

國立交通大學

電機與控制工程學系

碩士論文

機器人之表情辨識快速學習法則



A Fast Learning Algorithm for Robotic  
Facial Expression Recognition

研究生：洪濬尉

指導教授：宋開泰 博士

中華民國九十六年七月

# 機器人之表情辨識快速學習法則

## A Fast Learning Algorithm for Robotic Facial Expression Recognition

研究生：洪濬尉

Student: Jung-Wei Hong

指導教授：宋開泰 博士

Advisor: Dr. Kai-Tai Song

國立交通大學

電機與控制工程學系



**Submitted to Department of Electrical and Control Engineering**

**College of Electrical and Computer Engineering**

**National Chiao Tung University**

**in Partial Fulfillment of the Requirements**

**for the Degree of Master**

**in**

**Electrical and Control Engineering**

**July 2007**

**Hsinchu, Taiwan, Republic of China**

中華民國九十六年七月

# 機器人之表情辨識快速學習法則

學生:洪濬尉

指導教授:宋開泰 博士

國立交通大學電機與控制工程學系

## 中文摘要

應用在機器人的表情辨識系統，會因為使用者呈現表情的方式有所不同，而產生系統無法辨識的新人臉表情。為了使機器人能夠適應新人臉之表情，本篇論文提出了一套能夠學習新臉孔之表情辨識系統。主要的想法是利用調整支持向量機(Support vector machine, SVM)的切割平面(Hyperplane)係數，來達到辨識新表情資料的目的。支持向量追蹤學習法(Support vector pursuit learning, SVPL)的概念被引入在高斯核空間(Gaussian kernel space)中來調整切割平面。為了加快訓練學習的速度，只有錯誤的表情資料和一定數量的關鍵舊集合被拿來重新訓練，藉以產生新的SVM分類器。經過調整切割平面後，不僅可以辨識之前無法辨識之新臉孔表情，並且還可以保持對舊有資料的辨識率。另外，我們使用蓋伯小波(Gabor wavelet)特徵擷取的方法來強化擷取之效能，以確保欲學習表情特徵值之正確性。所提出的表情學習演算法已成功的應用在實驗室之娛樂機器人的平台上，線上測試結果顯示即使是新的表情資料，也可以透過學習系統把辨識率由58%提升到81.3%，並且還可以對舊有表情資料保持78.7%之辨識率。

# **A Fast Learning Algorithm for Robotic Facial Expression Recognition**

Student: Jung-Wei Hong

Advisor: Dr. Kai-Tai Song

Department of Electrical and Control Engineering

National Chiao Tung University

## **ABSTRACT**

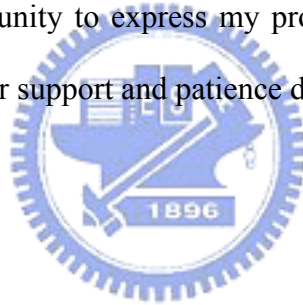
A robotic facial expression recognition system very often misclassifies data from a new face because different people may show their expressions in different ways. This thesis aims to study a facial expression recognition system that can learn new facial data and facilitate a robot to accommodate itself to various persons. The main idea of the proposed method is to adjust parameters of the hyperplane of support vector machine (SVM) for classifying new facial data. The concept of support vector pursuit learning (SVPL) is adopted to retrain the hyperplane in the Gaussian kernel space. To expedite the training procedure, we propose to retrain the new SVM classifier by using only samples classified incorrectly and the critical sets (CSs) from previous samples. After adjusting hyperplane parameters, the new classifier not only recognizes new facial data but also keeps acceptable performance of classifying previous data. Further, to obtain reliable facial features, we adopted Gabor wavelet to develop a feature extraction method in the system. The proposed algorithms have been successfully implemented on an entertainment robot platform. On-line experimental results show that the proposed system learns new facial data with a recognition rate of 81.3% increased from an original recognition rate of 58%. The proposed method also keeps satisfactory recognition rate of old facial samples with a recognition rate of 78.7%.

# ACKNOWLEDGMENT

First of all, I would like to express my deepest sense of gratitude to my advisor Dr. Kai-Tai Song for his patient guidance, advice encouragement and excellent advice throughout this study. Further, I would like to thank Dr. Li-Chen Fu, Dr. Kuu-Young Young and Dr. Ching-Chih Tsai, for their comments and suggestions for the editing of my thesis.

I also thank Dr. Fuh-Yu Chang and the cooperation project with Industrial Technology Research Institute (ITRI). I like to thank my labmates in ISCI lab, Meng-Ju, Chia-How, Chi-Yi, Fu-Sheng, Chen-Yang, Chun-Wei and Zhi-Sheng for sharing experiences and knowledge during the time of study.

Finally, I take this opportunity to express my profound gratitude to my beloved parent, and my friends for their support and patience during my study in NCTU.



# Contents

中文摘要.....	i
<b>ABSTRACT.....</b>	<b>ii</b>
<b>ACKNOWLEDGMENT.....</b>	<b>iii</b>
<b>Contents.....</b>	<b>iv</b>
<b>List of Figures.....</b>	<b>vi</b>
<b>List of Tables.....</b>	<b>ix</b>
<b>Chapter 1 Introduction.....</b>	<b>1</b>
1.1. Preface.....	1
1.2. Related Works.....	2
1.2.1 Feature-Based Approach.....	3
1.2.2 Template-Based Approach.....	4
1.3. Problem Statement.....	7
1.4. System Overview and Organization of the Thesis .....	8
<b>Chapter 2 SVM-based Classifiers.....</b>	<b>10</b>
2.1 Support Vector Classification.....	10
2.1.1 Linear SVM.....	10
2.1.1.1 Support Vectors.....	12
2.1.2 The Non-Separable SVM.....	14
2.1.3 Non-linear SVM.....	15
2.2 Incremental SVM.....	15
<b>Chapter 3 Facial Expression Classification and Learning.....</b>	<b>18</b>
3.1 The Hierarchical SVM Classifier.....	19
3.2 Fast Learning for Robotic Applications.....	19
3.2.1 Support Vector Pursuit Learning.....	21

3.2.2 Critical Sets.....	22
3.2.3 The Proposed Learning Algorithm.....	24
3.2.4 Decomposing Kernel Function.....	27
<b>Chapter 4 Feature Extraction Under Illumination Variation.....</b>	<b>29</b>
4.1 Face Detection.....	29
4.2 Face Tracking.....	32
4.3 Face Image Preprocessing.....	33
4.3.1 Face Normalization.....	33
4.3.2 Feature Region Localization.....	34
4.4 Gabor Wavelet Transformation.....	35
4.5 Feature Points Extracting.....	40
4.5.1 Training Phase of Feature Points Extracting.....	41
4.5.2 Testing Phase of Feature Points Extracting.....	42
4.6 Feature Extraction Evaluation.....	43
4.7 The Experimental Results of Facial Points Extraction.....	45
<b>Chapter 5 The Experimental Results.....</b>	<b>52</b>
5.1 Experimental Results of Facial Expression Recognition.....	52
5.2 Experimental Results of the Proposed Learning Algorithm.....	53
5.3 On-Line Experiment using a Robot Platform.....	59
5.3.1 The Hardware Architecture of the Robot Platform.....	59
5.3.2 HRI Procedure of the Pet Robot Platform.....	61
5.3.3 Results of On-Line Testing.....	61
<b>Chapter 6 Conclusions and Future Work.....</b>	<b>68</b>
6.1 Conclusions.....	68
6.2 Future Work.....	68
<b>References.....</b>	<b>70</b>

## List of Figures

Figure 1-1 The process to automatically recognizing facial expression.....	2
Figure 1-2 The architecture of the designed system.....	9
Figure 2-1 The optimal hyperplane of SVM.....	11
Figure 2-2 A maximal margin hyperplane with its support vectors.....	14
Figure 2-3 Linear separating hyperplane for the non-separable case.....	15
Figure 3-1 The structure of hierarchical SVM classifier.....	19
Figure 3-2 An example of classifying neutral expression.....	20
Figure 3.3 The diagram of SVPL retraining strategy.....	22
Figure 3-4 (a) a SVM trained with all samples (b) a similar SVM trained with critical sets.....	23
Figure 3-5 An example of the retraining with critical sets by SVPL. (a) is the original SVM hyperplane, (b) one possible error of retraining all new samples by SVPL, (c) retraining only the two erroneous black points combined with critical sets by SVPL.....	25
Figure 3-6 The flowchart of proposed learning algorithm.....	28
Figure 4-1 The flowchart of face detection.....	30
Figure 4-2 The results of face detection (a) is the testing image. (b) is the result of color segmentation and (c) is the result of closing operation. (d) is the candidate of face regions. (e) is the final result via the attentional cascade.....	31
Figure 4-3 The lower and upper thresholds of each color channel (a) Y1 and Y2 of Y channel (b) Cb1 and Cb2 of Cb channel (c) Cr1 and Cr2 of Cr channel.....	33
Figure 4-4 The 14 facial regions of interest.....	33



Figure 4-5 Image interpolation.....	34
Figure 4-6 (a) the diagram of locating three reference points and (b) is the result.....	36
Figure 4-7 Different frequencies and orientations of Gabor wavelet filters.....	38
Figure 4-8 original face image .....	38
Figure 4-9 Six selected filtered facial images .....	39
Figure4-10 (a) is the image by summing 6 Gabor jets. (b) is binary image of facial edges.(c) is the intersection of 14 ROIs and the binary image of edge regions.....	40
Figure 4-11 14 facial points.....	40
Figure 4-12 The overall procedure of feature point extraction in the training phase and testing phase.....	42
Figure 4-13 The list of AUS related to five expressions [4].....	44
Figure 4-14 Examples of the AR Face Database (a)neutral expression (b) smile (c)scream (d)left light on (e)right light on (f)all lights on.....	47
Figure 4-15 Accurate extraction of all facial points.....	48
Figure 4-16 Five categories of facial expressions under four lighting conditions.....	49
Figure 5-1 Five categories of facial expressions.....	55
Figure 5-2 Sample images of a new person in the experiment.....	56
Figure 5-3 Sample images of other five new persons in the experiment.....	58
Figure 5-4 Comparison of recognition rate of proposed critical sets training algorithm and the conventional method using all SVs in training.....	58
Figure 5-5 The real-time vision system.....	60
Figure 5-6 Hardware architecture of the imaging system on Momobear.....	60
Figure 5-7 HRI procedure of the proposed emotion recognition system.....	62
Figure 5-8 The designed actions of Momobear.....	62

Figure 5-9 The interaction scenario.....63

Figure 5-10 An example of interaction with Momobear.....67



## List of Tables

Table 3-1 The relationship of decision value and Lagrangian multiplier.....	23
Table 3-2 The overall procedure of proposed learning algorithm.....	28
Table 4-1 The detailed descriptions of 14 ROIs.....	36
Table 4-2 The definitions of 14 facial points.....	41
Table 4-3 The association of 5 facial expressions to AU combination.....	43
Table 4-4 Feature points based descriptions for AUS.....	44
Table 4-5 The detailed descriptions of these 16 feature values.....	46
Table 4-6 The results of facial point extraction.....	47
Table 4-7 The results of facial point extraction in neutral expression.....	49
Table 4-8 The results of facial point extraction in happiness expression.....	50
Table 4-9 The results of facial point extraction in anger expression.....	50
Table 4-10 The results of facial point extraction in surprise expression.....	51
Table 4-11 The results of facial point extraction in sadness expression.....	51
Table 5-1 Average recognition results of training subjects.....	53
Table 5-2 Average recognition results of testing subjects.....	53
Table 5-3 Average recognition results of normal light.....	54
Table 5-4 Average recognition results of left light on.....	54
Table 5-5 Average recognition results of all light on.....	54
Table 5-6 Average recognition results of right light on.....	54
Table 5-7 The recognition result of the original four trained person.....	56
Table 5-8 The recognition rate of the 1 <sup>st</sup> new facial data.....	56
Table 5-9 The recognition rate of the 1 <sup>st</sup> new facial data after learning.....	57
Table 5-10 The recognition rate of all new face learning.....	58
Table 5-11 The recognition result of five trained persons .....	63

Table 5-12 Recognition rate of the first new person before online learning.....63

Table 5-13 Recognition rate of the first new person after online learning.....64

Table 5-14 Recognition rate of the second new person before online learning.....64

Table 5-15 Recognition rate of the second new person after online learning.....64

Table 5-16 Recognition rate of the third new person before online learning.....65

Table 5-17 Recognition rate of the third new person after online learning.....65

Table 5-18 Recognition rate of the fourth new person before online learning.....65

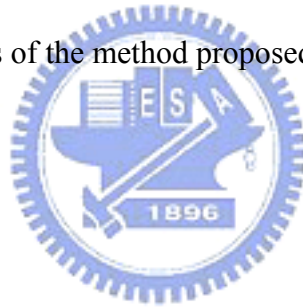
Table 5-19 Recognition rate of the fourth new person after online learning.....65

Table 5-20 Average recognition results of four new persons before online learning...65

Table 5-21 Average recognition results of four new persons after online learning .....66

Table 5-22 The recognition results of five trained person after online learning.....67

Table 5-23 Recognition results of the method proposed in [41].....67



# Chapter 1

## Introduction

### 1.1 Preface

In recent years, many service robots, household robots and entertainment robots have been developed for various applications [1]. One of the most important features of these robots is their human-centered functions. In the near future, intelligent service robots will have human-like interactions and work with us in our daily life. Therefore, the research of human-robot interaction (HRI) has become increasingly popular in robotics area.

In current HRI design, a variety of interaction modes including gesture, facial expression and audio have been adopted in a more natural manner. For human beings, facial expressions reveal a person's emotion and provide important communicative cues during social interaction. This implies that facial expressions form a major modality in human robot communication.

Several research efforts on facial expression recognition (FER) have been reported recently. A set of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) were defined by psychologists [2]. In order to make the analysis of FER more standard, facial action coding system (FACS) was created to describe the sets of facial muscle movements[3][4]. In a form of rules, FACS provides a linguistic description of all possible facial changes in terms of 44 Action Units (AUs), which can be combined to form all kinds of facial expressions.

For applications of smart human-robot interfaces, automatic recognition of facial

expression is an essential step to make the interaction more natural and efficient. In an automatic FER system, there are three major components to achieve this goal (See Figure 1-1). First, a face is detected and localized in a camera scene. Next, relevant facial feature information from the detected face region is extracted. Finally, the facial expression category is classified based on the extracted features.

In general, most image-based recognition systems need to collect certain number of facial dataset to train the emotion classifiers. However, appearance variations of human facial expression would be too great to be represented by fixed number of samples. Thus, high recognition rate in practical robotic applications can hardly be expected with only limited datasets. Accordingly, we propose a novel solution to this problem (termed subject-dependent) by making an entertainment robot to learn incrementally through HRI and to adapt to incoming samples from a new face, which are wrongly recognized by using previous emotion classifier.

Furthermore, inaccurate feature extraction very often results in erroneous recognition of facial expression. Environmental factors such as illumination variation may cause FER system to extract feature inaccurately. In this thesis, Gabor wavelet based feature extraction is also employed to detect facial features robustly under uncertain environments.

## 1.2 Related Works

In past decades, many techniques have been presented in the literature related to facial expression recognition. A survey of this area can be found in [5]. The report

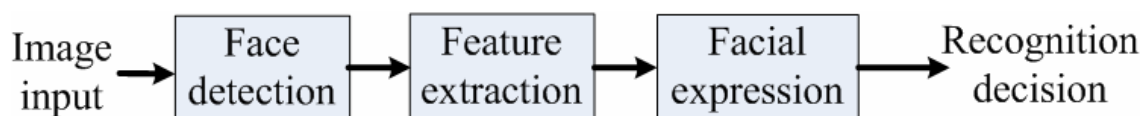


Figure 1-1 The process to automatically recognizing facial expression

approaches can be categorized into two main directions, the feature-based methods and the template-based methods, according to the way that the facial information is extracted. Feature-based methods mainly use local spatial analysis or geometrical information as facial features. Template-based methods use holistic spatial analysis, 2D or 3D face models as templates to represent facial expression information.

### **1.2.1 Feature-Based Approach**

A method of facial expression recognition based on selective feature extraction was proposed in [6]. Active appearance model (AAM) was applied to extract the contour of eyes, eyebrows and mouth. The displacement information of these features was classified based on selective feature rule. Another method proposed in [7] overcame some limitations of current HRI techniques, such as their sensitivity to partial occlusion and noisy data. This method features a representation of facial expressions which combines spatially-localized geometric facial model with state-based model of facial motion. Through the transition of feature states, this method reduces influences of noise and occlusion.

Gabor filters are popular in representing facial features. The results in [8] revealed that Gabor features from different Gabor channels have different contributions to the facial expression recognition. Therefore, [8] proposed a method to improve the performance of recognizing facial expressions by combining features belonging to different channels.

A fast and robust facial expression recognition design using video stream was proposed in [9]. In the first frame of the face video, 20 facial points were extracted from each region of interest in face region by using GentleBoost templates built from Gabor wavelet features. Then, a particle filter was exploited to track points in the subsequent image frames. AdaBoost selected the most informative spatio-temporal

features to train a support vector machine (SVM) classifier.

Facial expression recognition using small number of training samples was reported in [10]. An algorithm called feature selection via linear programming (FSLP) was proposed to select features efficiently. In [10], probability distribution based learning methods were compared with margin based methods for the case of small number of training examples. It was reported that the margin methods such as FSLP and SVM [11-12] have more accurate facial classifying results than those based on probability distribution, such as Bayes classifier and AdaBoost in the small sample cases.

Another spatio-temporal approach to recognize six facial expressions from visual data attempted to measure levels of interest from video [13]. It used projected optical flow vectors as facial features and applied principal component analysis (PCA) to reduce the dimension of optical flow. Discrete hidden Markov Models were used to learn the classifying model for each facial expression. Finally, the third dimensional affect space was also adopted to combine facial motion information around apex frame in order to guarantee the performance of facial expression recognition. In [14], a set of multi-scale and multi-orientation Gabor wavelet coefficient extracted from face images at fiducial points were adopted to represent the facial features. These features were fed to the input units of the two layer classifier.

### **1.2.2 Template-Based Approach**

In [15], facial expression recognition system was reported to be subject-independent and robust against illumination variation and image deformation. In their method, Gabor features at each lattice of expression image constructed an elastic graph. The amplitudes of those feature vectors tended to be larger at some key lattices. Expression recognition was performed by template matching of features and



position of key-points.

An approach to recognizing gender and facial expression was proposed in [16]. AAM was used to extract facial features [17]. Features extracted by a trained AAM were classified by a SVM. Gender-specific classification cascading facial expression classification were considered to yield better performance of classification.

Two real-time methods for facial expression recognition in image sequences were proposed in [18]. Some of Candide grid's points in face region were depicted manually at the first image sequences. The grid tracking system based on deformable models tracks all grids of expressional face in the remainder image sequences until the image frame reached the greatest facial expression intensity. The geometrical displacement of those Candide nodes were used as input to a multiclass SVM classifiers, which recognize the six basic facial expressions and a set of Facial Action Units.

The related works mentioned above mainly focus on algorithms or methods to improve the recognition performance. Few papers involve in the research of learning new facial data in online robotic applications. Recently, a method of expression learning was proposed by imitating the process of a baby's learning process [19]. A penguin robot started learning three facial expressions by recognizing human actions through CdS light sensors equipped on the robot's head. After learning new facial expressions, the penguin robot could identify the emotions without using the CdS light sensor.

Although few papers concern about expression learning, incremental learning for online face recognition, a similar procedure to facial expression recognition, has drawn much attention in recent years. In [20], a new approach to face recognition in which not only a face classifier but also a feature space was learned incrementally to adapt to new incoming samples. To learn a feature space and an optimal decision

boundary in online, an extended version of incremental principal component analysis (IPCA) and resource allocating network with long-term memory (RAN-LTN) were combined to achieve this goal. For using IPCA, a feature space was updated to a new training number by rotating its eigen-axes and increasing its dimensions. In RAN-LTM, a small number of training samples called memory items were selected for retraining a new classifier. The face recognition system was adapted to a new face incrementally by learning a new feature space and a face classifier online.

The architecture suitable for real time robotic learning of face recognition was proposed in [21]. A face tracking coupled to a clustering technique was utilized to learn a person's face appearance when the system interacted with a user. The learning approach is similar to the partial memory incremental learning method, where the representative samples of similarity measurement can be updated for new incoming faces.

According to the survey of facial expression and learning, we hope a robot to learn facial expressions incrementally. Further, a good and adaptive classifier is essential for online learning. Support vector machine (SVM) has been an effective method for designing facial expression recognition systems, especially when the training samples are small [10]. Therefore, online robotic learning of new facial data has become the problem of SVM-based incremental learning in this study.

A popular method to expedite SVM learning is to reduce the number of training data. Vanik [11][12] showed that the training result of SVM depends only on a small set of samples, termed support vector (SV) sets. Chunking algorithm [12] solved the SVM containing all SV sets plus some new samples which violate the Karush-Kuhn-Tucker (KKT) condition. Decomposition method [12] is similar to the Chunking algorithm. The main difference is that the size of retraining samples in decomposition method is fixed. In addition to these conventional methods,  $\alpha$ -ISVM

[22] provides an efficient incremental algorithm based on the discard factor  $\alpha$ . Through the adjustable parameter, it is possible to discard samples optimally. Thus retraining time of SVM can be saved greatly. Another incremental learning algorithm I-SVM [23] also discarded part of history samples, both the pre-extracting SVs algorithm and the iteration algorithm have been used in the retraining of the SVM. SeqSVM [24] trained a SVM classifier on a small part of training samples and selected the so-called convex hull samples, which were wrongly classifier by the current SVM, to retrain a new SVM.

### **1.3 Problem Statement**

Note that the incremental learning methods described above need old training data of the SV sets. In order to make an original SVM emotional classifier to adapt to new subjects' facial expressions, one intuitive incremental learning approach is to store old SV samples in memory and retrain all of them, including the new samples. However, this approach will lead to a problem that if the number of retraining data is large, it will require much time to solve quadratic optimization (QP) to obtain a SVM classifier[11][12]. The training time will increase dramatically with the number of training dataset, resulting in inefficiencies in practice from both memory requirement and real time criteria. To assure that the robot can accommodate its emotion recognition system to a new human face effectively, a fast incremental learning algorithm for emotion recognition is desired.

Inaccurate feature extraction will result in erroneous recognition of facial expressions. In practical applications, feature extraction may fail due to lighting conditions of the environment because gray values suffer from huge ambiguities as well as slight changes in illumination. Therefore, how to extract facial features robustly under illumination variation conditions is still a problem in facial expression

recognition.

## **1.4 System Overview and Organization of the Thesis**

The architecture of the proposed system is shown in Figure 1-2. The system consists of five units, namely face detection, facial feature extraction, facial expression classification, human-robot interaction and facial classifier learning.

First, a face region is detected under illumination variation conditions. Next, in order to improve the efficiency of feature extraction, the front-view face region is preprocessed by normalizing the face and localizing the feature region before feature extraction. Gabor wavelet based features regardless of illumination variation are used to detect the facial fiducial points in the relevant region of interest. After extracting those feature points, the feature values are evaluated from the geometric displacement of facial points according to FACS. Third, five categories of facial expression, neutral, anger, happiness, sadness and surprise are classified by hierarchical SVM classifiers. An entertainment robot can response to the estimated emotions. If an entertainment robot recognizes the user's facial erroneously, a fast SVM-based learning algorithm is employed to adjust the SVM classifier. The incremental procedure is executed to learn facial expressions of new individuals. A novel algorithm using support vector pursuit learning (SVPL) coupled to critical historical sets to restrain a new SVM classifier in Gaussian feature space is designed to make the robot achieve this goal.

The rest of this thesis is organized as follows. In chapter 2, SVM-based classifier is introduced. Chapter 3 describes the classification of facial expressions and the proposed learning algorithm of SVM classifier. In chapter 4, a robust facial feature extraction insensitive to illumination is introduced. In chapter 5, the experimental results of the proposed algorithm are presented and discussed. Chapter 6 summarizes the contribution of this thesis and gives directions of further research.



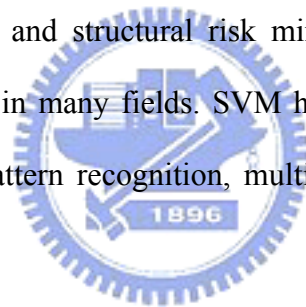
# Chapter 2

## SVM-Based Classifiers

The emotional classification and the learning algorithm of an entertainment robot have been developed based on SVM, a statistical learning theory. We will introduce the basic principle of SVM in this chapter.

### 2.1 Support Vector Classification

Support Vector Machine [11][12] is a popular machine learning technique based on statistical learning theory and structural risk minimization principle. SVM has become increasingly popular in many fields. SVM has been successfully applied in many applications such as pattern recognition, multi-sensor information fusion and bio-sequence analysis.



#### 2.1.1 Linear SVM

The simplest model of Support Vector Machine is the maximal margin classifier. It works only for data which is linearly separable in the feature space.

Given a set of training samples belonging to two classes,  $\{x_i, y_i\}, i = 1, \dots, l, x_i \in R^n, y_i \in \{\pm 1\}$ , the goal of SVM is to find a hyperplane  $w \bullet x + b = 0$  to separate the data successfully. In general, SVM classifiers separate the data with the maximal margin hyperplane. Figure 2-1 shows that there are many possible hyperplanes to separate the data, but only one can maximize the margin between two classes.

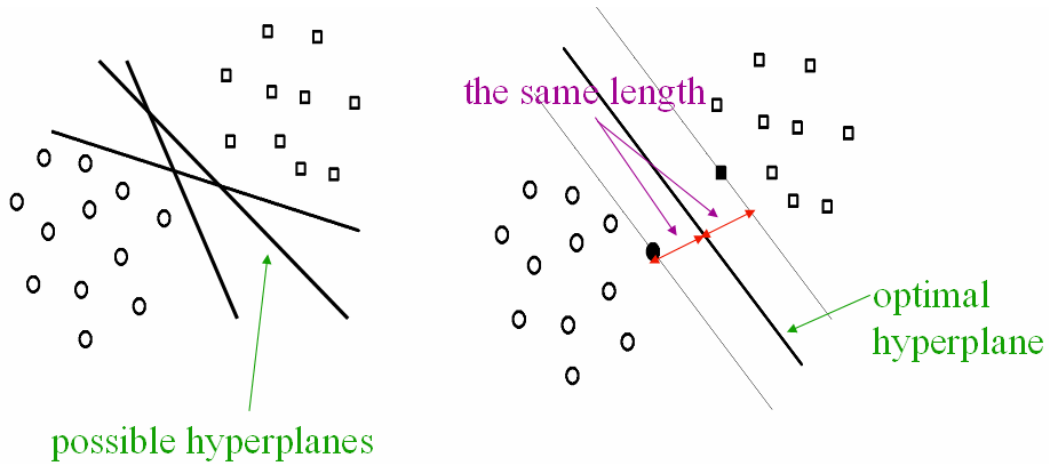


Figure 2-1 The optimal hyperplane of SVM

For the linearly separable case, all the training data satisfy the following constraints [12] :

$$w \bullet x^+ + b \geq 1 \quad \text{for } y_i = +1 \quad (2-1)$$

$$w \bullet x^- + b \leq -1 \quad \text{for } y_i = -1 \quad (2-2)$$

The geometric margin  $\gamma$  can be maximized by computing the following function :

$$\gamma = \left( \frac{w \bullet x^+}{\|w\|^2} - \frac{w \bullet x^-}{\|w\|^2} \right) = \frac{1}{\|w\|^2} (\langle w \bullet x^+ \rangle - \langle w \bullet x^- \rangle) \propto \frac{1}{\|w\|^2} \quad (2-3)$$

Hence, an optimal hyperplane which gives the maximum margin by minimizing  $\|w\|^2$  can be found. Similarly, the hyperplane (w,b) can be solved by the following optimization problem :

$$\text{Minimize}_{w,b} \quad \langle w \bullet w \rangle, \quad (2-4)$$

$$\text{Subject to } y_i (w \bullet x_i^+ + b) \geq 1, \quad i=1, \dots, l, \quad (2-5)$$

This optimization problem can be solved by finding the saddle point of the Lagrangian. The primal Lagrangian is

$$\max_{L(w,b,\alpha)} = \|w\|^2 - \sum_{i=1}^l \alpha_i \{y_i (w \bullet x_i + b) - 1\} \quad (2-6)$$

where  $\alpha_i \geq 0$  are the Lagrange multipliers. The corresponding dual is found by differentiating w and b, and the solution can be written as follows :

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2-7)$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2-8)$$

Substituting the relations into (2-6) to obtain the following dual objective function:

$$\max L(w, b, \alpha) = \frac{1}{2} \langle w \bullet w \rangle - \sum_{i=1}^l \alpha_i [y_i (\langle w \bullet x_i \rangle + b) - 1] \quad (2-9)$$

$$= \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i \bullet x_j \rangle - \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i \bullet x_j \rangle + \sum_{i=1}^l \alpha_i \quad (2-10)$$

$$= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i \bullet x_j \rangle = \alpha^T - \frac{1}{2} \alpha^T D \alpha \quad (2-11)$$

To maximize the equation, the Lagrange multipliers  $\alpha_i$  can be obtained by differentiating (2-11). After substituting  $\alpha_i = D^{-1} \bullet \hat{1}$  into (2-7), the parameter of SVM hyperplane  $w$  can be solved. Although the value of  $b$  does not appear in the dual problem,  $b$  can be solved by using the constraints :

$$b = -\frac{\max_{y_i=-1} (\langle w \bullet x_i \rangle) + \min_{y_i=1} (\langle w \bullet x_i \rangle)}{2} \quad (2-12)$$

### 2.1.1.1 Support Vectors

In addition, the Karush-Kuhn-Tucker (KKT) conditions play an important role in the structure of the solution [12]. The KKT conditions state that the optimal solutions  $\alpha_i, (w, b)$  must satisfy :

$$\alpha_i [y_i (\langle w \bullet x_i \rangle + b) - 1] = 0, \quad i = 1, \dots, l. \quad (2-13)$$

This equation implies that  $x_i$ , which decision value is one, has corresponding non-zero  $\alpha_i$ . According to (2-7), the parameter  $w$  is involved mostly with these non-zero  $\alpha_i$ , so these non-zero  $\alpha_i$  are called support vectors (SVs). Figure 2-2



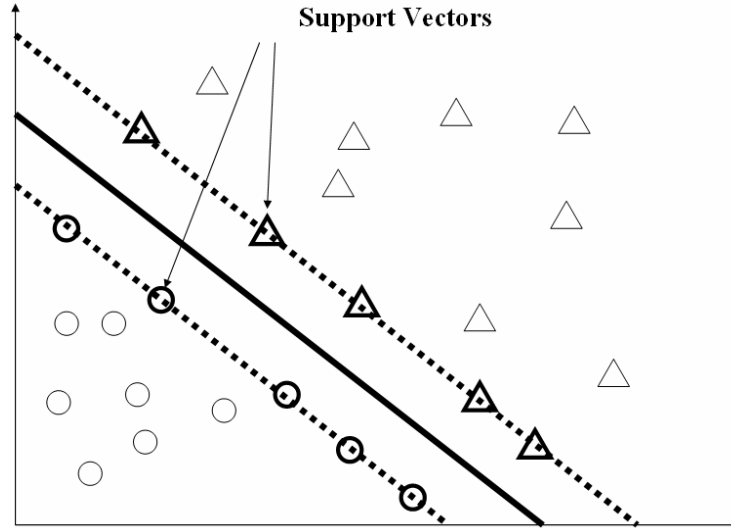


Figure 2-2 A maximal margin hyperplane with its support vectors

shows that there are many vectors in the plane, but only few lie on the dotted line and form the SVM hyperplane.



### 2.1.2 The Non-Separable SVM

The above algorithm is for separable data, but it will find no feasible solution when the data in the feature space are non-separable [12]. 2-Norm Soft Margin SVM is introduced to separate the data that are not linearly separable in the feature space. In order to solve this non-separable case, positive slack variables  $\xi_i, i = 1, \dots, l$  are used to make the constraints to be violated (See Figure 2-3) :

$$\text{Minimize}_{w,b} \quad \langle w \bullet w \rangle + C \sum_{i=1}^l \xi_i^2, \quad (2-14)$$

$$\text{Subject to} \quad y_i(w \bullet x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l, \quad \xi_i > 0 \quad (2-15)$$

where  $\xi_i$  is slack variables and C is a constant which determines the trade-off between training error and VC-dimension. Then, this optimization problem can be solved by finding the saddle point of the Lagrangian :

$$\max L = \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i^k \{y_i^k (w^k \bullet x_i + b^k) - 1 + \xi_i\} \quad (2-16)$$

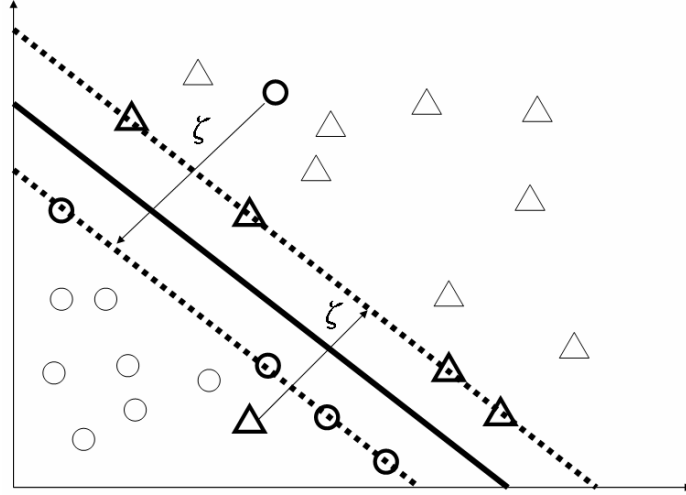


Figure 2-3 Linear separating hyperplane for the non-separable case.

where  $\alpha_i \geq 0$  are the Lagrange multipliers. The solution can be written as follows :

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2-17)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2-18)$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow c\xi - \alpha = 0 \quad (2-19)$$

Substituting the relations into (2-16) to obtain the following dual objective function

$$\begin{aligned} \max D(\alpha^k) &= -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^l \alpha_i + \frac{1}{2C} (\alpha \cdot \alpha) - \frac{1}{C} (\alpha \cdot \alpha) \\ &= -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^l \alpha_i - \frac{1}{2C} (\alpha \cdot \alpha) \end{aligned} \quad (2-20)$$

Hence, maximizing the above objective function over  $\alpha$  is equivalent to maximize

$$L(w, b, \xi, \alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j + \frac{1}{C} \delta_{ij}) \quad (2-21)$$

$$\text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{others} \end{cases} \quad (2-22)$$

After substituting  $\alpha_i^k$  in (2-17), the parameter of the hyperplane  $w$  is solved.

The value of  $b$  is chosen using the relation  $\alpha_i = C\xi_i$  and referred by

Karush-Kuhn-Tucker complementarily condition. The condition is as below

$$\alpha_i(y_i(w \cdot x_i + b) - 1 + \xi_i) = 0 \quad \forall i \quad (2-23)$$

### 2.1.3 Non-linear SVM

SVM also can generalize to the case where the decision function is not a linear function. This problem can be solved by mapping the data from the feature space to some other higher dimension space (possibly infinite dimension). In most cases of SVM, several kernel functions have been used for nonlinear mapping. Three kernel functions are the most commonly used, and those are

$$\text{Polynomial : } K(x_1, x_2) = (\gamma x_1^T x_2 + r)^d, \gamma > 0 \quad (2-24)$$

$$\text{Sigmoid : } K(x_1, x_2) = \tanh(\gamma x_1^T x_2 + r) \quad (2-25)$$

$$\text{Gaussian radial basis function : } K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{c}\right) \quad (2-26)$$

where  $\gamma$ ,  $d$ ,  $c$  are kernel parameters. In these kernel functions, kernel space is infinitely dimensional, so we do not need to know the mapping space of a kernel function explicitly. In conclusion, for a given kernel function, the non-linear decision function of SVM classifier can be now written as

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(x \bullet x_i) + b\right) \quad (2-27)$$

where  $x_i$  are the support vectors.

## 2.2 Incremental SVM

SVM requires much time to solve quadratic optimization to obtain a SVM

classifier. The key point is that obtaining  $\alpha_i$  of the dual objective function is time consuming because it involves in the computation of matrix inverse in the programming procedure. So the training time will increase dramatically with the number of training dataset. On the other hand, SVM also suffers from the problem of large memory requirement when training on a large data set. Therefore, incremental learning techniques are developed to make the SVM training faster and to reduce the storage cost over very large data sets.

According to (2-13), the training result of SVM only depends on a small set of samples, termed support vector (SV) sets [11-12]. Therefore, Chunking algorithm [12] solves the SVM incrementally containing all SV sets plus some of new samples which violate the Karush-Kuhn-Tucker (KKT) condition. But the shortcoming is that the algorithm is limited by the maximal number of support vectors and still requires the procedure of the quadratic optimization. Decomposition method [12] is similar to the Chunking algorithm. The main difference is that the size of retraining samples in decomposition method is fixed.

Except for these conventional methods,  $\alpha$ -ISVM [22] provides an efficient incremental algorithm based on the discard factor  $\alpha$ . Through the adjustable parameter, it is possible to discard samples optimally. Thus retraining time of SVM can be saved greatly. Another incremental learning algorithm I-SVM [23] also discards part of history samples, both the pre-extracting SVs algorithm and the iteration algorithm have been used in the retraining of the SVM.

SeqSVM[24] trains a SVM classifier on a small part of training samples and selects the so-called convex hull samples to retrain a new SVM. Convex hull samples are wrongly classified by the current SVM and furthest from the current SVM solution. It implies that these samples are the most possible to be the support vectors in the future

SVM.

Some other traditional methods of SVM-based incremental learning [25-26] also expedite the retraining by reducing the number of the training data. Besides reducing the number of retraining samples for the purpose of speeding up the retraining procedure, Sequential Minimal Optimization (SMO) [12] provides an alternative approach to avoid the overall procedure of QP. SMO is derived by taking the idea of decomposition method to its extreme and just optimizes two multipliers  $\alpha_1$  and  $\alpha_2$  in each QP sub-problem. We can compute new values of the two multipliers by

keeping the linear constrain  $\sum_{i=1}^l \alpha_i y_i = 0$  and the new multipliers have the following relationship :

$$\alpha_1 y_1 + \alpha_2 y_2 = \text{constant} = \alpha_1^{old} y_1 + \alpha_2^{old} y_2 \quad (2-29)$$

The choice of updating the two multipliers is determined by a heuristic. The procedure of SMO is closed while all Lagrange multipliers satisfy the KKT conditions.

Note that the incremental learning methods described above mostly need old training SV sets. Therefore, if we want to train a SVM which can classify new learning datasets and keep the performance of recognizing old ones, retraining new data together with SV set of the original samples or many historical sets would be necessary. This would be insufficient for real-time criteria and memory requirement in robotic applications. To this end, we will propose a fast SVM-based learning algorithm to expedite the training of SVM.

# Chapter 3

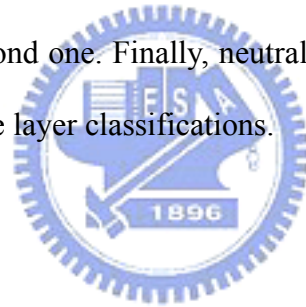
## Facial Expression Classification and Learning

This chapter describes the developed classifier and learning algorithms of a robotic facial expression recognition system. SVM is applied to the proposed system for recognizing five facial expressions. Generally, an image-based facial expression recognition system needs to collect certain number of facial datasets to train the emotion classifiers. This will lead to a problem that appearance variations of human faces would be too great to be represented by limited number of samples in practical applications. This means that high recognition rate in practical robotic applications can hardly be expected with only few dataset. Therefore, we propose a novel incremental learning algorithm based on SVM to overcome the problem (termed *subject-dependent*) of facial expression recognition.

The basic idea of the proposed learning algorithm is to adjust parameters of the SVM hyperplane for learning facial expressions of a new face, which is defined as the data recognized incorrectly through previous trained SVM classifiers. Support vector pursuit learning (SVPL, see 3.2.1) [27-28] is applied to retrain the hyperplane iteratively in the Gaussian-kernel space. To expedite the retraining procedure, only erroneous samples are combined with critical historical sets (see 3.2.2) to restrain a new SVM classifier. After the fast retraining using the proposed method, the new classifier will learn to recognize new facial data with improved correct rate.

### 3.1 The Hierarchical SVM Classifier

It is well known that SVM has been an effective method for designing facial expression recognition system [10], especially when a small number of training data are considered. For using SVM classifiers, the categories of two-class data can be decided by computing the sign of the decision function. The facial expression classifier needs to recognize five categories of emotional expressions consisting of anger, happiness, neutral, sadness and surprise. Consequently, a hierarchical SVM classifier is designed to categorize five emotional expressions. The structure of hierarchical SVM classifier is as shown in Figure 3-1. An example of classifying a neutral expression is also illustrated in Figure 3-2. First, neutral, anger and sadness are performed through the first layer classification. Next, anger expression assumes sadness expression in the second one. Finally, neutral expression can assume sadness and be outputted through three layer classifications.



### 3.2 Fast Learning for Robotic Applications

A facial expression recognition system needs to collect some representative

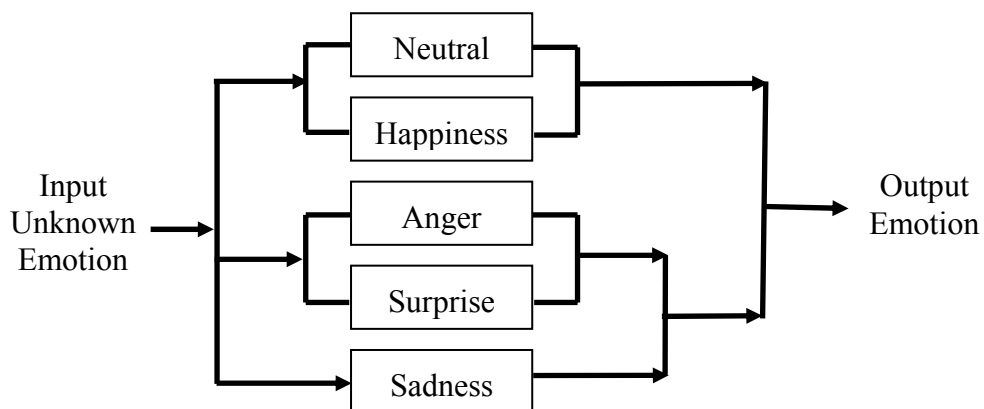


Figure 3-1 The structure of hierarchical SVM classifier.

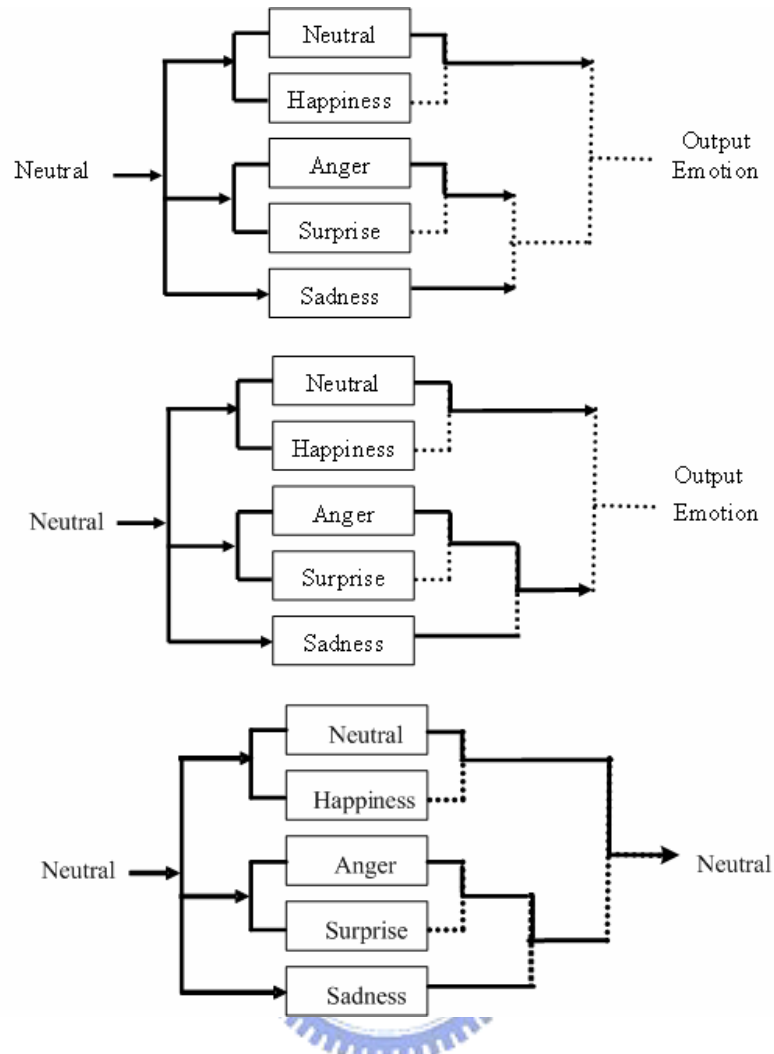


Figure 3-2 An example of classifying neutral expression

samples to train SVM classifier beforehand. The subject-dependent problem in facial expression recognition can hardly be avoided because different people may show a categorized expression in different ways. If the facial expressions of a test person are not collected previously in the database, high recognition rates are difficult to obtain. Therefore, a solution to this problem is investigated to accommodate the robot to various persons' facial expressions.

The objective of this study is to design a system for emotional interaction of entertainment robots. An entertainment robot equipped with a facial expression recognition system and learning algorithm can recognize the facial expression of its



host. For a newly purchased robot, if the robot recognizes new faces correctly, then the current parameters of the recognition system will be good enough. On the contrary, if the owner perceived that the robot recognizes new facial expressions erroneously by observing the emotional responses of the robot, the owner would inform the robot about the wrong recognition results through a simple input device. Consequently, the entertainment robot can start to retrain a new SVM hyperplane immediately.

As mentioned in Chapter 2, large sizes of retraining samples will require much time to solve quadratic optimization (QP) of a SVM classifier. This would not be efficient in practice from both memory requirement and real time criteria. So a fast SVM-based incremental learning algorithm will be required for practical applications.

### 3.2.1 Support Vector Pursuit Learning

We adopt the concept of support vector pursuit learning (SVPL) [27-28] to develop the emotion learning system. Previous parameters of the hyperplane together with the new data are employed to restrain a new SVM classifier. The main idea of SVPL is that the old hyperplane  $w^{k-1}$  shifts a minimal distance to a new hyperplane  $w^k$  in order that the new one can separate new data correctly (See Figure 3-3). The data in Figure 3-3 are all new training samples. Because the distance between the new and previous hyperplane is minimal, the new hyperplane is also expected to separate the old data. When the new classifier is obtained, the new training data of the current step can be discarded after completing the training procedure. Hence SVPL effectively reduces the competition and the memory requirement in learning new data.

In practical applications, facial expressions of new faces, which will be learned incrementally, are usually far different from the original ones. It will result in that the

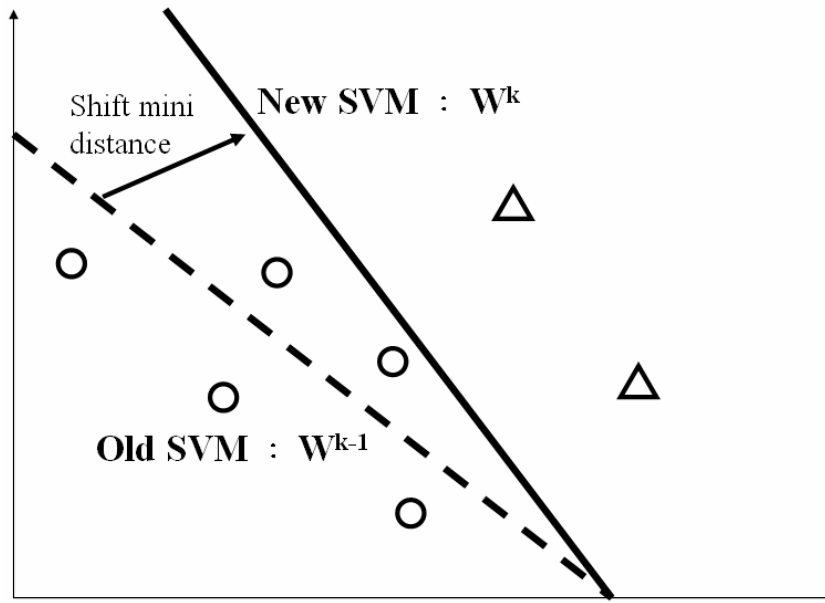
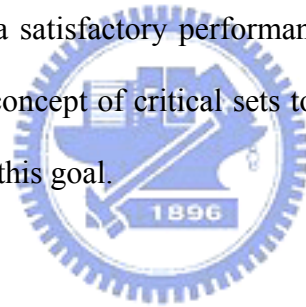


Figure 3-3 The diagram of SVPL retraining strategy

new SVM will not maintain a satisfactory performance of recognizing the old data. So we propose to use a new concept of critical sets to couple with SVPL to design a learning algorithm to achieve this goal.



### 3.2.2 Critical Sets

Gaussian-kernel mapping method is added in the first stage in the developed algorithm to map the feature space to kernel space where facial data can be retrained easily. Gaussian radial basis function is written as follows :

$$K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{c}\right) \quad (3-1)$$

Subsequently, as described in (2-7), the hyperplane  $w$  is constructed from the summation of feature values and the class category. Lagrange multiplier  $\alpha_i$  determines weights of the training samples to form the hyperplane  $w$ . In general, the samples which are nearest to the SVM hyperplane have greatest values of Lagrange multipliers. Namely, these samples play critical roles in forming the SVM hyperplane. This concept can be explained using Figure 3-4. Figure 3-4 (a) shows that a SVM is

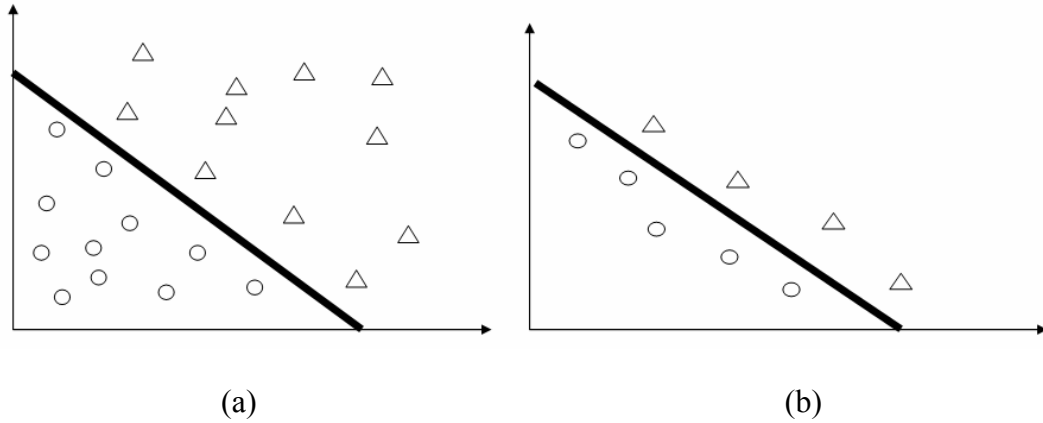


Figure 3-4 (a) a SVM trained with all samples (b) a similar SVM trained with critical sets

Table 3-1 The relationship of decision value and Lagrangian multiplier

Decision value	Lagrangian multiplier
0.7894	0.2132
-0.91	0.0875
-0.975	0.0225
-0.9791	0.0183
1.0053	-0.0027
1.0694	-0.0669

trained with all samples existing in the feature space, but one can also train an almost the same SVM with few sets that are as shown in Figure 3-4 (b). On the other hand, Table 3-1 shows an example that the sets whose absolute decision values are smaller have bigger Lagrangian multipliers.

As a result, a few historical and important samples, which are nearest to the SVM hyperplane, are reserved to restrain a new SVM classifier for keeping the recognition of old data. We define these samples as critical sets (CSs).

$$\text{CSs: } X_i = \arg \min |w \bullet X_i + b| \quad (3-2)$$

The size of critical sets is determined empirically. There is a trade-off between training time and the size of critical sets. Thus, we can adjust the size of critical sets to

meet the seesaw between learning time and non-forgetting learning.

The effectiveness of critical sets to maintain the performance of recognizing old data is shown in Figure 3-5. Assume that the hyperplane trained with the white circles and the white triangle, as shown in Figure 3-5 (a) and the black samples are new samples. It can be seen that there are two new black circles classified erroneously by the original hyperplane. Figure 3-5 (b) shows one possible retraining error risk with traditional SVPL, where the new hyperplane can classify the new samples correctly, but still one old sample (white triangle) is misclassified. Next is a new strategy to retrain SVM, only the two erroneous black points combined with critical sets are used to retrain a new hyperplane with SVPL. One black triangle becomes the new critical set after online updating the CSs. The result of the new hyperplane is shown in Figure 3-5 (c). It can separate both the new and old samples correctly.

### 3.2.3 The Proposed Learning Algorithm

This section introduces the proposed learning algorithm that utilizes CSs and SVPL to retrain a SVM. We denote critical sets as  $TS_0 = \{x_i^0, y_i^0\}_{i=1}^L$  which are nearest to the initial hyperplane. At t-th incremental learning step, some new samples  $TS_k = \{x_i^t, y_i^t\}_{i=1}^m$  come in, and we assume that the initial SVM cannot classify these new samples correctly. The parameters of the new hyperplane that can separate new sample successfully can be changed to  $(w', b')$  by proposed algorithm. The following steps are similar to the conventional SVM procedure, the quadratic programming problem of the incremental updating can be expressed as follows :

$$\text{Objective function : } \min \frac{1}{2} \|w^t - w^{t-1}\|^2 + C \sum_{i=1}^{(L+m)} \xi_i \quad (3-3)$$

$$\text{s.t. } y_i(w^t \bullet x_i^t + b^t) \leq 1 - \xi_i \quad \text{for } y_i = -1 \quad (3-4)$$

$$y_i(w^t \bullet x_i^t + b^t) \geq 1 + \xi_i \quad \text{for } y_i = +1 \quad (3-5)$$

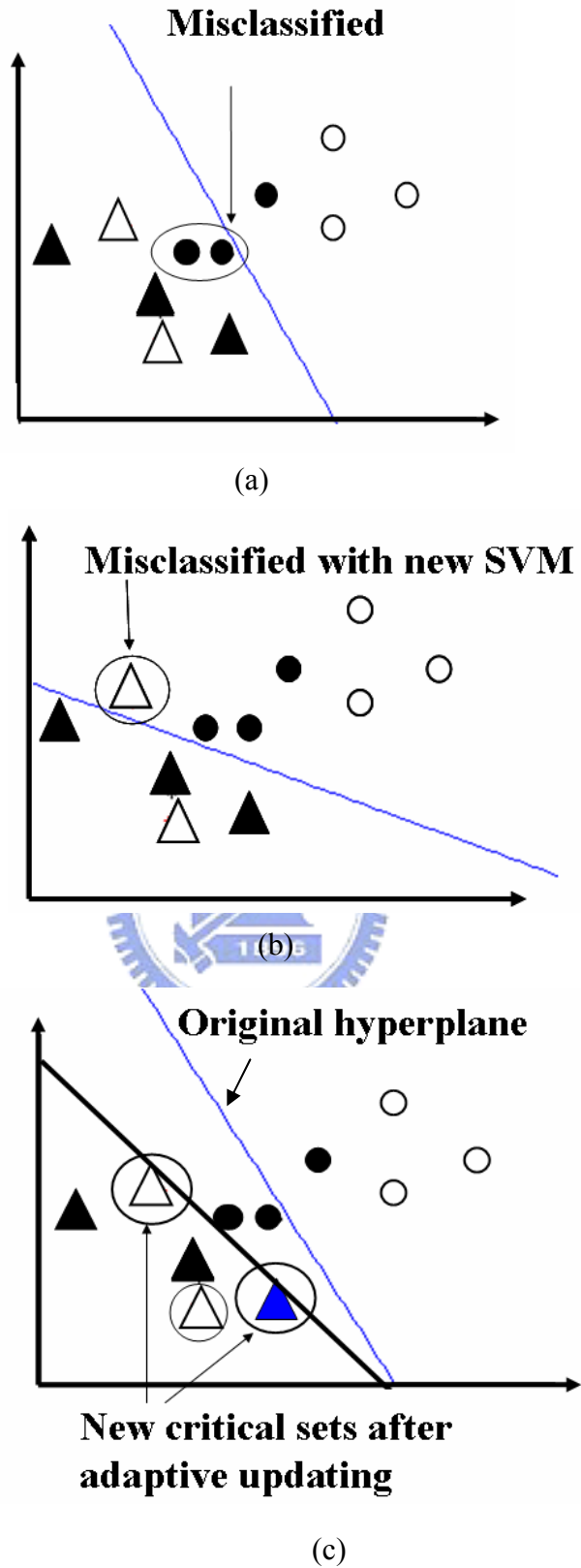


Figure 3-5 An example of the retraining with critical sets by SVPL. (a) is the original SVM hyperplane, (b) one possible error of retraining all new samples by SVPL, (c) retraining only the two erroneous black points combined with critical sets by SVPL

where  $\xi_i$  is slack variables in the sets of  $TS_0$  and  $TS_k$ .  $C$  is a trade-off between training error and VC-dimension. Let  $w^{t-1}$  denotes the previous trained parameter and a fixed constant. This optimization problem can be solved as follows :

$$\frac{\partial L}{\partial w^t} = 0 \Rightarrow w^t = w^{t-1} + \sum_{i=1}^{(L+m)} \alpha_i^t y_i^t x_i^t \quad (3-6)$$

$$\frac{\partial L}{\partial b^t} = 0 \Rightarrow \sum_{i=1}^{(L+m)} \alpha_i^t y_i^t = 0 \quad (3-7)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow 0 \leq \alpha_i^t \leq C, i = 1, \dots, m \quad (3-8)$$

The procedure to solve  $\alpha^t$  is similar to 2-Norm Soft Margin SVM [12]. After substituting  $\alpha^t$  in (3-6), the parameter of the new hyperplane  $w^k$  is solved. (3-6) indicates that the new parameter of SVM can be derived iteratively through the parameter of previous SVM and the retraining sets comprising  $TS_0$  and  $TS_k$ . Consequently, a new SVM, which not only recognizes new facial data but also keeps an acceptable recognition rate of original facial data, will be achieved by the proposed learning algorithm. Further, the proposed algorithm speeds up the QP procedure of SVM dramatically by reducing the retraining datasets. This is beneficial to the real-time requirement in practically robotic applications.

Finally, the decision function can be written as below :

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(x \bullet x_i) + b\right) \quad (3-9)$$

where  $K(\cdot)$  is a Gaussian kernel function which can be used to map the input space to some higher dimensional kernel space. Suppose that  $\phi: X \rightarrow F$  is a non-linear mapping from the input space to some higher kernel space, the decision function can be rewritten as :

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i \langle \phi(x_i) \bullet \phi(x) \rangle + b\right) \quad (3-10)$$

where  $w = \sum_{i=1}^l \alpha_i y_i \phi(x_i)$ , the parameter  $w$  can be saved for further learning.

### 3.2.4 Decomposing Kernel Function

Kernel function is an implicit mapping from feature space to kernel space, the underlying feature mapping is not known explicitly. Hence,  $w$  can not be derived directly for the purpose of retraining SVM hyperplane incrementally.

We can take the feature mapping to be any linear transformation  $x \rightarrow AX$ , for some matrix  $A$ . In this case the kernel mapping is given by

$$K(x, z) = \langle Ax, Az \rangle = x' A' A z = K \quad (3-11)$$

where  $K$  is a square symmetric matrix,  $K$  can be diagonalized into the following expressions :

$$K = V \Lambda V' = V \Lambda^{\frac{1}{2}} (\Lambda^{\frac{1}{2}} V') = V \Lambda^{\frac{1}{2}} (V \Lambda^{\frac{1}{2}})' = X \bullet X' \quad (3-12)$$

where  $X = V \Lambda^{\frac{1}{2}}$ . We assume that  $X$  is one possible result of kernel mapping, thus the parameter of  $w$  can be derived explicitly. The final decision function in kernel space is given below :

$$f(x) = \text{sign}(\sum_{i=1}^l \alpha_i y_i K(x \bullet x_i) + b) = \text{sign}(\sum_{i=1}^l \alpha_i y_i X \bullet X_i + b) = \text{sign}(\sum_{i=1}^l \alpha_i y_i X_i) \bullet X + b = w \bullet X + b \quad (3-13)$$

where  $w = \sum_{i=1}^l \alpha_i y_i X_i$ . Diagonalizing kernel matrix to know the kernel space, the proposed learning algorithm can be accomplished even though the kernel function is an implicit mapping function. The architecture of the proposed learning algorithm is illustrated in Figure 3-6. Table 3-2 summarizes proposed learning algorithm. First, features of facial expression sample are collected. They are mapped to Gaussian kernel space. Next, a hierarchical SVM classification is used to categorize five emotion expressions. In the meanwhile, the parameter of SVM and critical sets are reserved for future learning. If new data are supplied, a new SVM classifier will be learned by the proposed algorithm, which uses only erroneous data and critical sets to

update the SVM. As soon as the new SVM is adjusted, critical sets are updated accordingly. The learning procedure is continued until no new data exists.

Table 3-2 The overall procedure of proposed learning algorithm

Initial : Train the SVM classifier **S1** with training facial datasets **IS** (initial sets) and save **CS** (critical sets);  
 Step1 : Test the new facial data **NS** with classifier **S1**,  
     **if erroneous data=Nil**  
     then :  
         do nothing and exit;  
     **else erroneous data=ES(erroneous sets)**  
     then :  
         retrain new SVM classifier **S** with proposed learning algorithm using **ES and CS** ;Update **CS**;  
 Step2 : New SVM classifier **S1=S**; **IS=NS+IS**;  
 Step3 : Test **IS** with **S1**;

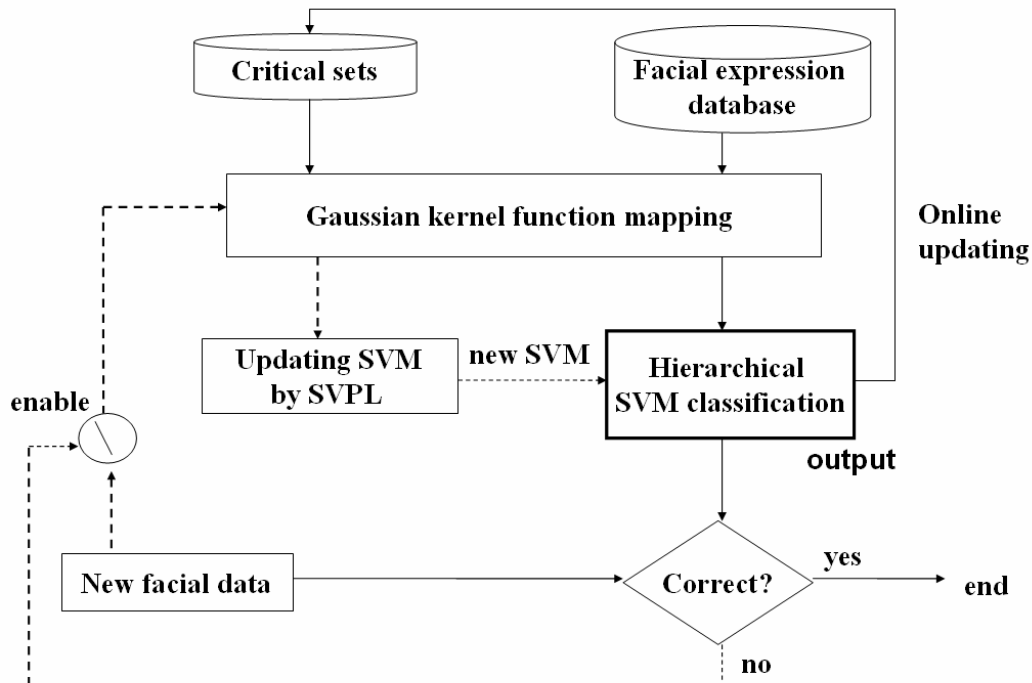


Figure 3-6 The flowchart of proposed learning algorithm



# Chapter 4

## Feature Extraction Under Illumination Variation

To assure that feature values are extracted accurately for classification and learning, we adopt Gabor wavelet to develop a robust feature extraction method. Facial points can be detected under various lighting conditions. Before extracting facial features, a face should be localized and segmented from digitized images sequences. Face preprocessing stage consists of face normalization and feature region localization steps to extract facial features efficiently. While regions of interest corresponding to relevant features are determined, we apply Gabor jets based on Gabor wavelets transformation to extract the facial points. Gabor jets are more invariable and reliable than the gray values which suffer from huge ambiguities as well as slight changes in illumination while representing local features. Each feature point can be matched by a phase-sensitivity similarity function in the relevant regions of interest. As long as the feature points are extracted under illumination variety, we can evaluate the geometric displacements of these points as emotional feature values.

### 4.1 Face Detection

Color is a direct cue for face localization, but skin color is easily suffered from illumination uncertainty. In this design, a YCrCb 3D color distribution model is applied to segment skin color under illumination variation [29]. The method of face localization consists of face detection and face tracking. Face detection module is an initial state which contains the location and size of face information for latter face

tracking. Face region is detected in the first image of image sequences by executing face color segmentation, morphology, color region mapping and attentional cascade (see Figure 4-1).

As shown in Figure 4-2 (b), the binary image is obtained by using skin color segmentation in YCrCb color space. Then, closing operation of morphology [30] is used to eliminate the discontinuous interference in the skin color region(see Figure 4-2 (c)). After obtaining the binary image of color segmentation, the 2-D histogram mapping is utilized to find possible areas of faces and Figure 4-2 (d) illustrates the result. Finally, attentional cascade assigns several rules to confirm if the candidate region is a real human face. If the following conditions are satisfied, the face detection is regarded successful :

- a. The ratio of mapping length to mapping width is between 1 and 2.
- b. The sum of grayscale in the upper area is smaller than the sum of the lower one.
- c. The sum of grayscale of eye areas is smaller than the sum of the center area of two eyebrows.

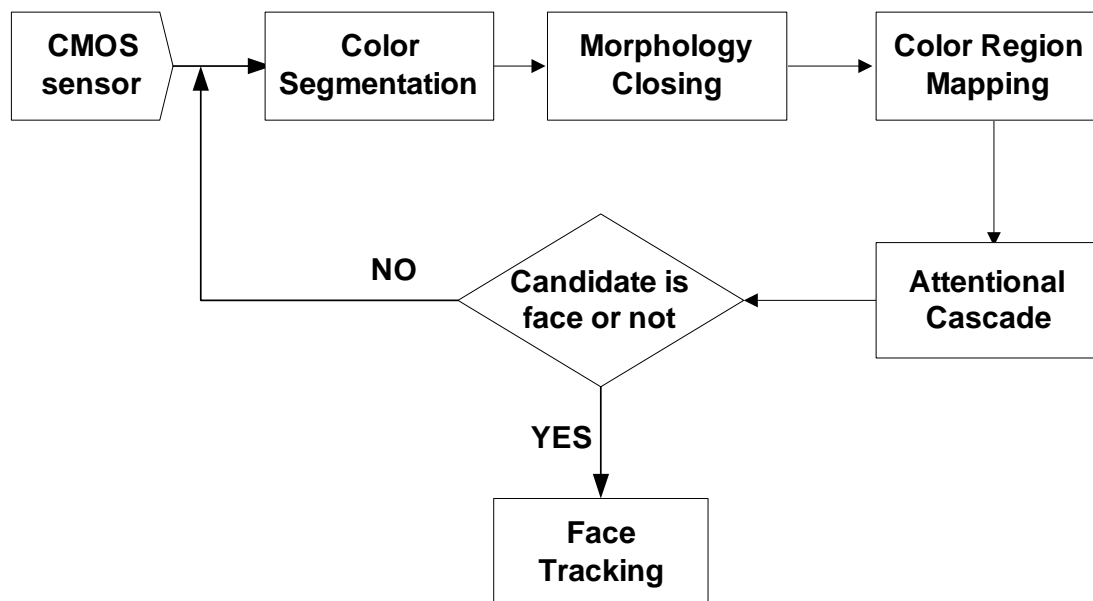


Figure 4-1 The flowchart of face detection

- d. The sum of grayscale of the adjacent cheek areas is less than one fixed threshold.

One successful example is shown in Figure 4-2 (e).

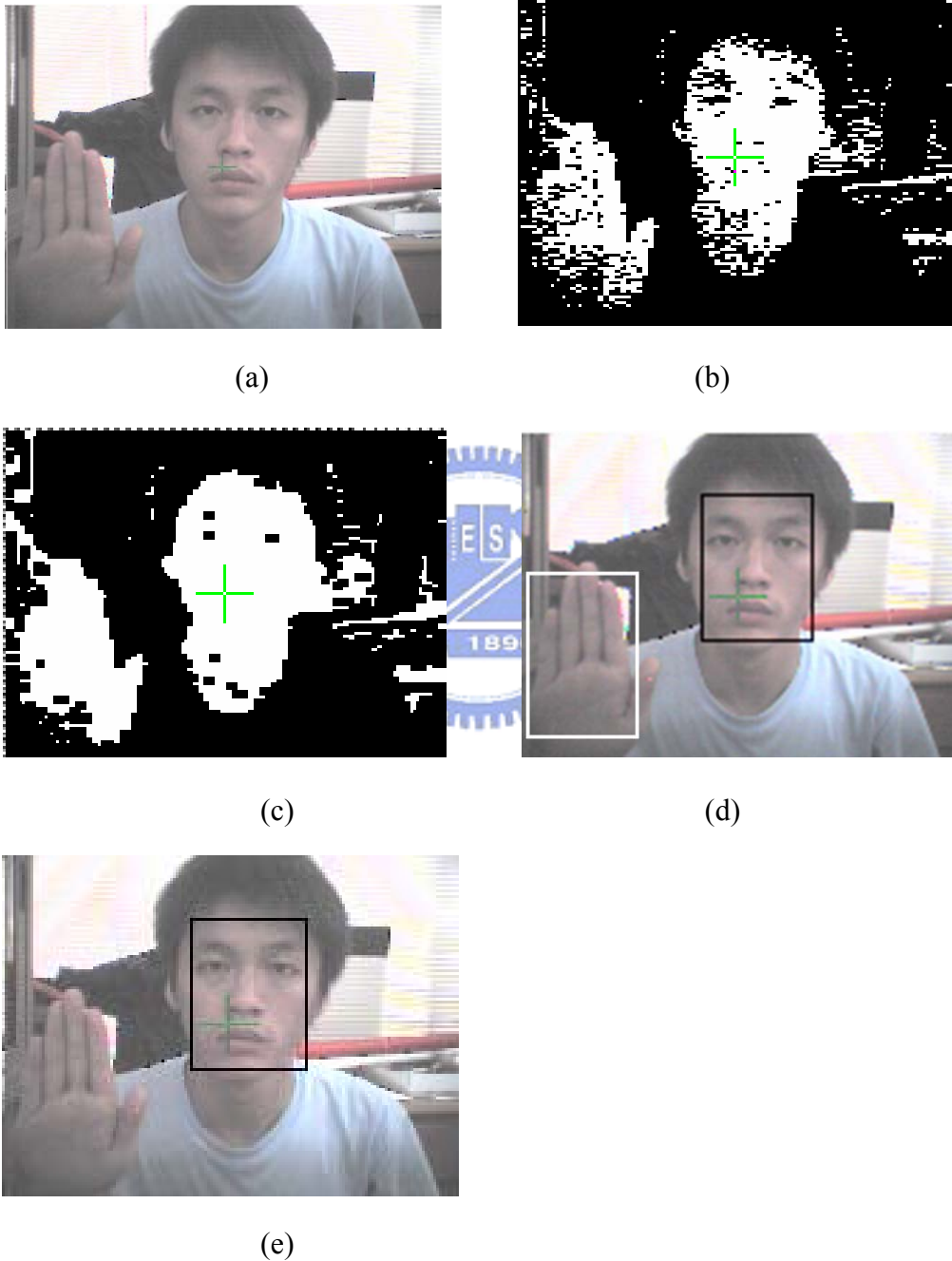


Figure 4-2 The results of face detection (a) is the testing image. (b) is the result of color segmentation and (c) is the result of closing operation. (d) is the candidate of face regions. (e) is the final result via the attentional cascade.

## 4.2 Face Tracking

After a face is detected in the initial state, face tracking can be accomplished in the subsequent images by applying the adaptive YCrCb 3D color distribution model. This statistical model determines the proper threshold values for each color channel of the face image in the tracking mode. Let  $Y_1(Y_2)$  be lower(upper) threshold of skin color of Y channel,  $Cr_1(Cr_2)$  be lower(upper) threshold of skin color of Cr channel and  $Cb_1(Cb_2)$  be lower(upper) threshold of skin color of Cb channel. The threshold values used to segment the face regions are updated by the following equations[29] :

$$C_2 = \arg \min_i \left[ \left( \sum_{i=C_{\min}}^{C_{\max}} C_{hist}(i) - N_{total} \times S \right) > 0 \right] \quad (4-1)$$

$$C_1 = \arg \min_i \left[ \left( \sum_{i=C_{\min}}^{C_{\max}} C_{hist}(i) - N_{total} \times S \right) > 0 \right] \quad (4-2)$$

Where  $C$  denotes the channel of color spaces, Y, Cr or Cb and  $C_{hist}$  is the histogram of the  $C$  channel.  $S$  is a scale factor from 0 to 1 to reject the undesired histogram such as eye or eyebrow; in this case the scale factor is set to 0.1.  $N_{total}$  is the total pixel number of previous face region. Figure 4-3 shows an example of computing the threshold values. The face tracking method utilizes the color distribution model to update the threshold values of skin color from the previous face region of sample instant  $t-1$ . The threshold values can be adjusted dynamically to accommodate the unexpected changes of lighting conditions.

## 4.3 Face image preprocessing

To improve the efficiency of extracting facial features, we perform face image preprocessing including face normalization and feature region localization. The segmented front-view face region is firstly normalized to 160x120 grayscale image.

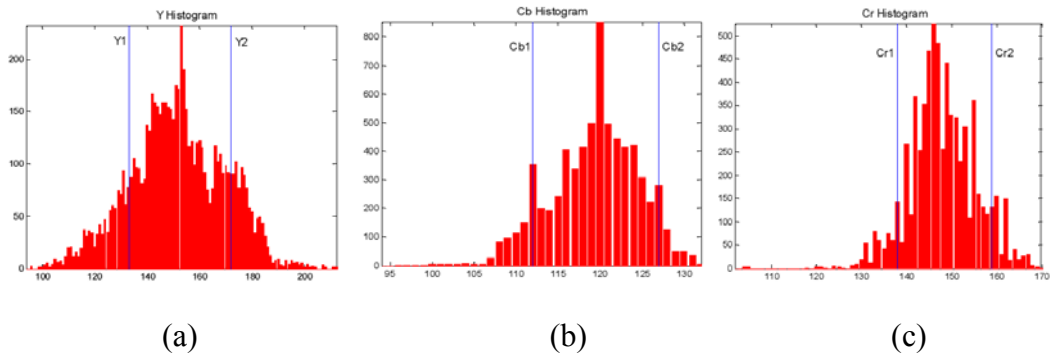


Figure 4-3 The lower and upper thresholds of each color channel (a) Y1 and Y2 of Y channel (b) Cb1 and Cb2 of Cb channel (c) Cr1 and Cr2 of Cr channel

Three reference points are found on both pupils of the eyes and at the center of the mouth. As shown in Figure 4-4, the face region is divided into 14 relevant regions of interest (ROI), each containing one feature point.

### 4.3.1 Face normalization

Face region is normalized to a 160x120 gray-value image by using bilinear interpolation [30]. Image bilinear interpolation estimates function value  $f(x', y')$  based on the known values of nearby pixel  $f(x, y)$ . As shown in Figure 4-5, assume that the black points are unknown values and we estimate them linearly by using the

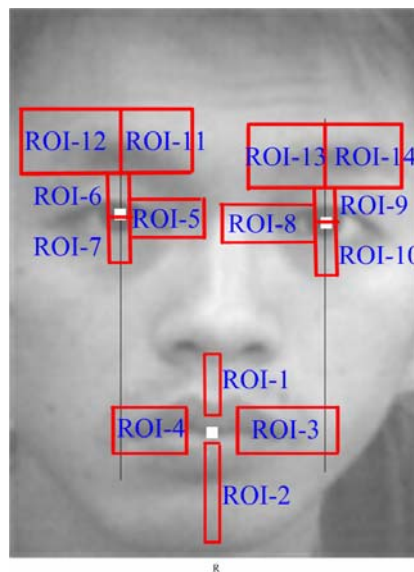


Figure 4-4 The 14 facial regions of interest

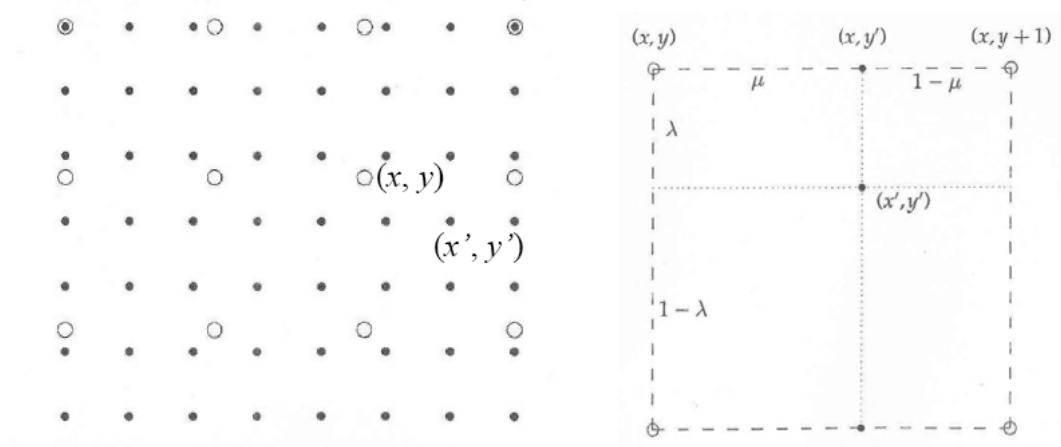


Figure 4-5 Image interpolation

known white points. The bilinear interpolation function can be expressed as follows :

$$f(x', y') = (1 - \lambda)[(1 - u)f(x, y) + uf(x, y + 1)] + \lambda[(1 - u)f(x + 1, y) + uf(x + 1, y + 1)] \quad (4-3)$$

where,  $\lambda = \frac{x' - x}{x + 1 - x}$ ,  $u = \frac{y' - y}{y + 1 - y}$

### 4.3.2 Feature Region Localization

For feature extraction, the relevant regions of interest are determined first. Three reference points located on both pupils of the eye and at the center of the mouth are detected for this purpose by using an adaptive threshold method, integral optical density (IOD) [31]. The pupils are supposed to be the darkest in the upper face regions, so the binary image of the pupils can be segmented by selecting a proper threshold. However, the entire gray-values of the face may vary from image to image, and a fixed threshold does not acquire the binary image of pupils successfully. Consequently, to overcome this problem, IOD is employed to adaptively determine the threshold. The IOD is defined as :

$$IOD = \frac{\iint Y_k(x, y) dx dy}{\iint_D dx dy} \quad (4-4)$$

where  $Y_k(x, y)$  is binary image with a certain threshold value  $k$ .

In this design, the binary image of the pupils is segmented with the value of IOD set to 0.05 in the rectangle area of 30x30 pixels. Namely, we choose 5% of the darkest gray-value to locate the positions of the pupils. The positions of these two rectangle regions are illustrated in Figure 4-6 (a). In this case, the coordination of the left pupil region is from pixel(15, 40) to pixel(45, 70). The coordination of the right pupil region is defined from pixel(75, 40) to pixel(105, 70). After obtaining the binary image of the pupils, we use 2D-histogram mapping to locate the positions of the pupils accurately. On the other hand, the center of the widest peak will define the vertical position of the middle of the mouth between pix100 to pix150 in the vertical direction, and the horizontal position of the mouth is the center of two pupils. In this way, three reference points can be found. Figure 4-6 (b) shows typically detected result of these points.

Subsequently, we utilize these three reference points to further divide the face region into 14 regions where each facial point is extracted individually. The detailed descriptions of 14 ROIs are given in Table 4-1.

#### 4.4 Gabor Wavelet Transformation

Features based on Gabor filters have been used in image processing due to their powerful properties [32-35]. In particular, the most favorite one is to remove the variability in lighting, rotation, small shift or deformation in the local feature areas. In this thesis, we will apply Gabor wavelets-based features instead of grayscale features to extract facial points.

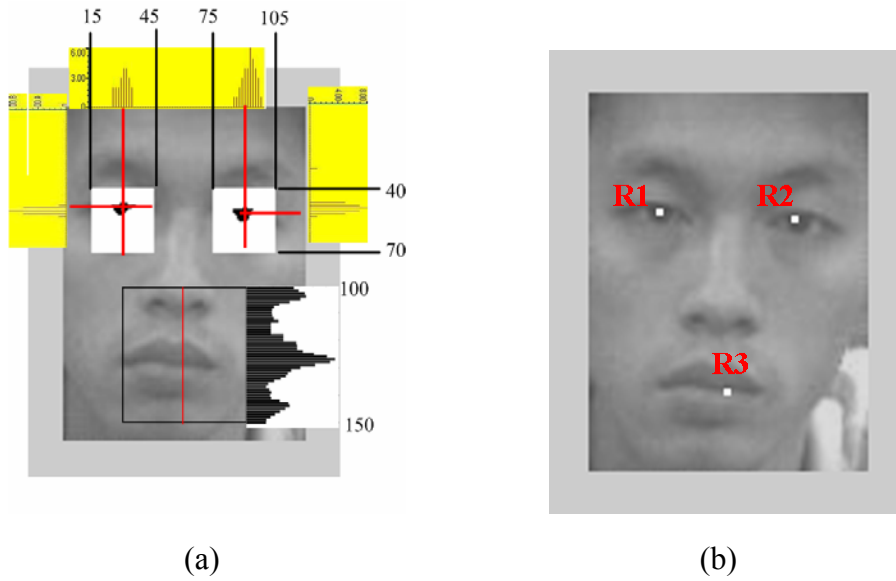


Figure 4-6 (a) the diagram of locating three reference points and (b) is the result

Table 4-1 The detailed descriptions of 14 ROIs

ROI-1	Range from pixel( $R3\_x-8, R3\_y-18$ ) to pixel( $R3\_x+8, R3\_y-3$ ).
ROI-2	Range from pixel( $R3\_x-8, R3\_y+3$ ) to pixel( $R3\_x+8, R3\_y+30$ ).
ROI-3	Range from pixel( $R3\_x+15, R3\_y-10$ ) to pixel( $R3\_x+40, R3\_y+10$ ).
ROI-4	Range from pixel( $R3\_x-40, R3\_y-10$ ) to pixel( $R3\_x-15, R3\_y+10$ ).
ROI-5	Range from pixel( $R1\_x+3, R1\_y-5$ ) to pixel( $R1\_x+15, R1\_y+5$ ).
ROI-6	Range from pixel( $R1\_x-3, R1\_y-10$ ) to pixel( $R1\_x+3, R1\_y$ ).
ROI-7	Range from pixel( $R1\_x-3, R1\_y$ ) to pixel( $R1\_x+3, R1\_y+10$ ).
ROI-8	Range from pixel( $R2\_x-15, R2\_y-5$ ) to pixel( $R2\_x-3, R2\_y+5$ ).
ROI-9	Range from pixel( $R2\_x-3, R2\_y-10$ ) to pixel( $R2\_x+3, R2\_y$ ).
ROI-10	Range from pixel( $R2\_x-3, R2\_y$ ) to pixel( $R2\_x+3, R2\_y+10$ ).
ROI-11	Range from pixel( $R1\_x+5, 10$ ) to pixel( $R1\_x+30, R1\_y$ ).
ROI-12	Range from pixel( $10, 10$ ) to pixel( $R1\_x, R1\_y-20$ ).
ROI-13	Range from pixel( $R2\_x-30, 10$ ) to pixel( $R2\_x-5, R2\_y$ ).
ROI-14	Range from pixel( $R2\_x, 10$ ) to pixel( $110, R2\_y-20$ ).



In its general form, a 2D Gabor wavelet kernel function can be described as [33] :

$$\psi_j(\vec{k}_j, \vec{x}) = \frac{\vec{k}_j^2}{\sigma^2} \exp\left(-\frac{\vec{k}_j^2 x^2}{2\sigma^2}\right) [\exp(i\vec{k}_j \bullet \vec{x}) - \exp(-\frac{\sigma^2}{2})] \quad (4-5)$$

$$\vec{k}_j = \begin{pmatrix} k_v \cos \phi_v \\ k_v \sin \phi_v \end{pmatrix}, \quad k_v = 2^{\frac{v+2}{2}} \pi, \quad \phi_u = u \frac{\pi}{N}$$

where  $i$  denotes a complex number and  $\sigma$  is the standard deviation of the Gaussian envelope,  $\vec{k}_j$  is the wave vector and different Gabor kernel functions are controlled by different  $\vec{k}_j$ .  $u$  is the orientation and  $v$  is the frequency of the

filter. The multiplicative factor  $\frac{\vec{k}_j^2}{\sigma^2}$  ensures that filters tuned to different spatial frequency bands have approximately equal energies. The term  $\exp(-\frac{\sigma^2}{2})$  is subtracted to render the filters insensitive to illumination.  $\vec{x}$  represents the coordinates of a given pixel. In addition, because facial expression features are mainly described by high frequency components, 3 frequencies ( $v=0,1,2$ ) and  $N=8$  orientations ( $u=1, \dots, 8$ ) are used to yield  $3 \times 8$  filters. But, in the application of robotic emotional recognition, evaluating all 24 filters to convolving the face image is quite time consuming. Consequently, only six important filters are used to extract facial points.

While Gabor wavelets kernel functions have been defined in (4-5), a single Gabor feature is obtained by convolving one of the filters (a specific  $v$  and  $u$ ) with the original image. For pixel  $\vec{x}$  in image  $I(x)$ , a Gabor jet  $J$  is defined such that :

$$J_j(\vec{x}) = I(\vec{x}) * \psi_j(\vec{k}_j, \vec{x}) = \int I(\vec{x}) \psi_j(\vec{k}_j, \vec{x}) dx dy \quad (4-6)$$

where  $I(\vec{x})$  represents gray-value image. When we apply Gabor filters at a specific point  $\vec{x}$ , we get multiple filter responses for that point. The results of transformation

are described by complex values, so their amplitude can be calculated as the final transformation results. Thus, each feature point can thus be represented by such a Gabor jet in place of just its gray value.

Figure 4-7 shows examples of response of Gabor filters (size is 11x11) and Figure 4-8 is the original face image. Obviously, the facial features of filtered images are generated with different frequencies and orientations of Gabor wavelet filters. Features of the filtered images have less illumination variation compared with the original grayscale image, which has a light from left. In order to reduce the computing cost of convolution of all 24 filters, we just use 6 filters to extract Gabor jets. 3 frequencies ( $v=0,1,2$ ) and 2 orientations( $u=7,8$ ) are the best choices to filter out the original face image. It is observed that these six filters can filter out the edges of eyebrows, eyes and mouth obviously. Six filtered facial images are illustrated in Figure 4-9.

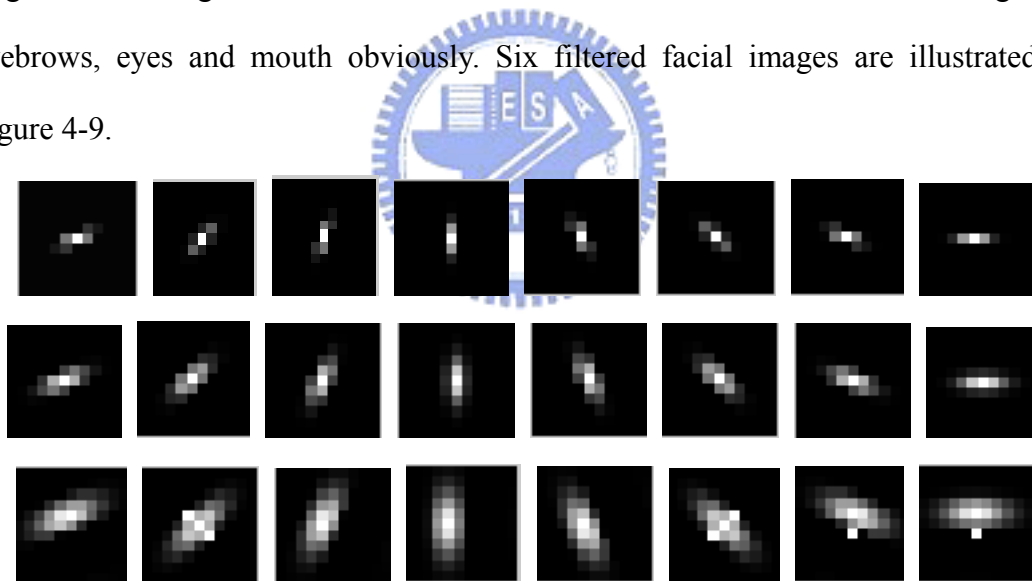


Figure 4-7 Different frequencies and orientations of Gabor wavelet filters



Figure 4-8 the original face image

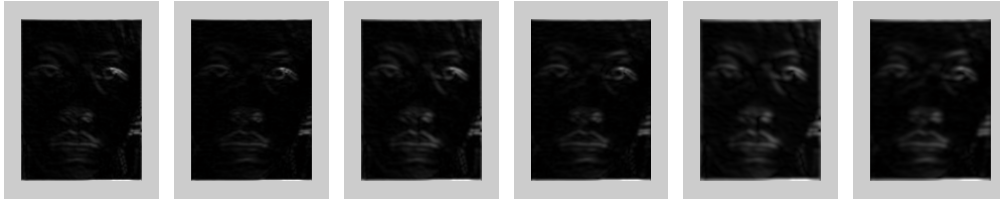


Figure 4-9 Six selected filtered facial images

We observe that the edges of eyebrows, eyes and mouth can be emphasized after performing Gabor filters. Since these edges are the regions where facial points are located, binary images of the edges can be approximated by setting a proper threshold of the filtered images. As a result, we can narrow the search regions of facial points validly only in the intersection of the binary image of the edges and the relevant ROIs.

In this design, a binary image of facial edges will be obtained by summing Gabor jets of six selected filtered images into one image, and then we set  $IOD=0.2$  to separate the facial edges in the upper face region and set  $IOD=0.2$  in the mouth region. The upper face region ranges from pixel(10, 10) to pixel(110, 70) and the mouth region ranges from pixel( $R3\_x-40$ ,  $R3\_y-20$ ) to pixel( $R3\_x+40$ ,  $R3\_y+30$ ).

By summing the selected 6 filtered images, these filtered images interfered with illumination can be weakened available. An example of the binary image of the face is shown in Figure 4-10. Figure 4-10 (a) is the result of summing six Gabor jets into one image, we see that the facial edges can be detected apparently. Accordingly, the binary image of edge is obtained by setting a proper threshold and the result is illustrated in Figure 4-10 (b). Eventually, we identify that the intersection of 14 ROIs and the binary image of edge regions are our final ROIs where facial points are located (See Figure 4-10 (c)). Afterwards, we use similarity function to match facial points efficiently only in these areas, and the cost of searching facial points can be saved greatly.

## 4.5 Feature Points Extracting

The proposed method of extracting facial points is composed of two phases, training phase and testing phase. In the training phase, some Gabor jets on defined facial points are collected for extracting facial points automatically in the testing phase. Then, for a certain searching point in the testing phase, 5x5 pixels sliding window with 6 different jets are applied to find the most matching point by using a phase-sensitive similarity function. The facial points involving in the changes of facial expressions will be extracted. The exact positions of these points are shown in Figure 4-11. Table 4-2 depicts the definitions of 14 facial points.

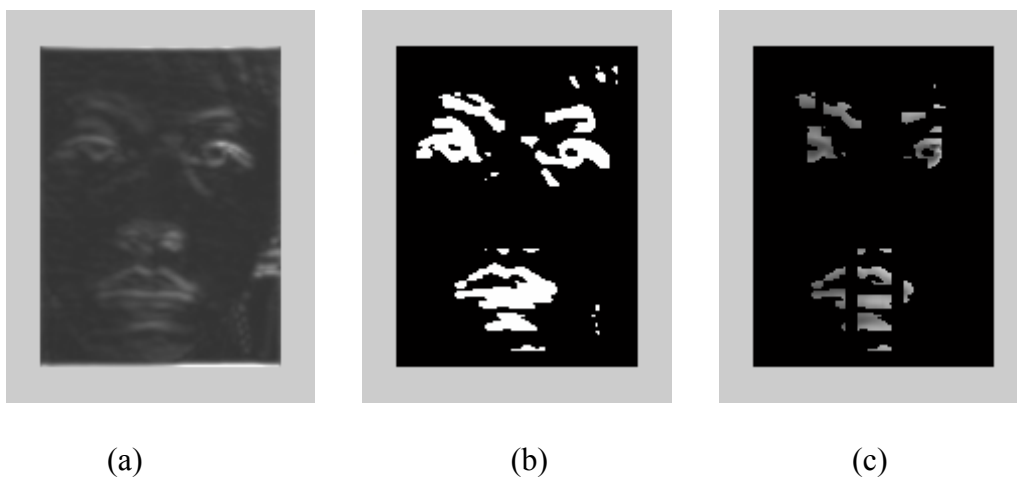


Figure4-10 (a) is the image by summing 6 Gabor jets. (b) is binary image of facial edges.(c) is the intersection of 14 ROIs and the binary image of edge regions on

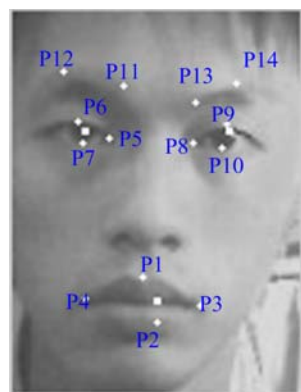


Figure 4-11 14 facial points

Table 4-2 The definitions of 14 facial points

	Descriptions
P1	Mouth top
P2	Mouth bottom
P3	Left mouth corner
P4	Right mouth corner
P5	Inner corner of the right eye
P6	Top of the right eye
P7	Bottom of the right eye
P8	Inner corner of the left eye
P9	Top of the left eye
P10	Bottom of the left eye
P11	Inner corner of right eyebrow
P12	Top of right eyebrow
P13	Inner corner of left eyebrow
P14	Top of left eyebrow

#### 4.5.1 The Training Phase of Feature Points Extracting

The overall procedure of feature point extraction in the training phase is given in left part of Figure 4-12. First, some expressionless images are collected into database, and then we select these facial points manually and evaluate Gabor jets both on and around these points. A 5x5 pixels patch around the facial points is considered to extract their Gabor jets by filtering the defined point with a bank of 6 Gabor filters at 2 orientations and 3 frequencies. In other words, 150 (5x5x6) Gabor features are used to represent one facial point. Finally, Gabor jets of 14 feature points are collected to the feature point database in order to extract facial points automatically in the testing phase.

### 4.5.2 The Testing Phase of Feature Points Extracting

In the testing phase(see right part of Figure 4-12), the final ROIs are first defined by the intersection of 14 relevant regions of interest and the binary image of edge regions; subsequently, we evaluate each Gabor features of the 5x5 pixels patch by scanning the overall ROIs. A phase-sensitive similarity function is applied to match these Gabor features of possible facial pixels, and finally the position with the highest response reveals the exactly facial position in question. The phase-sensitive similarity function [33] :

$$S_{\phi}(J, J') = \frac{\sum_j a_j a_j' \cos(\phi_j - \phi_j' - \vec{d} \vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} \quad (4-7)$$

The Gabor vector of each point can be written as  $J_j = a_j \exp(i\phi_j)$ , where  $a_j$  is

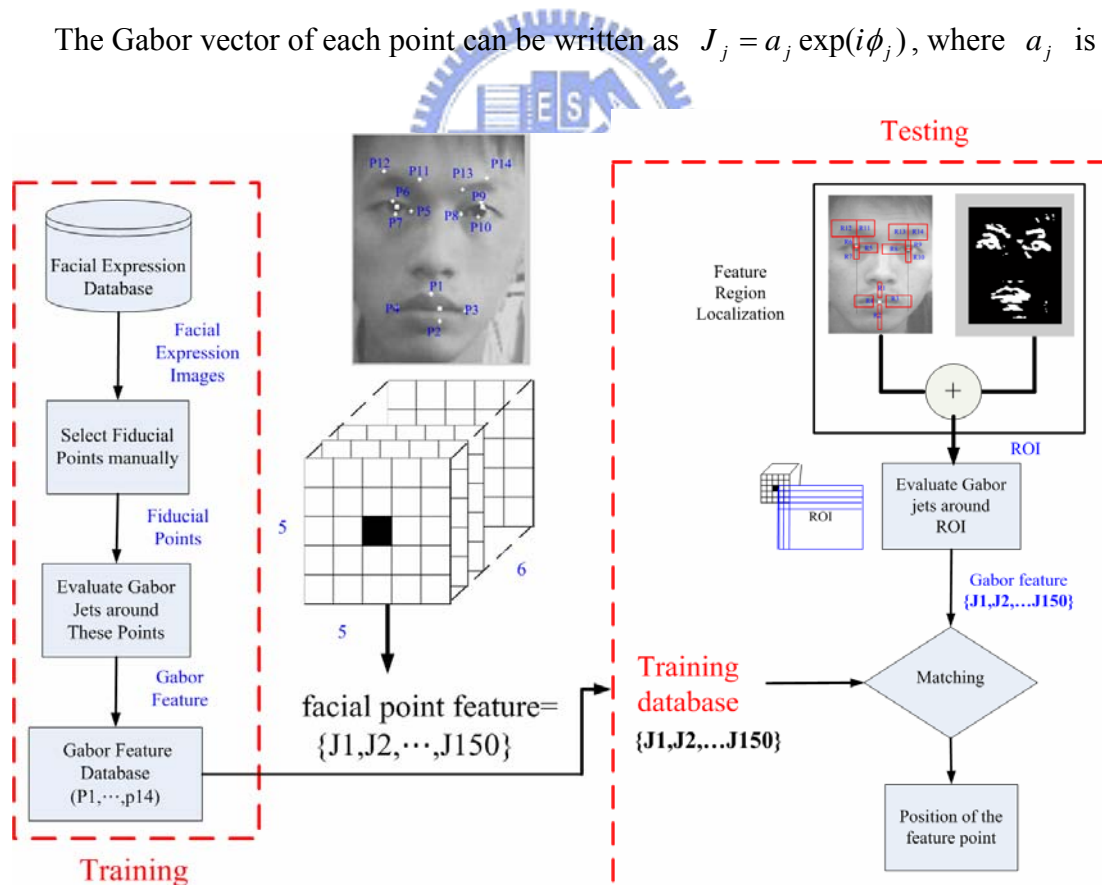


Figure 4-12 The overall procedure of feature point extraction in the training phase and testing phase

the amplitude which slowly varies the position and  $\phi_j$  is the phase.  $\vec{d}$  is a small relative displacement between two jets J and J'. The displacement  $\vec{d}$  can be estimated by maximizing  $S_\phi$  in its Taylor expansion around  $\vec{d}=0$ , which is a constrained fit of the two-dimensional  $\vec{d}$  to the 6 phase differences  $\phi_j - \phi_{j'}$ .

#### 4.6 Feature Extraction Evaluation

The displacement information of these key points is taken as the feature values in the proposed facial expression recognition system. To analyze expressional feature values systematically, we adopted AU-coded descriptions of facial expression in the FACS to describe the relationship between AU and the displacement of facial points [36]. Designed for human's observations to subtle muscle changes of facial appearances, facial action coding system (FACS)[3][4] is a well-known method to analyze facial activities. FACS provides a linguistic description of all possible facial changes in terms of 44 Action Units (AUs).

Referring to [36], Table 4-3 shows the association of facial AUs of four expressions, and the facial AUs pertaining to five expressions are also illustrated in Figure 4-13 [4]. According to these AUs, we can establish the association of the AUs and the displacement of the fiducial points as shown in Table 4-4. We can measure scientifically the facial changes of AUs via displacement of the facial points. At last,

Table 4-3 The association of 5 facial expressions to AU combination

Emotional category	Visual Cues			
Anger	AU4	AU7	AU23	AU24
happiness	AU7	AU12	AU25	
Sadness	AU1	AU15		
Surprise	AU1	AU5	AU7	AU27

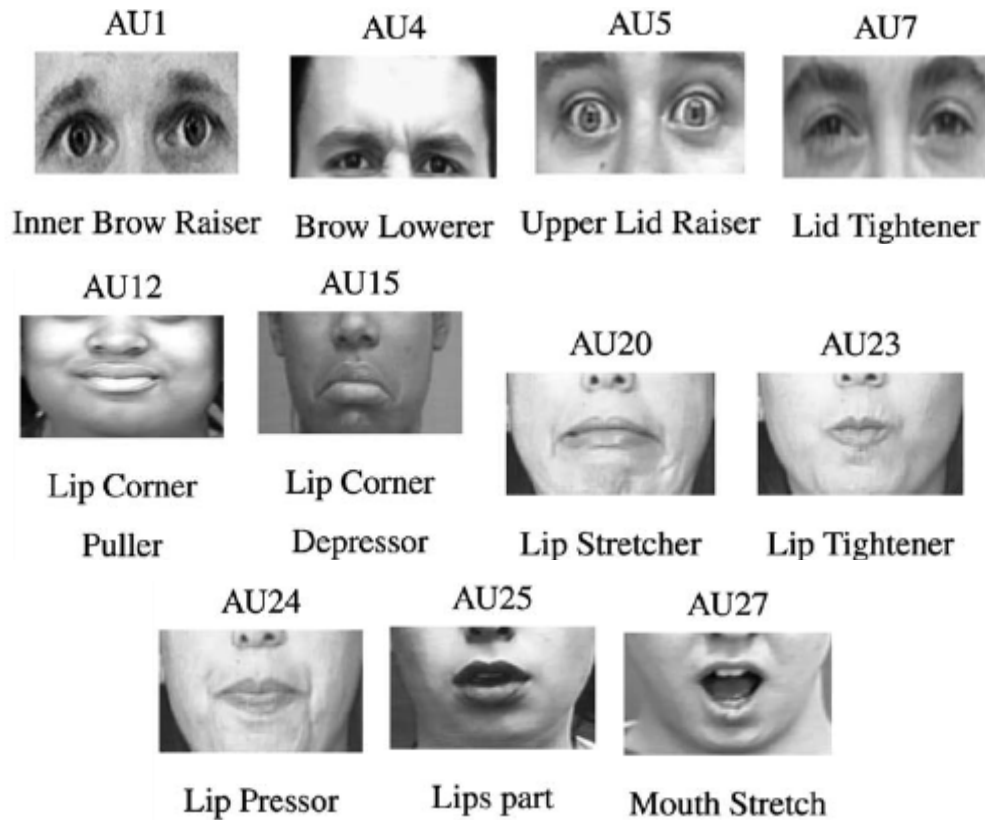


Figure 4-13 The list of AUS related to five expressions [4]

Table 4-4 Feature points based descriptions for AUS

AUs	Facial Visual Cues	Description
AU1	$\overline{P9P15}$ increased	Inner brow raiser
AU4	L1, $\overline{P9P15}$ , $\overline{P10P16}$ decreased,	Brow lower
AU5	$\overline{P10P12}$ increased	Upper eyelid raiser
AU7	$\overline{P10P12}$ decreased ,	Eyelid tighter
AU12	L2 decreased, $\overline{P3P4}$ increased, $\overline{P3P11}$ decreased	Lip corner puller
AU15	L2 increased, $\overline{P3P11}$ decreased, L3 decreased	Lip corner depressor
AU20	L2 non-change, $\overline{P3P4}$ increased	Lip stretcher
AU23	$\overline{P1P2}$ , $\overline{P3P4}$ decreased	Lip tightener
AU24	$\overline{P1P2}$ decreased, $\overline{P3P4}$ non-change	Lip pressor
AU25	$\overline{P1P2}$ , $\overline{P3P4}$ increased	Lips part**
AU27	$\overline{P1P2}$ increased, $\overline{P3P4}$ decreased	Mouth Stretch

\*L1 : the vertical displacement of right brow upper center and right eyelid lower center.

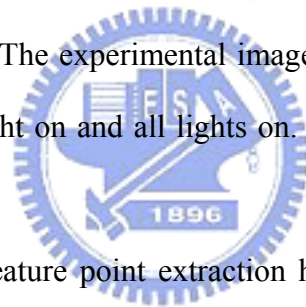
L2 : the vertical displacement of right eyelid inner corner and right lip corner.



we summarize 16 feature values as AUs for recognizing 5 emotional expressions. The detailed descriptions of these 16 feature values associated to defining facial points are shown in Table 4-5.

#### **4.7 The Experimental Results of Facial Points Extraction**

For evaluating the efficiency of Gabor wavelet based feature extraction, a set of images containing different expressions and different lighting conditions from AR database [37] were used. The AR Face database was created by Alex Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains frontal view faces with different facial expressions, illumination conditions, and occlusions, but only the images of different expressions and different illumination conditions are experimented. The experimental images are neutral expression, smile, scream, left light on, right light on and all lights on. Figure 4-14 shows examples of the images.



Gabor wavelets based feature point extraction has been tested on a set of 180 images from 30 persons and each person has six data. Neutral expressions of ten persons are used as the matching data in facial point extraction database. Therefore, neutral expression has 20 images and each of other expressions has 30 images. To evaluate the performance of facial point matching, each of the extracted facial points is also compared to the true point. We regard the points displaced within 5 pixels of true points as the successful points. The recognition rates of facial point extraction under various expressions and illumination variety are shown in Table 4-6. Typical results of successful extractions are also illustrated in Figure 4-15.

Table 4-6 shows that most facial points can be detected accurately under lighting conditions variation and appearing expressions. It also reveals that the proposed method of feature points matching could detect vast majority of features automatically

Table 4-5 The detailed descriptions of these 16 feature values

Dimension	Physical description
d1	The vertical displacement of left brow inner corner and left eyelid inner corner. $\overline{P5P13}$
d2	The vertical displacement of right brow inner corner and right eyelid inner corner. $\overline{P9P15}$
d3	The vertical displacement of left brow upper center and left eyelid upper center. $\overline{P6P14}$
d4	The vertical displacement of right brow upper center and right eyelid upper center. $\overline{P10P16}$
d5	The vertical displacement of left eyelid upper center and left eyelid lower center. $\overline{P6P8}$
d6	The vertical displacement of right eyelid upper center and right eyelid lower center. $\overline{P10P12}$
d7	The vertical displacement of left eyelid inner corner and left eyelid upper corner. $\overline{P6P5}$
d8	The vertical displacement of right eyelid inner corner and right eyelid upper corner. $\overline{P10P9}$
d9	The vertical displacement of upper lip corner and the average position of two pupils. $P1 \frac{R1 + R2}{2}$
d10	The vertical displacement of lower lip corner and the average position of two pupils. $P2 \frac{R1 + R2}{2}$
d11	The horizontal displacement of left lip corner and right lip corner. $\overline{P3P4}$
d12	The vertical displacement of upper lip corner and lower lip corner. $\overline{P1P2}$
d13	The vertical displacement of left lip corner and upper lip corner. $\overline{P1P4}$
d14	The vertical displacement of right lip corner and upper lip corner. $\overline{P1P3}$
d15	The vertical displacement of left lip corner and lower lip corner. $\overline{P2P4}$
d16	The vertical displacement of right lip corner and lower lip corner. $\overline{P2P3}$



Figure 4-14 Examples of the AR Face Database (a)neutral expression (b) smile (c)scream (d)left light on (e)right light on (f)all lights on [37]

Table 4-6 The results of facial point extraction

	Neutral (20*1)	Left light on (30*1)	Right light on(30*1)	All lights on(30*1)	Smile (30*1)	Scream (30*1)
p1	92%	91%	97%	97%	76%	69%
p2	75%	91%	82%	88%	79%	82%
p3	88%	91%	100%	97%	85%	76%
p4	79%	97%	88%	94%	85%	76%
p5	100%	100%	97%	97%	94%	85%
p6	100%	100%	97%	97%	100%	88%
P7	100%	100%	97%	97%	100%	88%
P8	100%	100%	100%	100%	100%	76%
P9	100%	100%	100%	100%	100%	88%
p10	100%	100%	100%	100%	100%	82%
p11	100%	100%	97%	97%	100%	97%
p12	100%	100%	97%	91%	97%	97%
p13	100%	100%	100%	100%	100%	97%
p14	100%	97%	100%	97%	100%	88%



Figure 4-15 Accurate extraction of all facial points

from the Gabor filter components. Most detection rates of facial points are between 80% and 100% on the average.

Notwithstanding the AR Face Database provides various face images, only neutral expression is under light changing conditions. These images will not adequately demonstrate that the proposed method can robustly extract facial points under illumination variation. Therefore, we have built in the lab a database of five facial expressions, each of which is under four lighting conditions. These emotional expressions are anger, happiness, neutral, sadness and surprise, and four lighting conditions are normal light, right light on, left light on and all lights on. Figure 4-16 shows examples of the database.

The results of facial point extraction are as shown in Table 4-7~ 4-11. From these experimental results, we find that Gabor based matching method is quite robust against the changes of lightings and the transitions of facial expressions. The performances of extracting most facial points in various lighting conditions are as accurate as in normal light except for a few points on p2, p3 and p8. Extraction rates of most points are above 90% and some ones reach 100%. We conclude that Gabor

wavelets based feature extraction method provides a more general scenario to extract facial features under illumination variation.

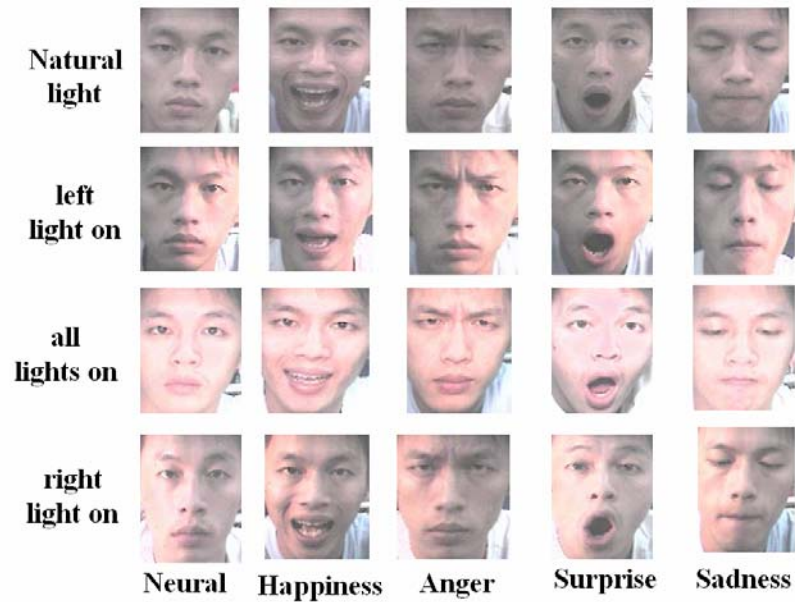


Figure 4-16 Five categories of facial expressions under four lighting conditions



Table 4-7 The results of facial point extraction in neural expression

neutral	natural light	left light on	all lights on	right light on
p1	100%	100%	100%	100%
p2	100%	90%	90%	100%
p3	100%	80%	90%	100%
p4	100%	100%	100%	100%
p5	100%	100%	100%	100%
p6	100%	100%	100%	100%
p7	100%	100%	100%	100%
p8	80%	100%	80%	80%
p9	100%	100%	100%	100%
p10	100%	100%	100%	100%
p11	100%	100%	90%	100%
p12	100%	100%	100%	100%
p13	100%	100%	100%	100%
p14	100%	80%	100%	100%

Table 4-8 The results of facial point extraction in happiness expression

happiness	natural light	left light on	all lights on	right light on
p1	100%	90%	90%	90%
p2	90%	90%	100%	90%
p3	90%	100%	50%	90%
p4	90%	90%	100%	100%
p5	100%	100%	100%	100%
p6	100%	100%	100%	100%
p7	100%	100%	100%	100%
p8	100%	100%	100%	90%
p9	100%	100%	100%	100%
p10	100%	100%	100%	100%
p11	100%	100%	100%	100%
p12	100%	100%	100%	100%
p13	100%	100%	100%	100%
p14	100%	100%	100%	100%

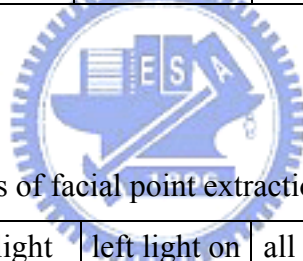


Table 4-9 The results of facial point extraction in anger expression

anger	natural light	left light on	all lights on	right light on
p1	100%	100%	100%	90%
p2	100%	90%	70%	90%
p3	100%	100%	100%	100%
p4	100%	100%	100%	90%
p5	100%	100%	100%	100%
p6	100%	100%	100%	90%
p7	100%	100%	100%	100%
p8	100%	80%	100%	100%
p9	100%	100%	100%	100%
p10	100%	100%	100%	100%
p11	90%	100%	100%	90%
p12	100%	100%	100%	100%
p13	90%	100%	90%	80%
p14	100%	70%	80%	100%

Table 4-10 The results of facial point extraction in surprise expression

surprise	natural light	left light on	all lights on	right light on
p1	90%	100%	100%	90%
p2	100%	100%	80%	90%
p3	100%	90%	90%	80%
p4	90%	80%	80%	80%
p5	100%	90%	100%	100%
p6	80%	90%	100%	100%
p7	100%	100%	100%	100%
p8	70%	80%	60%	100%
p9	100%	100%	90%	100%
p10	100%	100%	100%	100%
p11	100%	100%	100%	100%
p12	100%	100%	100%	100%
p13	90%	90%	100%	100%
p14	100%	90%	100%	90%



Table 4-11 The results of facial point extraction in sadness expression

sadness	natural light	left light on	all lights on	right light on
p1	90%	90%	100%	90%
p2	80%	80%	60%	60%
p3	100%	100%	100%	100%
p4	100%	100%	100%	90%
p5	100%	100%	100%	100%
p6	100%	100%	100%	100%
p7	100%	100%	100%	100%
p8	100%	80%	90%	90%
p9	100%	100%	100%	100%
p10	100%	100%	100%	100%
p11	100%	100%	100%	90%
p12	100%	100%	100%	100%
p13	100%	90%	100%	100%
p14	100%	90%	100%	90%

# Chapter 5

## Experimental Results

This chapter presents several experimental results of the proposed algorithm in facial expression recognition, including emotion classification and learning. The feature values are used to category facial expressions using a SVM classifier. In addition, we have built a specific learning database to verify the proposed learning algorithm. We have also implemented the proposed algorithm on an entertainment robot platform, the Momobear, which gives a user active interactions corresponding to estimated expressions.

### 5.1 Experimental Results of Facial Expression Recognition

In this thesis, Gaussian kernel function is used to map the facial data to a high dimensional space [12] :

$$K(x_1, x_2) = \exp\left(\frac{-\|x_1 - x_2\|^2}{c}\right) \quad (5-1)$$

where  $c$  is a kernel parameter. In the experiments, the parameter  $c$  of Gaussian kernel of facial expression recognition system is set to 0.1. The average recognition results of facial expression on the AR Face Database are shown in Table 5-1 and Table 5-2. Table 5-1 shows the recognition rate of the subjects whose features were used to train the SVM classifiers and Table 5-2 is the result of testing subjects. We observe that the recognition rates of recognizing facial expressions under illumination variation are, respectively 86.6% for the training subjects and 81.7% for the testing subjects.

The recognition rates of the database built in the lab are shown in Tables 5-3~ 5-6. The SVM classifiers were trained by ten persons under the normal lighting conditions,



Table 5-1 Average recognition results of training subjects.

	Neutral	Scream	Smile	left light on	right light on	all lights on
Neutral	9	0	1	9	7	9
Scream	1	10	1	0	1	0
Smile	0	0	8	1	2	1
average recognition rate	90%	100%	80%	90%	70%	90%

Table 5-2 Average recognition results of testing subjects

	Neutral	Scream	Smile	left light on	right light on	all lights on
Neutral	24	2	4	23	27	23
Scream	5	26	2	6	2	1
Smile	1	2	24	1	1	6
average recognition rate	80%	86%	80%	77%	90%	77%

and we will verify if the SVM classifiers can recognize facial expressions correctly as illumination varies. From the results, we see that the proposed algorithm maintains sufficient recognition rates even the lighting conditions change heavily. In summary, the SVM classification results based on Gabor feature extraction give the satisfactory recognition rates under light changing conditions (78% left light on, 80% all lights on, 86% right light on) as well as the rate of the normal lighting condition (82% on the average).

## 5.2 Experimental Results of the Proposed Learning Algorithm

The developed FER system not only recognizes the master's emotion in a natural environment but also accommodates itself to new users. In other words, the proposed FER system can learn facial data of new person incrementally through the developed

Table 5-3 Average recognition results of normal light

natural light	neutral	happiness	anger	surprise	sadness
neutral	9	1	1	0	1
happiness	0	8	1	1	0
anger	1	0	7	0	1
surprise	0	1	0	9	0
sadness	0	0	1	0	8
recognition rate	90%	80%	70%	90%	80%

Table 5-4 Average recognition results of left light on

left light on	neutral	happiness	anger	surprise	sadness
neutral	8	2	1	1	1
happiness	1	8	0	0	0
anger	0	0	7	0	1
surprise	0	0	0	8	0
sadness	1	0	2	1	8
recognition rate	80%	80%	70%	80%	80%

Table 5-5 Average recognition results of all light on

all lights on	neutral	happiness	anger	surprise	sadness
neutral	8	1	1	1	0
happiness	0	8	0	0	1
anger	0	0	8	0	0
surprise	1	0	1	9	1
sadness	1	1	0	1	8
recognition rate	80%	80%	80%	80%	80%

Table 5-6 Average recognition results of right light on

right lights on	neutral	happiness	anger	surprise	sadness
neutral	8	0	1	0	0
happiness	0	8	0	0	1
anger	1	0	8	0	0
surprise	0	2	0	10	0
sadness	1	0	1	0	9
recognition rate	80%	80%	80%	100%	90%

HRI interface. This section will demonstrate how the proposed learning algorithm adjusts its SVM classifier to specific facial expressions of a new subject.

To verify the proposed learning algorithm, we have built a learning database. The images were acquired by using an embedded vision system [29] and a CMOS image sensor [38] in an arranged environment. Images from ten persons are stored in the database and each person gives ten expression data for each emotion category, so there are 100 data sets for each emotion category. The SVM classifier was trained by using a set of 20 data for each facial expression. These 20 data come from four persons as shown in Figure 5-1 and each person has five samples for each emotional expression. We put the other 20 data obtained from the same four persons into the classifier for recognizing the emotional category. The test results are shown in Table 5-7. The recognition rate is 86%.

Now the system is tested on a new facial data (See Figure 5-2) gathered from a new person by the same recognition system. The experimental results are shown in Table 5-8. It can be seen that the correct rate is 52% (13/25) and this is much lower than the previous result (86%). It means that the previous-trained facial expression recognition system cannot recognize the test data of the new person correctly. In order

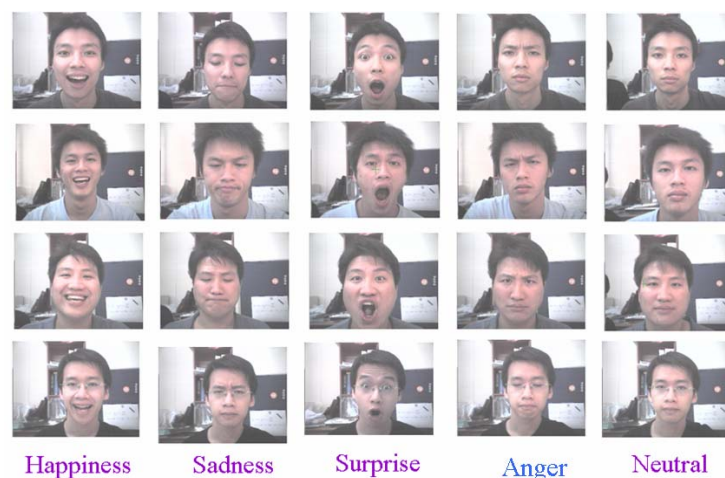


Figure 5-1 Five categories of facial expressions

to improve this result, we used the proposed learning algorithm to retrain a new SVM. The experimental results are shown in Table 5-9. The recognition rate of these new facial data is increased from 52% to 96%. Moreover, it also gives 80% of the recognition rate for the other 20 test data of the same person. Table 5-9 shows that the recognition system after 1<sup>st</sup> new face learning could also keep the recognition rate 80% for classifying the old facial data of the original four persons.

We kept on adding the other five new faces incrementally to the facial database and tested the recognition rate. These five samples are depicted in Figure 5-3. We applied a new SVM classifier after each learning to test these new facial data in turn. The results of testing these new faces are shown in Table 5-10. The second row of Table 5-10 is the recognition rate of each new person. It is observed that the correct

Table 5-7 The recognition result of the original four trained person

	Surprise	Sadness	Neutral	Anger	Happiness
Number of original test data	20	20	20	20	20
Number of correct data	20	18	17	15	16



Figure 5-2 Sample images of a new person in the experiment

Table 5-8 The recognition rate of the 1<sup>st</sup> new facial data

	Surprise	Sadness	Neutral	Anger	Happiness	Average recognition rate
Number of first test data	5	5	5	5	5	
Number of correct data	0	5	3	5	0	52%

Table 5-9 The recognition rate of the 1<sup>st</sup> new facial data after learning

	Surprise	Sadness	Neutral	Anger	Happiness	Average recognition rate
Number of the new facial test data	5	5	5	5	5	
Correct number of recognition after learning the new face	5	5	4	5	4	96%
Correct number of the previous test data after learning the new face	5	5	1	4	5	80%
Rate of correct recognition of old data after learning	19/20	18/20	15/20	13/20	15/20	80%

rate of each new face increase dramatically using the proposed learning algorithm, see row 3 of Table 5-10. Row 4 of Table 5-10 summaries the average recognition rate of other test data to confirm that the new SVM classifier has the general performance on new faces. The experimental results reveal that the proposed emotion recognition algorithm effectively improves the recognition rate of new faces.

The solid line of Figure 5-4 illustrates that the new classifier obtained by the proposed learning algorithm keeps acceptable performance for classifying the previous facial data of both the original and last newly learned faces. We also compare the performance of recognizing facial expression using the proposed algorithm and that obtained from retraining with all the support vectors, as shown by the dotted line of Figure 5-4. It can be seen that, for classifying old data, the performance of the SVM retrained by all SVs is a little bit better than that of the proposed algorithm, but the retraining time for the proposed algorithm is greatly reduced. For instance, the retraining number of facial data for proposed algorithm is fixed at five at each learning step, on the contrary, the retraining number of the method with retraining all SVs is increased from 21.4 to 30 on the average



Figure 5-3 Sample images of other five new persons in the experiment

Table 5-10 The recognition rate of all new face learning

	1 <sup>st</sup> new face	2 <sup>nd</sup> new face	3 <sup>rd</sup> new face	4 <sup>th</sup> new face	5 <sup>th</sup> new face	6 <sup>th</sup> new face
Recognition rate of new face (using previous SVM)	52%	76%	56%	40%	68%	32%
Recognition rate of new face after learning(new SVM)	96%	92%	84%	92%	96%	96%
Recognition rate of previous test data after learning(new SVM)	80%	64%	100%	84%	88%	80%

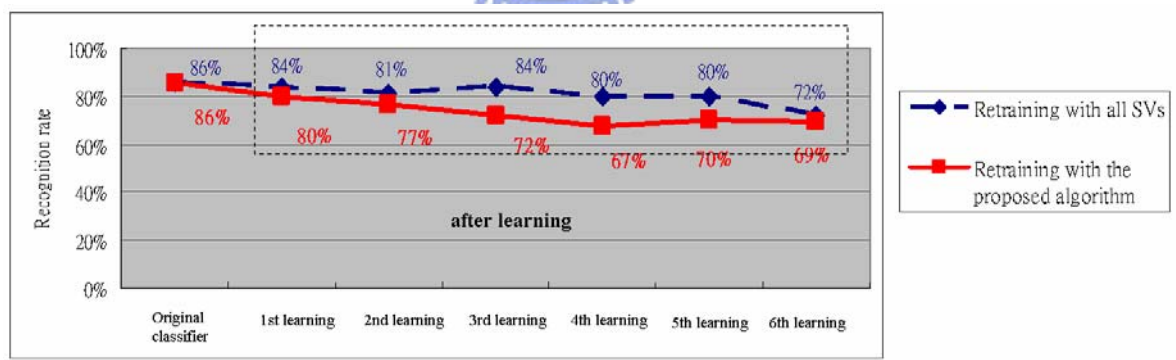


Figure 5-4 Comparison of recognition rate of proposed critical sets training algorithm and the conventional method using all SVs in training

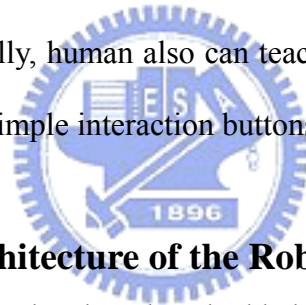
incrementally. The calculation to solve a QP problem in retraining a new SVM is proportional to the square of the training number [11-12]. Therefore, one can expect a significant time saving in comparing the proposed algorithm with the method of

retraining all SVs.

Experimental results show that the proposed algorithm recognizes new facial data with high correct rate after retraining the hyperplane. It also keeps the acceptable recognition rate of original facial data in the meantime and speeds up the QP procedure of SVM. This would be sufficient for real-time requirement in practical robotic applications.

### **5-3 On-Line Experiment Using a Robot Platform**

The proposed FER system has been successfully implemented on an embedded image processing platform to classify five facial expressions on-line. Moreover, the embedded emotion recognition system has also been integrated on an entertainment robot, called Momobear. Finally, human also can teach Monobear to learn new facial expressions of them through simple interaction buttons.



#### **5.3.1 The Hardware Architecture of the Robot Platform**

Figure 5-5 illustrates the developed embedded image processing system [29]. The image processing platform consists of a CMOS sensor board [38] and a DSK6416 board from Texas Instrument [39] as the main processing unit. The selection of the CMOS sensor module as the image sensor is due to its merits of low power consumption, easy to integrated with the DSP system, and less expensive compared with CCD sensors. Meanwhile, the selection of DSK6416 board as the image processing board is due to its high performance fixed-point calculation with 1 GHz clock rate and 1.67ns instruction cycle time.

The left part of Figure 5-6 illustrates the Momobear pet robot. The real-time vision system is installed at back of the Momobear and the CMOS image sensor is put on top in the black hat. Six RC servos are used to control the movement of ears, hands and

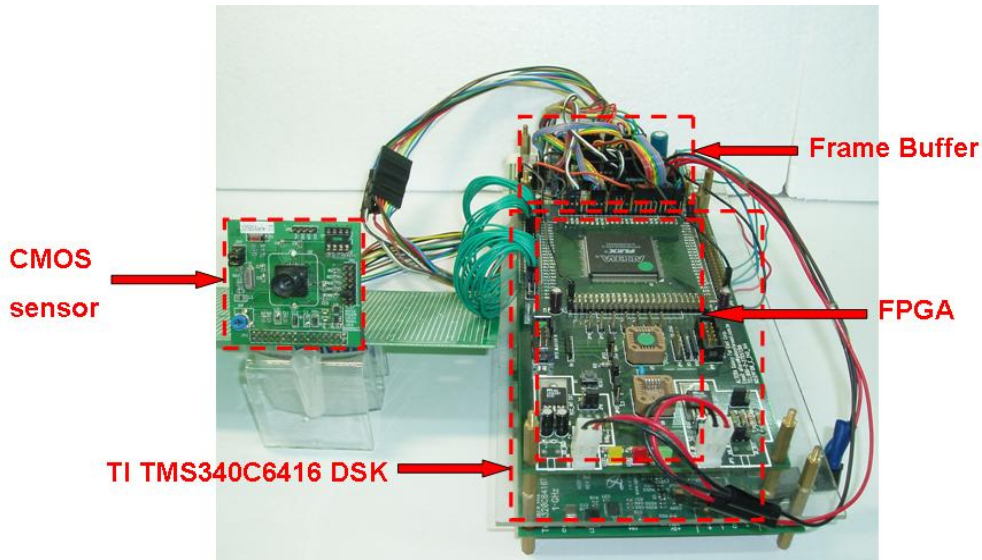


Figure 5-5 The real-time vision system

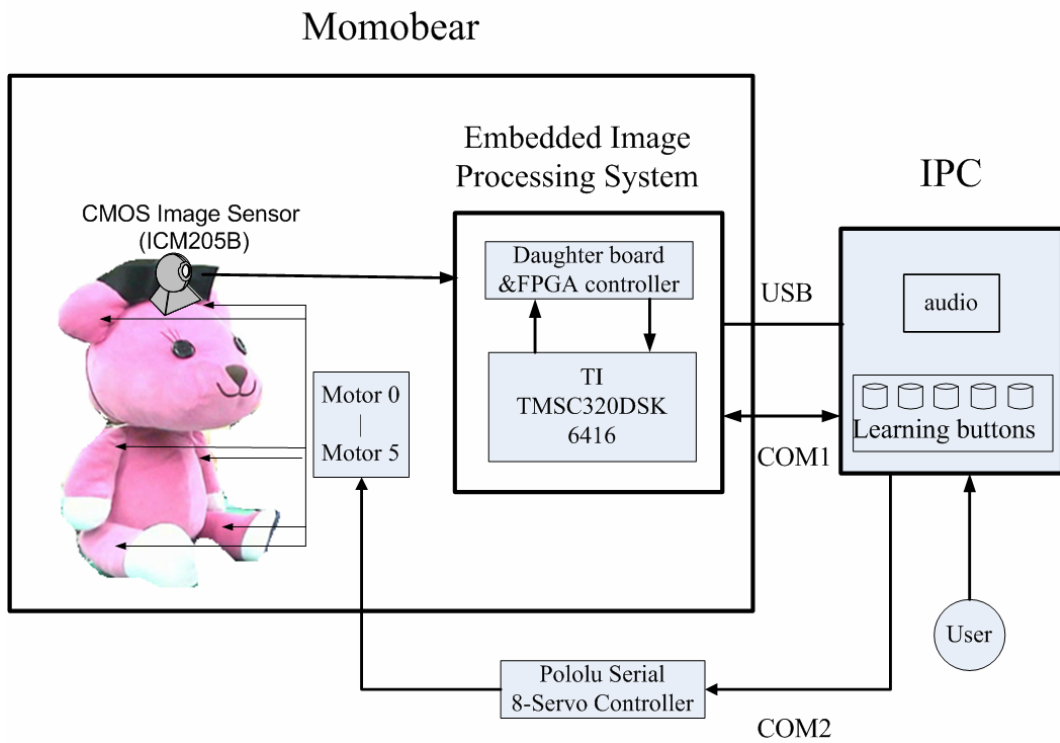


Figure 5-6 Hardware architecture of the imaging system on Momobear

legs of the pet robot. A Pololu serial 8-servo controller [40] is applied to control six RC servos on the robot. The hardware architecture of Momobear is also depicted in Figure 5-6. The embedded image processing system calculates the image data from CMOS image sensor and sends the emotion decisions to an embedded IPC. Based on



different emotions, IPC controls suitable responses of Momobear. If the master perceived that Momobear responses the wrong actions, he/ she would inform the robot to learn his/her unique facial expressions by simple learning buttons.

### **5.3.2 HRI Procedure of the Pet Robot Platform**

The software architecture of proposed FER for a pet robot is as shown in Figure 5-7. First, we choose the mode for the pet robot to recognize or to learn a facial expression through an interactive interface. Based on the mode which the users want to evaluate, IPC sends a flag to embedded image processing system through RS-232 link. If the flag is set to 0, face detection, feature extraction and SVM classification are carried out in turn, and then the final emotional category is transmitted to IPC for controlling the actions of the pet robot. Otherwise, if the flag lies between 1 and 5, it indicates that erroneous recognition occurs and the classifier is needed to retrain for the purpose of accommodating themselves to facial expressions of new faces. Erroneous SVM hyperplanes are adjusted adequately through the flag that stands for the need of learning. The learning procedure does not finish until the user can aware that the pet robot reacts properly. The designed actions of Momobear are displayed in Figure 5-8.

### **5.3.3 Results of On-Line Testing**

Online experiments were carried out using the embedded vision system, IPC and an pet robot as shown in Figure 5-9. We collected facial data of five persons to train the emotion classifiers. The online recognition result of five trained persons is shown in Table 5-11. The average recognition rate is 80.6%.

Subsequently, we invited four new persons to interact with Momobear, the pet

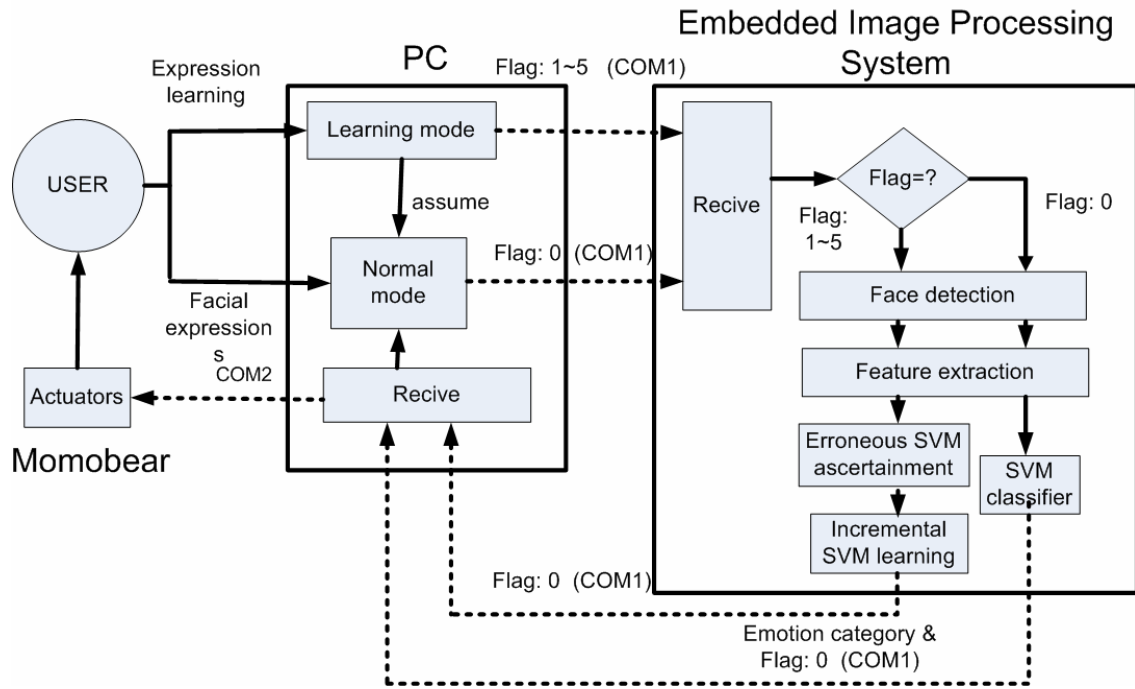


Figure 5-7 HRI procedure of the proposed emotion recognition system






Anger		Action : kick legs Audio : Come on! Relax.
Happiness		Action : swing hands Audio : Ha ha ha ...
Neutral		Action : swing ears Audio : Hi! Master!
Sadness		Action : ears down Audio : Oh oh! Cheer up!
Surprise		Action : swing ears swing hands Audio : Oh ! Really?

Figure 5-8 The designed actions of Momobear

robot. Every person expressed five time of each emotion category under three lighting conditions in front of the CMOS image sensor. Table 5-12 shows that the recognition rate of the first new person is lower in some facial expressions. But that was increased greatly by the proposed learning algorithm. Table 5-13 gives the new recognition result after online learning. The results reveal that the average recognition rate of the first new person can be raised from 57.3% to 82.7%.

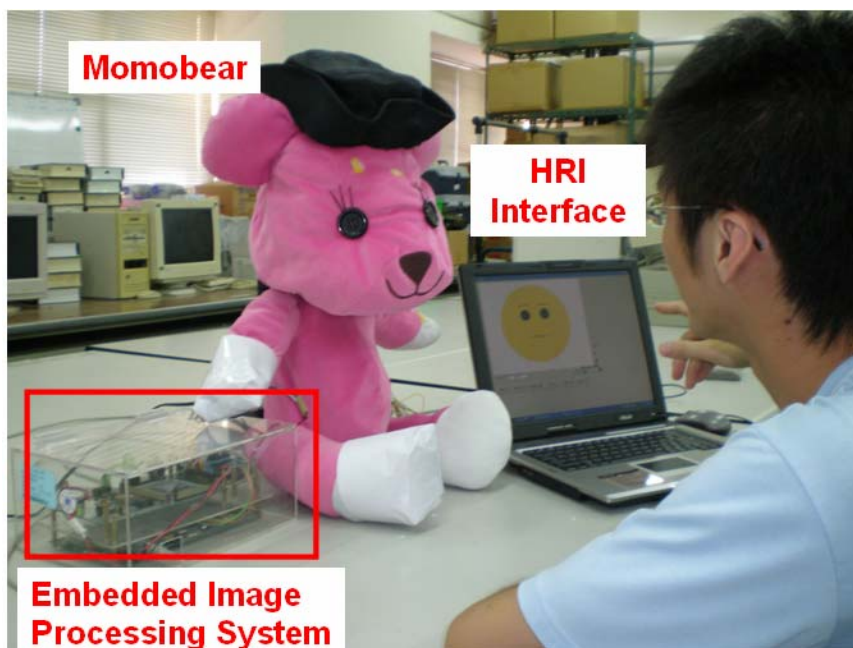


Figure 5-9 The interaction scenario

Table 5-11 The recognition result of five trained persons

	anger	happiness	neutral	sadness	surprise
recognition rate	80%	77%	81%	85%	80%

Table 5-12 Recognition rate of the first new person before online learning

	anger	happiness	neural	sadness	surprise
right light on	40%	0%	80%	80%	20%
all lights on	100%	0%	80%	80%	80%
left light on	40%	0%	80%	100%	80%

Table 5-13 Recognition rate of the first new person after online learning

	anger	happiness	neural	sadness	surprise
right light on	80%	100%	80%	80%	60%
all lights on	100%	80%	80%	80%	80%
left light on	80%	80%	80%	100%	80%

The experimental procedures of the next three persons are the same as the first one. The original recognition results of three persons are shown in Table 5-14, Table 5-16 and Table 5-18. It is observed that original recognition rates of new subjects are very low, but they can be increased dramatically by the proposed learning system. Table 5-15, Table 5-17 and Table 5-19 show the results after online learning. Each recognition result is evaluated on a new SVM classifier that is learned from the original trained one. The average rates of the four new subjects are shown in Tables 5-20~5-21. In summary, the average recognition rate of four new persons after each on-line learning can be increased from 58% to 81.3 %, which is as high a recognition rate as that of training persons (80.6%).

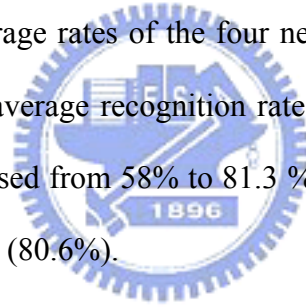


Table 5-14 Recognition rate of the second new person before online learning

	anger	happiness	neural	sadness	surprise
right light on	20%	0%	80%	80%	20%
all lights on	20%	0%	80%	80%	20%
left light on	80%	20%	80%	80%	80%

Table 5-15 Recognition rate of the second new person after online learning

	anger	happiness	neural	sadness	surprise
right light on	80%	80%	80%	80%	80%
all lights on	60%	80%	80%	80%	80%
left light on	80%	80%	80%	80%	80%

Table 5-16 Recognition rate of the third new person before online learning

	anger	happiness	neural	sadness	surprise
right light on	20%	0%	100%	80%	100%
all lights on	80%	40%	80%	80%	0%
left light on	80%	20%	80%	100%	80%

Table 5-17 Recognition rate of the third new person after online learning

	anger	happiness	neural	sadness	surprise
right light on	80%	80%	100%	80%	100%
all lights on	80%	60%	80%	80%	80%
left light on	80%	80%	80%	100%	80%

Table 5-18 Recognition rate of the fourth new person before online learning

	anger	happiness	neural	sadness	surprise
right light on	80%	80%	80%	100%	40%
all lights on	100%	80%	0%	80%	0%
left light on	80%	60%	80%	40%	40%

Table 5-19 Recognition rate of the fourth new person after online learning

	anger	happiness	neural	sadness	surprise
right light on	80%	80%	80%	100%	100%
all lights on	100%	80%	80%	80%	80%
left light on	80%	80%	80%	80%	80%

Table 5-20 Average recognition results of four new persons before online learning

	anger	happiness	neural	sadness	surprise
right light on	40%	20%	85%	85%	45%
all lights on	75%	30%	60%	80%	25%
left light on	70%	25%	80%	90%	60%

Table 5-21 Average recognition results of four new persons after online learning

	anger	happiness	neutral	sadness	surprise
right light on	80%	80%	85%	85%	85%
all lights on	85%	75%	80%	80%	80%
left light on	80%	75%	80%	90%	80%

Table5-22 shows that the recognition rates of previously data also can be maintained (78.67% on the average) even when the SVM classifiers are adjusted for the new individual. With the proposed learning, the pet robot not only accommodates itself to new facial data but also keeps the satisfactory performance of old data.

Table 5-23 presents the experimental results of comparing the proposed method with the previous method developed in [41]. We see that the performances of the previous method are much lower than that of the proposed method. This is because their method cannot accommodate the FER system to facial expressions of new subjects and feature extraction of the method may fail under illumination variation. Therefore, we conclude that the developed FER system with a learning function outperforms those without learning or accommodating functions while categorizing facial expressions of new subjects.

Figure 5-10 shows an example of emotional interaction with Momobear. Figure 5-10 (a) is a successful recognition with Momobear. Figure 5-10 (b) shows that Momobear misclassifies surprise expression. Figure 5-10 (c) illustrates that the user informs Momobear to learn surprise expression through learning buttons. Figure 5-10 (d) is a successful recognition of surprise expression after online learning.

Table 5-22 The recognition results of five trained person after online learning

	anger	happiness	neutral	sadness	surprise	Average
recognition rate after SVM classifier learning	82%	77%	82%	82%	72%	78.67%

Table 5-23 Recognition results of the method proposed in [41]

	anger	happiness	neural	sadness	surprise
right light on	45%	20%	65%	60%	80%
all lights on	70%	50%	90%	25%	75%
left light on	75%	60%	15%	70%	70%

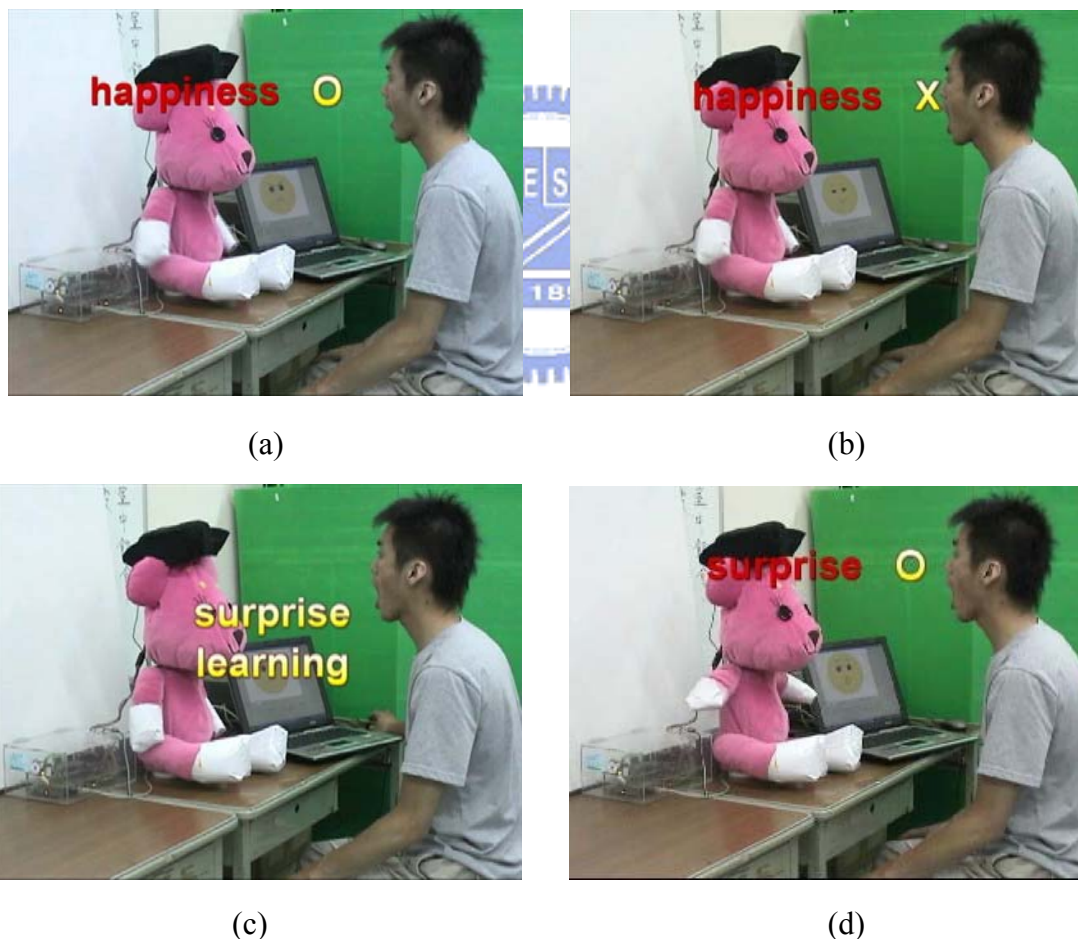


Figure 5-10 An example of interaction with Momobear.

# Chapter 6

## Conclusions and Future Work

### 6.1 Conclusions

This thesis presents a fast facial expression learning algorithm for a pet robot. An emotion recognition system can accommodate itself to new facial data. The proposed learning method adjusts parameters of SVM hyperplane. After adjusting hyperplane parameters, the new classifier not only recognizes new facial data but also keeps acceptable recognition rates of classifying previous old data. Because only new erroneous samples combined with historical critical sets (CSs) are used to restrain a new SVM classifier, the proposed algorithm can speed up the learning procedure. Further, to obtain facial features correctly, Gabor wavelet based feature extraction is employed in the FER system.

The proposed FER algorithm has been evaluated on the AR Face Database and the database built in the lab. These offline experimental results show that recognition rate is 81.7% on The AR Face Database and 81.5% on the lab database. Moreover, the learning algorithm also has been verified using self-built database and a robot platform. The average recognition rate of new persons after online learning can be raised from 58% to 81.3%. In the meantime, new SVM classifier also keeps satisfactory performance of recognizing old data (78.7%).

### 6.2 Future work

Gabor wavelet based feature extraction method has robust properties against the changes of lighting conditions, but the computation cost of extracting facial points is very high. Moreover, extracting feature points around the mouth is not stable enough.



In the future, some popular facial feature models such as active appearance model (AAM) can be combined with Gabor based feature extraction method to improve the performance of extracting facial points. One alternative solution is to use Gabor based feature matching to detect fiducial points in the first frame of image sequences. Then, in the subsequent image frames, AAM or other facial feature models are applied to reduce the computation cost and raise the recognition rates of detecting facial points. On the other hand, we will also work on new methods for improving the error rate of the original trained data with SVM learning in order to apply the proposed learning algorithm to practical robotic applications.



# References

- [1] M. Fujita, "On Activating Human Communications with Pet-type Robot AIBO," *Proc. of the IEEE*, vol. 92, no. 11, pp. 1804-1813, 2004.
- [2] P. Ekman and W. V. Friesen, *Emotion in the Human Face*. Englewood Cliffs, NJ : Prentice-Hall, 1975.
- [3] P. Ekman and W. V. Friesen : *The Facial Action Coding System* (Consulting Psychologists Press, San Francisco 1978).
- [4] <http://face-and-emotion.com/dataface/facs/manual/TitlePage.html>
- [5] M. Pantic and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions : The State of the Art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12., pp. 1424 - 1445, 2000.
- [6] Gengtao Zhou, Yongzhao Zhan and Jianming Zhang, "Facial Expression Recognition Based on Selective Feature Extraction," in *Proc. of IEEE 6<sup>th</sup> Int. Conf. on Intelligent Systems Design and Applications*, Shandong, China, 2006, pp. 412 – 417.
- [7] F. Bourel, C.C. Chibelushi, and A.A. Low, "Robust Facial Expression Recognition Using a State-Based Model of Spatially-Localised Facial Dynamics," in *Proc. of IEEE 5<sup>th</sup> Int. Conf. on Automatic Face and Gesture Recognition*, Washington, D.C, USA, 2002, pp. 106 – 111.
- [8] WeiFeng Liu and ZengFu Wang, "Facial Expression Recognition Based on Fusion of Multiple Gabor Features," in *Proc. of IEEE 18<sup>th</sup> Int. Conf. on Pattern Recognition*, Hong Kong, China, 2006, pp. 536 – 539.
- [9] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," in *Proc, IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshop*, New York, USA, 2006, pp. 149-156.
- [10] G. Guo and C.R. Dyer, "Learning from Examples in the Small Sample Case: Face Expression Recognition," *IEEE Trans. on Systems, Man and Cybernetics, Part B*, vol. 35, no. 3, pp. 477 – 488, 2005.
- [11] V. Vanik, *The Nature of Statistic Learning Theory*, New York : Springer-Verlag, 1995.

- [12] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, NY Cambridge University Press 2000.
- [13] M. Yeasin, B. Bulot, and R. Harma, "Recognition of Facial Expressions and Measurement of Levels of Interest From Video," *IEEE Trans. on Multimedia*, vol.8, no.3, pp.500-508, 2006.
- [14] Zhengyou Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perception," in *Proc. of 3rd IEEE Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp. 454 – 459.
- [15] Y. Z. Zhan, J. F. Ye, D. J. Niu, and P. Cao, "Facial Expression Recognition Based on Gabor Wavelet Transformation and Elastic Templates Matching," in *Proc. of 3rd Int. Conf. on Image and Graphics*, Hong Kong, China, 2004, pp. 254 – 257.
- [16] Y. Saatci and C. Town, "Cascaded Classification of Gender and Facial Expression Using Active Appearance Models," in *Proc. of 7th Int. Conf. on Automatic face and Gesture Recognition*, Southampton, UK, 2006, pp. 393 – 398.
- [17] G. J. Edwards, T. F. Cootes and C. J. Taylor, "Face Recognition Using Active Appearance Models," in *Proc. of European Conf. Computer Vision*, Freiburg, Germany, 1998, pp.581-695.
- [18] I. Kotsia and I. Pitas, "Facial Expression Recognition in Image Sequences Using Geometric Deformation Features and Support Vector Machines," *IEEE Trans. on Image Processing*, vol.16, no.1, pp. 172-187, 2007
- [19] T. Kobayashi, Y. Ogawa, K. Kato and K. Yamamoto, "Learning Systems of Human Facial Expression for a Family Robot," in *Proc. of the 6<sup>th</sup> IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Seoul, Korea, 2004, pp. 481-486.
- [20] S. Ozawa, S.L. Toh, S. Abe, Pang Shaoning and N. Kasabov, "Incremental Learning for Online Face Recognition," In *Proc. of Int. Joint Conf. on Neural Networks*, Montreal, Canada, 2005, pp. 3174 – 3179.
- [21] W. M. Huang, B. H. Lee, L. Y. Li and K. Leman, "Face Recognition by Incremental Learning," In *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, Washington, D.C., USA, 2003, pp.4728-4723.
- [22] R. Xiao, J. Wang and F. Zhang, "An Approach to Incremental SVM Learning Algorithm," in *Proc. of the 12th International Conference on Tools with Artificial Intelligence*, Vancouver, Canada, 2000, pp.268– 273.

- [23] J. L. An, Z. O. Wang and Z. P. Ma, “An Incremental Learning Algorithm for Support Vector Machine,” in *Proc. of Int. Conf. on Machine Learning and Cybernetics*, Xian,, China, 2003, pp. 1153 – 1156.
- [24] L. Xuchun, Z. Yan and E. Sung, “Sequential Bootstrapped Support Vector Machines - a SVM accelerator,” in *Proc. of IEEE Int. Joint Conf. Neural Networks*, Montréal, Canada, 2005, pp. 1437 – 1442.
- [25] W. H. Zeng and J. Ma, “A Novel Incremental SVM Learning Algorithm,” in *Proc. of Int. Conf. on CSCW in Design*, Xiamen, China, 2004, pp. 658 – 662.
- [26] P. Mitra, C. A. Murthy and S. K. Pal, “Data Condensation in Large Databases by Incremental Learning With Support Vector Machines,” In *Proc. of Intl. Conf. on Pattern Recognition*, Catalonia, Spain, 2000, pp. 708 – 711.
- [27] Y. Liu, Q. He and Q. Chen, “Incremental Batch Learning with Support Vector Machines,” in *Proc. of 5th world congress on Intelligent Control and Automation*, Hangzhou, China, 2004, pp. 1857– 1861.
- [28] Y. Liu, Q. He and Y. Tang, “Support Vector Pursuit Learning,” in *Proc. of IEEE Int. Conf. on System, Man and Cybernetics* , The Hague, The Netherlands, 2004, pp.5841– 5846.
- [29] C.M. Chou: *Real-Time Face Tracking Under Illumination Variation*, Master. Eng. Thesis, National Chiao Tung University, Hsinchu, Taiwan, R.O.C., 2005.
- [30] A. McAndrew, *Introduction to Digital Image Processing with Matlab*, Thomson Course Technology, 2004
- [31] J.H. Lai, P.C Yuen, W.S. Chen, S. Lao and M. Kawade, “Robust Facial Feature Point Detection Under Nonlinear Illuminations,” in *Proc. of IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*, Canada, 2001, pp.168-174.
- [32] J. K. Kamarainen and H. Kalviainen, “Invariance Properties of Gabor Filter-Based Features---Overview and Application,” *IEEE Trans. on Image Processing*, vol. 15, no 5, pp.1088-1099, 2006.
- [33] L. Wiskott, J. M. Fellous, N. Kruger and von der Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” *IEEE Trans. on Pattern analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.
- [34] E. J. Hokdeb and R. Owens, “Automatic Facial Point Detection,” in *Proc. of 5th Asian Conference on Computer Vision*, Melbourne, Australia, 2002, pp. 731-736.
- [35] D. Vukadinovic and M. Pantic, “Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers,” in *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, Hawaii, U.S.A., 2005, pp. 1692 – 1698.

- [36] Y. Zhang and Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligent*, vol. 27, no 5, pp. 699-714, 2005
- [37] [http://cobweb.ecn.purdue.edu/~aleix/aleix\\_face\\_DB.html](http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html)
- [38] IC-MEDIA Corp.: ICM205B Datasheet, Oct 2002.
- [39] Henry Andrian and K. T. Song, "Embedded CMOS Imaging System for Real-Time Robotic Vision," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Alberta, Canada, 2005, pp. 3694-3699.
- [40] <http://www.pololu.com/products/pololu/0727/>
- [41] J.H. Hsu: *Bimodal Emotion Recognition System Using Image and Speech Information*, Master. Eng. Thesis, National Chiao Tung University, Hsinchu, Taiwan, R.O.C., 2006.

