THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

Human Face Recognition based on a Single Front View

Lin Kwan Ho

A dissertation* submitted in partial fulfillment of the requirements for the degree of

Master of Philosophy

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University

April 2002

*This research work was supported by The Hong Kong Polytechnic University Grand A/C No. G-V498

Pao Yue-Kong Library
PolyU • Hong Kong

Abstract

Human face recognition is one of the most useful techniques for identifying or authenticating a person. Although research on this topic has been conducted for more than twenty years, many problems still remain, and better techniques for facial feature detection and face recognition are needed. Therefore, the objectives of this thesis are to devise and develop efficient methods for preprocessing facial images and recognizing human faces. In this thesis, different approaches for human face detection, facial feature extraction and human face recognition are reviewed. Human face detection and facial feature extraction are the preprocessing steps for automatic human face recognition. Their accuracy will directly affect the performance of the recognition system. However, since the location of a face, its facial expression and lighting conditions in an image are unknown, and considering that its size and orientation may be different, the recognition procedure is difficult and computationally intensive. Thus, human face recognition is a challenging research topic.

In this research, we propose a fast approach based on valley field detection and a modified fractal dimension to extract an eye pair in a complex background, which can then be used to represent a face region. Instead of searching the whole image space to determine the scale of a face, only possible eye pairs as detected by the valley field and their local properties are investigated. These possible eye pairs are then identified by means of the modified fractal dimension. Furthermore, in order to improve detection reliability, uneven lighting conditions on the two halves of a face are normalized by means of a histogram technique. The corresponding average fractal dimensions of the binarized eye-pair regions and the face regions are then used to verify whether the eye pairs selected are valid.

Human face recognition techniques focusing on whole face and facial features such as the eyes and mouth have been proposed. Due to the fact that different facial regions have different degrees of importance for face recognition, a new modified Hausdorff distance is proposed. This distance measure incorporates the *a priori* structure of a human face to emphasize the importance of facial regions. The face recognition technique proposed in this thesis is computationally simple and can provide a reasonable performance level.

Acknowledgements

I would like to thank my Chief Supervisor, Dr. K. M. Lam, and my co-supervisor, Prof. W.C. Siu,. for guiding and encouraging me from the outset of my degree studies. Without their support, this research work would not have been completed. They also offered me many invaluable ideas and suggestions in writing my thesis.

I would also like to thank all members of the DSP Research Laboratory, past and present. The countless discussions I had with them have proved to be both fruitful and inspiring.

I gratefully acknowledge the financial support of The Hong Kong Polytechnic University through the Research Grants Council of the Hong Kong Special Administrative Region (Project No. G-V498), which made possible this research by granting me a studentship over the course of my degree program.

Finally, I would also like to take this opportunity to thank the Centre for Multimedia Signal Processing of the Department of Electronic and Information Engineering.

Table of Contents

List of Figures

List of Tables

Author's Publications

The following technical papers have been published or submitted for publication based on the result generated from this work.

Journal paper:

1. Kwan-Ho Lin, Kin-Man Lam and Wan-Chi Siu, "Locating the Eye in Human Face Images Using Fractal Dimensions", *IEE Proceedings - Vision, Image and Signal Processing,* vol. 148, no. 6, pp. 413-421, 2001.

Conference papers:

2. Kwan-Ho Lin, Kin-Man Lam and Wan-Chi Siu, "A Reliable Method for Extracting the Chin Contour", *Proceeding The First IEEE Pacific-Rim Conference on Multimedia*, pp. 410-413, 2000.

3. Kwan-Ho Lin, Baofeng Guo, Kin-Man Lam and Wan-Chi Siu, "Human face recognition using a spatially weighted modified Hausdorff distance", *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 477-480, 2001.

4. Kwan-Ho Lin, Kin-Man Lam and Wan-Chi Siu, "Automatic Human Face Recognition System Using Fractal Dimension and Modified Hausdorff Distance", *Proceeding The Second IEEE Pacific-Rim conference on Multimedia,* pp. 277-284, 2001.

Chapter 1
Introduction

The objectives of this chapter are to introduce the general concepts of face recognition and its preprocessing steps, as well as its applications. The state-of-the art technology for human face recognition will be presented. An overview of the techniques for face recognition, facial feature extraction, and face recognition techniques proposed and developed in this thesis will also be given.

1.1 Motivation of Automatic Human Face Recognition

Nowadays, the term 'biometrics' is used to refer to the study of automated methods for the identification of individuals using physiological characteristics. Although many possible approaches can be used to identify a person, such as speech recognition, iris scanning, signature verification and fingerprint scanning, face recognition has the advantage of being non-intrusive and requiring little co-operation from the person to be identified to collect useful data for recognition. However, the human face can have many variations, including perspective variations, facial expressions, uneven lighting conditions, scales and orientations. Hence, automatic human face recognition is a challenging research topic that has been studied for more than 20 years.

1.2 Introduction to the Automatic Human Face Recognition Techniques

Automatic human face recognition can be divided into two steps: preprocessing and face recognition. The preprocessing step includes the procedures to detect the existence and position of the human face, and then those used to extract the important facial features. Face detection is an important step whereby the human face in a source image or video scene is located. The human face may appear at any position in an image, be of any size and orientation, and experience varied lighting conditions and facial expressions. The task of face detection is therefore a difficult and computationally intensive process. In order to perform face detection efficiently and accurately, algorithms based on the template-matching approach, knowledge-based approach, feature-based approach, shape information approach, etc. have been proposed. Facial feature extraction is part of the preprocessing step in automatic human face recognition. Face detection and facial feature extraction always mix with each other. In some approaches, important facial features are detected before the corresponding human face is located. In other approaches, a human face is detected, and then, based on the structure of a human face, the respective facial features can be extracted. Facial feature detection and extraction consist mainly of two approaches: the feature-based method and the deformable template. The feature-based method identifies the facial features based on their properties. However, this method is computationally intensive and non-flexible to scene variations. The deformable template is an approach to modeling the boundary shape of natural objects such as the eyes and mouth. However, the deformable template approach is associated with problems such as slow convergence and a lengthy processing time. Consequently,

efficient and reliable methods for face and facial feature detection are still under investigation.

In an automatic human face recognition system, a face region is extracted and then normalized based on the position of the two eyes. The normalized human face is aligned with those human faces in a database, and they are then compared. In other words, the accuracy of face detection and facial feature extraction will affect the performance of an automatic human face recognition system significantly.

Many techniques for human face recognition have been proposed, and can be divided into two categories: analytic-based and holistic-based approaches. In the analytic-based approach, recognition is based on the relationship between the human facial features such as the eyes, mouth, nose and face boundary. The analytic-based approach can provide high flexibility in handling non-rigid facial features such as the eyes and mouth. However, the accuracy of the facial feature extraction will affect greatly the success of this approach. The holistic-based approach takes into account the global properties of a pattern. One of the typical holistic-based techniques is principal component analysis (PCA), which represents a human face by a linear combination of weighted eigenvectors. However, to achieve a reasonable performance, the images in consideration must also be aligned to each other.

1.3 Our Methods for Automatic Human Face Recognition

The objectives of this research are to investigate and develop efficient techniques for automatic human face recognition and to construct an automatic human face recognition system. The crucial step for automatic human face recognition is to locate the position of the two eyes in a facial image. This is used for normalization and

alignment. The recognition rate of the system is related to the accuracy of locating the two eyes. Furthermore, the face database of the system may contain thousands of human faces. The efficiency of the method used for face recognition is therefore also an important issue for the practical use of this technology. Psychological studies showed that the edges of an object contain important information about its shape and structure, and can be used in face recognition. Hence, we use edge-like retinal images of faces as the input to our human face recognition system.

The system consists of two major parts: the first part is human face detection and facial feature extraction, while the second part is human face recognition based on the edges of faces and the properties of the facial features. In the first part, the respective locations of a human face and its two eyes are extracted by means of a modified fractal dimension, which is a measure of the surface roughness. The input facial image is then normalized and aligned to those facial images in a database based on the inter-distance of their two eyes. This normalized facial image is compared to each facial image in the database. Our objective is to devise an automatic human face recognition technique that is computationally simple and can provide a reasonable performance level.

1.4 Organization of the Thesis

The rest of this thesis will give an overview of existing techniques for face detection, facial feature extraction and human face recognition, as well as the respective techniques devised and developed in this thesis. Chapter 2 will present the state-of-the-art technology for human face detection, facial feature extraction, and human face recognition, as well as the problems associated with this technology. Chapter 3 outlines our efficient approach for human face detection and facial feature extraction.

The techniques used include valley field detection and fractal dimensions. A modified fractal dimension is proposed, which can measure the roughness of an image region, provide information about orientation, and is insensitive to lighting conditions. In this chapter, we locate the position of the two eyes, and thus the corresponding face region. In Chapter 4, an effective method for the extraction of a chin contour will be described. The deformable template and active contour model can extract the shape of facial features; however, they cannot extract a chin contour accurately and reliably. This is due to the fact that the contrast of the chin boundary is usually very weak, while the edge forces and the valley forces from the mouth and below the chin might be strong. In our approach, we apply a new template in the representation and use the information about the position of the mouth in determining the chin contour. In Chapter 5, our proposed efficient approach for human face recognition based on a modified Hausdorff distance will be presented. In comparing our proposed modified Hausdorff distance with other Hausdorff distance measures, experimental results show that our proposed method outperforms other Hausdorff distance measures. An automatic human face recognition system is described in Chapter 6. This system is actually the combination of our proposed techniques for face detection, facial feature extraction, and human face recognition. Finally, a summary of the work completed for this thesis and a conclusion will be provided in Chapter 7.

Chapter 2
Overview of Face Detection, Facial Feature Extraction and Face Recognition

2.1 Introduction

Feature and pattern detection is a classical computer vision problem with many potential applications, ranging from automatic target recognition to industrial tasks. One of the tasks in pattern detection is called facial feature detection. Face detection and facial feature extraction are important since these are usually the first steps in an automatic human face recognition system [4][12] and a low bit-rate video-conferencing system [28][29]. Human face recognition may seem to be an easy task for humans, but computerized face recognition systems still cannot achieve a completely reliable performance due to large possible variations in facial appearance, lighting conditions, head size and orientation. These are fundamental problems for human face recognition. For automatic human face recognition, the performance in terms of recognition rate is heavily dependent on the accuracy of face detection and facial feature extraction. In other words, a reliable method for detecting face regions and extracting facial features is requisite to such applications. Various approaches to face detection, facial feature extraction and face recognition will be presented in this chapter.

2.2 Face Detection Method

2.2.1 The Face Detection Problem

Face detection is difficult because face patterns can have significantly variable image

appearances; such as differences in facial expression, skin color, size, position and rotation. Furthermore, the human face is essentially a 3D object, different directions of a light source can create different shadows on the face and perspective variations can produce different facial images after being projected a face onto the 2D space. As a result, the true face in an image is very difficult to detect. In addition, face detection is a time-consuming process due to lack of information about the number, location, size and orientation of faces in an image or video scene.

2.2.2 Hierarchical Knowledge-Based Approach

The block diagram of a face detection system adopting the hierarchical knowledge-based approach [30] is shown in Fig. 2.1. The whole system consists of three levels. The first two levels use the mosaic images of different resolutions to detect and verify possible faces according to some pre-specified rules. The third level extracts edges of the facial components and applies the domain knowledge in further verification. A true face will be declared if the input region fulfils all the rules at each of the levels. However, this method is unable to detect a rotated human face in an image.

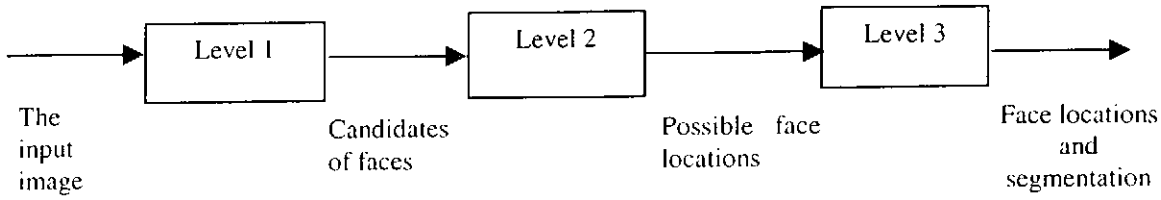


Figure 2.1: General block diagram of the hierarchical knowledge-based approach.

Miao *et al.* [31] presented another hierarchical technique based on the gravity-center template in still gray-level mosaicized images. This approach can save much time consumed in the rough detection of human faces when compared to the

traditional search method [30]. The procedure for this face detection method can be divided into four stages, as shown in Fig. 2.2.

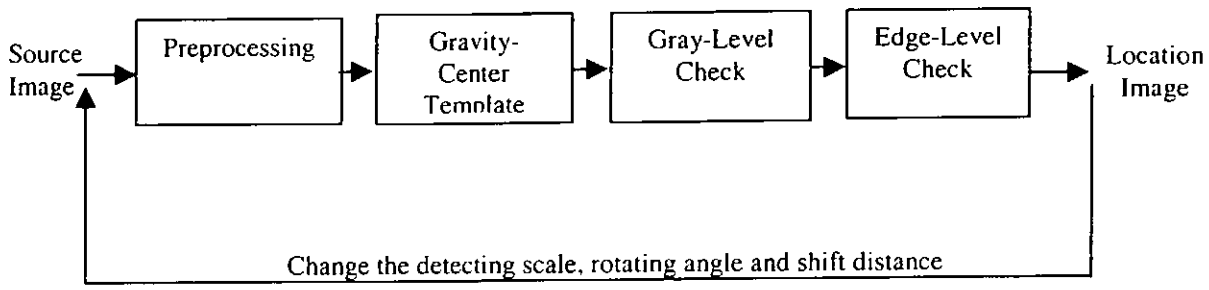
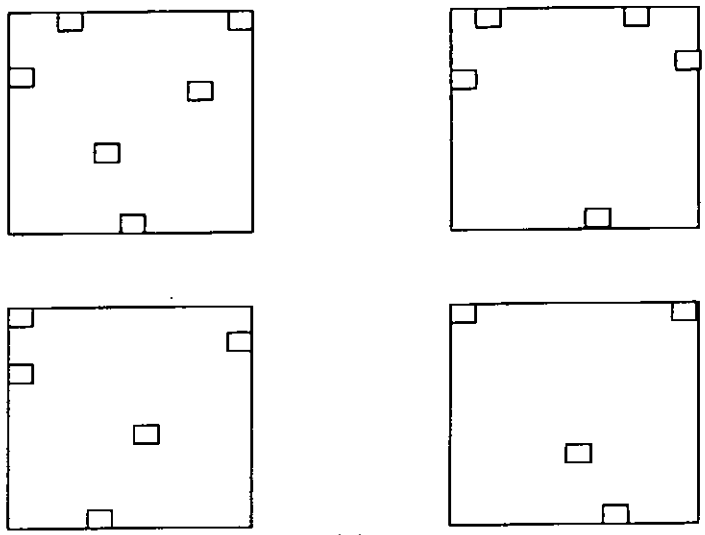
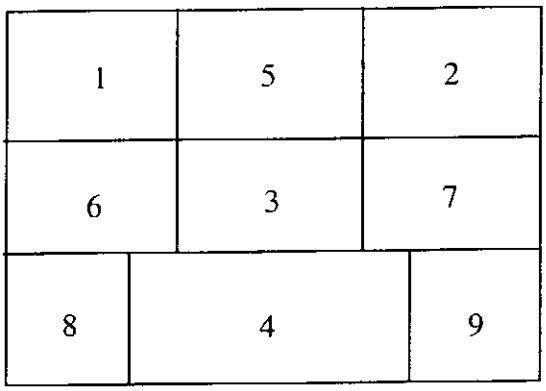


Figure 2.2: System architecture of the hierarchical-approach based on the gravity-center template.

The preprocessing stage is composed of four modules: rotating image, mosaicizing image, extracting mosaic edge and calculating gravity-center. In levels 2 and 3, the possible face candidates are verified by using the patterns and a set of rules, as shown in Fig. 2.3 (a) and (b). These patterns are defined according to the structure of a real human face. The detection rate based on this approach is 86.7%.



(a)



(b)

Figure 2.3: (a) Four gravity-center templates (point-patterns), and (b) a sub-areas of the used in the last stage.

2.2.3 Feature Based Approach

A feature based approach for the detection of human faces in a complex background was presented in [32][42]. The eyes are the most salient features in a human face. Due to the geometric properties and the low gray-level intensity of the iris, a valley exists at an eye region. The valley field is extracted by means of morphological operators [33]. Possible eye candidates are selected based on the valley field and their local properties extracted by using Gabor filter [20]. The possible eye candidates are paired to form possible eye pairs and the corresponding possible face candidates. Each

of the possible face candidates is normalized to a specific low resolution and is then compared to a face template. The overall hit rate based on this approach was reported to be 93.75% without head tilt and under head-on lighting. A similar method was presented in [36], which uses the Genetic Algorithm [34] and the Eigenface technique [35] to search possible face regions in an image. In order to perform the search efficiently using the genetic algorithm, the searching space for possible face regions is limited to possible eye regions. In addition, the lighting effect and orientation of the faces are considered and solved by means of a histogram normalization technique. Experiment results show that the method can achieve an overall hit rate of 90% without head tilt and under head-on lighting. The example-based learning approach [37] uses a number of "face" and "non-face" examples to train up a system, which can reliably locate a vertical and frontal human face in complex scenes. Six "face" clusters and six "non-face" clusters are obtained based on 4,150 normalized face patterns. Human faces are detected by matching the window patterns at different image locations and scales to the distribution-based face and non-face models. Even though this approach can achieve a high detection rate with a complex background, it is very computationally intensive.

2.2.4 Shape Information Approach

Shape information has been used for human face detection [38-40]. An ellipse is used as the shape representing human faces [38-39]. A triangle-based approach was also presented [40] for face detection, where potential face regions are identified according to a number of triangle criteria. Both these two shape-based methods can efficiently detect human faces in an image with a simple background. The block diagram for the

face detection method [38] using ellipse can be divided into four stages, as shown in Fig. 2.4.

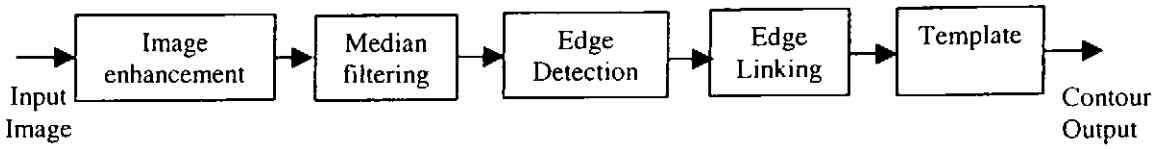


Figure 2.4: Basic flowchart of the algorithm.

This method considers that the overall shape of a face or a head is similar to an ellipse. An input image is first enhanced by means of a histogram equalization, which is followed by edge detection based on a multiple-scale filter. The extracted edges are then linked based on the minimization of an energy function. The face contour is finally extracted by using the direction information of the linked edges. This method was reported to be able to locate human faces with an accuracy of 87.5%. Due to the fact that the shape based method to locate the human face is computational intensive, the Genetic Algorithm (GA) [34] is employed to locate the position of human facial regions precisely without imposing any constraints on size and positions in [39].

2.2.5 Color Analysis Approach

The use of color information [44-46] has also been studied for face detection. Some techniques [47-49] based on the observation that skin color distributes over a very small area on the chrominance plane even if it differs from person to person and race to race. The Segmentation methods of the human face based on skin color are presented in [41,43]. In [41], the RGB model is combined with the HSV model to characterize the human skin. Each homogeneous skin colour region is segmented and a face-like pixel searcher is defined where it can sense its local position and evaluate the colour value of a pixel, mark the face-like point, and exhibit a number of

evolutionary behaviours. A method based on the face-texture model was reported in [43]. With this model, the *I* component of the *YIQ* color system is utilized to enhance regions close to orange, which will include the face areas. In [44], two natural visual cues were presented, the motion and the color, to locate face regions. The face motion is detected by means of the velocity vector field which is used for face extraction. According to the hue space and knowledge based process, the eye, eyebrow, and mouth regions of a face can be detected.

2.3 Facial Feature Extraction

Facial feature extraction is a complex issue for research although it might not be difficult for people to perceive human faces and facial features from an image. Research on facial feature extraction has been presented in [52-66]. In [52-55], the techniques used define a head model for extracting the facial features after the human faces have been detected. Different approaches to extracting the position of the facial features have been proposed; however, they can only roughly estimate the position. One approach is to compute the horizontal and vertical projections of an image [52][55][56][59][65] to determine the position of the eye and the mouth. As the eyes, nose and mouth regions appear darker than other regions in a face, the summation of the grey-level intensities for these regions will exhibit as a local minimum. Another approach [51][58][66] for detecting facial feature points is by means of the Gabor filter [20], which can extract local properties of the facial features. In [58], the valley features and corners of the eyes are used to identify possible eye candidates. Two eye candidates will be paired if their respective local properties are similar. Each eye candidate is further verified by comparing it to an eye template. However, the use of a Gabor filter to extract features is computationally intensive. Hence, an efficient

approach for automatic facial feature extraction was proposed in [50], which uses the Genetic Algorithm and the Region Growing method. The block diagram of this automatic feature extraction algorithm is shown in Fig. 2.5.

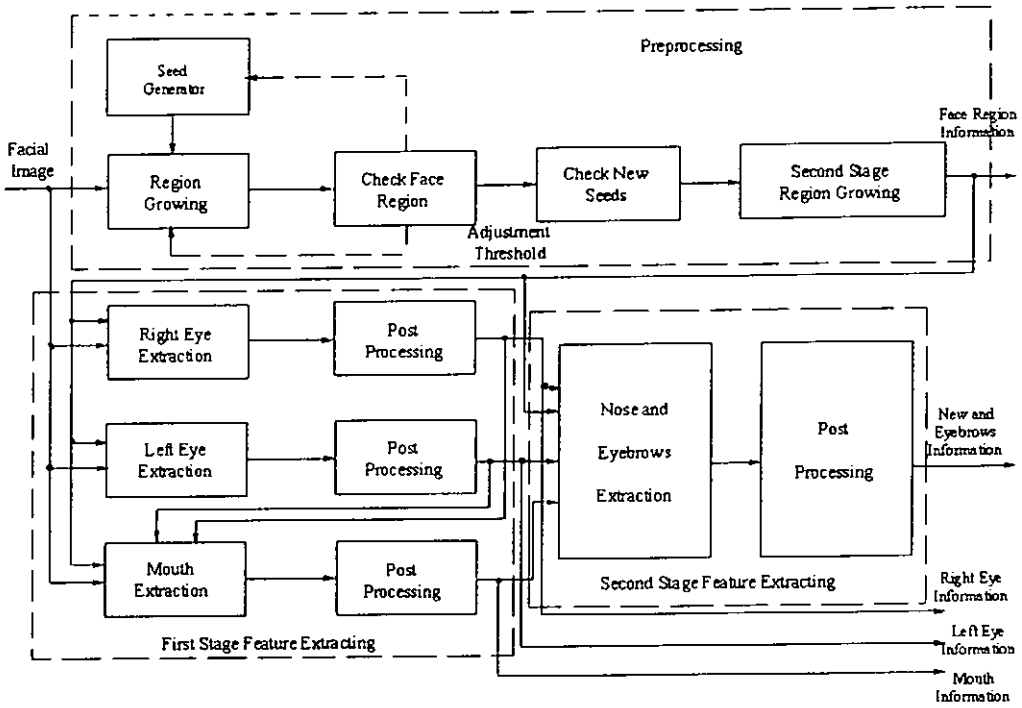


Figure 2.5: Block diagram of the automatic feature extraction algorithm [50].

The algorithm is composed of two main stages: face region estimation and feature extraction. In the face region estimation stage, a face is located roughly at the center of a head-and-shoulders image. The face usually exhibits higher contrast than its surroundings, so it can be extracted by means of a region growing and segmentation method. In the feature point extraction stage, the desired facial features will be extracted and searched within the face region by template matching and the genetic algorithm, respectively.

2.3.1 Deformable Template and Active Contour Model

Facial features can also be extracted by means of deformable templates [60-65, 67] and the active contour model (Snake) [68-71]. These two methods have been widely used to extract the shapes of the eyes and mouth, and to represent head boundaries [69], respectively. Yuille *et al.* [67] proposed deformable templates which are specified by a set of parameters. This technique allows *a priori* knowledge about the expected shape of features due to the fact that the general shapes of the human eye and mouth are more or less fixed. This knowledge of the shape guides the detection process. An eye template and a mouth template [67] are shown in Fig. 2.6(a) and 2.6(b), respectively.

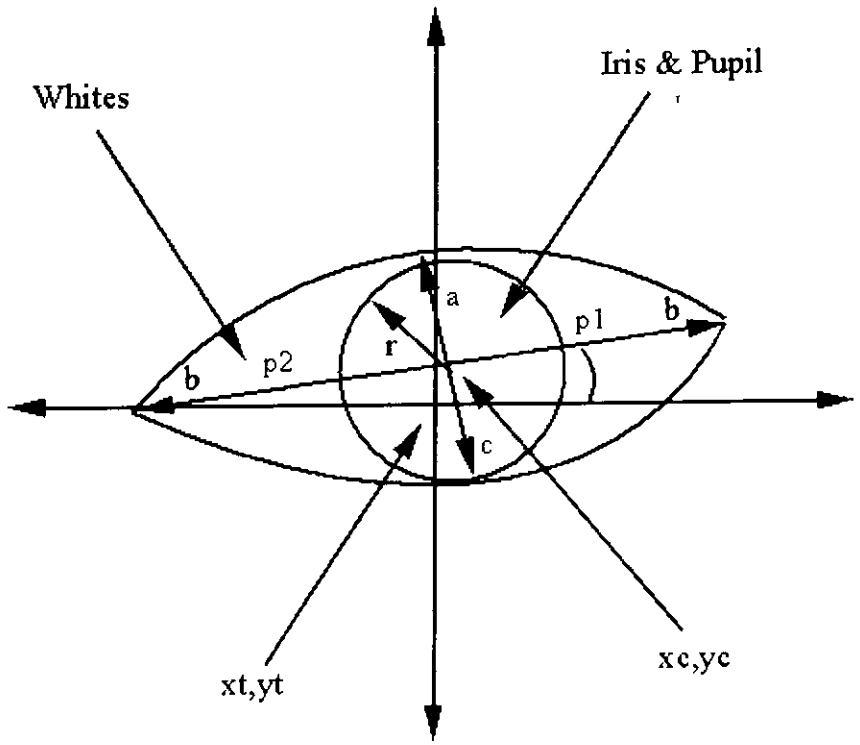


Figure 2.6: (a) A deformable template for a human eye [67].

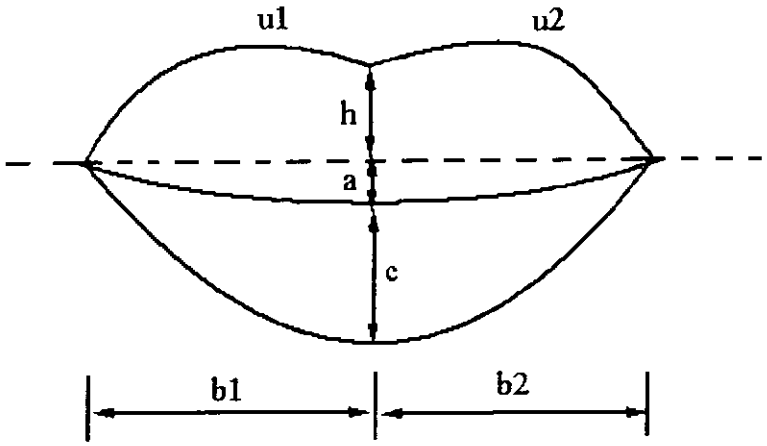


Figure 2.6: (b) A deformable template for a mouth [67].

The eye template has totally eleven parameters represented by ($\vec{x}_c$, $\vec{x}_e$, p_1, p_2, r, a, b, c, θ). This template is modeled by two parabolic curves representing the upper and lower parts of the boundary. It has a center $\vec{x}_e$, with a width of $2b$, maximum heights of a and c for the upper and lower boundaries, and an angle of orientation θ. r is the radius of the iris represented by a circle centered at $\vec{x}_c$. For the mouth template, its center is at point $\vec{x}_m$ and its orientation is θ. The widths of the left and right parts of the mouth template are b_1 and b_2 from $\vec{x}_m$, respectively. The lower two parabolas have maximum distances of a and $a+c$ from the central line. The intersection of the upper two parabolas, u_1 and u_2 occurs at a height of h above $\vec{x}_m$. These templates act on three representations of an image, which are the peak, valley, and edge, as well as on the image itself. An energy function is devised based on these four representations and used to guide the deformation of these templates. The final size, shape and orientation of the eye and mouth templates are obtained by determining the local minimum of the respective energy functions. It is a time consuming procedure to

determine all the parameters through the optimization process. Furthermore, in order to extract the eye, the template must be started at or below the eye. If it is started above the eye, the valley force from the eyebrows may cause problems. Hence, [60][61][65] and [62][63] introduced a reliable method to locate the eye and the mouth, respectively. In [60][61], the corners of an eye are first located by means of a corner detection scheme. Based on the corner position, the shape of the eye can be estimated accurately. The exact shape of the eye is then extracted by a new scheme which is similar to snake [68]. In [62][63], a mouth boundary curve is initially formed by three control points. The exact locations of these control points are then determined through an optimization process by using a set of cost functions. Variance projection function [65] is employed to locate landmarks of the human eye which are then used to guide the detection of the eye position and shape. Another approach to extracting the contour of facial features is the active contour model (Snake) which can stick to edges accurately. The snake model is an energy-minimizing spline that can be operated under the influence of internal contour forces, image forces and external force. A snake is represented as a parametric curve v(s) = [x(s), y(s)], where the arc length s is a parameter. The energy functional [70] of a snake is given as follows:

$$\begin{aligned} E^*_{snake} &= \int_0^1 E_{\text{snake}}[v(s)]ds \\ &= \int_0^1 E_{\text{internal}}[v(s)] + E_{\text{image}}[v(s)] + E_{\text{constraint}}[v(s)]ds \end{aligned} \tag{2.1}$$

The solution to the energy function can be found using variational calculus, but this method has the problem of numerical stability. The greedy algorithm [70] is a fast iteration method which allows a contour with controlled first- and second-order continuity to converge in an area with high image energy. Another fast approach based on the greedy algorithm was presented in [69][70]. In this approach, two

alternate search patterns are used, and a reduction in execution time of about 30% can be achieved with the same performance as the greedy algorithm.

2.4 Human Face Recognition

Psychophysics, neuroscience and engineers have conducted research on various aspects of human face recognition over the past 30 years. Some literature reviews of human face recognition have been conducted [1][2]. Human face recognition techniques can be categorized into: analytic and holistic approaches. The analytic approach [6-8][10-13][20] considers a set of geometrical features, such as the hairline, eyes, mouth and nose, and the relation between the human facial features in a human face. These geometrical features are obtained by means of deformable templates and snake which were presented in Section 2.3. This approach provides high flexibility in handling non-rigid facial features. However, its success heavily relies on the accuracy of facial feature extraction. The holistic approach [3-6][14-15][17][19] considers the global properties of a face pattern which is treated as a two dimensional pattern of intensity variation. However, the recognition performance will degrade if the face to be recognized is not aligned properly. Most of the techniques consider a human face as a two dimensional image (2D). A surface based (3D) approach for face recognition has also been proposed. This approach can eliminate the problems of these 2-D approaches such as illumination and perspective variation. In this section, applications of human face recognition will be given and a review of the analytic-and holistic-approaches will be presented.

2.4.1 Applications of Human Face Recognition

Human face recognition is a research topic that still remains challenging after having been studied for more than 20 years. The difficulty of handling variations resulting from facial expression [18][26][27], age, lighting condition, as well as the wide range of contexts in which the technology is employed, are the main reasons that the field still presents so many challenges. Commercial and law enforcement applications of face recognition technology range from static, controlled format photographs to uncontrolled video images. Examples of law enforcement application include the credit card, passport and personal identification. These applications usually examine the head-and-shoulders images which are usually under well controlled illumination. On the other hand, commercial application of face recognition includes bank/store security, which require this process to be performed in real-time. Applications for commercial enforcement become more complicated than those for law enforcement due to the fact that the input facial image is usually capture under a uncontrolled environment, which can greatly affect the accuracy of face recognition

2.4.2 Existing Human Face Recognition Method

2.4.2.1 Earlier Approaches

Some of the earlier approaches to human face recognition were presented in [21-24], which are based on the feature vectors extracted from the face profile. [21-23] described an automatic human face identification system based on profile analysis which reduces a face profile to an outline curve and then identifies the fiducial mark points. These points are significant in the characterization of the outline curve and a

feature vector is then formed for each face profile. The classification of a human face is based on a weighted distance measure.

2.4.2.2 Analytic Approaches

The analytic approach [6-8][10-13][20] uses the position of facial features and the outline of a face in the recognition process. These important facial features can be extracted by means of deformable templates and the active contour model. The face recognition systems in [11][12] are examples based on facial feature and face contour information. Chen *et al.* [11] used this approach for face recognition with a database of 12 faces. However, this work does not consider the problem of perspective variations. Lam *et al.* [12] proposed a new analytic method which consists of two stages. The first stage is to select similar faces from a database based on the position of 15 feature points in a human face. The position of these 15 feature points is illustrated in Fig 2.7.

Figure 2.7: The 15 feature points identified on the face.

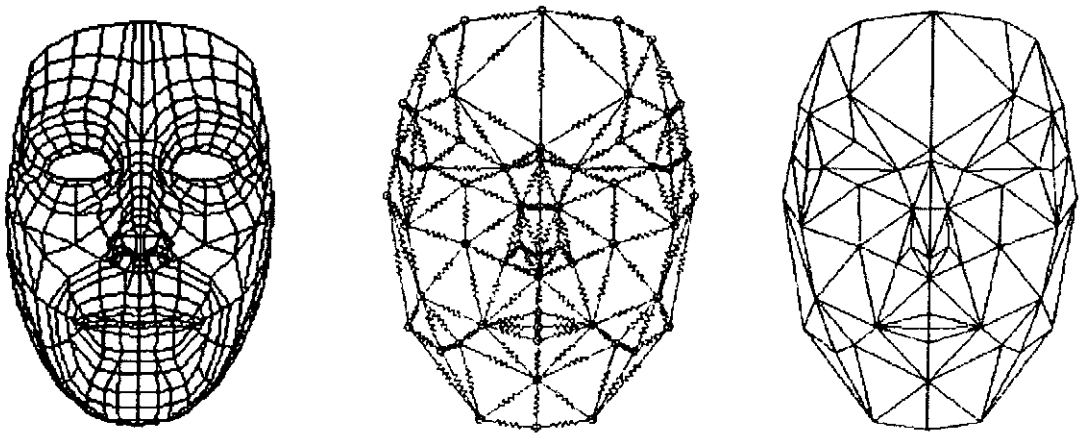
In order to compensate for the effect of perspective variation, the feature points are adjusted based on a simple head model. This model can be used to estimate the rotation of a face by geometrical measurements. The second stage of the system compares the eye windows and the mouth windows of the query input image and that of the images in the face database. This approach can achieve a high recognition rate under different perspective variations.

2.4.2.3 Holistic Approaches

Holistic approaches [3-6][14][15][17-19] consider the global features of a face. The problem formulation of this approach may be simple because it does not need to analyze the respective facial features. In order to maximize the correlation of the global features between different faces, the images are usually normalized and aligned to each other. The position of the two eyes are detected and used for alignment. One of the most well-known methods for this approach is based on the principal component analysis [1][3]. [3] proposed an eigenface technique which projects face images onto a feature space that spans the significant variation among the face images. Kondo *et al.* [4] described a system that detects and recognizes a human face in an image under non-uniform illumination. The experimental results showed that the scheme applied to face detection and face recognition has achieved a fairly robust performance under non-uniform lighting conditions. Another holistic approach for face recognition based on Hausdorff distance [79] has been proposed.

This method operates on edge maps and derives holistic similarity measures using a modified Hausdorff distance, namely "doubly modified Hausdorff distance", which introduces the notion of local neighborhood and associated penalties to a matching algorithm. This new Hausdorff distance can allow for large tolerance to local, non-

rigid distortions without point-to- point correspondence problem. The recognition rate of this method can attain an accuracy rate of about 90%. However, holistic approaches are usually unable to solve the problem of varying poses. An idea of transforming a face with an unknown pose into a frontal view in advance was proposed in [19]. Owing to the fact that the orientation and the main landmarks of the facial model roughly matched the original face image, the modified 3D facial model for view synthesis is adopted, as illustrated in Fig 2.8. A spectroface method [25] was then applied for recognition of virtual frontal-view images. The recognition accuracy of this method is 84.7%.



(a) (b) (c)

Figure 2.8: (a) 3-D facial model; (b) 3-D SB model used in [19]; (c) simplified 3-D SB model.

2.4.2.4 Other Approaches

One of the difficulties in designing a reliable method for face recognition is to recognize human faces appearing with different variations such as facial expressions, uneven lighting conditions, perspective variations, etc. One of the approaches to

handling these variations is the Dynamic Link Architecture (DLA) [20]. This architecture can recognize an object by using sparse graphs, where vertices are labeled by a multi-resolution description in terms of a local power spectrum, and whose edges are labeled by geometrical distance vectors. Object recognition can be formulated as elastic graph matching, which is performed by stochastic optimization of a matching cost function. These local features are extracted by using Gabor wavelet [20]. The graph matching process consists of two stages. The position of a face is located in a face image using a non-distorted grid in the first stage. The grid is positioned over the face and its structure is then allowed to deform until each node has achieved a minimum of the cost function. The total cost of matching is a weight of the cost of each node in the grid and the amount of the grid deformation, as shown in Eqn 2.2 [69].

$$\begin{aligned} C_{total}(\{x_i^I\}) &:= \lambda C_e + C_v \\ &= \lambda \sum_{(i,j)\in E} S_e(\vec{\Delta}_{ij}^I \cdot \vec{\Delta}_{ij}^M) - \sum_{i\in V} S_v(J^I(x_i^I), J_i^M) \end{aligned} \tag{2.2}$$

where $\vec{\Delta}_{ij}$ is the Euclidean distance vector of labeled edges between vertices $\vec{x}_i$ and $\vec{x}_j$, $S_e(\vec{\Delta}_{ij}^I \cdot \vec{\Delta}_{ij}^I)$ is the comparison function between the edge labels in the image graph to the corresponding ones in the model graph, and J^I and J^M represent a jet of the image graph and model graph, respectively. Each jet stores the feature vectors which are extracted by means of Gabor wavelet. $S_v(J^I, J_i^M)$ is a function used to measure the similarity of pairs of vertex label between image graph and model graph. Elastic matching of a model graph M to a variable graph I in the image domain amounts to a search for a set $\{x_i^I\}$ of vertex positions which optimizes the matching of vertex labels and of edge labels. The coefficient λ controls the rigidity of the image graph, large values penalizing distortion of the graph I with respect to the graph M. In

[7], a system based on this technique was proposed, which can handle larger galleries and have a higher matching accuracy. This method uses the phase of the complex Gabor wavelet coefficients to estimate the position of the nodes accurately and employs object adapted graphs for referring the nodes to the specific facial landmarks. The recognition rate was reported to be over 90% for frontal faces. However, the extraction of the feature vectors using the Gabor filter is very time-consuming. A dynamic link architecture based on multiscale morphological dilation-erosion was therefore proposed [9][16] for frontal face authentication. Morphological operations of different orientations and scales are employed to yield a feature vector at each node. The multiscale dilation-erosion of the image $f(x)$ by $g_\sigma(x)$ is defined as follows:

$$(f * g_\sigma)(x) = \begin{cases} (f \oplus g_\sigma)(x), & \text{if } \sigma > 0 \\ f(x), & \text{if } \sigma = 0 \\ (f \ominus g_{|\sigma|}(x), & \text{if } \sigma < 0 \end{cases} \qquad (2.3)$$

Let R and Z denote the set of real and integer numbers, respectively. Given an image $f(x)$: $D \subseteq Z^2 \rightarrow R$ and a structuring function $g(x)$: $G \subseteq Z^2 \rightarrow R$, the dilation of the image $f(x)$ by $g(x)$ is denoted by $(f \oplus g)(x)$. Its complementary operation, the erosion, is denoted by $(f \ominus g)(x)$. σ denotes the scale parameter of the structuring function. The outputs of multiscale dilation-erosion for $\sigma = -\sigma_m, \ldots, \sigma_m$ form the feature vector located at the grid node **x**

$$j(x) = ((f * g_{\sigma m})(x), \ldots, (f * g_1)(x), \ldots, (f * g_{-\sigma m})(x)) \qquad (2.4)$$

The linear projection technique is used to reduce the dimensionality of the feature vectors. The algorithm of feature selection and automatic weighting of the nodes according to their discriminatory power can increase the authentication capability of the method.

2.5 Summary of Review

Face detection is an important step for many applications, such as face recognition, object tracking, image coding, and video coding. However, the position of a face in an image is very difficult to locate. The performance to locate the human faces is affected by several variables, such as the presence of moustache, glasses, gender, skin color, facial expression, facial hair, as well as the background. In order to solve these problems, various approaches to face detection such as the shape information approach, knowledge-based method, example-based learning method and feature based approach have been introduced. In the shape information approach, the shape of a head is assumed to be elliptical. Thus, the size and orientation angle of a face is used to obtain the elliptical shape in an image. For the knowledge-based method, the prior knowledge about human faces is used to define the rules for face detection. The human face region is declared if the input image complies with all the rules. In the example-based learning method, a large number of face and non-face are used to obtain the distributions of face images. The input image is classified into the face clusters or non-face clusters according to the training results.

For facial feature extraction, one of the approaches is to compute the projections of a face region along the horizontal and vertical directions. However, this method can only roughly estimate the position of the facial features. Other approaches for facial feature extraction are deformable template and active contour model, which are based on *a priori* knowledge about the shape of the feature and are operated under the influences of different forces. Both analytic and holistic approaches to human face recognition have been presented. The analytic approach considers a set of geometrical features related to human facial features. On the other hand, the holistic approach considers global properties of the human face.

Chapter 3
Locating the Eye in Human Face Images Using Fractal Dimensions

3.1 Introduction

In Chapter 2, we presented the problem of face detection and facial feature detection. Various approaches to human face detection and facial feature extraction have been presented to solve the problem. In this Chapter, we will present a fast approach based on valley field detection and fractal dimensions (*FD*) [72] to extract eye pairs in a complex background, which can then be used to represent a face region.

Research in fractal analysis and its applications have grown rapidly in recent years. Fractals are highly detailed, complex geometric shapes. One measure of the complexity of an image is fractal dimension, which can be estimated using the box-counting [72-74] technique. Advance research and development in fractal analysis have practical applications in video coding, pattern recognition, natural texture analysis [75] and classification [76], etc. Fractal dimension has also been used with wavelet analysis in feature extraction [77].

The approach to locating eye pairs proposed consists of three levels, as shown in Fig. 3.1.

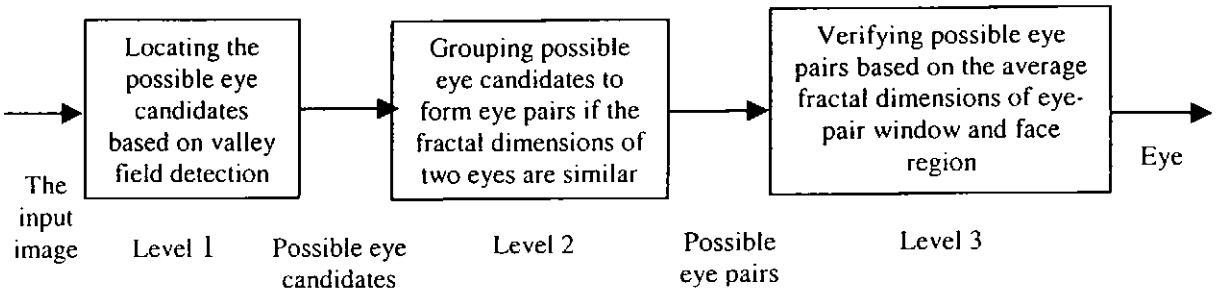


Figure 3.1: The block diagram of the eye extraction system.

The first level is to detect in an image the location of possible eye candidates, which exhibit themselves as valleys in the image space. These valleys are used to define the possible eye candidates. Valley positions that have features similar to the eyes are considered as possible eye candidates, and will be passed to level 2 to form possible eye pairs. Two possible eye candidates with similar fractal dimensions are grouped to form a possible eye pair in level 2. As fractal dimension is sensitive to lighting conditions, a normalization process is performed before estimating fractal dimension in order to reduce the effect of lighting. Furthermore, the two eyes of a valid pair should have similar orientation. An oriented fractal dimension is therefore devised to increase the matching accuracy. In stage three, in order to reduce the effect of uneven lighting conditions in an eye-pair window, its edge image is considered in computing the fractal dimension. Based on the input eye pair, the corresponding face region can be formed. The input eye-pair window and the corresponding face region are then binarized, and their respective fractal dimensions are measured. These two fractal dimensions are then compared with the average fractal dimensions of the eye-pair window and face region. The performance of this approach has been evaluated using the ORL database, MIT database, and various complex images. Experimental

results have shown that the eye pairs can be detected reliably under different orientations, lighting conditions, and slight perspective variations.

This chapter is organized as follows. Section 3.2 describes the box-counting method for estimating the fractal dimension. The details of the eye detection procedure based on valley field detection and measurement of fractal dimension are presented in Section 3.3. Experimental results will be given in Section 3.4. Finally, a conclusion will be drawn in Section 3.5.

3.2 Estimation of Fractal Dimension

In order to estimate the fractal dimension, a number of methods such as the Fourier power spectrum [75] and box-counting [73][74] can be used. The box-counting method is an efficient technique for estimating fractal dimensions, and it can be applied to gray-level images and binary images, as described below.

The differential box-counting approach [73] has been used to estimate fractal dimensions. This method counts the number of boxes that cover the image intensity surface. The surface can be considered as a 3D space in which the two coordinates (x,y) represent the 2D position and the third coordinate (z) represents the image gray-level intensity. For a given image of size $I \times I$, the image is partitioned into grids of size $S \times S$. The grids are numbered as (i,j), where $0 \leq i,j < r$ and $r = \lfloor I/S \rfloor$. Each grid is stacked with a column of boxes of size $S \times S \times S'$. Suppose that the maximum gray-level intensity is G, then $\lfloor G/S' \rfloor = \lfloor I/S \rfloor$. The boxes on a grid are assigned a number with the box at the bottom as box one and the one on the top as $\lfloor G/S' \rfloor$. If the minimum and maximum gray-levels of the image in the $(i,j)^{th}$ grid fall in the boxes numbered k and l, respectively, then $n_r(i,j) = l - k + 1$, where $r = \lfloor I/S \rfloor$ and $0 \leq i,j < \lfloor I/S \rfloor$. $n_r(i,j)$

represent the number of boxes covering the image intensity surface over the $(i,j)^{th}$ grid, as shown in Fig. 3.2. The total number of boxes required to cover the surface is $N_r = \sum_{i,j} n_r(i,j)$. With different grid size S, different values of r and N_r can be obtained. The fractal dimension can then be estimated from the least-square linear fit of $\log(N_r)$ against $\log(r)$. In Fig. 3.3, four points of $(\log(N_r),\log(r))$ for estimating the fractal dimension at the eye center are plotted and joined by solid lines. The corresponding dotted line represents the least-square linear fit of the four points. The fractal dimension is computed to be 2.115, and the mean square error between the dotted line and the solid line is 0.0029.

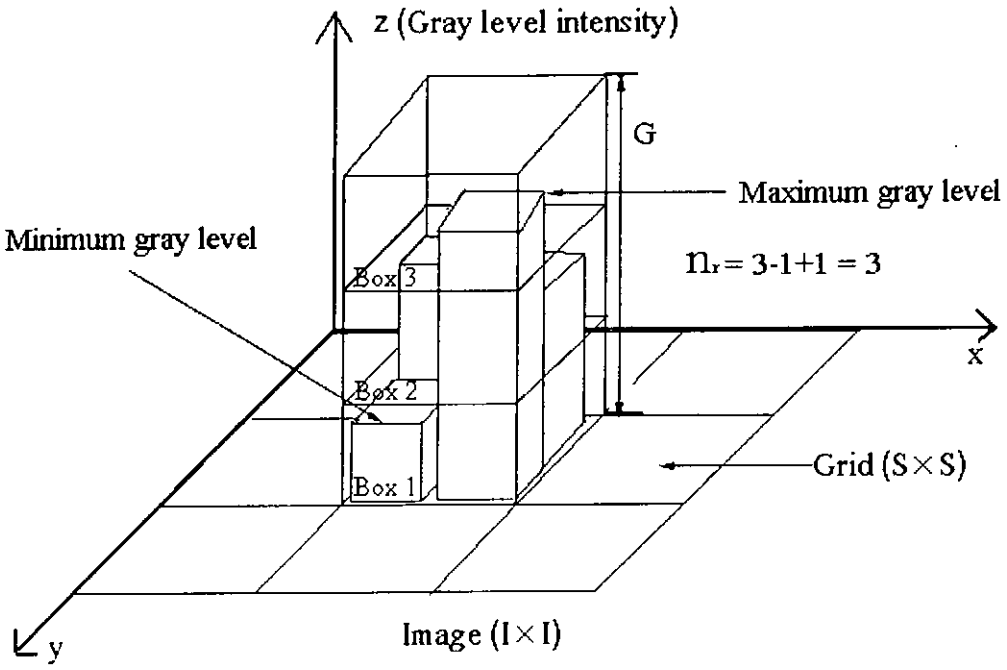


Figure 3.2: Estimation of fractal dimension for a gray-scale image using the box-counting technique.

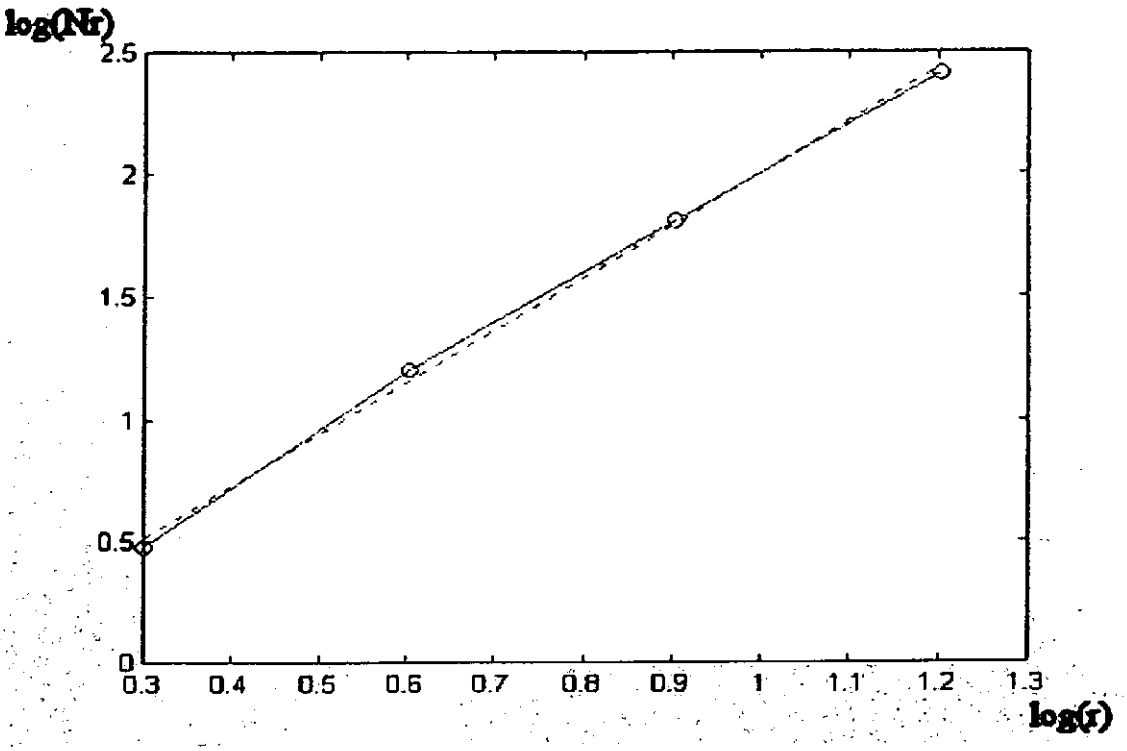
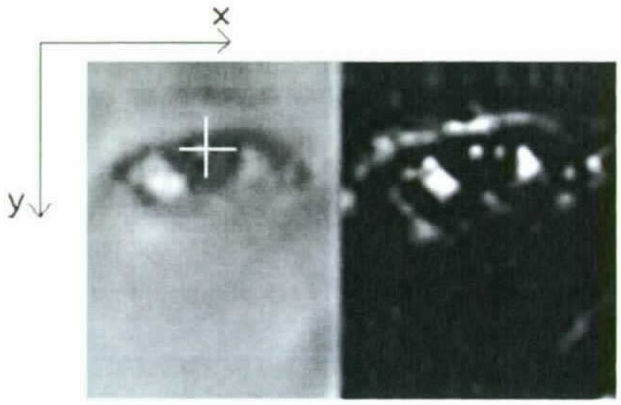


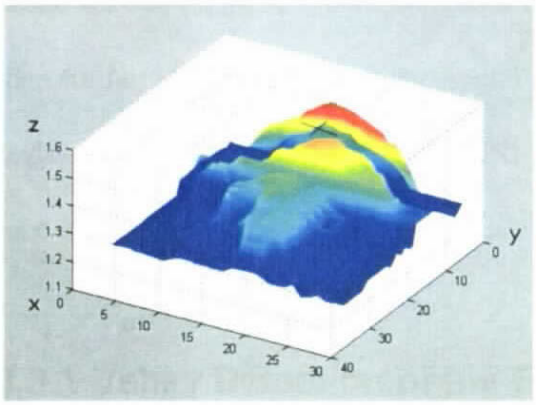
Figure 3.3: The curve of $\log(N_r)$ against log(r) and the corresponding least-square linear fit for an eye region.

For binary images, only two levels exist: black and white. A black pixel represents an image point or an edge point, while a white pixel is regarded as a point in the background. An image of size $I \times I$ is divided into grids of size $S \times S$. At each grid, the number of boxes, $n_r(i,j)$, that contain any black pixels is counted. The total number of boxes over all the grids is denoted as $N_r = \Sigma n_r(i,j)$, where $i,j, = 1,\ldots..,\ I/S$ and $r = \lfloor I/S \rfloor$. The grid size $S \times S$ is changed such that different r and the corresponding total box count N_r are obtained. Again, the fractal dimension can then be estimated by using the least-square linear fit of $\log(\Sigma n_r)$ against $\log(r)$.

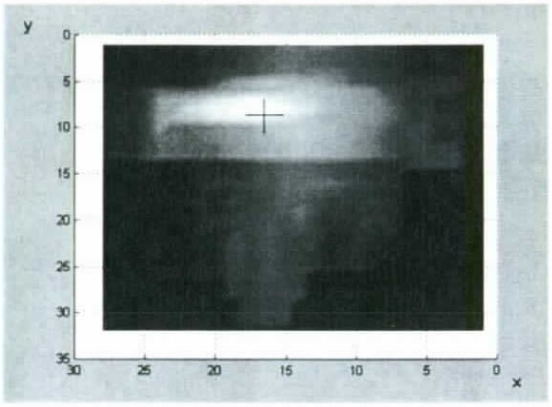
Figure 3.4 shows the distribution of fractal dimension for an eye region and its surrounding area. The fractal dimension at the center of the eye region is higher than that of its surrounding area due to the fact that the variations of the gray level intensities or the edges around the eye center are great. Fig. 3.4(d) and (e) illustrate that the magnitude of the fractal dimension is a maximum near the eye center.



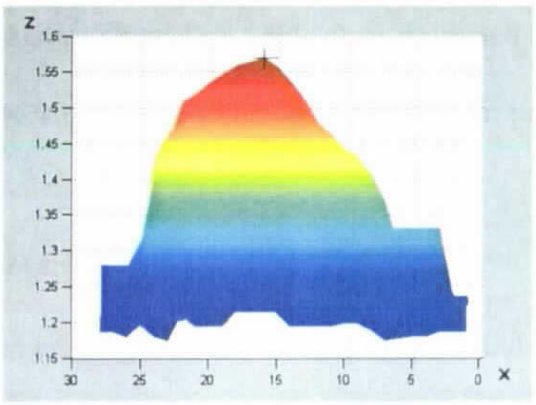
(a)



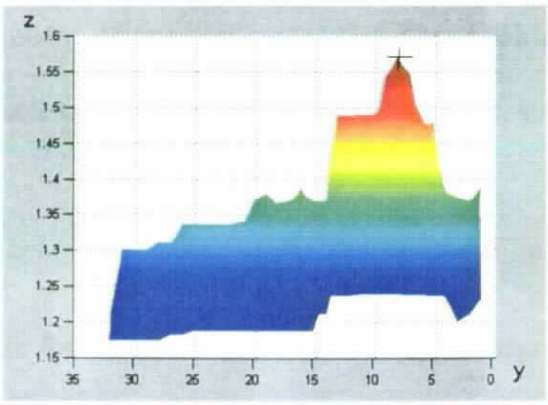
(b)



(c)



(d)



(e)

Figure 3.4: The distribution of fractal dimension of an eye region: (a) the eye region and its eye image, (b) 3D representation of fractal dimension, where the z-axis is the magnitude of fractal dimension, (c) projection of the fractal dimension onto the x-y plane, where the fractal dimension is represented by gray-level intensities, (d) projection of the fractal dimension onto the x-z plane, and (e) projection of the fractal dimension onto the y-z plane.

3.3 Detecting the Eye Pairs

The detection of eye features is an important step in human face recognition because the eyes are the most salient feature on a human face. In this section, we propose a new method for locating the eye pairs in an image with a complex background. The detection is divided into three stages. First, all the valley regions in an image are detected and tested for possible eye candidates, which are selected based on a number of criteria. In the second stage, possible eye pairs are formed from the eye candidates based on eye features and the oriented fractal dimension. The possible eye pairs are then further verified based on the fractal dimensions of the eye-pair windows and face regions in the third stage. The detailed procedure for locating the eye pairs and the face regions is described below.

3.3.1 Valley Detection of the Eyes

As the iris is of low gray-level intensity in a human face, a valley exists at an eye region. The valley field, Φ_v, is extracted by means of morphological operators [33]. A possible eye candidate is identified at position (x,y) if the following two criteria are satisfied:

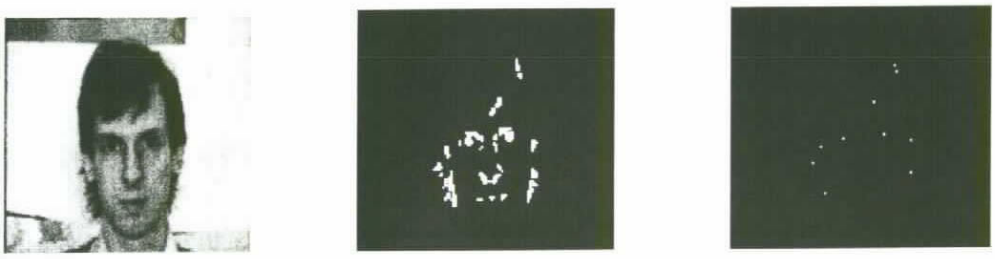
$$f(x,y) < t_l \text{ and } \Phi_v(x,y) > t_v \tag{3.1}$$

where $f(x,y)$ is a facial image, and t_l and t_v are thresholds. Fig. 3.5(a) and 3.5(b) show an image and its corresponding possible eye candidates, respectively. A number of regions of possible eye candidates are detected, and are then reduced to a point by choosing the best candidates in each of the regions. Two functions, $v_1(x,y)$ and $v_2(x,y)$, are used to locate the best eye candidate in each region. The two functions are defined as follows:

$$v_1(x,y) = C_{1,1}\Phi_{1,1}(x,y) + C_{1,2}\left(\frac{f(x-2,y)+f(x+2,y)}{2} - S_{1,1}(x,y)\right)$$

$$v_2(x,y) = C_{2,1}\Phi_{2,1}(x,y) + C_{2,2}\left(\frac{f(x-3,y)+f(x+3,y)}{2} - S_{2,1}(x,y)\right) \qquad (3.2)$$

where C's are weighting factors, $\Phi_{1,1}(x,y)$ and $S_{1,1}(x,y)$ are the average valley intensity and the average gray-level intensity inside a 3×3 window, respectively, while $\Phi_{2,1}(x,y)$ and $S_{2,1}(x,y)$ are the corresponding values inside a 5×5 window. This arrangement allows us to detect the eyes of different scales. The good eye candidates are those having large values in the functions $v_1(x,y)$ and $v_2(x,y)$. Fig. 3.5(c) illustrates the corresponding best eye candidate for the segmented regions in Fig. 3.5(b).



(a) (b) (c)

Figure 3.5: (a) Original image, (b) possible eye regions, and (c) best candidates for the eye regions.

3.3.2 Grouping the Best Eye Candidates to form Possible Eye Pairs

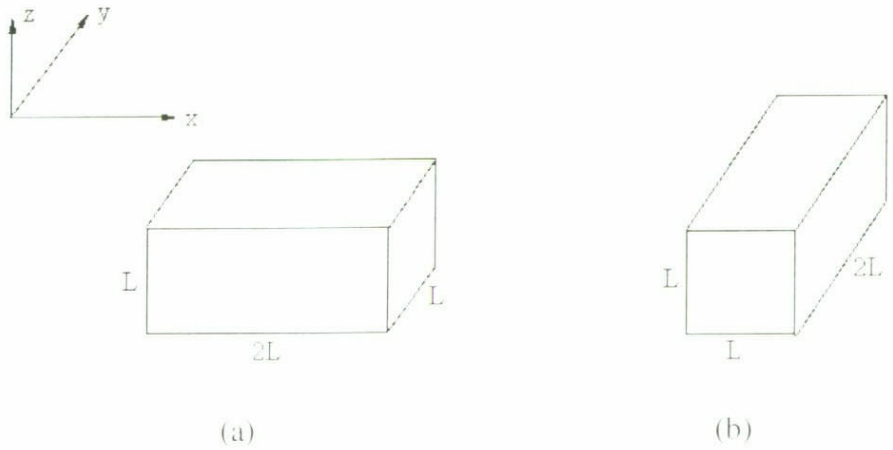
Two eye candidates are paired to form a possible eye pair. We assume that the rotation of the human face is less than 45°. The two eyes of a face should have similar orientations and be of similar size. Moreover, the inter-distance between the two eye candidates should not be less than a threshold, which also represents the size of the human face. If the face size is too small, it will not be feasible for face recognition. Moreover, the roughness of the two eye regions should be close to each other.

Based on the inter-distance between the two eye candidates, we can set up the corresponding windows to cover each of the two eye regions, as shown in Fig. 3.6. The height, h, and width, w, of the windows are equal to a quarter of the inter-distance, L. As fractal dimension is sensitive to lighting conditions, a histogram normalization technique [78] is used. An eye region is equalized in such a way that the histogram of the region is normalized to a specified histogram. This arrangement alleviates the effect of uneven lighting conditions on both halves of the face. When two eye candidates are paired, they should have similar fractal dimensions, as well as orientation. However, due to the use of square boxes in counting, fractal dimension lacks this information about orientation. In our approach, we use two rectangular boxes, as shown in Fig. 3.7, of different orientations instead of a square box in box-counting. The two measured fractal dimensions are called horizontal fractal dimension, FD_h, and vertical fractal dimension, FD_v, respectively. Table 3.1 shows the horizontal and vertical fractal dimensions of the left and right eyes under different orientations for a number of facial images. The results demonstrate that the FD_h of the left and right eyes of a human face are very similar at a given orientation. The same applies to the FD_v. The two fractal dimensions, FD_h and FD_v, for a human face are different from each other, and change according to the orientation of the eyes. For any given image region, the texture inside will remain more or less the same under different rotations, but both the FD_h and FD_v will change. Nevertheless, when the FD_h decreases, the FD_v will increase, and vice versa. This is due to the fact that the total number of boxes required to cover the image area or space should change only slightly even when the image is rotated. As shown in Table 3.1, the sums of FD_h and FD_v for the eye samples under different rotations remain fairly constant. Thus, a valid eye pair can be selected if the following criteria are satisfied.

$$\begin{aligned} &|FD_h(x_0, y_0) - FD_h(x_1, y_1)| < t_1 \text{ and } |FD_v(x_0, y_0) - FD_v(x_1, y_1)| < t_2 \text{ and} \\ &|(FD_h(x_0, y_0) + FD_v(x_0, y_0)) - M_{eye}| < t_3 \text{ and } |(FD_h(x_1, y_1) + FD_v(x_1, y_1)) - M_{eye}| < t_4 \end{aligned} \tag{3.3}$$

where (x_0, y_0) and (x_1, y_1) are the locations of the left- and right-eye candidates, M_{eye} is the average fractal dimension of the eye windows, and t_1, t_2, t_3, and t_4 are thresholds. In our experiments, t_1 and t_2 are equal and set at 0.001, t_3 and t_4 are set at 0.033, respectively. These thresholds are obtained on a set of 35 face samples selected randomly from different databases. For a valid pair, the respective difference should be less than certain thresholds. In Fig. 3.8(a), all the possible eye candidates in the face image are illustrated. These possible eye candidates are then validated by the conventional fractal dimension and the oriented fractal dimension to form possible valid eye pairs, as illustrated in Fig. 3.8(b) and Fig. 3.8(c), respectively. The use of oriented fractal dimension provides a better pairing result, since the conventional fractal dimension forms eight pairs while the oriented fractal dimension results in four pairs only. As the oriented fractal dimension matches eye pairs more accurately, the total amount of computation required in stage three can be reduced.

Figure 3.6: The two eye windows for fractal dimension measurement.



(a) (b)

Figure 3.7: Rectangular boxes used for measuring (a) FD_h, and (b) FD_v.

Sample	Upright					
	FD_h of left eye	FD_v of left eye	FD_h of right eye	FD_v of right eye	$FD_h + FD_v$ of left eye	$FD_h + FD_v$ of right eye
C1	1.1696	1.1450	1.1701	1.1464	2.3146	2.3165
C2	1.1744	1.1516	1.1716	1.1441	2.3260	2.2957
C3	1.1786	1.1524	1.1858	1.1608	2.3310	2.3466
C4	1.1701	1.1406	1.1613	1.1505	2.3107	2.3118
C5	1.1769	1.1572	1.1776	1.1464	2.3341	2.3240
C6	1.1747	1.1439	1.1837	1.1505	2.3186	2.3342
C7	1.1687	1.1533	1.1727	1.1562	2.3220	2.3289
C8	1.1827	1.1525	1.1760	1.1489	2.3352	2.3249
C9	1.1756	1.1479	1.1696	1.1440	2.3235	2.3136
C10	1.1810	1.1564	1.1772	1.1499	2.3374	2.3271

(a)

Sample	15° rotated right					
	FD_h of left eye	FD_v of left eye	FD_h of right eye	FD_v of right eye	$FD_h + FD_v$ of left eye	$FD_h + FD_v$ of right eye
C1	1.1715	1.1550	1.1635	1.1494	2.3265	2.3129
C2	1.1598	1.1428	1.1533	1.1494	2.3026	2.3027
C3	1.1757	1.1359	1.1693	1.1414	2.3116	2.3107
C4	1.1786	1.1432	1.1748	1.1393	2.3218	2.3141
C5	1.1808	1.1381	1.1759	1.1324	2.3189	2.3083
C6	1.1812	1.1323	1.1865	1.1369	2.3135	2.3234
C7	1.1793	1.1438	1.1801	1.1467	2.3231	2.3268
C8	1.1832	1.1502	1.1773	1.1481	2.3334	2.3254
C9	1.1764	1.1502	1.1722	1.1499	2.3266	2.3221
C10	1.1638	1.1559	1.1703	1.1583	2.3197	2.3286

(b)

	15° rotated left					
Sample	FD_h of left eye	FD_v of left eye	FD_h of right eye	FD_v of right eye	$FD_h + FD_v$ of left eye	$FD_h + FD_v$ of right eye
C1	1.1681	1.1435	1.1620	1.1471	2.3116	2.3091
C2	1.1723	1.1391	1.1751	1.1435	2.3114	2.3186
C3	1.1653	1.1490	1.1713	1.1473	2.3143	2.3186
C4	1.1697	1.1432	1.1753	1.1403	2.3129	2.3156
C5	1.1731	1.1381	1.1776	1.1353	2.3112	2.3129
C6	1.1750	1.1424	1.1717	1.1432	2.3174	2.3149
C7	1.1683	1.1480	1.1651	1.1510	2.3163	2.3161
C8	1.1667	1.1477	1.1701	1.1413	2.3144	2.3114
C9	1.1594	1.1501	1.1550	1.1514	2.3095	2.3064
C10	1.1803	1.1511	1.1751	1.1551	2.3314	2.3302

(c)

	30° rotated right					
Sample	FD_h of left eye	FD_v of left eye	FD_h of right eye	FD_v of right eye	$FD_h + FD_v$ of left eye	$FD_h + FD_v$ of right eye
C1	1.1626	1.1450	1.1665	1.1464	2.3076	2.3129
C2	1.1609	1.1516	1.1593	1.1441	2.3125	2.3034
C3	1.1608	1.1524	1.1627	1.1608	2.3132	2.3235
C4	1.1633	1.1406	1.1642	1.1505	2.3039	2.3147
C5	1.1685	1.1572	1.1699	1.1464	2.3257	2.3163
C6	1.1664	1.1439	1.1615	1.1505	2.3103	2.3120
C7	1.1555	1.1533	1.1600	1.1562	2.3088	2.3162
C8	1.1601	1.1525	1.1637	1.1489	2.3126	2.3126
C9	1.1749	1.1479	1.1693	1.1440	2.3228	2.3133
C10	1.1799	1.1564	1.1727	1.1499	2.3363	2.3226

(d)

	30° rotated left					
Sample	FD_h of left eye	FD_v of left eye	FD_h of right eye	FD_v of right eye	$FD_h + FD_v$ of left eye	$FD_h + FD_v$ of right eye
C1	1.1496	1.1326	1.1536	1.1385	2.2822	2.2921
C2	1.1489	1.1369	1.1533	1.1295	2.2858	2.2828
C3	1.1517	1.1290	1.1597	1.1251	2.2807	2.2848
C4	1.1653	1.1402	1.1642	1.1413	2.3055	2.3055
C5	1.1594	1.1305	1.1631	1.1397	2.2899	2.3028
C6	1.1610	1.1454	1.1615	1.1383	2.3064	2.2998
C7	1.1490	1.1325	1.1563	1.1392	2.2815	2.2955
C8	1.1523	1.1297	1.1592	1.1353	2.2820	2.2945
C9	1.1601	1.1292	1.1642	1.1254	2.2893	2.2896
C10	1.1590	1.1354	1.1564	1.1367	2.2944	2.2931

(e)

Table 3.1: The oriented fractal dimensions, FD_h and FD_v, and their corresponding sums, FD_h+FD_v, for a number of eye samples under different orientations: (a) upright, (b) 15° rotated right, (c) 15° rotated left, (d) 30° rotated right, and (e) 30° rotated left.

(a) (b) (c)

Figure 3.8: (a) Possible eye pairs in a facial image, (b) possible eye pairs based on the conventional fractal dimension, and (c) possible eye pairs based on the oriented fractal dimension.

3.4 Verification of Eye Pairs

The eye pairs selected in stage two are passed to the next stage for further verification. At this stage, the measurement of each possible eye-pair region and its corresponding face region will be computed by means of fractal dimension. In order to reduce the effect of lighting conditions, the edge images are used in the computation of the fractal dimension. Examples of eye pairs and human face regions, together with their

corresponding binary images, are shown in Fig. 3.9.

Based on the inter-distance L between two eye candidates, the eye-pair region and its corresponding face region can be extracted. The height, EH, and width, EW, of an eye-pair window are defined to be $0.5L$ and $1.67L$, respectively, as shown in Fig. 3.10(a). Furthermore, the corresponding width of the face region, WF, is also set to $1.67L$. The height of the face region is set in reference to the position of the eye pairs: the top and bottom lines of a face window are $L/3$ and $1.5L$, respectively, from the eye pair, as illustrated in Fig. 3.10(b).

(a) (b)

Figure 3.9: (a) Samples of eye pairs and their corresponding binary images, and (b) samples of the human face regions and their corresponding binary images.

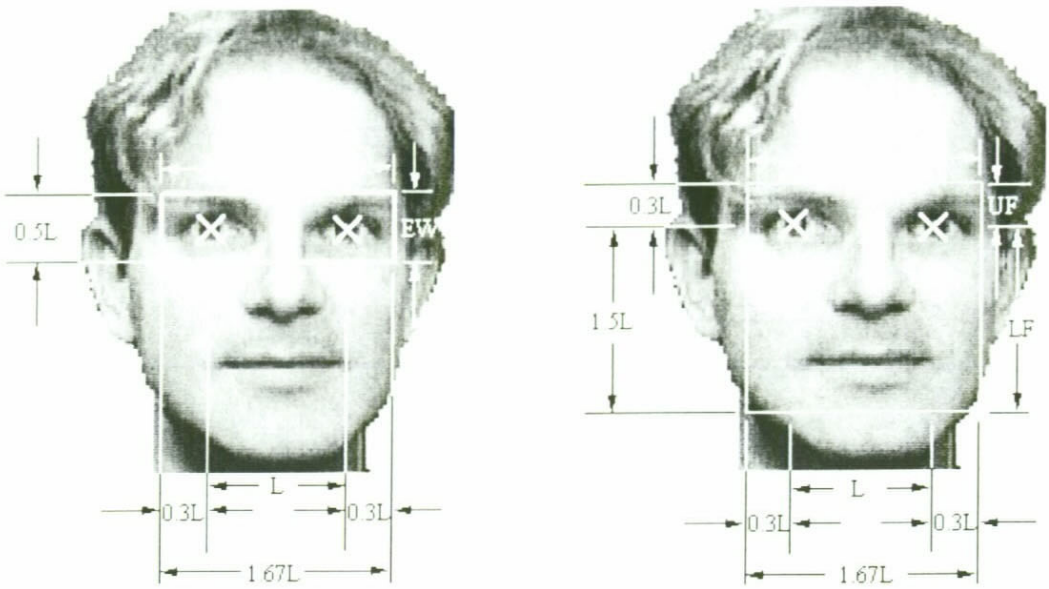


Figure 3.10: (a) An eye-pair candidate and its corresponding eye-pair window, and (b) an eye-pair candidate and its corresponding face region.

In order to verify whether a selected eye-pair candidate is valid, the average fractal dimensions of the eye-pair regions, M_{eye}, and the face regions, M_{face}, are computed. Tables 3.2 and 3.3 tabulate the fractal dimensions of a number of eye-pair windows and face windows. The corresponding means and variances of the fractal dimensions are 1.7341 and 0.0025, respectively, for the eye-pair region and 1.9404 and 0.003, respectively, for the face region.

Sample	Size	*FD* of eye pair
E1	55x20	1.7146
E2	55x20	1.7382
E3	55x20	1.7156
E4	55x20	1.7245
E5	55x20	1.7286
E6	40x15	1.7528
E7	40x15	1.7244
E8	40x15	1.7412
E9	40x15	1.7234
E10	40x15	1.7342
E11	55x20	1.7283

E12	55x20	1.7352
E13	55x20	1.7413
E14	55x20	1.7231
E15	55x20	1.7212
E16	40x15	1.7428
E17	40x15	1.7155
E18	40x15	1.7512
E19	40x15	1.7644
E20	40x15	1.7545

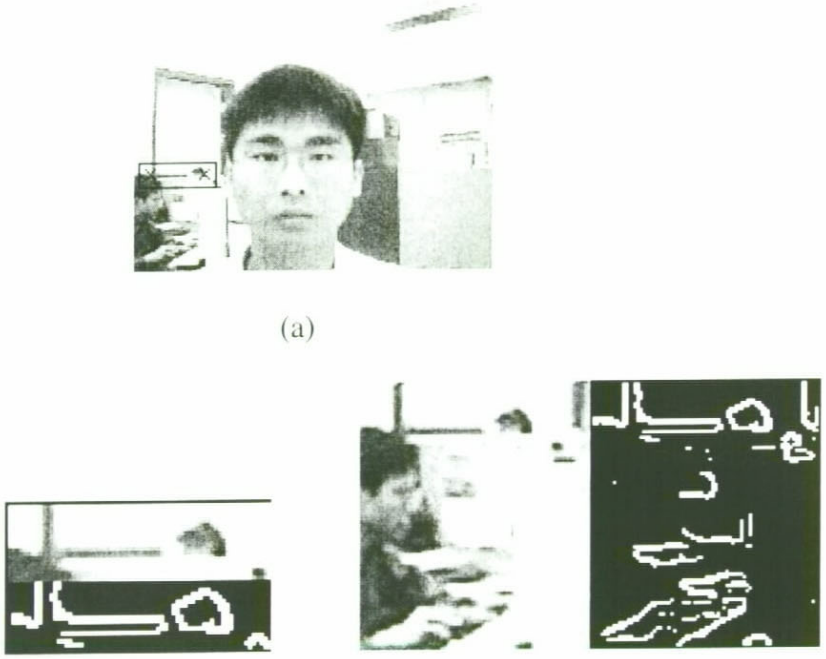
Table 3.2: Fractal dimensions of binarized eye-pair windows at different sizes.

Sample	Size	*FD* of face region
F1	50x60	1.9380
F2	46x59	1.9297
F3	52x65	1.9373
F4	50x62	1.9629
F5	48x62	1.9211
F6	50x62	1.9639
F7	52x64	1.9481
F8	54x66	1.9613
F9	48x60	1.9696
F10	50x64	1.9375
F11	52x62	1.9552
F12	46x59	1.9347
F13	48x60	1.9277
F14	52x64	1.9384
F15	48x62	1.9271
F16	50x64	1.9138
F17	52x64	1.9630
F18	54x64	1.9113
F19	48x62	1.9296
F20	52x62	1.9375

Table 3.3: Fractal dimensions of binarized face windows at different sizes.

The computed fractal dimensions of the eye-pair regions and face regions show a small variance under different scales, as shown in Tables 3.2 and 3.3. It is also found that there is a big difference between the fractal dimensions for facial images and

those for non-facial images. Fig. 3.11(a) illustrates a possible eye-pair candidate which is actually the background of an image. Fig. 3.11(b) and 3.11(c) show the invalidly selected eye-pair window and its corresponding face window. The corresponding fractal dimensions for these two binarized windows are 1.564 and 1.679, respectively, which is a big difference from the corresponding average fractal dimensions.



(a)

(b) (c)

Figure 3.11: (a) A possible eye pair is selected in background, (b) the invalid eye pair and its binarized image: fractal dimension of the binarized image is about 1.5638, and (c) the invalid face region and its binarized image: fractal dimension of the binarized image is about 1.6789.

Thus, a valid eye pair can be selected if the following criteria are satisfied.

$$| F_{face}(x, y) - M_{face}(x, y) | < t_5 \text{ and } | F_{eye}(x, y) - M_{eye}(x, y) | < t_6 \qquad (3.4)$$

where (x,y) represents the position of the eye-pair window and the face windows, F_{eye} and F_{face} are the fractal dimensions of the eye-pair and face regions, M_{eye} and M_{face} are the average fractal dimensions of the two windows, and t_5 and t_6 are the

thresholds. In our experiments, t_5 and t_6 are set at 0.038 and 0.035, respectively.

However, it is possible to detect more than one valid eye pair, which cluster around the valid eye-pair region, as shown in Fig. 3.12(a). In order to choose the overlapping valid eye pairs, the one with the lowest value of $|F_{eye}(x,y) - M_{eye}(x,y)| + |F_{face}(x,y) - M_{face}(x,y)|$ should be selected as the best eye pair in the overlapping region. Fig. 3.12(b) shows the best eye pair used in the representation.

(a) (b)

Figure 3.12: (a) Valid eye pairs, and (b) the best eye pair.

3.5 Experiment Results

The new approach for extracting the human eye pairs is tested using the MIT face database, ORL face database, and some complex images with a number of faces.

Some of the results are shown in Fig. 3.13. All the upright faces in the MIT face database and the complex images with a number of faces under head-on lighting are detected without any error. Table 3.4(a) shows the hit rates and miss rates under different conditions using the conventional fractal dimension. Table 3.4(b) shows the hit rates and miss rates using our new approach under different conditions. Table 3.4(c) shows the results of locating the human eye and face [32] using the Gabor wavelet. Our new approach achieves an overall hit rate of 100% without head tilt and under head-on lighting. When the heads tilt to the left or right, the hit rate is 90.6%. When

the light source to the faces is from 45°, the hit rates for the upright and tilted faces are 87.5% and 85.9%, respectively. However, without using our normalized and oriented fractal dimension, the hit rates for the upright faces with head-on lighting are only 75% and 78.1%, respectively. When the lighting is 90°, the hit rate for the upright face is 84.4% and the hit rate for the tilted face is 75%. Table 3.5 shows the hit rates and miss rates of our method based on the ORL database for upright faces and faces with perspective variations. Our approach achieves an overall hit rate of 92.9% for upright faces. When a face is subject to a slight perspective variation, the average hit rate is 82.4%. The experiment was performed on a Pentium II 450MHz PC. The processing time for locating eye pairs in an image of size 128×120 and 92×112 are less than 1.5s. In conclusion, our new approach can greatly improve the hit rate under a lighting source at 90°. The approach also works well even if the eye pairs and faces are under uneven lighting conditions, or subjected to slight perspective variations.

	Lighting	Head On		45°		90°	
	Head tilt	No	Tilt	No	Tilt	No	Tilt
Full Scale	Hit	16	28	13	26	13	17
	Miss	0	4	3	6	3	15
Medium Scale	Hit	14	26	11	24	9	14
	Miss	2	6	5	8	7	18

Table 3.4(a): Using the conventional fractal dimension.

	Lighting	Head On		45°		90°	
	Head tilt	No	Tilt	No	Tilt	No	Tilt
Full Scale	Hit	16	28	15	29	15	26
	Miss	0	4	1	3	1	6
Medium Scale	Hit	16	30	13	26	13	24
	Miss	0	2	3	6	3	8

Table 3.4(b): Using our new approach.

	Lighting	Head On		45°		90°	
	Head tilt	No	Tilt	No	Tilt	No	Tilt
Full Scale	Hit	16	28	14	24	14	19
	Miss	0	4	2	8	2	13
Medium Scale	Hit	14	26	12	26	10	15
	Miss	2	6	4	6	6	17

Table 3.4(c): Using the Gabor Wavelet.

Table 3.4: Experiment results based on the MIT face database using (a) the conventional fractal dimension, (b) our new approach, and (c) the Gabor wavelet.

Perspective Variation	No	Slight
Hit	70	51
Miss	5	9

Table 3.5: Experiment results based on the ORL face database using our new approach.

3.6 Conclusion

In this chapter, we have proposed an oriented fractal dimension for accurately extracting eye pairs in a complex image. Instead of searching the whole image space to determine the different scales of the eye pairs, only possible eye pairs are investigated. These possible eye pairs are identified by means of valley detection and measuring their fractal dimensions. In order to reduce the effect of uneven lighting conditions on the left-eye and the right-eye, eye windows are normalized by an histogram technique. In addition, when grouping a left eye candidate with a right eye candidate to form an eye pair, it is proposed that the oriented fractal dimension be used because it can provide a higher level of matching accuracy. The corresponding average fractal dimensions of the binarized eye-pair regions and the face regions are then used to verify whether the eye pairs selected are valid. The proposed method can successfully extract valid eye pairs under different scales, orientations, lighting conditions, and slight perspective variations.

Figure 3.13: Experiment results.

Chapter 4

A Reliable Method for Extracting the Chin Contour

4.1 Introduction

In Chapter 3, we presented an efficient and reliable approach for extracting the facial features such as eye. Based on the position of the facial features, we can locate the face region and the face contour. In this chapter, we will present a method to extract the chin contour, as the accuracy of its representation has a significant effect on the performance of the face recognition systems. There are many existing methods for the estimation of a chin contour. The active contour model and the deformable template are the typical examples. The active contour model (snake) is an energy minimizing spline which can be operated under the influence of both internal contour forces and edge forces. The edge forces lead the snake points to the boundary of a contour. However, the chin contour cannot be represented accurately using the model because the edge forces at the chin are usually weak, and the contour is not specified by *a priori* knowledge of the expected shape. The deformable template is used to estimate the chin contour. However, the template will be attracted by the valley forces and edge forces from the mouth if its initial position is not near the chin. In our approach, a reliable method for extracting a chin contour is proposed. The idea of the deformable template is employed and applied to chin contour extraction by considering two parabolic curves. These two curves are further represented by five control points, which are used to estimate the shape of the chin contour. We firstly evaluate the position of the mouth and the eyes, which provides useful information about the position of the chin. The initial position of the chin can then be located

approximately by the anthropometric measure. The chin model can be fitted into the chin region of a human face by minimizing a set of energy functions through an iteration process. The set of energy functions is devised in such a way that the salient features of the chin and *a priori* knowledge of the shape are used to guide the detection process. A proposed method is employed to extract the chin contour during the optimization process.

4.2 Locating the Eyes and Mouth Position

The positions of two eyes are located by means of fractal dimension and valley field detection which have been presented in previous chapter. Having located the eyes, the position of the mouth is then located by means of integral projection.

The horizontal and vertical projection functions are defined as follows:

$$H(y) = \sum_{x=x_1}^{x_2} f(x, y)$$

$$V(x) = \sum_{y=y_1}^{y_2} f(x, y) \qquad (4.1)$$

where f(x,y) is the gray level intensity. The position of the mouth can then be located by the two projections as shown in Fig 4.1a.

4.3 Extraction of the Chin Contour in Facial Images

Our method of chin contour extraction is divided into two stages. Firstly, the initial position of the chin contour is estimated and represented using a chin template. In our approach, five points are used to control the shape of the template. The initial position of the chin is estimated by means of anthropometric measure. Furthermore, the

position of the cheeks is approximated using projection techniques. A set of energy functions is then defined to guide the movement of the chin template in the optimization process. The details of our algorithm is described here.

4.3.1 Initial Position of the Chin Contour

The chin contour is represented by two parabolic curves, (x_1, x_2, x_3) and (x_1, x_4, x_5), which consist of five control points, as shown in Fig 4.2. The position of the point x_1 is estimated from the typical anatomy of the human face by considering the position of the mouth and eyes. As the mouth is used in the estimation, the effect of the edge forces and valley forces from the mouth can be avoided during the optimization process.

A method for locating the position of the eyes and the mouth has been described in Section 4.2. A window is then set up to determine the position of the cheeks (points x_2 and x_4) by applying vertical projection on the edge image, as shown in Fig 4.1b. In order to extract the chin contour accurately, points x_3 and x_5 are placed in between the points x_1, x_2 and points x_1, x_4, respectively. These five points are then allowed to move to the chin contour.

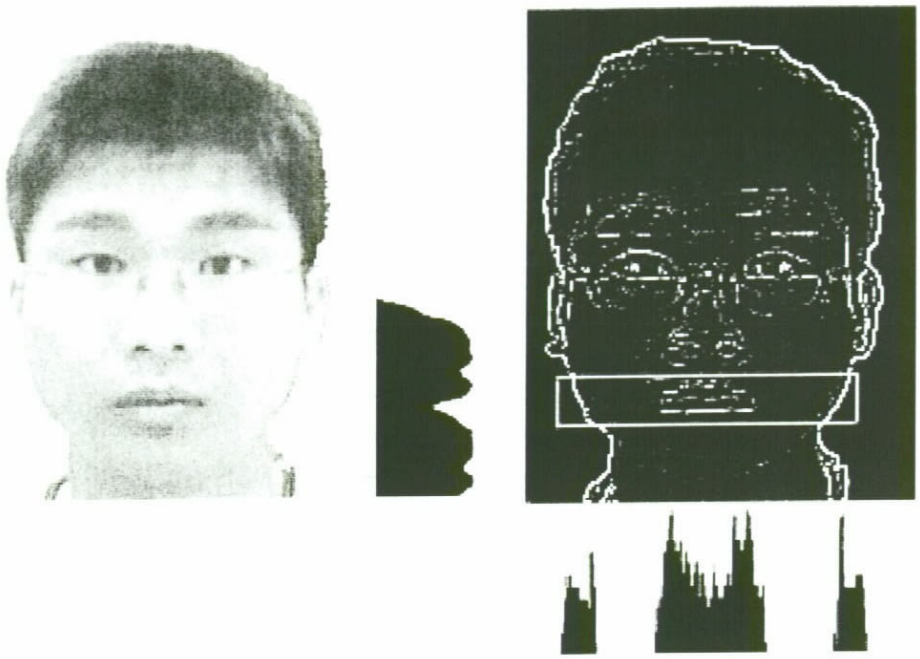


Figure 4.1. a) The mouth position extracted by using horizontal projection. b) Two boundary points extracted by using vertical projection in an edge image.

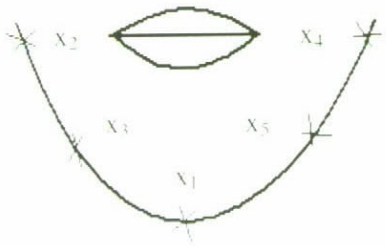


Figure 4.2. The initial points and labels for the chin contour

4.3.2 The Energy Functions for Estimation of the Chin Contour

The chin contour of a facial image can be extracted reliably through an optimization process. The template is composed of two parabolic curves which are described by

five control points. These points are allowed to move to adapt the template to the chin contour of the facial image. This matching process will minimize the sum of the energy functions. These energy functions are composed of the smoothed edge forces Φ_e, the image forces Φ_I, *a prior* shape of a chin, and the intensity difference. The smoothed edge forces can attract the distant control points towards the chin position. However, if the initial position of the control points is not close to the actual chin position, the edge forces from the mouth or the neck might dominate, and a wrong chin contour will result. Therefore, the estimation of the initial position of the control points is vital to the success of chin extraction. In addition, a salient feature of the chin is the difference in image intensity between the regions above and below the chin contour. Thus, the intensity difference term can make the template locate the chin contour more accurately. The different energy functions are defined as follows.

1. The edge energy potential, E_e, measures the blurred edge intensity value along a parabolic curve, as shown in the following equation.

$$E_e = -\frac{1}{|Lp|}\oint \Phi_e ds \tag{4.2}$$

where |Lp| represents the length of the parabolic curve. This term attempts to attract the control points to the boundary of the chin contour.

2. The intensity difference, E_d, measures the difference in image intensity between the regions above and below the chin template.

$$E_d = -D(R_i, R_O) \tag{4.3}$$

$$= -\left(\frac{1}{n_1}\sum_{Ri} I(x,y) - \frac{1}{n_2}\sum_{R_O} I(x,y)\right)$$

where R_i and R_o represent the regions in the inner part and the outer part of the chin template, respectively, while n_1 and n_2 represent the number of pixels in the

corresponding regions. If the template is located at the chin contour, this energy term will become a minimum.

There are three possible situations for the location of the template:

1) The template is located near the mouth. The value of E_d will become positive as the image intensity in the mouth region decreases.

2) The template is located between the mouth and chin contour. The value of E_d is close to zero as the pixel intensities in these two regions are very similar.

3) The template is located in the chin contour. The value of E_d becomes negative due to the fact that the gray-level intensity around the neck is lower than that above the chin contour due to the effect of lighting.

3. The valley energy term, E_v, defines the summation of valley intensity along the parabolic curve, which is then divided by its length.

$$E_V = -\frac{1}{|Lp|}\oint \Phi_V ds \tag{4.4}$$

This term is defined in such a way because the chin contour exhibits as a region of a valley,so the valley force along the chin contour is stronger. This term attempts to attract the control points to the boundary of the chin contour.

4. The *a priori* energy terms are defined to maintain the shape of a chin contour that is similar to the parabolic curve. Therefore, the Euclidean distance is used to measure the similarity. The equation is defined as follows:

$$D = \sum_{K=x2}^{x4} \| f_1(K) - f_2(K) \| \tag{4.5}$$

where f_1 represents the parabolic curve formed by the points x_2, x_1 and x_4 as shown in Fig 4.3, denoted as the line (a). f_2 represents the chin contour formed by combining two parabolic curves, as shown in Fig 4.3, denoted as the line (b). K

represents the position along the curves from point x_2 to x_4. This term D attempts to keep the shape of the chin contour close to a parabolic curve.

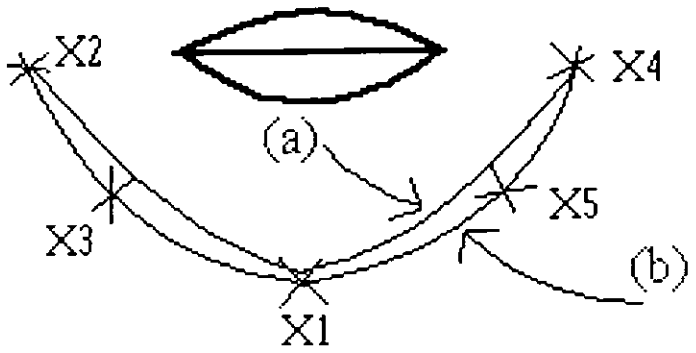


Figure 4.3. Line (a) is the parabolic curve formed by the three points (points x_2, x_1 and x_4). Line (b) is the curve formed by the two parabolic curves (points x_2, x_3, x_1 and the points x_1, x_5, x_4).

The Energy Terms Associated with the Control Points

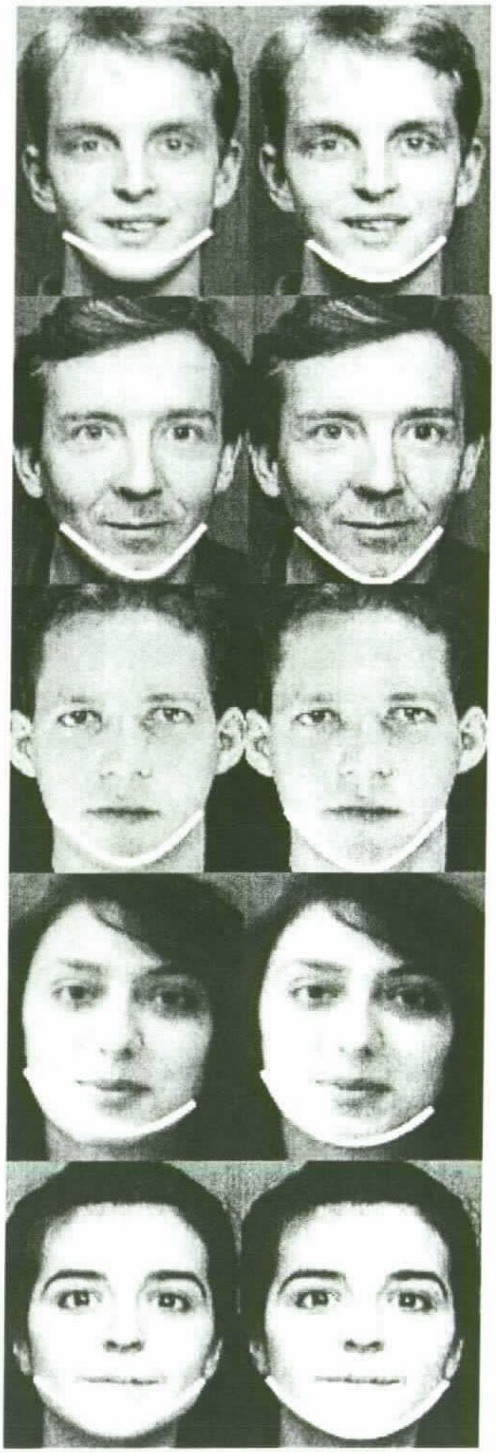
The new position of a control point is determined by the total energy potential at its current position and its eight neighbouring points. The new position is where the total energy potential is a minimum. The different control points are assigned with different energy terms, as shown in the following.

$$E_i(\vec{x}_i)=c_1\frac{E_v(\vec{x}_i)}{E_{v\max}}+c_2\frac{E_e(\vec{x}_i)}{E_{e\max}} \qquad \text{where i=1,2 and 4} \tag{4.6}$$

$$E_i(\vec{x}_3)=c_3\frac{E_v(\vec{x}_3)}{E_{v\max}}+c_4\frac{E_e(\vec{x}_3)}{E_{e\max}}+c_5\frac{E_d(\vec{x}_3)}{E_{d\max}}+c_6\frac{D(\vec{x}_3)}{D_{\max}} \tag{4.7}$$

$$E_i(\vec{x}_5)=c_7\frac{E_v(\vec{x}_5)}{E_{v\max}}+c_8\frac{E_e(\vec{x}_5)}{E_{e\max}}+c_9\frac{E_d(\vec{x}_5)}{E_{d\max}}+c_{10}\frac{D(\vec{x}_5)}{D_{\max}} \tag{4.8}$$

where $\vec{x}_i$ represents the position of a control point. The maximum value of each energy term is used for normalization. The interation will be stopped if the changes in energy of all five control points are less than a threshold.



(a) (b)

Figure 4.4. (a) The initial templates in facial images, (b): The final templates for the estimated chin contour in facial images.

4.4 Experimental Results

The proposed algorithm has been applied to over 40 head-and-shoulders images. The experiment shows that the chin contour can be extracted accurately in about 83.3% of the images. The remainder was unsuccessful due to adverse lighting effects. A chin extraction is considered to be successful if the two parabolic curves exactly placed precisely on the chin contour. The initial and final templates are illustrated in Fig 4.4a and Fig 4.4b, respectively. The required total runtime is less than 1 second: half of the total is spent locating the eye pair and the mouth, and the other half is for extracting the chin contour. The experiment was conducted on a Pentium II 450MHz personal computer.

4.5 Conclusion

In this chapter, a method for locating and extracting the chin contour of a facial image has been introduced. The head is first located and detected by an active contour model. The position of the eyes and mouth is then estimated. The initial location of the template is determined by anthropmetric measure, so the effect of the edge forces from the mouth and the neck can be avoided. The template is represented by two parabolic curves and described by five points. A set of energy terms is defined and applied to locate the chin contour accurately. Experiment results show that this method is accurate.

Chapter 5
New Modified Hausdorff Distances For Human Face Recognition

5.1 Introduction

Human face recognition has a wide range of applications such as criminal identification, security systems, credit card verification, automatic teller machine transactions, etc. Various approaches to human face recognition such as the holistic and analytic approaches will be presented in Chapter 2.

Psychological studies have revealed that humans can categorize a human face at a glance and recognize the line drawings of objects as quickly and almost as accurately as photographs [17]. It is suggested that the edge-like retinal images of faces can provide useful information for face identification at the level of early vision. Furthermore, when the template matching method [13] is used, edge-like maps can achieve a better performance than the feature-based techniques. The advantage of using edges as image features is that they can provide robustness to illumination change and simplicity of presentation. In this chapter, two new modified Hausdorff distances are proposed for face recognition. These new distance measures incorporate the spatial information of a human face. In order to evaluate this new method, three face databases, namely the MIT, Yale, and ORL, are used in the experiment. The results show that our proposed modified Hausdorff distance measures can achieve higher recognition rates than those of other existing Hausdorff distance measures.

This chapter is organized as follows. The basic concept of Hausdorff distance is introduced in Section 5.2. The two new modified Hausdorff distances, namely spatially weighted Hausdorff distance (SWHD) and spatially weighted 'doubly'

Hausdorff distance (SW2HD), are proposed and defined in Section 5.3. Section 5.4 presents the experimental results concerning the recognition rates of the various modified Hausdorff measures. Finally, a conclusion is given in Section 5.5.

5.2 Hausdorff Distance

One of the shape comparison methods is Hausdorff distance, which is based on a distance measure of the edge maps of two objects. A major advantage of this distance measure is that the distance can be calculated without the explicit pairing of points in their respective data sets. Given two finite point sets $A=\{a_1,.....,a_p\}$ and $B=\{b_1,.....,b_q\}$, Hausdorff distance is defined as follows:

$$H(A,B) = \max\{h(A,B),h(B,A)\} \tag{5.1}$$

$$\text{where } h(A,B) = \max_{a\in A} \min_{b\in B} \|a-b\|$$

$\|.\|$ is an underlying norm on the point sets A and B. The function $h(A,B)$ is called directed Hausdorff distance from point set A to B. It identifies the point $a \in$ A that is farthest from any point of B and measures the distance from point a to its nearest neighbor in B. Hence, Hausdorff distance measures the mismatch between the two point sets and can be used as a measure for shape comparsion. This shape comparsion method does not require any explicit correspondence between the model and the image data set, as it does not build a one-to-one pairing between the model and the image feature points. A number of modified Hausdorff distance measure methods [80-81] have been proposed which provide a more reliable and robust measure than the original one. Takács [17] further improved the modified Hausdorff distance for face recognition. This modified Hausdorff distance introduces the notion of neighborhood

function N_B^a and assoicated penalty (P), which is called 'doubly' modified Hausdorff distance (*M2HD*). This modified Hausdorff distance is defined as follows:

$$H(A,B) = \max\{h(A,B), h(B,A)\} \tag{5.2}$$

where $d(a,B) = \max(I \cdot \min_{b \in N_b^a} \|a-b\|, (1-I)\cdot P)$

and

$$h(A,B) = \frac{1}{N_a} \sum_{a \in A} d(a,B)$$

In this modified definition, N_B^a is the neighborhood of point a in set B. I is an indicator, which is equal to 1 if there exists a point $b \in N_B^a$, and 0 if otherwise. This formulation may have two different values for *h(A,B)*: when all matching pairs fall within a given neighborhood, its value will be the same as the original Hausdorff distance; however, if no matching pair is found, then the penalty value P is considered. It has been shown that the 'doubly' modified Hausdorff distance is suitable for face recognition, because it accounts for small and non-rigid local distortions.

5.3 Matching Face Images Using Spatially Weighted 'Doubly' Hausdorff Distances

In human face recognition, the different facial regions have different degrees of importance; especially the eye and the mouth regions. However, traditional Hausdorff distance [79] and the 'doubly' modified Hausdorff distance do not consider the relative importance between the different facial regions, and make no distinction between different parts of the face. We therefore devise a new modified Hausdorff distance, namely spatial weighted hausdorff distance (SWHD). This proposed Hausdorff distance is specific to face recognition and places more emphasis on the crucial facial features, such as the eye and mouth regions. The *SWHD* is also

combined with the *M2HD* to provide a distance measure which can achieve a higher recognition rate than that of both the *M2HD* and *SWHD*.

In the implementation, the position of a human face in an input facial image is detected and the corresponding position of the two eyes is located. The edge map of an input facial image is generated and then normalized to a specific size based on the inter-distance of the two eyes. This normalized edge facial image and the facial image in the database to be compared are then aligned based on the location of the eyes. Finally, distance transform [82] is applied to the input edge image and different Hausdorff distances are used to measure the similarity between the input human face and the face in the database. The input edge facial image is matched against a database of faces and the results are ranked according to the measure of similarity. Fig. 5.1 shows the whole process of our new proposed face recognition system.

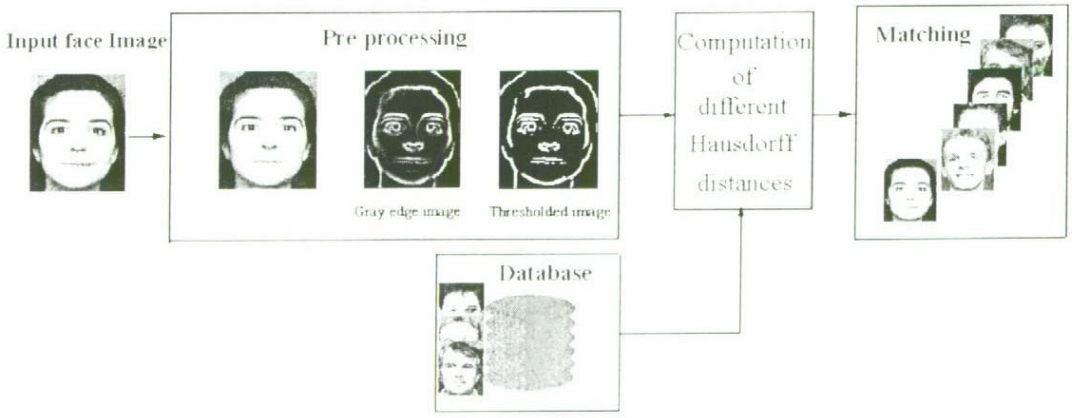


Figure 5.1: The architecture of our face recognition system.

5.3.1 Edge detection and thresholding

In applying Hausdorff distance to face matching, the shape comparison basically operates on edge maps. Consequently, the edge detection algorithm used will have significant affect on the matching performance based on Hausdorff distance measure. However, when converting a gray scale edge image to a binary edge image, a fixed threshold does not work well. This is due to the fact that the contrast in an edge image may vary significantly. In our approach, we use an adaptive threshold for each edge image to obtain the corresponding thresholded binary image. The threshold T is calculated adaptively for an image by considering the gray-level intensities of the 15% highest pixel intensities. Assume that the histogram of an edge image of size $h \times w$ is $his(i)$, where $i = 0,1,\ldots,255$. T is then determined as the largest value such that the following equation is satisfied:

$$\sum_{i=T}^{255} hist\ (i) \geq 0.15\ h \times w \qquad (5.3)$$

Fig. 5.2 shows some edge images based on this adaptive thresholding scheme. Fig. 5.2(a) and Fig. 5.2(d) are the original images, Fig. 5.2(b) and Fig. 5.2(e) are the gray-level edge images, and Fig. 5.2(c) and Fig. 5.2(f) are the thresholded edge images.

(a) (b) (c)

(d) (e) (f)

Figure 5.2: First column: original facial images; second column: gray-level edge images; third column: binary edge images by using adaptive thresholding method.

5.3.2 Spatially Weighted Modified Hausdorff distance

A new modified Hausdorff distance for face recognition is proposed in this paper based on the fact that the various facial regions have different degrees of importance, whereby the eyes and mouth play the most important role in face recognition. A weighted function is defined according to the spatial position of the respective regions of the facial features, hence this modified Hausdorff distance measure is called spatially weighted Hausdorff distance (SWHD). The following is the definition of this new Hausdorff distance:

Given two finite point sets $A=\{a_1,\ldots,a_p\}$ and $B=\{b_1,\ldots,b_q\}$,

$$H(A,B) = \max(h_{sw}(A,B), h_{sw}(B,A)) \tag{5.4}$$

The spatially weighted directed Hausdorff distance, $h_{sw}(A,B)$, is defined as follows:

$$h_{sw}(A,B) = \frac{1}{N_a} \cdot \sum_{a \in A} w(b) \min_{b \in B} \|a - b\| \tag{5.5}$$

where $\|\cdot\|$ is an underlying norm on the points of A and B; and N_a is the number of points in set A. $w(x)$ is a weighted function, whose definition is:

$$w(x) = \begin{cases} 1 & x \in R_{ie} \\ w_u & x \in R_u \\ w_{om} & R_{im} \cap R_{gm} \\ w_{nm} & (R_{im} \cap R_{gm}{}^{c}) \cup (R_{im}{}^{c} \cap R_{gm}) \\ 0 & x \in R_b \end{cases} \tag{5.6}$$

where R_{ie} represents the regions of two eyes, R_{im} and R_{gm} are the mouth regions of the input facial image (probe image) and the facial image in the database (gallery image), R^c represents the complement of the region itself, R_u is the facial region other than the eye and mouth regions, and R_b is the background region that contains no facial parts. As the two images are aligned based on the two eyes, only parts of the respective mouth regions will overlap. Different values are used in the weighted function $w(x)$ according to the position of a point in a facial image. Those points in the important facial regions, R_i and $(R_{im} \cap R_{gm})$, are fully counted (i.e. 1), while edge points that occur in the background are ignored. Those points in the less important facial regions, $(R_{im} \cap R_{gm}{}^{c}) \cup (R_{im}{}^{c} \cap R_{gm})$, and unimportant facial regions, R_u, will be suppressed by the weighted values of w_{nm} (0.7) and w_u (0.5), respectively, which are less than one. Also, the weight of the background is 0, so non-facial points will be ignored. Fig. 5.3 illustrates a 3D representation of the spatially weighted function obtained by superimposing the probe image and the gallery image.

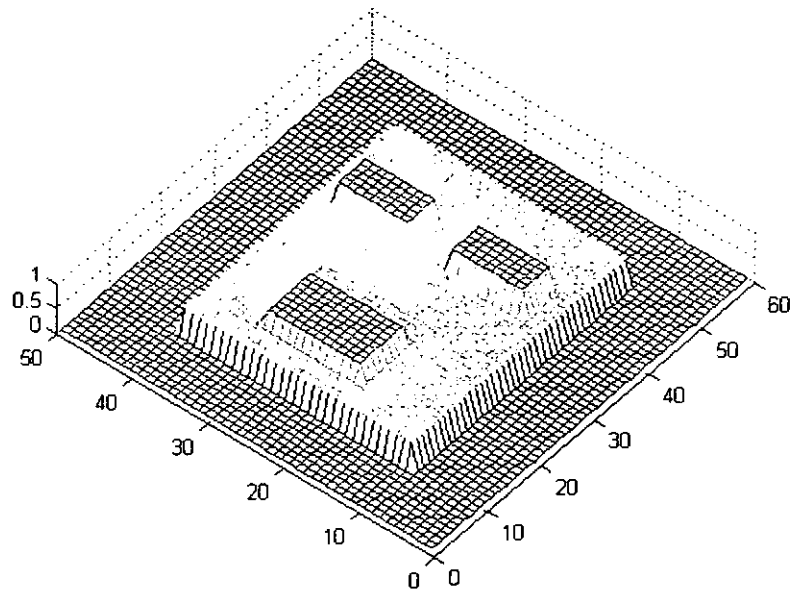


Figure 5.3: The 3D graph of a spatially weighted function w(x).

The *SWHD* takes into account the different degrees of importance of different regions in a facial image. However, differences in facial expressions will lead to differences in the eye windows and the mouth window. In order to alleviate this effect, we combine the *M2HD* with the *SWHD* to have a new modified hausdorff distance, namely spatially weighted 'doubly' Hausdorff distance (SW2HD). Its corresponding directed Hausdorff is then defined as follows:

$$d(a,B) = \max(I \min_{b \in N_B^a} \|a-b\|, (1-I)P) \tag{5.7}$$

$$h(A,B) = \frac{1}{N_a} \sum_{a \in A} w(a) \cdot d(a,B) \tag{5.8}$$

The position of the mouth region is determined based on the position of the two eyes and estimated by means of the integral projection technique [13]. Integral projection has been used successfully for the extraction of facial features. Based on the inter-distance between the two eye candidates, we can set up a window to cover the mouth region. The height, h, and width, w, of the window is half and equal to the inter-distance of the two eyes, L, respectively, as shown in the Fig. 5.4.

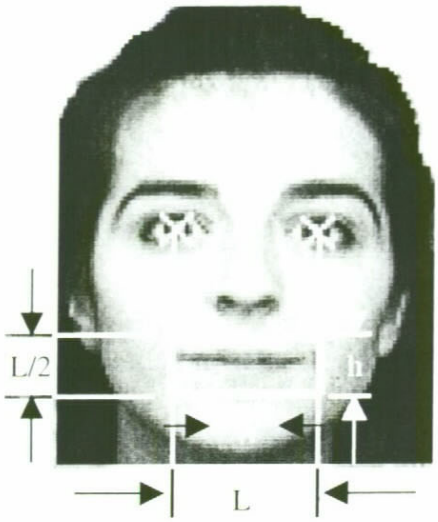


Figure 5.4: Two eyes and mouth window.

5.4 Experimental Results

The MIT, Yale (from the Yale University), and ORL (from the Oliver Research Laboratory in Cambridge, U.K) face databases were used in the experiment. The MIT database has 144 face images with 16 distinct subjects, while the Yale database has 150 face images with 15 distinct subjects. For the ORL database, there are 400 different face images corresponding to 40 distinct persons, but only 6 images for each of the 40 subjects were included in the testing set. The experiment setup is that an upright frontal view of each of the 71 subjects with suitable scale and normal facial expression was chosen to form a database consisting of 71 persons. Each of the testing images was then compared to each face in the database. The results were ranked and arranged in ascending order of their measured similarity with respect to the current input.

In the experiment, the principle component analysis (PCA), the Hausdorff distance (HD), the 'doubly' modified Hausdorff distance (M2HD), the spatially weighted Hausdorff distance (SWHD), and the spatially weighted 'doubly' Hausdorff distance (SW2HD) of a testing facial image and each face in the database were computed. The

results were ranked and arranged in ascending order according to their corresponding measured distances. To choose the optimum parameters for the two proposed Hausdorff distances, we conducted several experiments with different values of w_u, w_{nm}, and P. Table 5.1 tabulates the recognition rates for different values of w_u, w_{nm} and P. We found that the best recognition rate was achieved when the parameters w_u and w_{nm} are set at 0.5 and 0.7, respectively, and the penalty value P=4. Fig. 5.5 illustrates the cumulative recognition rates of the *PCA*, *HD*, *M2HD*, *SWHD*, and *SW2HD* based on the combined database, which in total has 71 subjects. Our proposed Hausdorff distances, *SWHD* and *SW2HD*, can both achieve a better performance than that of the *PCA*, *HD* and *M2HD* methods, while the *SW2HD* produced the best performance. The recognition rates for *PCA*, *HD*, *M2HD*, *SWHD* and *SW2HD* are *60%*, 55%, 74%, 80% and 81%, respectively. The *SW2HD* method can achieve recognition rates of 87%, 89%, and 97% for the first three, the first five, and the first ten likely matched faces, respectively, as shown in Fig. 5.5. Table 5.2 shows the corresponding recognition rates of the five methods based on the three individual databases and the combined database. The experiments were conducted on a Pentium II 450MHz computer system. The average runtime for our new method is less than two seconds.

P=4.0	w_u=0.3	w_u=0.5	w_u=0.7
w_{nm}=0.5	84%	87%	86%
w_{nm}=0.7	87%	89%	88%
w_{nm}=1.0	87%	87%	86%

P=5.0	w_u=0.3	w_u=0.5	w_u=0.7
w_{nm}=0.5	83%	87%	86%
w_{nm}=0.7	85%	88%	87%
w_{nm}=1.0	88%	87%	87%

P=6.0	w_u=0.3	w_u=0.5	w_u=0.7
w_{nm}=0.5	85%	87%	87%
w_{nm}=0.7	85%	88%	87%
w_{nm}=1.0	86%	88%	87%

Table 5.1: Recognition rates using different values of w_{nm}, w_u and P.

	PCA	HD	M2HD	SWHD	SW2HD
ORL	63%	46%	75%	82%	88%
Yale	50%	66%	80%	82%	83%
MIT	50%	49%	80%	84%	89%
Combined	60%	55%	74%	80%	81%

Table 5.2: Recognition rates of the five face recognition methods: PCA, HD, M2HD, SWHD, and SW2HD, with different databases. The number of faces in the MIT, Yale, ORL, and combined databases are 16, 15, 40, and 71, respectively.

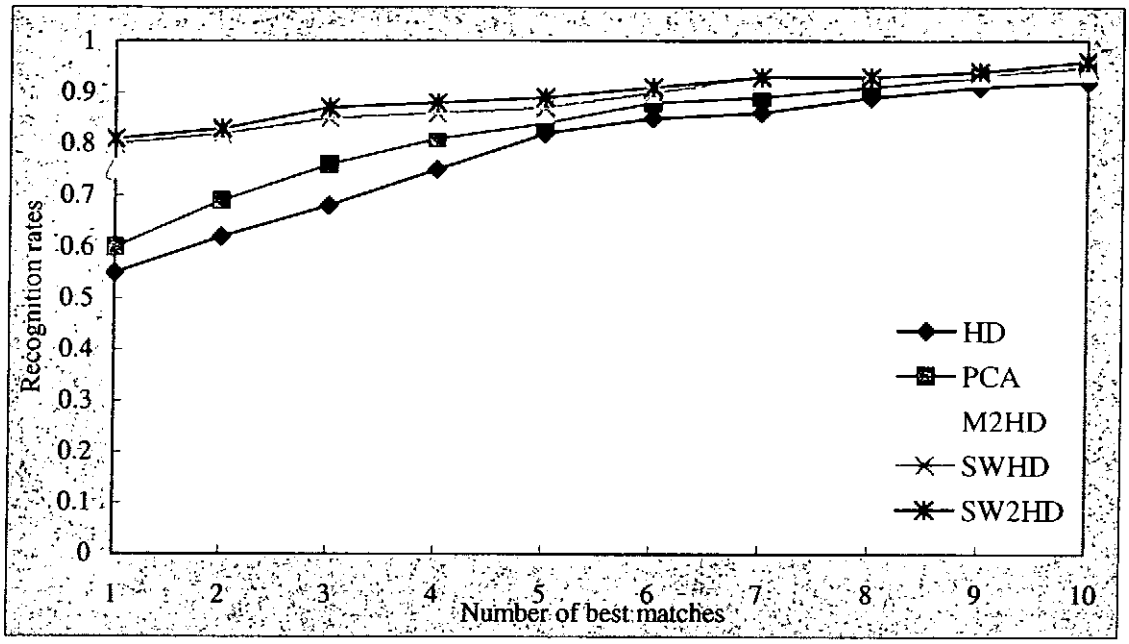


Figure 5.5: Comparison of the overall recognition rates using 534 testing face images.

5.5 Conclusion

In this chapter, new modified distances are proposed for human face recognition. Because different facial regions have different degrees of importance for face recognition, the distance measures proposed incorporate the *a priori* structure of a human face to emphasize the importance of facial regions. The two proposed Hausdorff distances are the spatially weighted Hausdorff distance (SWHD) and the spatially weighted 'doubly' Hausdorff distance (SW2HD), respectively. The *SW2HD* includes a penalty function when the measured distance for a point is larger than a certain value. This distance measure can reduce the effect of outliers on the image. Experimental results based on the combined face databases show that this new distance measure can achieve a higher recognition rate when compared to the *HD* and *M2HD* methods. The recognition rates of the *SWHD* method are 80%, 87% and 95% for the first one, the first five, and the first ten likely matched faces, respectively, while the corresponding recognition rates for *SW2HD* are 81%, 89% and 97%, respectively The techniques proposed in this chapter are computationally simple and

can provide a reasonable performance level. In practice, our approach can be used as the first stage of a human face recognition system which selects those similar faces to an input face from a large face database. Those selected faces are then further analyzed by means of more computationally intensive and more accurate methods.

Chapter 6
Automatic Human Face Recognition System Using Fractal Dimension and Modified Hausdorff Distance

6.1 Introduction

In this chapter, the efficient methods developed in this thesis for locating the two eyes and recognizing the human face are combined to form an automatic human face recognition system. The position of the two eyes are located based on the valley field detection and measurement of fractal dimensions, as described in Chapter 3. After the eyes are located, a modified Hausdorff distances introduced in Chapter 5 will be used to compare two human faces.

This chapter is organized as follows. Section 2 introduces an approach to locating the eyes. Spatial weighted Hausdorff distance, and spatial weighted "doubly" modified Hausdorff distance are described in Section 3. Section 4 presents the experimental results. Finally, conclusions are given in Section 5.

6.2 Detecting the Eye Pairs

We employ our new approach for locating the eye pairs in an image as described in Chapter 3. Our new approach is based on the surface roughness of facial features by means of fractal dimension. Box-counting is one of the methods which estimates the fractal dimension. This method is an efficient technique for estimating fractal dimensions, and can be applied to gray-level images and binary images. The experiment result shown that our new approach for locating the eye pairs can achieve better performance.

6.2.1 Valley Detection of the Eyes

As the iris has a relatively low gray-level intensity in a human face, a valley exists at an eye region. The valley field, Φ_v, is extracted by means of morphological operators. The criteria are then used to identify the best eye candidate in face region.

6.2.2 Grouping the Best Eye Candidates to form Possible Eye Pairs

Two eye candidates are paired to form a possible eye pair. We assume that the rotation of the human face is less than 45°. The two eyes of a face should have similar orientations and be of similar size. Moreover, the roughness of the two eye regions should be close to each other. In our approach, we use two rectangular boxes, namely a vertical rectangular box and a horizontal rectangular box of different orientations instead of a square box in box-counting. The two measured fractal dimensions are called horizontal fractal dimension, FD_h, and vertical fractal dimension, FD_v, respectively. The possible eye candidates are then validated by the oriented fractal dimension to form possible valid eye pairs. Our experiments show that by using the oriented fractal dimension we can form eye pairs more reliably than cases where the conventional fractal dimension is used as described in section 3.3.2 of Chapter 3.

6.2.3 Verification of Eye Pairs

The eye pairs selected in previous stage are passed to the next stage for further verification. At this stage, the measurement of each possible eye-pair region and its corresponding face region will be computed by means of fractal dimension. In order to reduce the effect of lighting conditions, the edge images are used in the computation of the fractal dimensions. In our study, we computed the fractal dimensions of a number of eye-pair windows and face windows. We observed that the computed fractal dimensions of the eye-pair regions and face regions show a small variance under different scales. A large difference is found between the fractal dimensions for facial images and those for non-facial images. Thus, a number of criteria are performed to select a valid eye pair as described in section 3.3.3 of Chapter 3.

6.3 Face Recognition Using Spatially Weighted Hausdorff Distance

In human face recognition, different facial regions have different degrees of importance. The eyes and mouth regions on a face are crucial features for identification and are therefore more important than other parts of the face. However, the traditional Hausdorff distance does not consider the different degrees of importance between different facial regions, and makes no distinction between different parts of the face. We therefore devise a new Hausdorff distance measure that is specific to face recognition and put more emphasis on the crucial facial features, including the eyes and mouth regions. The method is described in Chapter 5.

6.4 Experimental Results

The input to our automatic human face recognition system is a facial image. The system locates the position of the two eyes automatically. The face region is then normalized according to the inter-distance of the two eyes, and the weighted function for the input face can then be generated. Based on the position of the two eyes, we can also estimate the face region. The corresponding edge map of the face region is then produced and two human faces, one from the input and the other from the face gallery, are compared using the spatially weighted Hausdorff distance.

We used the ORL face database in the experiment. An upright frontal view of each of the 40 subjects with a normal facial expression was chosen in our experiment. Among the rest of the faces, 6 images for each of the 40 subjects were selected to form a pool of 240 faces as a testing set.

Figure 6.1 shows the cumulative recognition rates from our experiment based on spatially weighted Hausdorff distance and the spatially weighted doubly modified Hausdorff distance with automatic selection of two eyes. The automatic face recognition system using spatially weighted Hausdorff distance can achieve recognition rates of 76%, 84% and 92% for the first one, the first five, and the first ten likely matched faces, respectively, while using spatially weighted doubly modified Hausdorff distance can achieve recognition rates of 80%, 88%, 83% for the first one, the first five, and the first ten likely matched faces, respectively. The experiments were conducted on a Pentium III 733MHz computer. The average runtime for detecting the eyes and recognizing a human face input from a database of 240 faces is about two seconds: slightly less than one second for eye detection, and slightly more than one second for face recognition

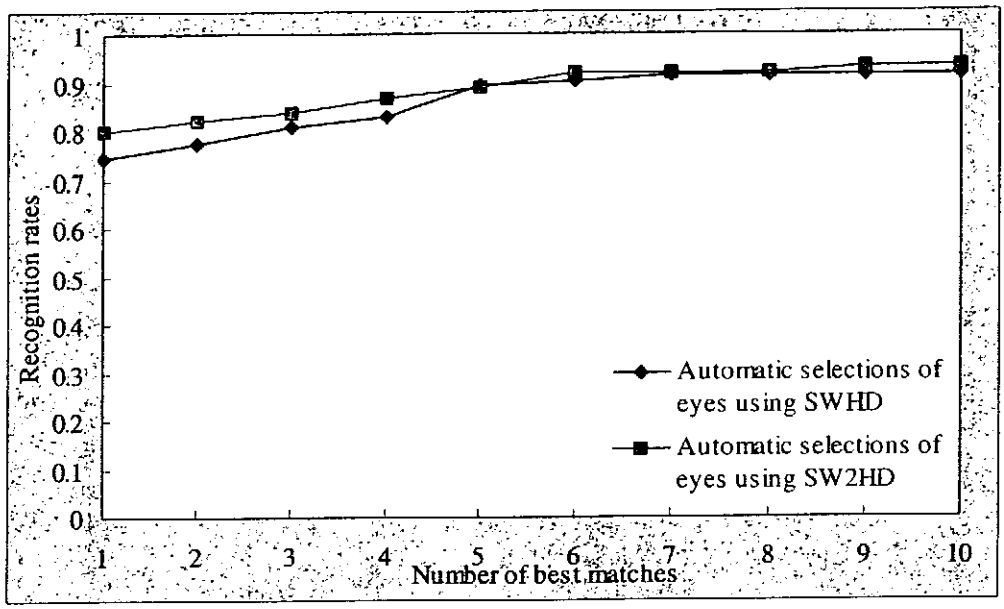


Figure. 6.1. Overall recognition rates using 240 testing face images.

6.5 Conclusion

An automatic human face recognition is proposed which combines the detection of the two eyes by fractal dimension, and face recognition by a modified Hausdorff distance. We have employed an oriented fractal dimension to accurately extract the eye pairs. The effect of uneven lighting conditions on the left-eye and the right-eye can be reduced by normalization to a specific gray-level intensity. Moreover, when grouping a left eye candidate with a right eye candidate to form an eye pair, the oriented fractal dimension provides a higher level of matching accuracy. A spatially weighted Hausdorff distance is also employed in this chapter for human face recognition. Since different facial regions have different degrees of importance for face recognition, this new distance measure incorporates the *a priori* structure of a human face by emphasizing the important facial regions. This can reduce the effect of outliers on the image. Experimental results based on the ORL database show that our automatic human face recognition can achieve recognition rates of 76%, 84%, and

92% for the first one, the first five, the first ten likely matched faces, respectively. The corresponding recognition rates manual selection of the two eyes are 82%, 95% and 98%, and the first one, the first five and the first ten likely matched faces respectively.

Chapter 7
Conclusion and Future Work

7.1 Conclusion

In this thesis, we have introduced various techniques for human face detection, facial feature extraction and human face recognition. For human face detection, the knowledge-based approach, shape information approach, example-based approach, feature-based approach and color analysis approach have been reviewed. Most of these approaches have a high computational complexity and are unable to detect rotated faces. Thus, these approaches might not be suitable for real-time applications. In addition, we have reviewed some approaches for facial feature extraction, such as the integral projection approach, the deformable template and the active contour model. The deformable template and the active contour model have been widely adopted; they can achieve a better performance in facial feature extraction, such as the mouth, the eyes and face contour. However, these methods are time-consuming due to the fact that a large number of parameters are involved during the optimization process. Various approaches for human face recognition have also been reviewed. One of them is called the analytic approach. This approach considers a set of geometrical features, such as eyes, mouth and face contour, which can be obtained by means of the deformable template. Although the approach can provide high flexibility in non-rigid facial features, its success relies heavily on the accuracy of facial feature extraction. Another approach which considers the global properties of a human face is

called the holistic approach. However, the recognition rate will be degraded if the face to be recognized is not aligned well.

In our research, we proposed efficient methods for human face detection, facial feature extraction, and human face recognition. To locate an eye pair and detect the human face in an image, we have devised a fast approach based on valley field detection and the measurement of fractal dimensions. Based on the position of the two eyes, a new template that can extract the chin contour reliably and accurately has also been proposed. Furthermore, an efficient approach for human face recognition based on a modified Hausdorff distance has been presented.

Our efficient approach for detecting the human face and locating eye pairs consists of three stages. Firstly, possible eye candidates in an image with a complex background are identified by means of the valley field detection. These possible eye candidates are passed to stage 2 to form possible eye pairs. Two possible eye candidates with similar fractal dimensions are grouped to form a possible eye pair in stage 2. As fractal dimension is sensitive to lighting conditions and lack of orientation information, our proposed fractal dimensions are less sensitive to lighting conditions and can provide information about the orientation of an image under consideration. Based on the input eye pair, the corresponding face region can be formed in stage 3. The input eye-pair window and the corresponding face region are then binarized, and their respective fractal dimensions are measured. These two fractal dimensions are then compared with the average fractal dimensions of the eye-pair window and face region. Experimental results have shown that the eye pairs can be detected reliably under different orientations, lighting conditions, and slight perspective variations. We have also presented a method to extract the chin contour based on the position of the two eyes and the mouth. There are many existing methods, such as the deformable

template and active contour model, for the estimation of a chin contour. However, the chin contour cannot be represented accurately using the model because the edge forces at the chin are usually weak, and the contour is not specified by *a priori* knowledge of the expected shape. Furthermore, the template will be attracted by the valley forces and edge forces from the mouth if its initial position is not near the chin. Therefore, a new template consisting of two parabolic curves is employed and applied to chin contour extraction. The two parabolic curves are represented by five control points. The chin model can be fitted into the chin region of a human face by minimizing a set of energy functions through an iteration process. Experimental results showed that the chin contour can be detected accurately. The human face recognition technique proposed in this thesis is based on a new modified Hausdorff distance method. This distance measure incorporates information about the location of important facial features comparing the edge maps of two facial images. The distance measure is weighted according to a weighted function derived from the spatial information of a human face. As a result, this new distance measure can achieve a higher recognition level than the original Hausdorff distance or the modified Hausdorff distance method. The major advantages of this method are that it is less sensitive to the lighting conditions and is computationally simple.

Finally, an automatic human face recognition system was developed, which combines our research work on human face detection, facial feature extraction and human face recognition. Firstly, the human face and the corresponding position of the two eyes in an image are detected and located. The edges of an input facial image are generated and normalized to a specific size based on the inter-distance of the two eyes. This normalized edge facial image and the facial image in the database to be compared are aligned based on the location of the two eyes. The input edge image and

our new modified Hausdorff distance are used to measure the similarity between the input human face and the face in the database. Experiment results show that method can provide a reasonable performance level.

7.2 Future Work

The employment of human face information to strengthen electronic security systems is an important application. In the previous chapters, we have proposed a new modified Hausdorff distance for human face recognition, which can achieve a reasonable performance level. However, there are still other problems to be resolved before a practical and reliable face recognition system is possible. These problems include perspective variations, facial expressions, etc, which will affect the recognition rate. A possible future task for this project is to solve the problem of variations in pose. The first possible way is to use a 3-D to have an approximate representation of a face. A face with perspective variation can be adjusted such that an approximation of its corresponding frontal view is obtained, so the recognition rate can be higher. The use of a 3-D face model to represent a face image relies on the accuracy of face detection and facial feature extraction. When the 3-D face model is aligned on the face image, other 2-D views can be constructed by rotating the 3-D mask and performing texture mapping. If this module is integrated into our approach for face recognition, the effect of perspective variation can be reduced. Another possible way to tackle the perspective variation is to integrate the elastic bunch graph matching approach into our system. This approach can handle large galleries and large variations in pose, and can further increase the matching accuracy.

References

[1] R. Chellappa, C. L. Wilson and S. Sirohey, " Human and Machine Recognition of Faces: A Survey", *Proc. of IEEE*, vol. 83, no. 5, pp. 705-740, May. 1995.

[2] M. A. Grudin, "On Internal Representations in Face Recognition systems", *Pattern Recognition*, vol. 33, no. 7, pp. 1161-1177, 2000.

[3] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. Cogn. Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[4] T. Kondo and Hong Yan, " Automatic Human face Detection and Recognition under non-uniform illumination", *Pattern Recognition*, vol. 32, no. 10, pp. 1707-1718, 1999.

[5] M. Tistarelli and E. Grosso, "Active Face Recognition with a Hybrid approach", *Pattern Recognition Letters*, vol. 18, no. 9, Sep. 1997.

[6] R. P. Wurtz, "Object Recognition Robust Under Translations, Deformations, and Changes in Background", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 769-775, July 1997.

[7] L. Wiskott, J. Fellous, N. Kruger and C. Malsburg, "Face Recognition by Elastic bunch graph matching", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, July 1997.

[8] B. Duc and S. Fischer, "Face Authentication with Gabor Information on Deformable Garphs", *IEEE Trans. on Image Processing*, vol. 8, no. 4, pp. 504-516, April 1999.

[9] C. Kotropoulos, A. Tefas and I. Pitas, "Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions", *Pattern Recognition*, vol. 33, no. 12, pp. 19351947, Dec. 2000.

[10] A. L. Yuille, "Deformable Templates for Face Recognition", *J. Cogn. Neuroscience*, vol. 3, no. 1, pp. 59-70, 1991.

[11] C. L. Huang and C. W. Chen, "Human Facial Feature Extraction for Face Interpretation and Recognition", *Pattern Recognition*, vol. 25, no. 12, pp. 1435-1444, Apr. 1992.

[12] K. M. Lam and H. Yan, "An Analytic-to-Holistic approach for Face Recognition based on a Single Front View", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 673-686, 1998.

[13] R. Brunelli and T. Poggio. "Face Recognition: features versus templates", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, Oct. 1993.

[14] F. Goudail, E. Lange and T. Iwamoto, "Face Recognition System Using Local Autocorrelations and Multiscale Integration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 1024 –1028, Oct. 1996.

[15] H. Liu, M. Wu, G. Jin and G. Cheng, "An Automatic Human Face Recognition system", *Optics and Lasers Eng.*, vol. 30, pp. 305-314, Sep. 1998.

[16] C. Kotropoulos and I. Pitas, "Frontal Face Authentication Using Morphological Elastic Graph Matching", *IEEE Trans. on Image Processing*, vol. 9, no. 4, pp. 555-560, Apr. 2000.

[17] B. Takács, "Comparing Face Images Using the Modified Hausdorff Distance", *Pattern Recognition*, vol. 31, no. 12, pp. 1873-1880, 1998.

[18] M. Lyons and S. Akamatsu, "Coding Facial Expressions with Gabor Wavelets", *Proc. Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.

[19] G.C. Feng and P.C. Yuen, "Recognition of Head-&-Shoulder Face Image Using Virtual Frontal-View Image", *IEEE Trans. On Systems, Man, and Cybernetics-part A: Cybernetics*, vol. 30, no. 6, pp. 871-883, Nov. 2000.

[20] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C V.D. Malsburg, R. P. Wurtz and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", *IEEE Tans. On Computers*, vol. 42, no. 3, pp.300-311, Mar. 1993.

[21] L. D. Harmon and W. F. Hunt, "Automatic Recognition of human face profiles", *Computer Graphics Image Processing* 6, pp. 135-156, 1978.

[22] L. D. Harmon, S. C. Kuo, P. F. Ramig and U. Raudkivi, "Identification of Human Face Profiles by Computer", *Pattern Recognition*, vol. 10, pp. 301-312, 1978.

[23] C. J. Wu and J. S. Huang, "Human Face Profile Recognition by Computer", *Pattern Recognition*, vol. 23, no. 3/4, pp. 255-259, 1990.

[24] Z. Liposcak and S. Loncaric, "Face Recognition from profiles using Morphological Operation", *Proc. Int. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pp. 47-52, 1999.

[25] J. H. Lai, P. C. Yuen and G. C. Feng, "Face Recognition using Holistic Fourier Invariant Features", *Pattern Recognition*, vol. 34, no.1, pp. 95-109, 2001.

[26] J. Lyons, J. Budynek and S. Akamatsu, "Automatic Classification of Single Facial Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, Dec. 1999.

[27] P. Kalocsai, H. Neven and J. Steffens, "Statistical Analysis of Gabor-filter Representation", *Proc. of third IEEE Int'l conf. on Automatic Face and Gesture Recognition*, pp. 360-365, 1998.

[28] Eleftheriadis and A. Jacquin, "Automatic Face Location Detection for Model-Assisted rate control in H.261-compatible coding of Video", *Signal Processing: Image Communication*, vol. 7, pp. 435-455, Nov. 1995.

[29] Eleftheriadis and A. Jacquin, "Automatic Face Location Detection and Tracking for model assisted coding of video teleconferencing sequences at low bit rates", *Signal Processing: Image Communication*, vol. 7, pp. 231-248, Sep. 1995.

[30] G. Yang and S. Huang, "Human Face Detection in a complex background", *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, Jan. 1994.

[31] J. Miao, B. Yin, K. Wang, L. Shen and X. Chen, "A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template", *Pattern Recognition,* vol. 32, no. 7, pp. 1237-1248, 1999.

[32] K. M. Lam, "A Fast Approach for Detecting Human Faces in A Complex Background", *Proc. of the 1998 IEEE. Int'l Sym. on Circuits and Systems*, vol. 4, pp. 85-88, 1998.

[33] P. Maragos, "Tutorial on Advances in Morphological Image Processing and Analysis", *Optical Engineering*, vol. 26, no. 7, pp. 623-632, 1987.

[34] D.E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", *Addison-Wesley*, 1999.

[35] M. Turk and A. Pentland, "Eigenfaces for recognition", *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71-86, 1991.

[36] K. W. Wong and K. M. Lam, "A Reliable approach for Human Face Detection using Genetic Algorithm", *Proc. of the 1999 IEEE Int'l Sym. on Circuits and Systems*, vol. 4, pp. 499-502, 1999.

[37] K. K. Sung and T. Poggio, "Example-Based Leaning for View-Based Human Face Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[38] J. Wang and T. Tan, "A new Face Detection method based on shape information", *Pattern Recognition Letters*, vol. 21, no. 6, pp. 463-471, Jun. 2000.

[39] Y. Yokoo and M. Hagiwara, "Human Faces Detection Method using Genetic Algorithm", *Proc. of IEEE Int. Conf. on Evolutionary Computation*, pp. 113-118, 1996.

[40] C. Lin, K. Fan, "Triangle-Based Approach to the Detection of Human Face", *Pattern Recognition*, vol. 34, no. 6, pp. 1271-1284, 2000.

[41] Y. Wang and B. Yuan, "Segmentation method for Face Detection in complex background", *Electronic Letters*, vol. 36, no. 3, Feb. 2000.

[42] C. C. Han, H. Y. Mark Liao, G. J. Yu and L. H. Chen, "Fast Face Detection via Morphology-base processing", *Pattern Recognition*, vol. 33, no. 10, pp. 1701-1712, 2000.

[43] Y. Dai and Y.Nakano, "Face Texture Model Based on SGLD and its application in Face Detection in a Color Scene", *Pattern Recognition*, vol. 29, no. 6, pp.1007-1016 1996.

[44] C. H. Lee, J. S. Kim and K. H. Park, "Automatic Human Face Location in a Complex Background using Motion and Color information", *Pattern Recognition*, vol. 29, no. 11, pp. 1877-1889, 1996.

[45] G. Wei and Ishwar K. Sethi, "Face Detection for image annotation", *Pattern Recognition Letters*, vol. 20, pp. 1313-1321, Nov. 1999.

[46] A. M. Alattar and S. A. Rajala, "Estimating Head's Measurements from Front-View Head and Shoulders Images", *Proc. of the 1999 IEEE Int. Sym. on Circuits and Systems*, vol. 4, pp. 114-117, 1999.

[47] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Processing: Image Communication*, vol. 12, no. 3, pp. 263-281, June 1998.

[48] Q.B. Sun, W.M. Huang, and J.K. Wu, "Face detection based on color and local symmetry information," *Proc. of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 130-135, 1998.

[49] K. Sobottka, and I. Pitas, "Segmentation and tracking of faces in color images," *Proc. of the Second International Conference on Automatic Face and Gesture Recognition*, pp. 236-241, 1996.

[50] C.H. Lin and J. L Wu, "Automatic Facial Feature Extraction by Genetic Algorithms", *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 834-845, June 1999.

[51] K. M Lam and Y. L. Lin, "An Efficient Approach for Facial Feature Detection", *IEEE of 14th Int. Conf. on Signal Processing Proceedings*, vol. 2, pp. 1100-1103, 1998.

[52] S. H. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern and Y. T. Liu, "Facial Feature Detection using Geometrical Face Model: an Efficient Approach", *Pattern Recognition*, vol. 31, no. 3, pp. 273-282, 1998.

[53] W. Haiyuan, C.Qian and M. Yachida, "Facial Feature Extraction and Face Verification", *Proc. of the 13th Int. Conf. on Pattern Recognition*, vol. 3, pp. 484-488, 1996.

[54] F. C. Wu, T. J. Yang and M. Ouhyoung, "Automatic Feature Extraction and Face Synthesis in Facial Image Coding", *Sixth Pacific Conference on Computer Graphics and Applications*, pp. 218-219,1998.

[55] A. M. Alattar, S.A. Rajala, "Facial Features Localization in Front View Head- and -Shoulders Images", *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3557-3560, 1999.

[56] G. B. De Natale, D. Giusto and F. Maccioni, "A Symmetry-Based Approach to Facial Features Extraction", *Proc. Of the 13th Int. Conf. on Digital Signal Processing*, vol. 2, pp. 521-525, 1997.

[57] F. Mu, H. Li and R. Forchheimer, "Automatic Extraction of Human Facial Features", *Signal Processing: Image Communication*, vol. 8, no. 4, pp. 309-326, May 1996.

[58] D. Pramadihanto, Y. Iwai and M. Yachida, "A Flexible Feature Matching for Automatic Face and Facial Feature Points Detection", *Proc. of the 14th Int. Conf. on Pattern Recognition*, vol. 1, pp. 92-95, 1998.

[59] S. J. McKenna, S. Gong, R.P. Wurtz, J. Tanner and D. Banin, "Tracking Facial Feature Points with Gabor Wavelet and Shape Models", *Proc. of the 1st Int. Conf. on Audio- and Video-based Biometric Person Authentication*, 1997.

[60] K. M. Lam and Hong Yan, "An Improved Method for Locating and Extracting the Eye in Human Face Images", *Proc. of the 13th Int. Conf. on Pattern Recognition*, vol. 3, pp. 411-415, 1996.

[61] K. M. Lam and Hong Yan, "Locating and Extracting the Eye in Human Face Images", *Pattern Recognition*, vol.29, no. 5, pp.771-779, 1996.

[62] A. R. Mirhosseini, H. Yan and K.M. Lam, "Adaptive Deformable Model for Mouth Boundary Detection", *Optical Engineering*, no. 27, vol 3, pp. 869-875, March 1998.

[63] A. R. Mirhosseini, C. Chen, K.M. Lam and H. Yan, "A Hierarchical and Adaptive Deformable Model for Mouth Boundary Detection", *Proc. Int. Conf. on Image Processing*, vol. 2, pp. 756-759, 1997.

[64] S. Bernogger, L. Yin, A. Basu and A.Pinz, "Eye Tracking and Animation for MPEG-4 Coding", *Proc. of the 14th Int. Conf. on Pattern Recognition*, vol. 2, pp. 1281-1284, 1998.

[65] G.C. Feng and P. C. Yuen, "Variance Projection Function and its application to Eye Detection for Human Face Recognition", *Pattern Recognition Letters*, vol. 19, no. 9, pp. 899-906, July 1998.

[66] T. Hamada, K. Kato and K. Kawakami, "Extracting Facial Features as in infants", *Pattern Recognition Letters*, vol. 21, pp. 407-412, 2000.

[67] Alan L. Yuille, Peter W. Hallinan and David S. Cohen, "Feature Extraction from Faces Using Deformable Templates", *Int. J. Of Computer Vision*, vol. 8, no. 2, pp. 99-111, 1992.

[68] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models", *Int. J. Computer Vision 9*, pp. 321-331, 1988.

[69] K. M. Lam and H. Yan, "Fast Algorithm for Locating Head Boundaries", *J. of Electronic Imaging*, vol. 3, no. 4, pp.351-359, Oct. 1994.

[70] K. M. Lam. and H. Yan, "Fast Greedy Algorithm for Active Contours", *Electronic Letter*, vol.30, no.5, pp.21-22, Jan. 1994.

[71] W. P. Choi, K. M. Lam and W. C. Siu, "An Adaptive Active Contour Model for Highly Irregular Boundaries", *Pattern Recognition*, vol. 34, no. 2, pp. 323-33, 2001.

[72] P.J. Saupe, and M. P. Yunker, "Fractals for the classrooms", *Strategic Activities Vol. One*, Spring Verlag.

[73] N. Sarkar, and B.B. Chaudhuri, "An Efficient Differential Box-Counting approach to Compute Fractal Dimension of image", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 24, no. 1, pp.115-120, 1994.

[74] J. Feng, W. C. Lin, and C. T. Chen, "Fractional Box-Counting Approach to Fractal Dimension Estimation", *Proc. of the 13th Int. Conf. Pattern Recognition*, no. 2, pp. 854-858, 1996.

[75] K.L. Chan, "Fractal Based Texture Analysis", *Singapore ICCS/ISITA '92. Comm. on the Move*, no. 1, pp. 102-106, 1990.

[76] T. Sato, M. Matsuoka, and H. Takayasu, "Fractal image Analysis of Natural Scenes and Medical images", *Fractals, World Scientific Publishing Company*, vol. 4, no. 4, pp. 463-468, 1996.

[77] Y. Y. Tang, and Y. Tao, "Feature Extraction by Fractal Dimension", *Proc. of the 5th Int. Conf. on Document Analysis and Recognition*, pp. 217-220, 1999.

[78] C. Gonzalez, and E. Woods, "Digital Image Processing", *Addison Wesley*, Sep. 1993.

[79] D. P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff distance", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.

[80] D. G. Sim, O.K. Kwon and R. H. Park, "Object Matching Algorithms Using Robust Hausdorff Distance Measures", *IEEE Trans. on Image Processing*, vol. 8, no. 3, pp. 425-429, Mar. 1999.

[81] M. P. Dubuisson and A. K. Jain, "A Modified Hausdorff distance for Object Matching", *Proc. of 12th Int Conf. Pattern Recognition*, pp. 566-568, Jerusalem, Israel, Oct. 1994.

[82] P.E. Danielsson, "Euclidean Distance Mapping", *Computer Graphics and Image Processing*, vol. 14, pp. 227-248, 1980.