Enhancing Facial Action Recognition Using Multikernel

Bombale G. R.¹, Prof. Rokade P. P.²

¹Department of Computer Engineering, SNDCOE & RC, Yeola
²Department of Information Technology, SNDCOE & RC, Yeola

Abstract—Facial expressions are depending on Facial Action Coding System (FACS) and action units (AUs), instead of six basic expressions, facial component models are proposed for tracking and modeling different facial features, including lips, eyes, brows, cheeks. For this we propose to use multi-kernel SVMs this help in selecting the most accurate information from each kernel function. For this we present first kernel an active appearance model (AAM) trained on face images from a publicly available database to represent shape and texture variation key to expression recognition. Parameters from the AAM are used as features for a classification scheme. Another kernel is Gabor-wavelet-based methods (LGBP) it can achieve high sensitivity and specificity for emotion specified expressions it evaluates a Gabor wavelet based method to recognize AU’s in image sequences of increasing complexity and the third kernel proposed feature in that we will be calculate the edges of land marking part from the face. These calculation of edges is first converted into gray color and calculate the edge. We combine these feature and compare with the database to deal with these three kinds of information and take the best output.

Keywords—Emotion Extraction, Facial Action, LGBP, AAM, Feature Extraction, Binary image.

I. INTRODUCTION

Everyday almost everyone in this world interact with other in one or another way either directly (for e.g. face to face) or indirectly (for e.g. phone calls). In some profession interaction with people are the main deed to perform like call centers, sale executives etc. With great advancement in technology in terms of different techniques of people interacting with each other it is quite necessary that one should be aware of current emotions of the person he/she is interacting. With the advancement of 3G technology in mobile communication field one may be capable to interact face to face with other while talking so if one is aware of mood of other in advance that interaction will certainly result in social as well as professional benefits. Firstly, what are emotions? A mental state that arises spontaneously rather than through conscious effort and is often accompanied by physiological changes and these physiological changes are recognized from outer world. Secondly, major emotions that humans face in day to day life: Happy, sad, anger, disgust.

Most of the existing approaches to detect AU’s fall into three categories depending on the type of feature used.

The first and oldest category includes geometric feature- based methods; the second includes appearance feature- based methods; whereas the last considers all the methods that use both geometric and appearance features. As geometric and appearance feature-based methods have their own advantages and drawbacks, i decide, as participant, to combine both. And given output of this pass as a input to proposed features to get the emotion.

In the face registration, first step is to normalize the image. Some simple normalization methods are used for this. The main complicated part of this step is detection of face. For this we remove extra part from image to get the exact emotion. Image registration relies on preliminary face detection and facial landmark localization. Face detection is usually based on the genetic algorithm and eigen face technique face detector it capture the eyes first and then classify them as face or non face.

The next step is on a classification framework. Detentions can be performed using an Multikernel Support Vector Machine (MK-SVM) to deal with features. After getting the cropping face from whole image Using landmark on the relative position we compare different AAM fitting algorithms to aware about the facial action. The landmark from the cropping image done manually before the analysis to get the action. So that the output from this process which is depends upon the land marking position on detected face. This is the main drawback of this approach, So the using single process it does not confirm that the which action is done on the face.

II. RELATED WORK

Very little facial expression analysis research has focused on recognizing specie facial actions like raising an eyebrow, squinting and depressing the lip corners. Cootes et al., in [5], has demonstrated a system to recognize some basic facial actions like mouth activity and eye-movements using Hidden Markov Models (HMMs) and multi dimensional receptive histograms. Her system on lyre cognizes basic facial actions like blinks/mouth open/eyes closed etc. And cannot recognize more subtle facial actions like eye-widening[21].

Priya Metri et al., in [13], have describe system to detect head nods and head shakes in real time by directly detecting and tracking the between-eyes” region.
The between-eyes” region is detected and track educating circle frequency “, which is a discrete Fourier transform of point slying on a circle, together with skin color information and templates [16]. P. Ekman et al., in [6]. Head nods and head shakes are detected dbase don predened rules applied to the positions of between-eyes” in consecutive frames. Motivated by Paul Ekman’s Facial Action Coding System(FACS), some of the approaches attempt to recognize action units (AUs) the fundamental muscle movements that comprise Paul Ekman’s Facial Action Coding System, which can be combined to describe all facial expressions [2]. These facial actions are essentially facial phonemes, which can be assembled to form facial expressions.

Ying li et al., in [9], have developed a system to recognize sixteen action units and any combination of those. The shape of facial features like eyes, eyebrow, mouth and cheeks are described by multi state templates. Takeo Kanade et al., in [10] proposed a parameters. The parameters of these multistate templates are used by a Neural Network based classier to recognize the action units. This system requires that the templates be initialized manually in the frame of the sequence, which prevents it from being fully automatic.

In an earlier work, Maja panic et al., in [14], describe a system that recognizes various action units based on dense, feature point tracking and edge extraction. Mehrabian[12] compared several techniques, which included optical, principal component analysis, independent component analysis, local feature analysis and Gabor wavelet representation, to recognize eight single action units and our action unit combinaions using image sequences that were manually aligned and free of head motions. They showed 95.5% recognition accuracy using Independent Component Analysis and Gabor wave letter presentations [11].

Zhao and Pietikainen, in[25], apply volume local binary patterns (LBPs), which are the temporal equivalent of an LBP. Micheal [19] and Zhang and Ji, in [24], use facial points or component shapes with features such as crow feet wrinkles and nasallabial furrows. Rakotomamonjy et al., in [15], use a constrained local model (CLM) to track the face and features and encode appearance using an LBP. SVMs are used to classify AUs.

Baltrusaitis et al., in [1], It is interesting to compare this database with standard databases such as Cohn-Kanade database or MMI database. After that we choose to use the AAM introduced by D. Keltner et al., in [7]. An AAM is a statistical model of the shape and gray-level appearance of the face that can generalize to almost any valid example.

With the help of this LGBPs, introduced by Zhang et al., in [23] for face recognition, exploit multi resolution and multi orientation links between pixels and are very robust to illumination variations and misalignment. to combine these kernels in a multikernel SVM framework is developed [17].

Finally, to deal with temporal aspects of AU display, they post process the classification outputs using a filtering and a thresholding technique. Face detection is usually based on the public OpenCV face detector designed by Viola and Jones, in [20], that applies an attentions cascade of boosted classifiers on image patches to classify them as face or nonface.

Koelstra et al., in [8], achieve 90.9 % accuracy in recognizing 6 single action units by combining holistic facial analysis and optical with local feature analysis. Both of the above approaches report their results on manually pre-processed image sequences of individuals deliberately making facial actions in front of a camera.

III. System Overview

Enhancing action recognition is system proposed to recognize exact action. It describes image coding and details the LGBP histogram, with AAM coefficient calculation and proposed texture feature matrix. It explains the process to detect Action Units (AU) in facial images and post processing temporal analysis. It details experiments to validate the choice of LGBP histograms and the histogram intersection kernel. It reports Action Units detection results on the trainy data set. it reports emotion recognition results. In addition to detailing the steps of the Action Units detector conception for the FERA challenge, this tries to emphasize with rigorous experiments the benefit of combining features and the AU labels compatibility between databases.

In this system, first capture the facial image through the source and detect the facial part, i concentrate on face to get the AU. After that, i crop the image and get cropped image after ignoring all noises. Then i apply AAM kernel. An AAM is a statistical model of the shape and gray-level appearance of the face that can generalize to almost any valid example. AAM can provide important spatial information of key facial landmarks but are dependent of an accurate matching of the model to the face images so i use second kernel LGBP.

Local Gabor binary pattern (LGBP) histograms generates with shape and edge calculation. LGBPs, introduced for face recognition and the use of histograms results in the loss of spatial information that really depends on identity. So to overcome these i used third proposed texture feature.
Curve calculation is the third kernel in that first accept the output which comes from previous kernel and then this convert into the gray scale, convert in binary code and calculate the edges, these calculation is compared with database and give the final output.

To perform AU detection, I select support vector machines (SVMs) for their ability to find an optimal separating hyper plane between the positive and negative samples in binary classification problems. Our method combines multiple visual descriptors with features for multimodal different classification of image. Extracted features are combined using Multiple Kernel Learning algorithm and they classified using an SVM into one of the seven emotion categories: Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise. The proposed method achieves competitive results, with an accuracy.

A. Project Aim
- The aim of this system is to use computer vision techniques to automatically detect and analyze the emotions from the digital images.
- To develop a system that is easy to use, can be easily adaptable, modified, reproduced, and even improved.

B. Motivation
The algorithms used in this project are very general in form. The idea behind this is to allow the system to be sensitive enough to detect the instances of facial regions as well as lip region which can occur in background of image. In such cases faces will be detected and that will result in chances where several false detection will be done. This will result from any lip colored marks in the image.

C. Motivation
The algorithms used for implement the system are very general in form. The idea behind this is to allow the system to be sensitive enough to detect the instances of facial regions as well as lip region which can occur in background of image. In such cases faces will be detected and that will result in chances where several false detection will be done. This will result from any lip colored marks in the image.

D. Applications
- Recognizing the expression of a man can help in many of the areas like in the field of medical science where a doctor can be alerted when a patient is in severe pain. It helps in taking prompt action at that time.
- Face detection and recognition, as an important component of pattern recognition, has a wide range of applications on virtual reality, videophone system and security system etc. [9].
- Facial expression recognition has been a subject of investigation in the last years due to the great number of potential day-to-day applications such as human–computer interaction (HCI), emotion analysis, automated tutoring systems, smart environments, operator fatigue detection in industries, interactive video, indexing and retrieval of image and video databases, image understanding, and synthetic face animation [10].
- Other applications that use emotion recognition are customer services, intelligent automobile systems, game and entertainment industry, automated systems that provide aid to psychologists, behavioral and neuro science researchers [5].
- Facial expression recognition systems can also provide a less intrusive method to apprehend the emotion activity of a person of interest [10].
- Apart from the two main applications, namely robotics and affect sensitive HCI, expression recognition systems find uses in a host of other domains like Telecommunications, Animations, Affect sensitive music juke boxes and Educational Software [11].

E. Emotions Consider In This Project
In this system we are considering six major emotions which are mainly centering toward lips in facial region. These emotions are:

Fig. 1. System Architecture
IV. FEATURES

A. Active Appearance Model (AAM)

To extract the AAM coefficients, we train two AAM local models, i.e., one for the mouth and one for both eyes. The reason behind taking two local models instead of one global one for the whole face comes from the fact.

B. Local Gabor Binary Pattern (LGBP):

I apply the LBP operator on the 18 Gabor magnitude pictures, resulting in 18 LGBP-maps per facial image. This combination of the LBP operator with Gabor wavelets exploits multi resolution and multi orientation links between pixels.

C. Skin Color Maintenance

First we contrast the image. Then we perform skin color segmentation after that finding of largest region is very important. After finding largest region we need to check for the probability that how much the connected region can be become as face.

D. Face Detection

For face detection, first we convert binary image from RGB(Red, Green, Blue) image. For converting binary image, we compute the mean value of RGB for every pixel and if the mean assessment is less than 110, we substitute it by black pixel and if not we substitute it by white pixel.

E. Eyes Detection

For eyes detection, we convert the RGB face to the binary aided face. Let us consider the width of the face by W. We scan from the W/4 to (W-W/4) to find the two eyes middle position. The maximum white continuous pixel along the height between the ranges is the middle position of the two eyes. Then we find the starting high or upper position of the two eyebrows by the process of searching vertically. For left eye, scan w/8 to mid and for right eye we scan mid to w – w/8. Here w is the width of the image and mid is the middle position of the two eyes.

F. Lips Detection

For lip detection, we determine the lip box. And we consider that lip must be inside the lip box. So, first we determine the distance between the forehead and eyes.
Then we add the distance with the lower height of the eye to determine the upper height of the box which will contain the lip. Now, the starting point of the box will be the \( \frac{1}{4} \) position of the left eye box and ending point will be the \( \frac{3}{4} \) position of the right eye box. And the ending height of the box will be the lower end of the face image. So, this box will contain only lip and may some part of the nose. Then we will cut the RGB image according the box.

G. Database & Training

In this we can consider the data base in tables. Here we consider two tables as person and position. Person table is for storing the name of the people and their emotions will be stored in the position table. For each index of position table there will be 6 control points for each lip Bezier curve, left eye Bezier curve and right eye Bezier curve, height and width of lip, height and width of left eye, and height and width of right eye. So, by this method we can know the emotion of the people.

H. Emotion Detection

The value which is found here is compared with the values that are present in the data base, then the written program will try to match with emotions.

V. ALGORITHMS

A. Face Extraction Algorithm

Firstly, a color image is taken as input and it is converted into RGB color space. The pixels which lie in the above boundaries of different models, they represent a skin. Then after detecting the skin, the required face region is cut. Figure 8(a) shows the flow chart for this module and Figure 8(b) shows the implementation of this module.

Fig. 8(a) Flowchart For Face Extraction Module

Fig. 8(b) Implementation of Face Extraction Module

B. Facial Feature Extraction Algorithm

After the face is detected from the color image, and then facial features like eyes and mouth are located and extracted. Firstly the eyes and mouth are localized within the input image. In this search mode, the search space for face in subsequent color image is reduced to the small area surrounding the eyes and mouth. The eye boxes and lip box is extracted from the detected face. Figure 9(a) shows the flow chart for this module and Figure 9(b) shows the implementation of this module.

Fig. 9(a): Flowchart of Feature Extraction Module
C. Simple MKL Algorithm

VI. CLASSIFIERS

Given training samples composed of LGBP histograms and AAM coefficient vectors, and texture feature associated with labels, the classification function of the SVM associates a score to the new sample. With the triple representation of the hyperplane’s normal vector \( k \) is the kernel function resulting from the dot product in a transformed high-dimensional feature space.

This is a new way of using multikernel learning, instead of combining different kinds of kernel functions. I combine different features. The AAM modeling approach takes the localization of the facial feature points into account and leads to a shape-free texture less dependent to identity. However, one of the severe drawbacks is the need for good accuracy for the localization of the facial feature points. The GEMEP-FERA database contains large variations of expressions that sometimes lead to in accurate facial landmarks tracking. In such cases, multikernel SVMs will decrease the importance given to AAM coefficients and increase the performance of the texture feature for the best analysis from database which used to compare the given input image and the database. So that it combines these features and calculate the ratio of histogram which is better performance from these. In the case of a multikernel SVM, the kernel \( k \) can be any convex combination of semi definite functions.

Thus,

\[
K = LGBP + AAM + \text{Proposed feature}
\]

Weights of 3 are set to have an optimum hyperplane in the feature space induced by \( k \). This hyperplane separates positive and negative classes and maximizes the margin (minimum distance of one sample to the hyper plane). This optimization problem has proven to be jointly convex in three object; therefore, there is a unique global minimum, which can be efficiently found. first represents the weight accorded to the LGBP features, and second is the one for the AAM appearance vector, and third represents the gray level co coherence matrix. Thus, using a learning database, the system is able to find the best combination of these three types of features that maximizes the margin.

VII. EXPERIMENTS AND RESULT

The expressions such as smile, sad, surprise and neutral are considered for the experiment of face recognition. The faces with expressions are compared against the model face database consisting of neutral faces.
All the face images are normalized using some parameters. The Bezier points are interpolated over the principal lines of facial features. These points for each curve form the adjacent curve segments. The Hausdorff distance is calculated based on the curve segments. Then, understanding and decision of facial emotion is chosen by measuring similarity in faces. The ground truth set for estimating the performance of the algorithm is provided with the categories in the experiments, which are correct if the decision is belonged to the correct category. Fig. 10.1 shows how to move the control points across different subjects (e.g., neutral and smile) and to interpret a tracking facial features with Bezier curve and extracted feature points interpolation. To categorize facial emotion, I need first to determine the expressions from movements of facial control points.

The third kernel algorithm is performed two major steps: one is a detection of facial region with skin color segmentation and calculation of feature-map for extracting two interest regions focused on eye and mouth. And the other is a verification of the facial emotion of characteristic features with the three different methods. Experimental results shows average successful ratio to recognize the facial expression, and this indicates the good performance and enough to applicable to mobile devices.

In this system I have just concentrated on the static images of the human. The extension of this work can be done in such a way that i can capture the video and the i can analyze the expression of the human present in the human and resolve the issues related with quality of image.

**REFERENCES**


