Abstract—A real-time face detection algorithm for locating faces in images and videos is one of the main research areas in computer vision. Here, we propose an efficient method for emotion recognition from facial expressions in static color images containing the frontal view of the human face. Our goal is to categorize the facial expression in the given image into four basic emotional states – Happy, Sad, Neutral and Surprise. Our method consists of three steps, namely face detection and localization, facial feature extraction and emotion recognition. First, face detection is performed using a novel skin-color extraction and extraction of location of facial component features such as the eye and the mouth analysis by using a knowledge based approach and projection. Next, the extraction of facial features distance is performed by employing an iterative search algorithm, on the edge information of the localized face region in gray scale. Finally, emotion recognition is performed by giving the extracted eleven facial features as input to different classifiers like PCA with Fuzzy C-means, Rule based classification and feed-forward neural network trained by back-propagation and analyzed based on accuracy.

Keywords—Artificial Neural Network, Fuzzy C-means, Expression classification, Facial features, Projection, Skin colour extraction.

I. INTRODUCTION

Computer vision is the branch of artificial intelligence that focuses on making computers to emulate human vision, including learning, making inferences and performing cognitive actions based on visual inputs, i.e. Images. Computer vision also plays a major role in Human Computer Intelligent Interaction (HCII) which provides natural ways for humans to use computer as aids and FACS(Facial Action Coding System)[1][2]. HCI will be much more effective and useful if computer can predict about emotional state of human being and hence mood of a person from supplied images on the basis of facial expressions. Mehrabian [3] pointed out that 7% of human communication information is communicated by linguistic language (verbal part), 38% by paralanguage (vocal part) and 55% by facial expression. Therefore facial expressions are the most important information for emotions perception in face to face communication.

For classifying facial expressions into different categories, it is necessary to extract important facial features which contribute in identifying proper and particular expressions. Recognition and classification of human facial expression by computer is an important issue to develop automatic facial expression recognition system in vision community.

This paper explains about an approach to the problem of facial feature extraction from a still frontal posed image classification and recognition of facial expression; emotion and mood of a person. Rule based classification, PCA with Fuzzy C-means clustering [4] and Feed forward back propagation neural network [5] is used as a classifier for classifying the expressions of supplied face into four basic categories like surprise, neutral, sad and happy. For face portion segmentation basic image processing operation like projection and skin color polynomial model are used. Facial features like eye, eyebrow and lip are extracted based on the projection and fine search method. Neural network is used as the most efficient method for facial expression.

II. DATA COLLECTION

Data acquisition for experimentation had done at NMIT campus and database is created for training and testing of classifiers. The database consists of 40 color images of 4 facial expressions (Happy, Sad, Surprise, and Neutral) posed by ten students, with four emotional adjectives. Few samples are shown in Figure 1.

![Figure 1. Few samples of facial expressions of persons](image-url)
III. PROPOSED METHOD

An overview of the proposed work contains three major modules: (i) face detection, (ii) facial feature extraction and (iii) emotion recognition from the extracted features as shown in Figure 2. The first step, namely face detection is performed on a color image containing the frontal view of a human subject.

![Diagram](https://example.com/diagram.png)

**Figure 2. Block Diagram for the Expression Recognition System**

The Real-time face expression recognition algorithm starts from detection of face in images and videos is done by the extraction of skin pixels based upon rules derived from a simple quadratic polynomial model [6]. The advantage of using this model if firstly saves computation and secondly, both extraction processes can be performed simultaneously in one scan of the image or video frame.

A. Face detection

Skin-color region extraction: For chromatic color space, each pixel is represented by two values, denoted by \( r \) and \( g \). The conversion from conventional RGB color space to chromatic color space is defined as follows:

\[
g = \frac{G}{R + G + B} \quad \quad r = \frac{R}{R + G + B}
\]

Where \( R, G, B \) denotes the intensities of pixels in red, green and blue channels, respectively.

The skin locus is formed on the \( r \_g \) plane is based on following two polynomials:

\[
f_{\text{upper}}(r) = -1.3767r^2 + 1.0743r + 0.1452
\]

\[
f_{\text{lower}}(r) = -0.776r^2 + 0.5601r + 0.1766
\]

The skin-color pixels, which fall within the area between these two polynomials, can be extracted by the following rule:

\[
R1: g > f_{\text{upper}}(r) \text{and} g > f_{\text{lower}}(r);
\]

In order to exclude the bright white pixels on the \( r \_g \) plane, falls around the point \((r,g)=(0.33,0.33)\).

\[
R2: W = (r - 0.33)^2 + (g - 0.33)^2 \leq 0.0004;
\]

The rules R1 and R2 are used for skin pixel extraction.

To improve the extraction result, the two simple rules are used.

\[
R3: R > G > B;
\]

\[
R4: R - G \geq 45;
\]

Rule R3 is defined in order to remove the blue pixel. And to remove the yellow-green pixels, Rule R4 is used.

\[
S = \begin{cases} 
1 \text{ if all } R1; R2; R3 \text{ and } R4 \text{ are true} \\
0 \text{ otherwise}
\end{cases}
\]

Where \( S=1 \) means the examined pixel is a skin pixel.

To improve the compactness of the formed skin-color region, the projection method is applied to extract the correct face part. The face part is cropped based on detected face centroid.

B. Facial Feature Extraction

Lip detection:

After a careful observation of the pixels of lips, it is observed that the colors of lips range from dark red to purple under normal light condition.

Two discriminate functions for lip pixels are:

\[
f_{\text{lower}}(r) = -0.776r^2 + 0.5601r + 0.1766
\]

\[
l(r) = -0.776r^2 + 0.5601r + 02123
\]

During the discrimination using \( l(r) \), the darker pixels on faces which are usually the eyes, eyebrows or nose holes, also need to be excluded from the discrimination.

The rule for detecting lip pixels is as follows:

\[
L = \begin{cases} 
1 \text{ if } f_{\text{lower}}(r) \leq l(r) \text{ and } R \geq 45 \text{ and} \\
G \geq 20 \text{ and } B \geq 45 \\
0 \text{ otherwise}
\end{cases}
\]

\( L = 1 \) indicates the examined pixel is a lip pixel.

The above rule may fail sometimes due to other darker features like eye. Testing of lip part can be done by vertical and horizontal projection[8].

Suppose \( I(x, y) \) be the input image. Vertical and horizontal projection vectors in the rectangle \([x1, x2] \times [y1,y2] \) are defined as –

\[
V(x) = \sum_{y=y1}^{y2} I(x, y)
\]

\[
H(y) = \sum_{x=x1}^{x2} I(x, y)
\]
Where \( V(x) \) is vertical projection and \( H(y) \) is horizontal projection.

The projection points are thicker only when there is more concentration of pixels. The distance between two corners of lip can be identified by applying horizontal projection. Similarly, distance between the upper and lower lip region can be identified by vertical projection.

**Eye detection:**

The eye components are extracted through a threshold operation (e.g. threshold = 20) on the histogram-equalized grayscale image converted from the original color image.

Testing of correct eye part can be done by cropping the eye part i.e. after detection of face part, apply the Edge detector (Sobel edge operator). For the binary image apply vertical projection \( V(x) \) and horizontal projection \( H(y) \). The peak value of projection will gives the location of an eye in a face. Crop the eye region [8] [9] based on the peak value. Total eleven features are extracted by eyes, eyebrows, and lip components.

**Facial Feature distance Ratio:**

Total eleven features extracted based on the projection mechanism.

i. Ratio of eye brow distance

\[
\frac{\text{Distance between point } b & c}{\text{Distance between point } a & d}
\]

ii. Ratio of number of skin and hair pixel is

\[
\frac{\text{No. of skin pixel}}{\text{No. of hair pixel}}
\]

iii. The distance between the corner of left eye and right corner eye is considered as (E). The distance between corners of mouth as (F).

Ratio of eye and lip distance is:

\[
\frac{\text{Distance E}}{\text{Distance F}}
\]

iv. Ratio of distance between left eye and eye brows to both(left and right) eye and eye brows:

\[
\frac{G}{G+H}
\]

The distance between left eye brow midpoint and left eye midpoint is considered as (G). The distance between right eye brow midpoint and right eye midpoint is considered as (H).

\[
\frac{H}{G+H}
\]

Ratio LE is:

Ratio of distance between right eye and eye brows to both (left and right) eye and eye brows:

The distance between left eye brow midpoint and left eye midpoint is considered as (G). The distance between right eye brow midpoint and right eye midpoint is considered as (H).

\[
\frac{H}{G+H}
\]

Ratio RE is:

**Ratio of Areas of features:**

The midpoint of a left eye brow is considered as (A1), the midpoint of an right eye brow is considered as (A2) and lower lip point is considered as (A3) and then find the area of an triangle A for the points A1, A2 and A3. The upper peak point of a left eye is considered as (B1), the upper peak point of a right eye is considered as (B2) and lower lip points is considered as (B3) and then find the area of a triangle B for the points B1, B2 and B3. The lower peak point of a left eye is considered as (C1), the lower peak point of a right eye is considered as (C2) and lower lip point is considered as (C3) and then finds the area of a triangle C for the points C1, C2 and C3.
C. Recognition of Emotion

After extracting the facial features, classification of expression done based on the following observations. As shown in Figure 4.

a. Surprise:

The eye brows are raised and curved. Eyelids are opened, white of the eye showing above and below. Jaw drops open and teeth are parted but there is not tension or stretching of the mouth.

b. Happiness (smile):

Corners of the lips are drawn back and up. Lips may or may not be parted, teeth exposed and Cheeks are raised.

c. Sadness:

Inner corners of the eyebrows are drawn up. Corner of the lips are drawn down and lower lip pouts out.

d. Neutral:

Inner corners of the eyebrows are normal. Corner of the lips are stretched. Eyelids are partially open.

The classification methods proposed in this paper are:

i. PCA with Fuzzy C-means clustering:

PCA [2] is widely used for dimensionality reduction method that finds the projection of data into a lower dimensional linear space such that the variance of the projected data is maximized, the features are reduced. Output of features values are given for Fuzzy C-means clustering.

ii. Rule based classification:

This classifier mainly called as - “if...then...” rules showed in Figure 5. Ranges of rules are generated based on features of an image. These feature rules are used for classification of an expression.

Rule: (Condition) → y

Where

- Condition is a conjunction of attribute tests (A1 = v1) and (A2 = v2) and ... and (An = vn)
- y is the class label

LHS: rule antecedent or condition
RHS: rule consequent.

But in Rule based classification, only limited number of features is used for decision making.

iii. Artificial Neural Network classifier:

Neural computing has re-emerged as an important programming paradigm that attempts to mimic the functionality of the human brain. This area has been developed to solve demanding pattern processing problems, like speech and image processing. These networks have demonstrated their ability to deliver simple and powerful solutions in areas that for many years have challenged conventional computing approaches. A neural network is represented by weighted interconnections between processing elements (PEs). These weights are the parameters that actually define the non-linear function performed by the neural network.
Back-Propagation Networks are most widely used neural network algorithm than other algorithms due to its simplicity, together with its universal approximation capacity. The back-propagation algorithm defines a systematic way to update the synaptic weights of multi-layer perceptron (MLP) networks [3][10].

The supervised learning is based on the gradient descent method, minimizing the global error on the output layer. The learning algorithm is performed in two stages feed-forward and feed-backward.

In the first phase the inputs are propagated through the layers of processing elements, generating an output pattern in response to the input pattern presented. In the second phase, the errors calculated in the output layer are then back propagated to the hidden layers where the synaptic weights are updated to reduce the error. As shown in Figure 6.

Eleven features of face values are given as an input to the neural network. Model uses an input layer, 11 hidden layers and 4 output layers.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, facial expressions are detected on Face Database developed at NMIT for frontal view facial images. Fig. 7 shows the results of facial feature extraction. Initially Face portion segmentation is done by skin polynomial model and the accurate face part extracted by

Cropped Facial image is divided into two regions based on the centre of an image and location of permanent facial feature by projection mechanism. Figure 7 shows localized permanent facial features which are used as input to classification.

The two phases of PCA with Fuzzy C-means clustering method are:

Training Phase: Features of all face expressions are given as input to a PCA. Reduced features obtained as output, is input to the Fuzzy C-means clustering. Four clusters obtained as output are neutral, smile, sad and surprise.

Testing Phase: Read unknown image. Calculate the features. Classification of face expression is done based on the closest distance between trained clusters.

For Rule based classification:

The ranges of all eleven features for all images are calculated. “ if..then rule” are applied for all range of values for classification. From the result it is observed that the output is database dependent.

The two phases of ANN classification method are:

Training Phase: In this work, supervised learning is used to train the back propagation neural network. The training samples are taken from NMIT database. This work has considered 18 training samples for all expressions. After getting the samples, supervised learning is used to train the network.

Testing Phase: The proposed system is tested with NMIT database. The Database consists of 40 sample images for all of the facial expressions. Fig. 8 shows the results of face expression.

Analysis:

Success rate of Expression recognition by ANN classifier is given in Table I.

ANN classifier will gives the maximum accuracy compare to Rule based classification and PCA with Fuzzy C-means clustering. The outcomes of different classification methods for different expressions are as shown in Table II.

Implementation details

The proposed approach is implemented on Windows 7 in Microsoft Visual Studio platform using VC++ language for programming.
Figure 7. Face feature extraction: a) input face part, b) face part extraction, c) cropped face part, d) left eye part, e) right eye part, f) lip part, g) combined features sample.

Figure 8. Expression recognition by ANN classifier

Table I. ANN classification

<table>
<thead>
<tr>
<th>Expression</th>
<th>Number of Image Experimented</th>
<th>Number of correct Recognition</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>10</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>Sad</td>
<td>10</td>
<td>9</td>
<td>90%</td>
</tr>
<tr>
<td>Surprise</td>
<td>10</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>Neutral</td>
<td>10</td>
<td>9</td>
<td>90%</td>
</tr>
</tbody>
</table>

Table II. Comparison of classifiers

<table>
<thead>
<tr>
<th>Expression</th>
<th>Fuzzy C-means(PCA)</th>
<th>Rule based classification</th>
<th>ANN classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>8(10)</td>
<td>10(10)</td>
<td>10(10)</td>
</tr>
<tr>
<td>Sad</td>
<td>4(10)</td>
<td>8(10)</td>
<td>9(10)</td>
</tr>
<tr>
<td>Surprise</td>
<td>8(10)</td>
<td>10(10)</td>
<td>10(10)</td>
</tr>
<tr>
<td>Neutral</td>
<td>6(10)</td>
<td>8(10)</td>
<td>9(10)</td>
</tr>
</tbody>
</table>
V. CONCLUSION

Faces are tangible projector panels of the mechanisms which govern our emotional and social behaviors. The automation of the entire process of facial expression recognition is a highly intriguing problem. The solution to which would be enormously beneficial for fields as diverse as medicine, law, communication, education, and computing. The proposed method uses the skin based polynomial model reduces the computation time, features are very effective for classification of the expression. Finally comparison of PCA based Fuzzy C-means clustering, rule based classification and back propagation neural network is made and it is found that the back propagation neural network is efficient for recognition facial expression as shown in Table II.

Acknowledgement

The authors acknowledge Prof N R Shetty, Director, Nitte Meenakshi Institute of Technology and Dr H C Nagaraj, Principal, Nitte Meenakshi Institute of Technology for providing the support and infrastructure to carry out our research. And Shilpa Ankalaki, Mtech, CSE, Nitte Meenakshi Institute of Technology for helping in back propagation neural network classification.

REFERENCES


