Recognition of Expressions on Human Face using AI Techniques

Arpita Nagpal[1], Astha Garg[2]

[1] **M.Tech. Scholar, Jaypee Institute Of Information & Technology**
Noida , Uttar Pradesh, India
er.arpitanagpal@gmail.com

[2] **M.Tech. Scholar, Jaypee Institute of Information and Technology**
Noida, Uttar Pradesh
garg.astha08@gmail.com

Abstract

Facial expressions convey non-verbal cues, which play an important role in interpersonal relations**.** Facial expressions recognition technology helps in designing an intelligent human computer interfaces. This paper discusses a three phase technique for the facial expression recognition of the Indian faces. In the first phase the faces are tracked using Haar classifier in the live videos of Indian student's community. In the second phase 38 facial feature points are detected using Active Appearance Model (AAM) technique. In the last step the support vector machine (SVM) is used to classify four primary facial expression. Integrating these broader techniques and obtaining a reasonably good performance is a very big challenge. The performance of the proposed facial expressions recognizer is 82.7%.
.

Keywords: *Facial Expression, Support Vector Machine, Active appearance Model, Haar classifier, Human Computer Interface.*

1. Introduction

Usually facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. Facial expression analysis refers to computer systems that attempt to automatically analyze and recognize facial motions and facial feature changes from visual information.

Facial expressions play an important role wherever humans interact with machines.

Facial Expression recognition is one of the topics of interest from Affective Computing. Affective computing powers the computer system the human-like capabilities of observation, interpretation and generation of affect features. People express the affects through a series of action on facial expression, body movements, various gestures, voice behavior, and other physiological signals, such as heart rate and sweat, etc. "Affect" was linked to lifeless machines, and was normally studied by psychologists. It is quite new in the recent years that the affect features were captured and processed by the computer. The facial expression recognition approaches have generally attempted to recognize expression of six basic emotions named as happiness, surprise, sadness, disgust, fear and anger[1] [2] or to recognize Action Units (AUs) of the Facial Action Coding System (FACS)[7][9].

FACS defines AUs, which are a contraction or relaxation of one or more muscles. It corresponds to individual facial muscles. It was originally developed by Paul Ekman and Wallace V. Friesen in 1978.

Moreover, this system divides the facial expression recognition system into face tracking, facial feature extraction and classification and there are different schemes to be followed in different parts of the system. Different facial expression recognition algorithms [2][5][6][12][13][14] have been introduced in the literature depending on the discriminative information extracted from the image and the classification procedure used.

Most previous systems assume presence of a full frontal face view in the image or the image sequence being analyzed, yielding some knowledge of the global face location. To give the exact location of the face, Viola & Jones [8] use the Adaboost algorithm to exhaustively pass a search sub-window over the image at multiple scales for rapid face detection.

Shenchuan Tai and Huang [6] first manually localize feature points in the first frame of an image sequence and then use cross correlation of optical flow to track features and mathematical model. This technique of manually marking is not suitable every time.

Negazz,Besnassi and Benyettou [16] they have used improved AAM and probabilistic Neural Network. They have used JAFFE database to evaluate their performance.

Chang and Chen [2] have used Neural Network as a Classifier. But difficulties using back propagation neural network is that they converge slowly. They learn slowly and stick to local minima in learning phase.

In this paper we have integrated the approaches of face detection, facial feature extraction and expression classification which have individually been implemented but have never been used together for facial expression recognition. Viola & Jones Haar-like features [8][4] have been used for face detection. For feature extraction Active Appearance Model (AAM)[3][10][11] has been used. Machine learning system of Support Vector Machines (SVM) [1][15]classifies the expression. The novelty of this approach lies in using a large variety of faces in training data set. Indian Faces vary in features. Faces in the video may not be only frontal faces. It can do classification even for a little tilted faces. In each frame it can detect any number of faces and can do classification for each detected face. It gives good performance even if the features of the detected face vary a lot.

The paper is organized as follows. Section 2 is collection of image database. Section 3 shows the how it is worked out. Section 4 is technique of face detection. Section 5 is explaining technique used for extracting the important features. Section 6 tells about how the face is classified using those extracted features. Section 7 depicts the result obtained. Section 8 is the concluding part.

2. Data Collection

In this work, Database of images of Indian faces was collected. Front view face images were acquired from a camera of the students of colleges from different Indian community. Very little work has been done on Indian faces. So, it is an effort made to use Indian database as the training set.

Indians are one of the most varied people; they do not look a specific way because they are so diverse. The Indians among themselves vary considerably in their features, in form of head and face. There are lots of variations in Indian faces. There are many communities and people from different background and community differ a lot in their appearance and features. Except for the skin colour difference there are other feature differences of people in different parts of country. So, it was a challenging task to work over this varied type of database.

There are 50 subjects in the database. Since for some subjects, not all of the six facial expression sequences were available. Subjects in database have ages between 18 and 50. They were 80% male and 20% female.

Fig. 1 Sample Images used as training images

3. Implementation

Here in this project Facial Expression Recognition deals in finding firstly the face from a given video and then the facial expression of the face are determined. This target is achieved in three stages. In the first stage, Viola & Jones features related to visual representations that are reminiscent of Haar basis functions are used and the face is detected from the video that is input. It can detect frontal faces and even the faces which are tilted left or right Now this face image is input to the second stage. Here by using model based approach features are taken out from the face. 6 features such as eyes, nose, and mouth were extracted from 38 feature points from the face using AAM (Active Appearance Model) approach. Why 38 points? (See section 6) These feature vectors were used to train SVM classifier along with labels of expression given to each feature vector. This face expression recognition system is clearly distinguished from previous approaches as these 3 algorithms named above have never been integrated for expression recognition purpose. Use of database from student's community of different parts of India was also never used. The video runs at 25 frames per second. It has been found that this approach perform well even at lower frame rates. We assume that faces in video are not always full frontal, but there can be variations in head poses and can be tilted left or right.

4. Face Detection

Face detection is the essential front end of any face recognition system, which locates and segregates face regions from cluttered images, either obtained from video or still image. The task of face detection is so trivial for the human brain, yet it still remains a challenging and difficult problem to enable a computer to do face detection. This is because the human face changes with respect to internal factors like facial expression, beard and moustache, glasses etc and it is also affected by external factors like scale, lightning conditions, contrast between face and background and orientation of the face.

The pixel values inform us only about luminance and color of a given point. Feature based detection is much easier than pixel based detection. It is therefore more interest to find a detector based on more global features of the object. This is the case of HAAR descriptors, where the functions allow the knowledge of the contrasts difference between several rectangular regions in image. Used Viola & Jones's face detector [6] [8] based on the Haar-like features.

Haar basis function uses 3 kinds of features:

ISSN (Online): 2231-5268
www.ijcsms.com

1) Two- rectangle feature- Difference between sums of pixels within two rectangular regions. Type-1 & 2 in fig.
2) Three rectangle features- Computes the sum within two outside rectangle and subtracts it with the sum in the center rectangle. Type 3 & 4 in fig.
3) Four-rectangle features- difference between diagonal pairs of rectangle. Type 5 in fig.

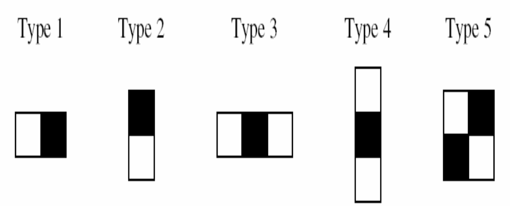

Fig. 2 Haar-like features [4]

Each feature results in a single value which is calculated by subtracting the sum of the white rectangle(s) from the sum of the black rectangle(s).
A window is passed over Image, part of image in window is called integral image. Integra Image (ii) [8] at location x, y contain sum of pixels above and to left of x, y. A 24X 24 sliding window scans the image and each sub window is classified as face or non-face using retinal approach.
A simple decision tree classifier is used. It is a weighted sum of weak classifier using boosting procedure. Finally a cascade is implemented; it is a cascade of boosted classifiers with increasing complexity.

5. Feature Extraction

Generally, for recognizing the expression we need to analyze the geometric features of the face. The geometric features like eyes, eyebrows, mouth need to be extracted from the detected face. For this, we have used a statistical model based approach named Active Appearance Model (AAM). A precondition for this method is to first determine the location of the features. These features are called landmarks. The landmarks are also called key points or facial characteristic points and represent distinguishable points on the face area like the inner and outer corners of the eyes or the left and right-most points of the mouth and eyebrows.
On the training set of face images, landmark points are manually marked and saved to generate a model of these points.
Each shape in training set is represented by a set of 'n' labeled landmark points. This 'n' is determined experimentally. As good choices for landmarks are points at clear corners of feature boundaries. Starting with a minimum of 3 points on eyebrow, 4 points on eye, 7 points on mouth, and 5 on nose contours found that a total of 26 points were sufficient but for eyebrows and mouth we need more intermediate points. Intermediate points are used at the curves to define boundary more precisely. Observed that the shape of the features were not exact when the AAM model is formed from the land marked training image. Then intermediate points on eyebrows and mouth were increased to 6 and 9 respectively.
By repeated experiments found that 38 points manually marked on facial features(5 points on each eye, 6 points on each eyebrows, 7 points on nose and 9 on mouth) are sufficient to approximate the geometry of original shape.. These points correspond with the peak and valleys of the curvature of the profile contour.
Procedure of finding facial feature

1) 38 feature points were manually marked with a computer-mouse around facial landmarks on each face image from an Indian database. See Figure
2) Build a training set of images in which corresponding `landmark' points have been manually marked. The marked points are stored in form of array and each image is stored in a form of a structure.
3) Then compute a statistical model of the shape variation, a model of the texture variation and a model of the correlations between shape and texture.
4) The shape model is built by aligning all of the shape vectors to a common coordinate frame and performing Principal Component Analysis (PCA) on these. The shape model is then controlled by $\mathbf{b}_s$, each shape generated by the model calculated using, $\mathbf{x} = \mathbf{x} + \mathbf{P}_s.\mathbf{b}_s$; where $\mathbf{P}_s$ contains the eigenvectors of the covariance matrix.
5) To construct a model of gray-levels, each training image is warped so that the control points match the mean shape **x**, using Delaunay triangulation to calculate the warp parameters. The texture data within the region of the face now bounded by the mean shape is sampled to form a texture vector.
6) In a similar manner to the shape model, the texture model is constructed from $\{\mathbf{g}_i\}$ using PCA.
7) Both the models of shape and appearance are combined to form one model of shape appearance.
8) Finally, this model is being mapped over the test image.

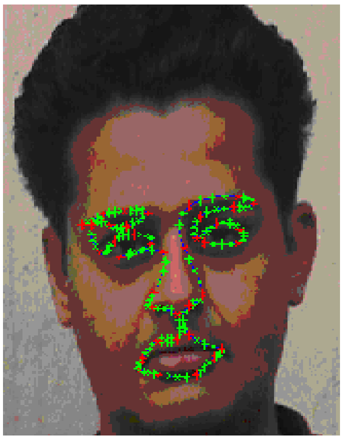

Fig. 3 Depicting one of the training images with points marked on features

6. Classification

The next task is to classify the data. The features extracted now need to be classified according to the set it belongs to. The detected face must belong to one of the categories: happy, sad, surprised or disgusted.
To do the same, Support Vector Machine (SVM) has been implemented. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other.
Training set faces have been labeled with name of the expression. SVM train is the function which is used to groups the training set into different categories. SVM train takes as parameters the training set and the labels. SVM classify classifies each of the test image into different categories based on the SVM train.
Using a random subset of the feature vectors, trained a *cascade* of SVM classifiers. Each binary SVM classifier in the cascade was trained to act as an *expression detector*, outputting a positive response if its expression is present and a negative response otherwise. So, for example, a binary SVM trained as a "happy detector" would classify between expressions which are *happy* and *not happy*.

7. Experimental Result

In this section we are going to explain the experiments carried out using an Indian student's community video, which was played for about 10 minutes. Frame rate was 25 frames per second and approximately 1000 faces it could detect. These faces can be of same person or different persons. It could also detect 2 or more faces in one frame. These detected faces are stored in the database and they are fed as to SVM, where expression of each face is recognized.
The result of recognition is depicted in the table below. This result is taken when approximately, 1000 frontal faces gets extracted from the video.

Table 1: Confusion matrix

Actual / *Predicted*	*Neutral*	*Happy*	*Surprise*	*Sad*	*Disgusted*
Neutral	180	11	9	8	7
Happy	12	220	14	5	7
Surprise	6	17	205	6	6
Sad	1	1	8	108	16
Disgusted	1	1	11	23	114

Testing Images Taken: 1000
Efficiency of each expression:
Neutral: 90% Happy: 88% Surprise: 82%
Sad: 72% Disgusted: 76%
Total accuracy: 82.7%
In paper [15] they have made an expression and gender recognition system that uses AAM for feature extraction and SVM for classification. They have recognized 4 expressions Happy, Sad, Angry, Neutral and the overall accuracy of the cascade was calculated as 76.4%.
The performance of our expression classification is 82.7%, which is comparable with papers using these techniques.

8. Conclusion

In this paper we have presented a system that is able to track a face from a video and classify the facial expression. Some of the very detailed techniques of computer vision and image processing domain have been integrated and worked upon. The recognizer presented here can be reconfigured, in the sense that it can be used to recognize any other set of expressions just by changing the training data. 4 expressions classified here can be extended to a larger number if we increase the number of images in training data and then its associated label in SVM.
It achieves an 82.7% recognition success on a set of 1000 sequences of faces extracted from a video.
This solution uses information from a set of images from an Indian database, to correctly train the classifier. The performance of the system can increase if we use more number of images in training set. Due to limitation of time, number of training images was limited. Video could run at 25 frames/sec, for about 10 minutes. In future, we can also increase the time for which the video is running and hence the number of frames. So, our testing database will

increase. We can also try to use Neural Network instead of Support Vector machine.

References

[1] Philipp Michel Rana El Kaliouby," Real Time Facial Expression Recognition in Video using Support Vector Machines", ACM 1-58113 ICMI'03, November 2003.
[2] Jyh-Yeong Chang , and Jia-Lin Chen, "Automated Facial expression Recognition System Using Neural Networks," Journal of the Chinese Institute of Engineers, Vol. 24, No. 3, pp. 345-356., 2001
[3] Jorgen Ahlberg, "An Active Model for Facial Feature Tracking", EURASIP Journal on Applied Signal Processing ,2002
[4] F. Abdat, C. Maaoui , and A. Pruski, "Real Time Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm", ISBN 978-953-307-051-3, pp. 288, February 2010
[5] W.K. Teo1, Liyanage C De Silva2 , and Prahlad Vadakkepat1" Facial Expression Detection , and Recognition System", Journal of The Institution of Engineers, Singapore Vol. 44 Issue 3 2004
[6] Shenchuan Tai , and Hungfu Huang, "Facial Expression Recognition in Video Sequences" W. Yu, H. He, , and N. Zhang (Eds.): ISNN 2009, Part III, LNCS 5553, pp. 1026–1033, 2009.
[7] Jeffrey F. Cohn, AdenaJ. Zlochower, James Lien, and Takeo Kanade, "Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding", Psychophysiology, 36 Cambridge University Press, 1999
[8] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features",Conference On Computer Vision And Pattern Recognition 2001
[9] Marian Stewart Bartlett, Gwen C. Littlewort, Mark G. Frank, Ian R. Fasel, Javier R. Movellan, "Automatic Recognition of Facial Actions in Spontaneous Expressions", Journal Of Multimedia, Vol. 1, No. 6, September 2006
[10] T.F. Cootes,G.J.Edwards,C.J Taylor, "Active Appearance Model", IEEE transactions on Pattern Analysis and Machine Intelligence. Vol 23. No 6. pp. 681-685. June 2001
[11] Ragib Morshed, Xiangxin Zhu, "Deformable Statistical Models for Face Detection: An Active Appearance Model(AAM) Perspective", European conference on computer Vision, 1998.
[12] Caifeng Shan, Shaogang Gong , andPeter W. McOwan "Robust facial Expression Recognition Using Local Binary Patterns", IEEE Trans.,2005
[13] Matti Taini, Guoying Zhao, and Matti Pietik ainen, "Weight-Based Facial Expression Recognition from Near-Infrared Video Sequences " Springer-Verlag Berlin Heidelberg 2009 pp. 239–248, 2009.
[14] He Li, Jose M. Buenaposada, Luis Baumela, "Real-Time Facial Expression Recognition with Illumination-Corrected Image Sequences", IEEE Trans.,2008
[15] Yunus Saatci , andChristopher Town, "Cascaded Classification of Gender and Facial Expression using Active Appearance Models", 2009
[16] N.Negazz, M. Besnassi, A. Benyettou, "Application of Improved AAM and Probabilistic Neural Network to facial Expression e-cognition", Journal of applied sciences, 2010
[17] Andrew Ryan, Jeffery F. Cohn, Simon Lucey, Adam Rossi, "Automated Facial Expression Recognition System", IEEE Trans., 2009
[18] Frank Y. Shih , Chao-Fa Chuang, "Automatic extraction of head and face boundaries and facial features", New Jersey Institute of Technology 10 March 2003
[19] Hadi Seyedarabi, Ali Aghagolzadeh "Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System", 2004 IEEE
[20] Buenaposada, J.M., Munoz, E., Baumela" Recognising Facial Expressions in Video Sequences.Pattern Anal. Applic" 11, 101–116 (2008)
[21] Jorge P. Batista, "Locating Facial Features Using an Anthropometric Face Model for Determining the Gaze of Faces in Image Sequences", Springer-Verlag Berlin Heidelberg pp. 839–853, 2007.