RESEARCH ARTICLE OPEN ACCESS

Recognition of Face Expression using Color Space

Brruna A , Siva Pradeepa S
1Brruna A. Author is currently pursuing M.E (Software Engineering) in Vins Christian college of Engineering. e-mail:brrunashiny@gmail.com.
2Siva Pradeepa S, Asst. professor, Department of Computer science and engineering, Vins Christian college of Engineering, Chunkankadai.

Abstract:
Face expression recognition can be stated as 'identifying the expression of an individual from images of the face'. Most of the existing systems of facial expression recognition focus on gray scale image features. This paper describes the novel approaches for effectively recognizing the facial expressions. In facial expression recognition (FER) framework, initially the face region of the image is detected using Ada boost learning algorithm. The RGB color image is transformed to another color space (YCbCr) to further improve face recognition performance. From this, the features are extracted by using Log-Gabor filter which results in a large amount of feature arrays. Furthermore, to improve the classification accuracy, the most desirable features are selected by using mutual information quotient (MIQ) technique. At last the multiclass linear discriminant analysis (LDA) classifier is used to classify these features. The experimental results evaluate the performance of using this framework under varying illumination and in low resolution images.
***Keywords*—** Log-Gabor filter, linear discriminant analysis, mutual information quotient, color space, facial expression recognition.

I. INTRODUCTION

Recently the interaction between human and computers are improved through many concepts. One of the concept used to improve Human-Computer Interaction (HCI) is automatic recognition of facial expressions. It has been developed over the past twenty years. The main application of Facial Expression Recognition (FER) framework is in the field of computer games, border security systems, machine vision, user profiling for customer satisfaction etc. Normally people's emotions are measured and classified into a six set of groups such as anger, disgust, fear, happy, surprise and sadness. The researcher Mehrabian made some study on human communication and he concluded that, in the communication information 7% is transferred by linguistic language, 38% by paralanguage and 55% by facial expressions. This proves that facial expressions play an important role in human communication.

Various methods have been introduced and used for FER system in recent years. They mostly focus on gray scale image features [2] than color image features [4], [5]. More efficient classification results can be obtained by considering the color feature data. The color information in face images improve the accuracy of recognition rate. The importance of color information is briefly explained and compared with using only luminance information [6]. A new color space for face recognition is introduced in [8] by Z. Liu and C. Liu from which, the color components are used to improve the performance of face retrieval. The facial color cues desirably increased the effectiveness of face recognition using low resolution images of face [9] which is stated by C.J. Young, R.Y. Man, and K.N. Plataniotis. However the RGB color improved FER performance, it did not perform well in different illuminations of images.

The performance of RGB color space lies on the type and angle of light source and sometimes it will make recognition as impossible. Therefore the processing of information in RGB color space may not always yield good results. To solve this problem the color components in RGB color space are transformed into various other color spaces like YCbCr, CIELab, CIELuv etc.

The two general approaches for Facial Expression Recognition (FER) using static images are of image based approach and model based approach. Without prior knowledge about the object of interest, the image based methods extract the features from the static images where as it is of fast and simple. The model based methods focus on the volumetric geometry of the image which is typically slow and complex. The geometric features consist of information about the location and shape of facial features (such as eyes, mouth, nose, and eyebrows). The appearance based features contains the appearance changes of the face image (such as bulges, wrinkles, and furrows) are extracted either from the face or sub regions of the faces by image filters.

This paper is presented as follows: the section II demonstrates the details of all components or modules used in the FER system. Section III describes the process of conversion of color transform. Section IV describes the experimental results and the section V gives some final conclusions for the whole system.

II. COMPONENTS OF FER SYSTEM

The holistic techniques of image based method using static images are applied for this system performance. This gives rise to various stages such as face detection, feature extraction, feature selection, and classification for FER system under varying illuminations. The figure 1 gives the system level design for FER system.

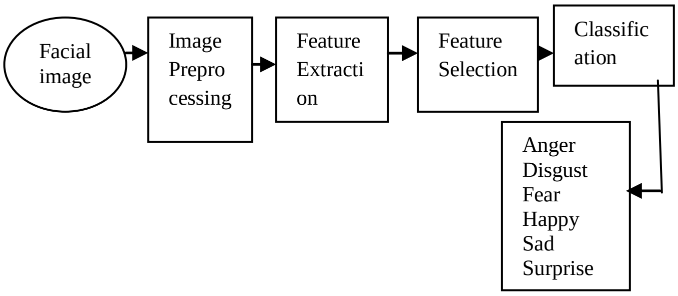

Figure 1: System level design

A. Image Preprocessing

The major goal of this module is to detect the face regions of the image and which has to be in uniform size and scale. The Ada Boost learning algorithm proposed is applied to derive the exact face area of image. This algorithm will detect the objects more accurately in the real time. A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called 2-rectangle feature.

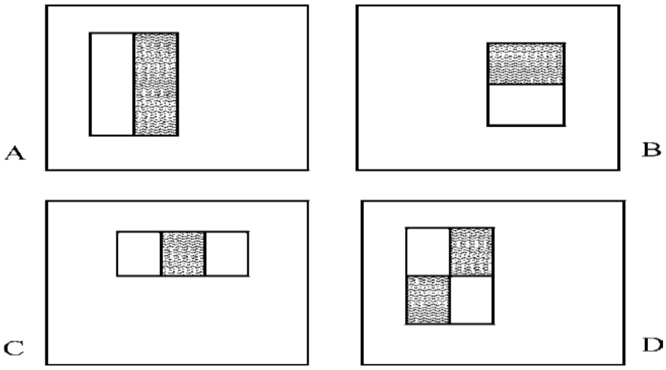

Figure 2: Haar-like rectangular features

The Haar-like features also includes 3-rectangle features and 4-rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region. A face image represented in the RGB color space is first translated, rotated, and rescaled to a fixed template, yielding the corresponding aligned face image. All facial images used in experiments were manually cropped from original images based on the locations of the two eyes.

The eye coordinates are those supplied with the original data set. Each cropped facial image was rescaled to the size of 64 × 64 pixels. After alignment, each of the facial images with a size of 64 × 64 is divided into the 64 different face local regions to extract the features. Thus, the size of each local region is 8 × 8 pixel. Subsequently, the aligned color image is converted into an image represented in another color space. Note that not only conventional linear or nonlinear color spaces but also new color spaces devised for the purpose of FER (eg : normalized color space proposed) can be used for color space conversion.

B. Feature Extraction

After the face detection step, human-face features are extracted from images. Directly using these features for face recognition have some disadvantages. First, each patch usually contains over 1000 pixels, which are too large to build a robust recognition system. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately.

Gabor filters are a traditional choice for obtaining localized frequency information. They offer the best simultaneous localization of spatial and frequency information. However they have two main limitations. The maximum bandwidth of a Gabor filter is limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization.

An alternative to the Gabor function is the log-Gabor function proposed by Field [1987]. Field suggests that natural images are better coded by filters that have Gaussian transfer functions when viewed on the logarithmic frequency scale whereas Gabor functions have Gaussian transfer functions when viewed on the linear frequency scale. Gabor filters can be constructed with arbitrary bandwidth and the bandwidth can be optimized to produce a filter with minimal spatial extent.

In frequency domain, the polar form of 2-D Log-Gabor filters is given by

$$H(f,\theta) = \exp\{\frac{-[\ln(\frac{f}{f_0})]^2}{2[\ln(\frac{\rho_f}{f_0})]^2}\}\exp\{\frac{-(\theta-\theta_0)^2}{2\rho_\theta^{\ 2}}\} \quad \ldots\ldots (1)$$

Where H(f, θ) is the frequency response function of the 2-D Log -gabor filters, f and θ denotes the frequency and the phase/ angle of the filter respectively. f_0 is the center frequency and θ_0 is the filter's direction. The constant ρ_f defines the radial bandwidth B in octaves and the constant ρ_θ defines the angular bandwidth $\Delta\Omega$ in radians

$$B = 2\sqrt{\frac{2}{\ln 2}} \times \left|\ln(\frac{\rho_f}{f_o})\right| \quad \ldots\ldots\ldots (2)$$

$$\Delta\Omega = 2\rho_\theta\sqrt{\frac{2}{\ln 2}} \quad \ldots\ldots\ldots (3)$$

the ratio ρ_f/f_0 is kept constant for varying f_0, *B* is set to one octave and the angular bandwidth is set to $\Delta\Omega = \pi/4$ radians. This left only ρ_f to be determined for a varying value of f_0. Six scales and four orientations are implemented to extract features from face images. This leads to 24 filter transfer functions representing different scales and orientations. The image filtering is performed in the frequency domain making the process faster compared with the spacial domain convolution. After the 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 $\{H_1, H_2, \ldots, H_{24}\}$, producing 24 spectral representations for each image[12]. The spectra are then transformed back to the spatial domain via the 2-D inverse FFT. This process results in a large number of the feature arrays, which are not suitable to build robust learning models for classification.

C. Feature Selection

Since the number of features resulted from the feature extraction process is fairly large, the feature selection module is required to select the most distinctive features. In other words, the feature selection module helps to improve the performance of learning models by removing most irrelevant and redundant features from the feature space. The optimum features are selected using minimum redundancy maximum relevance algorithm based on mutual information (MI). The mutual information quotient (MIQ) method is employed for feature selection. According to MIQ feature selection criteria, if a feature vector has expressions randomly or uniformly distributed in different classes, its MI with these classes is zero. If a feature vector is strongly different from 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,….*cj* }, and v_r denotes the vector of *M* observations for that feature

$$v_r = [v_r^1, v_r^2, \ldots\ldots v_r^M]^T \quad \ldots\ldots\ldots (4)$$

Where v_r is an instance of the discrete random variable V_r. The MI between features V_r and V_s is given by

$$I(V_r;V_s) = \sum_{v_r \in V_r}\sum_{v_s \in V_s} p(v_r, v_s)\log\frac{p(v_r, v_s)}{p(v_r)p(v_s)} \quad \ldots. (5)$$

Where $p(v_r, v_s)$ is the joint probability distribution function (PDF) of V_r and V_s , $p(v_r)$ and $p(v_s)$ are the marginal PDFs of V_r and V_s , respectively, for $1 \le r \le N f$, $1 \le s \le N f$, and $N f$ is the input dimensionality, which equals the number of features in the dataset. The MI between V_r and *C* can be represented by entropies

$$I(V_r;C) = H(C) - H(C \mid V_r) \quad \ldots\ldots. (6)$$

Where

$$H(C) = -\sum_{i=1}^{j} p(c_i)\log(p(c_i)) \quad \ldots\ldots\ldots (7)$$

$$H(C \mid V_r) = -\sum_{i=1}^{j}\sum_{v_r \in V_r} p(c_i, v_r)\log(p(c_i \mid v_r)) \quad \ldots\ldots\ldots (8)$$

where *H(C)* is the entropy of *C*, *H(C*| v_r*)* is the conditional entropy of *C* on V_r , and *j* is the number of classes (for six expressions, j = 6). The features (V_d) for desired feature subset, S, of the form (S; c) where $S \subset F$ and $c \in C$ are selected based on solution of following problems:

$$V_d = \arg\max \left\{ \frac{I(V_r;C)}{\frac{1}{|S|}\sum I(V_r;V_s)} \right\} V_r \in \bar{S}, V_s \in S \quad \text{……. (9)}$$

where $\bar{S}$ is the complement feature subset of S, $|S|$ is the number of features in subset S and I (v_r;v_s) is the MI between the candidate feature (v_r) and the selected feature (v_s). Based on (10), the MI between selected feature and intra-class features is maximized whereas the MI between the selected feature and inter-class features is minimized, respectively. These features are used for emotion classification.

D. Classification

The LDA classifier is used to classify the selected features of the image. It classifies the six different types of expressions. The transformation from a high dimensional space to low dimensional space is the one that maximizes the ratio of inter-class scatter (S_b) to the intra-class scatter (S_w) among several classes. The inter-class (S_b) and the intra-class (S_w) matrices for feature vector (X^f) are given by

$$S_b = \sum_{i=1}^{N_c} m_i (X^f_{\mu_i} - X^f_{\mu})(X^f_{\mu_i} - X^f_{\mu})^T \quad \text{.……. (10)}$$

$$S_w = \sum_{i=1}^{N_c} \sum_{X^f \in c_i} (X^f - X^f_{\mu_i})(X^f - X^f_{\mu_i})^T \quad \text{……. (11)}$$

where Nc is the number of classes (i.e., for six expressions, Nc = 6). Then m_i is the number of training samples for each class, ci is the class label, $X^f_{\mu_i}$ is the mean vector for each class samples (m_i), and X^f_{μ} is total mean vector over all training samples (m).

After obtaining the inter-class and intra-class values, the transformation W_{LDA} is computed. The predefined Euclidean distance measure is used to perform the classification. The new instance X^f_n is classified as:

$$c_n = \arg\min_i \left\| W_{LDA} X^f_n - W_{LDA} X^f_{\mu_i} \right\| \quad \text{……. (12)}$$

Where c_n is the predicted class label for X^f_n and the mean vector for each class samples $X^f_{\mu_i}$ is the centroid of i th class.

III. COLOR SPACE CONVERSION

The *YCbCr* color space was developed as part of the ITU-R Recommendation for digital video standards and television transmissions. It is a scaled and offset version of the *YUV* color space. The *YUV* is used by the NTSC or the PAL television/video standard. In *YCbCr*, the *RGB* components are separated into luminance (*Y*), chrominance blue (*Cb*) and chrominance red (*Cr*) by using (13). The *Y* component has 220 levels ranging from 16 to 235, while the *Cb,Cr* components have 225 levels ranging from 16 to 240:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \quad \text{……… (13)}$$

where the *R, G,B* values are scaled to [0,1].

IV. EXPERIMENTAL RESULTS

In this paper, the stock gallery static images are used for recognizing the facial expressions. There are six types of facial expressions such as happy, anger, disgust, fear, surprise, sadness are used.

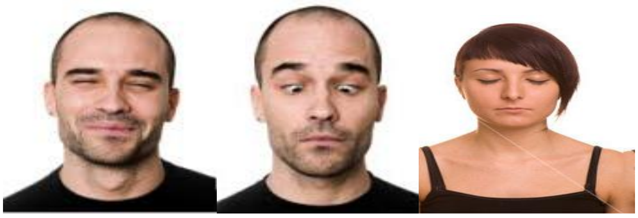

Figure 3: Sample input images

The input RGB color image is transformed into another color space as YCbCr and the face area of the image is detected and scaled to a size of (64 × 64). Then the features are extracted and classified with an expression.

Figure 4: Sample detected face images

V. CONCLUSION

In summary, it has been concluded that the proposed technique is giving much better results than the existing ones. Based on FER framework, the RGB color images were first transformed to YCbCr color spaces after which the features were extracted using a bank of 24 Log- Gabor filters, and the optimum features were selected based on MIQ algorithm. The images under slight illumination variation were used to test robustness of the FER system performance. Experimental results show that the color components provide additional information to achieve improved and robust performance of the system in terms of recognition rate for all expressions. Furthermore, the features are classified efficiently by multiclass LDA.

REFERENCES

[1] Seyed Mehdi Lajevardi and Hong Ren Wu, "Facial expression recognition in perceptual color space" , IEEE *Trans. on image processing*, vol. 21, no. 8, August 2012.

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

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

[4]. 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[11] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[12] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using* MATLAB. Upper Saddle River, NJ: Prentice-Hall, 2003.

[13] M. Thomas, C. Kambhamettu, and S. Kumar, "Face recognition using a color subspace LDA approach," in *Proc. 20th IEEE Int. Conf. Tools Artif. Intell.*, Dayton, OH, Nov. 2008, pp. 231–235.

[14] 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.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr.2004.

[16] S. M. Lajevardi and Z. M. Hussain, "Automatic facial expression recognition: Feature extraction and selection," *Signal, Image Video Process.*, vol. 6, no. 1, pp. 159–169, 2012.

[17] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Comput. Vis. Image Understand.*, vol. 115, no. 4, pp. 541–558, 2011.

Brruna .A received the B.Tech Degree in Software engineering from Anna University: Tirunelveli Region in 2012 and now studying M.E. Degree in Software engineering from Vins Christian College of engineering, Chunkankadai respectively. Her Area of Interest includes Data mining, Face Recognition.