

Automatic Facial Expression Recognition System using Orientation Histogram and Neural Network

Darli Myint Aung

University of Computer Studies, Mandalay,
Myanmar

Nyein Aye

University of Computer Studies, Mandalay,
Myanmar

ABSTRACT

Facial expression is the most challenging task in the field of computer vision. In this paper, an automatic facial expression recognition system from a still frontal posed image is presented. This system recognizes the human expression by observing the shape of the mouth. This paper uses color based segmentation followed by template matching for face detection and localization. For mouth segmentation, Canny_Template method is used. Orientation Histogram is used for feature extraction. Feed forward neural network is used as a classifier for classifying the expressions of supplied face into five basic expressions like surprise, neutral, sad, happy and angry. Experiments are carried out on Myanmar Facial Expression Database and give the correct performance in terms of 100% accuracy for training set and 70.71% accuracy for test set.

Keywords

Facial expressions, Canny_Template, Orientation Histogram, Feed Forward Neural Network.

1. INTRODUCTION

Facial expression is the most challenging task in the field of computer vision since the last two decades. Various researchers are still try to recognition the expressions in order to robustness and efficient to their system. A facial expression recognition system can be used to regard behaviors by involving certain facial expressions with attitudes. Although humans can interpret the facial expressions, computers can learn to understand human behaviours and be trained to respond appropriately. Facial expression recognition tools can be used to improve interaction between the human and computers. By noting face, one can decide whether a man is serious, happy, surprise, sad, anger and neutral. Recognizing the expression of a man can help in many of the areas.

Many applications such as virtual reality, videoconferencing, and user profiling and customer satisfaction studies for broadcast and web services, require efficient facial expression recognition in order to achieve the most wanted results [1] [2]. Therefore, the impact of facial expression recognition on the above mentioned application areas, is constantly increasing.

According to Mehrabian [3] pointed out that 7% of human communication information is communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. Therefore facial expressions are the most important information for emotions perception in face to face communication.

Hybrid Emotional Neural Network (HENN) for classification of emotions from facial expressions is described by [4]. In their paper, features are extracted from facial expressions by applying Gabor wavelet and Discrete Cosine Transform (DCT). And then, their result compared with the standard

Feed Forward Neural Network and Radial Basis Neural Network.

The feature extraction is the most significant stage that develops a successful expression recognition system. If insufficient features are used, the facial expression recognition system could be unsuccessful to achieve accurate recognition rate. In this paper, orientation histogram method is proposed to get the feature, and then feed forward neural network will be used as classifier to recognize facial expression.

There are several methods have been proposed for facial expression recognition system that is eigenfaces using neural network [5], the Eigenvector spaces that have been calculated independently for each class are considered; then the similarity between the images is obtained by measuring the mean-square error for the reconstructed image for each class [6], the classification of emotions using Support Vector Machine model [7] and the gabor filter with neural network is used to classify the five different facial expressions [8]. Skin pixel detection using RGB color image is presented by [9]. RGB, YCbCr, CEILAB and HSV color models is used to segment the skin color, and then skin region is tested whether the face or not [10].

In this paper, the automatic facial expression recognition system based on orientation histogram and neural network from Myanmar Facial Expression Database is presented. Unlike other existing facial expression recognition systems, the proposed system uses the geometric feature of human face that mainly emphasizes the shape of the mouth from the static image. The contributions of this system are – (1) Myanmar skin model is constructed by using color based segmentation such as YCbCr color space approach (2) Canny_Template approach is proposed for mouth segmentation phase (3) an efficient extraction method (Orientation Histogram) based on geometric feature such as mouth region is proposed for feature extraction phase.

The experimental results show that the feed forward neural network method can recognize the appropriate facial expression than other methods.

Details of this system are described in the remainder of this paper. The paper is organized as follows: Section 1 gives introduction, Section 2 highlights Myanmar Facial Expression Database, and Section 3 presents proposed system. In Section 4, experimental results and analysis are presented. Finally, conclusion and future works are summarized in Section 5.

2. Myanmar Facial Expression Database

Subjects in the available portion of the database were ten male subjects. They ranged in age from 18 to 30 years. All are the people of Myanmar. They posed for the five basic expressions: neutral, happy, surprise, sad and the anger face (see figure 1). The observation room was equipped with a chair for the subject and Nikon D80 camera. This camera was located directly in front of the subject. Each of the subjects posed with two or three examples per expression to make a

total of 152 images. The still images were captured in a controlled environment. The imaging and recording conditions (camera parameters, illumination setting, and camera distance) were carefully controlled and constantly recalibrated to ensure that settings are identical across subjects. The resulting RGB color images are 250×250 pixels in size.



Fig 1: Sample image from Myanmar Facial Expression Database

3. Proposed System

The facial expression recognition system is proposed to recognize facial expressions to identify five universal human expressions: neutral, happy, anger, sad and surprise using the face images. Figure 2 illustrates the flow chart of the facial expression recognition system. This system is divided into four phases: face detection, mouth segmentation, feature extraction from the mouth region and finally classification.

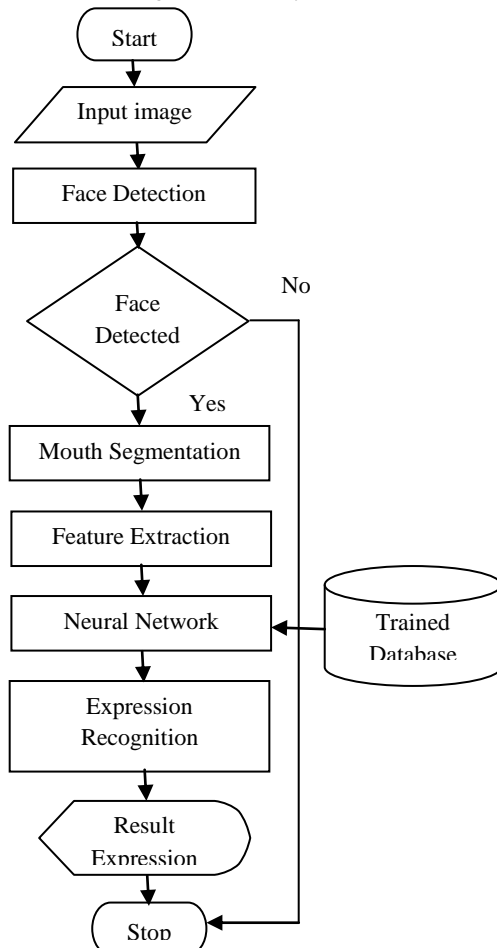


Fig 2: Flow chart of Facial Expression Recognition System

3.1 Face Detection

In this paper, the face detection system uses color based segmentation and template matching. It consists of three steps. The first step is to make a skin model that emphasizes on Myanmar skin. The images are converted from RGB to

YCbCr color space. The image is segmented into a luminosity component and chrominance components. To facilitate fit the data into Gaussian Model by computing these components and the covariance matrix. The second step is that adaptive threshold method segments the skin. After segmenting, morphological processing is used to make smooth the left over the noise in the background. It is necessary to remove the unnecessary specs in order to speed future processing. Finally, the template matching is used as not only for detecting a final result, but also for locating the centroid of the faces. This system uses a template face to determine if the segmented region that has at least one hole, and then a height to width ratio. The number of holes in a region is determined by using Euler method. After that, aspect ratio test and template matching test is performed. The centroid of each region is computed and the face region on the original image is marked. And then, the gray scale is converted for the next phase (see figure 3).

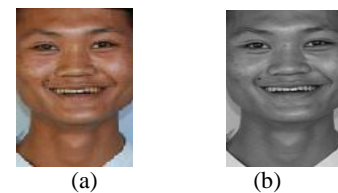


Fig 3: (a) Detected face (b) Gray face image

3.2 Mouth Segmentation

The second phase is mouth segmentation. Firstly, the gray face image is performed using the canny edge detection with static threshold value which after manual iterations chosen to be 0.4. And then, the mouth region is cropped manually for template. After that, take templates from at least 10 images and save them in a folder with names say "T1.jpg", "T2.jpg" and so on (see figure 4).

Canny_Template method calculates the correlation between test image and pre-selected mouth templates. It means the final decision comes from the similarity between input image and template. Lastly, it will invariably point to mouth region as shown in figure 5.

Template matching is ideally a good fast and simple step for detecting objects. It is scale dependent and rotation dependent. The computational cost of template matching is large. If the template is square and of size $m \times m$ and is matched to an image of size $N \times N$ then since the m^2 pixels are matched at all image points, the computational cost is $O(N^2 m^2)$ [11]. Therefore, we considered small size for template and is matched to an image with small size.

So the mouth detection time is taken 6 seconds approximately by running on Pentium Dual-Core CPU T4200 @ 2.00 GHz and 2.86 GB of RAM.

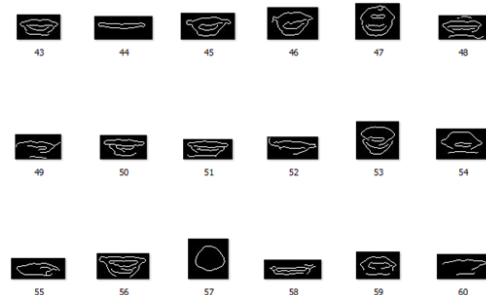


Fig 4: Sample template image folder

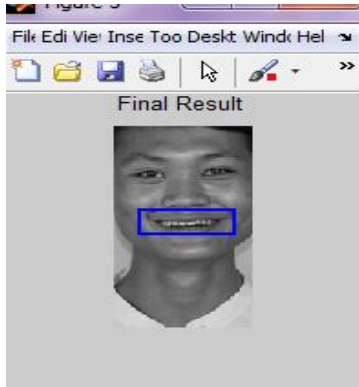


Fig 5: Detected mouth region using Canny_Template method

3.3 Orientation Histogram Feature Extraction

The orientation histogram is a simple and fast algorithm that can be developed to work on this system. It can be used as a feature vector for expression classification. This system computes how often each orientation element occurred in the image. Actually, however, a set of training feature vectors can be chosen with substantially different orientation histograms from each other.

Firstly, dimension reduction is made with resizing images to 25×72 image to reduce computational time in classification. And then, this system calculates the histogram using gradient for each pixel in the image and stores the angle of the gradient into bins. The angles and magnitudes of the edges are computed by using the horizontal and the vertical gradient data in equation (1).

$$m(x, y) = \sqrt{dy^2 + dx^2}$$

$$\theta(x, y) = \arctan\left(\frac{dy}{dx}\right) \quad (1)$$

where m is the magnitude and θ is the orientation of the gradient. Secondly, this system rearranges the image blocks into columns. And then, the angle histogram which is a polar plot showing the distribution of values grouped according to their numeric range is created as shown in figure 6. Thirdly, the vector for values can be ranged from 0° to 90° . The values only lie on the first and last quarter because for the elements of x , $\arctan(x)$ is between the range of $-\pi/2$ and $\pi/2$. Lastly, each bin was set to 10° , which the histogram bins to 19 bins. Therefore, orientation analysis should give robustness in illumination changes while histogram will offer translational invariance. Simultaneously, this feature vector would be the input to the network.

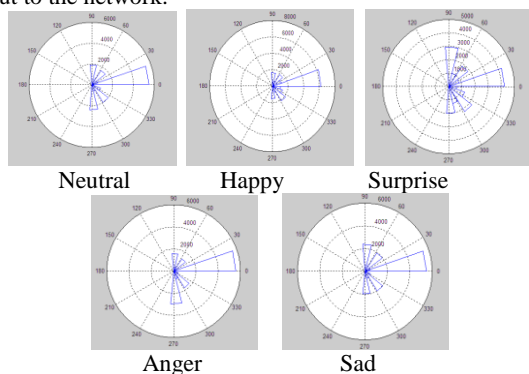


Fig 6: Orientation histogram examples

3.4 Expression Classification

This system can classify the five expressions such as neutral, happy, surprise, anger and sad by using feed forward neural network which is among the pattern recognition. Several different types of network structures exist in literature. Feed Forward Neural Network (FFNN) is used for classifier. It is the simplest structure in the neural network. Figure 7 is a kind of FFNN.

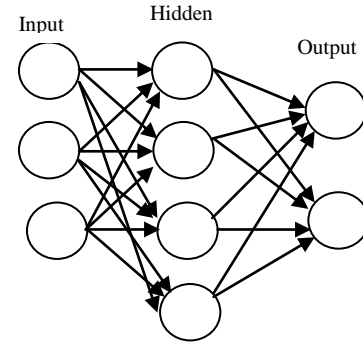


Fig 7: Multi Layer Network Structure

The information moves through input to output and does not perform any loop or cycle operations. The supervised learning is based on the Scale Conjugate Gradient method, minimizing the global error on the output layer. The system network properties are: input layer has 19 inputs, hidden layer has 10 neurons and output layer has 5 neurons. Output layer has 5 neurons since the number of expression in database is 5.

After generating, network must be trained to classify the given images with respect to facial expression database. Before testing, facial expression database is created. This system has 50 training sample. As a result of, training matrix has 950×50 size. And then, target matrix is generated to inform network vectors belonging to expressions.

Target vector is created with putting '1' with respect to order number of the name in database and other elements '0' for target vector. After that, training of NN can be performed. Training performance and goal errors are set to $1e^{-5}$ to classify given image correctly. The output values are compared with target values. The procedure is repeated to get the desired accuracy. The output is compared with training phase output to match the correct one.

4. Experimental Results and Analysis

In this paper Neural Network model is constructed for Myanmar Facial Expression Database for frontal view facial images. Initially, the face portion is segmented by using color based segmentation and template method, and hence face localization is achieved.

After getting the face region, the mouth is detected with the proposed method (Canny_Template). For feature extraction, the orientation histogram is used to extract the mouth feature vector.

Mouth image is resized to 25×72 because the smaller the vector the faster the processing time. Figure 8 shows localized permanent facial feature (mouth) which is used as input to neural networks.



Fig 8: Mouth image

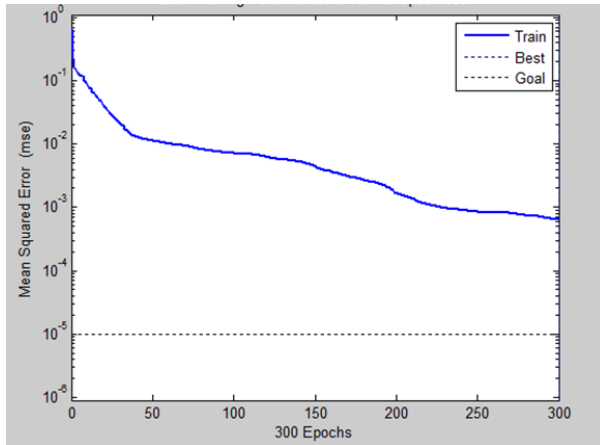


Fig 9: Performance Plot of Neural Network

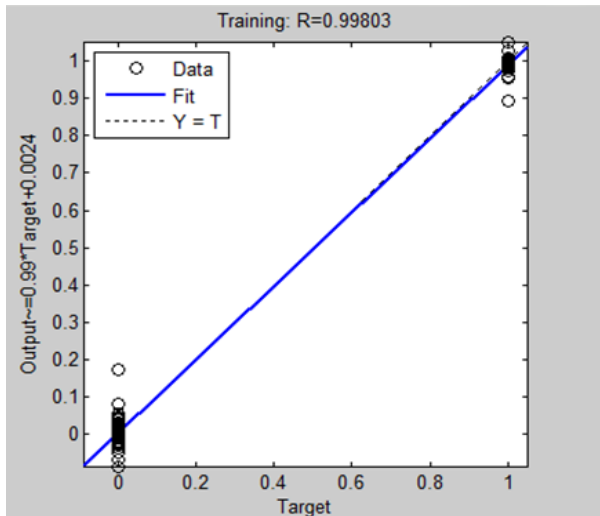


Fig 10: Regression Plot

Plot shows that the network learns gradually and reaches towards the goal. Performance plot of neural network is shown in figure 9. The regression plot is as shown in figure 10. The output has $0.99 * \text{target} + 0.0024$. The figure 10 shows the GUI for classification and recognition of the mouth and the results of face detection, extracted mouth feature with bounding box and its recognition and classification. In this figure, the sample image exhibits happy (HA) expression.

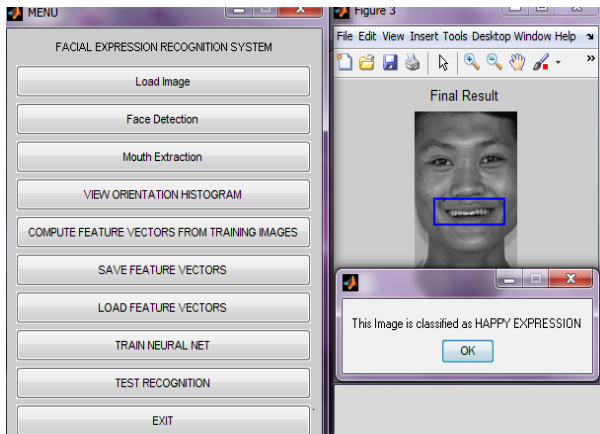


Fig 10: GUI for classification and recognition of the mouth and the result

Table 1. Confusion matrix of facial expression recognition system

	Surprise (SU)	Happy (HA)	Neutral (NE)	Anger (AN)	Sad (SA)	%
SU	19	1	-	-	-	95.00%
HA	-	20	-	1	-	95.24%
NE	-	3	16	-	1	80.00%
AN	-	-	6	10	4	50.00%
SA	-	1	4	9	7	33.33%
Total Classification Rate						70.71%

The experiment result can be observed evidently from the table 1. This system could classify happy, surprise, neutral and sad from the result. This system can classify neutral, happy and surprise as maximum rates. However, not only sad can detect 7 expressions but also anger can detect 10 expressions. There was a minimum rate as can be seen from the confusion matrix. The correct classification rate of the mouth phase is 100 % for training set and 70.71% for test set. The proposed feed forward neural network is simple and with low complexity. These results can be concluded the local mouth feature is useful for differentiating expressions.

5. CONCLUSION AND FUTURE WORK

In this paper, Canny_Template for mouth segmentation, Orientation Histogram for feature extraction and Feed forward neural network to tackle facial expressions classification problems is proposed. The mouth region provides confusing information for distinguishing the facial expressions. According to observation, there are the highest values ranging from 0^0 to 10^0 for all expressions when performing the orientation histogram of the mouth region because of a large pixel of the non-edge region. This system could strain the difficult expression classification efficiently. Angry expression and sad expressions may be ambiguous. These expressions are hard to discriminate. The future work will be the dealing with the images of eyes and eyebrows to test the recognition performance, therefore the recognition rate will be improved.

6. ACKNOWLEDGMENTS

I wish to thank Dr. Nyein Aye for his extreme support and guidance. I also thank my family and colleagues for their valuable feedback and thoughtful suggestions.

7. REFERENCES

- [1] Yi, J., R. Qiuqi et al. 2008. Gabor-based Orthogonal Locality Sensitive Discriminant Analysis for face recognition. Signal Processing, ICSP 2008. 9th International Conference on.
- [2] FrChing-Chih, T., C. You-Zhu et al. 2009. Interactive emotion recognition using Support Vector Machine for human-robot interaction. Systems, Man and Cybernetics, SMC 2009. IEEE International Conference on.
- [3] P.Ekman and W.Friesen. 1978. The Facial Action Coding System. Consulting psychologists Press, San Francisco, CA.
- [4] K.V. Krishna Kishore and G.P.S. Varma. 2011. Hybrid Emotional Neural Network for Facial Expression

- Classification. *International Journal of Computer Applications*, Vol.35-No.12, December 2011.
- [5] P.V.Saudagare and D.S. Chaudhari.2012. Human Facial Expression Recognition using Eigen Face and Neural Network. *International Journal of Engineering and Advanced Technology (IJEAT)*.Vol.1, Issue.5, June 2012.
- [6] Daw-Tung Lin. 2006. Facial Expression Classification using PCA and Hierarchical Radial Basis Function Network. *Journal of Information Science and Engineering*, Vol.22, pp.1033-1046.
- [7] S.Srivastava and K. Asawa.2012. Real Time Facial Expression Recognition using a Novel Method. *The International Journal of Multimedia & Its Applications*. Vol.4, No.2, April 2012.
- [8] Z. Abidin and A. Harjoko.2012. A Neural Network based Facial Expression Recognition using Fisherface. *International Journal of Computer Applications*. Vol. 59, December 2012.
- [9] C.Lin. 2005. Face Detection by Color and Multilayer Feed Forward Neural Network. *IEEE International Conference on Information Acquisition*. pp.518-523, 2005.
- [10] D. S. Raghuvanshi and D. Agrawal. 2012. Human Face Detection by using Skin Color Segmentation, Face Features and Regions Properties. *International Journal of Computer Applications*. Vol. 38, January 2012.
- [11] Mark S. Nixon, Alberto S. Aguado. *Feature Extraction and Image Processing*. Second edition.