Facial Emotion Recognition with a Neural Network Approach

By,

Wathsala Nayomi Widanagamaachchi

Registration Number : 2005CS155

A thesis presented to the University of Colombo School of Computing, in partial fulfillment of the requirements for the degree of

Bachelor of Science (Computer Science)

University of Colombo School of Computing
35, Reid Avenue, Colombo 07
Sri Lanka

August 2009.

Declaration

I, Wathsala Nayomi Widanagamaachchi (2005CS155), hereby certify that this thesis entitled 'Facial Emotion Recognition with a Neural Network Approach' is entirely my own work and it has never been submitted nor is currently been submitted for any other degree.

Candidate

. . . .
Wathsala Nayomi Widanagamaachchi

. . . .
Date

I, Anuja Dharmarathne, certify that I supervised this thesis entitled 'Facial Emotion Recognition with a Neural Network Approach' conducted by Wathsala Nayomi Widanagamaachchi. I recommend this thesis to the University of Colombo School of Computing in partial fulfillment of the requirement of the degree Bachelor of Science (Computer Science).

Supervisor

. . . .
Anuja Dharmarathne

. . . .
Date

Abstract

Behaviors, actions, poses, facial expressions and speech; these are considered as channels that convey human emotions. Extensive research has being carried out to explore the relationships between these channels and emotions. This paper proposes a system which automatically recognizes the emotion represented on a face. Thus a neural network based solution combined with image processing is used in classifying the universal emotions: Happiness, Sadness, Anger, Disgust, Surprise and Fear. Colored frontal face images are given as input to the system. After the face is detected, image processing based feature point extraction method is used to extract a set of selected feature points. Finally, a set of values obtained after processing those extracted feature points are given as input to the neural network to recognize the emotion contained.

To ... my beloved Amma, Thaththa, Akka, Malli
& Lalindra

Acknowledgements

I take this moment to express my heartfelt gratitude to my supervisor, Dr. Mrs. Anuja Dharmaratne for her guidance and assistance throughout the project. Then to Mr. Dulan Wathugala and Ms. Lasanthi De Silva for the effort put to improve the quality of our research. My parents, sister, brother and most of all Lalindra should be remembered for their continuous support and understanding. Finally, all of the lecturers at UCSC, friends and other people who assisted in different ways should be remembered.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

What is an emotion? An emotion is a mental and physiological state which is subjective and private; it involves a lot of behaviors, actions, thoughts and feelings.

Initial research carried out in emotions can be traced to the book 'The Expression of the Emotions in Man and Animals' by Charles Darwin. He believed emotions to be species-specific rather than culture-specific (19), but in 1969 after recognizing a universality within emotions despite the cultural differences, Ekman and Friesen classified six emotional expressions to be universal: happiness, sadness, anger, disgust, surprise and fear (6; 11; 15; 19). (Figure 1.1)

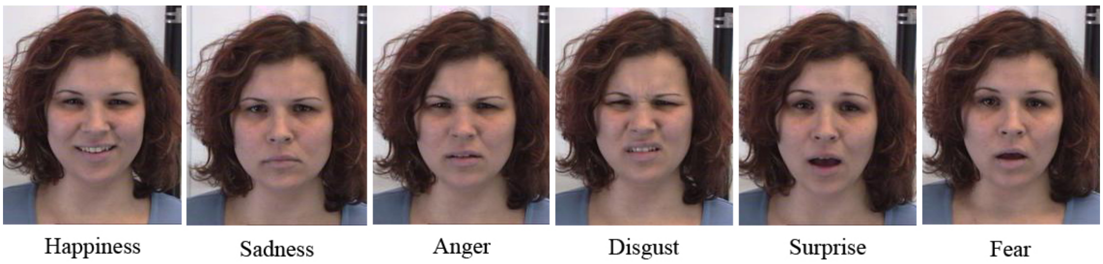

Figure 1.1: The six universal emotional expressions (43)

1.1 Problem Statement

Many factors contribute in conveying emotions of an individual. Pose, speech, facial expressions, behavior and actions are some of them. From these above mentioned factors facial expressions have a higher importance since they are easily perceptible.

In communicating with others humans can recognize emotions of another human with a considerable level of accuracy. If we can efficiently and effectively utilize heretofore found knowledge in computer science to find practical solutions for automatic recognition of facial emotions, we would be able to attain accuracy that is virtually comparable to the human perception.

Though numerous approaches have been taken on this topic, some limitations still exist. To obtain better results, some researchers have constrained their approaches by either using less feature points or by using a neutral image of the person in the emotion classification phase. If we can address this problem from an identified set of critical feature points and use straightforward feature detection techniques the overall performance can be augmented.

1.2 Motivation

Significant debate has risen in the past regarding the emotions portrayed in the world famous masterpiece of Mona Lisa (Figure 1.2). British weekly 'New Scientist' (3) has stated that she is in fact a blend of many different emotions, 83% happy, 9% disgusted, 6% fearful and 2% angry.

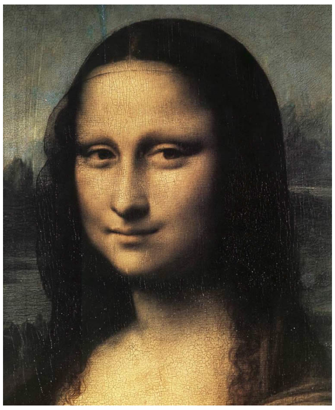

Figure 1.2: Mona lisa

In addition to Mona Lisa, Spielberg's film A.I. (Artificial Intelligence) can be considered as a significant attempt to portray the future prospects in this region. Modern drifts in emotion transferring to 3D game avatars, truthful evaluations and nonverbal communication interpretations have made this area more attractive.

1.3 Prospective Applications

The emotional frontier is in fact the next obstacle to be surmounted in understanding humans. Facial expressions can be considered not only as the most natural form of displaying human emotions but also as a key non-verbal communication technique (30). If efficient methods can be brought about to automatically recognize these facial expressions, striking improvements can be achieved in the area of human computer interaction. Research in facial emotion recognition has being carried out in hope of attaining these enhancements (9; 40). Moreover, there are other applications which can benefit from automatic facial emotion recognition. Artificial Intelligence has long relied on the area of facial emotion recognition to gain intelligence on how to model human emotions convincingly in robots. Recent improvements in this area have encouraged the researchers to extend the applicability of facial emotion recognition to areas like chat room avatars, video conferencing avatars. The ability to recognize emotions can be valuable in face recognition applications as well. Suspect detection systems and intelligence improvement systems meant for children with brain development disorders are some other beneficiaries (24).

1.4 Proposed Solution

In this work, a system which will efficiently recognize the six universal emotions from 2D color face images. The work has being limited to the universal emotions since classification and identification of other marginal emotions is problematic. The system can be broadly categorized in to three stages: face location determination stage, feature extraction stage and emotion classification stage. Two face detection algorithms are implemented for the face location determination stage. Eyes, mouth and eyebrows are identified as the critical features and their feature points are extracted to recognize the emotion. These feature points are extracted from the selected feature regions with the use of a corner point detection algorithm. After feature extraction is performed a neural network approach is used to recognize the emotions contained within the face.

1.5 Organization of the Thesis

The rest of the paper is organized as follows. Chapter 2 contains a brief overview of the recent work carried out in the area of facial emotion recognition. The system model of the proposed solution and detailed descriptions on each of it's stages (Face location determination, Feature extraction and Emotion classification) can be found in chapter 3. An evaluation of the overall system will follow in chapter 4. Finally the paper is concluded in chapter 6.

Chapter 2

Recent Work

The recent work relevant to the study can be broadly categorized into three steps: Face detection, Facial feature extraction and Emotion classification. The amount of research carried out in each of these categories is quite sizeable and noteworthy. These three categories are concerned with the central background pertaining to the issue of facial emotion recognition. Apart from them, another core area is the work carried out in forming an apposite facial database for such studies.

2.1 Face Detection

Given an image, detecting the presence of a human face is a complex task due to the possible variations of the face. The different sizes, angles and poses a human face might have within the image can cause this variation. The emotions which are deducible from the human face and different imaging conditions such as illumination and occlusions also affect facial appearances. In addition, the presence of spectacles, beard, hair and makeup have a considerable effect in the facial appearance (31; 32).

The approaches of the past few decades in face detection can be broadly classified in to four sections: knowledge-based approach, feature invariant approach, template-based approach and appearance-based approach (31; 32; 45; 46). (Table 2.1)

Table 2.1: Categorization of face detection approaches (46)

Knowledge-based Approach
Feature Invariant Approach
-Facial Features
-Texture
-Color
-Multiple Features
Template-based Approach
-Predefined Face Templates
-Deformable Face Templates
Appearance-based Approach
-Neural Network
-Support Vector Machine (SVM)
-EigenFace
-Distribution based
-Nave Bayes Classifier
-Hidden Markov Model (HMM)
-Information-Theoretical Approach

2.1.1 Knowledge-based Approach

Knowledge-based approach is based on rules derived from the knowledge on the face geometry. A typical face used in this approach is shown in figure 2.1. The most common way of defining the rules is by the relative distances and positions of facial features. By applying these rules faces are detected, then a verification process is used to trim the incorrect detections. Translating knowledge about the face geometry into effective rules is one difficulty faced in this approach, since strict rules may fail to detect faces but if the rules are too general it can increase incorrect detections. This approach is too limited since extending this to detect faces in all cases is impossible (31; 45; 46).

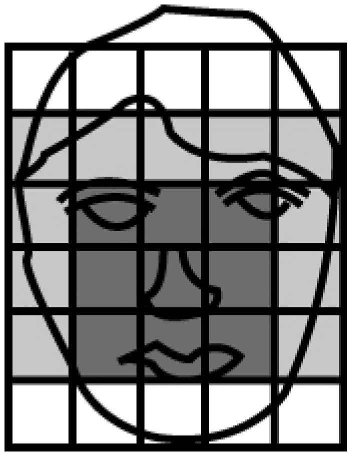

Figure 2.1: A typical face used in knowledge-based approach (46)

2.1.2 Feature Invariant Approach

In feature invariant approach, facial features are detected and then grouped according to the geometry of the face. Selecting a set of appropriate features is very crucial (20). This approach is not suitable for images with noise, illuminations and occlusions since they can weaken the feature boundaries. The main drawback of template-based approaches is the sensitivity to rotation and scaling. Since feature-based approach is not affected by this sensitivity, it provides a better solution to facial detection problem. A face model which is defined in terms of features or a face texture model which is defined in terms of a set of inequalities can be used for face detection in the feature invariant approach (5; 31; 45; 46). Recently human skin color has caught the attention of the researchers as a significant feature, since skin color reside in a small color range in different color spaces regardless of the race. Thus contemporary studies on various skin color detection techniques can be found (18; 25; 41; 42). Approaches which combine multiple facial features have also being proposed.

2.1.3 Template-based Approach

A standard pattern of a human face is used as the base in the template-based approach. The pixels within an image window are compared with the standard pattern to detect the presence of a human face within that window. Though the approach is simple, the scaling required for the image and template is a drawback. Besides, this approach is incapable of dealing with the variations of the human face. A predefined template based and deformable template based approaches are the two classes of the template-based approach (5; 31; 45; 46).

2.1.4 Appearance-based Approach

Appearance-based approach considers the human face in terms of a pattern of pixel intensities. Since non face patterns are not used in the training process of this approach it is not robust enough. Even the time taken is lengthy, as the number of patterns which needs to be tested is large. A neural network is the commonly used solution for capturing complex facial patterns from facial images. Both supervised and unsupervised learning approaches have being used to train the network. Since finding a sufficient training data set is questionable, unsupervised neural networks are more preferable. Apart from neural networks, Support Vector Machines (SVM), eigenfaces, Distribution based approaches, Nave Bayes classifiers, Hidden Markov Models (HMM) and Information theoretical approaches can also be used for face detection in the appearance-based approach (31; 32; 45; 46). Rather than minimizing the training error as in neural networks, SVM operate by minimizing the upper bound on the generalization error. Eigenfaces is a probabilistic visual learning method which uses eigenspace decomposition. Nave Bayes classifier provides a better assessment of the conditional density functions in facial sub-regions. The HMM does not require exact alignment as in template-based and appearance-based approaches. HMM usually see a face pattern as a sequence of observation vectors.

2.2 Facial Feature Extraction

Selecting a sufficient set of feature points which represent the important characteristics of the human face and which can be extracted easily is the main challenge a successful facial feature extraction approach has to answer (Figure 2.2). The luminance, chrominance, facial geometry and symmetry based approaches, template based approaches and Principal Component Analysis (PCA) based approaches are the main categories of the approaches available. Approaches which combine two or more of the above mentioned categories can also be found (20; 34).

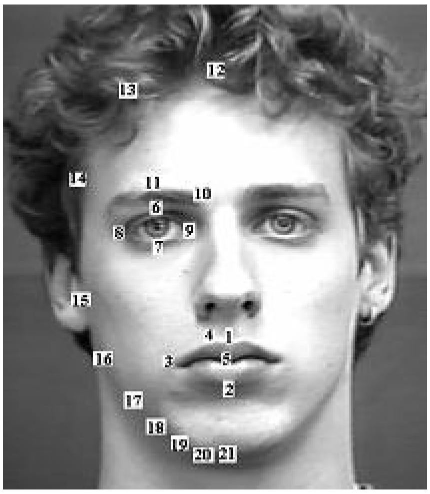

Figure 2.2: A selected set of 21 facial feature points (39)

2.2.1 Luminance, Chrominance, Geometry and Symmetry based Approach

When using facial geometry and symmetry for the extraction of features, the human visual characteristics of the face are employed. The human face is formed by a set of features. The symmetry contained within the face is helpful in detecting the facial features irrespective of the differences in shape, size and structure of features. Most common method of extracting features in geometry and symmetry approaches is the use of the feature shape. Moreover the knowledge on human face can be also be used in building relationships among facial features which will in turn assist the extraction process (Figure 2.3).

In extracting features based on luminance and chrominance the most common method is to locate the eyes based on valley points of luminance in eye areas. Observations have confirmed that under normal illumination conditions, facial features (eyes, nose, mouth etc.) contain relatively low gray levels and that the intensity histogram of the face produces a shape of two peaks: the peak in the lower grey levels of the histogram is caused by the darker parts of the face such as eyes; the other peak is caused by the lighter parts such as the cheeks and forehead. Thus the use of the intensity histogram has being extensively used in extracting features based on luminance and chrominance. The most important feature to extract is the eyes. Since the mid-

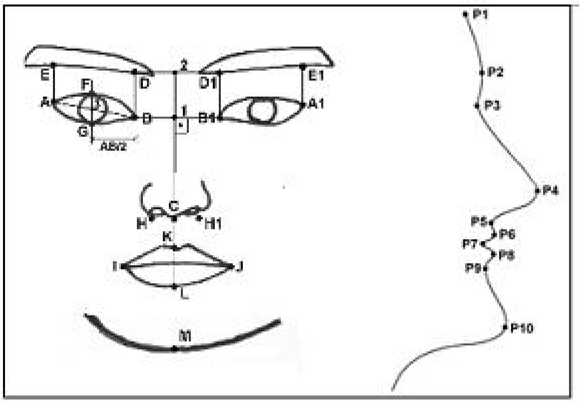

Figure 2.3: Relationship between of facial features (29)

point of the nostrils contain a higher stability than other feature points, despite of its relatively less importance compared to other facial features, its extraction is performed as well (4; 13; 20).

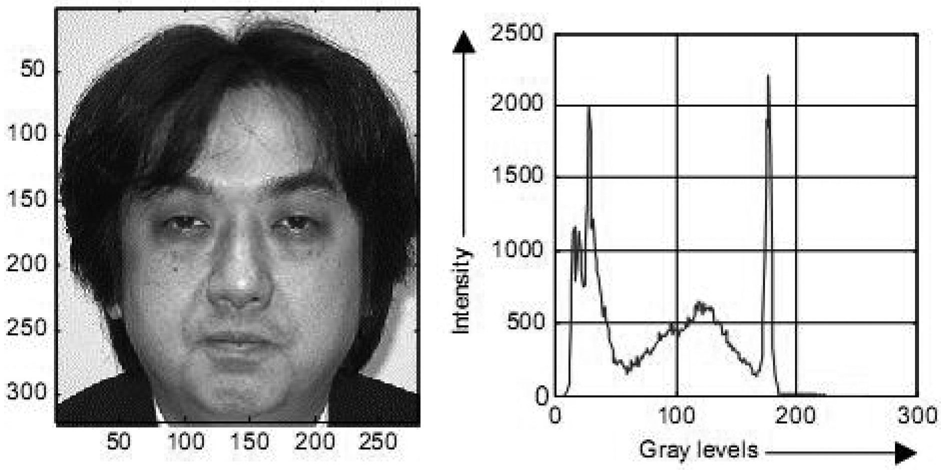

Figure 2.4: Intensity histogram of a face (4)

Hua Gu (20) has proposed a method to extract 9 feature points of the human face (2 eyeballs, 4 far and near points of the eye, 2 mouth points and the midpoint of nostrils) based on face geometry and symmetry. They used the Smallest Univalue Segment Assimilating Nucleus (SUSAN) operator to extract the edges and corner points of features. The valley point searching with directional projection of luminance and the symmetry of the eyeballs is used in extracting eyes from the facial image. This directional projection of luminance method can be used in extracting the nose and the mouth. (Figure 2.3)

The above mentioned directional projection method combined with skin color detection and window growing is used in the Deng and others' approach (13). Their method involved the use of color, shape and intensity. The Cao and others' approach (8) focuses on the extraction of the iris center with the use of region growing. The high speed and the notable performance achieved under pose and lighting variations are some characteristics of this approach.

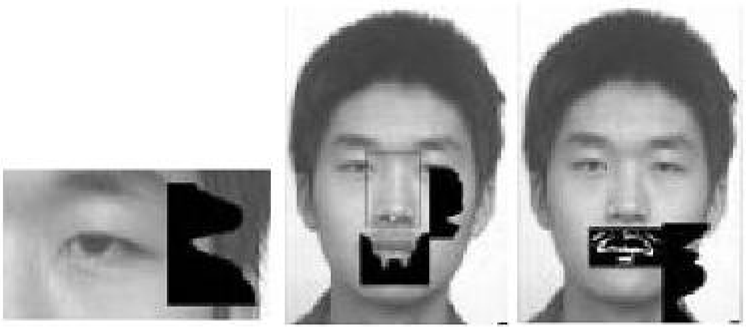

Figure 2.5: Localizing features with directional projections of luminance(20)

2.2.2 Template based Approach

Some researchers have based their approaches on the unique shape of features in hope of obtaining improved results. By defining each feature with a separate template, Craw and others' (10) have put forth an efficient approach for extraction of facial features. They have used deformable templates to extract the eyes, nose and face outline. A typical eye template used in most template-based approaches is shown in figure 2.6.

Kuo and Hannah (23) have also proposed a highly accurate and flexible approach for the extraction of eyes using basic shapes, eye corner detection and deformable templates. The iris is chosen as the first component to extract, because of its circular shape and high intensity contrast with the neighboring areas. In locating the eye center using a PCA based approach, a deformable circle template is used to search across to get the best fit. For a best fit, the area of the circle should be dark and a notable intensity contrast should be present along the circumference. After iris extraction, a simple and effective eye corner location technique is used for the location of eye corners. This approach has shown 94% accuracy in iris extraction and 88% accuracy in eye corner and eyelid fitting. The errors in extraction are mainly due to the strong reflections in the eye areas.

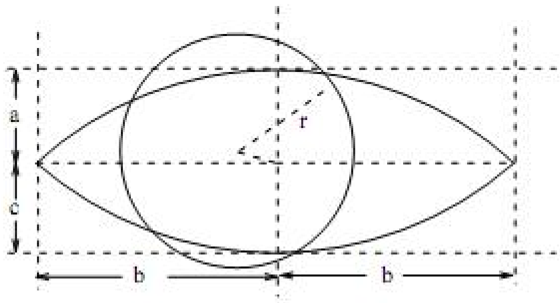

Figure 2.6: A template used for the extraction of single eye (10)

2.2.3 PCA based Approach

For extracting features based on PCA, unsupervised PCA and global PCA based approaches have been used. However PCA based approaches which consider the whole face image have the disadvantage of being sensitive to the face size, location, background. Furthermore larger illumination variations tend to skew the PCA learning process. Thus, interest has risen over more precise and localized PCA based approaches.

Such a localized PCA based learning approach for facial feature extraction has being put forth by King and Xu (22). The directed and guided training in the localized PCA learning combined with simple matching strategies has formed an efficient approach to extract features for matching face images. This approach results several masks for localized features. Several of such masks obtained for mouth are shown in figure 2.7. A batch algorithm has being used to implement the localized PCA learning. Result of this approach is a set of eigenfeatures useful for face recognition.

An approach for feature extraction and estimation of control points for facial image warping based on a PCA based statistical face model, has being presented by Xue and others (44). The PCA has being used to create full face model with contour points and control points.

Figure 2.7: Mouth masks obtained from PCA learning of mouth images (22)

The Luminance, chrominance, facial geometry and symmetry based approaches, template based approaches, Principal Component Analysis (PCA) based approaches apart from these, other facial feature extraction approaches also exist. Some of them involve eigenfaces, wavelets and discrete cosine transform (20; 34).

A wavelet network is a network which consists of a set of wavelets and associated networks. The geometric configurations of a wavelet network are defined with respect to a single coordinate system. Feris and others' approach (16) on two-level hierarchical wavelet networks localizes small features with the use of cascading sets of GWN features. Moreover they have tested their results with both one-level and two level hierarchical wavelet networks. In first level hierarchies the GWN features are trained for the whole face while GWN features are trained for each facial feature in the two level hierarchical networks.

A neural network can also be trained to locate the desired features from a facial image. To achieve this, either one neural network can be trained for all the features or several networks for each feature. The main challenge faced in using supervised neural networks is the acquisition of the training set. In the attempt of Paul Debevec, (12) a neural network is trained for the extraction of eyes, nose and mouth. The back propagation algorithm and a training set with 97 images have being used to train the network. The neural network suggested in the approach consisted of a hidden layer with ten neurons. Four such neural networks have being trained for each feature (left and right eyes, nose, mouth). For the visualization of the user, the neural network's suggestion for the feature has being mapped in to an image (also called a feature maps). Thus the accuracy of the neural network can be calculated by visual comparison of original image and the above mentioned image. A successful neural network would generate a complete white image except for the black area at the corresponding location of the feature. (Figure 2.8)

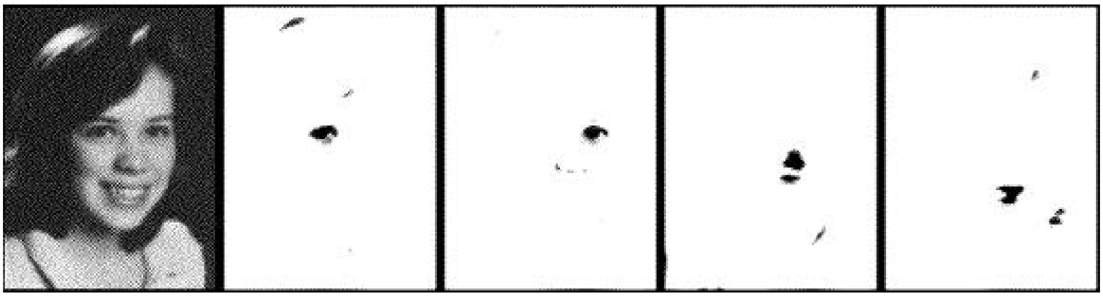

Figure 2.8: The neural network's corresponding feature maps for left eye, right eye, nose and mouth (12)

2.3 Emotion Classification

The research carried out by Ekman on emotions and facial expressions is the main reason behind the interest in the topic of emotion classification. His work has fascinated the researchers and urged them to analyze the emotions through image and video processing techniques. First, face location determination and feature extraction was done from images, then those extracted features were used as input to the classification system which in turn selects a predefined emotion category. Over the past few decades various approaches have been introduced for the classification of emotions. They differ only in the features extracted from the images and in the classification method used to distinguish between the emotions (14; 37).

All these approaches have focused on classifying the six universal emotions. The non-universal emotion classification for emotions like wonder, amusement, greed and pity is yet to be taken into consideration. A good emotional classifier should be able to recognize emotions independent of gender, age, ethnic group, pose, lighting conditions, backgrounds, hair styles, glasses, beard and birth marks.

In classifying emotions for video sequences, images corresponding to each frame have to be extracted, and the features extracted from the initial image have to be mapped between each frame. A sufficient rate for frame processing has to be maintained as well.

The most general method for answering this problem is through the use of a neural network which has obtained over 85% accuracy for emotion classification. Recent research has concentrated on techniques other than neural networks in hope of gaining higher accuracy (14; 37).

2.3.1 Neural Networks

Gargesha and Kuchi (17) has proposed a approach based on Multi Layer Perceptrons and Radial Basis Function Networks (MLP and RBF networks) which classify the seven basic types of emotions: Neutral, Happiness, Sadness, Anger, Fear, Surprise and Disgust. The geometric coordinates of the Facial Characteristic Points (FCPs) for the image, Euclidean distances for the contour points for the image and the differences in inter-feature distances are the only data fed into the neural network. Since the classification is done from the given image and the neutral face, the approach is based on the assumption that the neutral face image corresponding to each image is available to the system. This approach has shown 73% accuracy with the JAFFE database.

Lisetti and Rumelharts' neural network (24) classifies emotions based on signaled emotions and the level of expressiveness. The neural network is capable of dealing with the areas in the face which can carry out independent muscle movements: brow/forehead, eyes/lids and base of nose. The hidden layer in the neural network is connected with each part of the face and with each output unit. This approach works on two emotions: neutral and smiling in order to study how variations in expressiveness affect the overall performance of the system. The results with the FERET database has confirmed that by zooming areas in the face area better results can be obtained for emotion classification.

An ensemble of neural networks has being proposed by Padgett and others' (27; 28) in which each neural network contains a hidden layer with 10 nodes. Each neural network is trained independently with the use of on-line back propagation. The hidden layer with non-linear activation function is trained to map between input-output. The outputs of each network are combined to produce a percentage value for each emotion. In combining the outputs techniques like winner takes all, votes and weighted average output can be used but the technique used in this involves the use of Z scores from all networks.

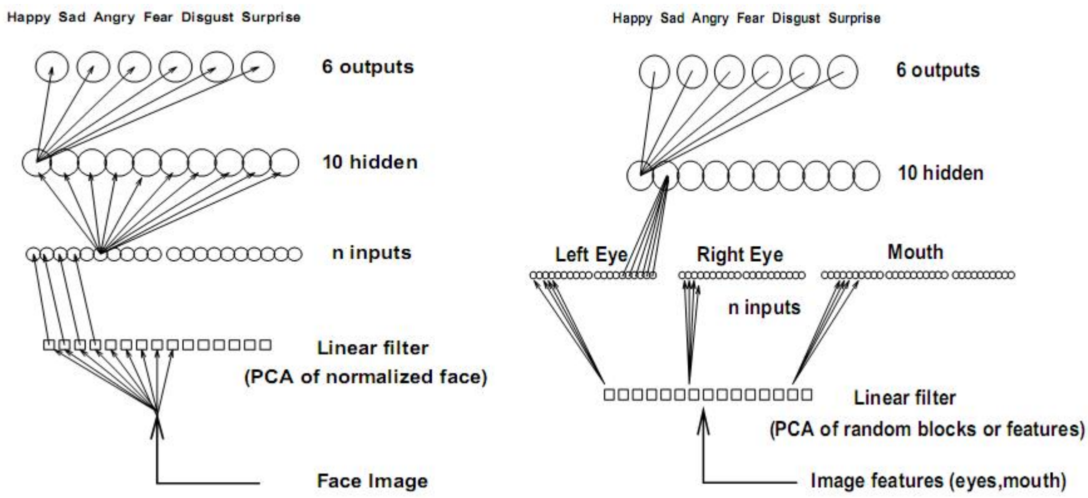

Figure 2.9: A neural network design for full-faced approach (left) and feature based approach (right) (27)

Moving further, the neuro-fuzzy network proposed by Raouzaiou and others' (33) have attempted at classifying primary and intermediate emotions. This classification is performed by translating the Feature Point (FP) movements in to Facial Animation Parameters (FAPs). FAPs have being designed in MPEG-4 for the definition of facial shape, texture and animation of faces. The challenge faced in this approach was the definition of FAP intensities and the definition of emotions in terms of FAPs. One such definition they made is shown in figure 2.10. Six such categories were made for the basic emotions and within each category, intermediate emotions were defined with different emotional and optical intensities.

Joy	*open_jaw (F_3), lower_t_midlip (F_4), raise_b_midlip (F_5), stretch_l_cornerlip (F_6), stretch_r_cornerlip (F_7), raise_l_cornerlip (F_{12}), raise_r_cornerlip (F_{13}), close_t_l_eyelid (F_{19}), close_t_r_eyelid (F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid (F_{22}), raise_l_m_eyebrow (F_{33}), raise_r_m_eyebrow (F_{34}), lift_l_cheek (F_{41}), lift_r_cheek (F_{42}), stretch_l_cornerlip_o (F_{53}), stretch_r_cornerlip_o (F_{54})*

Figure 2.10: FAP definition for the emotion 'Happiness' (33)

2.3.2 Support Vector Machines (SVMs)

y the use of SVM the classification problem can be viewed as a quadratic optimization problem. Since SVM classify data with a set of support vectors by minimizing the structural risk, the average error between input and their target vectors is reduced.

With the use of the SVM package LIBSVM, Dumas (14) has proposed an emotion classification approach. The LIBSVM has being developed by ChihChung Chang and Chih-Jen Lin. The objective of this study was to determine the highest possible accuracy attainable with SVM to classify the Pictures of Facial Affect (POFA) dataset. The POFA dataset is a static human facial image dataset.

Rather than concentrating on all features some researchers have concentrated on one feature and explored its ability to classify emotions. Li (36) has demonstrated that the mouth feature is enough to classify between happiness, sadness and surprise. But to distinguish between all basic emotions the information about the eyes and eyebrows should be obtained in addition to the mouth. He has classified emotions in video sequences with a lip tracking rate of 25 frames per second.

Seyedarabi and others' approach (39) demonstrates a novel FAP based approach to animate the facial emotions from video sequences. 21 predefined facial key points are extracted from the initial frame (Figure 2.2) and the other feature points are automatically extracted from the succeeding frames by the use of cross-correlation based optical flow (Figure 2.11). For animating purposes they have created three object models for eyes, mouth and eyebrows, in which each vertex of the triangle were determined by the feature points extracted.

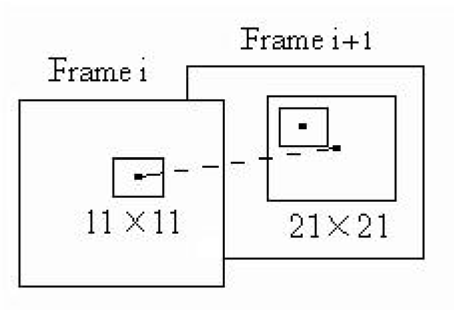

Figure 2.11: Use of cross-correlation based optical flow within frames (33)

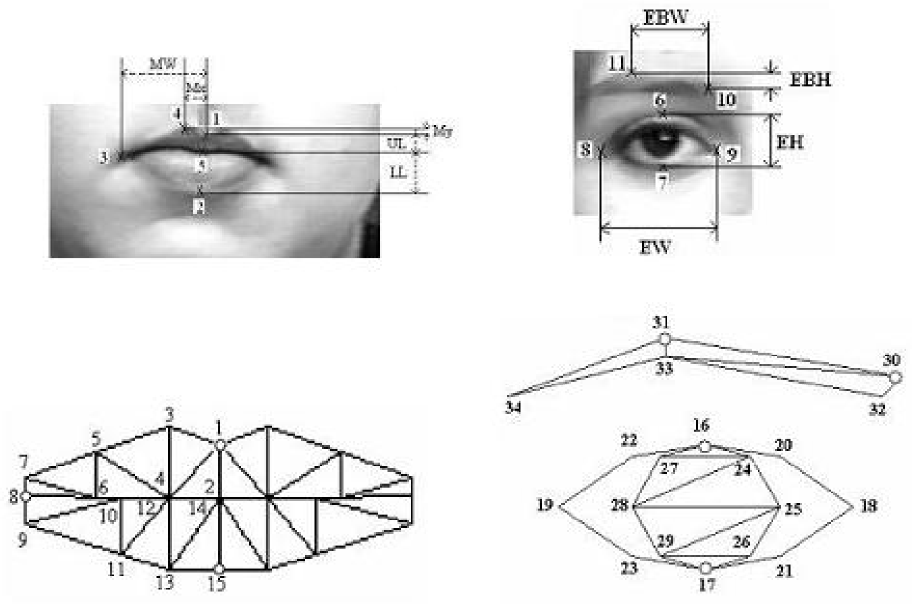

Figure 2.12: object models created for eyes, mouth and eyebrows (33)

The use of trajectories in eigenspace in the classification is demonstrated by Schwerdt and others' approach (35). After the face is tracked using color based techniques, the detailed movements of single features like eye brows were captured by eigenvectors with small eigenvalues. From these eigenvectors, those which do not contribute in distinguishing between emotions are removed and the remaining ones are used for the classification process. The observation that large eigenvalues are not the most reliable eigenvectors for the emotion classification was found from this study.

2.4 Facial Databases

Since a suitable database has to be found for the evaluation process, a survey on the available face databases was also performed. Due to the approach selected, certain constrains has to be met in selecting a database. The requirements were for the database to have color and frontal facial images. The subject, background variations and other variations like spectacles, beard and hair in images were favorable.

After some research we found the Caltech Faces, Georgia Tech Face Database and Facial Expressions and Emotion Database suitable for our study.

2.4.1 Caltech Faces (1)

This database contains 450 frontal face images with 896 x 592 pixels. The images are of 27 or more subjects with different expressions, backgrounds and lighting conditions. This database is provided by Markus Weber at the California Institute of Technology.

2.4.2 Georgia Tech Face Database (2)

This contains images of 50 people taken at the Center for Signal and Image Processing at Georgia Institute of Technology. All images are of 150 x 150 pixels each represented in 15 color JPEG images.

2.4.3 Facial Expressions and Emotion Database (43)

This is a database created by the Technical University at Munich. It contains images of 18 different individuals expressing the six basic expressions. Each individual has expressed each emotion three times. The images contain a simple and defined background.

Chapter 3

Emotion Recognition System

The system for emotion recognition is divided into 3 stages: face location determination, feature extraction and emotion classification. After locating the face with the use of a face detection algorithm, the knowledge in the symmetry and formation of the face combined with image processing techniques are used to process the face region to determine the feature locations. These feature areas are further processed to extract the feature points required for the emotion classification stage. From the feature points extracted, distances among the features are calculated and given as input to the neural network to classify the emotion contained. The neural network has been trained to recognize the 6 universal emotions. (Figure 3.1)

3.1 Face Location Determination Stage

The system offers two methods for face detection. Though various knowledge based and template based techniques can be developed for face location determination, a feature invariant approach based on the skin color is selected as the first method due to its flexibility and simplicity. When locating the face region with skin color, several algorithms can be found for different color spaces. The table 3.1 states several of those skin color segmentation algorithms.

After experimenting with a set of face images, the following condition 3.1 was developed to detect faces.

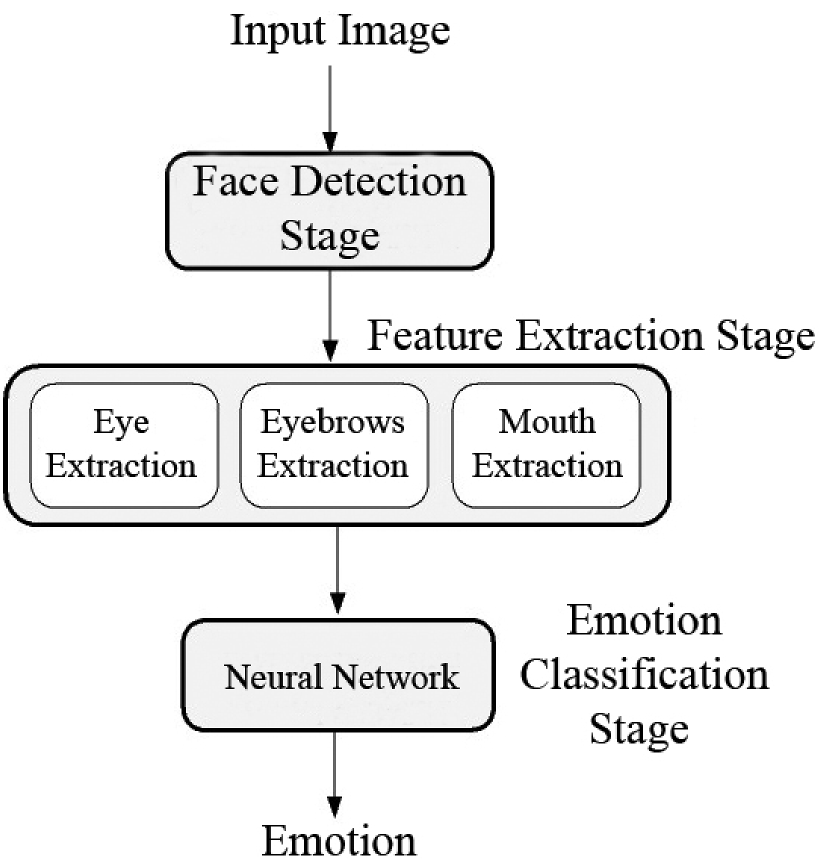

Figure 3.1: An overview of the emotion recognition system

$$(H < 0.1)\ OR\ (H > 0.9)\ AND\ (S > 0.75) \tag{3.1}$$

H and S are the hue and saturation in the HSV color space. For accurate identification of the face, largest connected area which satisfies the above condition is selected and further refined. In refining, the center of the area is selected and densest area of skin colored pixels around the center is selected as the face region. (Figure 3.2)

The second method is the implementation of the face detection approach by Nilsson and others' (26). Their approach uses local SMQT features and split up SNoW classifier to detect the face. This classifier results in more accurate results than the face detection method mentioned earlier. However in the final system, the user also has the ability to specify the face region with the use of the mouse.

Table 3.1: Skin color segmentation algorithms in different color spaces (38; 42)

Color Space	**Criteria**
RGB	R > 95 and G > 40 and B > 20 and {Max(R, G, B) - Min(R, G, B)} and |R-G| > 15 and R > G and R > B
RGB normalized	r/g > 1.185 and (r*b) / $(r+g+b)^2$ > 0.107 and (r*g) / $(r+g+b)^2$ > 0.112

HSV	I > 40 and
	If (13 < S < 110 and 0° < H < 28°) and
	332° < H < 360°
	If (13 < S < 75 and 309° < H < 331°)

YCbCr	77 ≤ Cb ≥ 127 and 133 ≤ Cr ≥ 173

3.2 Feature Extraction Stage

The face region extracted is further processed to extract the feature points required for the emotion classification stage. In this stage several feature points that were identified as important were extracted to use in the classification stage. (Figure 3.9) Feature extraction stage can be separated into two sections: Feature region extraction, Feature point extraction.

3.2.1 Feature Region Extraction

The face detected in the previous stage was further processed to identify eye, eyebrows and mouth regions. Initially, the likely Y coordinates of the eyes was identified then the area around that y coordinate is processed to identify the exact eye regions. Later eyebrow and mouth regions were also extracted based on the eye regions. Finally, a corner point detection algorithm is used to obtain the required corner points from the feature regions.

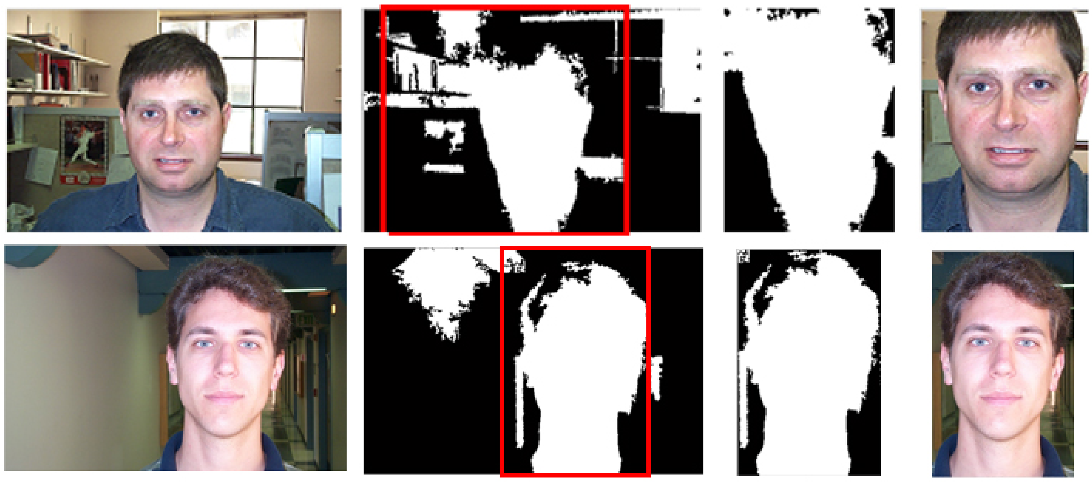

Figure 3.2: Steps in face detection stage a) original image b) result image from condition 3.1 c) region after refining d) face detected

3.2.1.1 Eye Extraction

The eyes display strong vertical edges (horizontal transitions) in images due to its iris and eye white. Thus, the sobel mask in Figure 3.4(a) can be applied to an image and the horizontal projection of vertical edges can be obtained to determine the Y coordinate of the eyes. From the experiments with eye images, it was observed that the use of one sobel mask alone is not enough to accurately identify the Y coordinate of the eyes. Hence, both sobel masks were applied.

To detect the Y coordinate of the eyes, sobel edge detection is applied to the upper half of the face image and the sum of each row is horizontally plotted. Then the top two peaks in horizontal projection of edges are obtained and the peak with the lower intensity value in horizontal projection of intensity is selected as the Y coordinate of the eyes.

Thereafter, the area around the Y coordinate is processed to find regions which satisfy the condition 3.2 and certain other geometric conditions. G in the condition is the green color component in the RGB color space. Figure 3.5(a) illustrates those found areas.

$$G < 60 \tag{3.2}$$

Figure 3.3: Face detection using Nilsson and others' approach (26)

-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0
-1	-2	1

Figure 3.4: Sobel operator a) detect vertical edges b) detect horizontal edges

Furthermore the edge image of the area round the Y coordinate is obtained by using the Roberts method. Then the edge image is dilated and holes are filled to obtain the figure 3.5(b). Thereafter the regions in the figure 3.5(a) grown till the edges in the figure 3.5(b) are included to result the figure 3.5(c). Finally a pair of regions that satisfy certain geometric conditions are selected as eyes from those regions. To achieve this, the largest four areas in the figure 3.5(c) are obtained and the pair of regions which are similar in size and y coordinate but different in x coordinate are selected as eye regions.

3.2.1.2 Eyebrows Extraction

Two rectangular regions in the edge image which lies directly above each of the eye regions are selected as initial eyebrow regions. Then the edge images of these two areas are obtained for further refinement. Here sobel method is used in obtaining the edge

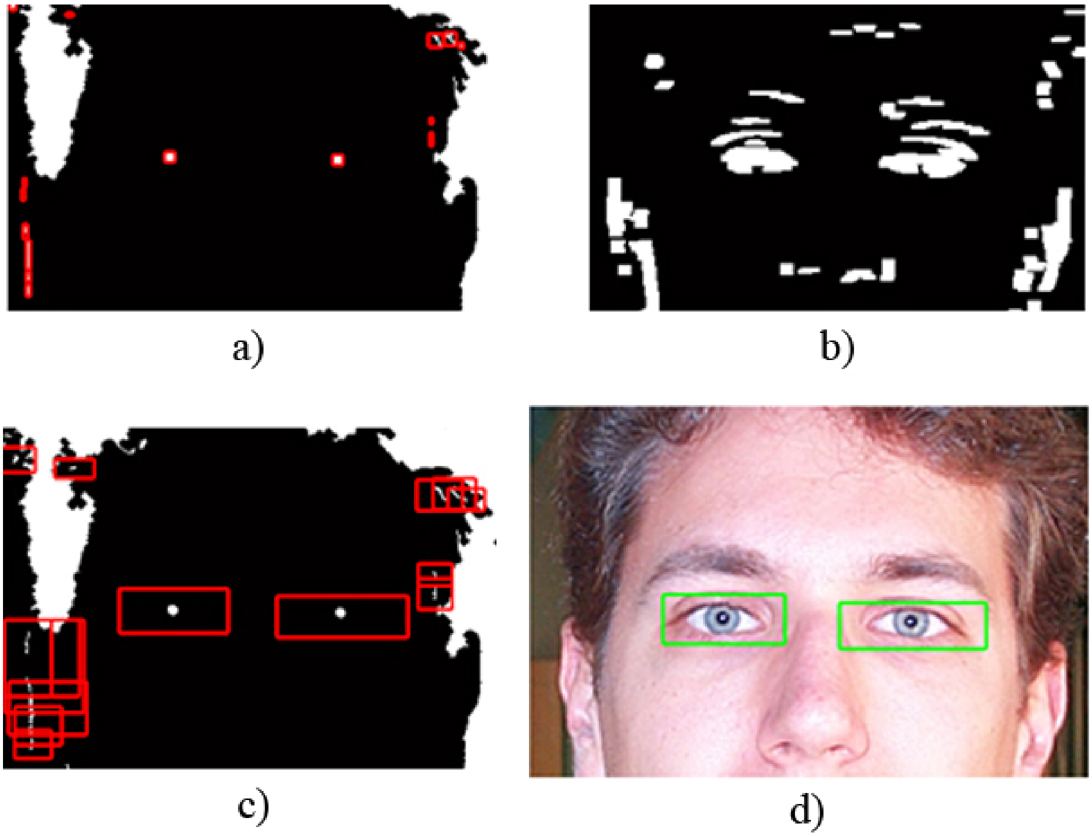

Figure 3.5: Steps in eye extraction a) regions around the Y coordinate of eyes b) after dilating and filling the edge image c) after growing the regions d) selected eye regions

image since it can detect more edges than roberts method. These obtained edge images are then dilated and the holes are filled. The result edge images are used in refining the eyebrow regions.

3.2.1.3 Mouth Extraction

Since the eye regions are known, the image region below the eyes in the face image is processed to find the regions which satisfy the following condition 3.3.

$$1.2 \leq R/G \leq 1.5 \tag{3.3}$$

Furthermore, the edge image of this image region is also obtained by using the roberts method. Then the edge image is dilated and holes are filled to obtain the

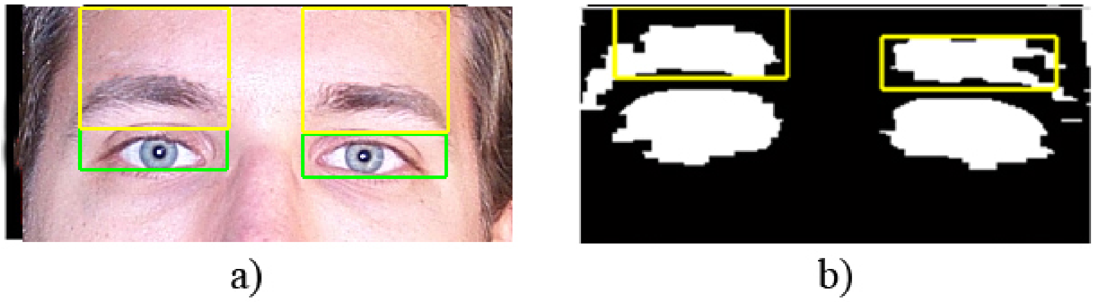

Figure 3.6: Eyebrow regions a) before refining b) after refining

figure 3.7(b). From the result regions of condition 3.3, a region which satisfies certain geometric conditions is selected as the mouth region. As a refinement step, this region is further grown till the edges in figure 3.7(b) are included.

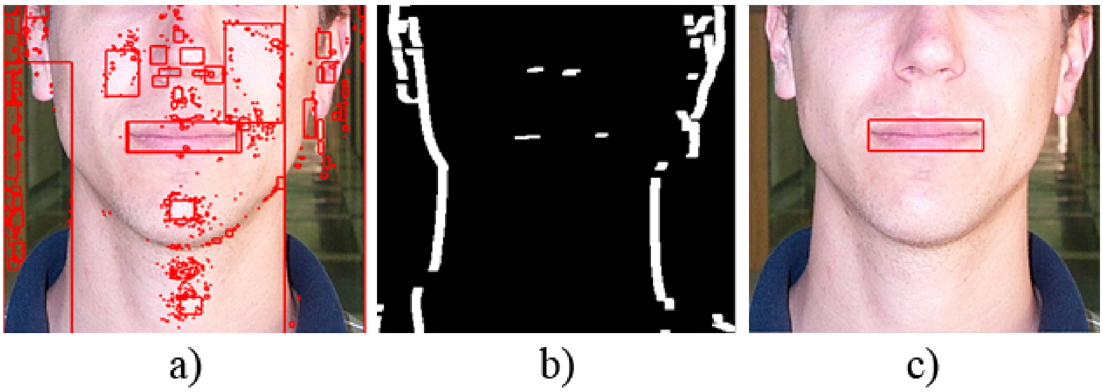

Figure 3.7: Steps in mouth extraction a) result image from condition 3.3 b) after dilating and filling the edge image c) selected mouth region

3.2.2 Feature Point Extraction

After feature regions are identified, these regions are further processed to extract the necessary feature points. The harris corner point detection algorithm (7; 21) is used to obtain the left and right most corner points of the eyes. Then the midpoint of left and right most points is obtained. This midpoint is used with the information in the figure 3.5(b) to obtain the top and bottom corner points. Finally after obtaining the top, bottom, right most and left most points, the centriod of the eyes is calculated.

Likewise the left and right most corner points of the mouth region is obtained with the use of the harris corner point detection algorithm. Then the midpoint of left and right most points is obtained so that it can be used with the information in the result image from condition 3.3 to obtain the top and bottom corner points. Again after obtaining the top, bottom, right most and left most points of the mouth, the centriod of the mouth is calculated.

The point in the eyebrow which is directly above the eye center is obtained by processing the information of the edge image displayed in figure 3.6.

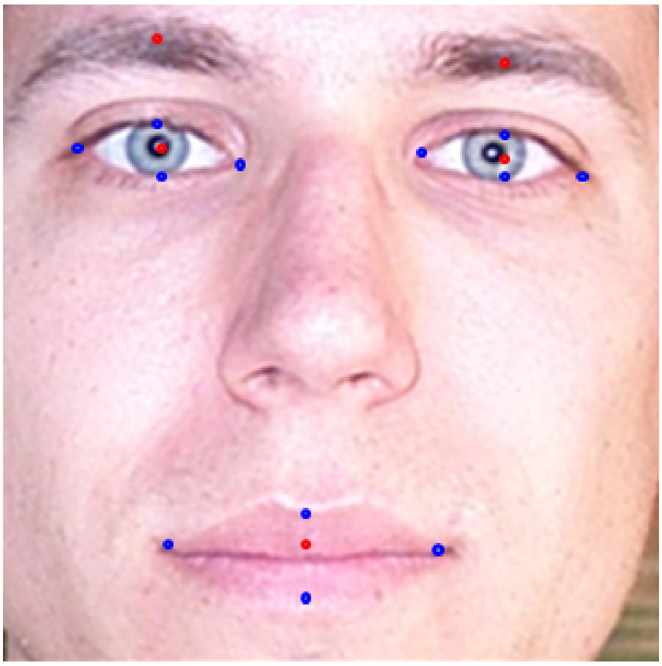

Figure 3.8: Extracted feature points

The above described feature point detection is an image processing based approach. In addition to that approach, the final system provides the ability to manually specify the feature points in the face region with the use of the mouse. This user specified feature point selection is helpful in cases where feature point detection based on the image processing provides inaccurate results.

3.3 Emotion Classification Stage

The extracted feature points are processed to obtain the inputs for the neural network. The neural network is trained so that the emotions neutral, happiness, sadness, anger, disgust, surprise and fear are recognized. 630 images of 15 individuals taken from Facial expressions and emotion database (43) are used to train the network. These 630 images consist of 42 images of an individual posing each of these 7 emotions (6 images of a person for an emotion). The inputs given the neural network are as follows.

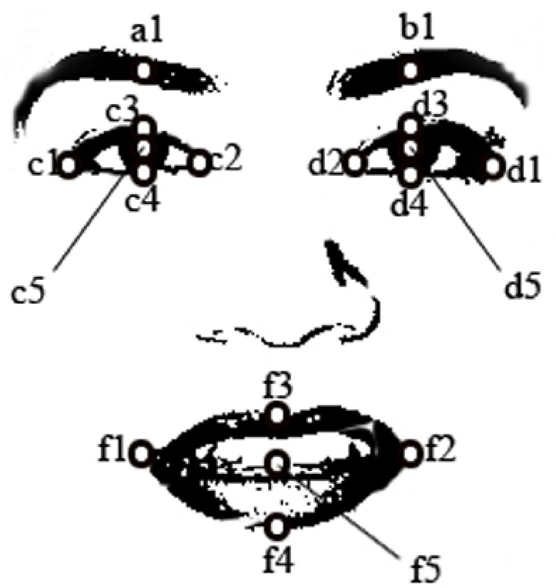

Figure 3.9: Feature points extracted by the feature point extraction stage

- Eye height = (Left eye height + Right eye height) / 2
  = [(c4 - c3) + (d4 - d3)] / 2
- Eye width = (Left eye width + Right eye width) / 2
  = [(c2- c1) + (d1 - d2)] / 2
- Mouth height = (f4 - f3)

- Mouth width = (f2 - f1)
- Eyebrow to Eye center height = [(c5 - a1) + (d5 - b1)] / 2
- Eye center to Mouth center height = [(f5 - c5) + (f5 - d5)] / 2
- Left eye center to Mouth top corner length
- Left eye center to Mouth bottom corner length
- Right eye center to Mouth top corner length
- Right eye center to Mouth bottom corner length
- Eye width/ Eye height
- Mouth width / Mouth height
- Eyebrow to Eye center height / (Eyebrow to Eye center height + Eye center to Mouth center height)

For training the network manually marked feature points for each of these 630 images and the emotion contained in each face were obtained. Then the obtained feature points were processed to calculate the inputs mentioned earlier. Finally the inputs along with the emotion was given to the supervised neural network.

The neural network contained a hidden layer with 30 neurons. The network was trained with a learning rate of 0.1 to achieve a goal of 0.01. After training the neural network was able to identify the training dataset with 89.21% accuracy.

Chapter 4

Evaluation

During the evaluation phase a naive bayes classification and a manual classification was used as a benchmark evaluation for the system.

Naive bayes is a simple probabilistic classifier based on the bayes theorem. The naive bayes classifier is trained using the data of the 630 training images in a supervised learning setting.

An image set containing 85 images considered in this evaluation stage containing images of different people with different emotion in different illumination and background conditions. Each of these images is given to the system and the results are observed. When obtaining the results, the local SMQT features and split up SNoW classifier based face detection method was used to detect the faces in the images. The emotion recognition results pertaining to both methods (image processing based feature points, user specified feature points) are considered in evaluating the system. For each of the results whether the face and features were accurately identified by the system was recorded as well. Then the naive bayes classifier is used to obtain the results for each of these 85 images. The naive bayes results were calculated for both cases of feature detection (image processing based and user specified feature point detection) Then the results were compared to analyze and evaluate the system.

The 25 individuals involved in the evaluation phrase were give the above mentioned 85 images to state the emotion contained. Thus 25 human judgments were obtained for each image. When evaluating the system, the two most prominent emotions were considered from the judgments.

First, we evaluated the system based on its accuracy in predicting the prominent human emotions. The system results were compared with that of the nave bayes classifier to highlight the system accuracy.

In cases where the local SMQT features and split up SNoW classifier based face detection approach was used with the image processing based feature extraction approach, the system was able to correctly classify 20% of images. The nave bayes classifier had only 18.82% accuracy in this case. The figure 4.1(a) illustrates the results in a graph. When the two most prominent human emotions were considered from the human judgments, the system had the ability to produce accurate results 60% of time while the nave bayes classifier had only a 48.24% accuracy (Figure 4.1(b)).

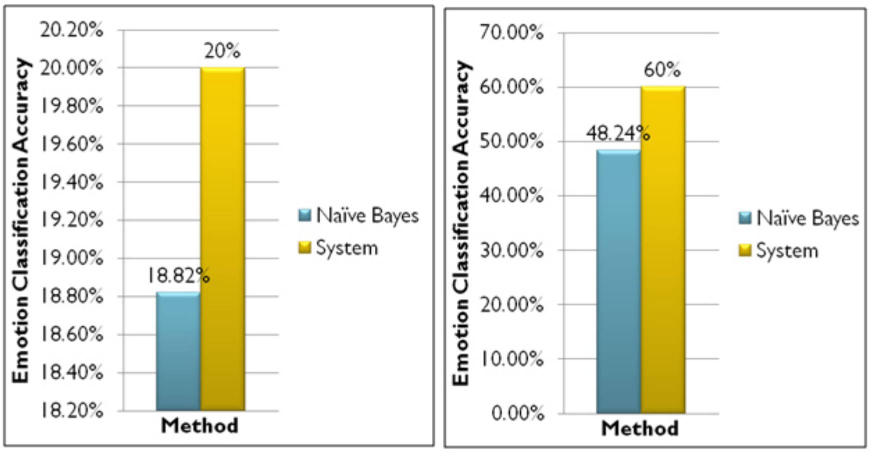

Figure 4.1: Accuracy when SNoW classifier based face detection and image processing based feature extraction was used a) most prominent emotion b) two most prominent emotions

In cases where the local SMQT features and split up SNoW classifier based face detection approach was used with the custom feature selection, the system had 27.06% accuracy. The nave bayes classifier had only 18.82% accuracy. The figure 4.2(a) illustrates the results in a graph. When the two most prominent human emotions were considered from the human judgments, the system had the ability to produce accurate results 55.29% of time while the nave bayes classifier had only a 30.59% accuracy (Figure 4.2(b)).

Thus the system provides a better accuracy in emotion classification than the nave bayes classification. Next, we focused out attention to the system's ability to predict

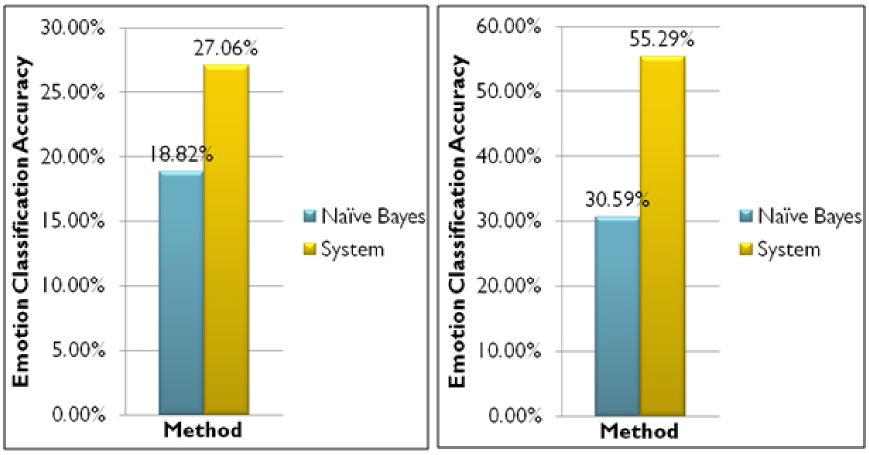

Figure 4.2: Accuracy when SNoW classifier based face detection and custom feature selection was used a) most prominent emotion b) two most prominent emotions

the correct emotion based on the feature detection accuracy.

The system could correctly classify emotions for 67.35% of time when the features are correctly classified and 50% of time when features are incorrectly classified (Figure 4.3(a)) while the nave bayes classifier could correctly classify emotions for 52.08% of time when the features are correctly classified and 43.24% of time when features are incorrectly classified (Figure 4.3(b)).

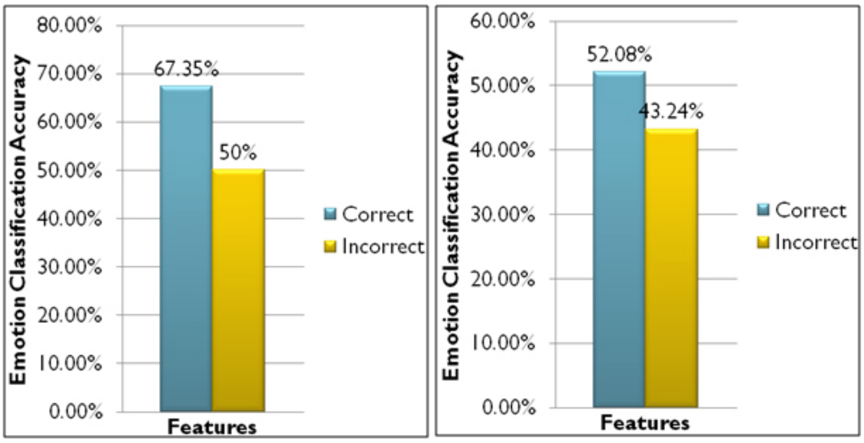

Figure 4.3: Emotion classification accuracy based on the feature extraction accuracy a) system classification b) naive bayes classification

Some individuals of the 25 people who participated in the evaluation had difficulties in identifying emotion in face images when the test dataset contained multiple images of the same person with minor changes. Some even tended to change their opinion about the face images when it appeared twice or more within the test dataset. Though the system provides both manual and automated face detection methods and feature detection methods, most people opted for a automate approach rather than a manual approach.

Chapter 5

Discussion and Future Works

When answering the complex problem of emotion recognition from facial expression, the main problem faced is the selection of suitable inputs for the neural network. The 13 inputs used in this system were chosen after countless experiments. Then the facial features which are required to attain those inputs are extracted. Prior to that, the selection of a particular set of preprocessing steps also required numerous experiments.

The third party face detection method used in the system (local SMQT features and split up SNoW classifier based face detection method) had a little latency in giving the detected face region. However, it was able to detect faces with a higher accuracy in spite of this latency. The accuracy of the image processing based feature detection method needs to be improved. Moreover, the effectiveness of the image processing based feature detection method has to be assessed as well. We used both the facial expressions and emotion database and the caltech database throughout the project. The proposed image processing based feature extraction method is effective for both of the databases. It's applicability to other face databases has to be examined. This feature detection approach could accurately identify features where beards, birth marks are present. Moreover, it could also produce correct results when gender, age and ethnic group varies.

The emotion recognition was performed by calculating a set of inputs to the neural network by using the extracted features. This neural network has a learning rate of 0.1 and a goal of 0.01. After training it obtained an accuracy of 89.21% in classifying the training dataset. In evaluating the proposed system we opted for a human based and a naive bayes based evaluation approach. After the evaluation, we found that the system classified emotions better than nave bayes based classifier. Whenever the features are correctly classified the system had 67.35% in identifying the correct emotion and it also had 50% accuracy in identifying when the features are incorrectly classified.

This system is only capable of identifying emotions of color frontal face images. We intend to extend the proposed system to recognize emotions from video sequences as a future.

Chapter 6

Conclusion

This work try to address the problem of emotion recognition from an image based and a neural network based approach. Emotion recognition is still a difficult and a complex problem in computer science. Most researches have tried to address this problem in various ways. Emotions are expressed by humans in different ways: facial expressions, pose, behavior, actions and speech. We chose facial expressions for the recognition of emotions in humans. Based on the feature points within the face a neural network was trained to identify the emotions contained. The proposed system is very promising and provides better accuracy in emotion recognition than a nave bayes classifier. It has ability in recognizing the correct emotion with 50% accuracy even when the features are incorrectly identified. The system is independent of factors like gender, age, ethnic group, beard, backgrounds and birthmarks. It even has some capability in accurate emotion classifications when poses and lighting conditions vary and also in cases where glasses are present.

Chapter 7

List of Publications

1. W.N.Widanagamaachchi and A. Dharmaratne, Emotion Recognizer : A neural network approach, to appear in Proceedings of 9th International Conference on Intelligent System Design and Applications, 2009.

2. W.N.Widanagamaachchi and A. Dharmaratne, Emotion Recognition with Image Processing and Neural Networks, to appear in Proceedings of 27th National Information Technology Conference, 2009.

Bibliography

[1] Caltech faces database, http://www.vision.caltech.edu/html-files/archive.html, last accessed date: 01st june 2009.

[2] Georgia tech face database, ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/, last accessed date: 01st june 2009.

[3] BBC. Mona lisa 'happy', computer finds, http://news.bbc.co.uk/1/hi/entertainment/4530650.stm, Dec. 2005.

[4] M. A.-A. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno. Face detection and facial feature localization for human-machine interface. *NII Journal*, (5):25–39, 2003.

[5] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[6] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211, New York, NY, USA, 2004. ACM.

[7] R. M. C. Schmid and C. Bauckhage. Evaluation of interest point detectors. In *Int. Journal of Computer Vision*, volume 37(2), pages 151–172, 2000.

[8] B. Cao, S. Shan, W. Gao, and D. Zhao. Localizing the iris center by region growing search. volume 2, pages 129–132, 2002.

[9] I. Cohen, A. Garg, and T. S. Huang. Emotion recognition from facial expressions using multilevel hmm. In *Neural Information Processing Systems*, 2000.

[10] I. Craw, D. Tock, and A. Bennett. Finding face features. In *ECCV '92: Proceedings of the Second European Conference on Computer Vision*, pages 92–96, London, UK, 1992. Springer-Verlag.

[11] M. N. Dailey, G. W. Cottrell, C. Padgett, and R. Adolphs. Empath: A neural network that categorizes facial expressions. *J. Cognitive Neuroscience*, 14(8):1158–1173, 2002.

[12] P. Debevec. A neural network for facial feature location. 1992.

[13] X. Deng, C.-H. Chang, and E. Brandle. A new method for eye extraction from facial image. In *DELTA '04: Proceedings of the Second IEEE International Workshop on Electronic Design, Test and Applications*, page 29, Washington, DC, USA, 2004. IEEE Computer Society.

[14] M. Dumas. Emotional expression recognition using support vector machines. Technical report, Machine Perception Lab, Univeristy of California, 2001.

[15] P. Ekman. Facial expression and emotion. *American Psychologist*, 48:384–392, 1993.

[16] R. S. Feris, J. Gemmell, and K. Toyama. Hierarchical wavelet networks for facial feature localization. In *In ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, pages 118–123, 2002.

[17] M. Gargesha, P. Kuchi, and I. D. K. Torkkola. Facial expression recognition using artificial neural networks. http://www.public.asu.edu/ pkuchi/ expressionrecognition.pdf. In *EEE 511: Artificial Neural Computation Systems*, 2002.

[18] M. R. Girgis, T. M. Mahmoud, and T. Abd-El-Hafeez. An approach to image extraction and accurate skin detection from web pages. *International Journal of Computer Science and Engineering*, 1(2).

[19] M. Grimm, D. G. Dastidar, and K. Kroschel. Recognizing emotions in spontaneous facial expressions. 2008.

[20] H. Gu, G. Su, and C. Du. Feature points extraction from faces. 2003.

[21] C. Harris and M. Stephens. A combined corner and edge detector. In *In Proceedings of the Fourth Alvey Vision Conference, Manchester*, 1988.

[22] I. King. Localized principal component analysis learning for face feature extraction. In *Proc. Workshop 3D Computer Vision*, pages 124–128, 1997.

[23] P. Kuo and J. Hannah. An improved eye feature extraction algorithm based on deformable templates. pages II: 1206–1209, 2005.

[24] C. L. Lisetti and D. E. Rumelhart. Facial expression recognition using a neural network. In *In Proceedings of the Eleveth International FLAIRS Conference. Menlo Park*, pages 328–332. AAAI Press, 1998.

[25] T. M. Mahmoud. A new fast skin color detection technique. *In Proceedings of world academy of Science, Engineering and Technology*, 33, 2008.

[26] M. Nilsson, J. Nordberg, and I. Claesson. Face detection using local smqt features and split up snow classifier. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007.

[27] C. Padgett and G. W. Cottrell. Representing face images for emotion classification. *Advances in Neural Information Processing Systems*, 9:894–900, 1997.

[28] C. Padgett, G. W. Cottrell, and R. Adolphs. Categorical perception in facial emotion classification. In *In Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pages 249–253. Erlbaum, 1996.

[29] M. Pantic and L. J. M. Rothkrantz. An expert system for multiple emotional classification of facial expressions. In *Proceedings, 11th IEEE International Conference on Tools with Artificial Intelligence, 1999, Page(s): 113*, pages 113–120, 1999.

[30] M. Pantic, L. J. M. Rothkrantz, and H. Koppelaar. Automation of nonverbal communication of facial expressions. In *in: EUROMEDIA 98, SCS International*, pages 86–93, 1998.

[31] T. Pham and M. Worring. Face detection methods: A critical evaluation. *ISIS Technical Report Series, University of Amsterdam*, 11, 2000.

[32] T. V. Pham, M. Worring, and A. W. M. Smeulders. Face detection by aggregated bayesian network classifiers. *Pattern Recogn. Lett.*, 23(4):451–461, 2002.

[33] A. Raouzaiou, S. Ioannou, K. Karpouzis, N. Tsapatsoulis, S. Kollias, and R. Cowie. An intelligent scheme for facial expression recognition. O. Kaynak (ed.), Artificial Neural Networks and Neural Information Processing, Lecture notes in Computer Science 2714, Springer, pp. 1109 1116, 2003., 2003.

[34] C. Sanderson and K. K. Paliwal. Fast feature extraction method for robust face verification. *Electronics Letters*, 8:1648 – 1650, 2002.

[35] K. Schwerdt. Visual recognition of emotional states.

[36] U. Sderstrm and H. Li. Emotion recognition and estimation from tracked lip features.

[37] N. Sebe, Y. Sun, E. Bakker, M. S. Lew, I. Cohen, and T. S. Huang. Towards authentic emotion recognition.

[38] I. S. S. Setiawan Hadi, Adang S. Ahmad and F. Wazdi. Dewa : A multiaspect approach for multiple face detection in complex scene digital image. *ITB Journal*, 1C(1), 2007.

[39] H. Seyedarabi, A. Aghagolzadeh, and S. Khanmohammadi. Facial expressions animation and lip tracking using facial characteristic points and deformable model. *International Journal of Information Technology*, (4).

[40] W. K. Teo, L. C. D. Silva, and P. Vadakkepat. Facial expression detection and recognition system. 2008.

[41] V. Vezhnevets and A. Andreeva. A comparative assessment of pixel-based skin detection methods. Technical report, Graphics and Media Lab. Moscow State University, Moscow, Russia, 2005.

[42] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *in Proc. Graphicon-2003*, pages 85–92, 2003.

[43] F. Wallhoff. Facial expressions and emotion database, http://www.mmk.ei.tum.de/ waf/ fgnet/feedtum.html, last accessed date: 01st june 2009., 2006.

[44] Z. Xue, S. Z. Li, and E. K. Teoh. Facial feature extraction and image warping using pca based statistic model. In *In Proceedings of 2001 International Conference on Image Processing. Thessaloniki*, pages 7–10, 2001.

[45] M. H. Yang, N. Ahuja, and D. Kriegman. A survey on face detection methods, 1999.

[46] M. H. Yang, D. J. Kriegman, S. Member, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:34–58, 2002.