Enhancing Facial Action Recognition Using Multikernel-Review

Bombale G.R.¹, Prof. Rokade P. P.²

¹PG Student, Department of Computer Engineering, SND COE & RC, Yeola, Savitribai Phule Pune University, Dist-Nashik, Maharashtra, India.
²Assistant Professor & HOD, Department of Information Technology, SND COE & RC, Yeola, Savitribai Phule Pune University, Dist-Nashik, Maharashtra, India.

Abstract—Multi-media technology especially image processing is going through a rapid development, the requirement to study how human beings reflect to environments. In effective computing, facial expression extraction and synthesis has become an important technological and research issue. Emotion extraction systems have been proposed to extract emotions from a human face. These systems play an important and effective role in Human-Computer Interaction and also find their applications in customer services, game and entertainment industries, security systems, animations, etc. This paper emotion extraction is proposes using three different approach. we get more accurate recognition rates and to achieve higher performance. The idea behind this paper is to extract emotions from eyes and lips along with the classification of different facial expressions into one of the five basic human emotions, viz. Smile, Angry, Sad, Neutral and Surprise.

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

I. INTRODUCTION

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. The face can express emotion and people realize their feelings. In the past decade, much progress has been made to build computer systems to understand and use this natural form of human communication. Most such systems attempt to recognize a small set of emotional expressions, i.e. joy, surprise, anger, sadness, fear, and disgust.

Emotion analysis is more difficult with using any single one method that is main purpose of this paper. Facial action is analysis by using no of actions which is already available in open source. Using these combinations we can not analysis the exact expression. Another one method which can analysis the exact part of face to know the expression which is know the AAM features present the shape and locations of facial components (including mouth, eyes, brows, nose etc.) with the help of land marking. The facial components or facial feature points are extracted to form a feature vector that represents the face geometry.

In appearance-based methods, image filters, such as LGBP kernel are applied to either the whole-face or specific regions in a face image to extract and to know the exact feature of face to recognize. The best method which is proposed feature to analysis facial action recognition using the edges calculation and curve calculation of different parts from face. To get the best answer we combine this three features. Combining is more difficult to get single output so, we use SVM classifier and simple MKL algorithm for this. For the proposed feature we pass the output of first two features for the best emotion.

The facial expression recognition system consists of four steps. First is face detection phase that detects the face from image. Second is normalization phase that removes the noise and crop the face then normalize the face against brightness and pixel position. In third phase features are extracted and irrelevant features are eliminated. In the final step basic expressions are classified into six basic emotions like anger, fear, disgust, sadness, happiness, and surprise.

II. RELATED WORK

The superiority of action clues is over voice tone or spoken word. It show that rapid facial movements convey information about people effective states, moods, and personality[1]. Then the original emotional taxonomy was defined in the early 70s by Ekman [3] as a set of six basic universal facial emotions (anger, disgust, fear, happiness, sadness, and surprise). Another standard way is to describe the set of muscle movements that produce each facial expression. These movements are called facial action units (AUs), and the corresponding code is called Facial Action Coding System (FACS)[4]. This paper mainly focuses on AU detection (although emotion recognition.)
It is interesting to compare this database with our own databases. After that we choose to use the AAM introduced by Cootes et al. [6]. 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. [5] 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[7]. Finally, to deal with temporal aspects of AU display, they post process the classification outputs using a filtering and a thresholding technique. We use a 3-D face tracker called piecewise Bezier volume deformation tracker (PBVD), and Cohn et al. [10] apply a cylindrical head model to analyze brow AUs and head movements. Valstar et al. encoded face motion into motion history images. Recently, Koelstra et al. [11] have used dynamic texture [12]. Zhao and Pietikainen [13] apply volume local binary patterns (LBPs), which are the temporal equivalent of an LBP. An et al. [14] or Zhang and Ji [15] use facial points or component shapes with features such as crow-feet wrinkles and nasal-labial furrows. Chew et al. [16] use a constrained local model (CLM) to track the face and features and encode appearance using an LBP. SVMs are used to classify AUs. In [17]

III. SYSTEM OVERVIEW

The image will be systematically broken down and analyzed by the series of algorithms to determine the pixels that represent facial region. After this a cropping algorithm is applied to first crop the facial region. Next algorithm will detect lips from facial region. The automatic algorithm must correctly identify all pixels correctly included in lips while not incorrectly classifying the other regions as lips or lip colored coat.

Use of emotion recognition from digital images has a large opportunity and upcoming market. This is the primary reason to adopt a general and easy to apply approach towards the entire process. The approach is based on the assumption that there are not multiple faces in the image.

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. Project Aim

- The aim of this project 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.

D. 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.
E. 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].

F. 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. 3 Emotions](image)

<table>
<thead>
<tr>
<th>Happy</th>
<th>Sadness</th>
<th>Disgust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surprise</td>
<td>Fear</td>
<td>Anger</td>
</tr>
</tbody>
</table>

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.

For the training phase, the mouth sub-model is formed of 36 points, which contain points from the lower part of the face and from the nose shape. The eyes sub-model contains both eyes and eyebrows with five landmarks.

![Fig. 4. Landmarks for the eyes and mouth models.](image)

The training process of the AAM is illustrated in Fig. For both sub models, shapes are normalized using analysis. PCA is performed on these shapes, which results in the shape parameters with 95% of the variation represented by them. Thus, the eyes and mouth shapes can be synthesized by landmark it contains the eight vectors of variation in shapes, and other are the shape parameters.

B. Local Gabor Binary Pattern (LGBP):
It describes the neighboring changes around the central point, is a simple yet effective way to represent faces? It is invariant to monotone transformation and hence is somewhat robust to illumination changes.

![Fig.5 LGB Histogram](image)

(a) Original image, (b) Gabor pictures, and (c) concatenation of resulting histograms after applying LBP.

Each area of the face contains different useful information for AU detection. Thus, we choose to divide the face into several areas and compute one histogram per area.
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. From above process we can get binary image. Now we have to find the forehead from the binary image. The process of identification of image will be started from the middle of the image; one constraint of this is we have to find the continuous consecutive white pixels after uninterrupted black 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 1/4 position of the left eye box and ending point will be the 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 height which is nearest the present height and the program will give the nearest emotion as output. In case if the emotion information is not available in the predefined set of data based, then the written program will calculates the average height for each emotion in the database for all the people and then takes a decision according to the average height.

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 5(a) shows the flow chart for this module and Figure 5(b) shows the implementation of this 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 6(a) shows the flow chart for this module and Figure 6(b) shows the implementation of this module.

![Flowchart of Feature Extraction Module](image)

**Fig 7(a): Flowchart of Feature Extraction Module**

![Implementation of Feature Extraction Module](image)

**Fig 8(b): Implementation of Feature Extraction Module**

### C. Simple MKL Algorithm

#### Algorithm 1 SimpleMKL algorithm

set $d_n = \frac{1}{2}$ for $n = 1, \ldots, M$

while stopping criterion not met do

compute $J(d)$ by using an SVM solver with $K = \sum_n a_n K_n$

compute $\frac{d_n}{\gamma_n^k}$ for $n = 1, \ldots, M$ and descent direction $D(12)$

set $\mu = \text{argmax} d_n, J(0, d, D) = D$

while $J' < J(d')$ do (descent direction update)

$d = d', D = D'$

$v = \text{argmin} -d_n |D_n|, \gamma_n \neq v$ 

$d' = d + \gamma_n D_n, D'_n = D_n - D_n^k D'_n = 0$

compute $J'$ by using an SVM solver with $K = \sum_n a_n K_n$

end while

line search along $D$ for $\gamma \in [0, \gamma_{\text{max}}]$, (calls an SVM solver for each $\gamma$ trial value)

$d = d - \gamma D$

end while

### VI. Conclusion

In this paper, we have proposed and implemented a simple approach for recognition of the facial expression analysis. AAM can be used to landmark the main part from the face. Kernel LGBP used, to calculate the histogram and pass the output of both feature as an input to third kernel. 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.
REFERENCES