ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

IJMECE International Journal of modern electronics and communication engineering

ISSN2321-2152
www.ijmece .com
Vol 5, Issue.3 Sep 2017

Real Time Automatic Facial Expressions Recognition Using Artificial Neural Networks

Tata Jagannadha Swamy*

Abstract *The most expressive way humans display emotions is through facial expressions. Enabling computer systems to recognize facial expressions and infer emotions from them in real time presents a challenging research topic. Most facial expression analysis systems attempt to recognize facial expressions from data collected in a highly controlled laboratory with very high resolution frontal faces (face regions greater than 200 x 200 pixels) and also cannot handle large head motions. But in real environments such as smart meetings, a facial expression analysis system must be able to automatically recognize expressions at lower resolution and handle the full range of head motion. This paper describes a real-time system to automatically recognize facial expressions in relatively low-resolution of face images (around 50x70 to 75x100 pixels). This system proves to be successful in dealing with complex real world interactions and is a highly promising method.*

Key Words: Facial recognition, Infer emotions, facial expressions, data collection, Complex interactions.

1. INTRODUCTION:

The human face possesses superior expressive ability and provides one of the most powerful, versatile and natural means of communicating motivational and affective state. We use facial expressions not only to express our emotions, but also to provide important communicative cues during social interaction, such as our level of interest, our desire to take a speaking turn and continuous feedback signaling understanding of the information conveyed. Facial expression constitutes 55 percent of the effect of a communicated message and is hence a major modality in human communication.Face Recognition and face expression recognition is the inherent capability of human beings. Identifying a person by face is one of the most fundamental human functions since time immemorial. Face expression recognition by computer is to endow a machine with capability to approximate in some sense, a similar capability in human beings. To impart this basic human capability to a machine has been a subject of interest over the last few years. Before we discuss how facial expressions can be recognized we will let u know what are the main problems that a face recognition system.

Professor, Department of Electronics and Communication Engineering
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad
Jagan.tata@griet.ac.in

ISSN2321-2152
www.ijmece .com
Vol 5, Issue.3 Sep 2017

In this paper we describe a new facial expression analysis system designed to automatically recognize facial expressions in real-time and real environments, using relatively low-resolution face images. Figure. 2 shows the structure of the tracking system. T. Kanade et.al, Z. Zhang, Y. Moses et.al and B. Fasel et.al discussed the input video sequence is used to estimate a background model, which is then used to perform background subtraction, as described in Section 3. In Section 4, the resulting foreground regions are used to detect the head. After finding the head, head pose estimation is performed to find the head in frontal or near-frontal views. The facial features are extracted only for those faces in which both eyes and mouth corners are visible. The normalized facial features are input to a neural network based expression classifier to recognize different expressions.

2. SYSTEM MODEL

The problem of recognizing and interpreting faces comprises four main sub problem areas:

- Finding faces and facial features: This problem would be considered a segmentation problem in the machine vision literature, and a detection problem in the pattern recognition literature.
- Recognizing faces and facial features: This problem requires defining a similarity metric that allows comparison between examples; this is the fundamental operation in database access.
- Tracking faces and facial features: Because facial motion is very fast (with respect to either human or biological vision systems), the techniques of optimal estimation and control are required to obtain robust performance.
- Temporal interpretation. The problem of interpretation is often too difficult to solve from a single frame and requires temporal context for its solution. Similar problems of interpretation are found in speech processing, and it is likely that speech methods such as hidden Markov models, discrete Kalman filters, and dynamic time warping will prove useful in the facial domain as well.

Current approaches to automated facial expression analysis typically attempt to recognize a small set of prototypic emotional expressions, i.e. joy, surprise, anger, sadness, fear, and disgust. Some group of researchers presented an early attempt to analyze facial expressions by tracking the motion of twenty identified spots on an image sequence. Some developed a dynamic parametric model based on a 3D geometric mesh face model to recognize 5 prototypic expressions. Some selected manually selected facial regions that corresponded to facial muscles and computed motion within these regions using optical flow. T. Kanade et.al, Z and B. Fasel et.al Some other group of researchers used optical flow work, but tracked the motion of the surface regions of facial features (brows, eyes, nose, and mouth) instead of the motion of the underlying muscle groups .

Limitations of Existing Systems:

The limitations of the existing systems are summarized as following:

- Most systems attempt to recognize facial expressions from data collected in a highly controlled laboratory with very high-resolution frontal faces (face regions greater than 200 x 200 pixels).
- Most system needs some manual preprocessing.
- Most systems cannot handle large out-of-plane head motion.
- None of these systems deals with complex real world interactions.
- Except the system proposed by Moses *et al.* [14], none of those systems performs in real-time.

In this paper, we report an expression recognition system, which addresses many of these limitations. In real environments, a facial expression analysis system must be able to:

- Fully automatically recognize expressions.
- Handle a full range of head motions.
- Recognize expressions in face images with relatively lower resolution.
- Recognize expressions in lower intensity.
- Perform in real-time.

Face Process	96x128	69x93	48x64	24x32
Detect?	Yes	Yes	Yes	Yes
Pose?	Yes	Yes	Yes	Yes
Recognize?	Yes	Yes	Yes	Maybe
Features?	Yes	Yes	Maybe	No
Expressions?	Yes	Yes	Maybe	No

Figure 1. A face at different resolutions.

Figure.1 shows a face at different resolutions. Most automated face processing tasks should be possible for a 69x93 pixel image. At 48x64 pixels the facial features such as the corners of the eyes and the mouth become hard to detect. The facial expressions may be recognized at 48x64 and are not recognized at 24x32 pixels. This paper describes a real-time system to automatically recognize facial expressions in relatively low-resolution face images (50x70 to 75x100 pixels). To handle the full range of head motion, instead of detecting the face, the head pose is estimated based on the detected head. For frontal and near frontal views of the face, the location and shape features are computed for expression recognition. This system successfully deals with complex real world interactions. We present the overall architecture of the system and its components: background subtraction, head detection and head pose estimation respectively. The method for facial feature extraction and tracking is also explained clearly. The method for recognizing expressions is reported and at we summarized our paper and presented future directions in the last part of our paper.

3. SEVERAL VERIFICATIONAL ASPECTS

BACKGROUND ESTIMATION AND SUBTRACTION:

The background subtraction approach presented is an attempt to make the background subtraction robust to illumination changes. The background is modeled statistically at each pixel. The estimation process computes the brightness distortion and color distortion in RGB color space. It also proposes an active background estimation method that can deal with moving objects in the frame. First, we calculate image difference over three frames to detect the moving objects. Then the statistical background model is constructed, excluding these moving object regions. By comparing the difference between the background image and the current image, a given pixel is classified into one of four categories: original background, shaded background or shadow, highlighted background, and foreground objects. Finally, a morphology step is applied to remove small isolated spots and fill holes in the foreground image.

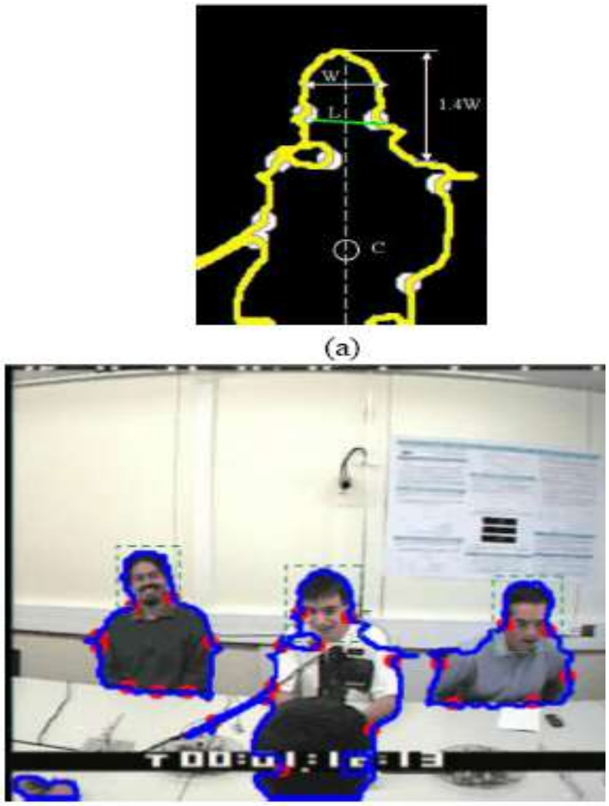
HEAD DETECTION:

In order to handle the full range of head motion, we detect the head instead of detecting the face. The head detection uses the smoothed silhouette of the foreground object as segmented using background subtraction. Based on human intuition about the parts of an object, a segmentation into parts generally occurs at the *negative curvature minima* (NCM) points of the silhouette as shown with small circles in Figure 3. The boundaries between parts are called cuts (shown as the line L in Figure 3(a). some researchers noted that human vision prefers the partitioning scheme which uses the shortest cuts. They proposed a shortcut rule which requires a cut: 1) be a straight line, 2) cross an axis of local symmetry, 3) join two points on the outline of a silhouette and at least one of the two points is NCM, 4) be the shortest one if there are several possible competing cuts.

In this system, the following steps are used to calculate the cut of the head:

- Calculate the contour centroid C and the vertically symmetry axis y of the silhouette.
- Compute the cuts for the NCMs, which are located above the contour centroid C.
- Measure the salience of a part's protrusion, which is defined as the ratio of the perimeter of the Part (excluding the cut) to the length of the cut.
- Test if the salience of a part exceeds a low threshold.
- Test if the cut crosses the vertical symmetry axis y of the silhouette.
- Select the top one as the cut of the head if there are several possible competing cuts.

After the cut of the head L is detected, the head region can be easily determined as the part above the cut. As shown in Figure 3(b), in most situations, only part of the head lies above the cut. To obtain the correct head region, we first calculate the head width W, then the head height H is enlarged to $\alpha * W$ from the top of the head. In our system, α = 1:4.



(a)

(b)

Figure 3. Head detection steps. (a) Calculate The cut of the head part. (b) Obtain the correct. Head region from the cut of the head part.

HEAD POSE DETECTION:

After the head is located, the head image is converted to gray-scale, histogram equalized and resized to the estimated resolution. Then we employ a three layer neural network (NN) to estimate the head pose. The inputs to the network are the processed head image by T. Kanade et.al, Z. Zhang, Y. Moses et.al and B. Fasel et.al. The outputs are the head poses. Here only 3 head pose classes are trained for expression analysis: 1) frontal or near frontal view, 2) side view or profile, 3) others such as back of the head or occluded face.

Poses	Frontal near frontal	Side view profile	Others
Definitions	Both eyes and lip corners are visible	One eye or one lip corner is occluded	Not enough facial features
Examples			

Figure 4. The definitions and examples of the 3 head pose classes: 1) frontal or near frontal View, 2) side view or profile, 3) others such as Back of the head or occluded faces.

Figure 4 shows the definitions and some examples of the 3 head pose classes. In the frontal or near frontal view, both eyes and lip corners are visible. Inside view or profile, at least one eye or one corner of the mouth becomes self-occluded because of the head turn. All other reasons cause more facial features to not be detected such as the back of the head, occluded face, and face with extreme tilt angles is treated as one class.

FACIAL FEATURE EXTRACTION FOR FRONTAL OR NEAR-FRONTAL FACE:

After estimating the head pose, the facial features are extracted only for the face in the frontal or near-frontal view. Since the face images are in relatively low resolution in most real environments, the detailed facial features such as the corners of the eyes and the upper or lower eyelids are not available to recognize facial expressions, however, we need to detect reliable facial features. We observe that most facial feature changes that are caused by an expression are in the areas of eyes, brows and mouth by T. Kanade et.al, Z. Zhang, Y. Moses et.al and B. Fasel et.al. In this paper, two types of facial features in these areas are extracted: location features and shape features .

LOCATION FEATURE EXTRACTION

In this system, six location features are extracted for expression analysis. They are eye centers (2), eyebrow inner endpoints (2), and corners of the mouth (2).

Eye centers and eyebrow inner endpoints: To find the eye centers and eyebrow inner endpoints inside the detected frontal or near frontal face, we have developed an algorithm that searches for two pairs of dark regions which correspond to the eyes and the brows by using certain geometric constraints such as position inside the face, size and symmetry to the facial symmetry axis. The algorithm employs an iterative threshold method to find these dark regions under different or changing lighting conditions.

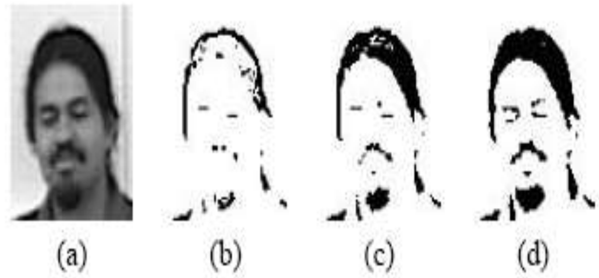


Figure 5. Iterative thresholding of the face to Find eyes and brows. (a) Gray-scale face image,
(b) Threshold = 30, (c) threshold = 42, (d) Threshold = 54.

Figure.5 shows the iterative thresholding method to find eyes and brows. Generally, after five iterations, all the eyes and brows are found. If satisfactory results are not found after 20 iterations, we think the eyes or the brows are occluded or the face is not in a near frontal view. Unlike to find one pair of dark regions for the eyes only, we find two pairs of parallel dark regions for both the eyes and eyebrows. By doing this, not only are more features obtained, but also the accuracy of the extracted features is improved. Then the eye centers and eyebrow inner endpoints can be easily determined. If the face image is continually in the frontal or near frontal view in an image sequence, the eyes and brows can be tracked by simply searching for the dark pixels around their positions in the last frame.

Mouth corners: After finding the positions of the eyes, the location of the mouth is first predicted. Then the vertical position of the line between the lips is found using an integral projection of the mouth region Finally, the horizontal borders of the line between the lips is found using an integral projection over an edge-image of the mouth. The following steps are use to track the corners of the mouth: 1) Find two points on the line between the lips near the previous positions of the corners in the image 2) Search along the darkest path to the left and right, until the corners are found. Finding the points on the line between the lips can be done by searching for the darkest pixels in search windows near the previous mouth corner positions. Because there is a strong change from dark to bright at the location of the corners, the corners can be found by looking for the maximum contrast along the search path.

4. RESULTS AND DISCUSSIONS:

LOCATION FEATURE REPRESENTATION:

After extracting the location features, the face can be normalized to a canonical face size based on two of these features, i.e., the eye-separation after the line connecting two eyes (eye-line) is rotated to horizontal. In our system, all faces are normalized to 90 *x* 90 pixels by re-sampling. We transform the extracted features into a set of parameters for expression recognition. We represent the face location features by 5 parameters, which are shown in Figure 6.

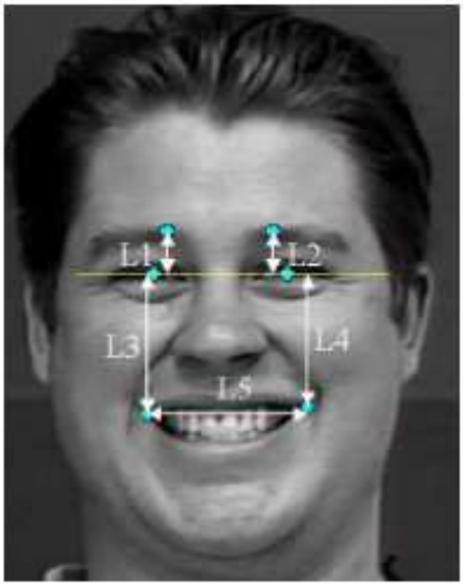


Figure 6. Face location feature representation for expression recognition.

These parameters are the distances between the eye-line and the corners of the mouth, the distances between the eye-line and the inner eyebrows, and the width of the mouth (the distance between two corners of the mouth).

SHAPE FEATURE EXTRACTION:

Another type of distinguishing feature is the shape of the mouth. Global shape features are not adequate to describe the shape of the mouth. Therefore, in order to extract the mouth shape features, an edge detector is applied to the normalized face to get an edge map. This edge map is divided into 3 *x* 3 zones as shown in Figure.7 (b). The size of the zones is selected to be half of the distance between the eyes. The mouth shape features are computed from zonal shape histograms of the edges in the mouth region.

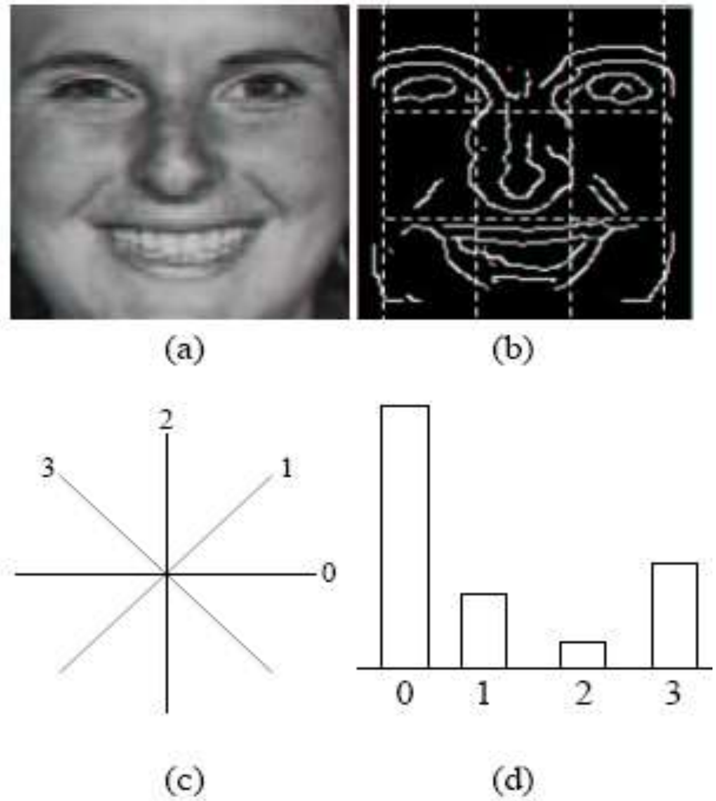


Figure 7. Zonal-histogram features. (a) Normalized face, (b) Zones of the edge map of the normalized face, (c) Four quantization levels for calculating histogram features, (d) Histogram corresponding to the middle zone of the mouth.

This system has neural network-based recognizer having the structure as shown in Figure 8. The standard back-propagation in the form of a three-layer neural network with one hidden layer was used to recognize facial expressions. The inputs to the network were the 5 location features (Figure 6) and the 12 zone components of shape features of the mouth (Figure7). Hence, a total of 17 features were used to represent the amount of expression in a face image. The outputs were a set of expressions. In this system, 5 expressions were recognized. They were neutral, smile, angry, surprise, and others (including fear, sad, and disgust). Researchers tested various numbers of hidden units and found that 6 hidden units gave the best performance.

ISSN2321-2152
www.ijmece .com
Vol 5, Issue.3 Sep 2017

5. CONCLUSION:

Automatically recognizing facial expressions is important to understand human emotion and paralinguistic communication so as to design multi modal user interfaces, and for related applications such as human identification. Incorporating emotive information in computer-human interfaces will allow for much more natural and efficient interaction paradigms to be established. It is very challenging to develop a system that can perform in real time and in real world because of low image resolution, low expression intensity, and the full range of head motion and the system that we have reported is an automatic expression recognition system that addresses all the above challenges and successfully deals with complex real world interactions. In most real word interactions, the facial feature changes are caused by both talking and expression changes. We feel that further efforts will be required for distinguishing talking and expression changes by fusing audio signal processing and visual image analysis. Also it will benefit the expression recognition accuracy by using the sequential information instead of using each frame.

REFERENCES:

T. Kanade, J.F. Cohn, and Y.L. Tian. Comprehensive database for facial expression analysis. In Proceedings of International Conference on Face and Gesture Recognition, pages 46–53, March 2000.

Z. Zhang. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron. International Journal of Pattern Recognition and Artificial Intelligence, 13(6):893–911, 1999.

Y. Moses, D. Reynard, and A. Blake. Determining facial expressions in real time. In Proc. Of Int. Conf.On Automatic Face and Gesture Recognition, pages 332–337, 1995.

B. Fasel and J. Luttin. Recognition of asymmetric facial action unit activities and intensities. In Proceedings of International Conference of Pattern Recognition, 2000.