International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

Emotion recognition based on contours of features

Gompanabilli Hema Latha & Prof. Kompella.Venkata Ramana
(M.Tech,) Department of CST, A.U.College of Engineering (A) Andhra University, Visakhapatnam
(B.E, M.E, Ph.D) Department of CST, A.U.College of Engineering (A) Andhra University, Visakhapatnam

Abstract

Facial gesture recognition is one of the important components of natural human-machine interfaces; it may also be used in behavioural science, security systems and inClinical practice. Automatic analysis of facial gestures is rapidly becoming an area of intense interest in computer science and human-computer interaction design communities. However, the basic goal of this area of research – translating detected facial changes into a human-like description of shown facial expression – is yet to be achieved. One of the main impediments to achieving this aim is the fact that human interpretations of a facial expression differ depending upon whether the observed person is speaking or not. A first step in tackling this problem is to achieve automatic detection of facial gestures that are typical for speech articulation. Although humans recognise facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge. The face expression recognition problem is challenging because different individuals display the same expression differently. This paper presents an overview of gesture recognition in real time using the concepts of correlations. Our approach to seizing this step in the research on automatic facial expression analysis. We consider the six universal emotional categories namely joy, anger, fear, disgust, sadness and surprise.The applications of gesture recognition are manifold, ranging from sign language through medicalrehabilitation to virtual reality. In this paper, various algorithms for gesture recognition have been investigated. Firststep in any gesture recognition process is face detection. We investigated algorithms like color segmentation,morphological Image Processing etc. for face detection, and Eigen faces for gesture recognition.

KEYWORDS

Object Recognition, Face Recognition, Image Sets, Canonical Correlation, Principal Angles, Gesture recognition; Cross correlation, Eigen Faces.

1. INTRODUCTION

Facial gestures (facial muscle actions) regulate our social interactions: they represent and they clarify whether our current focus of attention(e.g., a person or what has been said) is important, funny or unpleasant for us. They are direct, naturally preeminent means for humans to communicate their emotions [1, 2]. Automatic analyzers of subtle facial changes, therefore, seem to have a natural place in various vision systems including automated tools for psychological research, lip reading, bimodal analysis, affective computing, face and visual synthesis, and perceptual user interfaces. Thus, in recent years, there has been a tremendous interest in automating facial gesture analysis. Most approaches to automatic facial gesture analysis in face image sequences attempt to recognize a set of prototypic emotional facial expressions, i.e., happiness, sadness, fear, surprise, anger and disgust [3]. Yet, in everyday life such prototypic expressions occur rather infrequently; emotions are displayed more often by subtle changes in one or few discrete facial features, such as raising the eyebrows in surprise [1]. To detect such subtlety of human emotion, automatic recognition of facial gestures (i.e., fine-grained changes in facial expression) is needed.From several methods for recognition of facial gestures based on visually observable facial muscular activity, the FACS system [4] is the most commonly used in the psychological research. Following this trend, all of the existing methods for automatic facial gesture analysis, including the method proposed here, interpret the facial display information in terms of the facial action units (AUs) of the FACS system [3, 5]. Yet none automatic system is capable of encoding the full range of facial mimics, i.e., none is capable of recognizing all 44 AUs that account for the changes in facial display. From the previous works on automatic facial gesture recognition from face image sequences, the method presented in [6] performs the best in this aspect: it encodes 16 AUs occurring alone or in a combination in frontal-view face image sequences.However, even if an automatic detector of all possible facial muscle actions would be at hand, emotional interpretation of facial cues would remain by no meansa trivial task. This goal is made difficult by the rich shadings of affective/attitudinal states that humans recognize in a facial expression. Another major elementof difficulty is that a shown facial gesture may be easily misinterpreted. To date, however, automatic facial information analyzers do not perform usually user profiled interpretation of sensed data and virtually allapproaches to facial gesture analysis have largely avoided dealing with questions that involve whether the observed subject. The later is easy todo if one can limit the context. For example, if you know that except of the observed subject there is no other person in the area, then pursing the lips willprobably represent a facial signal of being bored or being in a mode of thinking and not a visible signal. But, as we move towards more generally competent perceptual user interfaces, which shouldfacilitate videoconferences, virtual visits to Internetsites, etc., we will have to directly confront the problemof distinguishing the facial gestures that are typical forspeech articulation from those attitude oraffect. Hence, both a reliable detector of whether theobserved subject is facial gestures which form the typical visiblespeech signals (to be treated as noise in affect-sensitiveanalysis of visual speech data) are needed for an (userprofiledor not) emotional interpretation of facial cues.Within our research on facial gesture analysis fromfrontal-view face image sequences, we investigated firstwhether and to which extent human facial gestures onset/offset can be recognized automatically.Hereafter, we investigated which facial gestures formtypical visualsignals. This paper presents thepreliminary results of our research. The devised methodfor rule-based recognition of 22 AUs from frontal-view

face image sequences is presented in section 2. Section3 gives an overview of a neural-network-based methodfor automatic determination of whether the observedsubject. Experimental evaluations ofthe two methods and an experimental study on facial

Muscle actions typical for speech articulation arepresented in section 4. Section 5 concludes the paper.

II. FACIALGESTURE RECOGNITION

The problem of automatic facial gesture recognitionfrom face image sequences is usually divided into threesub-problem areas (Fig. 1): detecting prominent facialfeatures such as eyes and mouth, representing subtlechanges in facial expression as a set of suitable midlevelfeature parameters, and interpreting these data in

Terms of facial gestures such as the AUs of the FACSsystem.

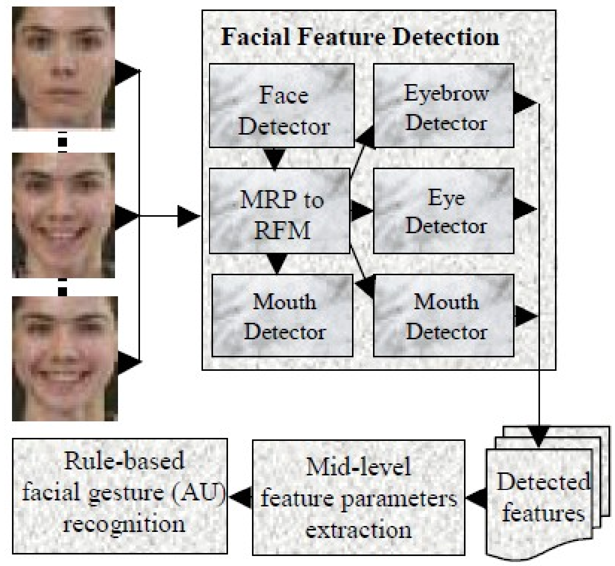


Fig. 1: Outline of the AU-recognition method

A. Facial Feature Detection

To reason about shown facial gestures, the face and its components (i.e., prominent facial features) should be detected first. In order to do so, we apply a multi-phase multi-detector processing of an input frontal-view face image sequence. The two phases of the proposed method for detection of prominent facial features are

coarse detection and fine detection.

In the first phase, we apply a HSV color-based segmentation of the face ("Face Detector" in Fig. 1). The face region is segmented from an input frame asthe largest connected image component with Hue, Saturation and Value within the range [5, 35], [0, 0.7] and [0.1, 0.9] respectively [7]. Then we use a simpleanalysis of image histograms ("MRP to RFM" in Fig. 1) to locate 7 regions of interest (ROI): two eyebrows, two eyes, nose, mouth and chin.

In the second phase, to spatially sample the contour of a certain permanent facial feature, we apply one or more facial-feature detectors to the pertinent ROI. Forexample, the contours of the eyes are localized in the ROIs of the eyes by using a single detector representing an adapted version of a hierarchical-perceptron featurelocation method [7]. On the other hand, the contour of the mouth is localized in the mouth ROI by applying both a 4-parameters deformable template and a methodthat fits three 2nd degree parabolas [8]. For further details about these and other detectors used to spatially sample the contours of the prominent facial features,readers are referred to [7, 8].

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

B. Parametric Feature Representation:

The contours of the facial features, generated by thefacial feature detection method (Fig. 1), are utilized forfurther analysis of shown facial gestures.First, we carry out feature points' extraction undertwo assumptions: (1) the face images are non-occludedand in frontal view, and (2) the first frame is in aneutral expression. We extract 22 fiducial points: 19are extracted as vertices or apices of the contours of thefacial features (Fig. 2), 2 represent the centers of theeyes (points X and Y), and 1 represents the the middlepoint between the nostrils (point C). We assign acertainty factor to each of the extracted points, basedon an "intra-solution consistency check". For example,the fiducial points of the right eye are assigned acertainty factor CF ∈ [0, 1] based upon the calculateddeviation of the actually detected inner corner Bcurrentfrom the pertinent point Bneutral localized in the firstframe of the input sequence. The functional form of thismapping is:CF = sigm(d(Bcurrent, Bneutral); 1, 4, 10)where d(p1, p2) is the block distance between points p1and p2 (i.e., maximal difference in x and y direction)while sigm(x; α, β, γ) is a Sigmoid function. The major

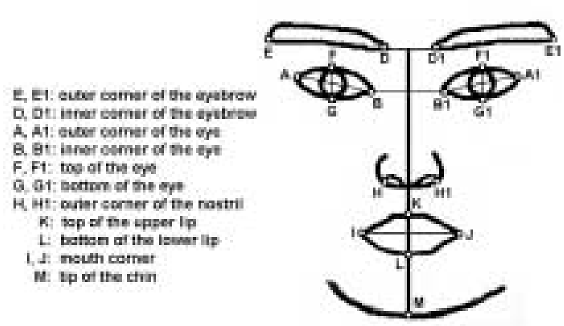


Fig. 2: Feature points (fiducials of the features' contours)

impulse for the usage of the inner corners of the eyes asthe referential points for calculating CFs of the fiducialpoints of the eyes comes from the stability of thesepoints with respect to non-rigid facial movements:

facial muscles' contractions do not cause physicaldisplacements of these points. For the same reason, thereferential features used for calculating CFs of thefiducial points of the eyebrows, nose/chin and mouthare the size of the relevant eyebrow area, the inner

corners of the nostrils and the medial point of themouth respectively. Eventually, in order to select thebest of sometimes redundantly available solutions (e.g.,for the fiducial points belonging to the mouth), an intersolutionconsistency check is performed by comparingthe CFs of the points extracted by different detectors ofthe same facial feature.AUs of the FACS system are anatomically related tocontractions of facial muscles [4]. Contractions offacial muscles produce motion on the skin surface andchanges in the shape and location of the prominentfacial features. Some of these changes are observablefrom changes in the position of the fiducial points. Toclassify detected changes in the position of the fiducialpoints in terms of AUs, these changes should berepresented first as a set of suitable feature parameters.Motivated by the FACS system, we represent thesechanges as a set of mid-level feature parametersdescribing the state and motion of the fiducial points.We defined a single mid-level feature parameter, whichdescribes the state of the fiducials. This parameter,which is calculated for each frame for various fiducialpoints by comparing the currently extracted fiducialpoints with the relevant fiducial points extracted fromthe neutral frame, is defined as:

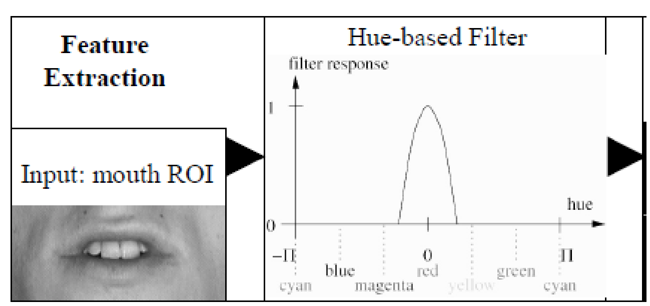
inc/dec(AB) = ABneutral – ABcurrent, where AB

$= \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$

If inc/dec(AB) < 0, distance AB increases.

C. Action Unit Recognition

The last step in automatic facial gesture analysis is to translate the extracted facial information (i.e., the calculated feature parameters) into a description of shown facial changes, e.g., into the AU codes. To achieve this, we utilize a fast-direct-chaining rulebasedmethod that encodes 22 AUs occurring alone orin a combination in the current frame of the input faceprofileimage sequence. A full list of the utilized rulesis given in Table 1. Motivated by the FACS system [4],each of these rules is defined in terms of the predicateof the mid-level representation and each encodes asingle AU in a unique way according to the relevantFACS rule.



International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

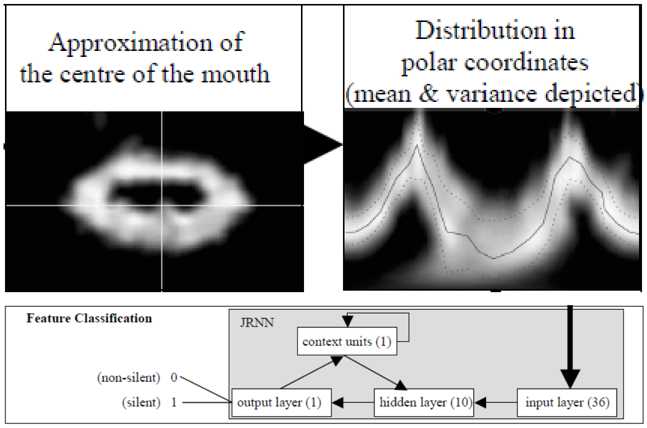


Fig. 3: The outline of the neural-network-based method for speech onset/offset detection

III. EXPERIMENTAL STUDIES

We conducted three experimental studies within our research on automatic facial gesture analysis. The first was aimed at evaluating the performance of our methodfor AU recognition. The second pertained to evaluating the proposed method for speech onset/offset detection. The third was aimed at discerning the facial muscleactions that are typical for speech articulation.

A. Image Database:

Most of the existing approaches to either facial gesture recognition or lip-reading assume that the presence of the face in the input image is ensured [3, 10]. However, in most of the real-life situations where such automated systems are to be employed (e.g., videoconferencing, human-computer interaction, etc.) the location of the face in the scene is not known a priori. The presence of a face can be ensured either by employing an existing method for automatic face detection in arbitrary scenes (e.g., see [11]) or by using a camera setting that will ascertain the assumption at issue. The two algorithms proposed here do not perform face detection in an arbitrary scene; they operate on frontal-view face image sequences acquired by a head-mounted CCD digital PAL camera (Fig. 4).

Fig. 4: Mounted-camera device and an example of an input frame

The face image sequences used in our experiments have been obtained with the help of six certified FACS coders drawn from college personnel. The acquired testimages represent a number of demographic variables including ethnic background (European, Asian and South American), gender (66% female) and age (20 to 35 years). Two datasets have been acquired:

- **Dataset 1:** 48 image sequences of subjects displaying series of facial expressions including single AUs and combinations of those. The first frame is in a neutral
expression and the length is from 95 to 250 frames. No movement of the lips due to a speech articulation is present.

- **Dataset 2:** 6 image sequences of subjects speaking a set of 5 sentences while maintaining a neutral facial expression. The sentences are from the POLYPHONE corpus [12] and contain all of the phonemes used in the Dutch language. The length of sequences varies from 850 to 1050 frames.

B. AU Recognition:

Dataset 1 has been used to evaluate the performance of the proposed method for AU recognition. Metadata were associated with the acquired test data in terms ofAUs that were scored by 5 certified FACS coders other than the one displaying the judged facial expressions. As the actual test data set, we used 40 image sequencesfor which the overall inter-coders' agreement about displayed AUs was above 75%. AU-coded descriptions of shown expressions obtained by human FACS coderswere compared further to those produced by our method. The results of this comparison, given in Table 2, show that in 93% of test cases, our method for AUrecognition coded the analyzed facial expression using the same AU codes as the human observers.

Table 2: Recognition results for the upper face AUs

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

(AU1, AU2, AU4, AU5, AU6, AU7, AU41), the AUs affecting the nose (AU38, AU39), the AUs affecting the jaw (AU26, AU27) and those affecting the mouth (AU8, AU12, AU13, AU15, AU18, AU20, AU23, AU24, AU25, AU28, AU35):
 # denotes the number of AUs' occurrences,
C denotes correctly recognized AUs' occurrences,
M denotes missed AUs' occurrences,
IC denotes incorrectly recognized AUs' occurrences.

	#	C	M	IC	Rate
upper face	54	50	4	0	92.6%
nose	13	12	0	1	92.3%
mouth	102	95	4	3	93.1%
jaw	23	21	1	1	91.3%
Total:	192	178	9	5	**92.7%**

Face Detection

The first step in our gesture recognition algorithm is face detection. The face is detected by using two steps:

- **Color segmentation,**
- **Morphological processing**

Color Segmentation:

Detection of skin color in color images is a very popular and useful technique for face detection. In the skin color detection process, each pixel was classified as skin or non-skin based on its color components values. We apply a simple rule to detect the skin pixels as fast as possible. Two methods in particular are explored.Firstly the RGB space was tried to locate face so as to avoid any calculations.

A pixel with color values (R, G, B) is classified asskin [33,] if:

- R > 95 and G > 40 and B > 20 and
- R > G and R > B and
- R-G > 15

Other widely used color segmentation methods [34] are based on Cr or Hue classifiers. A pixel is considered as skin if Cr Є [10 255]. As Cr component is easy tocompute from RGB and there are only two tests to perform, the classification is really fast, and gives good results. So, we adopted Cr classifier.

Morphological Image Processing

After color segmentation, a mask of non-skin pixels is obtained. However this mask is not perfect: some sparse non-skin pixels are still visible while some parts of the face can be masked. Morphological image processing is thus a good way to eliminate the non-skin visible pixels and regroup the skin pixels: First, erosion is performed to remove sparse non-skin pixels. Second, dilation is performed with a larger disk to regroup the skin regions and smooth their contours. Fig 2 shows the color segmentation and morphological processing stages EigenfaceTheEigenface scheme [35] is pursued as a dimensionality reduction approach, more generally known as principal component analysis (PCA), or Karhunen-Loevemethod.Such method chooses a dimensionality reducing linear projection that maximizes the scatter of allprojected images. Given a training set of N images

Gi (i = 1,2,.......N) each of size m × n, we could turn the set into a big matrix as

A = [F1F2 .....FN] (1)

where Fi 's are column vectors, each corresponding to an image as

Fi = fi - m

fi = reshape[Gi¢ ,(mn,1)]

mean(i) i

m = f

The total scatter matrix is defined asT

ST = AA (2)

Consider a linear transformation W mapping the image space into a p-dimensional feature space, p<=N<<mn. PCA chooses the projection Wopt that maximizes the determinant of the total scatter matrix of the projected images, i.e.,

argmax [1 2 ....] Topt W T P

W = W S W = w ww (3)

Where wi's are eigenvectors of ST corresponding to the p largest Eigen values. Each of them corresponds to an "Eigenface". The dimension of the feature space isthus reduced to p. The weights of the training set images and test images could be then calculated and the Euclidean distances are obtained. The test face is recognized as the gesture of training set with the closest distance, if such distance is below a certain distance.

IV. CONCLUSIONS AND FUTURE WORK

The presented method for automatic AU recognition extends the state of the art in automatic facial gesture analysis in face image sequences in terms of number ofAUs handled. The significance of this contribution is also in the performed experimental studies that suggest: (i) that it is possible to determine whether the observed subject is speaking or not from visual data only, and (ii) that at least 5 AUs are typical for articulation and could be, therefore, treated as noise in affect sensitive interpretation of visual data. The presented algorithm for automatic AU coding of face image sequences does not take into account thetemporal nature of facial gestures. Yet, the presented AU coder could greatly speed up the time-consuming (manual) process of acquiring AU-labeled data onwhich models that can capture the temporal nature of facial gestures (e.g., HMM) could be trained. Devising both a HMM-based AU coder and an affect-sensitiveanalyzer of AU-coded "silent" and "non-silent" facial data forms the main focus of our further research As part of the future work, we would like to develop anapplication that would allow the user to add/delete faceclasses in the training set. This would give users thefreedom to define their own user groups rather than a pre-defined set on the server. Another added feature willbe to run the application in real time to get its testdatabase from images with more than just one face in it.

REFERENCES

[1] J. Russell and J. Fernandez-Dols, The psychology of facial expression, Cambridge University Press, 1997.

[2] D. Keltner and P. Ekman, "Facial expression of emotion", Handbook of Emotions, Guilford Press, pp. 236-249, 2000.

[3] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art", IEEE TPAMI, vol. 22, no. 12, pp. 1424-1445, 2000.

[4] P. Ekman and W. Friesen, Facial Action Coding System, Consulting Psychologist Press, 1978.

[5] G. Donato, et al., "Classifying facial actions", IEEE TPAMI, vol. 21, no. 10, pp. 974-989, 1999.

[6] A. Pentland, "Looking at people", IEEE TPAMI, vol. 22, no. 1, pp. 107-119, 2000.

[7] M. Pantic and L.J.M. Rothkrantz, "Expert system for automatic analysis of facial expressions", Image and Vision Computing, vol. 18, no. 11, pp. 881-905, 2000.

[8] M. Pantic, et al., "A hybrid approach to mouth features detection", Proc. IEEE Conf. SMC, 2001, pp. 1188-1193.

[9] A. Adjoudani, et al., "A multimedia platform for audiovisual speech processing", Proc. Eurospeech, 1997, vol. 3, pp. 1671-1674.

[10] J. Wojdel and L. Rothkrantz, "Visually based speech onset/offset detection", Proc. Euromedia, 2000, pp. 156-160.

[11] R. Feraud, et al., "A fast and accurate face detector based on neural networks", IEEE TPAMI, vol. 23, no. 1, pp. 42-53, 2001.

[12] M. Damhuis, et al., "Creation and analysis of the Dutch polyphonecorups", Proc. ICSLP, 1994, pp. 1803-1803.