Emotion Recognition with Image Processing and Neural Networks

W.N. Widanagamaachchi & A.T. Dharmaratne
University of Colombo School of Computing, Sri Lanka
wathsy31@gmail.com, atd@ucsc.cmb.ac.lk

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 prototype 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 prototype 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.

1.0 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 on 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 [10]. In 1969, after recognizing a universality among emotions in different groups of people despite the cultural differences, Ekman and Friesen classified six emotional expressions to be universal: happiness, sadness, disgust, surprise and fear [6, 10, 9, 3]. (Figure 1)

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 [18]. 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 [4, 25]. In fact, there exist 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 and 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 [16].

The rest of the paper is as follows. Section 2 is devoted for the related work of the area while section 3 provides a detailed explanation of the prototype system. Finally the paper concludes with the conclusion in section 4.

2.0 RELATED WORK

The recent work relevant to the study can be broadly categorized in to three: Face detection, Facial feature extraction and Emotion classification. The number of research carried out in each of these categories is quite sizeable and noteworthy.

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 human face might have within the image 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, such as beard, hair and makeup have a considerable effect in the facial appearance as well [19, 20]. The approaches of the past few decades in face detection can be classified into four: knowledge-based approach, feature invariant approach, template-based approach and appearance-based approach [12, 13, 19,20].

Knowledge-based approaches are based on rules derived from the knowledge on the face geometry. The most common approach of defining the rules is based on 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 [19, 12, 13]. In feature invariant approaches, facial features are detected and then grouped according to the geometry of the face. Selecting a set of appropriate features is very crucial in this approach [11].

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 [19, 12, 13, 2]. Appearance-based approach considers the human face in terms of a pattern with pixel intensities.

2.2   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. The luminance, chrominance, facial geometry and symmetry based approaches, template based approaches, 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 [11, 23].

When using geometry and symmetry for the extraction of features, the human visual characteristics of the face are employed. The symmetry contained within the face is helpful in detecting the facial features irrespective of the differences in shape, size and structure of features.

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. Thus, the intensity histogram has being extensively used in extracting features based on luminance and chrominance [10, 1, 6]. When extracting features based on template based approaches, a separate template will be specified for each feature based on its shapes.

2.2   Feature Extraction

The research carried out by Ekman on emotions and facial expressions is the main reason behind the attraction for the topic of emotion classification. Over the past few decades various approaches have being introduced for the classification of emotions. They differ only in the features extracted from the images and in the classification used to distinguish between the emotions [17, 7].

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 initial features extracted from the initial frame have to be mapped between each of the frames. A sufficient rate for frame processing has to be maintained as well.

3.0   EMOTION RECOGNITION

The prototype system for emotion recognition is divided into 3 stages: face detection, 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 were used to process the enhanced face region to determine the feature locations. These feature areas were 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 was trained to recognize the 6 universal emotions.

3.1 Face Detection

The prototype system offers two methods for face detection. Though various knowledge based and template based techniques can be developed for face location determination, we opted for a feature invariant approach based on skin color 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.

After experimenting with a set of face images, the following condition 1 was developed based on which faces were detected.

$$(H < 0.1) \text{ OR } (H > 0.9) \text{ AND } (S > 0.75) \quad (1)$$

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 2)

The second method is the implementation of the face detection approach by Nilsson and others' [12]. Their approach uses local SMQT features and split up SNoW classifier to detect the face. This classifier results more accurate face detection than the hue and saturation based classifier mentioned earlier. Moreover within the prototype system, the user also has the ability to specify the face region with the use of the mouse.

3.2 Feature Extraction

In the feature extraction stage, the face detected in the previous stage is further processed to identify eye, eyebrows and mouth regions. Initially, the likely Y coordinates of the eyes was identified with the use of the horizontal projection. Then the areas around the y coordinates were processed to identify the exact regions of the features. Finally, a corner point detection algorithm was used to obtain the required corner points from the feature regions.

3.2.1 Eye Extraction

The eyes display strong vertical edges (horizontal transitions) due to its iris and eye white. Thus, the Sobel mask in Figure 3(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 our experiments with the images, we observed that the use of one Sobel mask alone is not enough to accurately identify the Y coordinate of the eyes.

Hence, the Sobel edge detection is applied to the upper half of the face image and the sum of each row is horizontally plotted. 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 2 and also satisfy certain geometric conditions. G in the condition is the green color component in the RGB color space. Figure 4(a) illustrates those found areas.

$$G < 60 \quad (2)$$

Furthermore the edge image of the area round the Y coordinate is obtained by using the Roberts method. Then it is dilated and holes are filled to obtain the figure 4(b). Then the regions in the figure 4(a) grown till the edges in the figure 4(b) are included to result the figure 4(c). Then pair of regions that satisfy certain geometric conditions are selected as eyes from those regions.

3.2.2 Eyebrow Extraction

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

3.2.3 Mouth Extraction

Since the eye regions are known, the image region

below the eyes is processed to find the regions which satisfy the condition 3.

$$1.2 \leq R/G \leq 1.5 \qquad (3)$$

Furthermore, the edge image of this image region is also obtained by using the Roberts method. Then it is dilated and holes are filled to obtain the figure 5(b). From the result regions of condition 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 5(b) are included.

After feature regions are identified, these regions are further processed to extract the necessary feature points. The harris corner point detection algorithm was 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 4(b) to obtain the top and bottom corner points. Finally after obtaining the top, bottom, right most and left most points, the centroid 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 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 section 3.2.2.

3.3 Emotion classification

The extracted feature points are processed to obtain the inputs for the neural network. The neural network has being trained so that the emotions happiness, sadness, anger, disgust, surprise and fear are recognized. 525 images from Facial expressions and emotion database [19] are taken to train the network. However, we are unable to present the results of classifications since the network is still being tested. Moreover we hope to classify emotions with the use of the naïve bias classifier as an evaluation step. The inputs to the neural network are

as follows.
- Eye height
= (Left eye height + Right eye height) / 2
= [$(c_4 - c_3) + (d_4 - d_3)$] / 2
- Eye width
= (Left eye width + Right eye width) / 2
= [$(c_2 - c_1) + (d_1 - d_2)$] / 2
- Mouth height = $(f_4 - f_3)$
- Mouth width = $(f_2 - f_1)$
- Eyebrow to Eye center height
= [$(c_5 - a_1) + (d_5 - b_1)$] / 2
- Eye center to Mouth center height
= [$(f_5 - c_5) + (f_5 - d_5)$] / 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

4.0 CONCLUSION

In this paper, we explored a novel way of classifying human emotions from facial expressions. Thus a neural network based solution combined with image processing was proposed to classify the six universal emotions: Happiness, Sadness, Anger, Disgust, Surprise and Fear. Initially a face detection step is performed on the input image. Afterwards an image processing based feature point extraction method is used to extract the feature points. Finally, a set of values obtained from processing the extracted feature points are given as input to a neural network to recognize the emotion contained.

The project is still continuing and is expected to produce successful outcomes in the area of emotion recognition. We expect to make the system and the source code available for free. In addition to that, it is our intention to extend the system to recognize emotions in video sequences. However, the results of classifications and the evaluation of the system are eft out from the paper since the results are still being tested. We hope to make the results announced with enough time so that the final system will be available for this paper's intended audience.

5.0 REFERENCES

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

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

3) Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S., 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.

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

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

6) Deng, X., Chang, C.-H., and Brandle, E., 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

7) Dumasm, M., Emotional expression recognition using support vector machines. Technical report,Machine Perception Lab, Univeristy of California, 2001.

8) Ekman, P., Facial expression and emotion. American Psychologist, 48:384–392, 1993.

9) Grimm, M., Dastidar, D. G., and Kroschel, K., Recognizing emotions in spontaneous facial expressions. 2008.

10) Gu, H., Su, G., and Du, C., Feature points extraction from faces. 2003.

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

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

13) Pantic, M., Rothkrantz, L. J. M., and Koppelaar, H., Automation of nonverbal communication of facial expressions. In: EUROMEDIA 98, SCS International, pages 86–93, 1998.

14) Pham, T., and Worring, M., Face detection methods: A critical evaluation. ISIS Technical Report Series, University of Amsterdam, 11, 2000.

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

16) Sanderson C., and Paliwal, K. K., Fast feature extraction method for robust face verification. Electronics Letters, 8:1648 – 1650, 2002.

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

18) Te,o W. K., Silva, L. C. D., and Vadakkepat, P., Facial expression detection and recognition system. 2008.

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

20) Yang, M. H., Ahuja, N., and Kriegman, D., A survey on face detection methods, 1999.

21) Yang, M. H., D. Kriegman, J., Member, S., and Ahuja, N., Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:34–58, 2002.
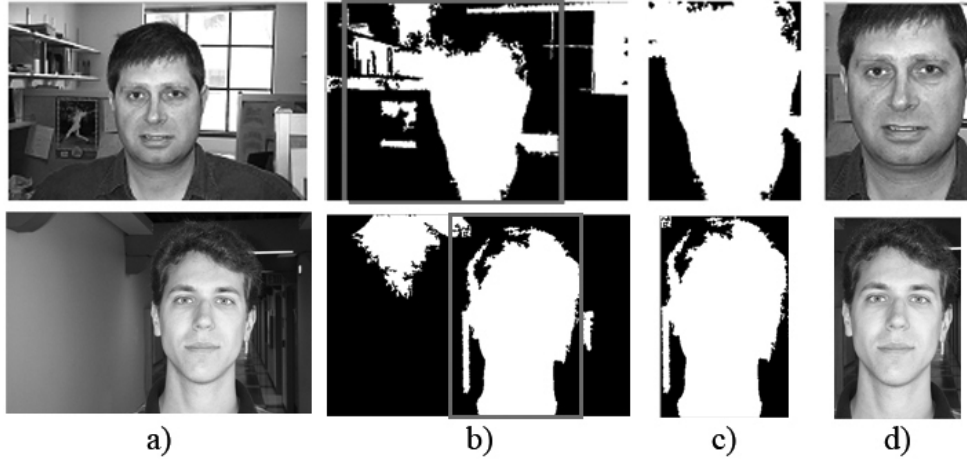
Fig 1 : The six universal emotional expressions [26]

Fig 2 : Steps in face detection stage.
a) original image b) result image from condition 1 c) region after refining d) face detected

Fig 3 : Sobel operator
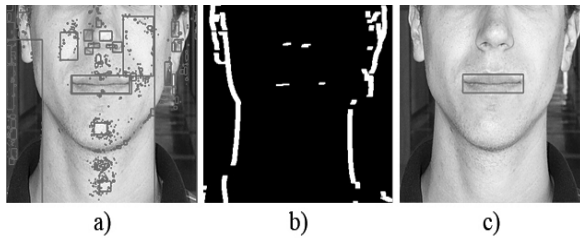a) detect vertical edges b) detect horizontal edges

Fig 5 : Steps in mouth extraction
a) result image from condition 3 b) after dilating and
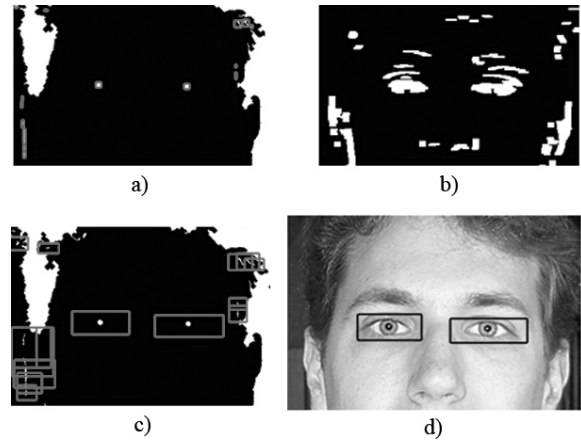filling the edge image c) selected mouth region

Fig 4 : Steps in eye extraction
a) regions around the y coordinate of eyes b) after
dilating and filling the edge image c) after growing
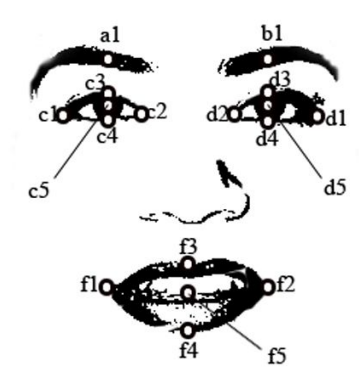the regions d) selected eye regions e) corner point
detection

Fig 6 : Extracted feature points