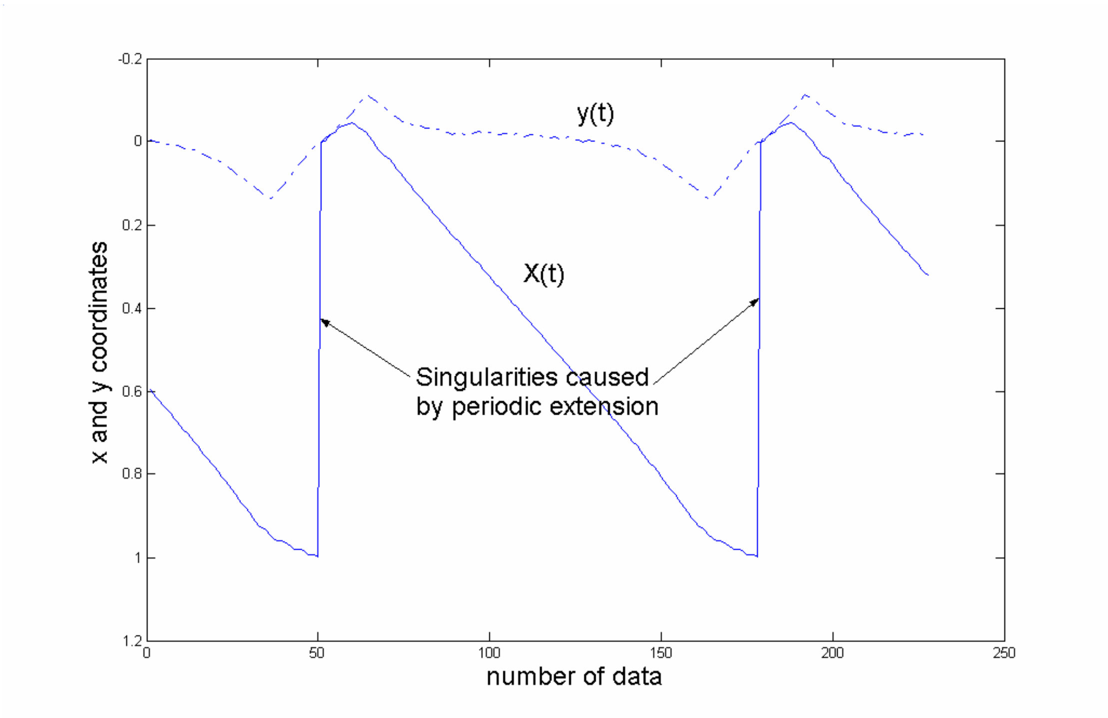


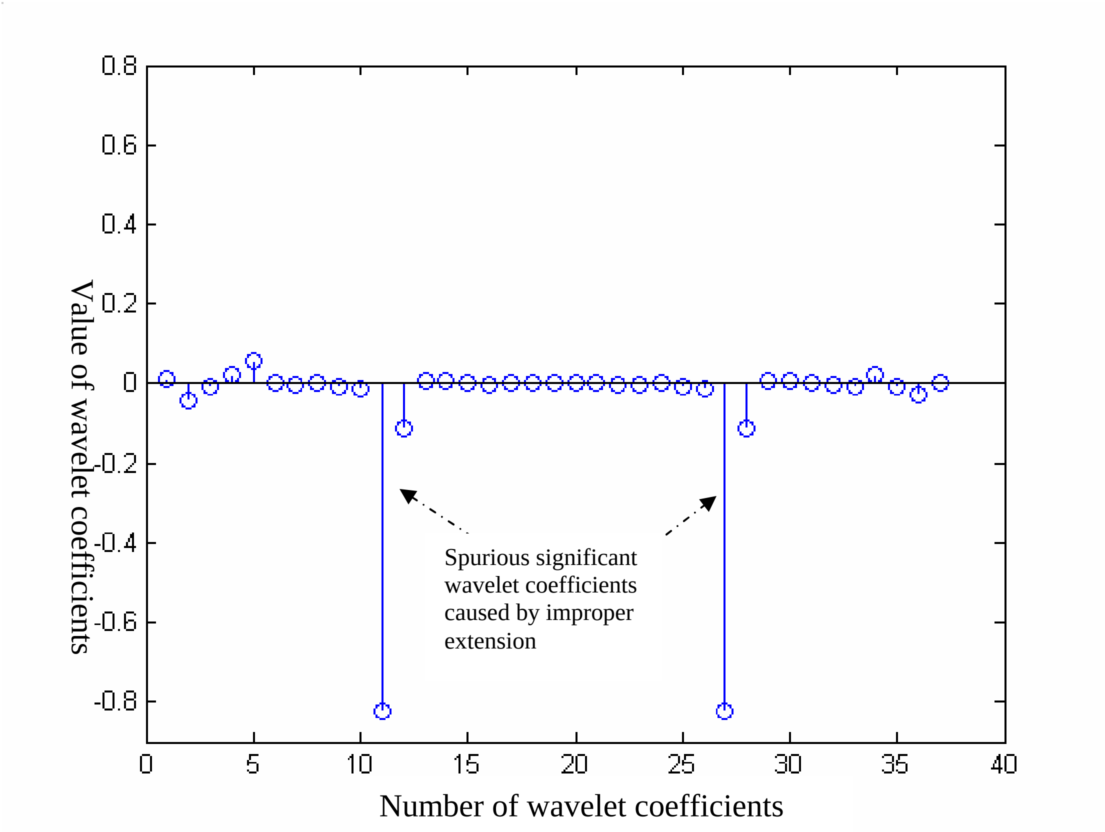
Fig. 5.6 Decomposition and reconstruction scaling and wavelet functions and their corresponded filters of Bior2.4 wavelet

5.4 Implementation consideration

Since Fast Wavelet Transform is implemented by convolution with the high- and low-pass filters followed by down-sampling, the discrete wavelet transform of a signal with limited length suffers from the border distortion. The simple way to deal with border distortion is to extend the signal on both sides, such as by zero-padding, smooth padding, periodic extension, or boundary value replication methods (Matlab Wavelet Toolbox, 2000). Normally, the signal can be extended by zero padding, symmetric extension, and periodic extension. However, zero padding and periodic extensions generate spurious singularities on both sides of the signal, which result in spurious significant wavelet coefficients on both sides as shown in Figure 5.7, the information of which is useless for recognition.



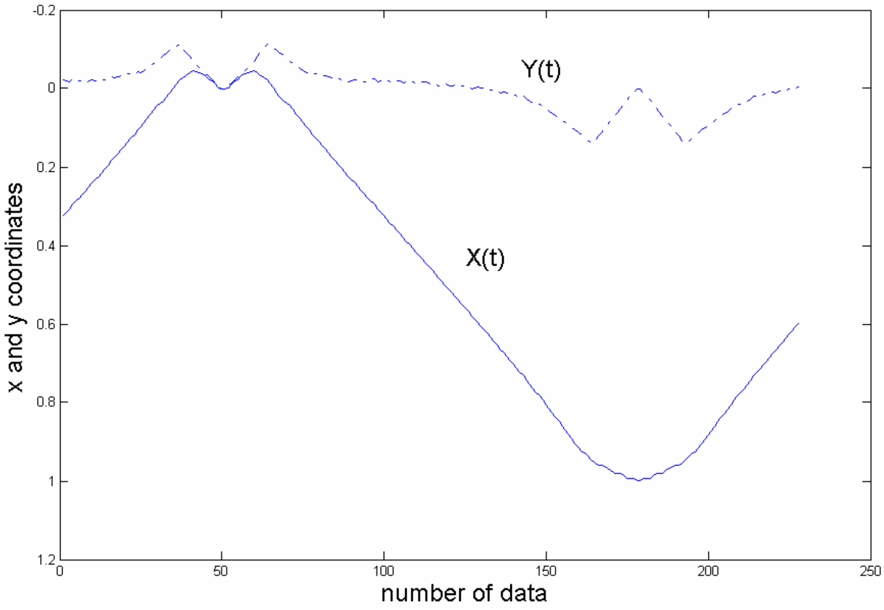
(a)



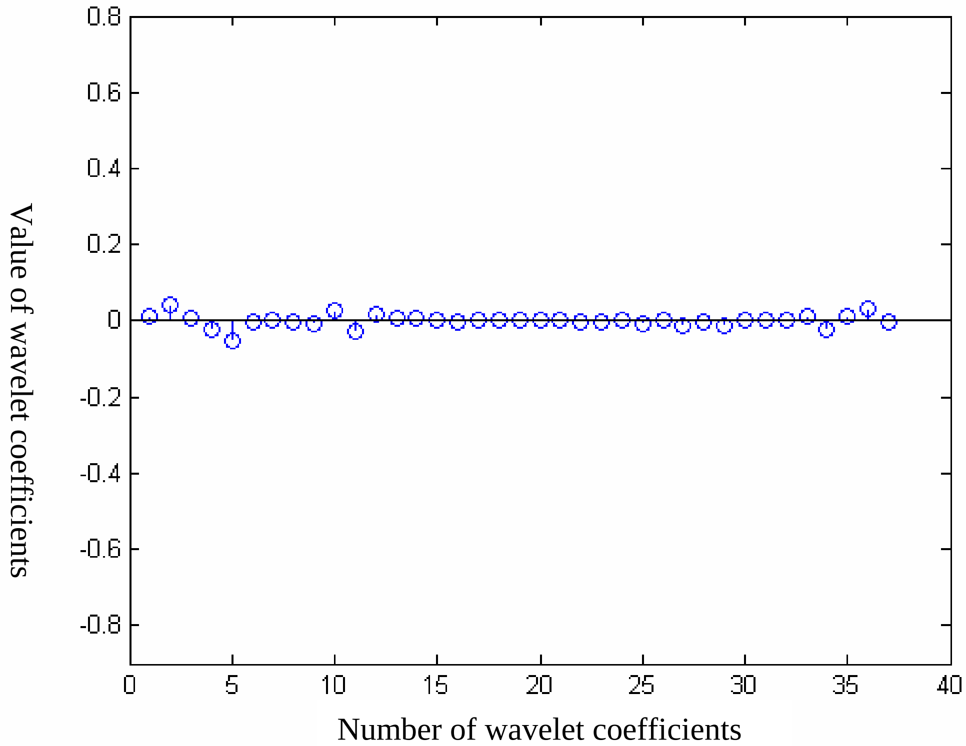
(b)

Fig. 5.7 (a) plot of the x and y coordinates of a curve segment after periodical extension (b)Spurious wavelet coefficients caused by improper extension (periodical extension)

This undesirable side effect can be dramatically reduced by using symmetric extension method as shown in Figure 5.8 (b). The border of the signal is much smoother (as shown in Figure 5.8 (a)) by using symmetric extension than using other extension methods, e.g. zero padding and periodic extension.



(a)



(b)

Fig. 5.8 (a) plot of the x and y coordinates of a curve segment after symmetrical extension

(b) Plot of the coarsest level wavelet coefficients of the x coordinates of the first segment using symmetric extension

5.5 Wavelet coefficients thresholding

The x and y coordinates of the curve segment are close to piece-wise linear signals as shown in Figure 5.8 (a). For piece-wise linear signal, wavelet transform creates a sparse representation of the signal. The signal energy is concentrated into a few wavelet coefficients, while most coefficients of the signal are zero or close to zero. On the other hand, the x and y coordinates of the curve segment contain noise and quantization error, since the coordinates are digitized and sampled. After wavelet transform of the signal representing the transformed coordinates of the extract curve segments, the noise and quantization error are spread widely and equally over all coefficients in wavelet domain. Hence, the coefficients with smaller amplitude are dominated by noise, while those with large amplitude are dominated by signal information (such as corner or deflection) rather than noise. Further-more, the signal can be reconstructed with only these notable coefficients using wavelet reconstruction without losing the essential signal characteristics. The above mentioned amazing properties of wavelet make it possible to eliminate the noise and quantization error by thresholding the wavelet coefficients. Donoho (1995) proposed a wavelet-base denoising algorithm which thresholds the wavelet coefficients called soft threshold. The idea of universal threshold is given by Donoho and Johnstone (1994).

In Donoho's paper (Donoho 1995), the objective is to remove noise from the signal, while the objective in this research is to extract only significant wavelet coefficients and set, as many as possible, insignificant wavelet coefficients to zero to make the representation compact. Therefore, the threshold used here borrows the idea from Donoho and Johnstone (1994), but is greater than the one proposed by them. After extensive experiments and computer simulation, the threshold T is chosen as:

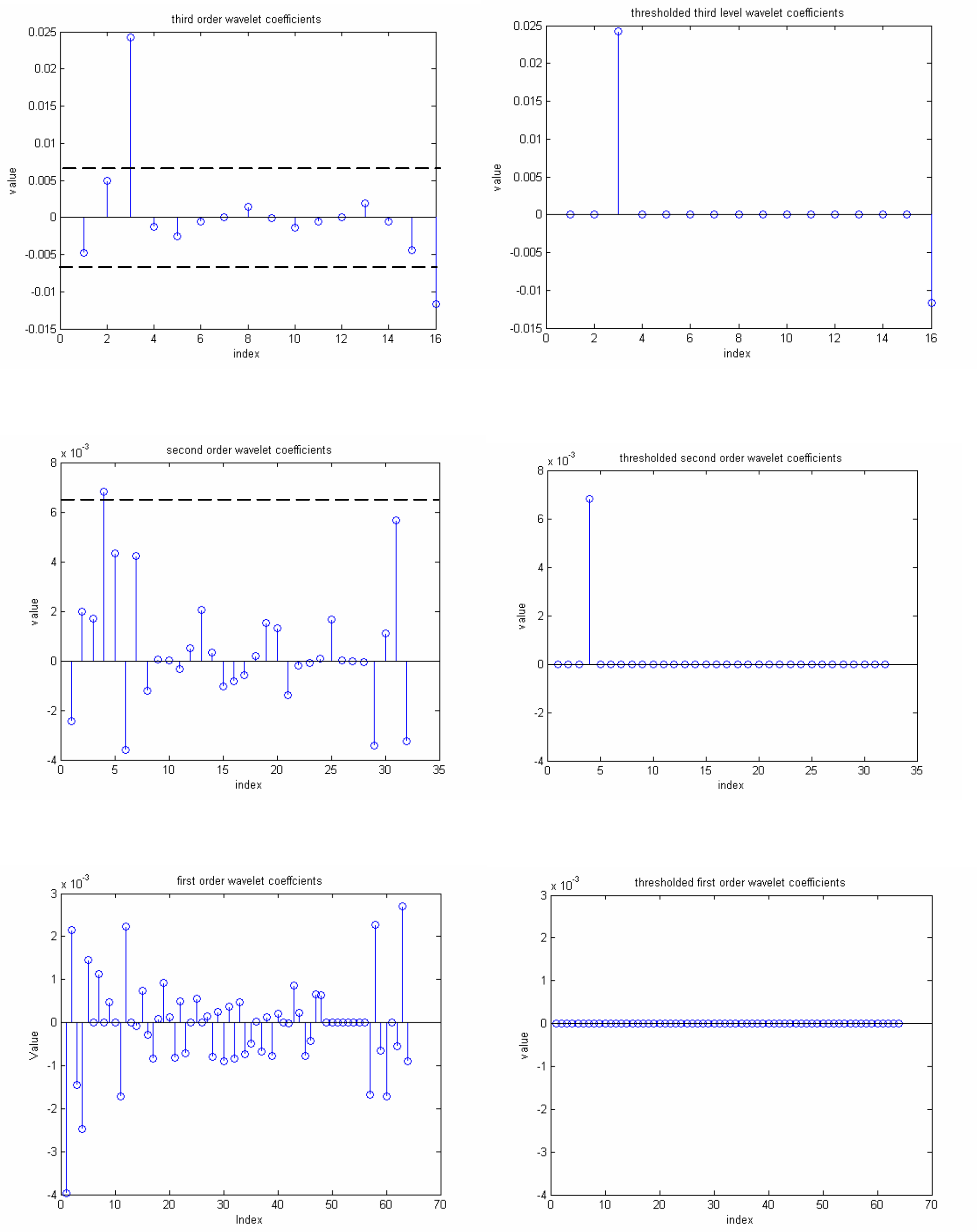
$$T = 3\sigma\sqrt{2\log_e n} \tag{5.6}$$

where n is the number of samples on the signal, and the variance σ^2 is estimated from the median of the finest scale wavelet coefficients by the following formula

$$\sigma = \frac{1}{0.6745} med(|x|) \quad x \in X \tag{5.7}$$

Here X denotes the set of the wavelet coefficients at the finest scale, and med denotes the median function, which is defined as follows: if we have a set of numbers, and the numbers are sorted into either ascending or descending order, the middle term is the median.

A thresholding operation is performed on the wavelet coefficients over all the resolutions in the following way: retain the coefficients which are greater than the threshold and set the rest to zero. An example of the wavelet coefficients before and after thresholding is shown in Figure 5.9. The plots of wavelet coefficients before thresholding are on the left hand side, while the plots of wavelet coefficients after thresholding are on the right hand side for comparison. From the graphs, we see that most of the wavelet coefficients have been set to zero after thresholding. We only retain the scaling coefficients and the non-zero wavelet coefficients as the representation of the curve segment. Hence, the representation of the curve segment is compact since the number of non-zero coefficients is much lesser in comparison with the original data.

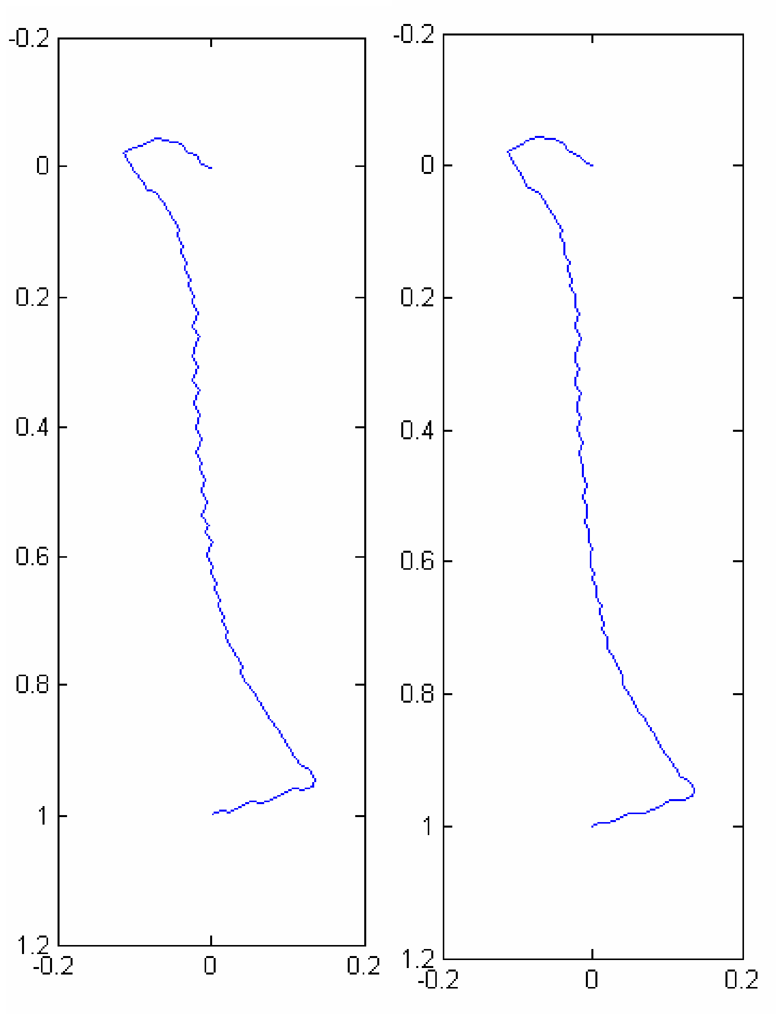


(a) (b)

Fig. 5.9 (a) plot of the wavelet coefficients before thresholding (the threshold is 0.007 represented by dash line)(b) plot of the wavelet coefficients after thresholding

By using the scaling coefficients and wavelet coefficients after thresholding, a close approximation of the original curve segment can be reconstructed using the thresholded wavelet coefficients as shown on the right of Figure 5.10, we can hardly see any difference from the original which is on the left (same with Figure 5.4). The energy difference between the original signal and the reconstructed signal which can be obtained by Equation (5.8) is very small, in the order of 10^{-16} when there is no noise contamination. Therefore, seting the threshold value according to equation (5.6) is an appropriate compromise between the compactness and the faithfulness of the wavelet representation.

$$\frac{(X_{resa}-X_{recons})^2+(Y_{resa}-Y_{recons})^2}{X_{resa}{}^2+Y_{resa}{}^2} \tag{5.8}$$



(a) (b)

Fig. 5.10 (a) Original curve segment (b) Reconstructed curve segment using wavelet coefficients after thresholding

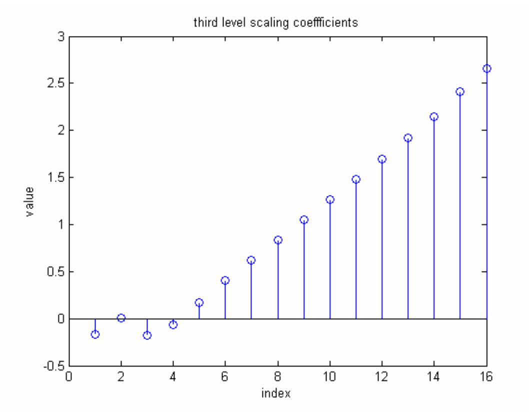
5.6 Object representation

The proposed object representation contains two portions, one is the wavelet representation F of the curve segments on the object boundary, and the other one is the similarity transformation information S of each segment.

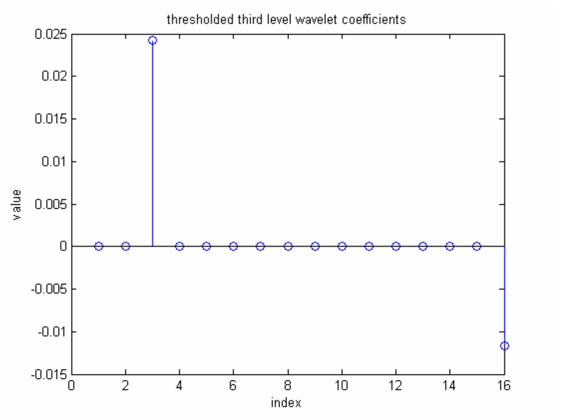
The wavelet representation of an object consists of m segments is defined as:

$$F = \begin{bmatrix} F^1 \\ \vdots \\ F^j \\ \vdots \\ F^m \end{bmatrix} = \begin{bmatrix} d^1{}_{n^1-1}(X,Y) & \cdots & d^1{}_{n^1-i}(X,Y) & \cdots & d^1{}_4(X,Y) & c^1{}_4(X,Y) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d^j{}_{n^j-1}(X,Y) & \cdots & d^j{}_{n^j-i}(X,Y) & \cdots & d^j{}_4(X,Y) & c^j{}_4(X,Y) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d^m{}_{n^m-1}(X,Y) & \cdots & d^m{}_{n^m-i}(X,Y) & \cdots & d^m{}_4(X,Y) & c^m{}_4(X,Y) \end{bmatrix} \quad (5.9)$$

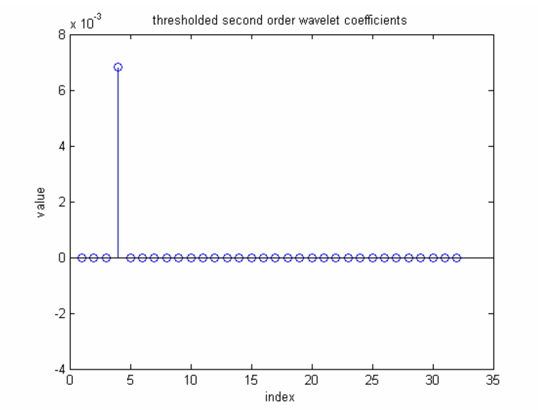
where F^j is the feature vector of curve segment j, which consists of the coarsest level scaling coefficients $c^j{}_4(X,Y)$ and the wavelet coefficients from the coarsest level $d^j{}_4(X,Y)$ to level $d^j{}_{n^j-1}(X,Y)$. According to the wavelet theory mentioned in Chapter 3, the scaling coefficients $c^j{}_4(X,Y)$ carry the approximation information, while the wavelet coefficients $d^j{}_4(X,Y)$ to $d^j{}_{n^j-1}(X,Y)$ carry the details from coarse level to finest level. An example is shown in Figure 5.11 to illustrate the representation of a curve segment using the scaling and wavelet coefficients. Note that the level of decomposition $n^j - 4$ depends on the original curve segment length.



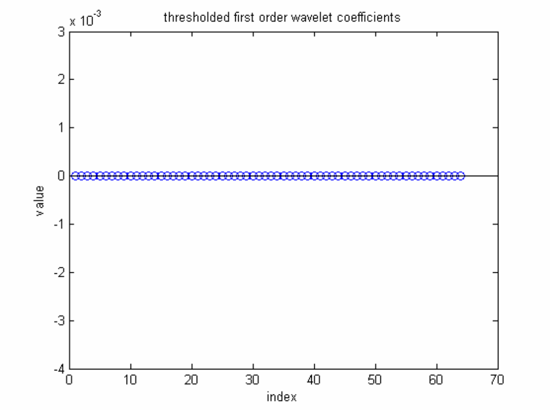
(a)



(b)



(c)



(d)

Fig. 5.11 Wavelet representation of the x coordinates of the segment of bull head as shown in figure 5.4. (a) scaling coefficients (b)-(d) wavelet coefficients at multiple scales

The similarity transformation information matrix S is defined as:

$$S = \begin{bmatrix} S^1 \\ \vdots \\ S^j \\ \vdots \\ S^m \end{bmatrix} = \begin{bmatrix} (x_0, y_0)^1 & \theta^1 & L^1 \\ \vdots & \vdots & \vdots \\ (x_0, y_0)^j & \theta^j & L^j \\ \vdots & \vdots & \vdots \\ (x_0, y_0)^m & \theta^m & L^m \end{bmatrix} \tag{5.10}$$

Where (x_0, y_0), θ and L are the start points coordinate, rotation angle and distance between start point to end point respectively mentioned earlier in section 5.2.

5.7 Evaluation of proposed object representation

In this section, the proposed object representation is evaluated according to the criteria of a good shape descriptor set in Section 1.4.

a) Invariance

The proposed object representation consists of the wavelet descriptors of isolated curve segments of the object boundary. Each curve segment undergoes normalization process including translation, rotation and scaling to standardize its position, orientation and scale as described in Section 5.2. Therefore, the proposed object representation is inherently invariant to similarity transform.

b) Stability

The proposed object representation algorithm includes a wavelet coefficients thresholding process. As mentioned in Section 5.5, the thresholding process removes the noise on the object boundary, if any. Therefore, the proposed object representation is robust to boundary noise contamination.

c) Compactness

Most of the wavelet coefficients have been set to zero after thresholding as shown in Figure 5.9. The number of scaling coefficients and non-zero wavelet coefficients is only 12-20% (depending on the length and the nature of object boundary) of the number of points on the object boundary (See chapter 6 and 7 for the details of the experimental results). Therefore, the proposed object representation is compact.

d) Completeness

Most of the existing object representation algorithms use some specific extracted features as the object representation, such as the geometrical primitives including corners, arcs, lines and holes and some abstract features e.g. curvature zero-crossings, wavelet transform zero-crossings, intersections of the predefined levels with the wavelet transforms as reviewed in Chapter 2. These features are unable to form a complete integrated object representation (Discussed in chapter 2). The proposed object representation can be used to reconstruct a good approximation of the object boundary as shown in Section 5.5. Therefore, the proposed object representation is more complete.

e) Hierarchical Representation

The proposed object representation consists of scaling coefficients and wavelet coefficients at multiple scales. The scaling coefficients carry the information on the approximation of the object, while the wavelet coefficients at multiple resolutions carry the information of the details of the object at multiple scales as described in Section 5.3. The proposed representation thus possesses the hierarchical representation property.

f) Generalization

Some object representations are only applicable for a particular category of objects, such as the polygonal approximation which is only applicable to objects with polygonal shape. The proposed object representation is applicable to any arbitrary shape, because wavelet transform can be applied on any arbitrary square integrable functions.

g) Efficiency

The main computational effort needed during feature extraction is the computing of the wavelet coefficients. Due to the use of fast wavelet transform algorithm which is computationally efficient, as reviewed in Section 3.4, the object feature can be extracted efficiently.

h) Uniqueness

The wavelet transform and inverse wavelet transform are one-to-one correspondence transformations as mentioned in Section 3.3. Therefore, for two objects with different shapes, the wavelet representations of these two objects are different. A proper designed classifier can readily distinguish two objects with different shape using our proposed representation.

i) Handling partial occlusion capability

The proposed feature extraction algorithm partitions the object boundary into independent curve segments. Therefore partial occlusion only affects the features of the segment(s) being occluded. The partially occluded object can still be recognized using the features from the intact portion.

Chapter 6

Hierarchical Matching

6.1 Introduction

Our pattern recognition system is model based. It is to recognize an object in the scene among the many model objects whose representations are known and stored in a database. Having the hierarchical wavelet representations of the object in the scene and the models in database, object recognition is equivalent to a pattern recognition (feature matching) problem.

There are several approaches for pattern recognition, including statistical pattern recognition, syntactic pattern recognition and neural network pattern recognition (Richard et al. 2000). Statistical pattern recognition assumes, as its name implies, a statistical-based classification algorithm. A set of characteristic measurements are extracted from the input data and are used to assign each feature vector (matrix) to one of the known classes. Features are assumed to be generated by a state of nature, and therefore the underlying model is of a state of nature or class-conditioned set of probabilities and probability density functions. Syntactic pattern recognition, also called structural pattern recognition, relies more on the interrelationships or interconnections of features. Neural network is a relatively newly emerged approach that attempts to draw on knowledge of how biological neural systems store and manipulate information. This leads to a class of artificial neural systems termed neural networks. This study involves an amalgamation of research in many diverse fields such as psychology, neuroscience, cognitive science, and systems theory. However, generally speaking, neural network systems need to be trained using large amount of training examples with known identities to represent the variation and distribution of the representation of each class in the feature space. Obtaining the training examples and the training process are inconvenient, computational expensive and sometimes impossible in practice. In this research, when constructing the model database, we only take one standard image of a model object. Therefore, neural network method is not applicable here.

As we have described in Chapter 5, the proposed object representation consists of the wavelet representations of the curve segments and the similarity transformation information which generate the interrelationship between segments. The wavelet representation of the curve segments F contains the shape information of constituent segments, while the similarity transformation information matrix S retains the interrelationships between constituent segments. Therefore, the matching process in this thesis is a combination of statistical and structural pattern recognition methods. The wavelet representation of each segment of the object in the scene is matched with the wavelet representation of the segments of the models in the database iteratively first using statistical approach as illustrated in Figure 6.1.

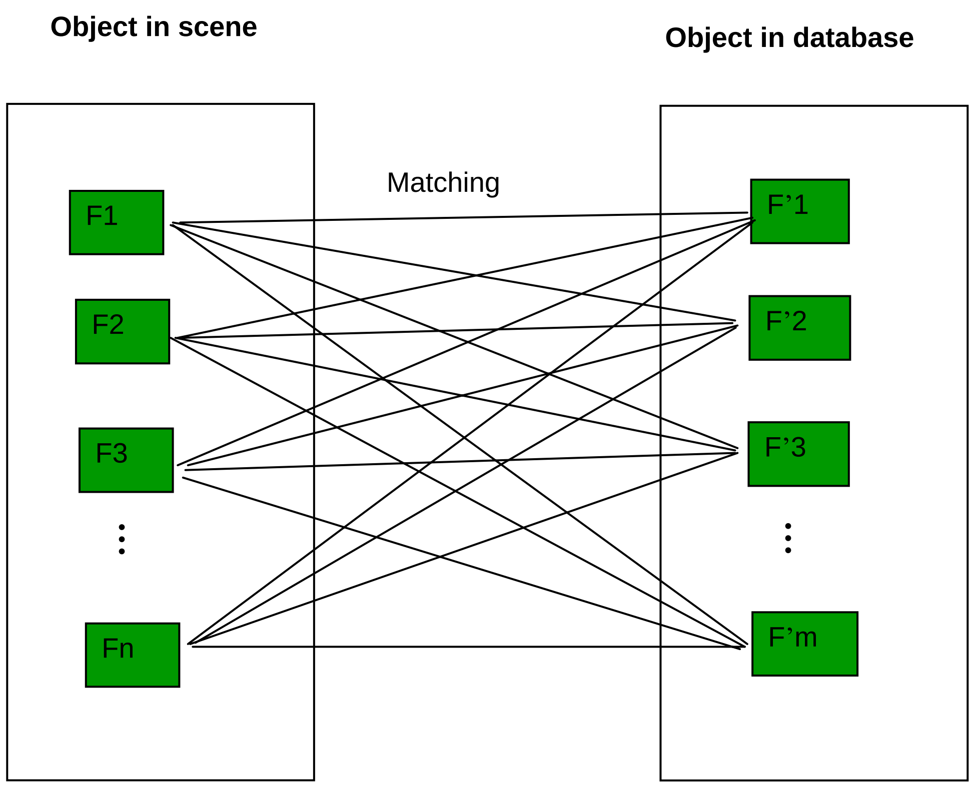


Fig. 6.1 Feature matching of object in scene with model object (m is the number of segments of the object in scene, n is the number of segments of the object in database)

If two or more separated segments between object in the scene and the model are matched, the relative position between the segments is used for further matching.

The matching process carries on iteratively between the object in the scene and the models in the database until the identity of the object in the scene is confirmed (as illustrated in Figure 6.2). Such iteratively matching is very computational intensive when the number of models in the database is large. We proposed a hierarchical matching algorithm to match the segments between object in the scene and the model objects which can tremendously reduce the computational load.

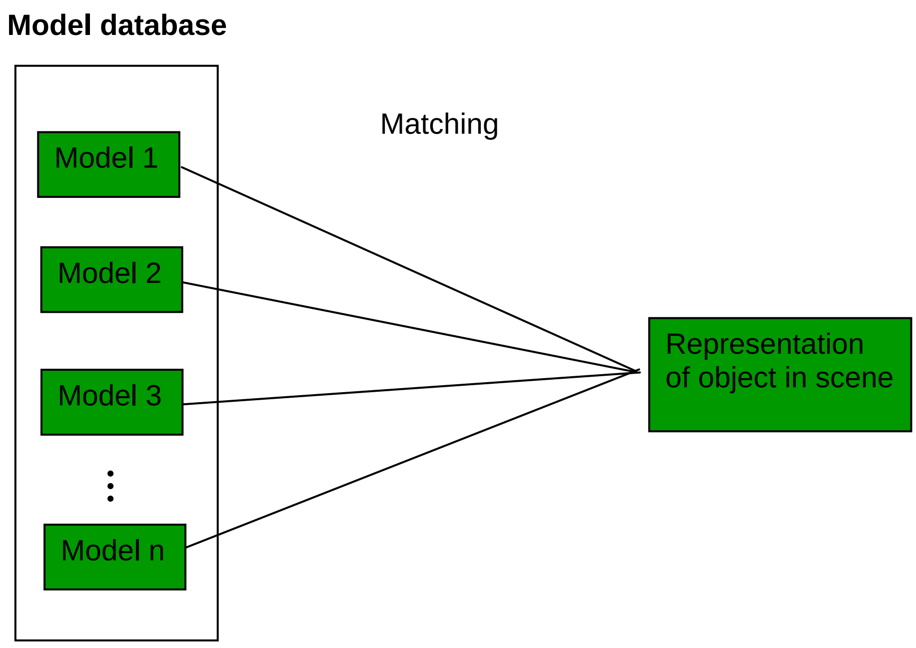


Fig. 6.2 Iteratively matching between object in scene with models in database

In the following sections, our proposed hierarchical segment matching algorithm and segment relative position matching method are described.

6.2 Hierarchical Matching of Segments

Firstly, we try to match the wavelet representations of segments between objects in the scene and the model objects to look for matched segment-pairs candidates. Matching in such an iterative manner can be very computational intensive, especially when the number of models in the database is large. To speed up the recognition process, a hierarchical matching algorithm which can increase the matching speed is proposed. The reason why a hierarchical matching strategy can be adopted is due to the multiscale representation structure of our proposed wavelet representation of object. From wavelet reconstruction theory, we can use c_4 in equation (5.9) to reconstruct an approximation of the original segment, and better and better approximation can be obtained by adding d_4 , d_5 $\cdots$ one by one using wavelet reconstruction. Hence, our feature matrix forms a multiresolution representation, with c_4 represents the approximation, and d_4 , d_5 ... represent the details in increasingly finer scale. This facilitates the use of the hierarchical coarse–to-fine matching strategy as illustrated in Fig.6.3. We match the manipulated scaling coefficients c_4 first, if the dissimilarity function (Equation 6.4) of c_4 is less than a proper chosen threshold value Tc_4 , the matching proceed to the next finer level wavelet coefficients d_4 , d_5 ... till the finest possible scale. If the dissimilarity value is greater than the properly chosen threshold at any level, the matching process between these two segments is terminated, and is considered as fails.

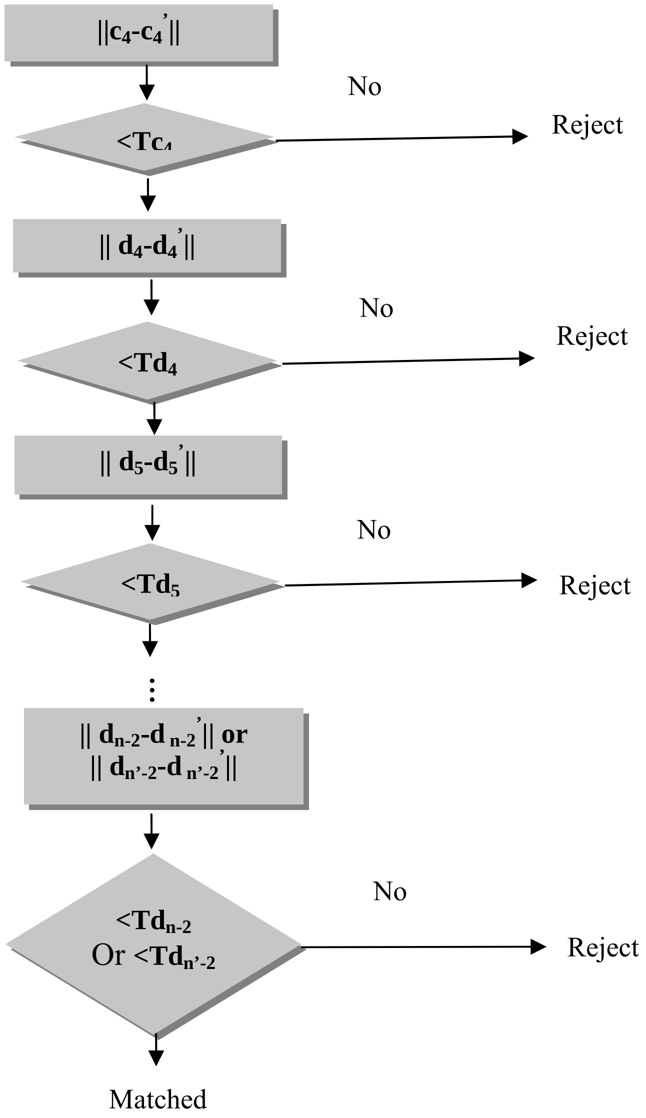


Fig. 6.3 Hierarchical matching flow chat

Using this hierarchical matching algorithm, most of the matching processes between two different segments are terminated at the first or the first few levels. Note that, the numbers of coefficients at the first few levels are less. In addition, most of the wavelet coefficients of the wavelet representation of the object have been set to zero after thresholding, and during matching process only non-zero wavelet coefficients need to be tested for matching. Therefore, the computational requirement for matching is greatly reduced in comparison with a thorough direct matching.

6.3 Matching of segments with different number of samples

As discussed in the previous chapter, the curve segments may have different number of samples after resampling in order to retain the original scale. As a result, the levels of wavelet decomposition for curve segments with different number of samples may also be different. To match the wavelet features of two segments with different level of decomposition, the highest possible scale of matching is taken to be the one with the lesser number of samples (lesser level of wavelet decomposition). As illustrated in Figure 6.3, the highest level of matching can be either $n-2$ or $n'-2$, which ever is smaller. Using such wavelet representation and matching strategy, the curve segments can be matched at their natural scales instead of a pre-designated scale, such as the Fourier descriptor (Gorman et al. 1988). Therefore, the matching accuracy is higher.

When we wish to match a segment with 2^n points with another one with $2^{n'}$ points, the scaling and wavelet coefficients need to be multiplied by a factor $\sqrt{2^{n-n'}}$ during matching to compensate the scale difference.

If we represent a segment with 2^n points by $\{x(t), y(t)\}\quad t = 1,2...2^n$, then its scaled version with $2^{n'}$ points would be represented by $\{X(k), Y(k)\}\ \ k = 1,2...2^{n'}$ in relation to $\{x(t), y(t)\}\quad t = 1,2...2^n$ by Equation (6.1).

$$\{X(k), Y(k)\} = \{2^{n'-n} * x(k/2^{n'-n}), 2^{n'-n} * y(k/2^{n'-n})\}\quad k = 1,2...2^{n'} \tag{6.1}$$

By substituting equations (6.1), (3.4) and (3.5) into equations (3.14) and (3.15), we obtain:

$$\begin{aligned}
c_4^{'}(k) &= \int X(k)\varphi_{n'-4}(k)dk \\
&= \int 2^{n'-n} x(k/2^{n'-n})\varphi_{n'-4}(k)dk \\
&= \int 2^{n'-n} x(t)\varphi_{n'-4}(2^{n'-n}t)dt \\
&= 2^{(n'-n)/2}\int x(t)2^{(n'-n)/2}\varphi_{n'-4}(t)dt \\
&= 2^{(n'-n)/2}\int x(t)\varphi_{n-4}(t)dt \\
&= 2^{(n'-n)/2}c_4(t)
\end{aligned} \tag{6.2}$$

$$\begin{aligned}
d_i^{'}(k) &= \int X(k)\psi_{n'-i}(k)dk \\
&= \int 2^{n'-n} x(k/2^{n'-n})\psi_{n'-i}(k)dk \\
&= \int 2^{n'-n} x(t)\psi_{n'-i}(2^{n'-n}t)dt \\
&= 2^{(n'-n)/2}\int x(t)2^{(n'-n)/2}\psi_{n'-i}(t)dt \\
&= 2^{(n'-n)/2}\int x(t)\psi_{n-i}(t)dt \\
&= 2^{(n'-n)/2}d_i\,(t)
\end{aligned} \tag{6.3}$$

Therefore, the dissimilarity function of the scaling and wavelet coefficients is shown in equation (6.4 and 6.5).

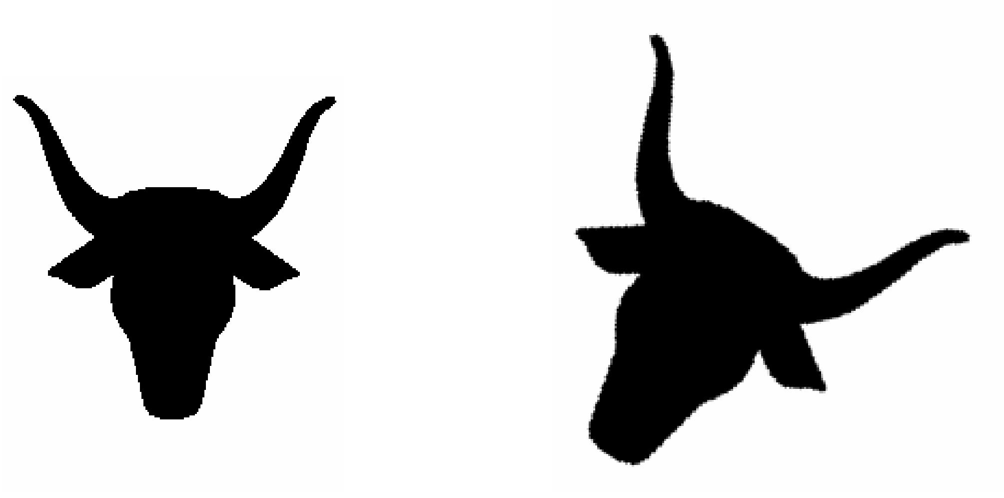
$$\left\|c_4 - c_4^{'}\right\| = \sum_{j=1}^{h}\left(\left|\sqrt{2^{n'-n}}c_{4j} - c_{4j}^{'}\right|_x + \left|\sqrt{2^{2^{n'-n}}}c_{4j} - c_{4j}^{'}\right|_y\right) \tag{6.4}$$

$$\left\|d_i - d_i^{'}\right\| = \sum_{j=1}^{h}\left(\left|\sqrt{2^{n'-n}}d_{ij} - d_{ij}^{'}\right|_x + \left|\sqrt{2^{2^{n'-n}}}d_{ij} - d_{ij}^{'}\right|_y\right) \tag{6.5}$$

where h is the number of coefficients, j is its index, subscripts x and y denote the scaling coefficients wavelet representation of x and y coordinates, and $|\;|$ is the Absolute value operation. Using Absolute operation instead of the conventional Euclidean distance which is the square root of the sum of square of the difference of each element is more computationally efficient.

6.4 Matching process

To summarize the steps of the segment hierarchical matching using its wavelet representation, an example of matching a image of bullhead with its scaled and rotated version (as illustrated in Figure 6.4) is shown below to demonstrate the hierarchical matching procedure:



(a) (b)

Fig. 6.4 (a) Original bull head (b) scaled and rotated bull head

Step1: Scaling wavelet coefficients matching

Table 6.1 shows the dissimilarity value of scaling coefficients $\left\| c_4 - c_4^{'} \right\|$ of the eight segments between the original bull head and it's scaled and rotated version. S_1, S_2 … S_8 denote the segments of the original bull head, while S_1', S_2' … S_8' denote the segments of the scaled and rotated bull head. The threshold for the scaling coefficients is set at 1 (decided experimentally). Therefore only the cells on the diagonal line marked with grey are eligible candidates to proceed to the next level matching, while the else are filtered out.

Table 6.1 dissimilarity value of scaling coefficients **$||c_4\text{-}c_4'||$**

$\|\|c_4\text{-}c_4'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1'	0.3246	8.6424	7.2334	24.034	10.567	7.0365	5.0535	30.754
S_2'	6.1133	0.5647	3.4635	25.963	2.4563	4.5256	6.9843	32.832
S_3'	5.1152	3.4245	0.3822	25.835	3.6767	2.3875	8.4433	33.337
S_4'	33.922	51.852	51.635	0.4348	51.676	53.734	31.763	32.744
S_5'	7.4624	2.4652	3.6764	25.634	0.4583	4.7344	8.3456	33.664
S_6'	4.9625	4.5244	2.3835	26.836	4.7867	0.3456	9.1645	32.836
S_7'	5.0784	9.8246	11.735	22.476	11.766	12.634	0.3874	30.475
S_8'	61.484	92.886	93.545	46.397	95.267	92.654	60.654	0.3837

Step2: Coarsest level wavelet coefficients matching

Table 6.2 shows the dissimilarity value of the manipulated coarsest level wavelet coefficients $\left\| d_4 - d_4' \right\|$ of the eight segments between the original bull head and its scaled and rotated version. Since all other cells except those on the diagonal line have been filtered out during last step, the calculation of the dissimilarity value of the coarsest level wavelet coefficients is only performed on a few eligible cells on the diagonal line. The threshold for the matching on this coarsest level wavelet coefficients is set as 0.2 (decided experimentally).

Table 6.2 dissimilarity value of the coarsest level wavelet coefficients **$||d_4\text{-}d_4'||$**

$\|\|d_4\text{-}d_4'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1'	0.0687							
S_2'		0.0764						
S_3'			0.0556					
S_4'				0.0753				
S_5'					0.0578			
S_6'						0.0875		
S_7'							0.0446	
S_8'								0.0466

Step3: Finer level wavelet coefficients matching till the finest

Table 6.3 shows the dissimilarity value of the threshoded wavelet coefficients d_5, $d_6 \ldots d_9$ of the eight segments between the original bull head and its the scaled and rotated version. The thresholds for these finer levels wavelet coefficients are set as 0.4, 0.3, 0.2, 0.15, 0.1, 0.08…respectively. Note that these values were obtained experimentally. To match the wavelet coefficients at finer levels is to confirm the shape of the curve segment at higher resolution. Therefore, the matching precision can be adjusted by fine tuning the threshold at each level or omit the matching above certain higher resolution if the precision requirement is low, noise level is high or to recognize some object with none-rigid shape. As shown in Table 6.3, the maximum level of matching is different among segments, from level d_7 some segments has reached their finest resolution, e.g. those marked with a tick "√". The maximum level of matching is on their original resolution, therefore the matching is precise and controllable.

Table 6.3 dissimilarity value of the finer level wavelet coefficients

$\|\|d_5-d_5'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1'	0.0234							
S_2'		0.0274						
S_3'			0.0124					
S_4'				0.0083				
S_5'					0.0129			
S_6'						0.0087		
S_7'							0.0192	
S_8'								0.0095

$\|\|d_6-d_6'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1'	0.0093							
S_2'		0.0000						
S_3'			0.0000					
S_4'				0.0082				
S_5'					0.0000			
S_6'						0.0000		
S_7'							0.0032	
S_8'								0.0042

$\|\|d_7-d_7'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1'	0.0000							
S_2'		√						
S_3'			√					
S_4'				0.0018				
S_5'					√			
S_6'						√		

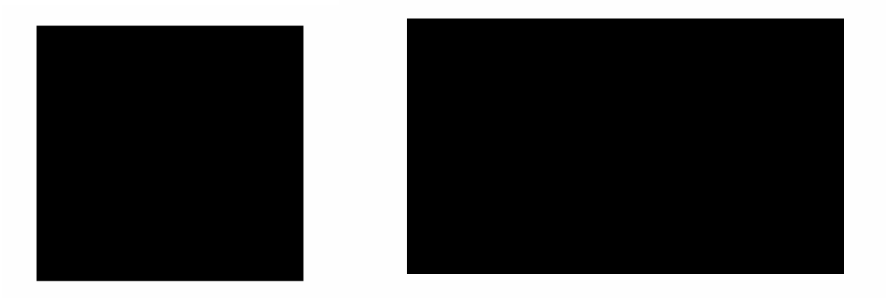
S_7’							0.0000	
S_8’								0.0023

$\|\|d_8-d_8'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1’	√							
S_2’		√						
S_3’			√					
S_4’				0.0000				
S_5’					√			
S_6’						√		
S_7’							√	
S_8’								0.0002

$\|\|d_9-d_9'\|\|$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1’	√							
S_2’		√						
S_3’			√					
S_4’				√				
S_5’					√			
S_6’						√		
S_7’							√	
S_8’								0.0000

6.5 Interrelationship verification

If two or more segments between the object in the scene and a model in the database match, the relative position information is used for verification. This is due to the fact that in some cases, the local shape information alone is insufficient to confirm the object identity. For example, the proposed wavelet representation (local shape information) is the same for a square and a rectangle as shown in figure 6.5, since it has lost its scale information during normalization process. Therefore, the position, scaling and orientation information can be added to confirm their interrelationship. For square and rectangle, the ratio between neighboring segments is different.



(a) (b)

Fig. 6.5 (a) Square (b) Rectangle

Consider an object with curve segments i and j in the scene, which is the similarity transformed of a model, with curve segments h and k, in the database. In addition, segments i and j correspond to segments h and k, respectively. Let δ be the distance between the starting points of segment i (x_0^i, y_0^i) and that of segment j (x_0^j, y_0^j) given by $\delta = \sqrt{(x_0^i - x_0^j)^2 + (y_0^i - y_0^j)^2}$; then, that between the starting point of segment h $(x_0^{'h}, y_0^{'h})$ and that of segment k $(x_0^{'k}, y_0^{'k})$ is given by $\delta^{'} = \sqrt{(x_0^{'h} - x_0^{'k})^2 + (y_0^{'h} - y_0^{'k})^2}$. Furthermore, let the length of segment i, j ,h and k be L^i L^j $L^{'h}$ and $L^{'k}$, respectively. Based on the Similarity Transform geometry (Scale), the ration $\frac{\delta}{\delta^{'}}, \frac{L^i}{L^{'h}}$ and $\frac{L^j}{L^{'k}}$ must be equal as shown in equation (6.6).

$$\frac{\delta}{\delta^{'}} = \frac{L^i}{L^{'h}} = \frac{L^j}{L^{'k}} \tag{6.6}$$

For the orientation, let us denote θ^i ,θ^j ,$\theta^{'h}$ and $\theta^{'k}$ to represent the angle defined by equation (5.3) for segments i, j ,h and k, respectively. On the same similarity transform basis, equation (6.7) must be true.

$$\theta^{i} - \theta^{'h} = \theta^{j} - \theta^{'k} \tag{6.7}$$

The relative relationship including angle difference, length ratio and distance ratio of the scaled and rotated bull head shown in figure 6.4(b) with the original bull head shown in figure 6.4(a) are listed in table 6.4, 6.5 and 6.6 respectively. From the result shown in these tables, we can confirm the identity of the similarity transformed bull head, as the results obey equations (6.6) and (6.7).

Table 6.4 Angle difference

$\theta^{i} - \theta^{'h}$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S1’	46.10							
S2’		47.63						
S3’			46.18					
S4’				46.65				
S5’					45.32			
S6’						45.43		
S7’							45.08	
S8’								45.63

Table 6.5 length ratio

$L / L^{'}$	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S1’	1.1572							
S2’		1.1541						
S3’			1.1558					
S4’				1.1598				
S5’					1.1588			
S6’						1.1566		
S7’							1.1577	
S8’								1.1599

Table 6.6 distance ratio

$\frac{\sqrt{(x_0^i - x_0^j)^2 + (y_0^i - y_0^j)^2}}{\sqrt{(x_0^{'h} - x_0^{'k})^2 + (y_0^{'h} - y_0^{'k})^2}}$	S_1, S_1’	S_2, S_2’	S_3, S_3’	S_4, S_4’	S_5, S_5’	S_6, S_6’	S_7, S_7’	S_8, S_8’
S_1, S_1’								
S_2, S_2’	1.155							
S_3, S_3’	1.1607	1.1541						
S_4, S_4’	1.1522	1.1584	1.1613					
S_5, S_5’	1.1597	1.1552	1.1588	1.1557				
S_6, S_6’	1.1617	1.1616	1.1541	1.1578	1.1529			
S_7, S_7’	1.1619	1.1593	1.1604	1.1565	1.1524	1.1557		
S_8, S_8’	1.1599	1.1561	1.1583	1.1524	1.1581	1.1583	1.1588	

6.6 Matching criteria

The task we are dealing with includes not only standalone object recognition, but also partially occluded object recognition. Partial occlusion causes part of the object to be invisible, and also may introduce new object contour from another overlapping object. Moreover, the degree of occlusion and place of occlusion are different from one case to another. Therefore, to judge whether a model object is present in the scene , we need to take all the possibilities mentioned above into consideration. Listed below are the possible matching result and our matching criteria:

i. All segments of an object in the scene match with all the segments of a model in sequence, and their relative position, scale and orientation are the same. For this case, the presence of the model object in the scene is definitive and is without occlusion.

ii. Some curve segments (two or more, but not all) of an object in the scene match some curve segments of a model, and their relative position, scale and orientation are the same. In this case, the model object is present in the scene but with partial occlusion. Those unmatched segments of model are occluded.

iii. Only one matched segment-pair is found between an object in the scene and a model object. In this case, the model object may be present in the scene; however, the evidence is weak. The model object is not considered present in the scene.

iv. Several segments of an object in the scene match with some segments of a model, but some of their relative scale and orientation differ. In this case, false matching should be eliminated. The largest cluster of segment-pair candidates which have the same relative position, scale and orientation implies true match.

6.7 Scale, position and orientation of the object in the scene

If the model object is found present in the scene, the position, scale and orientation of the object in the scene relative to the model object can be obtained using the following three equations (6.8), (6.9) and (6.10), respectively.

$$Scale = \frac{L^i}{L^{'h}} \tag{6.8}$$

$$(\Delta x, \Delta y) = (x_0^i / Scale - x_0^{'h}, y_0^i / Scale - y_0^{'h}) \tag{6.9}$$

$$\Delta\theta = \theta^i - \theta^{'h} \tag{6.10}$$

Chapter 7

Experimental Results and Discussion

7.1 Introduction

In this chapter, we present the experimental results to illustrate the performance of the proposed recognition system in the following aspects, followed by discussion to highlight its advantages and limitations.

i. Invariance to similarity transformation.

   Standalone objects with different position, scaling and rotation have been used to examine the proposed recognition system's performance under these conditions.

ii. Robustness to partial occlusion

   Occlusion is the major concern of this research project, therefore extensive experiments have been done to verify the performance of the proposed algorithm under partial occlusion. During the testing, two model objects are randomly positioned in a selected region so that they overlapped with each other and with arbitrary relatively position.

iii. Robustness to the combination of the similarity transform and partial occlusion

   The object recognition of object with both scaling and partial occlusion is considered as the most difficult problem in 2-D object recognition. We use this as the critical test of our recognition system.

7.2 Design of Experiment

i. System Configuration

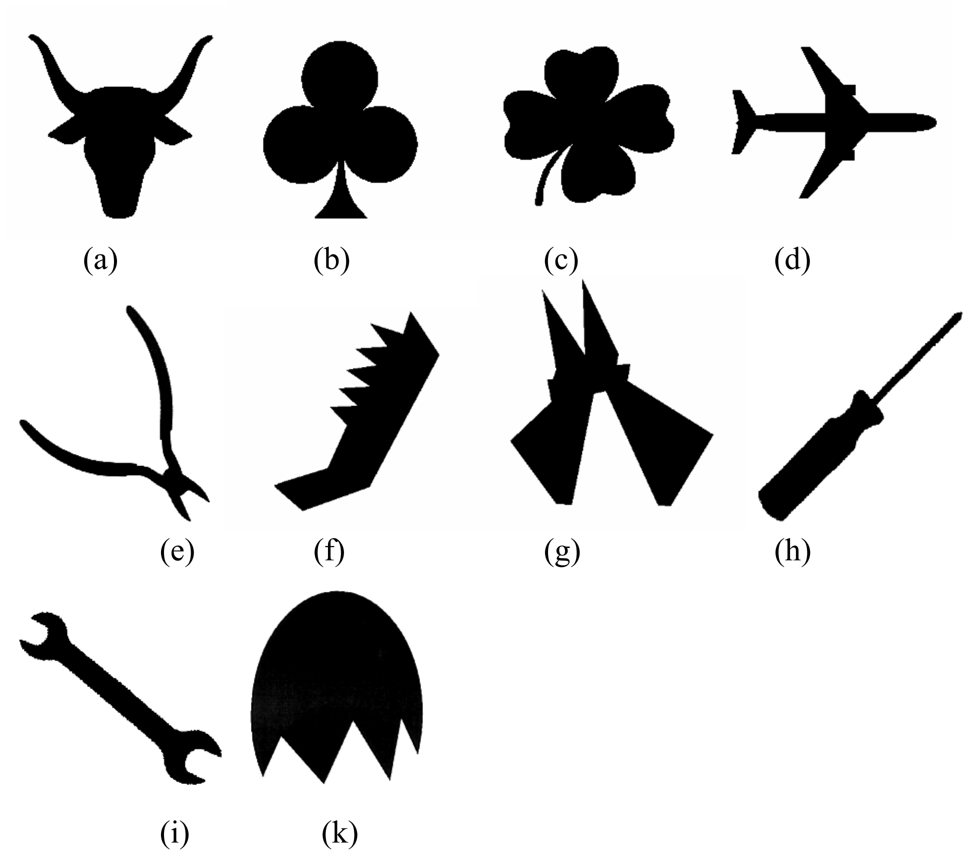
The image acquisition system is composed of a CCD (Charge-Coupled Device) camera, a PCI image capture card, which acts as a frame grabber and an IBM-compatible PC. The CCD camera is used to capture the object images. The PCI image capture card can be used to convert the captured image to a static image file in the format of TIFF (*Tag Image File Format*), a widely used format for storing image data. Then the IBM-compatible PC is used to implement our algorithms, and finally produces the results.

The proposed recognition algorithm, which consists of preprocessing, feature extraction and feature matching, is implemented and coded in Matlab 6.5 with the use of library functions from image processing toolbox, wavelet toolbox and statistics toolbox.

ii. Model images and testing images

The images of model objects to construct the model database are captured by the CCD camera or downloaded from internet. The model objects are not restricted to polygonal shape, in fact, the shape of model object can be arbitrary. However, since the proposed recognition system requires dominant points for boundary partitioning, the model objects are expected to have two or more dominants points. Some model objects in our database are shown in Figure 7.1.

For testing, the images are expected to have various variations in several aspects, e.g. scale, orientation, noise ratio and partial occlusion. In order to have large number of images to test the performance of the proposed recognition system under certain variation, the test images are mostly synthetic images generated by manipulating the images existing in the model database.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (k)

Fig. 7.1 Images to construct database

7.3 Database construction

To perform on-line recognition, a model database needs to be constructed first. For each image with known identity, an image is captured by placing the object in a specific known orientation. For example, the image shown in Fig 7.1(a) is captured by having the axis which bisects the bull head in the vertical position, we then extract its feature first using our proposed feature extraction algorithm, and then store its feature matrix together with its identity into the database as shown in Table 7.1. The construction of database is done offline in order to save time for later recognition.

Table 7.1 Model database

Index	Identity	Wavelet representation	Similarity transformation info
1	Bull head	F_1	S_1
2	Club	F_2	S_2
3	Flower	F_3	S_3
4	Plane	F_4	S_4
5	Pliers	F_5	S_5
6	Saw	F_6	S_6
7	Scissors	F_7	S_7
8	Screw driver	F_8	S_8
9	Wrench	F_9	S_9
10	Egg	F_{10}	S_{10}
…	…	…	…

7.4 Standalone object recognition with similarity transformation

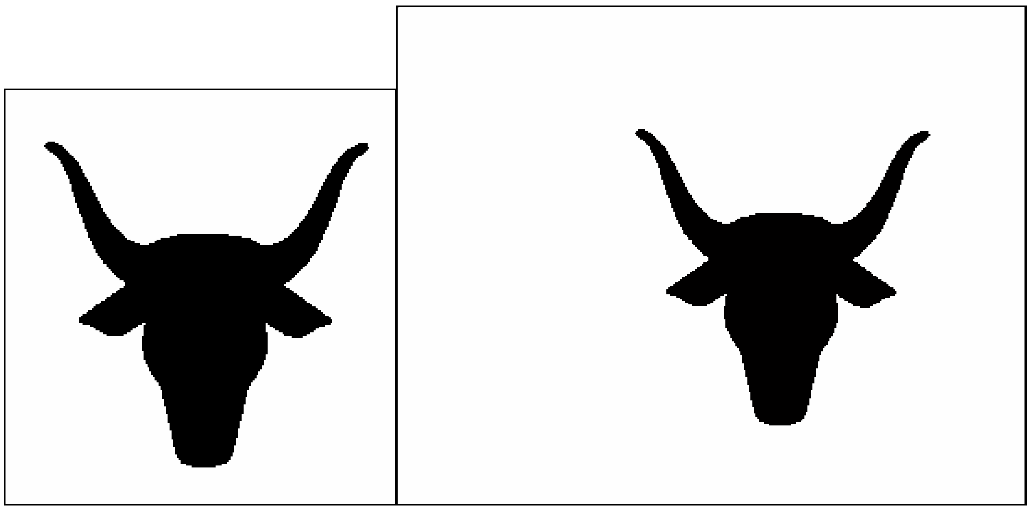
Invariant to translation, rotation and scaling (known collectively as Similarity Transform) is the basic requirement of a object recognition system. In this section, we present the experimental results of the proposed recognition system with the scene object under the said Similarity Transform

i. Translation invariance

The object in the scene may be at a location different from the model object to construct the database. The feature extracted by the proposed recognition system are invariant to translation, because the coordinates of the points on each curve segment are subtracted by the coordinates of it's starting points during normalization stage as described in chapter 5. To test the invariance to the translation of the proposed recognition system, the model objects are artificially shifted to a random new location. (The program for doing so is given in Appendix 1.)

Experiment 1.

An example is shown in Figure 7.2. The bull head is shifted by a certain distance towards upper-right with respect to the standard position of the model image in the database. The corner detection result is shown in Figure 7.3.



(a) (b)

Fig. 7.2 (a) model object-bull head (b) program generated bull head which is shifted by a distance towards upper right

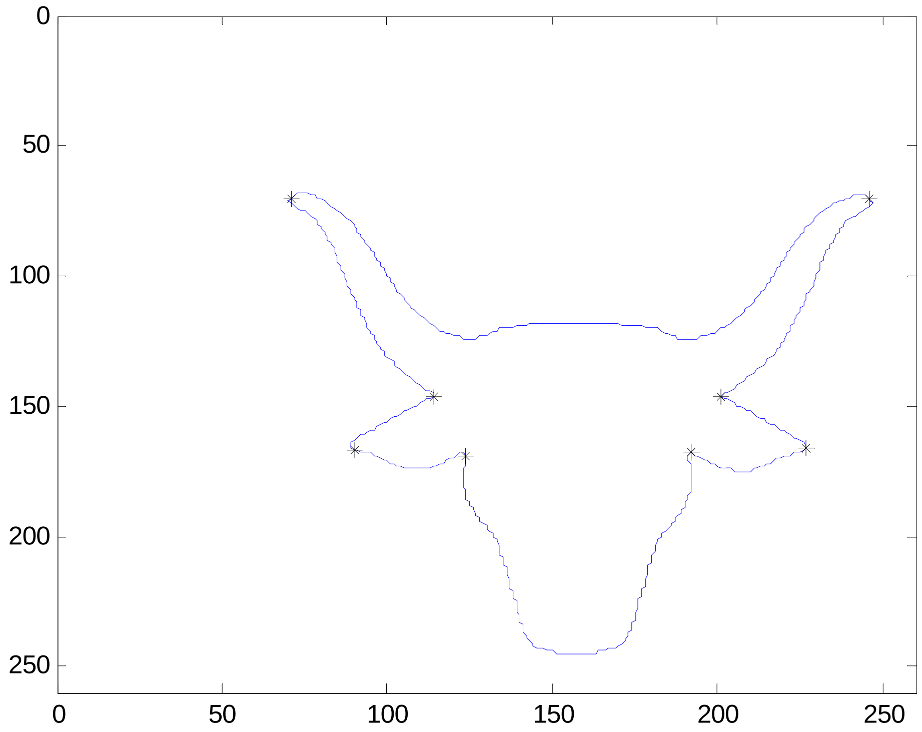


Fig. 7.3 Corner detection result

The dissimilarity values of the scaling coefficients between shifted bullhead and the original (Note: hereafter, "original" refers to the image of the model object in the database) one are shown in Table 7.2. The dissimilarity values of the wavelet coefficients are all zero. This is because translation does not change the proposed wavelet-based features at all. Extensive experiments to test the translation invariance have been done, the recognition rate are 100%. From the experimental result, we are confident that the proposed recognition algorithm is strictly translation invariant. Translation does not even change the feature values of the synthetically shifted images, because the amount of translation is integer number of pixels, it can be removed entirely by normalization.

Table 7.2 dissimilarity value of scaling coefficients $\|c_4 - c_4'\|$

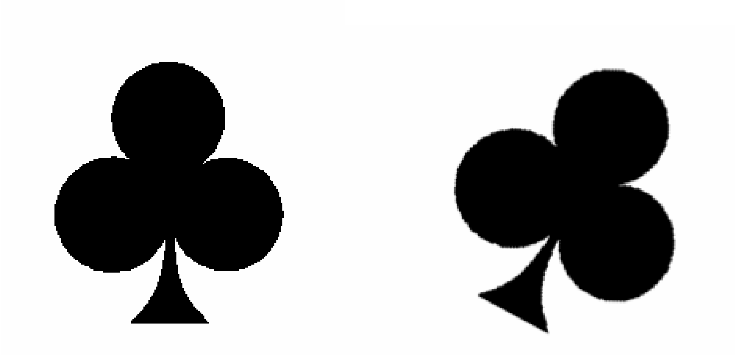
$\|c_4-c_4'\|$	S1	S2	S3	S4	S5	S6	S7	S8
S1’	0	8.6537	7.2327	24.018	10.569	7.0350	5.0793	30.702
S2’	6.1191	0	3.4904	25.931	2.4513	4.5239	6.9872	32.842
S3’	5.1143	3.4904	0	25.843	3.6784	2.3875	8.4008	33.000
S4’	33.967	51.863	51.685	0	51.294	53.765	31.737	32.799
S5’	7.4736	2.4513	3.6784	25.647	0	4.7827	8.2892	33.687
S6’	4.9745	4.5239	2.3875	26.882	4.7827	0	9.1177	32.820
S7’	5.0793	9.8813	11.880	22.442	11.722	12.894	0	30.443
S8’	61.405	92.891	93.338	46.385	95.283	92.830	60.886	0

ii. Rotation invariance

Objects placed in different orientations have been tested to verify the performance of our proposed recognition system under rotation. The proposed recognition algorithm is rotation invariant, since every curve segment is normalized prior to feature extraction so that the direction from its starting point to end point is along the x-axis as described in Chapter 5. To generate rotated images for testing, a program to rotate the images to a random orientation has been written which is given in appendix B.

Experiment 2.

In Figure 7.4, the image of a club is rotated at an unknown angle which is shown on the right, and the original one is on the left. The corner detection result is shown in Figure 7.5. Based on the corner detection result, we use our proposed boundary partition algorithm (described in Chapter 4) to partition the object boundary into 6 curve segments as shown in Figure 7.6. The dissimilarity value of the scaling coefficients and wavelet coefficients at each level are shown in Tables 7.3, 7.4 and 7.5. Finally, we can find the matched segment-pairs shown in Table 7.6, and the rotated club is found to be in an orientation of around 45 degree (from Table 7.7) rotated clockwise with respect to the original club in the database. Large number of rotated images have been tested, the recognition rate is always 100%. Note that the angle difference is calculated with θ values defined in equation (5.3). The experimental result shows that rotation only changes the wavelet representation by a small amount which is within the pre-defined threshold. The changes are due to the quantization error caused by the digitization of pixels' coordinates during rotation. From the experimental results, we can be confident that the proposed wavelet representation of the object is rotation invariant. Hence, the proposed recognition system is rotation invariant too.



(a) (b)

Fig. 7.4 (a) model object (b) program generated image which is rotated by a random angle

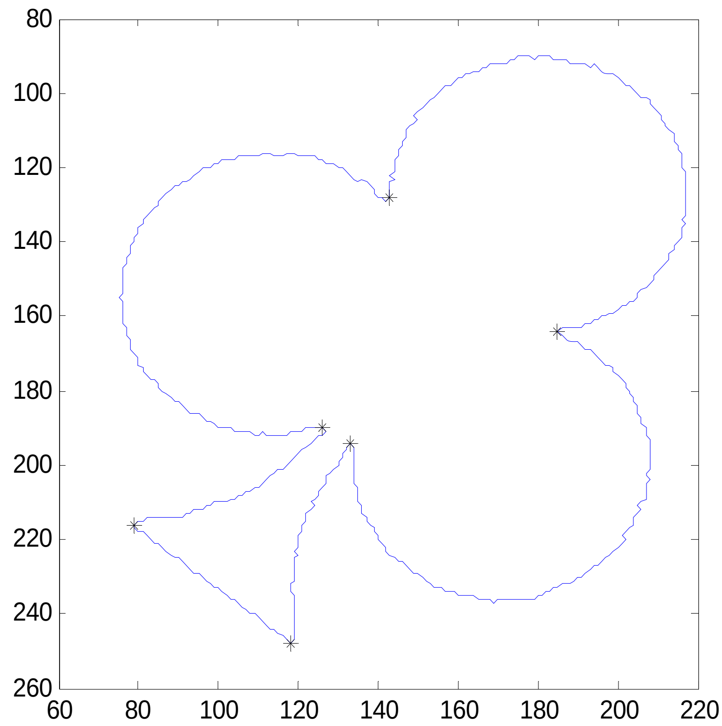


Fig. 7.5 Corner detection result of club

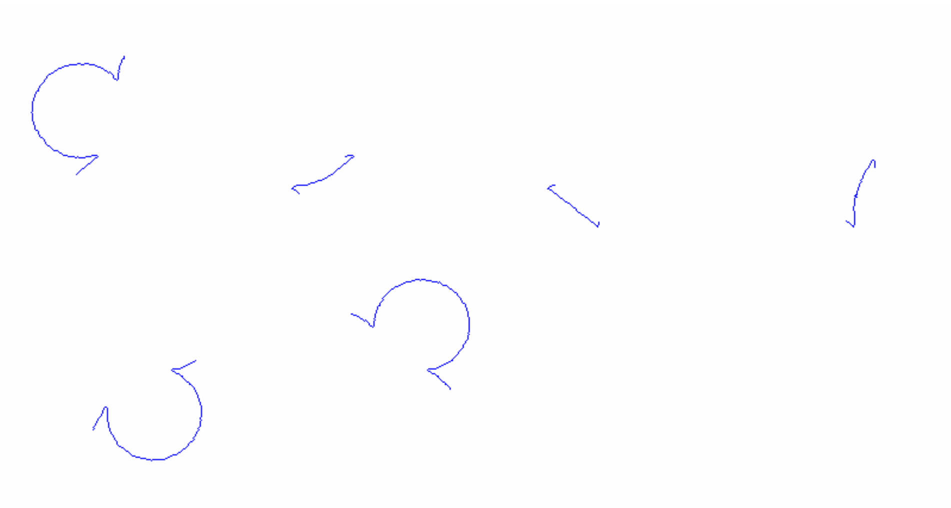


Fig. 7.6 Boundary partition result of club

Table 7.3 dissimilarity value of scaling coefficients **$||c_4\text{-}c_4^{'}||$**

$\|\|c_4\text{-}c_4^{'}\|\|$	S1	S2	S3	S4	S5	S6
S1'	0.7053	54.901	50.75	31.547	2.0605	1.5711
S2'	28.371	0.2279	5.0674	11.311	25.363	28.95
S3'	27.208	4.9134	0.2101	11.638	24.378	27.581
S4'	24.649	13.522	12.227	0.8515	21.819	25.022
S5'	1.6492	51.501	47.127	29.893	0.7324	1.9334
S6'	1.4574	59.789	54.468	34.734	1.8637	0.6227

Table 7.4 dissimilarity value of the coarsest level wavelet coefficients **$||d_4\text{-}d_4^{'}||$**

$\|\|d_4\text{-}d_4^{'}\|\|$	S1	S2	S3	S4	S5	S6
S1'	0.0446					
S2'		0.0252				
S3'			0.0311			
S4'				0.0273		
S5'					0.0482	
S6'						0.0382

Table 7.5 dissimilarity value of the finer level wavelet coefficients

$\|\|d_5\text{-}d_5^{'}\|\|$	S1	S2	S3	S4	S5	S6
S1'	0.0288					
S2'		0.0480				
S3'			0.0224			
S4'				0.0264		
S5'					0.0175	
S6'						0.0326

$\|\|d_6\text{-}d_6^{'}\|\|$	S1	S2	S3	S4	S5	S6
S1'	0.0266					
S2'		0.0254				
S3'			0.0088			
S4'				0.0300		
S5'					0.0016	
S6'						0.0266

$\|\|d_7\text{-}d_7^{'}\|\|$	S1	S2	S3	S4	S5	S6
S1'	0.0126					
S2'		√				
S3'			√			
S4'				0.0137		
S5'					0.0121	
S6'						0.0152

$\|d_8-d_8'\|$	S1	S2	S3	S4	S5	S6
S1’	0.0074					
S2’		√				
S3’			√			
S4’				√		
S5’					0.0047	
S6’						0.0146

Table 7.6 Final Matching result

Result	S1	S2	S3	S4	S5	S6
S1’	√					
S2’		√				
S3’			√			
S4’				√		
S5’					√	
S6’						√

Table 7.7 Angle difference

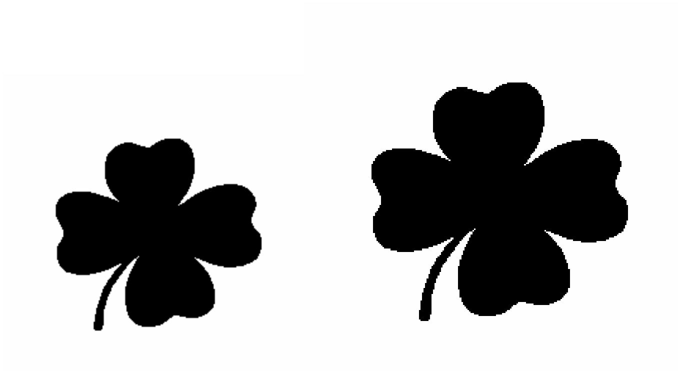
$\theta^i - \theta'^h$	S1	S2	S3	S4	S5	S6
S1’	44.54					
S2’		44.73				
S3’			45.34			
S4’				45.23		
S5’					45.93	
S6’						45.04

iii. Scale invariance

Scaling is one of the most critical problems in object recognition. For standalone object recognition, scale information can be obtained by the area or perimeter of the object, and then it can be normalized. However, in our work, both scaling and partial occlusion may happen simultaneously. It is impossible to obtain the scaling information by the area or perimeter anymore. The reason why scale is very critical in the recognition system is that it plays an important role in corner detection, like using curvature function to detection the corner, a proper size of Gaussian window must be applied according to the scale information. The proposed wavelet-based corner detection algorithm has been mathematically proven (Section 4.5) to be scale invariant (within a moderate range between 0.6 to 2.5). A program has been written to generate resized images which is shown in the Appendix. The original images are resized to scales between 0.4 to 4 to test the performance of the proposed recognition system.

Experiment 3.

Figure 7.7 shows a flower which has been magnified by 1.5 times from its original size. The corners are successfully detected as shown in Figure 7.8, and the flower are successful recognized subsequently as shown in Figure 7.9 and Tables 7.8-7.11. The scale information is also measured as shown in Table 7.12.



(a) (b)

Fig. 7.7 (a) model object-flower (b) program generated image which is resized by a random scale

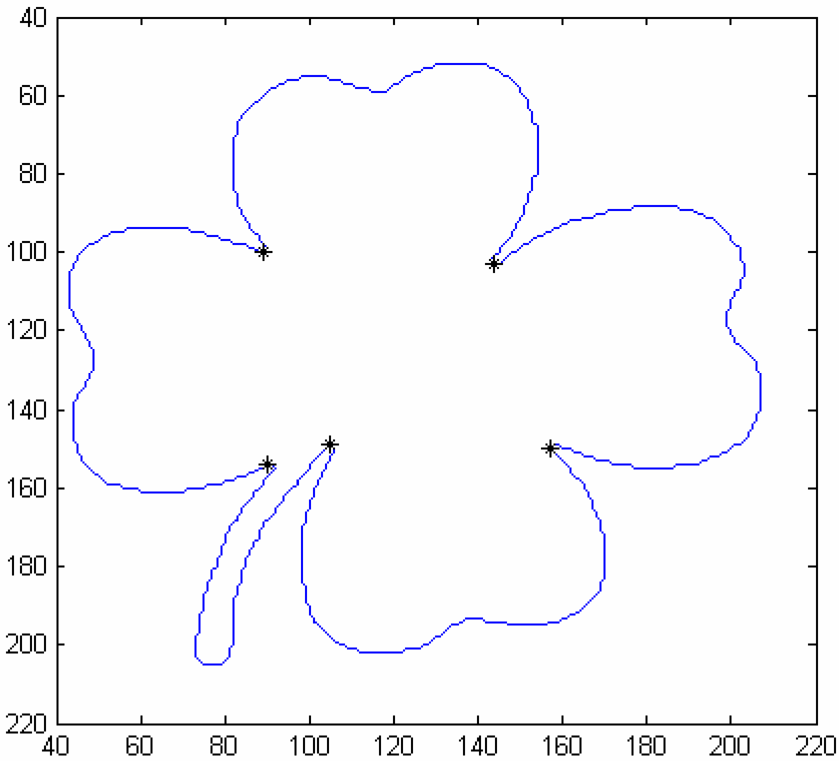


Fig. 7.8 Corner detection result of flower

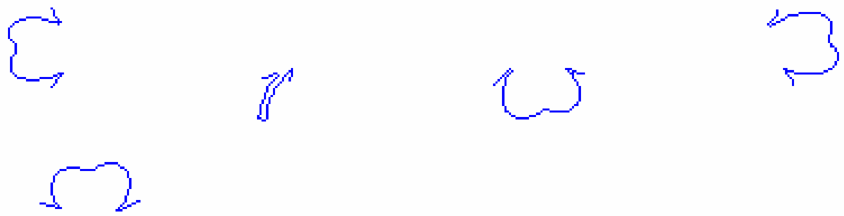


Fig. 7.9 Boundary partition result of flower

Table 7.8 dissimilarity value of scaling coefficients $||c_4-c_4'||$

$\|\|c_4-c_4'\|\|$	S1	S2	S3	S4	S5
S1’	0.4407	30.488	5.14	2.1907	3.2006
S2’	31.006	0.5947	30.44	29.628	30.232
S3’	5.3644	29.882	0.8952	6.0281	3.011
S4’	2.7321	29.001	5.4586	0.1517	3.7889
S5’	3.2393	29.966	2.3317	4.2542	0.5951

Table 7.9 dissimilarity value of the coarsest level wavelet coefficients **$||d_4-d_4^{'}||$**

$\|\|d_4-d_4^{'}\|\|$	S1	S2	S3	S4	S5
S1’	0.2033				
S2’		0.0092			
S3’			0.1767		
S4’				0. 1710	
S5’					0.1882

Table 710 dissimilarity value of the finer level wavelet coefficients

$\|\|d_5-d_5^{'}\|\|$	S1	S2	S3	S4	S5
S1’	0.1477				
S2’		0.1525			
S3’			0.1477		
S4’				0.1551	
S5’					0.1931

$\|\|d_6-d_6^{'}\|\|$	S1	S2	S3	S4	S5
S1’	0.1492				
S2’		0.1482			
S3’			0.1741		
S4’				0.1579	
S5’					0.1889

$\|\|d_7 d_7^{'}\|\|$	S1	S2	S3	S4	S5
S1’	0.1758				
S2’		0.1735			
S3’			0.1689		
S4’				0.1680	
S5’					0.1578

$\|\|d_8-d_8^{'}\|\|$	S1	S2	S3	S4	S5
S1’	0.1805				
S2’		0.0283			
S3’			0.1576		
S4’				0.0096	
S5’					0.1828

Table 7.11 Final segment matching result between resize flower and its original

Result	S1	S2	S3	S4	S5
S1’	√				
S2’		√			
S3’			√		
S4’				√	
S5’					√

Table 7.12 Scale difference between resize flower and its original

$L\ /L^{'}$	S1	S2	S3	S4	S5
S1’	1.5102				
S2’		1.5128			
S3’			1.4923		
S4’				1.5102	
S5’					1.4987

The proposed corner detection algorithm still tends to detect false corners when the image is excessively downsized. For instance, as shown in Figure 7.10, the size of the flower is reduced by a scale of 0.4 of its original size, a so call ‘false’ corner is detected at the end of the stem. On another extreme, when the image is overly upsized, some dull corners will not be detected. Such as the bull head shown in Figure 7.11 is enlarged 4 times of its original size. The two corners on its ears are not detected. Beside the extreme conditions, extensive experimental results show that our proposed corner detection algorithm detects consistent corners for scale within 0.6 to 2.5. The reason is because equation (4.5) is only applicable for small scale range of s (Mallat, 1992(b)), so all equations derived based on it will be subjected to this restriction.

Other than the corner detection, the proposed wavelet representation of the curve segment is strictly scale invariant. Large number of experimental results show that the proposed object recognition system performs well under moderate scaling between 0.6~2.5.

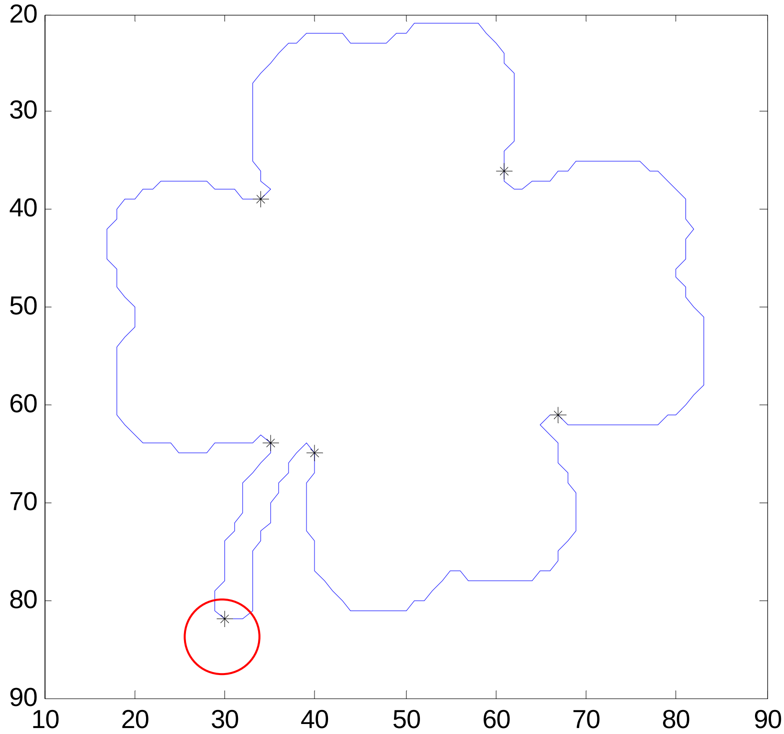


Fig. 7.10 Corner detection of flower which is downsized by 0.4

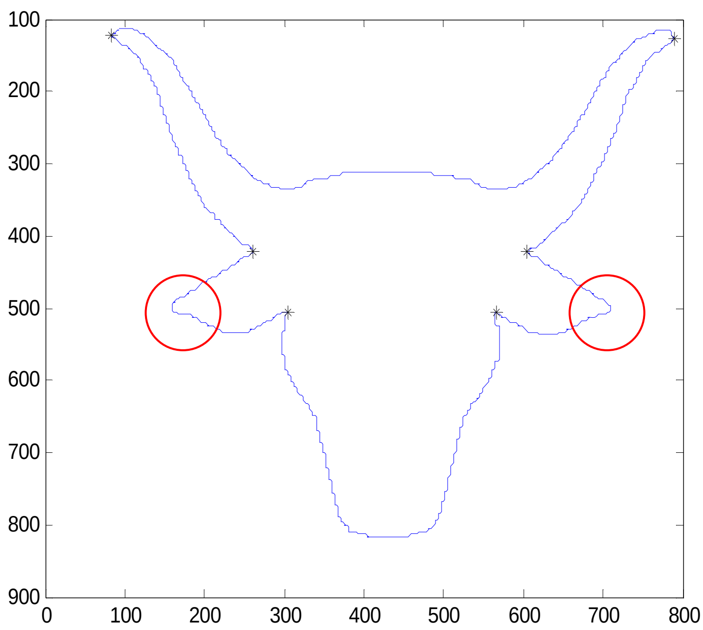
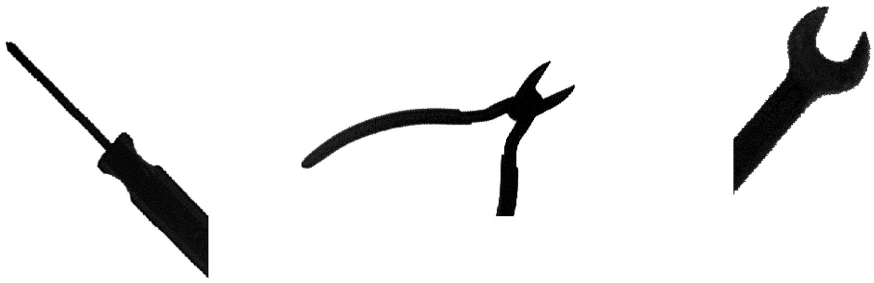


Fig. 7.11 Corner detection of bull head which is enlarged by 4

7.5 Partial occluded object recognition

Recognizing partially occluded objects is one of the main objectives of this research project, and it is also the most challenging task. Partial occlusion can be classified into two classes, one is part of the object is not visible such as the images shown in Figure 7.12, the other class is the object is overlapped by another object as shown in Figure 7.13. We have written two programs to generate these two classes of partially occluded object for testing. One for generating the images with part of the object is missing. It simply set a certain portion of the image to blank. And the other program puts two objects into one image randomly so that they are overlapped with random relative position.



(a) (b) (c)

Fig. 7.12 partial occluded objects which part of the object is unseen

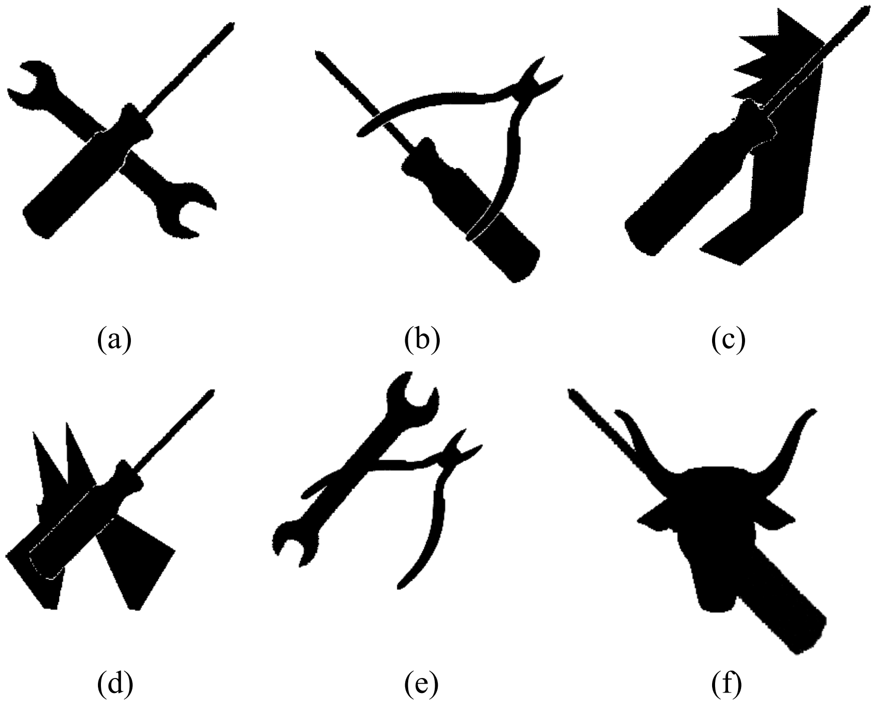


Fig. 7.13 partial occluded objects which are overlapped by each other

Experiment 4.

This experiment illustrates the performance of the recognition system on objects with part of them not visible. Figure 7.14 and 7.15 first show the corner detection and boundary partition result of model object-pliers. Figure 7.16 shows the corner detection result of the partially occluded pliers using proposed our wavelet-based corner detection algorithm. Figure 7.17 shows the boundary partition result based on the corners detected in Figure 7.16 using our proposed boundary partition algorithm.. Table 7.13 shows the dissimilarity values of the scaling coefficients between the model object-pliers and occluded pliers. All ineligible candidates are filtered out after only one stage. For space saving, the dissimilarity values of the wavelet coefficients are omitted. From Table 7.13, we can see that all intact segments are successfully matched with the model, while segment 3 and 4 are not matched because they are being occluded.

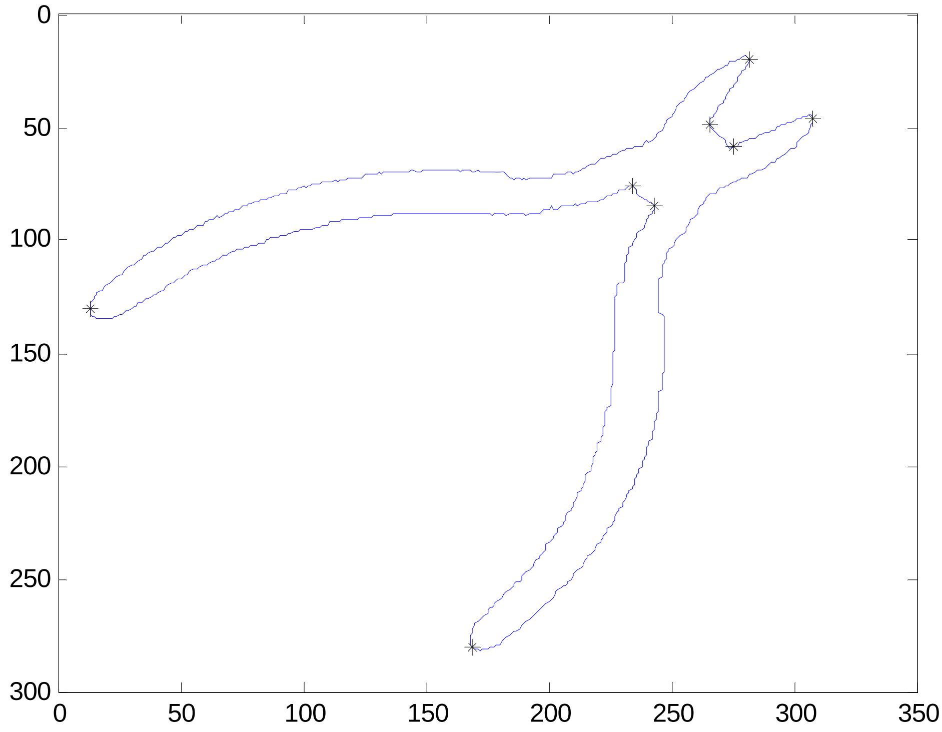


Fig. 7.14 Corner detection result of pliers

Fig. 7.15 Boundary partitioning result of pliers

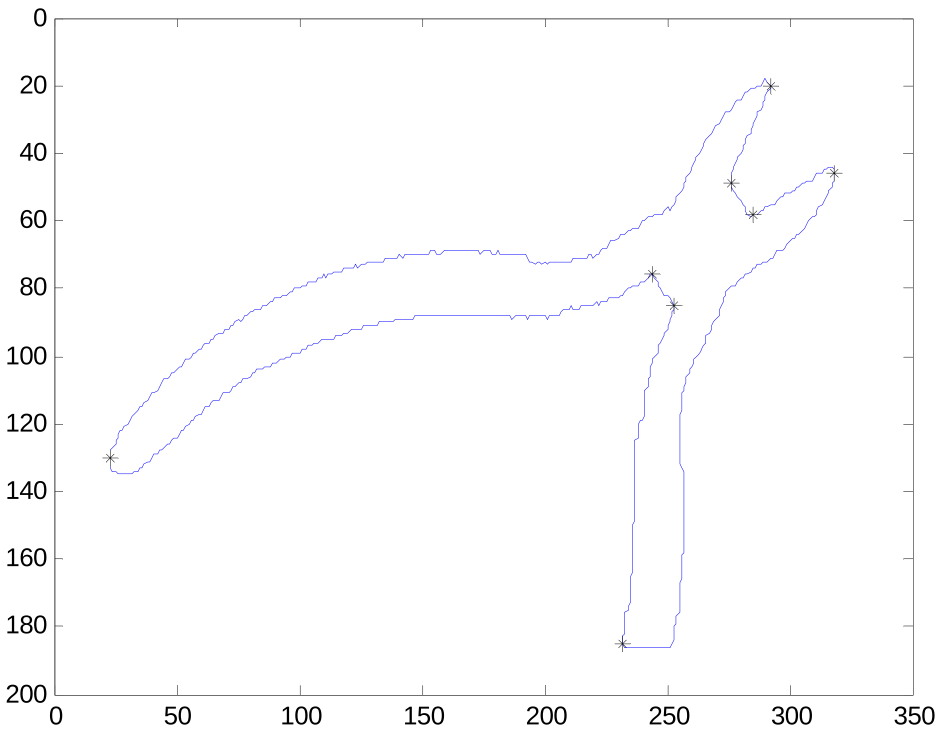


Fig. 7.16 Corner detection result of partial occluded pliers

Fig. 7.17 Boundary partitioning result of partial occluded pliers

Table 7.13 dissimilarity value of scaling coefficients ||c4-c4'|| between pliers and occluded pliers

$\|\|c_4-c_4'\|\|$	S1	S2	S3	S4	S5	S6	S7	S8
S1'	0	159.43	23.462	9.9398	75.83	159.64	74.514	11.473
S2'	28.184	0	21.193	30.551	6.4202	6.9373	7.9531	32.124
S3'	24.5	34.481	14.349	26.027	14.328	35.184	9.8125	26.99
S4'	18.109	76.259	9.058	19.152	32.825	76.72	33.22	22.965
S5'	26.81	12.84	20.031	28.827	0	13.548	5.8332	30.446
S6'	28.221	6.9373	21.195	30.695	6.7742	0	8.0593	32.275
S7'	26.345	15.906	17.688	28.413	5.8332	16.119	0	29.895
S8'	11.473	181.72	30.071	8.4901	86.115	182.58	84.555	0

Similar experiments have been done extensively, but we are unable to give a quantitive figure of recognition rate verses the percentage of occlusion. The reason is that the recognition rate does not only relying on the percentage of the boundary or area been occluded, but also more on how many consecutive corners pairs are intact. This is because the recognition system requires two consecutive corners to form a curve segment. For instance, the wrench shown below in Figure 7.18 only has two intact corners, which can only form one intact curve segment. According to the matching criteria set in Chapter 6, only one matched segment-pair can not confirm the presence of a model object in a scene. The proposed recognition system requires at least two intact segments to confirm the identity of the object in a scene.

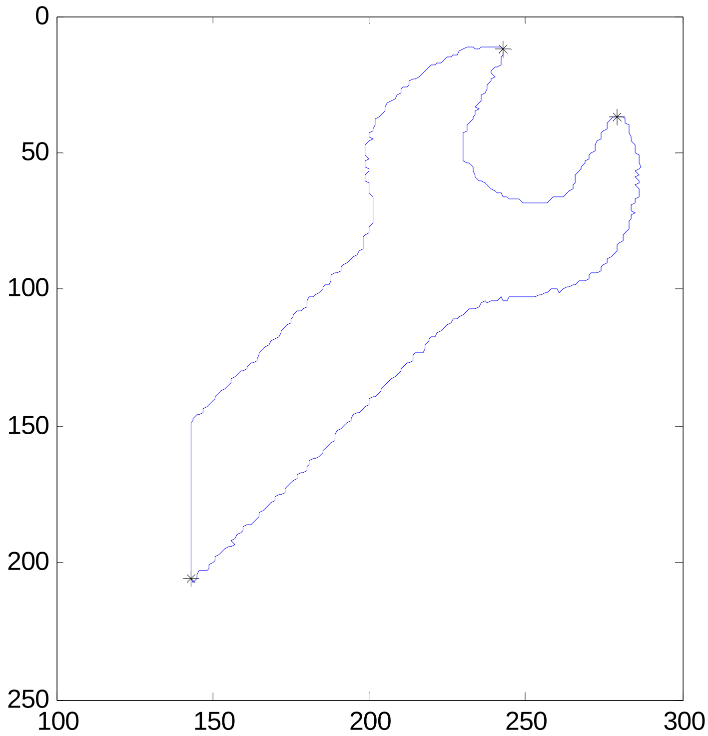


Fig. 7.18 Corner detection result of partial occluded wrench

Experiment 5.

To test the performance of the recognition system under arbitrary overlapping conditions, two model objects are randomly placed one on top of the other. We take Figure 7.13 (e) as an example which is a pliers overlapping a wrench. The corner detection result and boundary partitioning result are shown in Figures 7.19 and 7.20, respectively. The matching result between the overlapping objects with pliers and wrench are shown in Tables 7.14 and 7.15, matched segment pairs are found between overlapping objects with both pliers and wrench. Therefore, the recognition system can identify the objects in the scene which is a pair of pliers overlapping a wrench.

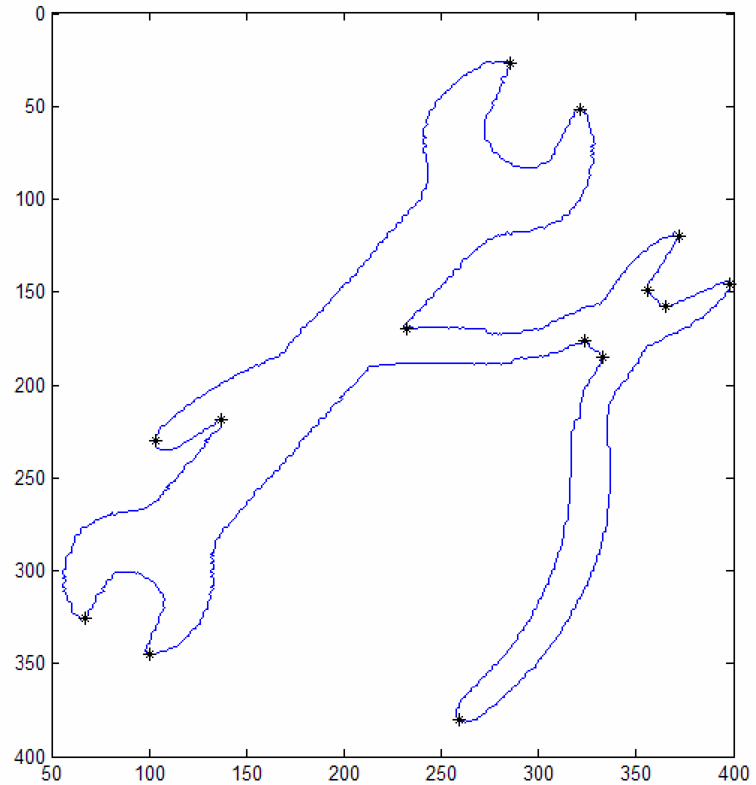


Fig. 7.19 Corner detection result of pliers overlapped with wrench

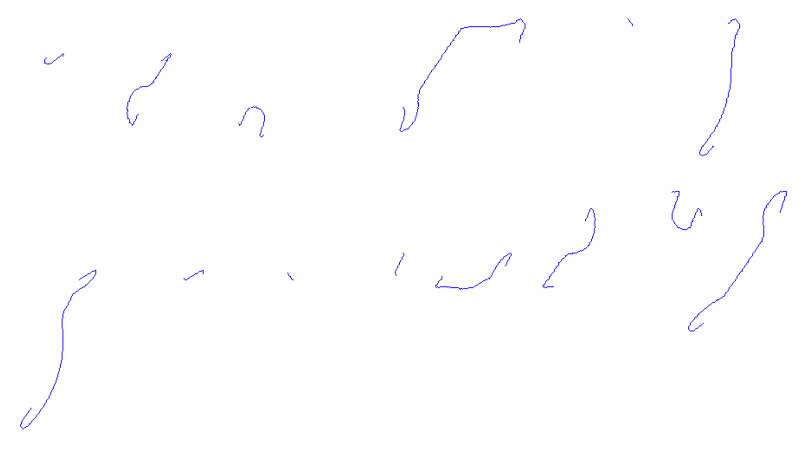


Fig. 7.20 Boundary partition result of pliers overlapped with wrench

Table 7.14 dissimilarity value of scaling coefficients ||c4-c4'|| between pliers and overlapping object

$\|\|c_4-c_4'\|\|$	S1	S2	S3	S4	S5	S6	S7	S8
S1'	27.345	17.624	20.823	28.512	2.9447	18.287	7.389	30.115
S2'	18.033	106.41	9.3922	18.144	49.613	106.6	46.287	17.803
S3'	23.194	42.951	13.37	27.495	19.592	42.953	14.085	29.399
S4'	7.3825	192.7	30.987	13.801	92.453	192.94	90.837	12.414
S5'	28.184	0	21.193	30.551	6.4202	0.9373	7.9531	32.124
S6'	16.59	84.773	0	19.144	40.063	84.781	35.375	21.263
S7'	9.9398	172.82	27.073	0	81.534	173.64	80.363	8.4901
S8'	26.81	12.84	20.031	28.827	0	13.548	5.8332	30.446
S9'	28.221	0.9373	21.195	30.695	6.7742	0	8.0593	32.275
S10'	26.345	15.906	17.688	28.413	5.8332	16.119	0	29.895
S11'	13.407	92.294	7.9989	21.083	41.887	92.503	41.116	22.717
S12'	17.188	102.43	7.1449	18.067	48.052	102.82	43.846	17.263
S13'	22.706	109.94	15.611	29.459	52.233	109.93	48.676	28.951
S14'	6.7199	157.99	23.543	5.1213	74.617	158.46	73.09	8.3588

Table 7.15 dissimilarity value of scaling coefficients ||c4-c4'|| between wrench and overlapping object

$\|\|c_4-c_4'\|\|$	S1	S2	S3	S4
S1'	108.34	84.344	95.858	90.719
S2'	91.538	89.014	96.963	47.179
S3'	0	85.488	95.768	51.864
S4'	62.41	81.35	103.46	48.47
S5'	79.954	37.656	91.75	84.881
S6'	60.24	60.461	93.711	101.75
S7'	93.617	78.582	35.491	66.39
S8'	70.633	25.832	60.414	19.058
S9'	49.933	47.11	89.786	67.761
S10'	108.3	19.794	44.789	75.858
S11'	99.878	92.367	91.181	73.869
S12'	36.768	39.968	85.072	94.259
S13'	81.025	24.309	0	62.724
S14'	63.449	20.762	18.572	87.649

We write a computer program to place one model object on top of another model object at random position. Using this program, we generate large number of partially occluded images to test the performance of our object recognition system. Table 7.16 shows the recognition rate of the object being occluded by another object at random position. We observe that, in general, the recognition rate is higher for objects with more number of corner points, e.g. pliers and plane. While objects with less number of corner points, in another word, less number of segments tend to be more difficult to identify.

Table 7.16 Recognition rate of object being overlapped by another object at random position

Occluded object	Overlapped by	Recognition rate
Bull head	Pliers	86.4
Pliers	Wrench	90.3
Wrench	Saw	51.8
Saw	Egg	78.9
Egg	Screwdriver	65.6
Screwdriver	Scissors	43.5
Scissors	Plane	68.3
Plane	Flower	86.9
Flower	Club	76.5
Club	Bull head	75.8

7.6 Partial occluded and scaled object recognition

Experiment 6.

Lastly, we use an experiment to demonstrate the performance of our proposed recognition system on an object which is partially occluded and also scaled. The combination of partial occlusion and scaling situation is difficult to solve in practice, and can be used to critically test our recognition system.

As our proposed corner detection algorithm is scaling and partial occlusion invariant, the corners on the intact portion of the bull head and screw driver can still be successfully detected as shown in Figure 7.21. There are also some corners generated by the intersection of the two objects detected. In addition, the boundary is partitioned using our proposed boundary partitioning method, and the result is shown in Figure 7.22. From this figure, we can see that only some intact complete segments are from the bull head. Table 7.17 shows the matching result between the bull head and the scaled and partially occluded bull head, 5 segment pairs between them are successfully found, their relative angle, position and length ratio are also confirmed as shown in Table 7.18, where the scaling factor of the bull head in the scene is about 1.3 times of the original one in the database. Therefore, we can conclude that the bull head is present in the scene, but with scaling and partial occlusion. However, the recognition system could not recognize the screwdriver due to insufficient consecutive corners belonging to the screwdriver can be detected.

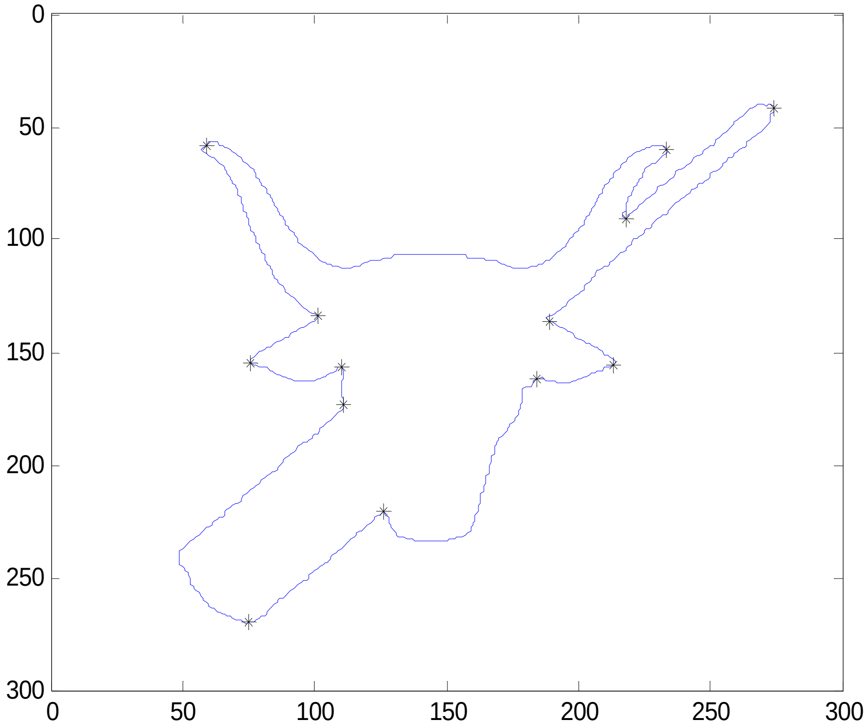


Fig. 7.21 Corner detection result of scaled bull head overlapped with screwdriver

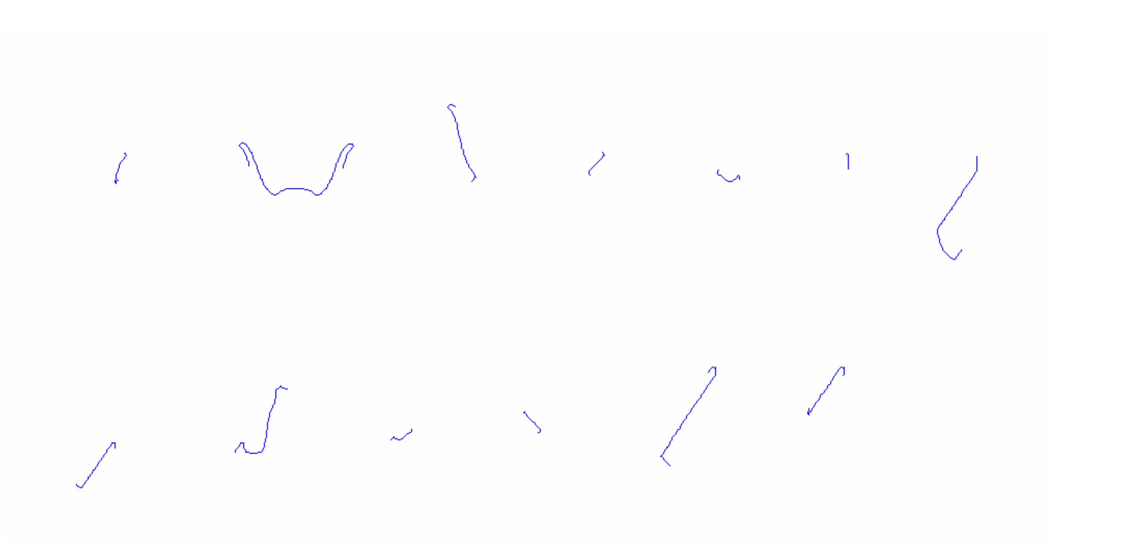


Fig. 7.22 Boundary partitioning result of scaled bull head overlapped with screwdriver

Table 7.17 dissimilarity value of scaling coefficients ||c4-c4'|| between model object-bull head and scaled and occluded bull head

$\|\|c_4-c_4'\|\|$	S1	S2	S3	S4	S5	S6	S7	S8
S1'	8.3395	6.0882	7.4976	28.869	6.7607	5.7597	10.305	33.584
S2'	60.93	92.203	92.65	45.925	94.595	92.142	60.4	0.5916
S3'	0.3684	8.4597	7.0756	24.201	10.111	7.1857	5.6266	31.059
S4'	6.6344	0.4028	4.2946	26.824	2.8006	3.8692	7.439	33.042
S5'	5.693	4.2676	0.8092	25.966	3.3519	2.3674	9.3457	33.258
S6'	9.0172	7.5243	8.4146	29.521	8.2003	6.9622	11.323	34.47
S7'	37.839	56.621	56.658	14.155	56.052	58.877	35.102	23.584
S8'	3.0497	12.148	10.123	22.591	13.562	9.6718	3.772	30.024
S9'	19.888	33.854	31.872	11.565	33.462	33.569	18.148	27.688
S10'	6.9636	2.5904	3.0439	25.887	2.2599	4.156	7.5615	33.415
S11'	5.3896	5.415	2.9929	26.974	5.1377	0.7489	9.7087	33.235
S12'	23.103	34.335	37.618	18.156	36.4	38.291	18.654	22.501
S13'	2.7864	11.16	8.8338	23.638	11.798	9.3932	6.309	31.491

Table 7.18 length ratio between the segments of the object in scene and the bull head in database

L/L'	S1	S2	S3	S4	S5	S6	S7	S8
S1'								
S2'								1.334
S3'	1.315							
S4'		1.305						
S5'			1.317					
S6'								
S7'								
S8'								
S9'								
S10'								
S11'						1.298		
S12'								
S13'								

7.7 Conclusion and discussion

The following conclusions can be derived from our experimental results:

1. The recognition algorithm is translation and rotation invariant.

   The recognition rate for translated and rotated images is 100%. This is due to:

   - Our proposed corner detection algorithm is invariant to translation and scaling.
   - Normalization process helps in the process.

2. The recognition algorithm is scaling invariant within the scale ranging from 0.6 to 2.5.

   The recognition algorithm can recognize scaled objects within the said range because:

   - Our proposed corner detection algorithm can tolerate scaling within a moderate range (0.6~2.5).
   - Normalization standardizes the curve length, and then re-samples it according to its original length.
   - Levels of wavelet decomposition of curve segments is determined by the size of the segment.
   - Dissimilarity function of curve segments with different lengths has been derived, and the scaling factor has been compensated.

   Our proposed corner detection algorithm may detect false corners and mis-detect corners when the scale is too small or too big. The reason is because equation

(4.5) only holds for small scale range of s, therefore our false corner elimination algorithm which relies on it only applicable for range between 0.6 to 2.5.

3. The recognition algorithm is computational efficient.

Our recognition system takes about 0.7 seconds to recognize an object in a 640x480 resolution image from a model database containing 20 model objects. (The program runs in Matlab 6.5 on a PentiumIV 1.6GHz PC). The reasons are:

- Fast wavelet transform algorithm is implemented.
- Hierarchical matching is adopted to eliminate most of the ineligible candidates at the early stage.

4. The recognition algorithm is robust to partial occlusion.

Our recognition algorithm is capable of recognizing both objects with missing portion and overlapping objects; it can also recognize partially occluded objects with similarity transformation due largely of the process we have adopted in work. We partition the object boundary into independent curve segments, and represent them separately. Partial occlusion only affects the feature of occluded segments, while the feature of intact segments remains the same.

5. The confidence level of our proposed recognition system is high.

As we use not only local features (wavelet descriptor) for recognition, but also the interrelationship between segments for verification, these two steps make sure that both local shape and relative position are the same for the object in the scene and the model object. Nevertheless, our algorithm fails in some cases involving occlusion (as the screw driver in experiment 7) since the recognition algorithm requires at least two intact segments from the object.

Chapter 8

Conclusion and Future Works

In this thesis, a complete wavelet based object recognition algorithm is presented and implemented. It can recognize not only standalone but also partially occluded two-dimensional objects. The algorithm consists of several components:

- A novel wavelet-based corner detection algorithm which is scaling invariant is proposed to facilitate boundary partitioning in order to the extract local features of standalone and partially occluded objects;
- A refined boundary segmentation method which is able to retain the corner information is specially designed;
- An adaptive normalization method is derived and implemented to make the curve segment similarity transformation invariant and to retain the original scale of the curve segment for precise matching.
- A multi-resolution wavelet-based feature extraction algorithm is proposed, which is also the core of our algorithm. We have shown that it is effective and the features extracted are compact in size.
- A hierarchical matching strategy is presented, in which the dissimilarity function is designed. This strategy enables the matching of segments at different size possible. The relative information between curve segments is used for verification, and the matching criteria are also defined.

From our experimental works, we have found that our algorithm exhibits several advantages over the existing systems. They include:

1) Fuller use of the localized shape information of the object, including the trend of the curve segment, dominant points information and the relative position information;
2) More consistent boundary partitioning results by using the proposed wavelet-based corner detection algorithm which uses the Lipschitz exponent as a measure to eliminate false corners.
3) Fast feature extraction due to the adoption of Fast Wavelet Transform which is very computational efficient.
4) Better accuracy because the object is represented and matched at its natural scale instead of a pre-defined scale;
5) More efficient matching, since most of the candidates are eliminated at early stage by our hierarchical matching method.

8.1 Contributions

The above advantages allow us to summarize our contribution from the work presented in this thesis:

1) We have developed a <u>novel wavelet based similarity transformation invariant corner detection algorithm</u>, which is applicable to both standalone and partial occluded objects.
2) We have specially designed a <u>new boundary partitioning method</u> requiring only two corners to form a segment, and hence resulting in smaller and hence more robust in occlusion situation.

3) A wavelet-based feature extraction algorithm is proposed to extract feature in multi-resolution, the coarse resolution carries more global information, and finer resolution feature retains the local details. These features are more robust to noise, and also facilitate the subsequent matching process, which is important in our work.

4) Making use of the multi-resolution nature of the features extracted, we have designed a Hierarchical Matching algorithm to match the curve segments of the object in a coarse to fine hierarchical strategy. Most of the ineligible curve segment candidates are filtered out at the coarse resolutions thereby reducing the computational load of our object recognition algorithm.

8.2 Future works

We see the following possible improvements to our algorithms, that we believe, if they could be realized, would further enhance the understanding and hence and applicability of our work.

1) Based on the current work, design an algorithm to handle object recognition problem in 3-dimension. The result would be very useful in industry such as sorting of parts stored in a bin, where large degree of object occlusion is present. In this regards, the new algorithm must take perspective transform into consideration.

2) We believe that we had not unleashed the potential of wavelet transform in solving the object recognition problem with occlusion. The ability of wavelets to handle image signal or representation with multi-resolution is extremely useful

in feature detection, followed by recognition. This thesis has only explored a small part this excellent property of wavelets.

3) To evaluate the performance of a partial occluded object recognition system, it is essential to have a quantitive measure of the degree of partial occlusion. The degree of the partial occlusion is a complicated term, it is not only related to the area of length of boundary been occluded, but also depends on the important features been occluded, such as corners. Therefore, developing a proper way to describe the degree of the partial occlusion is an important and challenging task in the future works priori to have a fair comparison with other algorithms.

BIBLIOGRAPHY

1. Anderson, I.M. and Bezdek, J.C., Curvature and tangential deflection of discrete arcs: a theory based on the commutator of scatter matrix pairs and its application to vertex detection in planar shapes, IEEE Trans. on Pattern Analysis and Machine Intell., 6, pp. 27-40, 1984

2. Ansari, N. and Delp, E. J. Partial Shape Recognition: A Landmark-based Approach, IEEE Trans. Pattern Analysis Mach. Intell. PAMI-125, 470-483 1990

3. Ansari, N. and Delp, E.J. On detecting dominant points, Pattern Recognition, 24, pp.441-451. 1991

4. Antoine, J.P., Barache, D., Cesar, J.R.M., Costa, L.D.F., Shape Characterization with the Wavelet Transform, Signal Processing, 62, pp. 265-290. 1997

5. Asada, H. and Brady, M. The curvature Primal Sketch, IEEE Trans. on Pattern Analysis and Machine Intell., 8, pp. 2-14, 1986

6. Attneave, F. Some informational aspects of visual perception, Psycholo. Rev., 61, pp.183-193. 1954

7. Ayache, N. and Faugera, O.D. HYPER: A New Approach for The Recognition and Position of Two-Dimensional Objects, IEEE, Trans. Pattern Analysis Mach Intell. PAMI-8, 44-54. 1986

8. Behis, G.N. and Papadourakis, G.M. Object recognition using invariant object boundary representation and neural network models, Pattern Recognition, 25, pp.25-44. 1992

9. Beus, H.L. and Tiu, S.S.H., An Improved Corner Detection Algorithm Based on Chain-Coded Plane Curves, Pattern Recognition, 20, pp. 291-296, 1987

10. Bhanu, B. and Faugeras, O.D. Shape Matching of Two-Dimensional Objects, IEEE Trans. Pattern Analysis Mach. Intell., PAMI-6, pp. 137-156. 1984.

11. Bhanu, B. and Ming, J.C. Recognition of Occluded Objects: A Cluster-Structure Algorithm, Pattern Recognition, 202, pp. 199-211, 1987

12. Bolles, R.C. and Cain, R.A. Recognizing and Locating Partially Visible Objects: The Local-Feature-Focus Method, Int. j. Robotics Res., 1, pp.57-81. 1982

13. Bookstein, F.L. A Statistical Method for Bilological shape Comparision, J. Theor. Biol. 107, pp. 75-520. 1984

14. Bui, T.D. and Chen, G. Invariant Fourier-Wavelet Descriptor for Pattern Recognition. M. Sc. Thesis, Computer Science Department, Concordia University, Canada, 1999

15. Burt, P.J. and Adelson, E.H., The Laplacian Pyramid as a Compact Image Code, IEEE Trans. on Comm., 31(4), pp. 532-540, 1983

16. Canny, J., A Computational Approach to Edge Detection, IEEE Trans. on Pattern Analysis and Machine Intell., 8(6), pp.679-698, 1986.

17. Chen, C.C. Improved Moment Invariants for Shape Discrimination, Pattern Recognition, Vol. 26, No. 5, 683-686, 1993

18. Chen, C.C., Improved moment invariants for shape discrimination, Pattern Recognition, 26, pp.167-174. 1993.

19. Cheng, F. and Hsu, W., Parallel algorithm for corner find on digital curves, Pattern Recognition Lett., 8, pp. 47-53, 1988

20. Chuange, G.C.H. and Kuo, C.C.J. Wavelet descriptor of planar curves: theory and applications, IEEE Tans. on Image Processing, 5, pp.56-70. 1996

21. Chui, C.K. An Introduction to Wavelets, Boston: Academic Press, 1992

22. Daubechies, I. Orthonormal Bases of Compactly Supported Wavelets, Comm. On Pure & Applied Mathematics, 41(7), pp.909-996, 1988

23. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

24. Donoho, D. L. and Johnstone I.M., Ideal denoising in an orthonormal basis chosen from a library of bases, Comptes Rendus Aca. Sci. Paris A , 319, pp. 1317–1322, 1994

25. Donoho, D. L., De-noising by soft-thresholding, IEEE Trans. Info. Theory, 41, pp. 613-627, 1995.

26. Eric, W. and Grimson, L. Correspondence: On the Recognition of Curved Objects, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-116, 632-643. 1989

27. Etesami, F. and Uicker, J.J. Jr., Automatic Dimensional Inspection of Machinepart Cross-Section using Fourier Analysis, Comput. Vision Graphics Image Process, 29, 216-247 1985

28. Ettinger, G. J. Large hierarchical object recognition using libraries of parameterized model sub-parts. in Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition, Ann Arbor, MI, June 5- 9,pp. 32-41. 1988.

29. Fischler, M.A. and Bolles, R.C. Perceptual organization and curve partitioning, IEEE Trans. on Pattern Analysis and Machine Intell., 8(1), pp. 100-105. 1986

30. Fischler, M.A. and Wolf, H.C., Locating perceptually salient points on planar curves, IEEE Trans. on Pattern Analysis and Machine Intell., 16(2), pp. 113-129. 1994

31. Freeman, H. and Davis, L.S. A corner finding algorithm for chain-code curves, IEEE Trans. Comput. 26, pp. 297-303, 1977

32. Fridtjof and Medioni, G. Structural Indexing: Efficient 2-D Object Recognition, IEEE Trans. Pattern. Anal. Mach. Intell. 14(12),pp.1198-1204. 1992

33. Gorman, J.W., Mitchell, O.R. and Kuhl, F.P. Parial shape recognition using dynamic programming, IEEE Trans. Pattern Analysis Mach. Intell., 102, pp.257-266. 1988

34. Goudail, F. and Refregier, P. Statistical Image Processing Techniques for Noisy Images: An Application-Oriented Approach, New York, Kluwer Academic / Plenum Publishers, 2004.

35. Grimon, W.E.L. On Recognition of Curved Objects. IEEE Trans. Pattern Anal. Machine Intell., 11(6), pp. 632-643, 1989

36. Grossman, A. and Morlet, J. Decmposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. SIAM J. Appl. Math., 15, pp. 723-736. 1984

37. Haig, T.D., Attikiouzel, Y. and Alder, M.D., Border following new definition gives improved borders, IEE Proc., 139(2), pp.206-211, 1992

38. Han, M.H. and Jang, D.S., The Use of Maximum Curvature Points for the Recognition of Partially Occluded Objects, Pattern Recognition, Vol. 23, No. ?, 21-33. 1990

39. Hu, M.K., Visual Pattern Recognition by Moment Invariants, IRE Trans. On Inf. Theory, Feb., 179-187, 1962

40. Hua, J.P. and Liao, Q.M. "Wavelet-based multiscale corner detection," Proc. ICSP 2000, Beijing, China, August, 2000.

41. Kalvin, A., Schonberg, E., Schwartz, J.T., and Sharir, M. Two-dimensional, model based, boundary matching using footprints. The International Journal of Robotics Research, 5(4), pp.38-55, 1986.

42. Karp. R.M. Complexity of computer computations, chapter Reducibility Among Combinatorial Problems, 85--103. Plenum Press, New York, 1972

43. Katzir, N., Lindenbaum, M. and Porat, M., Curve Segmentation Under Partial Occlusion, 16(5), pp.513-519, 1994

44. Kauppinen, H., Seppanen, T. and Pietikainen, M. An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification, IEEE Trans. on pattern analysis and machine intelligence, Vol. 17, pp.201-207. 1995.

45. Kauppinen, H., Seppanen, T. and Pietikainen, M. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. IEEE Trans. Pattern Anal. Machine Intell. 172, pp. 201-207. 1995

46. Kendall, D.G. Shape manifolds, Procrustean metrics, and complex projective spaces, Bull. London Math. Soc.16, pp. 81-121. 1984

47. Kendall, D.G. The Diffusion of Shape, Adv. Appl. Probab., 9, pp.428-430. 1977

48. Kenneth R. Castleman.Digital image processing / Englewood Cliffs, N.J. : Prentice Hall , c1996

49. Khalil, M.I. and Bayoumi, M.M. Invariant 2D Object recognition using the wavelet modulus maxima, Pattern Recognition Letters, 21 , pp.863-872. 2000

50. Khotanzad, A. and Hong, Y.H.Invariant Image Recognition by Zernike moment. IEEE Trans. Pattern Analysis and Machine Intelligence, 12, pp.489-497. 1990

51. Knoll, T. F. and Jain, R. C. Recognizing partially visible objects using feature indexed hypotheses. IEEE Journal of Robotics and Automation, 2(1), pp.3-13, 1986.

52. Koch M. W. and Kashyap, R.L. Using Polygons to Recognize and Locate Partially Occluded Objects, IEEE Trans. Pattern Anal. Machine Intell. 9(4), pp.483-494. 1987

53. Lamdan, Y., Schwartz, J.T.and Wolfson, H.J. Affine invariant model-based object recognition,IEEE Trans. On Robotics and Automation,6, pp.578-589,1990

54. Lee, J.S., Sun, Y.N. and Chen, C.H. Multiscale corner detection by wavelet transform, IEEE Trans. on Image Processing, 41, 100-104, 1995

55. Lee, J.S., Sun, Y.N., Chen, C.H., Tsai, C.T., Wavelet based corner detection. Pattern Recognition, 26, pp.853-865. 1993.

56. Lim, K.B. Du, T.H. and Zheng, H. 2-D Partially Occluded Objects Recognition using Curve Moments, In Conf. Proce. Of Seventh International Conference on Computer Graphics and Imaging, pg. 303-308, Hawaii USA, 2004

57. Lim, K.B., Xin, K. and Hong, G.S. Detection and Estimation of Circular arc Segments. Pattern Recognition Letters 16, pp.627-636. 1995(b)

58. Liu, H. C. and Srinath, M. D., Partial Shape Classification using Contour Matching in Distance Transformation, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1211, 1072-1079.1990

59. Mahmoud, S. Arabic character-recognition using fourier descriptors and character contour encoding, Pattern Recognition, 27, pp.815-824. 1994.

60. Mallat, S. A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 117, pp.674-693. 1989(a)

61. Mallat, S. A Wavelet Tour of Signal Processing, Academic Press, 1998

62. Mallat, S. and Hwang, W.L., Singularity Detection and Processing with Wavelets, IEEE Trans. on Info. Theory, 38(2), pp. 617-643, 1992(a)

63. Mallat, S. and Zhong, S. Characterization of Signals from Multiscale Edges, IEEE Trans. Pattern Analysis Mach. Intell. 147, 710-732, 1992(b)

64. Mallat, S. Multiresolution Approximations and wavelet orthonormal bases of $L^2(R)$, Trans. Amer. Math. Soc., 315, pp.69-87, 1989(b)

65. Mallat, S. Multiresolution representations and wavelets, Ph. D. thesis, Department of Electrical Engineering, University of Pennsyvania. 1988

66. Mokhtarian, F. and Mackworth, A.K. A Theory of Multi-scale, curvature-based representation for planar curves. IEEE Trans. Pattern Analysis and Machine Intelligence,14, pp.789-805.1992

67. Mokhtarian, F. and Mackworth, A.K. Scale-based Decription and Recognition of planar curves and Two-dimensional shapes, IEEE Trans. Pattern Analysis and Machine Intelligence,8, pp. 34-43.1986

68. Mokhtarian, F. Silhouette-Based Isolated Object Recognition through Curvature Scale Space. IEEE Trans. Pattern Analysis and Machine Intelligence, 17, pp. 539-544. 1995

69. Pavilidis, T. and Horowitz, S.L. Segmentation of palne curve, IEEE Trans. Comput., C-23, pp.860-870, 1974

70. Pei, S.C. and Lin, C.N. The detection of dominant points on digital curves by scale-space filtering, Pattern Recognition, 25, pp.1307-1314. 1992

71. Persoon, E. and Fu, K. S., Shape Discrimination using Fourier Descriptors, IEEE Trans. Syst. Man Cybern. 7, Mar., 170-179. 1977

72. Price, K.E. Matching Closed Contours. In Proc. Seventh Int. Conference on Pattern Recognition, July 30-Aug. 2, Montreal, P. Q., Canada. pp. 990-992. 1984.

73. Quddus, A. and Gabbouj, M., Wavelet-based corner detection technique using optimal scale. Pattern Recognition Letters, 23, pp. 215-220, 2002

74. Quddus, A., Fahmy, M.M., Fast wavelet-based corner detection technique. IEE Electron. Lett. 35, pp.287-288, 1999.

75. Rattarangsi, A. and Chin, R.T. Scale-based detection of corners for planar curves, IEEE Trans. Pattern Anal. Mach. Intell., 14, pp. 430-449. 1992

76. Richard, C. W. Jr. and Hamami, H., Identification of Three Dimensional Objects using Fourier Descriptors of the Boundary Curve, IEEE, Trans. Syst. Man Cybern, SMC-44, 371-378. 1974

77. Richard, O. D., Peter, E. H., and David G. S. Pattern Classification, Second Edition, New York, Wiley, 2000.

78. Rosenfeld, A. and Johnston, E. Angle detection on digital curves, IEEE Trans. Comput. C-22, pp. 875-878, 1973

79. Rosenfeld, A. and Weszka, J.S. An improved method of angle detection on digital curves, IEEE Trans. Comput. C-24, pp. 940-941, 1975

80. Salari,R. and Balaji, S. Recognition of Partially Occluded Objects Using B-Spline Representation, Pattern Recognition, 24 (7), pp.653-660, 1991

81. Sankar, P.V. and Sharma, C.V., A Parallel Procedure for the Detection of Dominant Points on a Digital Curve, Comput. Vision Graphics Image Process, pp. 403-412, 1978

82. Sarkar, B., Singh, L.K., and Sarkar, D., Approximation of Digital Curves with Line Segments and Circular Arcs using Generic Algorithms, Pattern Recognition Letters, 24, pp. 2585-2595, 2003

83. Stansfield, J.L. Conculsion from the commodity expert project. AI Memo No. 601, MIT AI Lab, Cambridge, Mass., 1980

84. Tang, Y.Y., Liu, J., Yang, L.H.and Ma, H. Wavelet Theory and Its Application to Pattern Recognition, 2000 - Singapore; River Edge, NJ: World Scientific

85. Teh, C.H. and Chin, R.T. On Image Analysis by the Method of Moments, IEEE Trans. Pattern Analysis and Machine Intelligence, 2, pp.583-588. 1980

86. Tieng Q.M. and Boles, W.W. Recognition of 2D object contours using the wavelet transform zero-crossing representation, IEEE Transaction on Pattern Analysis and Machine Intelligence, 19, pp. 910 – 916. 1997

87. Tsang, K.M. Recognition of 2D standalone and occluded objects using wavelet transform, IEEE Trans. Pattern Anal. Machine Intell., 154, pp.691-705. 2001

88. Tsang, P.W.M., Yuen, P.C. and Lam, F.K., Classification of Partially Occluded Objects Using 3-Point Matching and Distance Transform, 27(1), pp.27-40. 1994

89. Tsang, P.W.M., Yuen, P.C. and Lam, F.K., Recognition of Occluded Objects, Pattern Recognition 25, pp. 1107-1117. 1992

90. Turney, J.L., Mudge, T.N. and Volz, R.A. Recognizing Partially Occluded Parts, IEEE Trans. Pattern. Anal. Mach. Intell. 7(4),pp.410-421. 1985

91. Walter G.G. and Shen, X. Wavelet and other orthogonal systems, 2nd edition, CRC press, Boca Roton, 2000

92. Wang, S., Chen, P. and Lin, W. Invariant pattern-recognition by moment Fourier descriptor, Pattern Recognition, 27, pp.1735-1742. 1994.

93. Wei, X.F. Recognising Two-dimensional Object Shapes using Prominent Visual Features, Master Thesis, National University of Singapore,1998.

94. Wen, W. and Lozzi, A. Recognition and inspection of two-dimensional industrial parts using subpolygons, Pattern Recognition, 25, pp.1427-1434. 1992

95. Witkin, A.P. Scale-space filtering, in Proceedings of the International Joint Conference on Artificial Intelligence. Palo Alto: Kaufman, pp. 1019-1022, 1983.

96. Woods, J.W. and S.D. O'Neil, Subband Coding of Images, IEEE Trans. ASSP, 34, pp.1278-1288, 1986

97. Wuescher, D.M. and Boyer, K.L. Robust contour decomposition using a constant curvature criterion, IEEE Trans. on Pattern Analysis and Machine Intell., 13(1), pp. 41-51. 1991

98. Xin, K., Lim, K.B. and Hong, G.S. A Scale-Space Filter Approach for Visual Feature Extraction, Pattern Recognition, Vol. 28, No. 8, pp. 1145-1158. 1995(a)

99. Yoon, S.H., Kim, J.H., Alexander, W.E., Park, S.M. and Sohn, K.H. An Optimum Solution for Scale-Invariant Object Recognition based on the Multiresolution Approximation, Pattern Recognition, 31, pp.889-908. 1998

100. You, Z. and Jain, A.K. Performance evaluation of shape matching via chord length distribution, Comput. Vision, Graphics Image Process. 28, pp. 185-198. 1984

101. Yu, T., Lam,E.C.M. and Tang, Y.Y., Fracture Extraction using Wavelet and Fractal, Pattern Recognition Lett., 22, pp.271-287. 2001

102. Zhang, D. and Lu, G. Review of shape representation and description techniques, Pattern Recognition, 37, pp.1-19. 2004

103. Zhang, J., Zhang, X., Krim,H. and Walter, G.G. Object representation and recognition in shape spaces, Pattern Recognition, 36, pp. 1143-1154. 2003.

104. Zhao, D.M., Chen, J., Affine curve moment invariants for shape recognition, Pattern Recognition, Vol. 30, Issue 6, Jun., 895-901, 1997

Appendix

(a) Image Random Translation

```
%%This function translate a image to a random position deviate 0-100 pixels
%% on its x and y corrdinates relative to its original position

function I_trans=translation(image);
I=imread(image);
rand_x=ceil(rand(1)*100);
rand_y=ceil(rand(1)*100);
[size_I_x,size_I_y]=size(I);
I_trans=255*ones(size_I_x+rand_x,size_I_y+rand_y)
for i=1:size_I_x
    for j=1:size_I_y
        I_trans(i+rand_x,j+rand_y)=I(i,j);

    end
end
imshow(I_trans);
```

(b) Image Rotation

```
function  rotate( imagename,interval,name);
% Rotate(imagename,number,interval,name)
% imagename:    the image which are going to be rotated
% number:       number of images which are going be generated
% interval:     the rotate angle interval
% name          prefix of the name of rotated images
I = imread(imagename);
[s_x,s_y]=size(I);

for i=1:s_x
    for j=1:s_y

        I(i,j)=255-double(I(i,j));

    end
end

number = floor(360/interval);

for angle=1:number
    angle
    J = imrotate(I,angle*interval,'bilinear');
```

```
    [s_x_J,s_y_J]=size(J);
    for i=1:s_x_J
        for j=1:s_y_J

            J(i,j)=255-double(J(i,j));
        end
    end



    imname=strcat(char(name),'_',num2str(angle),'.tif');
    imwrite (J, imname,'tif');
end
```

(c) Image scaling

```
%%This function resize a image at a given scale with respect
%%to its original size
function I_resize=resize(image,scale);
I=imread(image);
I_resize = imresize(I,scale);
imshow(I_resize);
```

List of Publication:

1. 2-D Occluded Object Recognition Using Wavelets, The Fourth International Conference on Computer and Information Technology (CIT'04), pg. 227-232, Wuhan China, 2004
2. Comparison of the Support Vector Machine and Relevant Vector Machine in Regression and Classification, International Conference on Control, Automation, Robotics and Vision (ICARCV), KunMing China, 2004
3. 2-D Partially Occluded Objects Recognition using Curve Moments, Seventh International Conference on Computer Graphics and Imaging, pg. 303-308, Hawaii USA, 2004
4. Bayesian Kernel Inference for 2D Objects Recognition Based on Normalized Curvature, Proceeding, 12th International Multi-Media Modeling Conference, Beijing China, 2006
5. A Wavelet Approach for Partial Occluded Object Recognition, the 1st International Symposium on Digital Manufacture(ISDM'2006), Wuhan,, China, 2006 (Submitted)
6. Partial Occluded Object Recognition, Pattern Recognition (Submitted)