

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen



The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

FACIAL FEATURE REPRESENTATION AND RECOGNITION

by
Chao-fa Chuang

Facial expression provides an important behavioral measure for studies of emotion, cognitive processes, and social interaction. Facial expression representation and recognition have become a promising research area during recent years. Its applications include human-computer interfaces, human emotion analysis, and medical care and cure.

In this dissertation, the fundamental techniques will be first reviewed, and the developments of the novel algorithms and theorems will be presented later. The objective of the proposed algorithm is to provide a reliable, fast, and integrated procedure to recognize either seven prototypical, emotion-specified expressions (e.g., happy, neutral, angry, disgust, fear, sad, and surprise in JAFFE database) or the action units in Cohn-Kanade AU-coded facial expression image database.

A new application area developed by the Infant COPE project is the recognition of neonatal facial expressions of pain (e.g., air puff, cry, friction, pain, and rest in Infant COPE database). It has been reported in medical literature that health care professionals have difficulty in distinguishing newborn's facial expressions of pain from facial reactions of other stimuli. Since pain is a major indicator of medical problems and the quality of patient care depends on the quality of pain management, it is vital that the methods to be developed should accurately distinguish an infant's signal of pain from a host of minor distress signal. The evaluation protocol used in the Infant COPE project considers two conditions: person-dependent and person-independent. The person-

dependent means that some data of a subject are used for training and other data of the subject for testing. The person-independent means that the data of all subjects except one are used for training and this left-out one subject is used for testing. In this dissertation, both evaluation protocols are experimented.

The Infant COPE research of neonatal pain classification is a first attempt at applying the state-of-the-art face recognition technologies to actual medical problems. The objective of Infant COPE project is to bypass these observational problems by developing a machine classification system to diagnose neonatal facial expressions of pain. Since assessment of pain by machine is based on pixel states, a machine classification system of pain will remain objective and will exploit the full spectrum of information available in a neonate's facial expressions. Furthermore, it will be capable of monitoring neonate's facial expressions when he/she is left unattended. Experimental results using the Infant COPE database and evaluation protocols indicate that the application of face classification techniques in pain assessment and management is a promising area of investigation.

One of the challenging problems for building an automatic facial expression recognition system is how to automatically locate the principal facial parts since most existing algorithms capture the necessary face parts by cropping images manually. In this dissertation, two systems are developed to detect facial features, especially for eyes. The purpose is to develop a fast and reliable system to detect facial features automatically and correctly. By combining the proposed facial feature detection, the facial expression and neonatal pain recognition systems can be robust and efficient.

FACIAL FEATURE REPRESENTATION AND RECOGNITION

by

Chao-fa Chuang

A Dissertation

Submitted to the Faculty of

New Jersey Institute of Technology

**in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science**

Department of Computer Science

January 2006

Copyright © 2006 by Chao-fa Chuang

ALL RIGHTS RESERVED

APPROVAL PAGE

FACIAL FEATURE REPRESENTATION AND RECOGNITION

Chao-fa Chuang

Dr. Frank Y. Shih, Dissertation Advisor Professor of Computer Science, NJIT	Date
--	------

Dr. James A. M. McHugh, Committee Member Professor of Computer Science, NJIT	Date
---	------

Dr. Alexandros Gerbessiotis, Committee Member Associate Professor of Computer Science, NJIT	Date
--	------

Dr. Cristian M. Borcea, Committee Member Assistant Professor of Computer Science, NJIT	Date
---	------

Dr. Qun Ma, Committee Member Assistant Professor of Computer Science, NJIT	Date
---	------

Dr. Sheryl Brahnam, Committee Member Assistant Professor of Computer Information Systems, MSU	Date
--	------

BIOGRAPHICAL SKETCH

Author: Chao-fa Chuang
Degree: Doctor of Philosophy
Date: January 2006

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
New Jersey Institute of Technology, Newark, NJ, 2006
- Master of Science in Computer Science,
New Jersey Institute of Technology, Newark, NJ, 2001
- Master of Business Administration,
National Chung Hsing University, Taipei, Taiwan, 1991
- Bachelor of Business Administration,
National Chung Hsing University, Taipei, Taiwan, 1989

Major: Computer Science

Presentations and Publications:

Conference Papers:

Sheryl Brahnham, Chao-fa Chuang, Frank Y. Shih and Melinda R. Slack,
“SVM classification of neonatal facial images of pain,”
WILF 2005 6th International Workshop on Fuzzy Logic and Applications, Crema,
Italy, September 2005.

Frank Y. Shih, Shouxian Cheng and Chao-fa. Chuang,
“Face and eye detection using PCA and SVM in color images,”
7th Joint Conferences on Information Sciences, Cary, NC, pp. 731-734,
September 2003.

Journal Articles:

- Sheryl Brahnam, Chao-fa Chuang, Randal S. Sexton and Frank Y. Shih,
“Machine assessment of neonatal facial expressions of acute pain,”
Special Issue on “Decision Support in Medicine” in Decision Support Systems,
accepted in 2005.
- Sheryl Brahnam, Chao-fa Chuang, Frank Y. Shih and Melinda R. Slack,
“Machine recognition and representation of neonatal facial displays of acute
pain,”
Artificial Intelligence in Medicine, in press 2005.
- Frank Y. Shih, Chao-fa Chuang, and Vijayalakshmi Gaddipatti,
“A modified regulated morphological corner detector,”
Pattern Recognition Letters, vol. 26, no. 7, pp. 931-937, May 2005.
- Frank Y. Shih and Chao-fa Chuang,
“Automatic extraction of head and face boundaries and facial features,”
Information Sciences, vol. 158, pp. 117-130, January 2004.
- Chao-fa Chuang and Frank Y. Shih,
“Recognizing facial action units using independent component analysis and
support vector machine,”
submitted to *Pattern Recognition*, 2005.

This dissertation is dedicated to my beloved family

謹以此論文獻給我敬愛的雙親 莊嘉文先生和莊李淑敏女士，摯愛的妻子 林淑芬及
二個可愛的孩子 莊宏偉和莊育寧

ACKNOWLEDGMENT

I would like to express my deepest appreciation to Dr. Frank Y. Shih, who not only served as my research supervisor, providing valuable and countless resources, insight, and intuition, but also constantly gave me support, encouragement, and reassurance.

Special thanks are given to Dr. James A. M. McHugh, Dr. Alexandros Gerbessiotis, Dr. Cristian M. Borcea, Dr. Qun Ma and Dr. Sheryl Brahnam for actively participating in my committee.

Important thanks go to Dr. Michael J. Lyons for providing JAFFE database, Dr. Jeffrey F. Cohn for furnishing Cohn-Kanade AU-coded facial expression image database, Dr. Sheryl Brahnam for supplying Infant COPE database, and National Institute of Standards and Technology for offering FERET database.

Many of my fellow graduate students, Yi-Ta Wu, Shouxian Cheng, Peichung Shih, Yu Wang, Yan-Yu Fu, and Ming Qu, in the Computer Vision Laboratory are deserving of recognition for their support.

I also wish to thank my beloved wife, Shu-fen, and two lovely kids, Hung-Wei and Amanda, for their strongest support. I would like give special thanks to my parents; the love they give is unlimited.

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION.....	1
1.1 Objectives.....	1
1.2 Organization of This Dissertation.....	3
2 FACIAL EXPRESSION RECOGNITION-EMOTION SPECIFIED CASE.....	6
2.1 Introduction	6
2.2 The JAFFE Database and the Proposed Methods.....	8
2.2.1 Discrete Wavelet Transform (DWT).....	9
2.2.2 PCA and LDA	11
2.2.3 RBF and SVM	14
2.3 Experimental Procedure	16
2.3.1 Pre-processing	16
2.3.2 Feature Extraction	16
2.3.2 Expression Classification.....	16
2.4 Experimental Results	17
2.5 Conclusions	21
3 FACIAL EXPRESSION RECOGNITION-ACTION UNITS CASE.....	22
3.1 Introduction.....	22
3.2 Facial Action Coding System (FACS) and Face Expression Database	24
3.2.1 Facial Action Coding System (FACS)	24
3.2.2 Cohn-Kanade AU-Coded Face Expression Image Database.....	26

TABLE OF CONTENTS
(Continued)

Chapter	Page
3.3 Experimental Procedure	27
3.3.1 Independent Component Analysis	27
3.3.2 Support Vector Machine	30
3.3.3 Experimental Procedure	31
3.4 Experimental Results	32
3.5 Conclusions	37
4 NEONATAL PAIN CLASSIFICATION	38
4.1 Introduction.....	39
4.2 Study Design	42
4.2.1 Subjects	43
4.2.2 Apparatus	43
4.2.3 Procedure	43
4.3 Basic Concepts of PCA, LDA, and SVM	44
4.3.1 Principal Component Analysis	45
4.3.2 Linear Discriminant Analysis	49
4.3.3 Support Vector Machine	49
4.4 Experimental Procedure	56
4.5 Experimental Results	56
4.5.1 Person-Dependent Case	58
4.5.2 Person-Independent Case	61
4.6 Conclusions	66

TABLE OF CONTENTS
(Continued)

Chapter	Page
5 FACIAL FEATURE DETECTION.....	69
5.1 Automatic Extraction of Face and Facial Features	69
5.1.1 Introduction	70
5.1.2 Methodology	72
5.1.3 Finding Facial Features Based on Geometric Face Model	77
5.1.4 Experimental Results	81
5.1.5 Conclusions	81
5.2 Facial Feature Detection (Eyes Detection)	84
5.2.1 Introduction	85
5.2.2 Experimental Procedure	86
5.2.3 Conclusions	89
6 SUMMARY AND FUTURE RESEARCH	91
6.1 Summary	91
6.2 Future Research	93
6.2.1 Future Research on Pattern Recognition	93
6.2.2 Future Research on Neonatal Pain Classification	94
REFERENCES	95

LIST OF TABLES

Table		Page
2.1	Performance Comparisons in JAFFE Database	19
2.2	Performance Comparisons of Using Different Kernels of SVM	19
2.3	Comparisons of PCA, LDA, 2D-PCA, ICA, and 2D-LDA using SVM	19
2.4	Comparisons of PCA, LDA, 2D-PCA, ICA, and 2D-LDA using RBF.....	20
2.5	Confusion Matrix Using Cross-Validation Strategy	20
2.6	Confusion Matrix Using Leave-One-Out Strategy	20
3.1	The Upper Face Action Units	25
3.2	Miscellaneous Actions	25
3.3	The Lower Face Action Units	26
3.4	Recognition Rate for Upper Part of Face	34
3.5	Recognition Rate for Lower Part of Face	35
3.6	Recognition Rate for Whole Face	35
3.7	Recognition Rate Between Genders on Action Units	36
3.8	Performance Comparison of Different Systems	36
4.1	Pain versus Rest	59
4.2	Pain versus Cry	59
4.3	Pain versus Air Puff	59
4.4	Pain versus Fraction	60
4.5	Pain versus Non-pain	60
4.6	Average Method Classification Scores and Individual Experiment Scores....	64

LIST OF TABLES
(Continued)

Table		Page
4.7	95% Confidence Intervals Using t Distribution.....	65
4.8	Comparison of Classification Rates Using Person-dependent & -independent	65

LIST OF FIGURES

Figure		Page
2.1	Samples of two expressers containing 7 different facial expressions.....	9
2.2	Sub-band decomposition of an $N \times M$ image	10
2.3	First-level decomposition	10
2.4	Two-level decomposition of a happy image	11
2.5	The architecture of a radial basis function neural network	15
2.6	The experimental procedure	17
3.1	The experimental procedure.....	31
4.1	Examples of the five facial expressions in the Infant COPE dataset	45
4.2	The first ten eigenfaces of the images in the Infant COPE Database	47
4.3	Illustration of the linear combination of eigenfaces	47
4.4	The experimental procedure	57
5.1	The scheme diagram of double-threshold method	73
5.2	The outputs of Wechsler and Kidode's method	74
5.3	Edge detection masks	75
5.4	An image of facial feature candidates	76
5.5	The examples of repairing	77
5.6	Geometric face model	78
5.7	The procedure of finding facial features	79
5.8	Examples of geometric face model	80
5.9	The experimental results of double-threshold method	82

LIST OF FIGURES
(Continued)

Figure	Page
5.10 Partial experimental results	83
5.11 The procedure for detecting eyes	88
5.12 The partial experimental results	90

CHAPTER 1

INTRODUCTION

1.1 Objectives

Facial expression provides an important behavioral measure for studies of emotion, cognitive processes, and social interaction. Facial expression representation and recognition have become a promising research area during recent years. Its applications include human-computer interfaces, human emotion analysis, and medical care and cure.

A variety of systems have been developed to perform facial expression recognition. Most systems conduct two stages: feature extraction representation and expression classification. For feature extraction representation, Gabor filter, principal component analysis, and independent component analysis are used. For expression classification, linear discriminant analysis, support vector machine, two-layer perceptron, and Hidden Markov Model are applied. In this dissertation, the fundamental techniques will be first reviewed, and the developments of the novel algorithms and theorems will be presented later. The objective of the proposed algorithm is to provide a reliable, fast, and integrated procedure to recognize either seven prototypical, emotion-specified expressions (such as happy, neutral, angry, disgust, fear, sad, and surprise in JAFFE database) or action units in Cohn-Kanade AU-coded face expression image database.

A new application area developed by the Infant COPE project is the recognition of neonatal facial expressions of pain (e.g., air puff, cry, friction, pain, and rest in Infant COPE database). It has been reported in medical literature that health care professionals have difficulty in distinguishing newborn's facial expressions of pain from facial reactions of other stimuli. Since pain is a major indicator of medical problems and the

quality of patient care depends on the quality of pain management, it is vital that the methods to be developed should accurately distinguish an infant's signal of pain from a host of minor distress signal. In this dissertation, neonate pain classification problems is tackled by applying three different state-of-the-art face classification techniques (e.g., PCA, LDA, and SVM) to the task of distinguishing a newborn's facial expressions of pain. The evaluation protocol used in the Infant COPE project considers two conditions: person-dependent and person-independent. The person-dependent means that some data of a subject are used for training and other data of the subject for testing. The person-independent means that the data of all subjects except one are used for training and this left-out one subject for testing. In this dissertation, both evaluation protocols are experimented.

The Infant COPE research of neonatal pain classification is a first attempt at applying the state-of-the-art face recognition technologies to actual medical problems. Medical applications of face recognition technologies have been suggested but not tried with actual medical data. The objective of Infant COPE project is to bypass these observational problems by developing a machine classification system to diagnose neonatal facial expressions of pain. Since assessment of pain by machine is based on pixel states, a machine classification system of pain will remain objective and will exploit the full spectrum of information available in a neonate's facial expressions. Furthermore, it will be capable of monitoring neonate's facial expressions when he/she is left unattended. Experimental results using the Infant COPE database and evaluation protocols indicate that the application of face classification techniques in pain assessment and management is a promising area of investigation.

A robust automatic facial expression recognition system includes face detection, facial feature extraction and representation, and facial expression recognition. One of the challenging problems for building an automatic facial expression recognition system is how to automatically locate the principal facial parts since most existing algorithms capture the necessary face parts by cropping images manually. In this dissertation, two systems are developed to detect facial features, especially for eyes. The objective is to develop a fast and reliable system to detect facial features automatically and correctly. By combining the proposed facial feature detection system, the facial expression and neonatal pain recognition systems can be robust and efficient.

1.2 Organization of This Dissertation

The organization of this dissertation will be described as follows.

Chapter 1 Introduction

Chapter 2 Facial Expression Recognition-emotion specified case

Chapter 3 Facial Expression Recognition-Action Units case

Chapter 4 Neonatal Pain Classification

Chapter 5 Facial Features Detection

Chapter 6 Conclusions and Further Research

The brief statements of each chapter are given as follows.

In Chapter 2, various feature representation and classification schemes are investigated to distinguish seven different facial expressions such as happy, neutral, angry, disgust, sad, fear, and surprise, in homogeneous database, JAFFE database. Experimental results show that the method of using 2D-LDA (Linear Discriminant Analysis) and SVM

(Support Vector Machine) outperforms others. The recognition rate of the method is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. It takes only 0.0357 second for processing one input image of size 256×256 . The proposed system is fast, reliable, and can be applied to real time applications.

In Chapter 3, more complicated and heterogeneous facial database, Cohn-Kanade Facial Expression Database, is tackled. Instead of recognizing prototypical, emotional-specified expressions, action units which describe the subtle changes of face are classified. A technique using ICA (Independent Component Analysis) as feature extraction and representation method and SVM (Support Vector Machine) as classification technique is proposed. Experimental results are promising and comparable to published systems. The recognition rate of the proposed method is 97.06% for recognizing the upper part of face, 97.13% for lower part of face, and 100% for the whole face. The proposed system is fast and can be applied to real-time applications. It takes only 1.8 ms for processing one test image of size 40×90 . It is interesting to explore the gender effect on recognizing the action units. From the experimental results of Chapter 3, the male can express the action units more accurately than female do.

In Chapter 4, a machine recognition and assessment system of neonatal expressions of pain is developed to assist clinicians in diagnosing pain. The experiments reported in this chapter use the Infant COPE database. This database containing 204 facial expressions of 26 neonates (age 18-36 hours) was photographed experiencing the acute pain of a heel lance and three non-pain stressors including transport from one crib to another (a disturbance that can provoke crying that is not in response to pain), an air stimulus on the nose, and friction on the external lateral surface of the heel.. Three

algorithms were evaluated: PCA, LDA, and SVM. Experimental results indicate a high potential for developing a decision support system for diagnosing neonatal pain from images of neonatal facial displays.

In Chapter 5, a novel approach for the extraction of human head, face and facial features is presented. In the double-threshold method, the high-thresholded image is used to trace head boundary and the low-thresholded image is used to scan face boundary. After obtaining facial features candidates and eliminate noises, apply x - and y -projections is applied to extract facial features such as eyes, nostrils and mouth. Since low contrast of chin occurs in some face images, its boundary cannot be completely detected. An elliptic model is used to repair it. Because of noises or clustered facial features candidates, a geometric face model to locate facial features and an elliptic model to trace face boundary are adopted. FERET face database is tested and experimental results show that the proposed algorithm can perform the extraction of human head, face and facial features successfully.

In Chapter 6, summary of this dissertation and the further works are provided. The direction of the future research is to increase the recognition rate and decrease computational time simultaneously.

CHAPTER 2

FACIAL EXPRESSION RECOGNITION-EMOTION SPECIFIED CASE

Facial expression provides an important behavioral measure for studies of emotion, cognitive processes, and social interaction. Facial expression recognition has become a promising research area during recent years. Its applications include human-computer interfaces, human emotion analysis, and medical care and cure. In this chapter, various feature representation and classification schemes are investigated to recognize seven prototypical and emotion-specified facial expressions such as happy, neutral, angry, disgust, sad, fear, and surprise, in JAFFE database.

2.1 Introduction

Facial expression plays a principal role in human interaction and communication since it contains critical and necessary information regarding emotion. The task of automatically recognizing different facial expressions in human-computer environment is significant and challenging. A variety of systems have been developed to perform facial expression recognition [2, 12, 13, 21, 56, 73, 80, 97, 98]. These systems possess some common characteristics. First, they classify facial expressions using adult facial expression databases. For instances, the authors in [12, 21, 56, 73, 97] used JAFFE database to recognize seven main facial expressions: happy, neutral, angry, disgust, fear, sad, and surprise. Chen and Huang [13] used AR database to classify three facial expressions: neutral, smile, and angry. Second, most systems conduct two stages: feature extraction and expression classification. For feature extraction, Gabor filter [12, 56, 97], PCA

(Principal Component Analysis) [13, 21, 56], and ICA (Independent Component Analysis) [12] are used. For expression classification, LDA (Linear Discriminant Analysis) [1, 21, 56], SVM (Support Vector Machine) [12], two-layer perceptron [97], and HMM (Hidden Markov Model) [98] are used.

Recently, facial expression recognition is applied to medical treatment of patients. Dai *et al.* [19] proposed to monitor patients on bed by utilizing the facial expression recognition to detect the status of patients. Gagliardi *et al.* [28] investigated the facial expression ability for individuals with Williams syndrome. Sprengelmeyer *et al.* [77] explored facial expression recognition of emotion for people with medicated and unmedicated Parkinson's disease.

Since JAFFE database is benchmark and commonly used in measuring the performance of facial expression recognition systems, the proposed system is applied to this database and comparisons with other systems [12, 21, 56, 73, 97] are performed. Lyons *et al.* [56] made use of Gabor filter at different scales and orientations, and applied 34 fiducial points for each convolved image to construct the feature vector for representing each facial image. After that, PCA is applied to reduce the dimensionality of feature vectors and LDA is used to identify seven different facial expressions. Their recognition rate is 92% for JAFFE database. Zhang *et al.* [97] adopted Gabor wavelet coefficients and geometric positions to construct the feature vector for each image and applied two-layer perceptron to distinguish seven different facial expressions. Their recognition rate is 90.1%.

Buciu *et al.* [12] tested different feature extraction methods such as Gabor filter and ICA combined with SVM using three different kernels: linear, polynomial, and radial

basis function, to check which combination can produce the best result. From their experiments, the best recognition rate is 90.34% by using Gabor wavelet at high frequencies combined with the polynomial kernel SVM at degree 2. Dubuisson *et al.* [21] combined the sorted PCA as feature extractor with LDA as the classifier to recognize facial expressions. Their recognition rate is 87.6%. Shinohara and Otsu [73] used Higher-order Local Auto-Correlation (HLAC) features and LDA to test the performance. Unfortunately, their system is unreliable since the correct rate is only 69.4%.

In this chapter, different feature representation and expression classification methods are explored for achieving high accuracy and efficiency. The rest of this chapter is organized as follows. Section 2.2 describes the JAFFE database and the proposed methods. Section 2.3 presents the experimental procedure. Section 2.4 provides the experimental results and performance comparisons. Finally, conclusions are given in Section 2.5.

2.2 The JAFFE Database and the Proposed Methods

The image database used in our experiments is the JAFFE facial expression database [56]. This dataset is used as the benchmark database for researchers in [12, 21, 56, 73, 97]. The database contains ten Japanese females. There are seven different facial expressions such as neutral, happy, angry, disgust, fear, sad, and surprise. Each female has 2 - 4 examples for each expression. Totally, there are 213 grayscale facial expression images in this database. Each image is of size 256×256 . Figure 2.1 shows two expressers containing seven different facial expressions from the JAFFE database. In the following, the methods used in the proposed facial expression recognition system will be briefly

described.

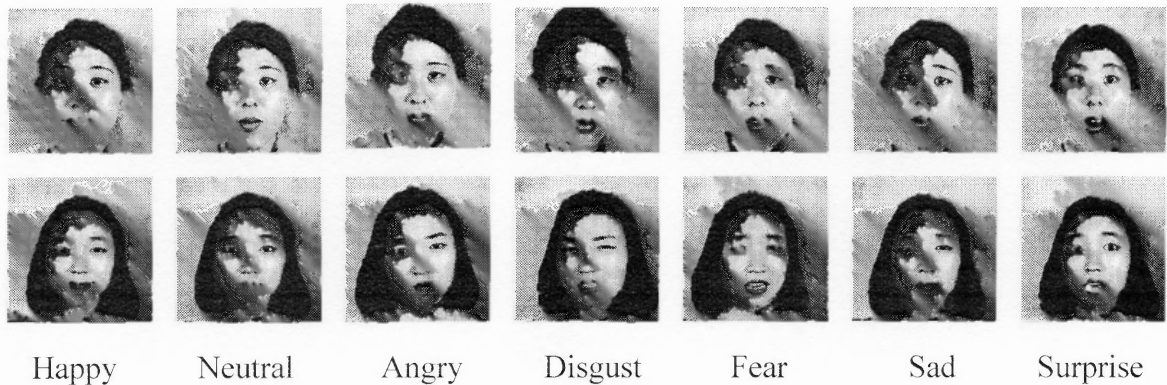


Figure 2.1 Samples of two expressers containing seven different facial expressions.

2.2.1 Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT) [14, 62, 96] is a suitable tool for extracting image features because it allows the analysis of images on various levels of resolution. Typically, low-pass and high-pass filters are used for decomposing the original image. The low-pass filter results in an approximation image and the high-pass filter generates a detail image. The approximation image can be further split into a deeper level of approximation and detail repeatedly according to different applications. Suppose that the size of an input image is $N \times M$. At the first filtering in the horizontal direction of down-sampling, the size of images will be reduced to $N \times (M/2)$. After that, further filtering and down-sampling in the vertical direction, four subimages are obtained, each being of size $(N/2) \times (M/2)$. Figure 2.2 shows the sub-band decomposition of an $N \times M$ image, where H and L respectively denote high-pass and low-pass filters, and $\downarrow 2$ denotes down-sampling by a factor of 2.

The outputs of these filters are given by Equations (2.1) and (2.2).

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} l[n-2p]a_j[n] \quad (2.1)$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] \quad (2.2)$$

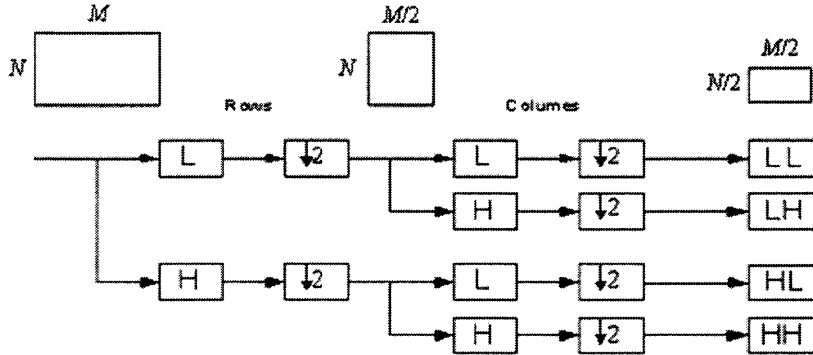


Figure 2.2 Sub-band decomposition of an $N \times M$ image.

where $l[n]$ and $h[n]$ are coefficients of low-pass and high-pass filters, respectively. Accordingly, four images denoted as LL, HL, LH and HH can be obtained. The LL image is generated by two continuous low-pass filters; HL is filtered by a high-pass filter first and a low-pass filter later; LH is created using a low-pass filter followed by a high-pass filter; HH is generated by two successive high-pass filters. Figure 2.3 illustrates first-level decomposition.

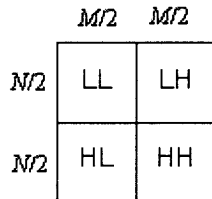


Figure 2.3 First-level decomposition.

Every sub-image can be decomposed further into smaller images by repeating the above procedure. The main feature of DWT is the multi-scale representation of a function. By using the wavelets, a given image can be analyzed at various levels of resolution. Since the LL part contains most important information and discards the effect of noises and irrelevant parts, the LL part is adopted for further analysis in this study. Features from the LL part of the second-level decomposition are extracted and analyzed. The reasons are the LL part keeps the necessary information and the dimensionality of the image is reduced sufficiently for the computation in the next stage. Figure 2.4 shows the two levels of DWT for a happy image from JAFFE database.

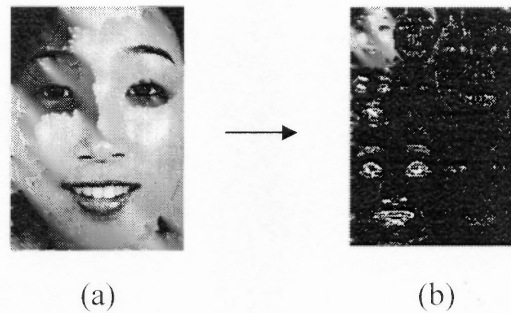


Figure 2.4 Two-level decomposition of a happy image (a) the original image and (b) the decomposed subimages.

2.2.2 PCA and 2D-LDA

Principal Component Analysis (PCA) is widely used for feature representation. In this chapter, PCA is used as a base-line method to compare with the *2D Linear Discriminant Analysis* (2D-LDA) [51]. The central idea behind PCA is to find an orthonormal set of axes pointing in the direction of maximum covariance in the data [46, 82, 83]. In terms of facial images, the idea is to find the orthonormal basis vectors or the eigenvectors of the covariance matrix of a set of images, with each image being treated as a single point in a

high dimensional space [47].

Since each image contributes to each of the eigenvectors which resemble ghostlike faces when displayed, it is referred to as *holon* [18] or *eigenface* [82, 83], and the new coordinates system is referred to as the *face space*. Individual images can be projected onto the face space and represented exactly as weighted combinations of the eigenface components. The resulting vector of weights that describe each face can be used in data compression and face classification. Data compression relies on the fact that the eigenfaces are ordered, with each one accounting for a different amount of variation among the faces. Compression is achieved by reconstructing images using only those few eigenfaces that account for the most variability [74]. It results in dramatic reduction of dimensionality. Classification is performed by projecting a new image onto the face space and comparing the resulting weight vector with the weight vectors of a given class.

Recently, Li and Yuan [51] introduced a new approach for image feature extraction and representation called *2D-LDA*. The difference between LDA and 2D-LDA is that in LDA, the image vector is constructed to compute the between-class and within-class scatter matrices, but in 2D-LDA, the original image matrix is adopted to compute the two matrices. They claimed that the 2D-LDA can achieve better results than other feature extraction methods, such as PCA, LDA, and 2D-PCA [94]. The idea of 2D-LDA is described below.

Suppose that M training samples are used belonging to L classes (L_1, L_2, \dots, L_L). The training samples in each class are denoted as N_i ($i=1, 2, \dots, L$). The size of each training image \mathbf{A}_j ($j = 1, 2, \dots, M$) is $m \times n$. The purpose is to find a good projection vector, \mathbf{x} , such that when \mathbf{A}_j is projected onto \mathbf{x} , the projected feature vector, \mathbf{y}_j of the image \mathbf{A}_j can

be obtained.

$$\mathbf{y}_j = \mathbf{A}_j \mathbf{x} \quad j=1, 2, \dots, M \quad (2.3)$$

Similar to LDA, the between-class scatter matrix, \mathbf{TS}_B , and the within-class scatter matrix, \mathbf{TS}_W , of the projected feature vectors can be computed by using training images. The criterion is to project the images onto a subspace that maximizes the between-class scatter and minimizes the within-class scatter of the projected data. Since the total scatter of the projected samples can be represented by the trace of the covariance matrix of the projected feature vectors, the Fisher linear projection criterion can be described as

$$\mathbf{J}(\mathbf{x}) = \frac{\text{tr}(\mathbf{TS}_B)}{\text{tr}(\mathbf{TS}_W)} = \frac{\mathbf{x}^T \mathbf{S}_B \mathbf{x}}{\mathbf{x}^T \mathbf{S}_W \mathbf{x}} \quad (2.4)$$

where $\mathbf{S}_B = \sum_{i=1}^L N_i (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})^T (\bar{\mathbf{A}}_i - \bar{\mathbf{A}})$ and $\mathbf{S}_W = \sum_{i=1}^L \sum_{\mathbf{A}_k \in I_i} (\mathbf{A}_k - \bar{\mathbf{A}}_i)^T (\mathbf{A}_k - \bar{\mathbf{A}}_i)$.

The optimal projection, \mathbf{x}_{opt} , can be decided when the criterion is maximized. That is, $\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x}} \mathbf{J}(\mathbf{x})$. The solution can be found by solving the generalized eigenvalue problem. Therefore, if d optimal projection axes $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d$ corresponding to the first d largest eigenvalues are chosen, then features for a given image \mathbf{I} can be extracted. The following equation can be used for feature extraction.

$$\mathbf{y} = \mathbf{I} [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d] \quad (2.5)$$

Since \mathbf{I} is an $m \times n$ matrix and $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d]$ is an $n \times d$ matrix, an $m \times d$ matrix, \mathbf{y} , can be formed to represent the original image, \mathbf{I} . The 2D-LDA representation of the original image is adopted as the input feeding into the classifier in the next stage.

2.2.3 RBF and SVM

Radial Basis Function Network (RBFN) is a two-layer, hybrid feed forward learning network. It is a fully connected network and generally used as a classification tool. It was first used for discrimination by Broomhead and Lowe in 1988 [10]. In a RBF model, the first layer from input nodes to hidden neurons is an unsupervised layer and the second layer from hidden neurons to output nodes is the supervised layer. This means that the first layer connecting input nodes and hidden neurons has a unit weight and does not need training; however, the weights between hidden neurons and output nodes must be trained.

The biggest difference between RBFN and traditional neural network is that RBFN only computes its optimal weights one time but traditional neural networks have to adjust their weights for a couple of times by using back propagation algorithm. In RBFN, each hidden neuron has a symmetric radial basis function as an activation function. Typically, Gaussian-like and direct or inverse multiquadric-like radial basis functions are mostly used as activation functions. The purpose of the hidden neurons is to cluster the input data and reduce dimensionality. The objective of the RBF network is to train input data in order to minimize the sum of square errors and find the optimal weights between hidden neurons and output nodes. Those optimal weights can classify effectively the test data into correct classes. Figure 2.5 shows the architecture of a traditional radial basis function network.

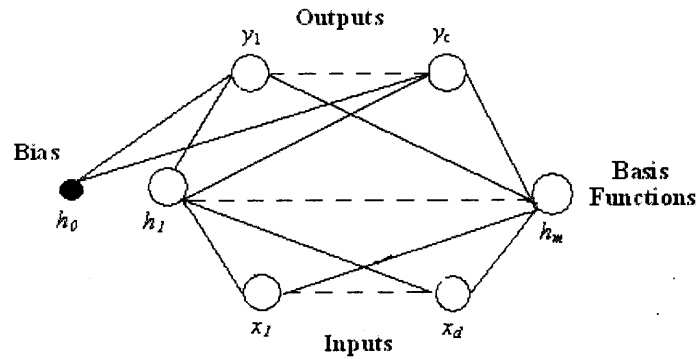


Figure 2.5 The architecture of a radial basis function neural network.

Support Vector machine (SVM) [86] is a learning system that separates a set of pattern vectors into two classes with an optimal separating hyperplane. The set of vectors is said to be optimally separated by the hyperplane if it is separated without an error and the distance between the closest vector to the hyperplane is maximal. SVM produces the pattern classifier by applying a variety of kernel functions (e.g. linear, polynomial, and radial basis function) as possible sets of approximation functions by optimizing the dual quadratic programming problem, and by using structural risk minimization as the inductive principle, as opposed to classical statistical algorithms that maximize the absolute value of an error or a squared error.

SVM is designed to handle dichotomic classes of input vectors. Recently, researchers have expanded from two-class classification to multi-class classification [32, 33]. Different types of SVM are used depending upon the distribution of input patterns. A linear maximal margin classifier is used for linearly separable data, a linear soft margin classifier is used for overlapping classes, and a nonlinear classifier is used for classes that are overlapped as well as separated by nonlinear hyperplanes. Since there are 7 different facial expressions to be recognized, tree-based one-against-one SVMs [35] are adopted to

perform multi-classes classification.

2.3 Experimental Procedure

The experimental procedure is categorized into three stages: preprocessing, feature extraction, and expression classification, as illustrated in Figure 2.6.

2.3.1 Pre-processing

Similar to [12, 21, 56, 73, 97], the original image of size 256×256 are cropped manually into 168×120 by removing the background influences. Since the illumination condition is varied to the images in JAFFE database, histogram equalization is applied to eliminate lighting effect.

2.3.2 Feature Extraction

DWT is applied to the cropped images two times. After that, the LL area is extracted for further analysis. Then 2D-LDA is used to extract important features from each LL part of DWT image. For comparison, other feature representations such as PCA, LDA, ICA, and 2D-PCA are tested.

2.3.3 Expression Classification

The linear SVM is performed to identify seven facial expressions in JAFFE database. To handle the multi-class problem, the tree-based one-against-one SVMs are constructed as in [35]. Different kernel types of SVM such as linear, polynomial, and radar basis function are also tested in this stage to compare the performance with the linear SVM.

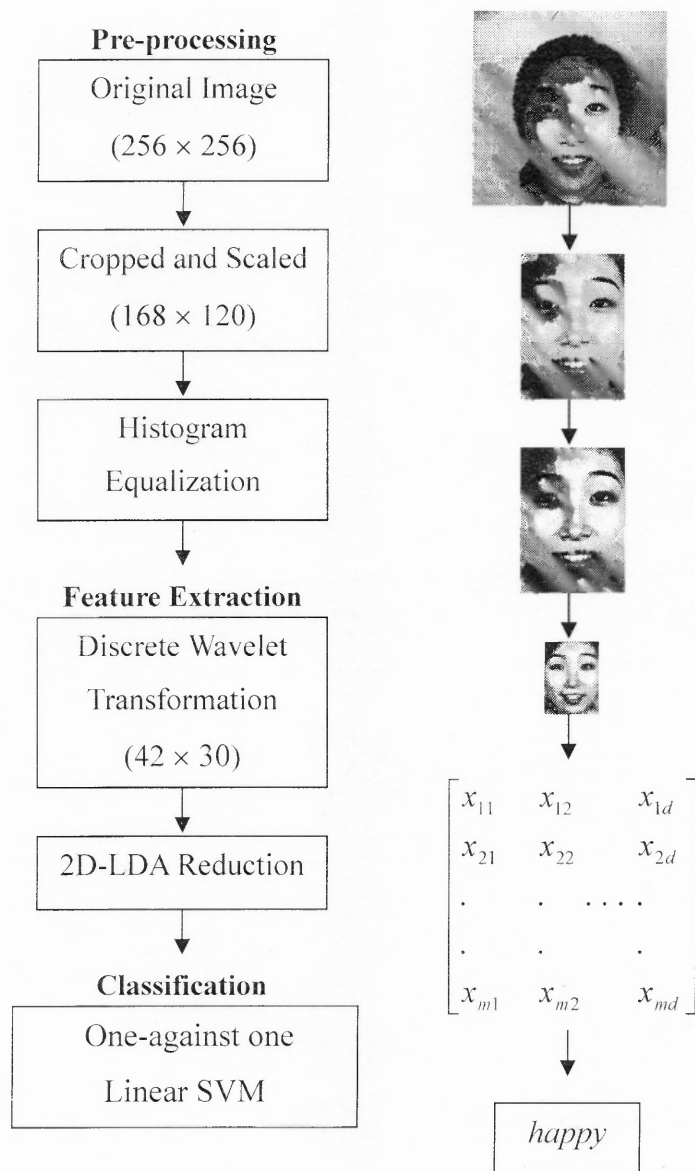


Figure 2.6 The experimental procedure.

2.4 Experimental Results

The objective of the proposed system is to identify different facial expressions accurately and efficiently. Since the JAFFE database contains only 213 images, the cross-validation strategy as in [56, 73, 97] and the leave-one-out strategy as in [12] are used to perform comparisons with other existing systems.

For the cross-validation strategy, the database is randomly divided into 10 segments in terms of different facial expressions. Each time, nine out of the ten segments is trained and the remaining segment is tested. The same procedure of training and testing is performed repeatedly for 30 times. At last, all the 30 recognition rates are averaged to obtain the final performance of the proposed system. For the leave-one-out strategy, each time only one image is tested in each class, and the remaining images are used for training.

The experimental results show that the proposed system can successfully meet the criteria of accuracy and efficiency for identifying different facial expressions. For accuracy, the proposed method outperforms other existing systems based on the same database. The recognition rate of the proposed system is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. For efficiency, it takes only 0.0357 second to process one input image of size 256×256 .

Table 2.1 shows the performance comparisons among the proposed system and the existing systems using the same JAFFE database. From Table 2.1, it is confirmed that no matter which strategy is used, the proposed system outperforms the others. The effects of different kernels of SVM, such as polynomial and radial basis functions, are tested for comparison. In those experiments, linear SVM is most suitable for the JAFFE database. It is because the feature vectors extracted by 2D-LDA are clustered in each class and can be separated by the linear SVM. The results are shown in Table 2.2. From those experimental results, Li and Yuan's conclusion [51] that 2D-LDA is superior to PCA, LDA, and 2D-PCA as the feature extraction method is conformed. The performances among different feature extraction with different classifier are also tested.

Table 2.1 Performance Comparisons in JAFFE Database

The Existing Systems	Strategy	Generalization Rate
Lyons <i>et al.</i> [56]	Cross-validation	92.00%
Zhang <i>et al.</i> [97]	Cross-validation	90.10%
Buciu <i>et al.</i> [12]	Leave-one-out	90.34%
Dubuisson <i>et al.</i> [21]	N/A	87.60%
Shinohara and Otsu [73]	Cross-validation	69.40%
The proposed system	Cross-validation	95.71%
	Leave-one-out	94.13%

Tables 2.3 and 2.4 show the performance comparisons among different feature extraction methods (such as PCA, LDA, 2D-PCA, ICA and 2D-LDA) along with SVM and RBFN, respectively. From Tables 2.3 and 2.4, SVM is the best classifier; however, RBFN is unreliable since the recognition rate is unsatisfied.

Table 2.2 Performance Comparisons of Using Different Kernels of SVM

Kernel Functions	Recognition Rates	
	Cross Validation	Leave-One-Out
Linear	94.13%	95.71%
Polynomial with degree 2	91.43%	92.38%
Polynomial with degree 3	92.86%	94.29%
Polynomial with degree 4	92.22%	93.33%
Radial basis function	85.71%	87.14%

Table 2.3 Comparisons of PCA, LDA, 2D-PCA, ICA and 2D-LDA Using SVM

Feature Extraction Method	Recognition Rates		Testing Speed per Image (second)
	Cross Validation	Leave-One-Out	
LDA + SVM	91.27%	91.90%	0.0367
2D-PCA + SVM	92.06%	93.33%	0.0357
ICA + SVM	93.35%	93.81%	0.0359
PCA + SVM	93.43%	94.84%	0.0353
2D-LDA + SVM	94.13%	95.71%	0.0357

2.5 Conclusions

In this chapter, different feature representation and expression classification schemes are investigated to recognize seven different facial expressions on JAFFE database. Experimental results show that the proposed system using DWT and 2D-LDA as feature representation technique, and linear one-again-one SVMs as classification method outperforms other systems. For accuracy, the recognition rate of the proposed system is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. For efficiency, it takes only 0.0357 second to process one input image of size 256×256 . That means the proposed system can be applied to real-time applications.

Since the proposed system is robust in feature extraction and pattern recognition, it can not only apply to facial expression recognition but also face recognition. In the aforementioned systems, the original images are cropped manually. How to automatically align and crop the image without degrading the recognition rate is the challenging problem. This problem will be addressed in Chapter 5.

CHAPTER 3

FACIAL EXPRESSION RECOGNITION-ACTION UNITS CASE

3.1 Introduction

Facial expression plays a principal role in human interaction and communication since it contains critical and necessary information regarding emotion analysis. The task of automatically recognizing different facial expressions in human-computer environment is significant and challenging. A variety of systems [1, 6, 8, 9, 12, 13, 20, 21, 56, 73, 80, 97, 98] have been developed to perform facial expression recognition in terms of database and expression targets. Some researchers [12, 21, 56, 73, 97] used homogeneous subject database, such as JAFFE database [56], to recognize seven prototypical, emotion-specified facial expressions, such as happy, neutral, angry, disgust, fear, sad, and surprise.

Lyons *et al.* [56] made use of Gabor filter at different scales and orientations, and applied 34 fiducial points for each convolved image to construct the feature vector for representing each facial image. After that, PCA is applied to reduce the dimensionality of feature vectors and LDA is used to identify seven different facial expressions. Their recognition rate is 92% for JAFFE database. Zhang *et al.* [97] adopted Gabor wavelet coefficients and geometric positions to construct the feature vector for each image and applied two-layer perceptron to distinguish seven different facial expressions. Their recognition rate is 90.1%.

Buciu *et al.* [12] tested different feature extraction methods such as Gabor filter and ICA combined with SVM using 3 different kernels: linear, polynomial, and radial basis function, to check which combination can produce the best result. From their

experiments, the best recognition rate is 90.34% by using Gabor wavelet at high frequencies combined with the polynomial kernel SVM at degree 2. Dubuisson *et al.* [21] combined the sorted PCA as feature extractor with LDA as the classifier to recognize facial expressions. Their recognition rate is 87.6%. Shinohara and Otsu [73] used Higher-order Local Auto-Correlation (HLAC) features and LDA to test the performance. Unfortunately, their system is unreliable since the correct rate is only 69.4%.

Some researchers [6, 20, 81] explored heterogeneous subject database, such as Cohn-Kanade expression database [44], to classify upper or lower face Action Units (AUs). Tian *et al* [81] developed an Automatic Face Analysis (AFA) system to analyze individual action unit or combinations based on both permanent and transient facial features in frontal face image sequence. The recognition rate of their AFA system is 95.6% on Cohn-Kanade expression database. Donato [20] compared different techniques for classify 6 single upper and lower face action units, on Ekman-Hager facial action exemplars [23], respectively. They found that the best performance is obtained by adopting Gabor wavelet decomposition or ICA representation. The recognition rate is 96.9% for their system. Bazzo and Lamar [20] invented a new pre-processing step based on the neutral face average difference. Their system used neural network based classifier combined with Gabor wavelet to categorize seven upper and lower face single action unit or combinations, respectively. The recognition rates are 86.55% and 81.63% for the upper and lower face, respectively.

In previous chapter, homogeneous database, JAFFE, is processed. In this chapter, heterogeneous database, Cohn-Kanade expression database, is applied to recognize the individual action unit or combinations. The rest of this chapter is organized as follows.

Section 3.2 describes Facial Action Coding System (FACS) and Face Expression Database. Section 3.3 presents the experimental procedure. Section 3.4 provides the experimental results and performance comparisons. Finally, conclusions are given in Section 3.5.

3.2 Facial Action Coding System (FACS) and Face Expression Database

3.2.1 Facial Action Coding System (FACS)

In 1978, Ekman and Friesen [22] designed the Facial Action Coding System (FACS) for characterizing facial expressions by action units. This system is a human observed system developed to explain the subtle changes of facial expressions.

Totally, they defined 44 FACS action units (AUs). Of those AUs, 30 are related to facial muscular contraction. There are 12 for upper face and 18 for lower face. For example, Action Unit 1 is related to frontalis and pars medialis describing inner corner of eyebrow raised and Action Unit 27 is related to pterygoids and digastric depicting mouth stretched open. The remainders of AUs are attributed to miscellaneous actions. For example, Action Unit 21 portrays the status of neck tighten. Tables 3.1-3.3 show the description of all Action Units including upper face, miscellaneous action units and lower face, respectively.

The action units can appear individual or in combinations. The combination of action units has additive or non-additive effects. Additive combination means that the combination does not alter the appearance of the comprised AUs. The example of additive combination is AU 12 + AU 25 displaying smiling and mouth open. Non-additive combination means that the appearance of the comprised AUs is modified. It represents the difficulty and complication for the expression recognition. The example of

non-additive combination is AU 12 + AU 15 showing that the lip corner of AU 12 is changed by the downward motion of AU 15.

Table 3.1 The Upper Face Action Units

Action Units	Upper Facial Muscle	Description of Motion Changes
1	Frontalis, pars medialis	Inner corner of eyebrow raised
2	Frontalis, pars lateralis	Outer corner of eyebrow raised
4	Corrugator supercilii, Depressor supercilii	Eyebrows drawn medially and down
5	Levator palpebrae superioris	Eyes widened
6	Orbicularis oculi, pars orbitalis	Cheeks raised; eyes narrowed
7	Orbicularis oculi, pars palpebralis	Lower eyelid raised and drawn medially
41	Relaxation of levator palpebrae superioris	Upper eyelid droop
42	Orbicularis oculi	Eyelid slit
43	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Eyes closed
44	Orbicularis oculi, pars palpebralis	Eyes squinted
45	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Blink
46	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Wink

Table 3.2 Miscellaneous Actions

Action Units	Description of Motion Changes
8	Lips toward
19	Tongue show
21	Neck tighten
29	Jaw thrust
30	Jaw sideways
31	Jaw clench
32	Bite lip
33	Blow
34	Puff
35	Cheek suck
36	Tongue bulge
37	Lip wipe
38	Nostril dilate
39	Nostril compress

Table 3.3 The Lower Face Action Units

Action Units	Lower Facial Muscle	Description of Motion Changes
9	Levator labii superioris alaeque nasi	Upper lip raised and inverted; superior part of the nasolabial furrow deepened; nostril dilated by the medial slip of the muscle
10	Levator labii superioris	Upper lip raised; nasolabial furrow deepened producing square-like furrows around nostrils
11	Levator anguli oris (a.k.a. Caninus)	Lower to medial part of the nasolabial furrow deepened
12	Zygomaticus major	Lip corners pulled up and laterally
13	Zygomaticus minor	Angle of the mouth elevated; only muscle in the deep layer of muscles that opens the lips
14	Buccinator	Lip corners tightened. Cheeks compressed against teeth
15	Depressor anguli oris (a.k.a. Triangularis)	Corner of the mouth pulled downward and inward
16	Depressor labii inferioris	Lower lip pulled down and laterally
17	Mentalis	Skin of chin elevated
18	Incisivii labii superioris and Incisivii labii inferioris	Lips pursed
20	Risorius w/ platysma	Lip corners pulled laterally
22	Orbicularis oris	Lips everted (funneled)
23	Orbicularis oris	Lips tightened
24	Orbicularis oris	Lips pressed together
25	Depressor labii inferioris, or relaxation of mentalis, or orbicularis oris	Lips parted
26	Masseter; relaxed temporal and internal pterygoid	Jaw dropped
27	Pterygoids and digastric	Mouth stretched open
28	Orbicularis oris Lips sucked	Lips sucked

3.2.2 Cohn-Kanade AU-Coded Face Expression Image Database

The facial expression image database used in experiments is the Cohn-Kanade AU-Coded Face Expression Image Database [44]. This database is representative, comprehensive and robust test-bed for comparative studies of facial expression.

It contains image sequences from 210 adult subjects between the ages of 18 and 50 years. For gender classification, there are 69% female and 31% in this database. For

racial group, there are 81% Euro-American, 13% Afro-American, and 6% other groups. Over 90% of the subjects had no prior experience in FACS. Subjects were instructed by an experimenter to perform single AUs and AU combinations.

Subjects sat directly in front of the camera and performed a series of facial behaviors which was recorded in an observation room. Lighting conditions and context were relatively uniform. The image sequences in this database also include in-plane and out-of plane head motion from small to mild. The resolution of the image sequence is 490×640 for grayscale and 480×640 for 24-bit color images. The image sequences began with a neutral face and ended with the maximum of the corresponding action units. Face size varies between 90×80 and 220×200 pixels.

3.3 Experimental Procedure

In the proposed system, except using histogram equalization for adjust lighting condition, Independent Component Analysis is adopted as feature extraction and representation method and Support Vector Machine is applied as classification measure. The following section will briefly describe basic concept of the two methods.

3.3.1 Independent Component Analysis

Independent component analysis (ICA) is a statistical and computational technique for finding out hidden factors that are representative and favorable for separating the different sets of images, sounds, telecommunication channels or signals. ICA was originally designed to process the cocktail-party problem [16, 43] and belongs to a class of *blind source separation* (BSS) methods for separating data into underlying representative components.

ICA is a general-purpose statistical and unsupervised technique where observed random vector are linearly transformed into components that are minimally dependent from each other. The concept of ICA is an extension of the principal component analysis (PCA), which can only impose independence up to the second order and, consequently, defines directions that are orthogonal. The formulation of ICA in this section is based on [55].

Suppose that $y \in \mathbb{R}^n$ is a n dimensional random vector. The covariance matrix of y is defined as

$$\Sigma_y = \varepsilon\{[y - \varepsilon(y)][y - \varepsilon(y)]^t\} \quad (3.1)$$

where $\varepsilon(\bullet)$ is the expectation operator, t denotes the transpose operation, and $\Sigma_y \in \mathbb{R}^{n \times n}$ is a real and symmetric square matrix. The ICA of random vector y factorizes the covariance matrix Σ_y into

$$\Sigma_y = F \Gamma F^t \quad (3.2)$$

where $\Gamma \in \mathbb{R}^{n \times n}$ is diagonal real positive and $F \in \mathbb{R}^{n \times n}$ transforms the original random vector $y \in \mathbb{R}^n$ to a new one $Z \in \mathbb{R}^{m \times n}$, where $y = FZ$, such that the m components ($m \leq n$) of the new random vector Z are independent or the most independent possible [16].

Let $P_z(u)$ be the probability density function (pdf) of the random vector Z . Vector Z has mutually independent components if and only if its joint density is equal to the product of its marginal densities

$$p_z(u) = \prod_{i=1}^m p_{z_i}(u_i) \quad (3.3)$$

To derive the ICA transformation, Comon [16] developed an optimization criterion for measuring the independence of the components of the random vector Z . This criterion calculates the Kullback–Leibler divergence (or relative entropy) of the two pdf's corresponding to the left and the right side of Equation (3.3)

$$I(p_z) = \int p_z(\mathbf{u}) \log \frac{p_z(\mathbf{u})}{\prod p_{z_i}(u_i)} d\mathbf{u} \quad (3.4)$$

$I(P_z)$ thus specifies the average mutual information of Z . Equations (3.3) and (3.4) show that the mutual information vanishes if and only if the random vector has mutually independent components. Equation (3.4) can be rewritten as follows [15]:

$$I(p_z) = J(p_z) - \sum J(p_{z_i}) + \frac{1}{2} \log \frac{\prod V_{ii}}{|V|} \quad (3.5)$$

where V is the covariance matrix of Z , and $J(P_z)$ the negentropy, a measure of similarity between a density $P_z(u)$ and the Gaussian density $\phi(P_z)$

$$J(p_z) = - \int p_z(\mathbf{u}) \log \frac{\phi_z(\mathbf{u})}{p_z(\mathbf{u})} d\mathbf{u} \quad (3.6)$$

By means of Equations (3.5) and (3.6), which provide a way to approximate the mutual information, Comon [16] developed an optimization procedure (minimization of the mutual information) that consists of three major steps:

- i) a whitening procedure, which involves only second-order statistics, cancels the last term of Equation (3.5);
- ii) a number of rotation transformations, which apply highorder statistics by means of k -statistics, minimize the second term on the right side of Equation (3.5) while keeping the others constant;
- iii) a normalization procedure, which standardizes the column vectors F in Equation (3.2) in terms of order, norm, and phase (sign), defines a unique ICA representation.

The potential applications of ICA can be applied to separation of artifacts in MEG data, finding hidden factors in financial data, reducing noise in natural images, telecommunications [40], face recognition [5], and biomedical signal processing, and so on. Bartlett *et al* [5]. thought that in face recognition, much of important information is contained in the high-order statistics of the images, so ICA representation is superior than PCA representation that is contained second-order statistics.

3.3.2 Support Vector Machine

SVMs, introduced by Vapnik [86], are learning systems that separate sets of input pattern vectors into two classes with an optimal separating hyperplane. The set of vectors is said to be optimally separated by the hyperplane if it is separated without error and the distance between the closest vectors to the hyperplane is maximal. SVMs produce the pattern classifier by 1) applying a variety of kernel functions (linear, polynomial, radial basis function (RBF), and so on) as the possible sets of approximating functions, 2) optimizing the dual quadratic programming problem, and 3) using structural risk minimization as the inductive principle, as opposed to classical statistical algorithms that maximize the absolute value of an error or of an error squared.

Different types of SVM classifiers are used depending upon the type of input patterns: a linear maximal margin classifier is used for linearly separable data, a linear soft margin classifier is used for linearly nonseparable, or overlapping, classes, and a nonlinear classifier is used for overlapped classes. The promising applications of SVM can be used on object and face detection, text categorization, face and facial expression recognition, optical character and speech recognition, and medical fields such as breast cancer diagnosis and protein structure prediction.

3.3.3 Experimental Procedure

A complete automatic facial expression analysis system includes face detection, facial component extraction and representation, and facial expression recognition stages. The purpose of facial detection is to automatically detect and catch the region of faces in the still images or sequences. Since there are a lot of algorithms and methods developed for detecting faces [69, 79, 87], in this study, it is assumed that the exact face is automatically located and ready to feed into the next stage.

The purpose of the facial extraction and representation is to find out the most valuable, favorable, and representative information that is derived from facial expression changes to represent the original detected faces. The implication of this stage is to reduce the dimensionality of the detected faces from the previous stage and to speed up the computation of the next stage, facial expression recognition.

The purpose of the facial expression recognition is to accurately and promptly identify different facial expressions. The facial expressions to be recognized can be divided into two types. The first type is emotion-specified expressions, such as happy, angry, surprise, and so on [12, 21, 56, 73, 97]. The other is facial action units [6, 20, 81]. Since in previous chapter, emotion-specified expression recognition is performed, in this paper, facial action unit recognition is intended. Figure 3.1 describes the proposed recognition system in this chapter.

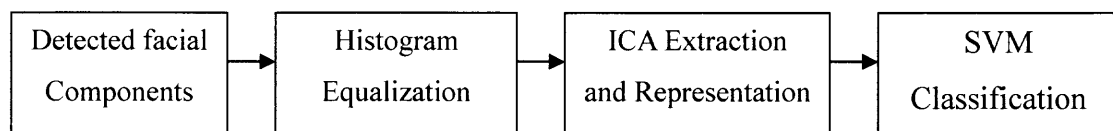


Figure 3.1 The experimental procedure.

Since the exact faces that can be located by face detection algorithms [69, 79, 87] are assumed, the first step of this proposed system is to divide the detected face into upper and lower part. The illumination condition is uneven, so histogram equalization is applied to the detected images to eliminate the lighting effect. To extract the subtle changes of the facial expressions, Independent Component Analysis is implemented. Linear Support Vector Machine classifier is adopted to recognition the individual action unit or in combination. The proposed system is fast and can be applied to real time application. All the experiments are processed in the Matlab 7 environment under XP and Pentium IV with 2.80 GHz. It only takes about 1.8 ms for processing a testing image.

3.4 Experimental Results

There are four experiments to be performed in this Chapter. The first experiment intends to recognize 6 individual or in combinations of the upper face action units including AU 4, AU 6, AU 1+AU 2, AU 1+AU 4, AU 4+AU 7, and AU 1+AU 2+AU 5. The second experiment is constructed to classify six individual or in combinations of the upper face action units containing AU 17, AU 9+AU 17, AU 12+ AU 25, AU 15+AU 17, AU 20+AU 25, and AU 25+AU 27.

The third experiment is designed to categorize four combination of action units on the whole face including neutral, AU 1+AU 2+AU 5+AU 25+AU 27, AU 6+AU 12+AU 25, and AU 4+AU 17. The last experiment is schemed to analyze the effect of gender factor on the previous three experiments. 27 subjects are randomly selected from the Cohn-Kanade AU-Coded Face Expression Image Database in terms of the action units to be recognized for those experiments.

Of 27 subjects, 20 are female and 7 male. Since the image sequences start from neutral and end at maximum of the corresponding action units, the first three and last three images are picked as the neutral and corresponding action units, respectively. Totally, there are 141 images for upper part of action units, 126 for lower part, and 135 for the whole face, respectively.

Since the dataset is limited, 90% of dataset is selected randomly as training set and the rest of 10% as test set. The same procedure is performed ten times, the recognition rates of ten independent tests are averaged, and the final performance for each experiment are obtained. Tables 3.4-3.6 describe the recognition results for different action units. Table 3.7 compares the difference of performance on action units between genders.

From Table 3.4, most misclassification is derived from the combination of AU4 + AU7. This category at times is misclassified as AU6. The cause is the combination of AU4 + AU7 make the eyes narrow as the AU6 does. From Table 3.5, most misclassification stems from the combination of AU20 + AU25. This group, sometimes, is recognized as the group of AU12 + AU25, since the difference between the two groups is the motion of lip corner. From Tables 3.4-3.6, there is no significant difference to our proposed system for recognizing the upper and lower part of face.

The recognition rates of 97.06% and 97.13% are obtained for upper part and lower part, respectively. 100% of recognition rate is obtained for classifying the action units appearing on the whole face. It means that for expression recognition, the combination of facial feature components, such as eyes and mouth, can produce better results than only single component is used.

It is interesting to explore the gender effect on recognizing the action units. From the results on Table 3.7, the males can express the action units more accurately than the females do. From the performance point of view, the proposed system is reliable and comparable to other published systems. Table 3.8 summarizes the performance of published systems with our proposed method.

Table 3.4 Recognition Rate for Upper Part of Face







Patterns	Images	Number of Samples	Recognition Rate
AU4		12	19/20 = 95%
AU6		48	50/50 = 100%
AU1 + AU2		12	19/20 = 95%
AU1 + AU4		9	10/10 = 100%
AU4 + AU7		12	17/20 = 85%
AU1 + AU2 + AU5		48	50/50 = 100%
Total		141	165/170 = 97.06%

Table 3.5 Recognition Rate for Lower Part of Face







Patterns	Images	Number of Samples	Recognition Rate
AU17		12	20/20 = 100%
AU9 + AU17		9	10/10 = 100%
AU12+ AU25		39	40/40 = 100%
AU15 + AU17		9	10/10 = 100%
AU20 + AU25		12	16/20 = 80%
AU25+ AU27		45	40/40 = 100%
Total		126	136/140 = 97.13%

Table 3.6 Recognition Rate for Whole Face





Patterns	Images	Number of Samples	Recognition Rate
Neutral		48	100%
AU4 + AU17		15	100%
AU6 + AU12 + AU25		39	100%
AU1+AU2+AU5+AU25+AU27		33	100%
Total		135	100%

Table 3.7 Recognition Rate Between Genders on Action Units

Action units on different part of face	Recognition rate	
	Male	Female
Action units on upper face	100%	95.38%
Action units on lower face	100%	93.33%
Action units on whole face	100%	99.29%

Table 3.8 Performance Comparison of Different Systems

Systems	Database	AUs to be recognized	Recognition rate
Tian <i>et al.</i> [81]	Cohn-Kanade	AU 9, 10, 12, 15, 17, 20, 25, 26, 27, 23+24	95.6%
Donato [20]	Ekman-Hager	AU 1, 2, 4, 5, 6, 7	96.9%
Bazzo and Lamar [6]	Cohn-Kanade	Upper AU 0, 6, 1+2, 4+7, 1+2+5, 4+6+7+9, 4+7+9	86.55%
		Lower AU 0, 25, 26, 27, 12+25, 15+17, 20+25	81.63%
Proposed system	Cohn-Kanade	Upper AU 4, 6, 1+2, 1+4, 4+7, 1+2+5	97.06%
		Lower AU 17, 9+17, 12+25, 15+17, 20+25, 25+27	97.13%
		Whole face AU neutral, 4+17, 6+12+25, 1+2+5+25+27	100%

3.5 Conclusions

In this chapter, recognizing action units, the subtle change of face, instead of the prototypical and emotion-specified facial expressions, such as happy, surprise, and so on, is analyzed. The comprehensive and heterogeneous subject database, Cohn-Kanade AU-coded face expression image database, is tested. The proposed system is reliable, efficient and comparable to other published system. It is also fast and can be applied to real-time application. The recognition rate of the proposed method is 97.06% for recognizing the upper part of face, 97.13% for lower part of face, and 100% for the whole face. The gender effect on recognizing the action units is explored. Two important findings are obtained. First, for recognition purpose, the combination of facial features is better than individual features. Second, for gender effect, the males can express the action units more accurately than the females do.

CHAPTER 4

NEONATAL PAIN CLASSIFICATION

A new application area developed by the Infant COPE project and database developed by Dr. Brahnam [9] at Missouri State University is the recognition of neonatal facial expressions of pain. It has been reported in medical literature that health care professionals have difficulty distinguishing a newborn's facial expressions of pain from facial reactions to other stimuli. Although a number of pain instruments have been developed to assist health professionals, studies demonstrate that health professionals are not entirely impartial in their assessment of pain and fail to capitalize on all the information exhibited in a newborn's facial displays. In this chapter these problems are tackled by applying three different state-of-the-art face classification techniques to the task of distinguishing a newborn's facial expressions of pain.

This chapter, in full, is a collaborative summary and reprint of materials that has been accepted for publication in *Artificial intelligence in Medicine*, Brahnam S, Chuang C.-F., Shih F. Y., and Slack M. R., in press 2005 [9] and *Special Issue on "Decision Support in Medicine" in Decision Support System*, Brahnam S, Chuang C.-F., Sexton R. S., and Shih F. Y., in press 2005 [8]. Computer Vision Laboratory at NJIT is participated in the Infant COPE project that is designed, developed, and led by Dr. Brahnam at Missouri State University. All the experiments were done in Missouri State University and the Computer Vision Laboratory at NJIT.

4.1 Introduction

The assessment of pain in newborns is considered one of the most challenging problems in neonatology [26]. Pain assessment is difficult because neonates cannot articulate their pain experiences and vary tremendously in their responses to pain and other stimuli [26, 88]. Since pain is a major indicator of medical problems [58] and the quality of patient care depends on the quality of pain management [33], it is vital that methods be developed that accurately distinguish an infant's signals of pain from a host of minor distress signals [88].

At present, assessment of neonate pain takes into consideration a number of physiological and behavioral factors. Among the many physiological indicators of pain are changes in heart and respiratory rates, blood pressure, vagal tone, and palmar sweating [15]. The consensus in the reference literature, however, is that physiological measures are insufficient and unreliable indices of pain. Physiological measures vary significantly from newborn to newborn, and they fail to reflect the intensity of pain [58]. Moreover, the physiological parameters associated with pain are not easily distinguishable from parameters associated with fear and anxiety [85].

Significant neonate behavioral responses to pain include body movement, crying, and facial expressions [58]. Facial expressions play a central role in pain assessment, as attested by the fact that most pain instruments developed for infants, toddlers, and older children, including COMFORT [4], CRIES [48], FLACC (Face, Legs, Activity, Cry, Consolability) [60], MIPS (Modified Infant Pain Scale) [11], and CFACS (Child Facial Coding System) [29], rely in whole or in part on facial displays. The facial characteristics associated with pain in infants include prominent forehead, narrowed eyes, deepening of

the nose-lip furrow, and an angular opening of the mouth [30]. Facial expressions are a critical factor in the assessment of infant pain because they are the most specific and frequent indicators of pain [78]. Body movement and crying are behaviors that are associated with other states, such as hunger, fright, and discomfort; and neonates do not always respond to pain by crying and moving. Sleep, for instance, inhibits bodily movement; yet, in sleep, an infant's face will often register the experience of pain [59]. This is an issue that is particularly relevant to neonates since they spend between 14 and 17 hours a day sleeping [90].

Even though the facial characteristics of infant expressions of pain have been studied extensively [30], there are serious problems with pain assessment instruments that utilize facial displays. The primary problem is that these tools rely on the observations of health professionals, and health professionals have been shown to be biased in their observations and less competent than nonhealth professionals in recognizing facial expressions of pain [68, 92]. Xavier Balda *et al.* [92] theorize that health professionals become desensitized because of their constant exposure to suffering. The findings of Xavier Balda *et al.* [92] corroborate other studies demonstrating that the greater the clinical experience of the health professional the more likely he or she is to underestimate patient pain [68]. A repeated refrain in the reference literature, therefore, is that neonate pain assessment tools need to be developed that alleviate or circumvent the problem of observer desensitization and bias [39, 92].

In order to evaluate and assess the reliable machine recognition of acute pain, the Infant COPE database is used. This database containing 204 facial expressions of 26 neonates (age 18-36 hours) was photographed experiencing the acute pain of a heel lance

and three non-pain stressors including transport from one crib to another (a disturbance that can provoke crying that is not in response to pain), an air stimulus on the nose, and friction on the external lateral surface of the heel..

The objective of Infant COPE project is to tackle these problems by applying state-of-the-art face classification techniques to the task of distinguishing a newborn's facial expressions of pain from facial expressions that are similar but not triggered by pain. Since the assessment of pain by machine is based on pixel states, the development of a machine classification system of pain will offer the following advantages: it will remain objective, exploit the full spectrum of information available in a neonate's facial expressions, and not degrade over time. A machine classification system of pain will offer the additional benefit of monitoring a neonate's facial expressions when the patient is left unattended.

As described more fully in Section 4.3, the face classification techniques used in this study are PCA (principal component analysis), LDA (linear discriminant analysis), and SVM (support vector machines). Although these techniques have succeeded in classifying faces according to identity [7, 82, 83], gender [61, 65], age [84], race [64], and emotions [12, 97], they have yet to be applied to medical problems that involve the face. Dai *et al.* [19] have proposed a method for observing the facial expressions of patients in hospital beds, but their facial images were not of actual patients but rather of subjects responding to verbal cues suggestive of medical procedures and conditions. To date, no work has employed face classification techniques to the task of classifying actual facial expressions of pain.

In Infant COPE project, PCA, LDA, and SVM, are trained and tested using facial photographs of neonates experiencing four noxious stimuli: transport from one crib to another, air puff on the nose, friction from cotton and alcohol rubbed on the lateral surface of the heel, and the puncture of a heel lance. The experimental design is described more completely in Section 4.2, and the three classifiers are reviewed in Section 4.3. Section 4.4 discusses the methods and the procedures used in the classification experiments, and Section 4.5 presents the experimental results. The paper is concluded, in Section 4.6, by noting some of the contributions and limitations of this study and by offering directions for future research.

4.2 Study Design

Warnock and Sardin [88] have recently stressed the importance of including a variety of contrasting stimuli in studies on infant pain assessment. Early research focused mostly on neonate responses to two stimuli: a pain inducing stimulus (pin prick or puncture of a lancet) and friction on the heel [31, 41]. Contemporary studies tend to include additional stressors, such as exposure to bright light [92] and diaper change [88].

Infant COPE database follows contemporary research by including two stressors in addition to the puncture of a lancet. Since PCA, LDA, and SVM classifiers are sensitive to light changes, an air puff stimulus was introduced in lieu of a bright light stimulus. Exposure to a puff of air on the face is similar to exposure to bright light in that it causes the eyes to squeeze tightly together, producing a facial expression that is similar, yet distinct, from the facial expression of pain. At the suggestion of hospital personnel, infants were transported from one crib to another before taking the photographs. This

change was welcomed as it supplied a stressor similar to diaper change: neonates can respond to both experiences by crying. It afforded the opportunity of contrasting classifier recognition rates of neonate crying expressions that were in response to pain to those crying expressions that were in response to a less noxious stimulus.

4.2.1 Subjects

This study complied with the protocols and ethical directives for research involving human subjects at St. John's Health System, Inc. Informed consent was obtained from a parent, usually the mother in consultation with the father. Most parents were recruited in the neonate unit of a St. John's Hospital sometime after delivery. Only mothers who had experienced uncomplicated deliveries were approached.

A total of 204 color photographs were taken of 26 Caucasian neonates (13 boys and 13 girls) ranging in age from 18 hours to 3 days old. Six males had been circumcised the day before the photographs were taken, and the last feeding time before the photography session ranged from 45 minutes to 5 hours. All infants were in good health.

4.2.2 Apparatus

All photographs were taken using a Nikon D100 digital camera under ambient lighting conditions in a room separated from other newborns.

4.2.3 Procedure

The facial expressions of the newborns were photographed experiencing four distinct stimuli: transport from one crib to another, air puff on the nose, friction from cotton and alcohol rubbed on the heel, and the puncture of a heel lance. The state of the infant after being transported to another crib was further evaluated at the time the photographs were

taken into one of two states: crying or resting.

All stimuli were administered by an attending nurse. Following the example of [92] and the requirements of standard medical procedures, photographs of the facial expressions of the four stimuli were taken in the following sequence:

1. Transport from one crib to another (Rest/Cry): After being transported from one crib to another the neonate was swaddled and not handled for 1 minute. After this minute of rest, a photograph was taken and the state of the neonate noted as either crying or resting.
2. Air Stimulus: After another minute of rest, the neonate's nose was exposed to a puff of air emitted from a squeezable plastic camera lens cleaner. A series of pictures of the neonate's face was taken immediately after the air puff contacted the infant's face.
3. Friction: After another minute of rest the neonate received friction on the external lateral surface of the heel with cotton wool soaked in 70% alcohol for 10 to 15 seconds. During the friction movements the face of the neonate was photographed.
4. Pain: After one minute of rest the external lateral surface of the heel was punctured when collecting blood. Several continuous photographs of the neonate's face were taken starting immediately after introduction of the lancet and while the skin of the heel was squeezed for blood samples.

Of the 204 facial photographs, 67 are Rest, 18 are Cry, 23 are Air Stimulus, 36 are Friction, and 60 are Pain. Figure 4.1 provides two example sets, with backgrounds removed, of the five neonate facial expressions of rest, cry, air puff, friction, and pain.

4.3 Basic Concepts of PCA, LDA, and SVM

In Infant COPE project, three types of classifiers are used to train and to test infant facial expressions: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM). The basic concepts behind these classifiers are presented in this section.

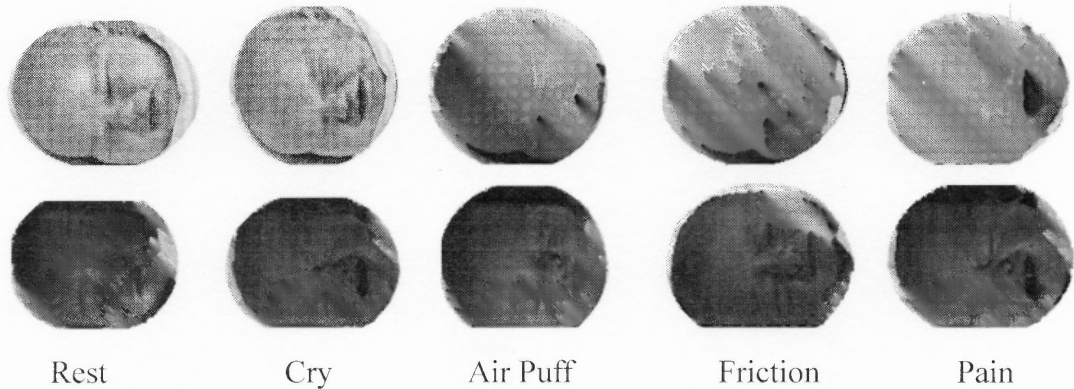


Figure 4.1 Examples of the five facial expressions in the Infant COPE dataset.

4.3.1 Principal Component Analysis

The central idea behind PCA is to find an orthonormal set of axes pointing in the direction of maximum covariance in the data. In terms of facial images, the idea is to find the orthonormal basis vectors, or the eigenvectors, of the covariance matrix of a set of images, with each image treated as a single point in a high dimensional space. It is assumed that the facial images form a connected subregion in the image space. The eigenvectors map the most significant variations between faces and are preferred over other correlation techniques that assume every pixel in an image is of equal importance, (see, for instance, [47]).

Since each image contributes to each of the eigenvectors, the eigenvectors resemble ghostlike faces when displayed. For this reason, they are oftentimes referred to in the literature as *holons* [18] or *eigenfaces* [82, 83], and the new coordinate system is referred to as the *face space* [82, 83]. Examples of eigenfaces are shown in Figure 4.2. Individual images can be projected onto the face space and represented exactly as weighted combinations of the eigenface components (see Figure 4.3).

The resulting vector of weights that describe each face can be used both in face classification and data compression. Classification is performed by projecting a new image onto the face space and comparing the resulting weight vector to the weight vectors of a given class [82, 83]. Compression is achieved by reconstructing images using only those few eigenfaces that account for the most variability [74].

PCA Classification

The principal components of a set of images can be derived directly as follows. Let $\mathbf{I}(x, y)$ be a two-dimensional array of intensity values of size $N \times N$. $\mathbf{I}(x, y)$ may also be represented as a single point, a one-dimensional vector Γ of size N^2 . Let the set of face images be $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$. Let

$$\Phi_k = \Gamma_k - \Psi \quad (4.1)$$

represent the mean normalized column vector for a given face Γ_k , where

$$\Psi = \frac{1}{M} \sum_{k=1}^M \Gamma_k \quad (4.2)$$

is the average face of the set.

PCA seeks the set of M orthonormal vectors, \mathbf{u}_k , and their associated eigenvalues, λ_k , which best describes the distribution of the image points. The vectors \mathbf{u}_k and scalars λ_k are the eigenvectors and eigenvalues, respectively, of the covariance matrix

$$\mathbf{C} = \frac{1}{M} \sum_{k=1}^M \Phi_k \Phi_k^T = \mathbf{A} \mathbf{A}^T, \quad (4.3)$$

where the matrix $\mathbf{A} = [\Phi_1, \Phi_2, \dots, \Phi_M]$ (Turk and Pentland [81, 82]).

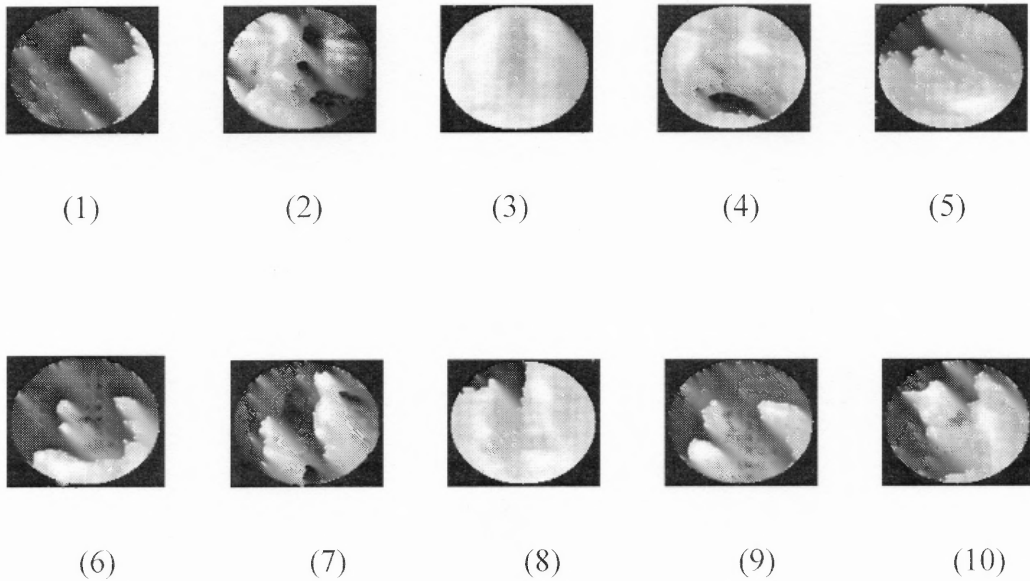


Figure 4.2 The first ten eigenfaces of the 204 neonate images in Infant COPE database. The eigenfaces are ordered by magnitude of the corresponding eigenvalue.

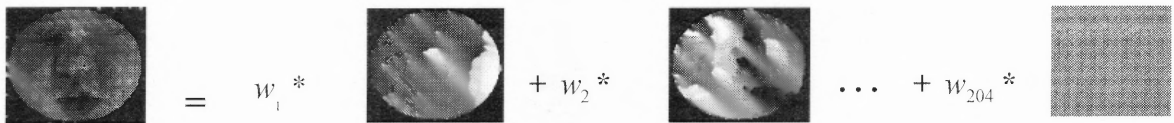


Figure 4.3 Illustration of the linear combination of eigenfaces. The face to the left can be represented as a weighted combination of eigenfaces.

The size of \mathbf{C} is N^2 by N^2 which for typical image sizes is an intractable task (Turk and Pentland, [82, 83]). However, since typically $M < N^2$, that is, the number of images is less than the dimension, there will only be $N - 1$ non-zero eigenvectors. Thus, the N^2 eigenvectors can be solved, in this case, by first solving for the eigenvectors of an $M \times M$ matrix, followed by taking the appropriate linear combinations of the data points

Φ (Turk and Pentland, [82, 83]).

PCA is closely associated with the **Singular Value Decomposition** of a data matrix and can be decomposed as

$$\Phi = \mathbf{USV}^T. \quad (4.4)$$

Where \mathbf{S} is a diagonal matrix whose diagonal elements are the singular values, or eigenvalues, of Φ , and \mathbf{U} and \mathbf{V} are unary matrices. The columns of \mathbf{U} are the eigenvectors of $\Phi\Phi^T$, and are referred to as *eigenfaces*. The columns of \mathbf{V} are the eigenvectors of $\Phi^T\Phi$ and are not used in this analysis.

Faces can be classified by projecting a new face Γ onto the face space as follows:

$$\omega_k = \mathbf{u}_k^T (\Gamma_k - \Psi) \quad (4.5)$$

for $k = 1, \dots, M'$ eigenvectors, with $M' \ll M$, if reduced dimensionality is desired. The weights form a vector $\mathbf{\Omega}_k^T = [\omega_1, \omega_2, \dots, \omega_{M'}]$, which contains the projections onto each eigenvector. Classification is performed by calculating the distance of $\mathbf{\Omega}_k$ from $\mathbf{\Omega}$, where $\mathbf{\Omega}$ represents the average weight vector defining some class [81, 82].

Data Compression

Since the eigenfaces are ordered, with each one accounting for a different amount of variation among the faces, images can be reconstructed using only those few eigenfaces, $M' \ll M$ in equation 4.3.4, that account for the most variability [74]. Because PCA results in a dramatic reduction of dimensionality and maps the most significant variations in a dataset, it is typically used to represent faces when performing other classification procedures.

4.3.2 Linear Discriminant Analysis

While PCA is optimal for reconstructing images from a low dimensional space, it is not optimal for discrimination. PCA yields projection directions that maximize the total scatter across all classes. LDA, or Fisher's linear discriminants, in contrast, is a supervised learning procedure that projects the images onto a subspace that maximizes the between-class scatter and minimizes the within-class scatter of the projected data. A classical technique in pattern recognition, LDA is an example of a *class specific method* in that it shapes the scatter in order to make it more reliable for classification [7]. There has been a tendency to prefer LDA to PCA because LDA deals directly with discrimination between classes, whereas PCA aims at faithfully representing the data. It has been shown that LDA outperforms PCA only when large and representative training data sets are given [57].

4.3.3 Support Vector Machine

Support Vector Machines (SVM), introduced by Vapnik [86], are learning systems that separate a set of input pattern vectors into two classes with an optimal separating hyperplane. The set of vectors is said to be optimally separated by the hyperplane if it is separated without error and the distance between the closest vectors to the hyperplane is maximal. SVM produces the pattern classifier by applying a variety of kernel functions (linear, polynomial, radial basis function, and so on) as the possible sets of approximating functions, by optimizing the dual quadratic programming problem, and by using structural risk minimization (SRM) as the inductive principle, as opposed to classical statistical algorithms that maximize the absolute value of an error or of an error squared.

Originally, SVM was designed to handle dichotomic classes. Recently, research has concentrated on expanding two-class classification to multi-class classification. Since the objective in this paper has been to distinguish neonate facial displays of pain from other facial displays, this discussion of SVM will be limited to dichotomic classification.

Different types of SVM classifiers are used depending upon the type of input patterns: a linear maximal margin classifier is used for linearly separable data, a linear soft margin classifier is used for linearly non-separable, or overlapping, classes, and a nonlinear classifier is used for classes that are overlapped as well as separated by nonlinear hyperplanes. All three classifiers are discussed in more detail below. It should be noted, however, that the linearly separable case is rare in real world problems and was not explored in our experiments.

4.3.3.1 Linear Maximal Margin Classifier. The case where the training patterns can be linearly separated by a hyperplane, $w \cdot x + b = 0$, is the simplest case and provides a good foundation for the other two cases. The purpose of the SVM is to find the optimal values for w (e.g., w_0) and b (e.g., b_0). After finding the optimal separating hyperplane, $w_0 \cdot x + b_0 = 0$, an unseen pattern, x_i can be classified by the decision rule $f(x) = \text{sign}(w_0 \cdot x_i + b_0)$, as shown below.

Suppose, there is a set of training data, x_1, x_2, \dots, x_k , where $x_i \in \mathbf{R}^n$ with $i = 1, 2, \dots, k$. Each x_i , belonging as it does to one of two classes, has a corresponding value y_i , where $y_i \in \{-1, 1\}$. The goal in this case is to build the hyperplane that maximizes the minimum distance between the two classes. Because the hyperplane is $w \cdot x + b = 0$, the

training data can be divided into 2 classes such that

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \quad \text{if } y_i = 1, \quad (4.6)$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad \text{if } y_i = -1.$$

where $\mathbf{w} \in \mathbf{R}^n$ and $b \in \mathbf{R}$.

Combining the equations in (4.6), Equation (4.7) is obtained.

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall \mathbf{x}_i, i = 1, 2, \dots, k. \quad (4.7)$$

The distance between a point \mathbf{x} and the hyperplane is $d(\mathbf{w}, b; \mathbf{x}) = |\mathbf{w} \cdot \mathbf{x} + b| / \|\mathbf{w}\|$.

According to (4.6), the minimum distance between one of the two classes and the hyperplane is $\frac{1}{\|\mathbf{w}\|}$. The margin, M , which is the distance between the two classes, is $\frac{2}{\|\mathbf{w}\|}$.

Finding the optimal separating hyperplane having a maximal margin requires that the following minimization problem be solved:

$$\text{Minimize:} \quad \frac{1}{2} \mathbf{w} \cdot \mathbf{w}$$

$$\text{Subject to} \quad y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall \mathbf{x}_i, i = 1, 2, \dots, k.$$

This nonlinear optimization problem with inequality constraints can be solved by the *saddle point* of the Lagrange function:

$$\mathbf{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} - \sum_{i=1}^K \alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1), \quad (4.8)$$

where $\alpha_i \geq 0$ are the Lagrange multipliers.

By minimizing the Lagrange function with respect to \mathbf{w} and b , as well as by maximizing with respect to α_i , the minimization problem above can be transformed to its dual problem, called the quadratic programming problem:

$$\frac{\partial L(\mathbf{w}, b, \alpha_i)}{\partial \mathbf{w}} \Big|_{\mathbf{w} = \mathbf{w}_0} = \left(\mathbf{w}_0 - \sum_{i=1}^K \alpha_i y_i \mathbf{x}_i \right) = \mathbf{0} \quad (4.9)$$

$$\frac{\partial L(\mathbf{w}, b, \alpha_i)}{\partial b} \Big|_{b = b_0} = \sum_{i=1}^K y_i \alpha_i = 0 \quad (4.10)$$

Equation (4.11) can be obtained by plugging (4.9) and (4.10) into (4.8).

$$\mathbf{L}(\alpha) = \sum_{i=1}^K \alpha_i - \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (4.11)$$

The dual problem can be described as follows:

$$\begin{aligned} \text{Maximize:} \quad & \sum_{i=1}^K \alpha_i - \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{Subject to:} \quad & \sum_{i=1}^K y_i \alpha_i = 0, \quad \alpha_i \geq 0 \end{aligned}$$

By solving the dual problem, the optimal separating hyperplane is determined by (4.12) and (4.13).

$$\mathbf{w}_0 = \sum_{i=1}^K \alpha_i y_i \mathbf{x}_i \quad (4.12)$$

$$b_0 = y_i - \mathbf{w}_0 \cdot \mathbf{x}_i \quad (4.13)$$

where \mathbf{x}_i belongs to support vectors, $y_i \in \{-1, 1\}$.

The unseen test data \mathbf{x}_t can be classified, therefore, by simply computing (4.14).

$$f(x) = \text{sign}(\mathbf{w}_0 \cdot \mathbf{x}_t + b_0) \quad (4.14)$$

By examining (4.12), it can be seen that the hyperplane is determined by all the training data, \mathbf{x}_i , that have the corresponding attributes of $\alpha_i > 0$. This kind of training data is called *support vectors*. Thus, the optimal separating hyperplane is not determined by the training data per se but rather by the support vectors.

4.3.3.2 Linear Soft Margin Classifier. As mentioned above, patterns that are linearly separable are rare in real world problems. In this sub-session, SVM is expanded to handle input patterns that are overlapping, or linearly non-separable. In this case, our objective is to separate the two classes of training data with a minimal number of errors.

To accomplish this, some non-negative slack variables ξ_i is introduced, $i = 1, 2 \dots k$ to the system. Thus, (4.6) and (4.7), in the linearly separable case above, can be rewritten as (4.15) and (4.16):

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1 - \xi_i \quad \text{if } y_i = 1, \quad (4.15)$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 + \xi_i \quad \text{if } y_i = -1.$$

$$y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2 \dots k \quad (4.16)$$

Just as the optimal separating hyperplane in the linearly separable case, obtaining the soft margin hyperplane in the linear non-separable case requires that the following minimization problem be solved:

$$\text{Minimize: } \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \left(\sum_{i=1}^k \xi_i \right),$$

$$\text{Subject to: } y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2 \dots k.$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots k$$

where C is a penalty or regularization parameter.

By minimizing the Lagrange function with respect to \mathbf{w} , b , and ξ_i , as well as by maximizing with respect to α_i , the minimization problem above can be transformed to its dual problem, described as follows:

$$\begin{aligned} \text{Maximize:} \quad & \sum_{i=1}^K \alpha_i - \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{Subject to:} \quad & \sum_{i=1}^K y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C \end{aligned}$$

Solving the dual problem, the soft margin hyperplane is determined by (4.17) and (4.18).

$$\mathbf{w}_0 = \sum_{i=1}^K \alpha_i y_i \mathbf{x}_i \quad (4.17)$$

$$b_0 = y_i - \mathbf{w}_0 \cdot \mathbf{x}_i \quad (4.18)$$

where \mathbf{x}_i belongs to margin vectors, $y_i \in \{-1, 1\}$.

By examining (4.3.17), it can be seen that the hyperplane is determined by all the training data, \mathbf{x}_i that have the corresponding attributes of $\alpha_i > 0$. These support vectors can be divided into two categories. The first category has the attribute of $\alpha_i < C$. In this category, $\xi_i = 0$, and these support vectors lie at the distance $\frac{1}{\|\mathbf{w}\|}$ from the optimal separating hyperplane. These support vectors are called *margin vectors*. The second category has the attributes of $\alpha_i = C$. In this category, the support vectors are correctly classified with either a distance smaller than $\frac{1}{\|\mathbf{w}\|}$ from the optimal separating hyperplane (if $0 < \xi_i \leq 1$) or they are misclassified (if $\xi_i > 1$). The support vectors in the second category are regarded as errors.

4.3.3.3 Non-linear Classifier. Sometimes, the input vectors can not be linearly separated in the input space. In this case, kernel functions, such as the polynomial or RBF kernel functions are used to transform the input space to a feature space of higher dimensionality. In the feature space, a linear separating hyperplane is sought that separates the input vectors into two classes.

If $\mathbf{x} \in \mathbf{R}^n$ is in the input space, the input vector \mathbf{x} can be mapped from the n -dimension input space to a corresponding N -dimension feature space through a function, ϕ . After the transformation, it is known $\phi(\mathbf{x}) \in \mathbf{R}^N$. Following the steps described in the case of linearly separable training patterns (section 4.3.3.1) and the case of linearly non-separable training patterns (section 4.3.3.2), the hyperplane and decision rule for nonlinear training patterns can be established.

So, similar to obtain the hyperplane for the linearly separable training patterns in equations (4.12) and (4.13) and the hyperplane for the linearly non-separable training patterns in equations (4.17) and (4.18), the hyperplane can be obtained for the non-linear training pattern as in equation (4.19).

$$\mathbf{w}_0 \cdot \phi(\mathbf{x}) + b_0 = \sum_{i=1}^K \alpha_i y_i \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) + b_0 \quad (4.19)$$

From the equation (4.19), the original dot products of input variables are replaced by a function, ϕ . That is kernel function, $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})$. The decision rule for nonlinear training patterns can be established as shown in equation (4.20):

$$f(x) = \text{sign}\left(\sum_{i=1}^K \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b_0\right) \quad (4.20)$$

4.4 Experimental Procedure

As illustrated in Figure 4.4, the experimental procedures can be divided into the following stages: preprocessing, feature extraction, and classification.

In the preprocessing stage, the original images are cropped, rotated, and scaled such that the eyes lie roughly along the same axis. The original 204 images, size 3008×2000 pixels, are also reduced to 100×120 pixels.

In the feature extraction stage, facial features are centered within an ellipse and color information is discarded. The rows within the ellipse are concatenated to form a feature vector of dimension 8383 with entries ranging in value between 0 and 255. PCA is used to reduce the dimensionality of the feature vectors further. In our experiments, the first 70 principle components resulted in the best classification scores.

Finally, in the classification stage, PCA, LDA, and SVM are used to classify the feature vectors into the following category pairs: pain/non-pain, pain/cry, pain/air puff, and pain/friction.

4.5 Experimental Results

This section describes three face classification experiments. In experiment 1, PCA, LDA, and SVM classify faces into the following classification pairs: a) pain/rest, b) pain/cry, c) pain/air puff, and d) pain/friction. In experiment 2, PCA, LDA, and SVM classify faces into the classification pair of pain/non-pain. The set of non-pain images was obtained by combining the rest, cry, air puff, and friction images into one category of 140 images. The remaining 60 images were of pain. The two experimental cases above are person-dependent where part of the data for each subject is used as training set, and another part as test set. The cross-validation technique is used in person-dependent case.

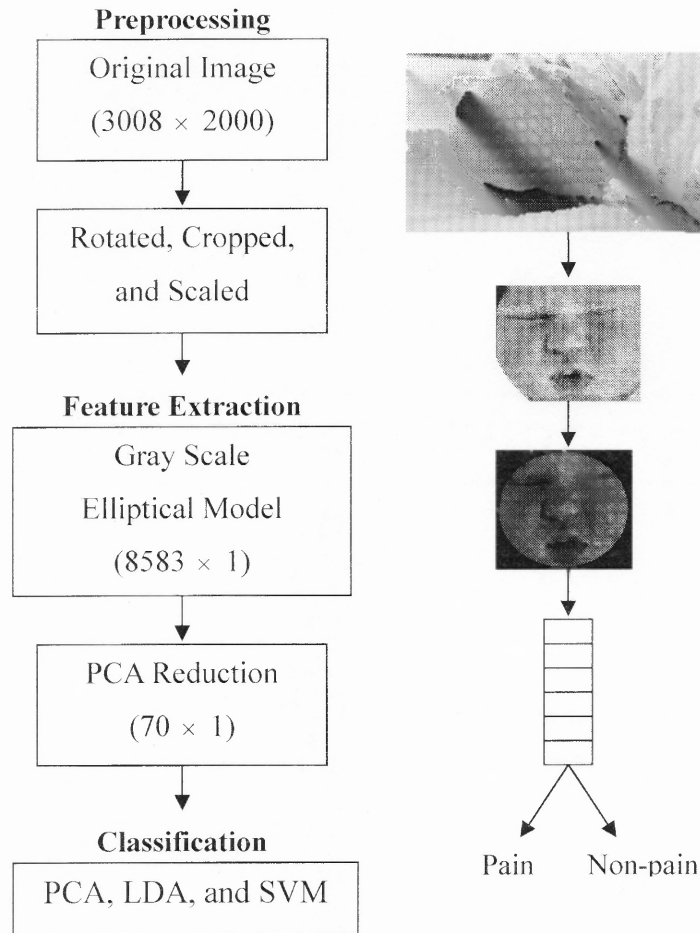


Figure 4.4 The experimental procedure.

As the intention of the initial study was to examine classification differences between pain expressions and a variety of similar non-pain expressions, the images used in the experiments were divided into training and testing sets based on facial expression categories, not subjects. As a result, the training and testing sets contained multiple samples of each subject in each expression category. While it is true that ideally, as is the case with speech recognition software, samples of individual subjects would be used to personalize the classifier, in a clinical setting, it is more realistic to assume that the classifier would be trained previously on one set of subjects and then applied out of the

box to future newborns. The experiment 3, person-independent case, is designed to examine the more realistic evaluation protocol of requiring that subjects in the testing sets not be included in the training sets. Furthermore, only two categories of facial expressions are examined: pain and non-pain, a choice that is discussed further in Section 4.5.2.

4.5.1 Person-dependent Case

Because the number of images in the dataset is small, a cross-validation technique was applied in all experiments. The images in each experiment were randomly divided into ten segments, and nine out of the ten segments were used in training. The remaining segment was used in testing. This procedure was repeated ten times. The ratio of right to wrong classifications was used as the classification index, and the ten classification indexes were averaged for a final classification score.

4.5.1.1 Experiment 1. Tables 4.1-4.4 compare the classification scores of PCA (using the L_1 metric), LDA (using the L_1 metric) and SVM (using linear, polynomials of degree 2, 3, and 4, and RBF kernel functions) for each of the following expression pairs: pain/rest, pain/cry, pain/air puff, and pain/friction. Referring to Tables 4.1 and 4.2, the best classification score of pain versus rest is 94.62% and pain versus cry is 80.00%, using a SVM with polynomial kernel of degree 3. In Table 4.3, the best classification score of pain versus air puff is 90.00% using a linear SVM. In Table 4.4, most SVM systems separate pain from friction. The best result is 96.00% using a SVM with polynomial kernel of degree 2.

Tables 4.1-4.4 demonstrate that different SVM systems have distinct effects on recognizing pain from other facial expressions. For example, a linear SVM has a stable

recognition rate of 90.00% in pain versus all facial expressions except cry. In general, however, a SVM with polynomial kernel of degree 3 has the best overall classification performance.

Table 4.1 Pain versus Rest

	Recognition rate
Linear	90.77%
Polynomial with degree = 2	84.62%
Polynomial with degree = 3	94.62%
Polynomial with degree = 4	86.15%
RBF Kernel	53.85%
PCA with L1 distance	87.18%
LDA with L1 distance	93.08%

Table 4.2 Pain versus Cry

Type of SVM	Recognition rate
Linear	71.25%
Polynomial with degree = 2	78.75%
Polynomial with degree = 3	80.00%
Polynomial with degree = 4	76.25%
RBF Kernel	75.00%
PCA with L1 distance	68.75%
LDA with L1 distance	70.42%

Table 4.3 Pain versus Air Puff

Type of SVM	Recognition rate
Linear	90.00%
Polynomial with degree = 2	77.78%
Polynomial with degree = 3	83.33%
Polynomial with degree = 4	78.89%
RBF Kernel	66.67%
PCA with L1 distance	81.48%
LDA with L1 distance	89.63%

Table 4.4 Pain versus Friction

Type of SVM	Recognition rate
Linear	90.00%
Polynomial with degree = 2	96.00%
Polynomial with degree = 3	93.00%
Polynomial with degree = 4	92.00%
RBF Kernel	60.00%
PCA with L1 distance	74.00%
LDA with L1 distance	91.00%

4.5.1.2 Experiment 2. In experiment 2, PCA, LDA, SVM are used to classify faces into two categories: pain and non-pain. The non-pain set consisted of all the air puff, cry, friction, and rest images. Table 4.5 shows the results of the PCA, LDA and several SVM systems. The best classification score (88.00%) is obtained using SVM with polynomial kernel of degree 3. Recall that SVM with polynomial kernel of degree 3 provided the best overall classification score in experiment 1 as well. SVM with polynomial kernel of degree 3, therefore, is probably the best selection for classifying neonate facial expressions of pain.

Table 4.5 Pain versus Non-pain

Type of SVM	Recognition rate
Linear	83.67%
Polynomial with degree = 2	86.50%
Polynomial with degree = 3	88.00%
Polynomial with degree = 4	82.17%
RBF Kernel	70.00%
PCA with L1 distance	80.33%
LDA with L1 distance	83.67%

4.5.2 Person-independent Case

The evaluation protocol used in this study called for 26 experiments to be performed by each classifier. For each of the 26 subjects, the set of facial images for that subject formed the testing set, and the facial images of the remaining 25 subjects formed the training set. The 204 photographs were divided into two categories: pain and non-pain. The set of non-pain images was obtained by combining the rest, cry, air puff, and friction images into one category of 144 images. The set of pain images consisted of the remaining 60 images. For each experiment, classification scores were computed for each classifier by averaging the number of correct classifications made.

This section describes person-independent experimental results using PCA, LDA, and SVM. Tables 4.6 and 4.7 present the classification scores and 95% confidence interval. SVM with linear kernel has the highest classification rate of 82.35% accuracy, with a 95% confidence interval of $\pm 6.20\%$. Given this result, it is believed that there is a high potential for developing a decision support system for diagnosing neonatal pain from images of neonatal facial displays. Classification results for the 26 experiments performed by each method are presented in table 4.6. It is clear by examining the results in this table that certain subjects are easy to classify and others are more difficult. For instance, all four methods correctly classified the image sets associated with subjects 2, 21, and 25, whereas subjects 1, 9, and 26 proved difficult for most methods.

The remainder of this section provides a more detailed discussion of method parameters and classification results. PCA and LDA used the sum of absolute differences, or L1, distance metric in all experiments. Referring to Table 4.6, the average classification score for PCA was 80.30% and for LDA 76.96%. However, from Table 4.7

there is no statistical difference in performance between PCA and LDA. This is not unexpected, as it has been shown that LDA outperforms PCA only when large and representative training data sets are given [57]. All PCA and LDA experiments were processed in the MATLAB environment under the Windows XP operating system using a Pentium 4 – 2.80 GHz processor.

SVMs with five different Kernels (linear, RBF, polynomial degree 2, polynomial degree 3, and polynomial degree 4) performed the 26 experiments defined by the evaluation protocol. The regularization parameter, C , for the SVMs was determined using a grid search. Since the recognition rates in those experiments were not significantly different in terms of different values for C , the regularization parameter $C=1$ is adopted. The bandwidth parameter, $\gamma = 1.2$, in SVM using RBF kernels was also optimized using a grid search. Referring to Table 4.6, an SVM with linear kernel provided the best recognition rate of 82.35%. However, examining Table 4.7, there is no statistical significant between the classification rates of the various SVM methods. All SVM experiments were processed in the same MATLAB environment used for the PCA and LDA experiments. SVM was implemented using the OSU SVM Classifier MATLAB Toolbox developed by Ohio State University.

From Tables 4.5 and 4.6, results of the two experiments, person-dependent and – independent cases were contradictory in terms of the best kernel to use with SVM. An SVM with polynomial kernel of degree 3 obtained the best classification score (88.00%) in the person-dependent study, and an SVM with a linear kernel obtained the best classification score (82.35%) in the person-independent study. Sampling error caused by the small number of images in the sample pool is one possible explanation for this

discrepancy.

Table 4.8 compares the average classification scores obtained using the two cases. Referring to Table 4.8, the average classification score for PCA was 80.36% and for LDA 80.32%. SVM, as expected, outperformed both PCA and LDA, except in the case of RBF kernel. Given previous reports in facial expression classification using SVM (see, for instance, [54]), it is not expected the RBF kernel performance to be as low as it was.

There are several possible explanations for the kernel performance differences in the two studies. The most likely cause for the discrepancy is sampling error due to the small number of images in the sample pool. The average performance of the SVMs using the two kernels, for instance, is very close, the difference being only 1.18%. However, since the data in the training sets used in the two sets of experiments differ only in a few inputs (approximately 15%), this assumption is questioned. To determine if the difference in kernel performance is the result of sampling error, experiments that varied the size of the sample pool are implemented. This can be achieved by comparing SVM classification of pain expressions to each of the other four facial displays.

Table 4.6 Average Method Classification Scores and Individual Experiment Scores

Method Exp. No.	SVM with Linear Kernel	SVM with Polynomial degree =2	SVM with Polynomial degree = 3	SVM with Polynomial degree = 4	SVM with RBF Kernel	PCA L ₁ distance	LDA L ₁ distance
1	77.78%	44.44%	77.78%	55.56%	55.56%	88.89%	66.67%
2	100%	100%	100%	100%	100%	100%	100%
3	87.50%	75.00%	75.00%	75.00%	75.00%	62.50%	75.00%
4	60.00%	100%	100%	100%	60.00%	60.00%	60.00%
5	75.00%	83.33%	75.00%	75.00%	75.00%	66.67%	75.00%
6	85.71%	57.14%	85.71%	57.14%	85.71%	85.71%	85.71%
7	88.89%	88.89%	77.78%	77.78%	77.78%	77.78%	77.78%
8	88.89%	66.67%	77.78%	66.67%	66.67%	77.78%	77.78%
9	50.00%	66.67%	33.33%	66.67%	83.33%	16.67%	33.33%
10	70.00%	80.00%	80.00%	70.00%	90.00%	60.00%	80.00%
11	100%	100%	100%	87.50%	50.00%	100%	100%
12	75.00%	87.50%	87.50%	87.50%	87.50%	75.00%	62.50%
13	60.00%	60.00%	90.00%	60.00%	60.00%	80.00%	70.00%
14	90.91%	100%	81.82%	72.73%	72.73%	81.82%	81.82%
15	83.33%	83.33%	83.33%	83.33%	83.33%	66.67%	66.67%
16	83.33%	83.33%	83.33%	83.33%	66.67%	91.67%	91.67%
17	100%	88.89%	77.78%	66.67%	55.56%	100%	100%
18	90.00%	80.00%	80.00%	70.00%	60.00%	100%	100%
19	85.71%	85.71%	85.71%	85.71%	85.71%	100%	100%
20	75.00%	83.33%	66.67%	50.00%	58.33%	91.67%	58.33%
21	100%	100%	100%	100%	100%	100%	100%
22	87.50%	75.00%	75.00%	62.50%	62.50%	62.50%	62.50%
23	50.00%	83.33%	100%	100%	66.67%	66.67%	33.33%
24	100%	85.71%	85.71%	42.86%	57.14%	100%	85.71%
25	100%	100%	100%	100%	100%	100%	100%
26	100%	50.00%	50.00%	50.00%	50.00%	66.67%	66.67%
Average	82.35%	79.90%	80.39%	72.06%	70.10%	80.39%	76.96%

Table 4.7 95% Confidence Intervals Using t Distribution ($\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$)

Method	95 % Confidence Interval	Standard Deviation
SVM with Linear Kernel	82.35% \pm 6.20%	15.34%
SVM with Polynomial degree = 2	79.90% \pm 6.36%	15.74%
SVM with Polynomial degree = 3	80.39% \pm 6.23%	15.41%
SVM with Polynomial degree = 4	72.06% \pm 7.03%	17.41%
SVM with RBF Kernel	70.10% \pm 6.32%	15.64%
PCA with L1 distance	80.39% \pm 7.98%	19.75%
LDA with L1 distance	76.96% \pm 7.81%	19.34%

Table 4.8 Comparison of Classification Rates Using Person-dependent and -independent

Type of SVM	Person-dependent	Person-independent	Average
Linear	83.67%	82.35%	83.01%
Polynomial degree = 2	86.50%	79.90%	83.20%
Polynomial degree = 3	88.00%	80.39%	84.20%
Polynomial degree = 4	82.17%	72.06%	77.12%
RBF	70.00%	70.10%	70.05%
PCA with L1 distance	80.33%	80.39%	80.36%
LDA with L1 distance	83.67%	76.96%	80.32%

This resulted in pool sizes of 83 images for pain versus air stimulus, 78 images for pain versus cry, 96 images for pain versus friction, and 127 images for pain versus rest. Only protocol person-dependent case was used in these experiments, as splitting expressions for each subject (person-independent case) resulted in pool sizes that were too small for training.

Tables 4.1-4.4 present the results of the set of experiments. The average performance of the four experiments using SVM with a linear kernel is 85.51%, and the average performance of SVM with a polynomial kernel of degree 3 is 87.74%. The

difference in kernel performance (2.23%) is half that in [9] (4.33%), which also used person-dependent. This leads us to believe that sample error is most likely the cause of kernel performance differences. As far as neonatal facial expressions are concerned, the results of the new set of experiments suggest that there is no significant classification difference in SVMs using a linear kernel versus a polynomial kernel of degree 3. This conclusion is consistent with [54], which examined SVM expressing classification performance using a number of adult facial databases.

4.6 Conclusions

Reported in this chapter were the classification results of three face classification techniques, PCA, LDA, and SVM, applied to the Infant COPE database. Person dependent and independent cases are discussed. The facial expressions of 26 neonates experiencing the puncture of a heel lance, transport from one crib to another, air stimulus to the nose, and friction on the external lateral surface of the heel were photographed. The state of the infant after being transported from one crib to another was further noted as being in one of two states: resting or crying. A series of experiments compared the recognition rates of PCA, LDA, and SVM classifying the following pairs: pain/non-pain, pain/rest, pain/cry, pain/air puff, and pain/friction.

For person-dependent case, SVM with a polynomial kernel of degree 3 produced the best overall recognition rates of pain versus non-pain (88.00%), pain versus cry (80.00%), pain versus rest (94.62%), pain versus air puff (83.33%), and pain versus friction (93.00%) is inferred. For person-independent, the classification rates of SVM were promising. SVM with a Linear Kernel provided the best SVM classification rate of

82.35% accuracy. PCA and LDA using an L1 distance metric produced an average classification rate of 80.39% and 76.96% respectively.

The Infant COPE project makes a number of contributions. It is a first attempt at applying state-of-the-art face recognition technologies to actual medical problems. As noted in the introduction, medical applications of face recognition technologies have been suggested [19] but not tried with actual medical data. The results of this project are promising and suggest that face recognition technologies could prove useful in neonate pain assessment.

Moreover, even though machine classification of emotion has long been an area of active investigation, research that includes the machine classification of pain experiences is unaware. The Infant COPE project not only addresses pain, but the dataset in this project also includes facial expressions in response to several stressors that result in expressions that are similar to the facial displays of pain. Infants typically respond to pain by crying, for instance, but they also cry in reaction to a number of minor disturbances; this project has included in the dataset expressions of crying that were not triggered by pain experiences.

Finally, this project is one of the first to investigate machine classification of neonate facial displays. Most work in facial classification has focused on adult faces. Rarely have the faces of children been included in these studies, and certainly not the faces of infants.

The current experimental design has a number of limitations that need to be addressed. First, only reactions to acute pain experiences were included in the dataset. This project does not address chronic pain—pain experiences that are thought to have

long term psychological and neurological consequences [58]. Second, because this project uses photographs, it does not take into account the dynamic nature of facial expressions. It is possible that temporal changes in expressions include significant information regarding a neonate's state. Third, this study does not compare human assessment of neonate pain with machine assessment, nor does it speculate on the practicality of implementing these technologies within a hospital setting.

In terms of future research possibilities, the author of this dissertation is working with the other members of the Infant COPE project to design a study that will compare human recognition rates of pain with machine classification rates, and in future studies investigating the dynamic nature of facial displays is planned. Another research direction would be to combine machine recognition of *physiological indices* with machine recognition of facial expressions. Lindh, Wiklund, and Håkansson [53], for instance, have had some success classifying pain as it relates to heart rate variability using PCA. Machine recognition of *behavioral indicators*, however, offers the advantage of monitoring neonates without the attachment of sensors. Since there is some evidence that the temporal frequency and intensity information in cry can discriminate pain (see [50]), combining sound classifiers with face recognition is yet another area of potential research.

CHAPTER 5

FACIAL FEATURES DETECTION

One of the challenging problems for facial expression or face recognitions is how to automatically locate the principal facial parts from an original image since most existing algorithms capture the necessary face parts by manually cropping the original images. In this chapter, two novel methods are presented to automatically locate the face features especially for eyes. After locating the facial feature, eyes, the face region can be easily obtained from the original image by using the geometric relationship of the face. The advantage of the first method described in section 5.1 is that it can locate the face boundary and facial features without the training of a lot of images. However, it is time consuming. The benefit of the second method described in section 5.2 is that it is fast. However, it has to use a lot of training samples.

5.1 Automatic Extraction of Face and Facial Features

This section presents a novel approach for the extraction of human head, face and facial features. In the double-threshold method, the high-thresholded image is used to trace head boundary and the low-thresholded image is used to scan face boundary. After facial features candidates are obtained and noises are eliminated, x- and y-projections are applied to extract facial features such as eyes, nostrils and mouth. Because low contrast of chin occurs in some face images, its boundary cannot be completely detected. An elliptic model is used to repair it. Because of noises or clustered facial features candidates,

a geometric face model is applied to locate facial features and an elliptic model to trace face boundary. The Gabor filter algorithm is adopted to locate two eyes.

5.1.1 Introduction

Automatic extraction of human head and face boundaries and facial features is critical in the areas of face recognition, criminal identification, security and surveillance systems, human computer interfacing [71], and model-based video coding [2, 3]. In general, the computerized face recognition includes four steps [75]. First, the face image is enhanced and segmented. Second, the face boundary and facial features are detected. Third, the extracted features are matched against the features in the database. Fourth, the classification into one or more persons is achieved. In order to detect faces and locate the facial features correctly, researchers have proposed a variety of methods which can be divided into two categories. One is based on gray-level template matching, and the other is based on computation of geometric relationship among facial features.

In the geometric relationship aspect, Jeng *et al.* [42] proposed an efficient face detection approach based on the configuration of facial features. Using their method, one can detect the images with frontal-view faces as well as with tilted faces. However the method will fail on the images with face sizes smaller than 80×80 pixels or with multiple faces. Wong *et al.* [91] developed an algorithm for face detection and facial features extraction based on genetic algorithms and eigenface. Lin and Fan [52] presented a face detection algorithm to detect multiple faces in complex background. They assume that in the frontal-view face images, the centers of two eyes and the center of mouth form an isosceles triangle, and in the side-view face images, the center of one eye, the center of one ear hole, and the center of mouth form a right triangle. The algorithm will fail when

the images are too dark or eyes are occluded by hair.

In the template matching aspect, Ryu and Oh [70] proposed an algorithm based on eigenfeatures and neural networks for the extraction of eyes and mouth using rectangular fitting from gray-level face images. The advantage is that it does not need a large training set by taking advantage of eigenfeatures and sliding window. However, their algorithm will fail on the face images with glasses or beard. Besides the aforementioned two categories, some researchers used motion active contour or snakes to detect the face contours [63, 75, 93, 95]. Kass *et al.* [45] proposed the Snake algorithm in 1988, and since then it has been widely used to detect contours. Yow and Cipolla [95] used the active contour model to enhance the feature-based approach to detect the face boundary. Sobottka and Pitas [75] used snakes to trace face contour on a number of image sequences. Because each method has its own advantages, Nikolaidis and Pitas [63] developed a combined approach using adaptive Hough transform, template matching, active contour model, and projective geometry properties. They used adaptive Hough transform for curve detection, template matching for inner facial features location, active contour model for inner face contour detection, and projective geometry properties for accurate pose determination.

This section is organized as follows. Section 5.1.2 presents our methodology. Section 5.1.3 describes the facial features extraction based on the geometric face model. Section 5.1.4 presents our experimental results. Conclusions are made in Section 5.1.5.

5.1.2 Methodology

In this section, the proposed method is introduced to process the frontal-view face images for the extraction of head boundary, face boundary, and facial features including eyes with eyebrows, nostrils, and mouth. Head boundary is the outer profile of head including shoulders. Face boundary is the face contour that excludes hair, shoulders, and neck. Rectangular boxes are adopted to locate facial features.

5.1.2.1 Smoothing and Thresholding. The scheme diagram of the double-threshold method is shown in Figure 5.1. The first step is to reduce noise by using a 3×3 median filter. After that, an edge operator is applied. The edge detection technique by Wechsler and Kidode [89] is tested and the result is shown in Figure 5.2. The edge output appears too thin and the top-face boundary is too weak to be detected in the later thresholding procedure. In order to obtain fat boundary, Hu *et al.* [36] proposed four masks (horizontal (size 3×7), vertical (7×3), 45° (9×3), and 135° (9×3)) to detect image edges and select the maximum as the edge strength. Instead of using large sizes, smaller sizes of masks are developed as shown in Figure 5.3 for edge detection.

Experimental results show that the proposed method is as good as their method. Thresholding is performed to transform the gray-level edge images into binary. The high threshold value is determined by choosing all the high intensity pixels that occupy 5% of the entire pixels. The low threshold value is decided by choosing all the high intensity pixels that occupy 25% of the entire pixels. The high thresholded image is used to obtain the head boundary and the low thresholded image is used to produce the face boundary. These thresholding percentages are determined based on our empirical data in order to achieve the best results.

5.1.2.2 Track Head and Face Boundary. In order to trace the head boundary, the high-thresholded image is divided into left and right halves. When the edge image is scanned from top to bottom, the first layer of the contour occurred is the head boundary,

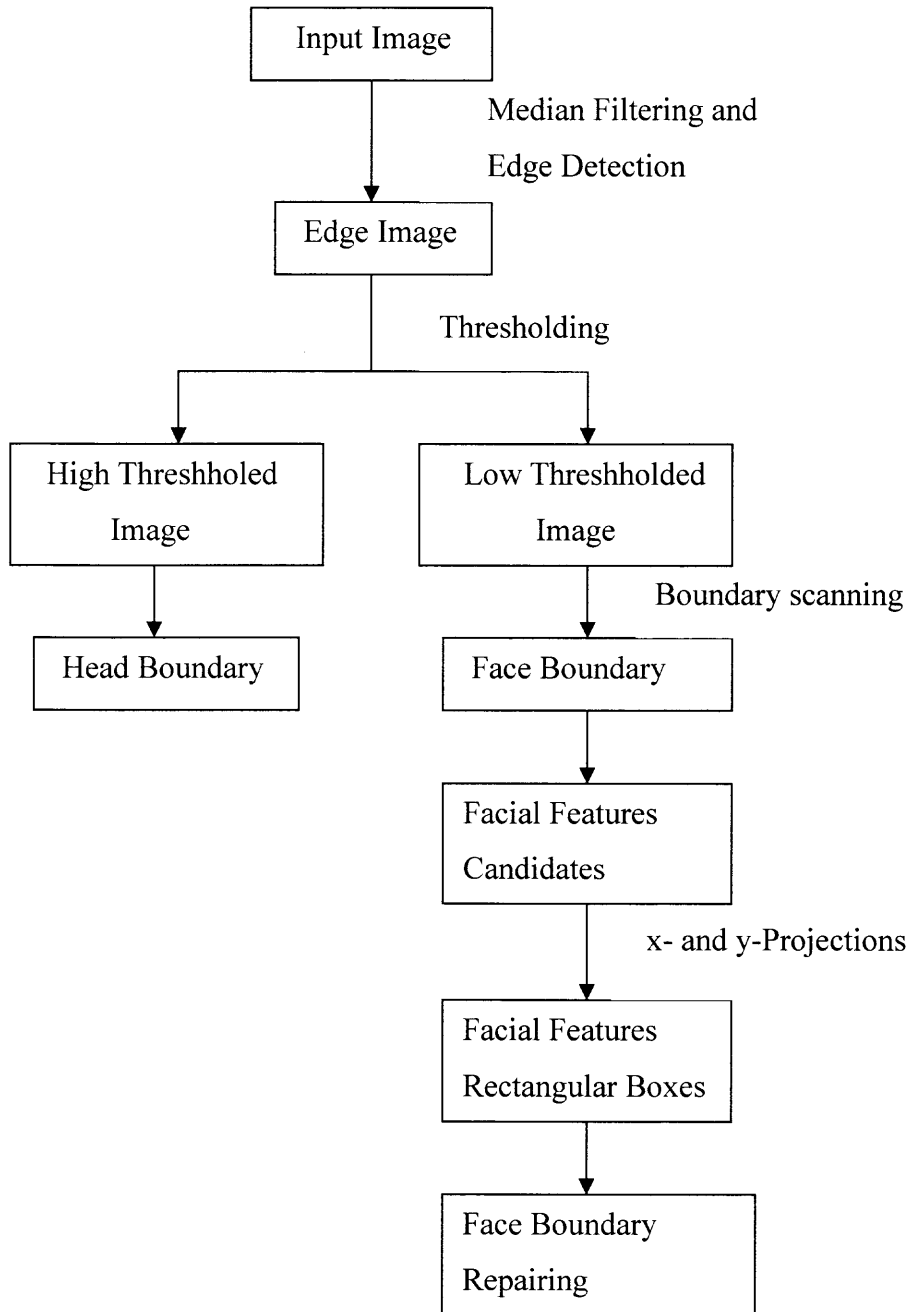


Figure 5.1 The scheme diagram of double-threshold method.

and the second layer of the contour occurred is the face boundary. For tracing the head boundary, a starting point is located as the first white pixel on the first layer of the left half. From the starting point, the left and right profiles of head are traced. Because the outer boarder of the first layer is shifted outwards from the actual boundary for a few pixels (say, p), the edge of the left half to the right and the edge of the right half to the left is adjusted by p pixels respectively.

Because some face profiles disappear in the high thresholded image, the low thresholded image is used to trace face boundary. The head borders are removed and a morphological opening is used to eliminate unnecessary noises. After that, the image is scanned from four directions (right to left, left to right, top to bottom, and bottom to top) to produce the face boundary.

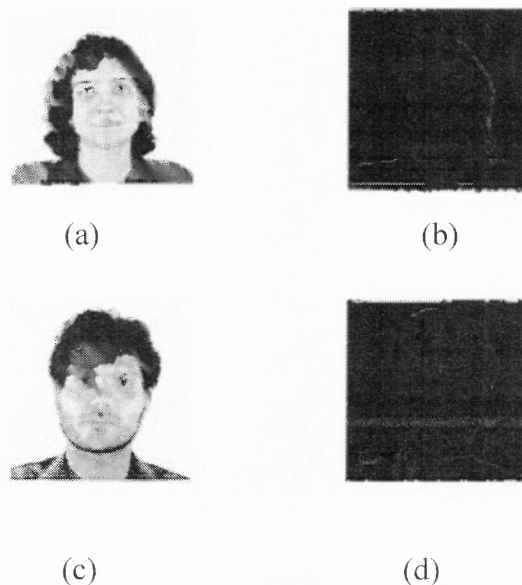


Figure 5.2 The outputs of Wechsler and Kidode's method. (a) and (c) are original images. (c) and (d) are edge detection images.

5.1.2.3 Locate Facial Features. In order to identify facial features, first the facial feature candidates are selected. The candidates are extracted by overlaying the face boundary obtained from the previous section on the binary edge image, and converting all the white pixels in the binary edge image that are on or outside the face boundary to black. After that, x- and y-projections are applied to locate facial features. In the candidate image, x-projection is used to obtain the facial features' horizontal locations and y-projection is used to obtain their vertical locations. By combining the horizontal and vertical locations, four rectangular boxes are obtained: two for eyes, one for nostrils, and one for mouth.

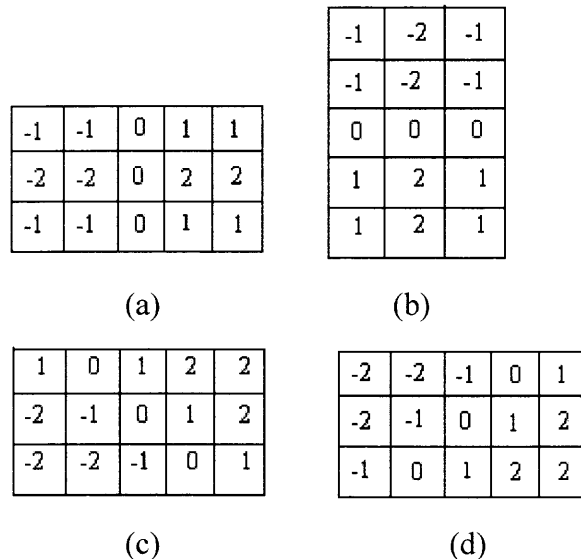


Figure 5.3 Edge detection masks: (a) vertical mask, (b) horizontal mask, (c) 135° mask and (d) 45° mask.

5.1.2.4 Face Boundary Repairing. Sometimes, the chin edge is too low-contrasted to be completely detected by edge detection. Two approaches are performed to repair the chin line of face boundary. In the first approach, the center point of chin is utilized (i.e., the average point of the available chin edge pixels) as the initial point in the grayscale

image to trace the chin line. The algorithm for tracing the chin line is described below.

(1) From the center point of the chin to its right:

Let the center point of the chin be $f(x, y)$, $\max \{f(x+1, y+1), f(x+1, y), f(x+1, y-1)\}$ are chosen to be the next connected point and repeat the procedure until it reaches the right part of the face boundary. As an example, Figure 5.4 (a) shows an image of facial features candidates, Figure 5.4 (b) shows the unrepaired face boundary, and Figure 5.4 (c) shows the repaired face boundary.

(2) From the center point of the chin to its left:

$\max \{f(x-1, y+1), f(x-1, y), f(x-1, y-1)\}$ are chosen to be the next connected point and the procedure is repeated until it reaches the left part of the face boundary.

In the second approach, an elliptic model is adopted where the ratio of the major axis to minor axis is 1.5. Figure 5.5 (a) and (c) illustrate the missing chin lines due to the face shadows. By using the lower-end points of the left and right face boundaries, the horizontal distance of the two end points can be computed as the minor axis and fit in the elliptic model to repair the chin line. Figure 5.5 (b) and (d) show the results after repairing.

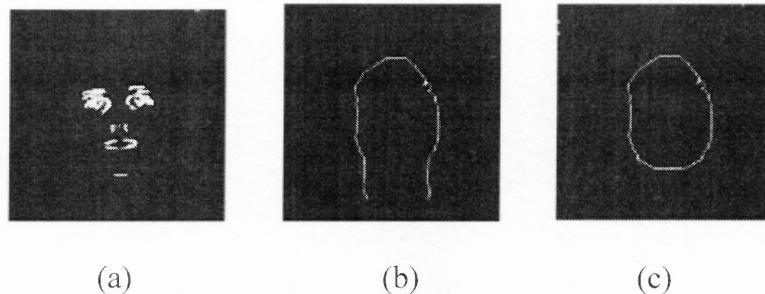


Figure 5.4 (a) An image of facial features candidates, (b) the unrepaired face boundary and (c) shows the repaired face boundary.

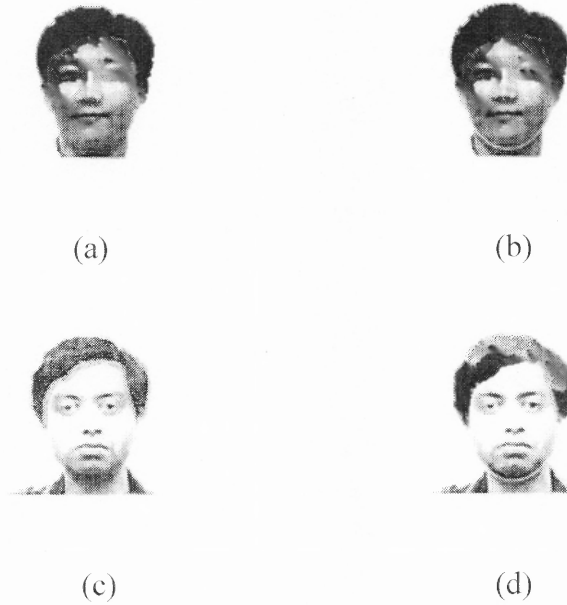


Figure 5.5 The examples of repairing (a) and (c) are images before repairing (b) and (d) are images after repairing.

5.1.3 Finding Facial Features Based on Geometric Face Model

5.1.3.1 Geometric Face Model. Sometimes, the facial features candidates are too clustered, the x- and y-projections cannot work well. Under these circumstances, the geometric face model is applied to locate facial features. The model utilizes the configuration among eyes, nostrils, and mouth since in most of faces, the vertical distances between eyes and nose and between eyes and mouth are proportional to the horizontal distance between the two centers of eyes.

Referring to Figure 5.6, let the distance between the two centers of eyes be D . The geometric face model (Figure 5.6) and related distances are described below [63].

- (1) The vertical distance between two-eyes and the center of mouth is D .
- (2) The vertical distance between two-eyes and the center of the nostrils is $0.6D$.
- (3) The width of the mouth is D .

(4) The width of nose is $0.8D$.

(5) The vertical distance between eyes and eyebrows is $0.4D$.

The procedure of locating facial features by the geometric face model is illustrated in Figure 5.7.

Figure 5.8 shows three examples of using the geometric face model. Figures 5.8 (a) is the original images. Figures 5.8 (b) is the facial features candidates obtained by using the method in Section 5.2. Because the facial feature candidates are clustered, x- and y-projections are performed. Figures 5.8 (c) is the facial features candidates after Sobel operation and thresholding. Figures 5.8 (d) shows the results after applying the geometric face model.



Figure 5.6 Geometric face model.

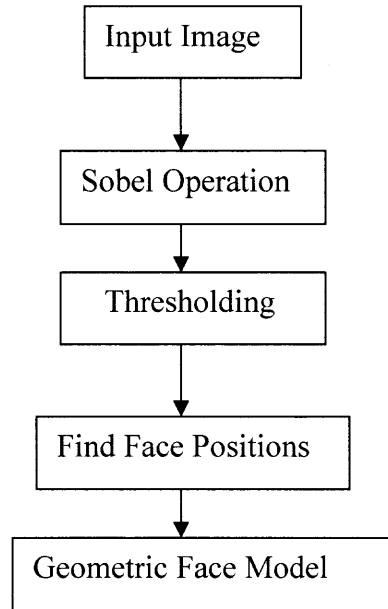


Figure 5.7 The procedure of finding facial features.

5.1.3.2 Geometrical Face Model Based on Gabor Filter. For particularly difficult cases such as poor lighting and shadows, two eyes can be located using Gabor filter [24, 34] and then apply the elliptic model to extract face boundary. The three steps to locate two eyes are described as follows:

- (1) Apply Gabor filter to the original image
- (2) Apply Gaussian weighting to the filtered image
- (3) Locate peaks in the image and find the locations of eyes

After two eyes are located, the geometric face model is applied to extract other facial features. Let the distance between two eyes be D . The center of the ellipse is placed at $0.4D$ below the midpoint of the two eyes. The length of the major axis is given $3D$ and of the minor axis is given $2D$.

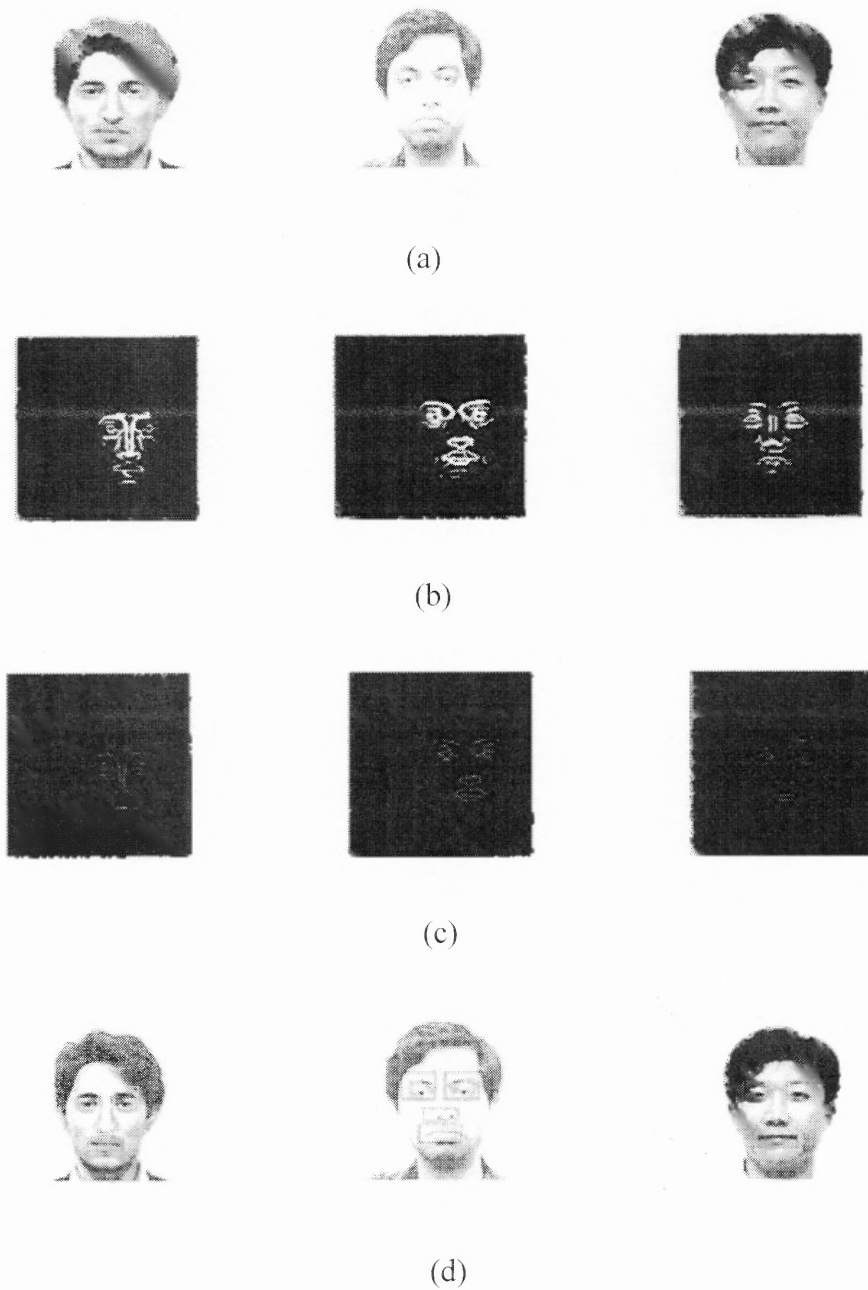


Figure 5.8 Examples of geometric face model (a) original images (b) facial features candidates after thresholding (c) facial features candidates after Sobel operation (d) extractions of facial features.

5.1.4 Experimental Results

Figure 5.9 shows the results of using the double-threshold method. Figure 5.9 (a) is the original image, (b) is the edge detection image, and (c) and (d) are high and low thresholded images respectively. Figure 5.9 (e) is the head boundary and (f) is the face boundary. Figure 5.9 (g) is the facial features candidates. Figure 5.9 (h) is the rectangular boxes for facial features. Figure 5.9 (i) shows the overlay of the head and face boundaries with the original image. Figure 5.9 (j) shows the overlay of facial features with the original image. The proposed algorithms are tested on more than 100 FERET face images. A partial set of our experimental results using the geometric face model is shown in Figure 5.10. By comparing with [63], the proposed approaches demonstrate better results in head boundary tracing, chin boundary repairing, and facial features detection. The programs using MATLAB are implemented on SUN ULTRA 60 workstation. The execution time is about 170 s including 15 s for the geometric face model to process a 384 by 256 image on FERET face database.

5.1.5 Conclusions

A novel approach is presented for the extraction of human head, face and facial features. Two thresholds are used: the high-thresholded image for tracing head boundary and the low-thresholded image for scanning face boundary. X- and y-projections are applied to extract facial features such as eyes, nostrils and mouth. Because low contrast of chin occurs in some face images, its boundary cannot be completely detected. An elliptic model is used to repair it. Because of noises or clustered facial features candidates, geometric face model is implemented to locate facial features and an elliptic model to trace face boundary. The Gabor filter algorithm is adopted to locate the positions of two

eyes. The proposed algorithms are tested on more than 100 FERET face images. Experimental results show that the proposed algorithm can perform the extraction of human head, face and facial features successfully.

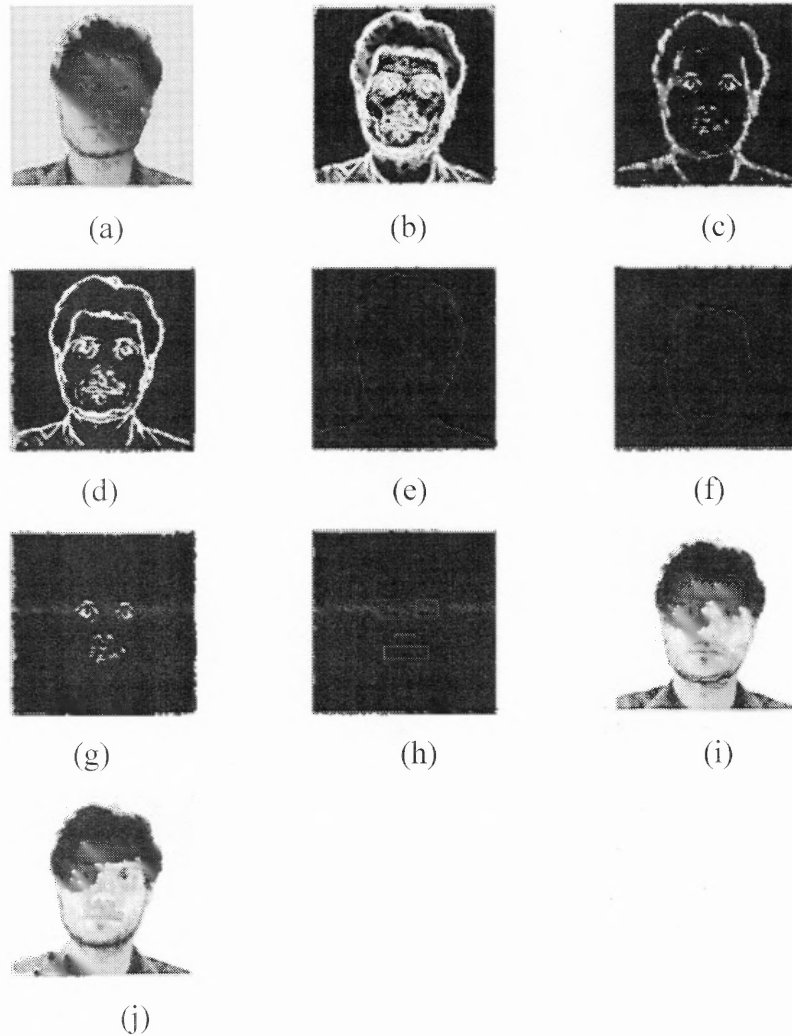


Figure 5.9 The experimental results of double-threshold method (a) original image (b) edge detection image (c) high thresholded image (d) low thresholded image (e) head boundary (f) face boundary (g) facial features candidates (h) rectangular boxes for facial features (i) detection of contours (j) extraction of facial features.

In this chapter, only frontal-view face images are considered that have not glasses. When the face is tilted intensively or rotated greatly, our two-dimensional geometric model cannot fit the facial features completely. Under the circumstance, the future work is emphasized on using a three-dimensional model to adjust for the three-dimensional rotations of faces.

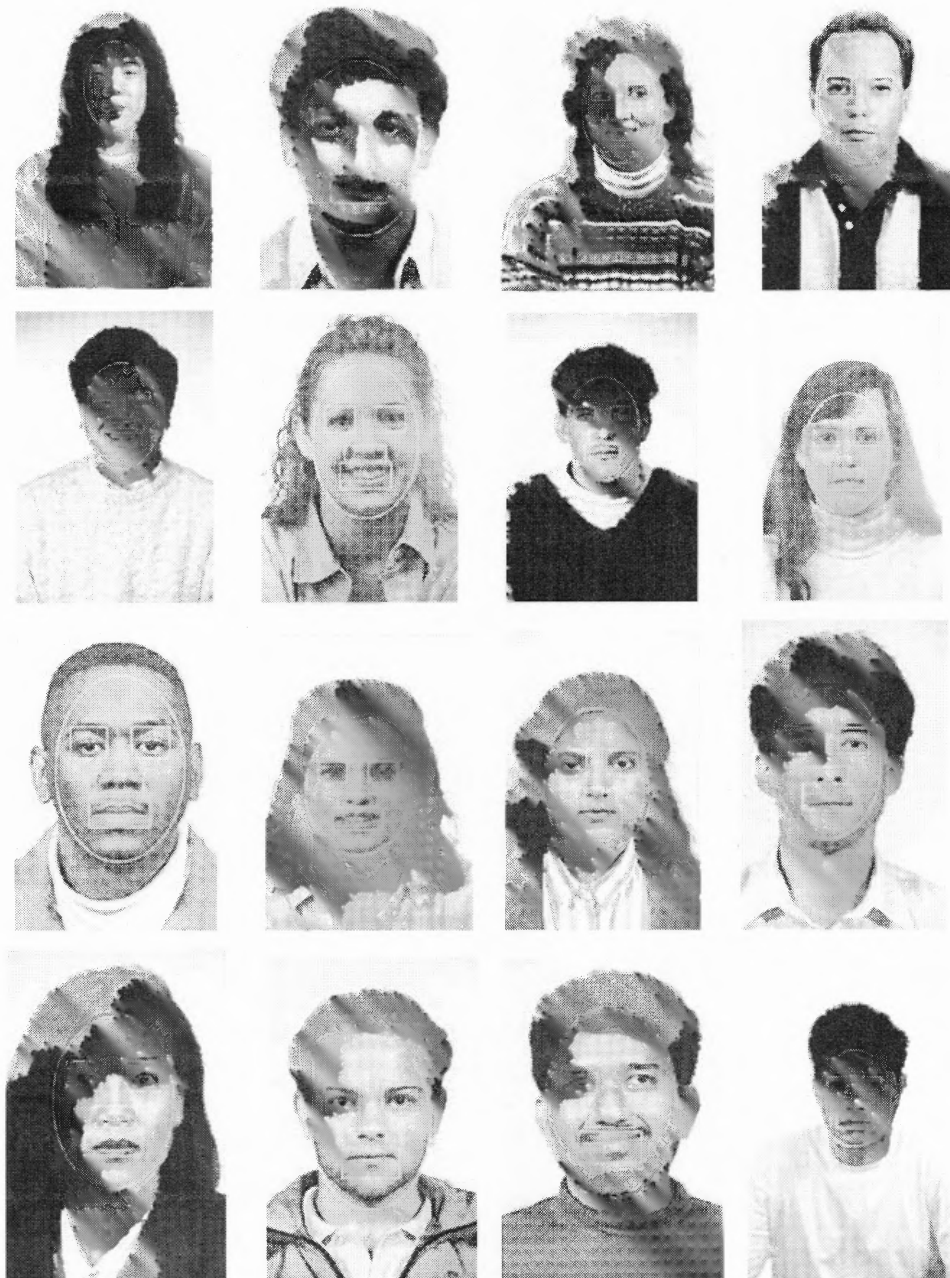


Figure 5.10 Partial experimental results.



Figure 5.10 Partial experimental results (cont.).

5.2 Facial Feature Detection (Eyes Detection)

In this session, locating the eyes automatically in the single-face frontal and profile (15° and 25°) images is implemented. The proposed procedure that makes use of edge information as the potential candidates of the eyes is a novel research. FERET face database is tested for the proposed procedure and the experimental results are promising.

5.2.1 Introduction

Researchers have endeavored to develop the eye detection algorithms in the last decade. Lam and Yan [49] proposed a method for locating and extracting eyes based on snake, corner detection, and cost functions. There are disadvantages in their algorithm, such that snake requires manual initialization to locate head boundary and their system can only handle the plain background and one-person images. Feng and Yuen [25] proposed a three-cue eye detection algorithm. The three cues are the relatively low intensity of the eye regions, the direction of the line connecting the centers of eyes, and the response of convolving the eye variance filter with the face image. They experimented on face images from MIT AI Laboratory to obtain the accuracy rate 92.5%. However, snake's manual initialization and handling only one-person images are their shortcomings.

Huang and Wechsler [38] developed their hybrid approach using optimal wavelet packets for eyes representation and Radial Basis Functions Neural Networks (RBFNN) for classification between eye and non-eye regions. There are some interesting results from their experiments. If they directly feed the images into RBFNN and do not use the optimal wavelet packets to reduce the dimensionality of the original image, the correct rate for eye detection is 70% and for non-eye detection is also 70%. If the test images are preprocessed by optimal wavelet packets and are reconstructed under the best 16 or 64 wavelet coefficients using Shannon entropy minimization, the performance becomes better. The correct rate for eye detection is 82.5% and for non-eye is 80% under the best 16 wavelet coefficients. Under the best 64 wavelet coefficients, the correct rate for eye is 85% and for non-eye is 100%. However, they only used RBFNN to discriminate eyes from non-eyes and did not test it on the whole face images for eye detection.

Peng *et al.* [66] went a step further by using RBFNN to locate the eye locations in the face images. In their experiments, 97.41% of the test samples received the correct detection results, with 1.69% of the test samples being rejected. Huang *et al.* [37] detected eyes using Support Vector Machine (SVM). In their experiments, the best result is achieved by using polynomial kernel at degree of 2. The correct rate is 96%. From the results, SVM outperforms RBFNN. Similar to [72], they just applied SVM to discriminate eyes from non-eyes and did not test on face images. In this section, SVM is adopted to locate the eye positions in the face images. Moreover, more statistical methods on eye detection can be referred in [17, 67]

5.2.2 Experimental Procedure

The experimental procedure is illustrated on Figure 5.11. After original FERET face images are fed into the proposed procedure, the Canny edge detection is performed to find out the potential candidates of eyes features. SVM is used to verify the candidates of facial features, and finally confirm the facial feature candidates by some rules. There are two experimental results on this session. First experiment only distinguishes eyes from non-eyes. The results are presented in section 5.2.2.1. In second experiment, the proposed procedure is applied to the FERET face database to examine the performance. The experimental results are illustrated in section 5.2.2.2.

5.2.2.1 The Classification of Eyes and Non-eyes. In this section, eyes and non-eyes classification is performed and the training procedure for eyes using SVM is discussed. Note that the training of mouth also follows the same procedure as eyes training. For training of SVM, 800 eyes and 10000 non-eyes from the FERET face database are selected. According to the observed approximated ratio of height-to-width of eyes, i.e., 1:

2, the size of eyes and non-eyes is fixed to be 19 by 39 pixels. In SVM, the bigger training set does not guarantee the better classification result. 5 sizes of the training set: 3000, 4000, 6000, 8000 and 10000 samples are tested. After the training stage, their classification accuracy are tested by means of different data. In these experimental results, 4000 training set is suitable and obtain the best performance among different training sets. The accuracy is 99.85%.

The processing above is time-consuming because the dimensionality of each sample is huge. In order to reduce the dimensionality and in the meantime to maintain as much information as possible, the technique of PCA is adopted. Various sizes of eigenvalues: 20, 30, 40, 50, 60, 70, and 80 eigenvalues, are tested in order to search for the best dimensionality of samples that are based on the aforementioned 4000 training samples. New unseen 4000 samples including 200 eyes and 3800 non-eyes are examined to measure the performance of SVM. From the experimental results, it is observed that the more eigenvalues used produce the higher accuracy rates as it is expected. To reduce the computational time and data storage, but still maintain as high accuracy rate as possible, the first large 60 eigenvalues are selected as the dimensionality for eyes and non-eyes.

Furthermore, different degrees of polynomial kernel functions: 2, 3 and 4 using 60 eigenvalues on 4000 samples are explored, and the accuracy rates respectively are 98.42%, 98.48%, and 98.51%. From the experimental results, the performance of linear SVM (98.52%) is almost the same as that of nonlinear SVM with degree = 4 (98.51%). It indicates that the linear SVM can be used to save the computational time. Compared our results with reference [37], the proposed method can obtain better performance.

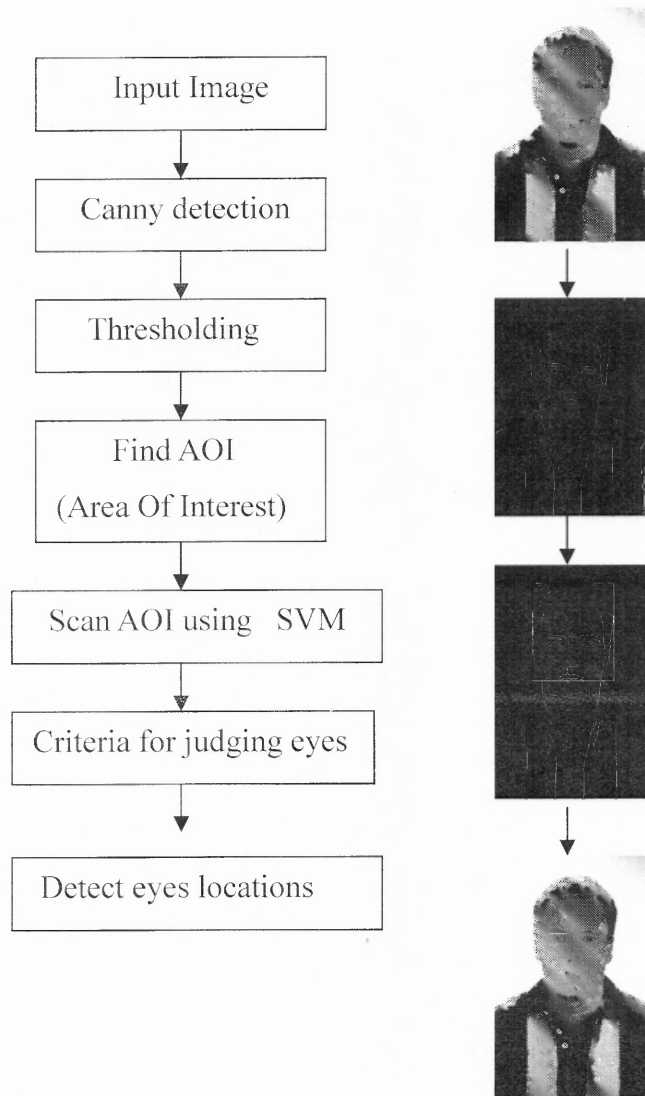


Figure 5.11 The procedure for detecting eyes.

5.2.2.2 Eye Detection on FERET Face Database. As illustrated in Figure 5.11, after original FERET face images are fed into the proposed procedure, the Canny edge detection is performed to find out the potential area of interest (AOI). Because the information of head boundary can be displayed by the canny edge detection, the AOI can be easily located in terms of head boundary information. Since facial features such as eyes and mouth are darker than other parts on the face, it must be existed some edges

around the facial features in the AOI. In order to locate eyes correctly and efficiently and to avoid unnecessary false positive, trained SVM is implemented to scan the edge points only inside the AOI. In some cases, the edge points appear near eyes, and the center of the eyes does not present by the edge points. In order to remedy this situation, 1 pixel up and down in the vertical positions of the edge point is included. For example, if the position of an edge pixel is (x, y) , the positions of $(x+1, y)$ and $(x-1, y)$ are also verified. The trained SVM is used to scan all the eyes candidates to locate the true location of eyes.

After the scan procedure, there may still have false positives. Some rules are investigated to eliminate the false positives. Sobottka and Pitas [76] proposed some requirements and Feng and Yuen [25] proposed some rules for validating eyes. Both requirements are combined to develop the following rules to remove false positives.

1. The two eyes have similar gray level values.
2. Basically, the distance between the two eyes is within a certain range of the head width.
3. Generally, the line connecting the two eyes is horizontal or within some angle ranges.

The experimental results are illustrated in Figure 5.12. From Figure 5.12, the proposed procedure not only can locate the eyes area correctly but also can detect the face images with right and left profile including 15° and 25° .

5.2.3 Conclusions

In this section, a feasible eye detection system is built. The presented eye detection technique combines image processing technique and Support Vector Machine (SVM). Compared with reference [37] using the polynomial kernel with degrees 2, 3, and 4 respectively, the proposed method shows the higher eye detection rates.



Figure 5.12 The partial experimental results.

CHAPTER 6

SUMMARY AND FUTURE RESEARCH

Facial expressions are the facial feature changes in response to a person's internal emotional states, intentions, cognitive process, or social interactions and communications. Facial expression recognition refers to make use of computer systems that attempt to automatically analyze and recognize facial motions and facial features changes from visual information including still images and video sequences. Automatic facial expression recognition system has a number of potential applications, such as emotion and paralinguistic communication, clinical psychology, medical care and cure, psychiatry, neurology, pain assessment, lie detection, human computer interface.

6.1 Summary

In this dissertation, different components of an automatic facial expression recognition system are discussed. For face detection, two methods are developed to locate the face and eyes. For feature extraction and representation, various techniques, such as PCA, ICA, LDA, 2D-PCA, and 2D-LDA, are explored to find out the suitable representation for the underlying information. For facial expression recognition, classifiers, such as RBF, SVM, PCA, and LDA, are performed to check their recognition abilities. A robust recognition system has the ability of handling homogeneous and heterogeneous database. In this dissertation, homogeneous database (e.g., Japanese female facial expression dataset, JAFFE) and heterogeneous database (e.g., Cohn-Kanade AU-coded face expression image database) are used. JAFFE is used to classify seven prototypical and emotion-specified expressions such as happy, surprise, sad, disgust, fear, angry, and

neutral. Cohn-Kanad database is used to recognize the action units, the subtle changes of facial expressions.

Time complexity and recognition rate are two main factors in performance evaluation of a proposed facial expression recognition algorithm. The recognition rate of the proposed method applying to JAFFE is 95.71% by using leave-one-out strategy and 94.13% by using cross-validation strategy. It takes only 0.0357 second for processing one input image of size 256×256 . The recognition rate of the proposed system applying to Cohn-Kanade AU-coded face image database is 97.06% for recognizing the upper part of face, 97.13% for lower part of face, and 100% for the whole face. It takes only 1.8 ms for processing one test image of size 40×90 . Both systems are reliable, efficient, and can be applied to real-time applications.

In this dissertation, a machine recognition and assessment system of neonatal expressions of pain is developed to assist clinicians in diagnosing pain. The Infant COPE database is used. This database is designed and developed by Dr. Brahnam at Missouri State University for the Infant COPE project that is led and supervised by Dr. Brahnam. The Infant COPE research of the neonatal pain classification is a first attempt at applying state-of-the-art face recognition technologies to actual medical problems. Medical applications of face recognition technologies have been suggested but not tried with actual medical data. Two different protocols, person-dependent and person-independent, are investigated.

The objective of Infant COPE project is to bypass these observational problems by developing a machine classification system to diagnose neonatal facial expressions of pain. Since assessment of pain by machine is based on pixel states, a machine

classification system of pain will remain objective and will exploit the full spectrum of information available in a neonate's facial expressions. Furthermore, it will be capable of monitoring a neonate's facial expressions when the patient is left unattended and will not degrade over time. The results indicate that the application of face classification techniques in pain assessment and management is a promising area of investigation.

6.2 Future Research

6.2.1 Future Research on Pattern Recognition

Time complexity and recognition rate are two main factors in performance evaluation of a proposed pattern recognition algorithm. The two main factors are influenced by pre-processing, face detection, feature extraction, feature selection, and classification methods. The future research focuses on two issues: one is to select the most important features, and the other is to search more useful feature extraction methods. The purpose is to increase the accuracy rate and decrease computational time simultaneously.

Feature extraction and selection are two techniques for picking up the most important features from the original variables to reduce dimensionality. Feature extraction provides a transformation of the original variables such that the resulting features are a set of new variables in the transformed space. PCA, LDA, and ICA are the examples of feature extraction techniques. Feature selection focuses on selecting features in the original space such that the chosen features are a subset of the original variables. Genetic algorithm, sequential forward selection, and sequential backward elimination are the examples of feature selection.

Although different feature extraction methods such as PCA, LDA, and ICA are investigated, they are in spatial domain. The feature extraction methods in frequency domain, such as discrete Fourier transform and discrete cosine transform, will be the main research direction.

In this dissertation, various feature extraction methods such as PCA, LDA, and ICA are investigated but not the feature selection algorithms. In the future research, the influence of different feature selection techniques, such as genetic algorithm and sequential forward selection, will be explored on the performance of the pattern recognition algorithm. The promising concept may be combining the feature extraction and selection approaches to form the best features for the input variables. For example, the feature extraction of PCA can be fused with the feature selection of genetic algorithm to form representative features for the original variables.

6.2.2 Future Research on Neonatal Pain Classification

Since there is some evidence that the temporal frequency and intensity information in cry can discriminate pain, combining sound classifiers with face recognition is yet another area of potential research.

The neonate face database is color images. In this dissertation, only grayscale space is used by transforming the color images to grayscale images directly. In the future research, different color space such as $YCbCr$ and HSV will be explored to check whether the color information is useful or not for our developed system.

REFERENCES

1. Abboud, B., Davoine, F., and Dang, M., "Statistical modeling for facial expression analysis and synthesis," *Proc. IEEE International Conference on Image Processing*, pp. 653-656, September 2003.
2. Ahlberg, J., "Facial feature extraction using deformable graphs and statistical pattern matching," *Proc. SSAB Annual Symposium on Image Analysis*, Göteborg, Sweden, pp. 89-92, March 1999.
3. Ahlberg, J., "A system for face localization and facial feature extraction," *Technical report no. LiTH-ISY-R-2172*, Linköping University, July 1999.
4. Ambuel, B., Hamlett, K. W., Marx, C. M., and J. L. Blumer, "Assessing distress in pediatric intensive care environments: The COMFORT scale," *Journal of Pediatric Psychiatry*, vol. 17, pp. 95-109, 1992.
5. Bartlett, M. S., Movellan, J. R., Sejnowski, T. J., "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2000.
6. Bazzo, J. J. and Lamar, M. V., "Recognizing facial actions using gabor wavelets with neutral face average difference," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.
7. Belhumeur, P., Hespanha, J., and Kriegman, D., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no. 7, pp. 711-720, July 1997.
8. Brahnam, S., Chuang, C., Sexton R. S. and Shih, F.Y., "Machine assessment of neonatal facial expressions of acute Pain," *Special Issue on "Decision Support in Medicine in Decision Support System*, in press.
9. Brahnam, S., Chuang, C., Shih, F.Y., and Slack, M.R., "Machine recognition and representation of neonatal facial displays of acute pain," *Artificial Intelligence in Medicine*, (in press and available on publisher website).
10. Broomhead, D. S. and Lowe, D., "Multivariable functional interpolation and adaptive networks," *Complex Systems*, pp. 321-355, 1988.
11. Buchholz, M., Karl, H. W., Pomietto, M., and Lynn, A., "Pain scores in infants: A modified infant pain scale versus visual analogue," *Journal of Pain and Symptom Management*, vol. 15, no. 2, pp. 117-124, 1998.

12. Buciu, I., Kotropoulos, C., and Pitas, I., "ICA and Gabor representation for facial expression recognition," *Proc. IEEE International Conference on Image Processing*, pp. 855-858, September 2003.
13. Chen, X.-W. and Huang, T., "Facial expression recognition: A clustering-based approach," *Pattern Recognition Letter*, vol. 24, pp. 1295-1302, 2003.
14. Chien, J.-T. and Wu, C.-C. "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1644-1649, December 2002.
15. Coffman, S., Alvarez, Y., Pyngolil, M., Petit, R., Hall, C., and Smyth, M., "Nursing assessment and management of pain in critically ill children," *Heart & Lung: The Journal of Acute and Critical Care*, vol. 26, no. 3, pp. 221-228, 1997.
16. Comon, P., "Independent component analysis, A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287-314, April 1994.
17. Cootes, T. F., Edwards, G.J., and Taylor, C.J., "Active appearance model," *Proc. 5th European Conference on Computer Vision*, Freiburg, Germany, 1998.
18. Cottrell, G. W. and Fleming, M. K., "Face recognition using unsupervised feature extraction," *Proc. International Conference on Neural Networks*, pp. 322-325, 1990.
19. Dai, Y., Shibata, Y., Ishii, T., Katamachi K., Noguchi, K., Kakizaki, N., and Cai, D., "An associate memory model of facial expression and its application in facial expression recognition of patients on bed," *Proc. IEEE International Conference on Multimedia and Expo*, pp. 772-775, August 2001.
20. Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., and Sejnowski, T. J., "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-985, 1999.
21. Dubuisson, S., Davoine, F., and Masson, M., "A solution for facial expression representation and recognition," *Signal Processing: Image Communication*, vol. 17, pp. 657-673, 2002.
22. Ekman, P. and Friesen, W. V., "The facial action coding system: A technique for the measurement of facial movement," *San Francisco: Consulting Psychologists Press*, 1978.
23. Ekman, P., Hager, J., Methvin, C. H., and Irwin, W., "Ekman-Hager Facial Action Exemplars," unpublished data, Human Interaction Laboratory, Univ. of California, San Francisco.

24. Fasel, I. R, Smith, E. C., Bartlett, M. S., and Movellan, J. R., "A comparison of Gabor filter methods for automatic detection of facial landmarks," *Proc. IEEE International Conference on Automatic face and Gesture Recognition*, pp. 231-235, May 2002.
25. Feng, G. C. and Yuen, P. C., "Multi cues eye detection on gray intensity image," *Pattern Recognition*, vol. 34, no. 5, pp. 1033-1046, 2001.
26. Fitzgerald, M., "The developmental neurobiology of pain," *Proc. The VIth World Congress on Pain*, pp. 253-261, Oxford, 1991.
27. Franck, L. S. and Miaskowski, C., "Measurement of neonatal responses to painful stimuli: A research review," *Journal of Pain and Symptom Management*, vol. 14 no. 6, pp. 343-378, December 1997.
28. Gagliardi, C., Frigerio, E., Buro, D. M., I. Cazzaniga, Pret, D., and Borgatti, R., "Facial expression recognition in Williams Syndrome," *Neuropsicologia*, vol. 41, pp. 733-738, 2003.
29. Gilbert, C. A., Lilley, C. M., Craig, K. D., McGrath, P. J., Court, C. A., Bennett, S. M., and Montgomery, C. J., "Postoperative pain expression in preschool children: Validation of the child facial coding system," *The Clinical Journal of Pain*, vol. 15, no. 3, pp. 192-200, 1999.
30. Grunau, R. E., Grunau, R. V. E., and Craig, K. D., "Pain expression in neonates: Facial action and cry," *Pain*, vol. 28, no. 3, pp. 395-410, 1987.
31. Grunau, R. V. E., Johnston, C.C., and Craig, K.D., "Neonatal facial and cry responses to invasive and non-invasive procedure," *Pain*, vol. 42, no. 3, pp. 295-305, 1990.
32. Guo, G, Li, S. Z. and Chan, K. L., "Support vector machines for face recognition," *Image and Vision Computing* , vol. 19, pp. 631-638, 2001.
33. Harrison, D., Evans ,C., Johnston, L., and . Loughnan, P, "Bedside assessment of heel lance pain in the hospitalized infant," *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, vol. 31, no. 5, pp. 551-557, 2002.
34. Hjelmås, E., "Feature-based face recognition," *Proc. Norwegian Image Processing and Pattern Recognition Conference*, 2000.
35. Hsu, C.-W. and Lin, C.-J., "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, March 2002.
36. Hu, J., Yan, H., and Sakalli, M., "Locating head and face boundaries for head-shoulder images," *Pattern Recognition*, vol. 32, pp. 1317-1333, 1999.

37. Huang, J., Li, D., Shao, X., and Wechsler, H., "Pose discrimination and eye detection using support vector machines (SVM)," *Proceeding: NATO-ASI on Face Recognition: From Theory to Applications*, pp. 528-536, 1998.
38. Huang, J. and Wechsler, H., "Eye detection using optimal wavelet packets and radial basis functions (RBFs)," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 7, pp. 1009-1026, 1999.
39. Hultgren, M. S., "Assessment of postoperative pain in critically ill infants," *Progress in Cardiovascular Nursing*, vol. 5, no. 3, pp. 104-112, 1990.
40. Hyvärinen, A. and Oja, E., "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, June 2000.
41. Izard, C. E., Huebner, R. R., Risser, D., McGinnes, G. C., and Dougherty, L. M., "The young infant's ability to produce discrete emotion expressions," *Developmental Psychology*, vol. 16, pp. 418-426, 1980.
42. Jeng, S.-H., Liao, H. Y. M., Han, C. C., Chern, M. Y., and Liu, Y. T., "Facial feature detection using geometrical face model: An efficient approach," *Pattern Recognition*, vol. 31, pp. 273-282, 1998.
43. Jutten, C. and Herault, J., "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp 1-10, July 1991.
44. Kanade, T., Cohn, J., and Tian, Y., "Comprehensive database for facial expression analysis," *Proceedings of International Conference on Face and Gesture Recognition*, pp. 46-53, Mar. 2000.
45. Kass, M., Witkin, A., and Terzopoulos, D., "Snakes: Active contour models," *International Journal Computer Vision*, pp. 321- 331, 1988.
46. Kirby, M. and Sirovich, L., "Applications of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103-108, 1990.
47. Kosugi, M., "Human-face search and location in a scene by multi-pyramid architecture for personal identification," *Systems and Computers in Japan*, vol. 26, no. 6, pp. 27-38, 1995.
48. Krechel, S. W. and Bilder, J., "CRIES: A new neonatal postoperative pain measurement score: Initial testing of validity and reliability," *Paediatric Anaesthesia*, vol. 5, no. 1, pp. 53-61, 1995.

49. Lam, K.-M. and Yan H., "Locating and extracting the eye in human face images," *Pattern Recognition*, vol. 29, no. 5, pp. 771-779, 1996.
50. Levine, J. D. and Gordon, N. C., "Pain in prelingual children and its evaluation by pain-induced vocalizations," *Pain*, vol. 14, pp. 85-93, 1982.
51. Li, M. and Yuan, B., "2D-LDA: A statistical linear discriminant analysis for image matrix," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 527-532, 2005.
52. Lin, C. and Fan K.-C., "Triangle-based approach to the detection of human face," *Pattern Recognition*, vol. 34, pp. 1271-1284, 2001.
53. Lindh, V., Wiklund, U., and Håkansson, S., "Heel lancing in term new-born infants: An evaluation of pain by frequency domain analysis of heart rate variability," *Pain*, vol. 80, pp. 143-148, 1999.
54. Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., and Movellan, J. Dynamics of Facial Expression Extracted Automatically from Video. The First IEEE Workshop on Face Proc-essing in Video. Washington, DC. (2004)
55. Liu, C., "Enhanced independent component analysis and its application to content based face image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34, no. 2, April 2004.
56. Lyons, M. J., Budynek, J., and Akamatsu, S., "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357-1362, December 1999.
57. Martinez, A. M. and Kak, A. C., "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.
58. McGrath, P. A., *Pain in children: Nature, assessment and treatment*, Guildford Press, New York, 1989.
59. Merkel, S. I., Voepel-Lewis T., and Malviya S., "Pain assessment in infants and young children: The FLACC scale," *American Journal of Nursing*, vol. 102, no. 10, pp. 55-58, 2002.
60. Merkel, S. I., Voepel-Lewis, T., Shayevitz, J. R., and Malviya, S., "The FLACC: A behavioral scale for scoring postoperative pain in young children," *Pediatric Nursing*, vol. 23, no. 3, pp. 293-297, 1997.
61. Moghaddam, B. and Yang, M.-H., "Gender classification with support vector machines," *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 306-311, March 2000.

62. Morales, E. and Shih, F. Y., "Wavelet coefficients clustering using morphological operations and pruned quadrees," *Pattern Recognition*, vol. 33, no. 10, pp. 1611-1620, October 2000.
63. Nikolaidis, A. and Pitas, I., "Facial feature extraction and pose determination," *Pattern Recognition*, vol. 33, pp. 1783-1791, 2000.
64. O'Toole, A. J., Abdi, H., Deffenbacher, K. A., and Bartlett, J.C., "Classifying faces by race and sex using an autoassociative memory trained for recognition," *Proc. 13th Annual Conference on Cognitive Science*, Hillsdale, NJ, pp. 847-851, 1991.
65. O'Toole, A. J. and Deffenbacher, K. A., "The perception of face gender: The role of stimulus structure in recognition and classification," *Memory & Cognition*, vol. 26, no. 1, pp. 146-160, 1997.
66. Peng, H., Zhang, C., and Bian, Z., "Human eyes detection using hybrid neural method," *Proc. International Conference on Signal Processing*, 1998.
67. Pentland, A., Moghaddam, B., Starner, T., Oliyide, O., and Turk, M., "View-Based and Modular Eigenspaces for Face Recognition," *Technical Report 245*, MIT Media Lab, 1993.
68. Prkachin, K. M., Solomon, P., Hwang, T., and Mercer, S. R., "Does experience influence judgments of pain behaviour? Evidence from relatives of pain patients and therapists," *Pain Research & Management*, vol. 6, no. 2, pp. 105-12, 2001.
69. Rowley, H., Baluja, S., and Kanade, T., "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.
70. Ryu, Y.-S. and Oh, S.-Y., "Automatic extraction of eye and mouth field from a face image using eigenfeatures and multilayer perceptrons," *Pattern Recognition*, vol. 34, pp. 2459-2466, 2001.
71. Samal, A. and Iyengar, P. A., "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern Recognition*, vol. 25, pp. 65-77, 1992.
72. Schneiderman, H. and Kanade, T., "Probabilistic modeling of local appearance and spatial relationships for object recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 45-51, 1998.
73. Shinohara, Y. and Otsu, N., "Facial expression recognition using fisher weight maps," *Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

74. Sirovich, L. and Kirby, M., "Low dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, vol. 4, no. 3, pp. 519-524, 1987.
75. Sobottka, K., and Pitas, I., "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.
76. Sobottka, K. and Pitas, I., "Looking for faces and facial features in color images," *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, Russian Academy of Sciences, vol. 7, no. 1, 1997.
77. Sprengelmeyer, R., Young, A.W., Mahn, K., Schroeder, U., Woitalla, D., Butter, T., Kuhn, W., and Przuntek, H., "Facial expression recognition in people with medicated and unmedicated parkinson's disease," *Neuropsychologia*, vol. 41, pp. 1047-1057, 2003.
78. Stevens, B. J., Johnston, C., Petryshen, P., and Taddio, A., "Premature infant pain profile: Development and initial validation," *The Clinical Journal of Pain*, vol. 12, no. 1, pp. 13-22, 1996.
79. Sung, K. and Poggio, T., "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39-51, 1998.
80. Thomaz, C. E., Gillies, D. F., and Feitosa, R. Q., "Using mixture covariance matrices to improve face and facial expression recognitions," *Pattern Recognition Letter*, vol. 24, pp. 2159-2165, 2003.
81. Tian, Y.-L., Kanade, T., and Cohn, J. F., "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, 2001.
82. Turk, M.A. and Pentland, A.P., "Face recognition using eigenfaces," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586-591, Silver Spring, MD, 1991.
83. Turk, M. A. and Pentland, A. P., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
84. Valentin, D., Abdi, H., O'Toole, A. J., and Cottrell, G. W., "Connectionist models of face processing: A survey," *Pattern Recognition*, vol. 27, no. 9, pp. 1209-1230, 1994.

- 85 Van Cleve, L., Johnson, L., and Pothier, P., "Pain responses of hospitalized infants and children to venipuncture and intravenous cannulation," *Journal of Pediatric Nursing*, vol. 11, no. 3, pp. 161-168, 1996.
- 86 Vapnik, V.N., *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- 87 Volia, P. and Jones, M., "Robust real-time object detection," Proceedings of International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, 2001.
- 88 Warnock, F. and Sandri, D., "Comprehensive description of newborn distress behavior in response to acute pain (newborn male circumcision)," *Pain*, vol. 107, no. 3, pp. 242-255, 2004.
- 89 Wechsler, H. and Kidode, M., "A new edge detection technique and its implementation," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-7, no. 12, pp. 827-835, December 1977.
- 90 Williams, R. L., Karacan, I., and Hirsch, C. J., *Electroencephalography (EEG) of human sleep: Clinical applications*, Wiley, New York, 1974.
- 91 Wong, K.-W., Lam, K.-M., and Siu, W.-C., "An efficient algorithm for human face detection and facial feature extraction under different conditions," *Pattern Recognition*, vol. 34, pp. 1993-2004, 2001.
- 92 Xavier Balda, R., Guinsburg, R., de Almeida, M. F., Peres, C., Miyoshi M. H., and Kopelman, B. I., "The recognition of facial expression of pain in full-term newborns by parents and health professionals," *Archives of Pediatrics & Adolescent Medicine*, vol. 154, no. 10, pp. 1009-1016, 2000.
- 93 Xu, C. and Prince, J. L., "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359-369, March 1998.
- 94 Yang, J., Zhang, D., Frangi, A. F., and Yang, J.-Y., "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26 no. 1, pp. 131-137, January 2004.
- 95 Yow, K. C. and Cipolla, R., "Enhancing human face detection using motion and active contours," *Proc. 3rd Asian Conference on Computer Vision*, vol. 1, pp. 515-522, Hong Kong, 1998.
- 96 Zhang, B.-L., Zhang, H., and Ge, S. S., "Face recognition by applying wavelet subband representation and kernel associative memory," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, January 2004.

97. Zhang, Z., Lyons, M., Schuster, M., and Akamatsu, S., "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 454-459, April 1998.
98. Zhu, Y., De Silva, L. C., and Ko, C. C., "Using moment invariants and HMM in facial expression recognition," *Pattern Recognition Letter*, vol. 23, pp. 83-91, 2002.