Dhaka Univ. J. Sci. 58(2): 175-181, 2010 (July)

Human Facial Expression Recognition Using Region-based Motion Estimation

Ahmedul Kabir and Shaily Kabir

Department of Computer Science and Engineering,Dhaka University, Dhaka-1000, Bangladesh

Email: ahmedulkabir@hotmail.com, shailykabir2000@yahoo.com

Received on 22.03.2009. Accepted for Publication on 13.12.2009.

Abstract

Facial expressions are the indicators to a person's emotional state and also the principle means of human non-verbal communication. If computers would have the ability to perceive and respond to facial expressions, the human-computer interaction would become much more spontaneous. Although several approaches have been proposed to recognize human facial expressions, none have gained universal acceptability due to their time complexity and difficulty of implementation. This paper proposes a simple method based on region-based motion estimation that can identify facial expressions with acceptable precision in real time. The proposed method identifies six regions of the human face that incur significant visual change caused by the facial motion of expressions. The amount of motion in each region, estimated by the amount of variation in intensity level, is observed and recorded. The results of the observation are used to describe a rule-based classifier that identifies a given emotion as belonging to one of the six universal expressions. The performance of the proposed method is tested on several participants, showing 76% accuracy in recognizing facial expressions.

Keywords: Facial expression, Facial regions, Motion estimation, Region population

I. Introduction

Due to the significant involvement of computers in today's life, the human-computer interaction has become a major field of research for the computer scientists. Emotions can be determined using various features including facial expressions, speech modulation, body language etc. Among them, facial expressions are the most definitive means of recognizing human emotions in an authentic and believable manner. Though a person's portrayal of emotion largely depends on his/her personality and cultural background, psychological research has classified six facial expressions that correspond to distinct universal emotions, which are disgust, sadness, happiness, fear, anger and surprise [1]. It has been observed that these emotions result in similar facial expressions irrespective of the person's nature or culture. The facial movement and timing involved with each of these expressions have been identified. This paper presents a real-time vision-based system that recognizes one of the six universal expressions on a person's face. Based on the position of the eyes, the human face is segmented in several regions that are relevant to the expressions. Through the calculation of facial movement in different regions, the proposed method determines which of the six expressions is being displayed at a given time.

This paper is organized into several sections. In Section II, a brief summary of the related work in the field of expression recognition is presented. Section III provides a detailed explanation of the proposed method for human facial expression recognition. The results obtained from the proposed method are discussed in Section IV. Finally in Section V, the performance of the method is analyzed and conclusions are drawn.

II. Related Work

Paul Ekman and his colleagues have conducted extensive studies of human facial expressions, where they found evidence to support universality in facial expressions [1]. Ekman and Friesen developed the Facial Action Coding System (FACS) where movements on the face are described by a set of action units (AUs) [2]. Each facial expression may be described by a combination of AUs. Ekman's work inspired many researchers to analyze facial expressions by means of image and video processing. By tracking facial features and measuring the amount of facial movement, they attempt to categorize different facial expressions.

Recent work in computer-assisted quantification of facial expressions started with Mase who used optical flow to recognize facial expressions [3]. Lanitis et al. used a flexible shape and appearance model for person identification, gender recognition and facial expression recognition [4]. Black and Yacoob used local parameterized models of image motion to obtain non-rigid motion [5]. Once obtained, these parameters are supplied to a rule-based classifier to recognize the six basic facial expressions. Yacoob and Davis computed optical flow and used similar rules to classify the six facial expressions [6]. Rosenblum et al. also computed optical flow of regions on the face. In order to classify expressions, he applied a radial basis function network [7]. Essa and Pentland used an optical flow region-based method to recognize expressions [8]. Cohen et al. describe classification schemes for facial expression recognition in two types of settings: dynamic and static [9]. More recently, Sebe et. al. described an algorithm using Bayesian network that specifically identifies authentic emotions [10]. Pantic and Patras focused on the dynamics of facial muscles, and proposed a system that recognizes facial actions based on a sequence of facial images [11].

These methods are similar in the general sense that they first extract some features from the images, then these features are fed into a classification system, and the outcome is one of the pre-selected emotion categories. However, the procedures of feature extraction and classification are considerably different.

III. Proposed Method for Facial Expression Recognition

Certain facial expressions cause certain amount of movement in different parts of the face, especially the eyes, eyebrows, lips and jaws. Due to these movements, the visual characteristic of the face undergoes some change, which enables other human beings to identify the expression manifested. A computer with a vision system, i.e. a camera, can capture the image of the human face and track the changes in intensity levels that occur whenever some sort of expression is shown. In this study, these particular changes are recorded, which are later applied for the calculations leading to the identification of the expression.

At first, the human face is detected using FaceVACS, a face recognition software [12]. The facial region is then converted to grayscale and normalized to a fixed resolution, thus making it easier to manipulate the facial image. The eye positions and eye distance in the normalized image are calculated. Based on these values, six regions of the face are specified that are generally affected by the facial expressions. The areas for these regions vary from person to person based on their eye positions in the input image.

Given reasonable light condition, the facial motions cause variations in the intensity levels of a subset of the pixels contained in the facial image. In order to detect the amount of such variation, the current input image of the person is compared with a previously captured image showing his/her normal face. The two images are subtracted and their pixel-by-pixel absolute difference is found. A threshold value is then applied on the resulting image, thus converting it into a bi-level image having black and white color pixels only. The threshold value is chosen based on observation, and is equal to L/10 where L is the number of gray levels.

The number of white pixels in each facial region is counted. This count represents an estimation of the facial motion for that particular expression. Based on the pixel count in each region, calculations are carried out that determine which of the six universal expressions is currently being exhibited. The whole procedure is shown in the block diagram of Figure 1.

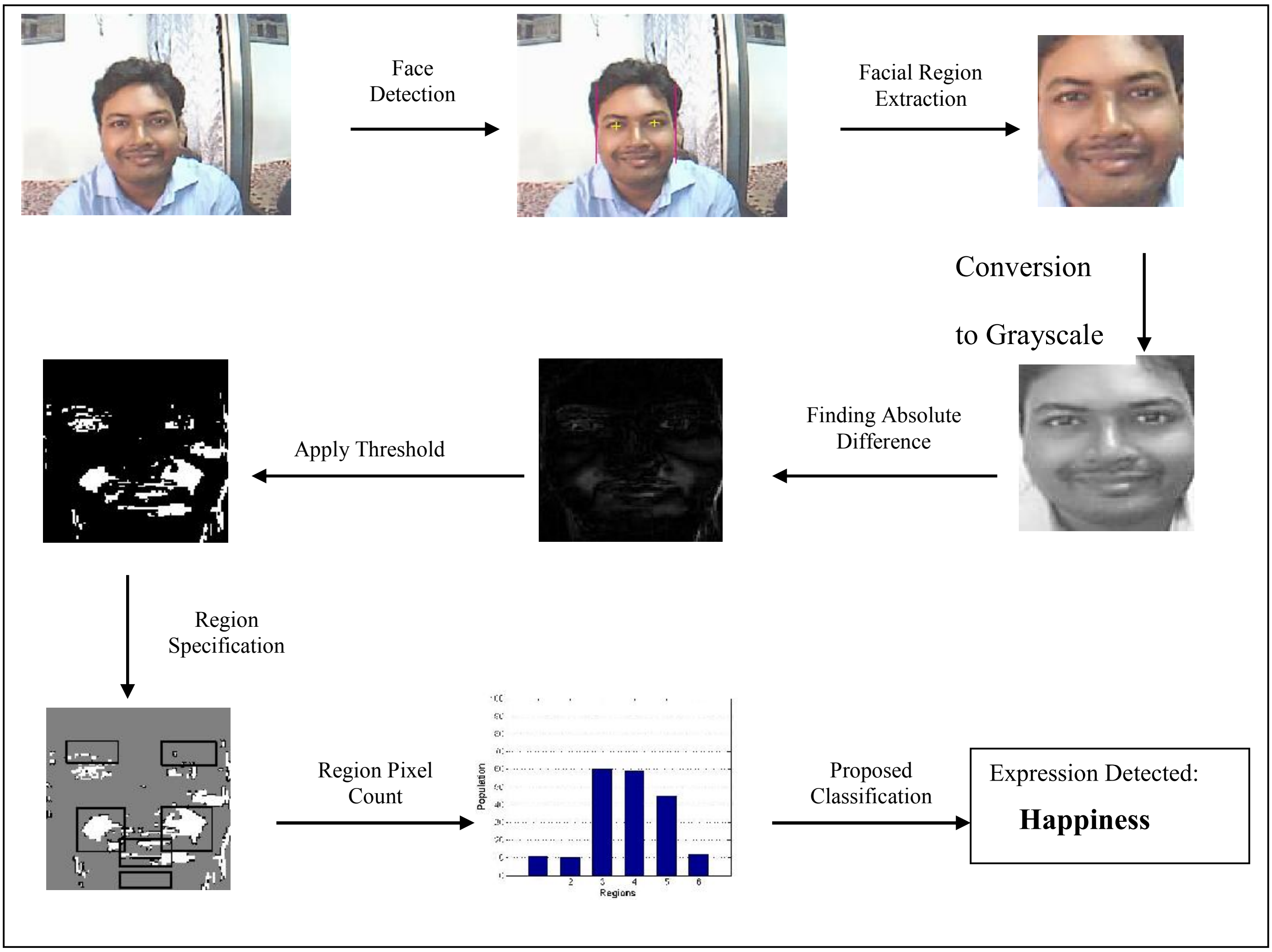


Fig. 1. Block Diagram of the Proposed Method for Expression Recognition

Detection of Areas of Facial Motion

Facial motion here refers to the movement in different areas of the face caused by facial expressions. For example: raising of eyebrows, expansion of lips, opening of mouth etc. Each expression causes some feature of the face to move, therefore inducing a change in the gray level intensity of the image. This intensity change is tracked by the subtraction of two images: one a real-time snapshot of the face, and the other a pre-stored image of the face in its normal state. Every four frames, the pixel-by-pixel absolute difference between these two images are calculated. If the difference exceeds a predefined threshold, it is stored as a white pixel with a value of one, otherwise as a black pixel with a value of zero. Figure 2 shows the procedure with some sample images.

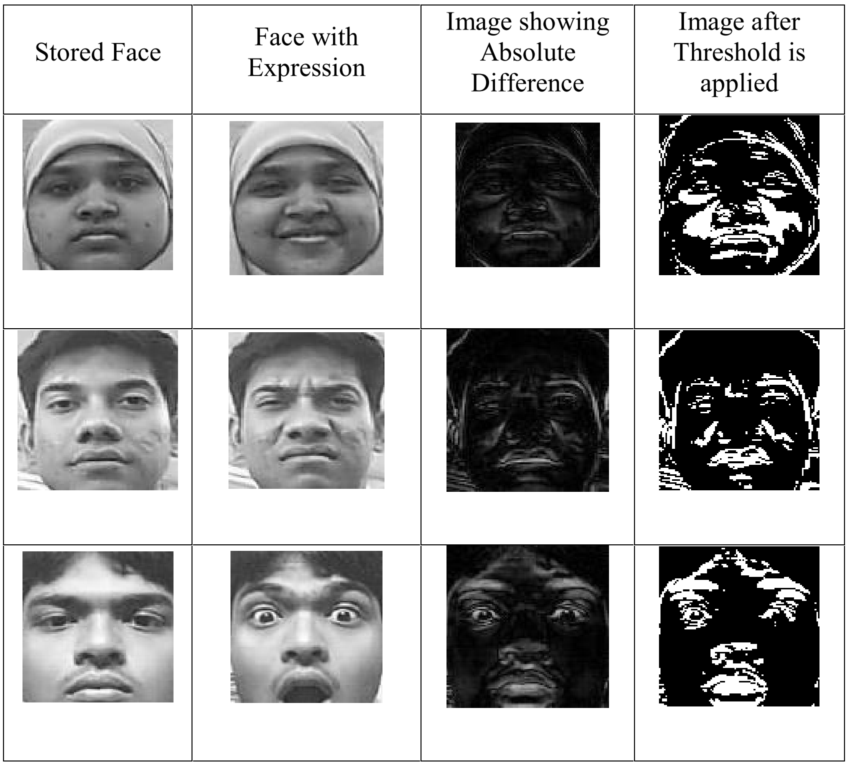


Fig. 2. Detection of the Areas of Facial Motion for some Sample Images

Specification of Facial Regions for Expression Recognition

Through observation of the samples, six regions of the face are identified where substantial changes occur. Their locations are as follows: Two regions near the two eyebrows, Two regions on the right and left sides of the lips, one region starting from the upper lip down to just below the lower lip, and one region near the chin. These regions are shown in Figure 3 on the face of a person with their corresponding region indices.

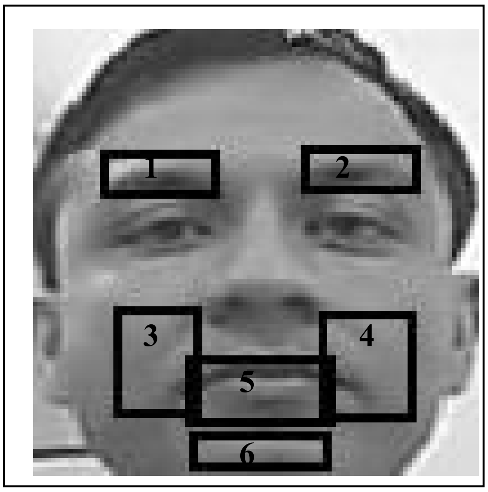


Fig. 3. Six Regions where Most Variations Occur

The positions of the facial regions vary according to the size and shape of the individual's face as well as his/her distance and orientation from the camera. These regions are detected according to the position of the eyes and distance between them. Let (E_{LX} , E_{LY}) and (E_{RX} , E_{RY}) be the coordinates of the left and right eyes respectively. The Eucledian distance between the eyes is calculated using the following formula:

$$D = \sqrt{(E_{LX} - E_{RX})^2 + (E_{LY} - E_{RY})^2}$$

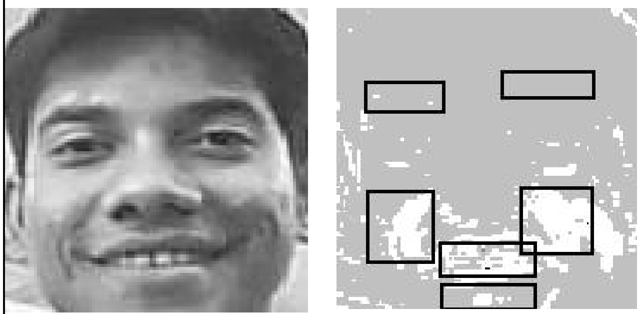
The top left and bottom right corners of each rectangular region, with respect to the eyes, are specified below in Table 1.

Motion Estimation in Regions

In the proposed method, the amount of facial motion is estimated by the amount of variation in the intensity values. The intensity variations are subjected to a threshold function reducing them to either black or white values. The number of white pixels in each region, corresponding to the amount of area that incurred intensity change, is calculated. We introduce a term "Region population" which refers to the percentage of white pixels in a given region. Figure 4 shows the Regions on a face with happy expressions, along with the population of each region for that expression.

Table. 1. Specification of the Coordinates of each Region

Region	Left Abscissa	Right Abscissa	Top Mantissa	Bottom Mantissa
1	$E_{RX} - D/3$	$E_{RX} + D/3$	$E_{RY} - D/2$	$E_{RY} - D/3$
2	$E_{LX} - D/3$	$E_{LX} + D/3$	$E_{LY} - D/2$	$E_{LY} - D/3$
3	$E_{RX} - D/4$	$E_{RX} + D/6$	$E_{RY} + D/2$	$E_{RY} + D/2 + 2D/5$
4	$E_{LX} - D/6$	$E_{LX} + D/4$	$E_{LY} + D/2$	$E_{LY} + D/2 + 2D/5$
5	$E_{RX} + D/8$	$E_{LX} - D/8$	$E_{RY} + 2D/3$	$E_{LY} + D$
6	$E_{RX} + D/8$	$E_{LX} - D/8$	$E_{RY} + 6D/5$	$E_{LY} + 7D/5$

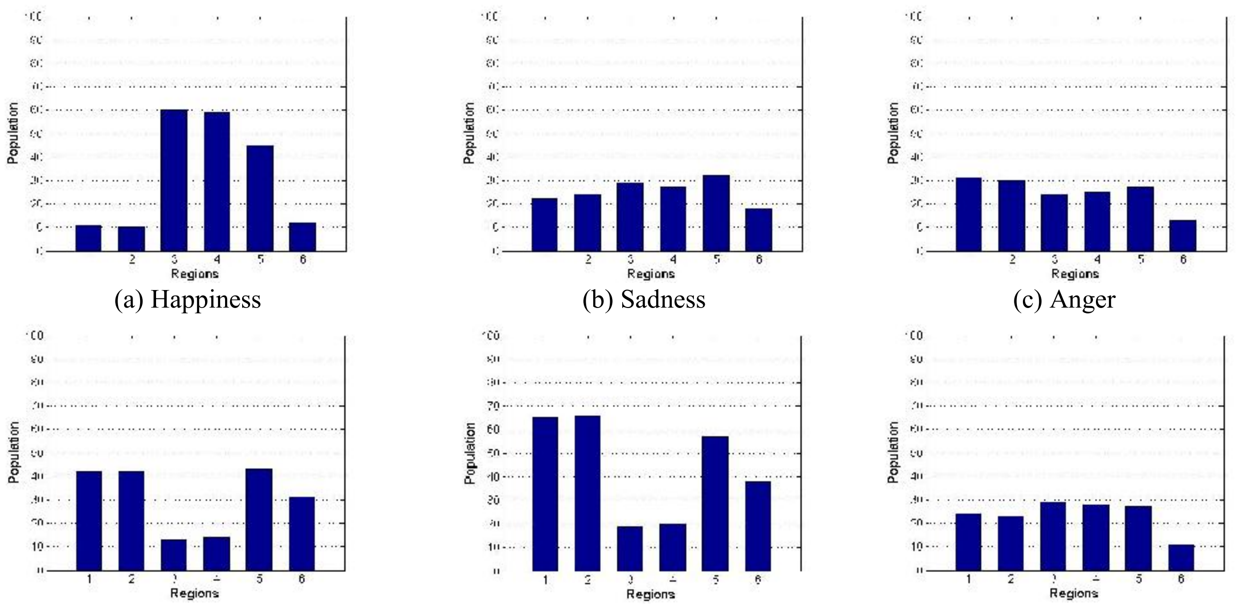


Region 1	4
Region 2	0
Region 3	55
Region 4	81
Region 5	58
Region 6	21

(a) (b) (c)

Fig. 4. (a) A typical Happy Expression (b) Bi-level Image Showing the Six Regions (c) Population of each Region

Different facial expressions were taken from twenty seven volunteers using camera, and the region populations were recorded at the peak of each expression. The numbers varied quite significantly from person to person, mainly due to the manner and severity of expression. The average region population for all the expressions found in the experiment are shown in Figure 5.



(a) Happiness (b) Sadness (c) Anger

(d) Fear (e) Surprise (f) Disgust

Fig. 5. (a-f): Average Values in each Region for the Six Universal Expressions

Classification of Expressions based on Region Population

Through observation of the various regions, a rule-based expression classifier is developed. The classifier takes into account the population of each region in order to identify the associated expression. A particular expression may show the properties of different types of expressions. For example: a Surprise expression is manifested by eyebrow raise and jaw drop, which can also be related to Fear. Therefore, the probability of a certain expression of belonging to a category is identified, depending on the severity of the expression shown. At first, scores for each expression are calculated depending on the region population according to the rules of Table 2.

Table. 2. Rules for Expression Classification

Region	Population	Score					
		Happiness	Sadness	Anger	Fear	Surprise	Disgust
1 or 2	> 60					3	
	40 – 60			2	1	2	
	20 – 40		1	1	2		2
3 or 4	> 70	5					
	50 – 70	4					
	30 – 50	3	1	1			2
5	> 40				1	2	
	30 – 40		2	1	2	1	1
	20 – 30		1	2	2	1	2
6	> 30				1	3	
	20 – 30		1		2	2	

The scores of each expression are added to find the total score. The expression with the highest score is output as the recognized expression followed by a percentage of probability calculated using the equation below:

Percentage of Probability = Score of recognized expression / Total score * 100%

IV. Results

The proposed method identifies the expression of the person in front of the camera, and classifies the expression as belonging to one of seven categories: the six Universal Expressions or ‘No Expression’. Results of some test cases are shown in Figures 6 and 7. The performance of the proposed method is shown in Confusion Matrix form of Table 3. Rows in the matrix show the actual expression performed, the number in bracket indicates the number of expressions taken. Columns in the matrix show the results identified by the method. In each case, the expression with the highest probability is considered.

Fig. 6. Snapshot of the System Recognizing Expression of Anger

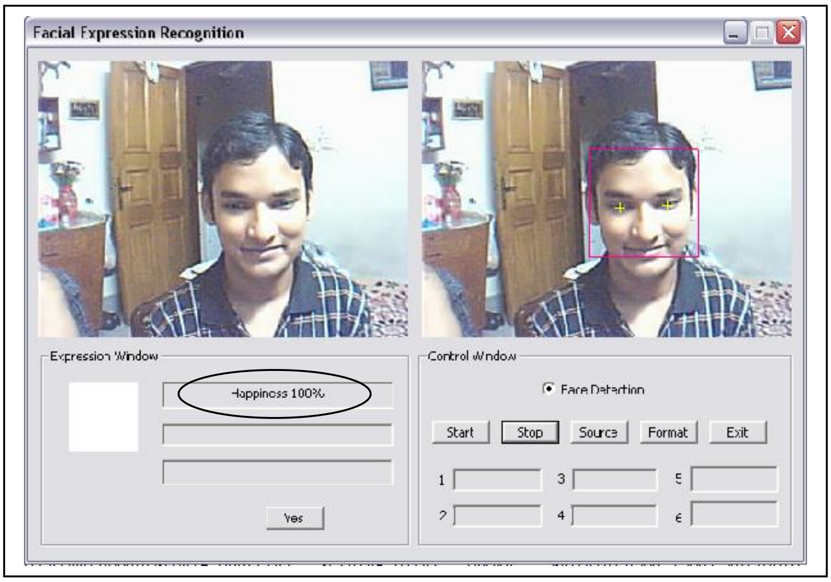


Fig. 7. Snapshot of the System Recognizing Expression of Happiness

Table. 3. Confusion Matrix for all participants

	Happiness	Sadness	Anger	Fear	Surprise	Disgust	No Expression	Success Rate
Happiness (87)	**87**	0	0	0	0	0	0	**100%**
Sadness (59)	0	**32**	4	8	1	7	7	**54.2%**
Anger (72)	1	2	**49**	5	1	13	1	**68.1%**
Fear (65)	0	1	6	**46**	8	2	2	**70.8%**
Surprise (83)	1	0	2	8	**72**	0	0	**86.7%**
Disgust (75)	1	3	9	2	0	**58**	2	**77.3%**

The proposed method works at an overall accuracy of 76%. The recognition of Happiness is the highest since this expression has a very distinguishable effect on Regions 3 and 4. Sadness is poorly recognized since there is no particular region of the face that is uniquely characterized by a sad expression. Sadness is often confused with other expressions and sometimes shown as ‘No expression’. This indicates that the expressions of Sadness by participants was less animated than other expressions. Anger, Disgust and Fear are often confused with one another. The reason is that all three expressions involve change in Regions 1, 2 and 5 which are regions for the eyebrows and the lips. Surprise is well-recognized though is sometimes confused as Fear. This happens when the person is reluctant to show an action of surprise.

V. Conclusions

This paper describes the method of recognizing human facial expressions in real time using region-based motion estimation. The proposed technique takes video input containing a human face via camera, and outputs the expression that is conveyed in the person’s face. To arrive at the desired output, several steps are taken. The face region is extracted from the input video frame, normalized to a defined resolution, and compared with a previously stored face image for comparison. The difference between the two images provides clues to the areas where facial motion occurs. The amount of motion in each region is determined, and fed to a rule-based classifier that identifies the given expression.

The proposed algorithm is tested upon twenty seven participants who performed six different facial expressions: happiness, sadness, anger, fear, surprise and disgust. The results show that happiness was the easiest expression to detect since 100% of smiling faces were correctly recognized. This is due to the fact that a smile affects two regions of the face that no other expression has any significant effect on. The Surprise expression, when performed in a demonstrative manner, is also satisfactorily detected. But a more subtle expression of surprise is often confused with fear. Fear is also confused with disgust since these expressions incur similar type of intensity change on the facial image. Anger is well-recognized when expressed by raising eyebrows, but confused with Fear and Disgust when expressed with contracted eyebrows. Sadness is found to be difficult to recognize, largely because the participants provided restrained expressions. The overall accuracy of the proposed method is 76%, which is satisfactory considering the simplicity of the algorithm and the real time nature of the application.

For expression recognition, the proposed method considers the conditions of only 10 to 15 percent of the total number of pixels in the input face image. Therefore, the response time is very low. Another advantage is that it works with simple 2D camera with no external equipments, tools, or markers to assist expression recognition, resulting in high usability and portability.

In the future, the proposed method can be expanded to adjust itself to light conditions of a wider range. Instead of being fixed values, the threshold values can vary according to the light. This study focuses on the position and quantity of facial motion, but overlooks the direction of motion. By incorporating the direction of motion, the algorithm can be improved to a great extent since the confusion between some expressions can be eradicated. Furthermore, methods of differentiating facial motion from rigid head movement could be implemented, allowing for greater flexibility.

----

1. Ekman P., "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique,"*Psychological Bulletin* 115(2), pp. 268–287, 1994.
2. Ekman P. and W. Friesen, *Facial Action Coding System: Investigator's Guide*, Consulting Psychologists Press, 1978.
3. Mase K., "Recognition of facial expression from optical flow," *IEICE Trans.* E74(10), 3474–3483, 1991.
4. Lanitis A., C. Taylor, and T. Cootes, "A unified approach to coding and interpreting face images," in Proc. International Conf. on Computer Vision, 368–373, 1995.
5. Black M. and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in Proc. International Conf. on Computer Vision, pp. 374–381, 1995.
6. Yacoob Y. and L. Davis, "Recognizing human facial expressions from long image sequences using optical flow," IEEE Trans. on Pattern Analysis and Machine Intelligence 18(6), pp. 636–642, 1996.
7. Rosenblum M., Y. Yacoob, and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," IEEE Trans. on Neural Network **7**(5), pp. 1121–1138, 1996.
8. Essa I. and A. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," IEEE Trans. on Pattern Analysis and Machine Intelligence **19(7)**, 757–763, 1997.
9. Cohen I., N. Sebe, A. Garg, L. Chen, and T. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," Computer Vision and Image Understanding 91(1-2), 160–187, 2003.
10. Sebe N, MS Lew, Y Sun, I Cohen, T Gevers, TS Huang, "Authentic facial expression analysis" - Image and Vision Computing, Elsevier, 2007.
11. Pantic M, I Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences", IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2006.
12. FaceVACS-SDK, Version 1.9.0, Cognitec Systems [Online], URL: *http://www.cognitec-systems.de*. Retrieved on 12 February, 2008.