International Journal of Research
Available at https://journals.pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 04
April 2019

Expression Recognition with Appearance Based Features of Facial Landmarks

B.Priyanka[1], P. Srinidhi [2], Ch. Sathwik[3]

[1]Assistant Professor: Dept. of ECE, Sreenidhi Institute of Science and Technology, Hyderabad
[2]UG Scholar: Dept. of ECE, Sreenidhi Institute of Science and Technology, Hyderabad
[3]UG Scholar: Dept. of ECE, Sreenidhi Institute of Science and Technology, Hyderabad

ABSTRACT:-

In this paper, Local Binary Patterns (LBP) is used for Facial Expression Recognition (FER). The concept of LBP is based on the information that is present in color images of face. Multi-linear image analysis can be done in different color spaces using LBP and it can be seen that the color content gives additional information about faces which can lead to an efficient FER. Using LBP, the components present in various color spaces such as RGB, YCbCr and CIELuv or CIELab, are made into two dimensional binary's using multi-linear algebra and concepts of binary's and then Log-Gabor filters are used to extract these features. For the selection of the features, mutual information quotients method is used. Multiclass linear discriminate analysis classifiers are used to classify the features that were extracted.

Keywords: **LBP, CIELab, CIELuv, FER, Log-Gabor filters**

I. INTRODUCTION

A goal of the Human-Computer-Interaction (HCI) systems is to enhance the communication between the computer and user by making it user friendly and user's needs. In [1] proposes the important of the automatic facial expression recognition (FER) plays an important role in the HCI system and it has been studied extensively over the past twenty years. Since the late 1960s use of the facial expression for measuring people's emotions has dominated psychology. Paul Ekman reawakened the study of emotion by linking expressions to a group of basic emotions (i.e., anger, disgust, fear, happiness, sadness and surprise) [2]. The research study by Megrabian [3] has indicated that 7% of the communication information is transformed by linguistic language, 55% by facial expression and 38% by paralanguage in human face-to-face communication. It shows that facial expression provides a large amount of information in human communication. Many approaches have been proposed for the FER in the past several decades [1],[4]. Current state-of-art techniques mainly focused on the gray-scale image features [1], rarely it consider the color image feature [5]-[7].

Color feature mat provides more robust classification results. Research reveals that the color information enhances the face recognition and image retrieval performance [8]-[11]. In [8] it was first reported in that taking color information enhance the reorganization rate as compared with the same scheme using only the luminance information. Liu and Liu in [10] proposed a new color space for face recognition. In [11] Young, Man and Plataniotis demonstrated that the facial color cues express the improved face recognition performance using the low-resolution face image. The RGB color binary has enhanced the FER performance but it does not consider the different illumination was reported in [7]. Recent research shows the improved performance by embedding the color components. The capability of the color information in the RGB color space in terms of the recognition performance depends upon the type and angle of the light source, often making recognition impossible. Thus the RGB may not be always be the most desirable space for processing color information. In [12] this issue can be addresses using perceptually uniform color system. In this paper a novel Local Binary Pattern (LBP) for FER is introduced which provides the information about the color facial images and investigates performance contained in the color facial images and investigates performance in perceptual color space under slight variation in the illumination.

This paper is organized as follows Section II provides the brief detail about the components of the FER systems used for this investigation. Section III defines and examines the binary-based representation of color facial images in different color space and explains the proposed LBP technique. Section IV presents the experimental result and Section V presents final conclusion.

II. CONSTRUCTION OF AN IMAGE-BASED FER SYSTEM

The principal approaches (i.e., image-based and model based) to FER using static images are explained in [1]. Image-based extract feature form the image without extensive knowledge about the object of interest, which are fast and simple. The model based methods attempt to recover the volumetric geometry of the scene, which are slow and complex [1]. Geometric features present the shape and location of facial components (including mouth, eyebrows, eyes and nose). The facial feature points or facial components are obtained from the feature vector that represents the face geometry.

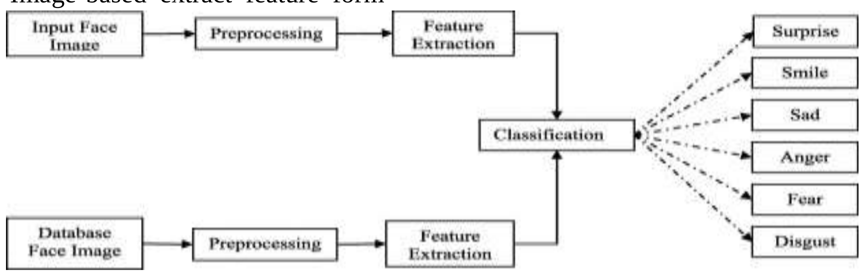


Fig 1: System Level Diagram

The appearance feature can be taken from either the whole face or specific regions in a face image. This paper focused on the static color image and a holistic technique of the image-based method is used for feature extraction. Image based FER systems consist of several components or modules, including face detection and normalization, feature extraction, classification and feature selection. The system level diagram of FER system shown in Figure 1 The following section will describe briefly about YCbCr, CIELab, and CIELuv [13].

A. Face Detection and Normalization

In this module is to obtain face images, which have normalized intensity, are uniform in shape and size and depict only the face region. Face area of an image is detected using the Viola-Jones method based on the Haar-like features and the AdaBoost learning algorithm [14]. The Viola and Jones method is an object detection algorithm provides competitive object detection in the real-time. Features used by Viola and Jones are derived from pixels selected from rectangle area imposed over the picture and exhibit high sensitivity to the vertical and horizontal lines. After face detection the image is scaled into some size (e.g.,64 × 64 pixels). Color values in the face image are then normalized with respect to RGB values of the image.

Color normalization is used to reduce the lighting effect because the normalization process is actually a brightness elimination process. Input image of $N_1 \times N_2$ pixels represented in the RGB color space,

$$X = \{X^{n_3}[n_1, n_2] \,|\, 1 \leq n_1 \leq N_1, 1 \leq n_2 \leq N_2, 1 \leq n_3 \leq 3\},$$

the normalized values, $X^{n_3}_{norm}[n_1, n_2]$, are defined by

$$X^{n_3}_{norm}[n_1, n_2] = \frac{X^{n_3}[n_1, n_2]}{\sum_{n_3=1}^{3} X^{n_3}[n_1, n_2]}$$

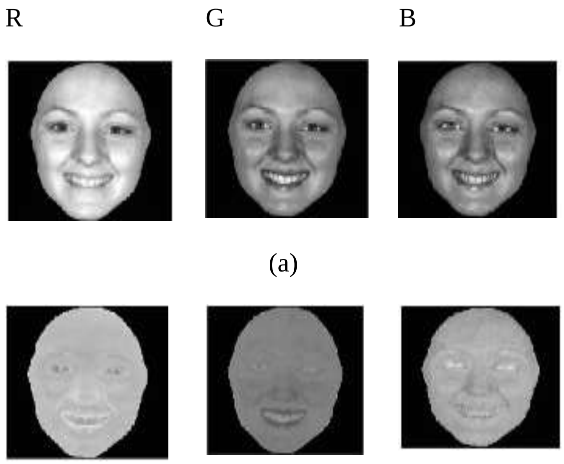
----------(1)

where $X^{n_3}_{norm}[n_1, n_2]$ for $n_3 = 1,2,3$ corresponding to red, green, and blue (or R, G, and B) components of the image X.

It is obvious that

$$\sum_{n_3=1}^{3} X^{n_3}_{norm}[n_1, n_2] = 1$$

----------(2)



(b)

Fig 2: Facial expression images: (a) the original color components (b) the normalized color components.

B. Feature Extraction

Feature extraction have been studied and compared in terms of their performance, including principal components analysis, independent components analysis, linear discriminates analysis (LDA), the Gabor filter bank, etc. In [1] presents the Gabor filter has better performance than the rest. The Gabor filters model the receptive field profiles of cortical simple cells quite good [1], [15]. Gabor filter have two major drawbacks i.e., the maximum bandwidth of Gabor filter the maximum bandwidth is limited to approximately one octave, and the Gabor filter are not optimal to achieve broad spectral information with the maximum spatial localization [16]. The Gabor filter are band pass filters, which may suffers from loss of the low and the high-frequency information is reported in [17]. To overcome the bandwidth limitation of the traditional Gabor filter, Field proposed Log-Gabor filter [17]. Response of the Log-Gabor filter, is Gaussian when viewed on a logarithmic frequency scale instead of a linear. It allows more information to be capture in the high-frequency area with desirable high pass characteristics. A bank of 24 Log-Gabor filter is employed to extract the facial features. Polar form of 2-D Log-Gabor filters in frequency domain is given by

$$H(f,\theta) = exp\left\{\frac{-\left[\ln(\frac{f}{f_0})\right]^2}{2\left[\ln(\frac{\sigma_f}{f_o})\right]^2}\right\} exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\}$$

----------(3)

where $H(f,\theta)$ is frequency response function of the 2-D Log-Gabor filter, f and θ denotes the frequency 2-D Log-Gabor filters, f and θ denotes the frequency and the phase/angle of the filter.f_o is the filter center frequency and θ_0 the filter's direction. The constant σ_f defines the radial bandwidth B in octaves and the constant σ_θ angular bandwidth $\Delta\Omega$ in radians.

$$B = \sqrt[2]{\frac{2}{ln2} \times \left|ln\left(\frac{\sigma_f}{f_0}\right)\right|}, \Delta\Omega = 2\sigma_\theta\sqrt{\frac{2}{ln2}}$$

----------(4)

In this paper describes here, the ratio σ_f/f_0 is kept constant for varyingf_0, B is set to one octave and the angular bandwidth is set to one octave and the angular bandwidth is set to $\Delta\Omega = \pi/4$ radians. σ_f is be determined for a varying value of f_0. Six scales and four orientations are implemented to extract features from face images. It leads to 24 filter transfer functions representing different scales and orientations. Image filtering is performed in the frequency domain making the process faster compared with the special domain convolution. After 2-D fast Fourier transform (FFT) into the frequency domain, the image arrays, X, are changed into the spectral vectors X and multiplied by the log-Gabor transfer functions $\{H_1, H_2, \dots H_{24}\}$ producing 24 spectral representations for each image [17]. Spectra are then transformed block to the spatial domain via the 2-D inverse FFT. In this process results are obtained in the large numbers which are not suitable to build the robust learning models for classifications.

C. Feature Selection

Feature selection module have a distinctive features of image and it help us to improve the performance of the learning models by removing the most relevant and redundant features from the feature space. Optimum features are selected using minimum redundancy maximum relevance algorithm based on mutual information (IM). In [18] presents a mutual information quotient (MIQ) method for feature

International Journal of Research
Available at https://journals.pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 04
April 2019

selection and adopted to select the optimum features. As per the MIQ features selection if a feature vector has expression randomly or uniformly distributed in different classes and its MI with these classes is zero. If a feature vector is different from the other features for different classes, it will have large MI. Let F denotes the feature space; C denotes a set of classes $C = \{c_1, c_2, \dots, c_k\}$, and v_t denotes the vector of N observation for that feature.

$$v_t = [v_t^1, v_t^2, \dots, v_t^N]^T \quad \text{----------(5)}$$

where v_t is an instance of the discrete random variable V_t. The MI between features V_t and V_s is given by

$$I(V_t; V_s) = \sum_{v_t \in V} \sum_{v_s \in V_s} p(v_t, v_s) log \frac{p(v_t, v_s)}{p(v_t)p(v_s)}$$

----------(6)

where $p(v_t, v_s)$ is the joint probability distribution function (PDF) of V_t and V_s, $p(v_t)$ and $p(v_s)$ are the marginal PDFs of V_t and V_s, for $1 \leq t \leq N_f$,$1 \leq s \leq N_f$, and N_f is the input dimensionality, which equals the number of features in the dataset. The Mi between the V_t and C can be represent by entropies [19]

$$I(v_t; C) = H(C) - H\big(C|v_t\big) \quad \text{----------(7)}$$

where

$$H(C) = -\sum_{i=1}^{k} p(C_i) \log\big(p(C_i)\big)$$

----------(8)

$$H(C\big| V_t = -\sum_{i=1}^{k} \sum_{v_t \in V_t} p(C_i, v_t) \log\big(p(C_i|v_t)\big)$$

----------(9)

where H(C) is the entropy of C, $H(C\big|V_t)$ is the conditional entropy of C on V_t, and k is the numbers of classes (for six expression, $k = 6$). The features (V_d) for desired feature subset, S, of the form $(S; c)$ where $S \subset F$ and $c \in C$ is selected based on solution of following problems:

$$V_d = arg \max_{V_t} \left\{ \frac{I(V_t; C)}{\frac{1}{|S|} \sum I(V_t; V_s)} \right\} V_t \in \bar{S}, V_s \in S$$

----------(10)

where $\bar{S}$ is the complement features subset of S, $|S|$ is the number of features in subset S and $I(V_t; V_s)$ is the MI between the candidate features (V_t) and the selected feature and intra-class features is maximized. MI between the selected feature and inter-class features is minimized. These features are used for emotion classification.

D. Classification

The LDA classifier was studied for the same database and provides the better result than other classifiers [5]. The selected features using the aforementioned MIQ techniques are classified by a multiclass LDA classifier. In [20] proposes a natural extension of Fisher linear discriminant that deals with more than two classes which uses multiple discriminant analysis. Projection from the high dimensional space to a low-dimensional space and the transformation descried to maximize the ratio of inter-class scatter (S_b) to the intra-class (S_w) scatter. The S_b can be viewed as the sum of square of distance between each class mean and the mean of all training samples. S_w can be regarded as the average class-specific covariance. Intra-class (S_w) and inter-class (S_b) matrices for feature vectors (X^f) are given by

$$S_b = \sum_{i=1}^{N_c} m_i \big(X_{\mu_i}^f - X_\mu^f\big)\big(X_{\mu_i}^f - X_\mu^f\big)^2$$

----------(11)

$$S_w = \sum_{i=1}^{N_c} \sum_{X^f \in c_i} \big(X^f - X_{\mu_i}^f\big)\big(X^f - X_{\mu_i}^f\big)^T$$

----------(12)

where N_c is the number of classes (i.e., for six expression, $N_c = 6$), m_i is the number of training samples for each class. c_t is the class label, $X_{\mu_i}^f$ is the mean vector for each class samples (m_i), and X_μ^f is the total mean vector over all training sample (m) defined by

$$X^f_{\mu_i} = \frac{1}{m_i}\sum_{X^f \in C_i} X^f \quad \text{----------(13)}$$

$$X^f_{\mu} = \frac{1}{m}\sum_{i=1}^{N_c} m_i X^f_{\mu_i} \quad \text{----------(14)}$$

After obtaining S_w and S_b based on Fisher's criterion the linear transformation, W_{LDA}, can calculated by solving the generalized Eigen value (λ) problem

$$W^T_{LDA} S_b = \lambda W^T_{LDA} S_w \quad \text{----------(15)}$$

The transformation W_{LDA} is given the classification can be performed in the transformed space based on preformed distance measure such as the Euclidean distance, $\|\bullet\|$. The instance, X^f_n, is classified to

$$C_n = arg\min_i \left\| W_{LDA} X^f_n - W_{LDA} X^f_{\mu_i} \right\| \quad \text{----------(16)}$$

where c_n denotes the predicted class-label for X^f_n and X^f_n is the centriod of the ith class.

III. COLOR SPACES

Several image representation models in the color space used for image processing [21]. The RGB color space is used in the image processing and pattern recognition systems. Color space can be used to generate the other alternative color formats including: YC_bC_r, CIELab, and CIELuv. The YC_bC_r color space is a ditial and offset version of the YUV used by the NTSC or the PAL television/video standard [13]. Conversion function between RGB and YC_bC_r is defined by

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.774 & -74.159 & 111.934 \\ 111.958 & -93.751 & -18.207 \end{bmatrix} \begin{bmatrix} X^1_{norm} \\ X^2_{norm} \\ X^2_{norm} \end{bmatrix}$$

----------(17)

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \quad \text{----------(18)}$$

where $X^{n_3}_{norm}$, $1 \le n_3 \le 3$, is defined in the (1).

To Convert the PGB to perceptual color spaces (CIELab or CIELuv), the RGB is first converted to XYZ color space, which than converted to perceptual color spaces. Components L are same for both CIELab and CIELuv color spaces. Conversion procedure is as follows in [13]

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{bmatrix} \begin{bmatrix} X^1_{norm} \\ X^2_{norm} \\ X^3_{norm} \end{bmatrix} \quad \text{----------(19)}$$

$$L = \begin{cases} 116 \times \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \frac{Y}{Y_n} > 0.008856 \\ 903 \times \left(\frac{Y}{Y_n}\right), & \frac{Y}{Y_n} \le 0.008856 \end{cases}$$

----------(20)

$$a = 500 \times \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right) \quad \text{----------(21)}$$

$$b = 200 \times \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right) \quad \text{----------(22)}$$

where X_n, Y_n, and Z_n are the reference white tri-stimulus value which are defined in CIE chromaticity diagram [21] and

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & t > 0.008856 \\ 7.787 \times t + \frac{16}{116}, & t \le 0.008856 \end{cases}$$

----------(23)

for u and v color components, the conversion is defined by

$$u = 13 \times L \times (u' - u'_n) \quad v = 13 \times L \times (v' - v'_n)$$

----------(24)

The equation for v' and u' are given below

$$u' = \frac{4X}{X + 15Y + 3Z} \quad \text{----------(25)}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \quad \text{----------(26)}$$

The quantities u'_n and v'_n are the (u', v') chromaticity coordinates of a specific white object defined by

$$u'_n = \frac{4X_n}{X_n+15Y+3Z} \quad v'_n = \frac{9Y}{X_n+15Y_n+3Z}$$

----------(27)

IV. EXPERIMENTAL RESULTS

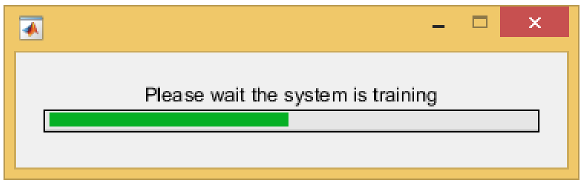


Fig 3: Training of the Dataset

Fig 4: Select Image from User

Fig 5: Input Image of Expression for Testing

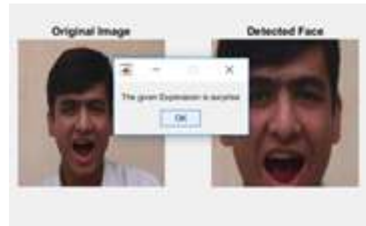


Fig 6: Obtained result as surprise

V. CONCLUSION

Local Binary Pattern facial expression approach is proposed in this paper, the present is evaluated with the Indian face data base under different color transformations and resolution and it is shown that the CIE-Lab and CIE-LUV transformation outperforms the highest recognition rate and the proposed method is also compared against the conventional Gabor based approaches which fall short of more than 2 % in recognition rate .This work can be further extended with more frontal facial image data base for more accurate results.

REFERENCES

[1] B. Fasel and J. Luettin, “Automatic facial expression analysis: A survey,”*Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, 2003.

[2] P. Ekman, E. T. Rolls, D. I. Perrett, and H. D. Ellis, “Facial expressions of emotion: An old controversy and new findings discussion,” *Phil. Trans. Royal Soc. London Ser. B, Biol. Sci.*, vol. 335, no. 1273, pp. 63–69, 1992.

[3] A. Mehrabian, *Nonverbal Communication*. London, U.K.: Aldine, 2007.

[4] M. Pantic and I. Patras, “Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences,” *IEEE Trans. Syst., Man, Cybern. B*, vol. 36, no. 2, pp. 433– 449, Apr. 2006.

[5] J. Wang, L. Yin, X. Wei, and Y. Sun, “3-D facial expression recognition based on primitive surface feature distribution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1399–1406.

[6] Y. Lijun, C. Xiaochen, S. Yi, T. Worm, and M. Reale, “A high-resolution 3-D dynamic facial expression database,” in *Proc. 3rd Int. Conf. Face*

Gesture Recognit., Amsterdam, The Netherlands, Sep. 2008, pp. 1–6.

[7] S. M. Lajevardi and Z. M. Hussain, "Emotion recognition from color facial images based on multilinear image analysis and Log-Gabor filters," in *Proc. 25th Int. Conf. Imag. Vis. Comput.*, Queenstown, New Zealand, Dec. 2010, pp. 10–14.

[8] L. Torres, J. Y. Reutter, and L. Lorente, "The importance of the color information in face recognition," in *Proc. Int. Conf. Imag. Process.*, vol. 2. Kobe, Japan, Oct. 1999, pp. 627–631.

[9] P. Shih and C. Liu, "Comparative assessment of content-based face image retrieval in different color spaces," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 7, pp. 873–893, 2005.

[10] Z. Liu and C. Liu, "A hybrid color and frequency features method for face recognition," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1975– 1980, Oct. 2008.

[11] C. J. Young, R. Y. Man, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1217–1230, Oct. 2009.

[12] M. Corbalán, M. S. Millán, and M. J. Yzuel, "Color pattern recognition with CIELab coordinates," *Opt. Eng.*, vol. 41, no. 1, pp. 130–138, 2002.

[13] G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data, and Formulae* (Wiley Classics Library). New York: Wiley, 2000.

"Feature selection based on mutual