International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 2 Issue: 12
ISSN: 2321-8169
3890 – 3895

A Novel Tensor Perceptual Color Framework based Facial Expression Recognition

P Amith Kumar,
E.C.E. Department, University College of Engineering, JNTUK,
Kakinada, Andhra Pradesh, India.
amithkumar.p@gmail.com

Abstract—The Robustness of Facial Expression Recognition (FER) is based on information contained in color facial images. The Tensor Perceptual Color Framework (TPCF) enables multilinear image analysis in different color spaces. This demonstrates that the color components provide additional information for robust FER. By using this framework color components RGB, YCbCr, CIELab or CIELuv space of color images are unfolded to 2-D tensors based on multilinear algebra and tensor concepts. The features of this unfolded image are extracted by using log-Gabor filter. The optimum features are selected based on mutual information quotient method in feature selection process. These features are classified using a multiclass linear discriminant analysis classifier. Experimental results demonstrate that color information has significant potential to improve emotion recognition performance due to the complementary characteristics of image textures.

Keywords-*facial expression recognition (FER), Log-Gabor filters, multilinear image analysis, Tensor Perceptual color Framework (TPCF).*

*****

I. INTRODUCTION

Basic goal of the human–computer-interaction (HCI) system is to improve the interactions between users and computers by making computers more users friendly and receptive to user's needs. Automatic facial expression recognition (FER) plays an important role in HCI systems and it has been studied extensively over the past twenty years. Border security systems, forensics, virtual reality, computer games, robotics, machine vision, user profiling for customer satisfaction, broadcasting, and web services are but a few different real world applications. The use of facial expression for measuring people's emotions has dominated psychology since the late 1960s when 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 Mehrabian [3] has indicated that 7% of the communication information is transferred by linguistic language, 38% by paralanguage, and 55% by facial expressions in human face-to-face communication. This, therefore, shows that facial expression provides a large amount of information in human communication.

Several approaches have been proposed for FER in the past several decades [1], [4]. The current state-of-the-art techniques for facial expression classification mostly focus on gray-scale image features [1], while rarely considering color image features [5]-[7]. Taking account of color feature data may lead to more robust classification results. Recent research reveals that color information improves face recognition and image retrieval performance [8]-[11]. It was first reported in that taking account of color information improves Recognition rate when compared with the same scheme using only the luminance information. It was shown in [9] that color components helped to improve face retrieval performance.

Liu and Liu proposed a new color space for face recognition [10]. Young, Man, and Plataniotis demonstrated that facial color cues significantly improved face recognition performance using low-resolution face images [11]. It was reported in that the RGB color tensor has improved the FER performance; however, it did not consider the images with different illuminations. Although recent research has shown improved performance by embedding the color components, the effectiveness of color information in the RGB color space in terms of recognition performance depends on the type and angle of light source, often making recognition impossible. Therefore, the RGB color space may not always be the most desirable space for processing color information. This issue can be addressed using perceptually uniform color systems [12].

This paper introduces a novel tensor perceptual color framework (TPCF) for FER based on information contained in color facial images. This paper is organized as follows. Section II describes the components of the FER system used for this investigation. Section III defines and examines the tensor-based representation of color Facial images in different color spaces and describes the proposed TPCF technique. Section IV presents and analyzes the main experimental results. Section V describes the conclusion.

II. THE PROPOSED IMAGE-BASED FER SYSTEM

Image-based methods extract features from images without relying on extensive knowledge about the object of interest, which are typically fast and simple, whereas model based methods attempt to recover the volumetric geometry of the scene, which are typically slow and complex [1]. The geometric features present the shape and locations of facial components (including mouth, eyes, eyebrows, and nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. The appearance features present the appearance (skin texture) changes of the face, such as wrinkles and furrows. The appearance features can be extracted from either the whole face or specific regions in a face image. This paper focuses on the static color images and a holistic technique of the image based method is used for feature extraction. The image based FER systems consist of several components or modules, including face detection and normalization, feature extraction, feature selection, and classification. Fig. 1 shows the system level diagram of the FER system. The following sections will describe each module in detail.

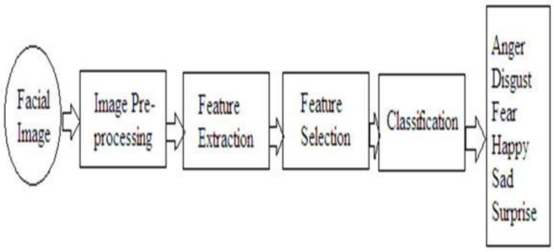


Fig. 1. System level diagram

A. Face Detection and Normalization

The aim of this module is to obtain face images, which have normalized intensity, are uniform in size and shape and depict only the face region. The 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 providing competitive object detection rates in real-time. It was primarily designed for face detection. The features used by Viola and Jones are derived from pixels selected from rectangular areas imposed over the picture, and exhibit high sensitivity to the vertical and the horizontal lines. After face detection stage, the face images are scaled to the same size (e.g., 128×128 pixels). The color values of face images are then normalized with respect to RGB values of the image. The purpose of color normalization is to reduce the lighting effect because the normalization process is actually a brightness elimination process. Given an input image of $N_1 \times N_2$ pixels represented in the RGB color space X = $\{X^{n3}[n_1, n_2] | 1 \leq n_1 \leq N_1 , 1 \leq n_2 \leq N_2 , 1 \leq n_3 \leq 3\}$, the normalized values,$X_{norm}^{n3}[n_1, n_2]$, are defined by

$$X_{norm}^{n3}[n_1, n_2] = \frac{X^{n3}[n_1, n_2]}{\sum_{n3=1}^{3} X^{n3}[n_1, n_2]} \quad (1)$$

Where $X_{norm}^{n3}[n_1, n_2]$ 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

$$\sum_{n_3=1}^{3} X_{norm}^{n3}[n_1, n_2] = 1 \quad (2)$$

B. Feature Extraction

Various methods of feature extraction have been studied and compared in terms of their performance, including principal component analysis, independent component analysis, linear discriminant analysis (LDA), the Gabor filter bank, etc. Fasel and Luettin reported that the Gabor filter bank has better performance than the rest [1]. The Gabor filters model the receptive field profiles of cortical simple cells quite well [1], [15]. However, the Gabor filters have two major limitations, i.e., the maximum bandwidth of Gabor filters is limited to approximately one octave, and the Gabor filters are not optimal to achieve broad spectral information with the maximum spatial localization [16]. Furthermore, the Gabor filters are bandpass filters, which may suffer from lost of the low and the high-frequency information [17]. To achieve the broad spectral information and to overcome the bandwidth limitation of the traditional Gabor filter, Field proposed Log-Gabor filter [17]. The response of the Log-Gabor filter is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high-frequency areas with desirable high pass characteristics. In this contribution, a bank of 24 Log-Gabor filters is employed to extract the facial features. The polar form of 2-D Log-Gabor filters in frequency domain is given by

$$H(f, \theta) = exp\left\{\frac{-\left[ln\left(\frac{f}{f_0}\right)\right]^2}{2\left[ln\left(\frac{\sigma_f}{f_0}\right)\right]^2}\right\} exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\} \quad (3)$$

whereH(f, θ) is the frequency response function of the 2-D Log-Gabor filters, f and θ denote the frequency and the phase/angle of the filter, respectively, f_o is the filter's 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 defines the angular band width in radians.

$$B = 2\sqrt{\frac{2}{\ln 2}} \times \left|\ln\left(\frac{\sigma_f}{f_o}\right)\right| , \Delta\Omega = 2\sigma_\theta\sqrt{\frac{2}{\ln 2}} \quad (4)$$

In this project described here, the ratio σ_f / f_o is kept constant for varying f_o, B is set to one octave and the angular bandwidth is set to ΔΩ= π/4 radians. This left only σ_f to be determined for a varying value of f_o. 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. 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 previously discussed 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 (IM). The mutual information quotient (MIQ) method for feature selection is adopted to select the optimum features [18]. 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 = $\{c_1, c_2, . . . ,c_k\}$, and $v_t$denotes the vector of N observations for that feature

$$v_t = [v_t^1, v_t^2, v_t^3, \ldots, v_t^N]^T \quad (5)$$

Where$\mathbf{v_t}$is an instance of the discrete random variable $\mathbf{V_t}$. The MI between features $\mathbf{V_t}$and $\mathbf{V_s}$is given by

$$I(V_t; V_s) = \sum_{v_t \in V_t} \sum_{v_s \in V_s} p(V_t, V_s) log \frac{p(V_t, V_s)}{p(V_t)p(V_s)} \quad (6)$$

where p($\mathbf{v_t}$, $\mathbf{v_s}$) is the joint probability distribution function (PDF) of $\mathbf{V_t}$and $\mathbf{V_s}$, p($\mathbf{v_t}$) and p($\mathbf{v_s}$) are the marginal PDFs of $\mathbf{V_t}$and $\mathbf{V_s}$, respectively, 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 $\mathbf{V_t}$ and C can be represented by entropies [19].

$$I(V_t; C) = H(C) - H(C|Vt) \quad (7)$$

Where

$$H(C) = -\sum_{i=1}^{k} p(c_i)\log(p(c_i)) \quad (8)$$

$$H(C|V_t) = -\sum_{i=1}^{k}\sum_{v_t \in V_t} p(c_i, v_t)\log(p(c_i|v_t)) \quad (9)$$

where H(C) is the entropy of C, H(C|**Vt**) is the conditional entropy of C on **Vt**, and k is the number of classes (for six expressions, k = 6). The features ($\mathbf{V_d}$) for desired feature subset, S, of the form {S; c}where S ⊂F and c ∈C are selected based on solution of following problems: $v_d = arg$

$$max\left\{\frac{I(V_t;C)}{\frac{1}{|s|}\sum I(V_t;V_s)}\right\} \boldsymbol{v_t} \varepsilon\, \overline{S}, \boldsymbol{v_s}\, \varepsilon\, S \quad (10)$$

Where $\overline{S}$ is the complement feature subset of S, |S|is the number of features in subset S and I ($\mathbf{V_t}$; $\mathbf{V_s}$) is the MI between the candidate feature ($\mathbf{V_t}$) and the selected feature ($\mathbf{V_s}$). Based on above equation, 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

For classification, the LDA classifier was studied for the same database and achieved better result than other classifiers [5]. Therefore, the selected features using the aforementioned MIQ technique are classified by a multiclass LDA classifier. The linear classifier based on discriminant analysis is used to classify the six different expressions. A natural extension of Fisher linear discriminant that deals with more than two classes exists [20], which uses multiple discriminant analysis. The projection is from a high-dimensional space to a low-dimensional space and the transformation sought is the one that maximizes the ratio of inter-class scatter ($\mathbf{S_b}$) to the intra-class ($\mathbf{S_w}$) scatter. The maximization should be done among several competing classes. The $\mathbf{S_b}$can be viewed as the sum of squared distances between each class mean and the mean of all training samples and $\mathbf{S_w}$can be regarded as the average class-specific covariance. The inter-class ($\mathbf{S_b}$) and the intra-class ($\mathbf{S_w}$) matrices for feature vector (x^f) are given by

$$S_b = \sum_{i=1}^{N_c} m_i (X_{\mu_i}^f - X_\mu^f)(X_{\mu_i}^f - X_\mu^f)^T \quad (11)$$

$$S_w = \sum_{i=1}^{N_c} \sum_{X^f \in c_i} (X^f - X_{\mu_i}^f)(X^f - X_{\mu_i}^f)^T \quad (12)$$

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

$$X_{\mu_i}^f = \frac{1}{m_i}\sum_{X^f \in c_i} X^f \quad (13)$$

$$X_\mu^f = \frac{1}{m}\sum_{i=1}^{N_c} m_i X_{\mu_i}^f \quad (14)$$

After obtaining $\mathbf{S_w}$and $\mathbf{S_b}$, based on Fisher's criterion the linear transformation, $\mathbf{W}_{LDA}$, can be obtained by solving the generalized eigenvalue (λ) problem

$$W_{LDA}^T S_b = \lambda\, W_{LDA}^T S_W \quad (15)$$

Once the transformation $\mathbf{W}_{LDA}$ is given, the classification can be performed in the transformed space based on predefined distance measure, such as the Euclidean distance, ||•||. The new instance, x_n^f, is classified to

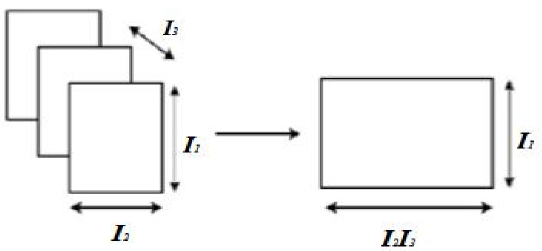
$$c_n = arg\, min_i \left\| W_{LDA} x_n^f - W_{LDA} x_{\mu i}^f \right\| \quad (16)$$

III.TENSOR-BASED REPRESENTATION OF COLOR FACIAL IMAGES

Each color image can be represented as a 3-D, (i.e., horizontal, vertical, and color) data array. There is a technical challenge to proceed with applying a 2-D filtering process to a 3-D matrix, which represents the color image. It can either process a single channel of the color image (e.g., luminance image) or perform the filtering operation on each color channel individually. The latter approach is to employ the 2-D filters three times over three component images, respectively. Instead of implementing the filter for each component of the color image, a tensor of the color image is generated and the filtering operation is directly applied to this tensor [7], [21], [22]. A tensor is a higher order generalization of a vector (first-order tensor) and a matrix (second-order tensor). Tensors are multilinear mappings over a set of vector spaces. A color image represented by τ is a tensor of order 3 and $\tau \epsilon R \prod_1^3 I_n$ where I_1 is the height of the image, I_2 is the width of the image, and I_3 represents the number of color channels. In this project, I_1 and I_2 vary from 128 to 196 and I_3 is 3. Tensor can be unfolded to n-mode mathematical objects. In this project, there are three modes for tensor $\tau^{(n-mode)}$ which defined by

$$\tau \epsilon R \prod_1^3 I_n \rightarrow \tau^{(1)} \epsilon R^{I_1 \times (I_2 \times I_3)}$$
$$\tau \epsilon R \prod_1^3 I_n \rightarrow \tau^{(1)} \epsilon R^{I_2 \times (I_1 \times I_3)}$$
$$\tau \epsilon R \prod_1^3 I_n \rightarrow \tau^{(1)} \epsilon R^{I_3 \times (I_1 \times I_2)} \quad (17)$$

The 3-D color image is unfolded to obtain 2-D tensors based on multilinear analysis criteria [21], which are suitable for 2-D feature extraction filters. These tensors are used for feature extraction and classification.



 ISSN: 2321-8169
3890 – 3895

Fig. 2. Horizontal unfolding of facial images

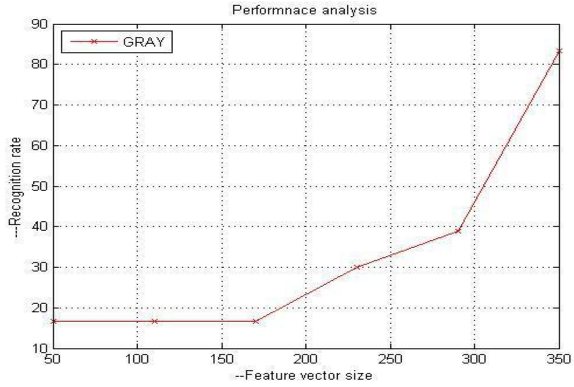
All modes are tested and the best one is unfolding mode 1 [17]. In this project, the image $XI_1 \times I_2 \times I_3$ is unfolded to $XI_1 \times I_2 I_3$, which is called horizontal unfolding.

IV. EXPERIMENTAL RESULTS

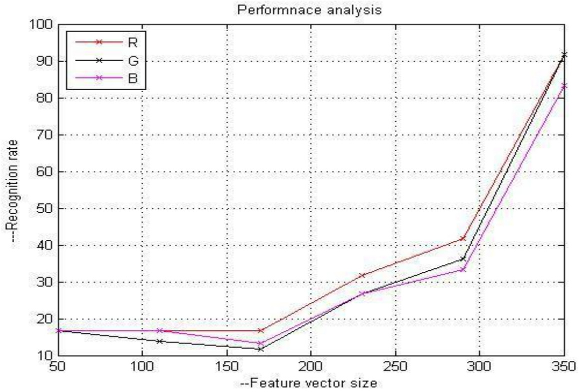
The results are shown in Figs. 3–5 using images of 128 ×128, 144×144, and 198×198 resolutions. The average recognition rate is improved as well when the resolution is increased from 128 × 128 to 198 × 198. Table II summarizes the average recognition rates under different resolutions. Furthermore, it has a comparable or slightly superior performance to that of others in terms of recognition rate for images.

TABLE I COMPARISON OF AVERAGE RECOGNITION RATES FOR DIFFERENT RESOLUTIONS AND COLOR SPACES

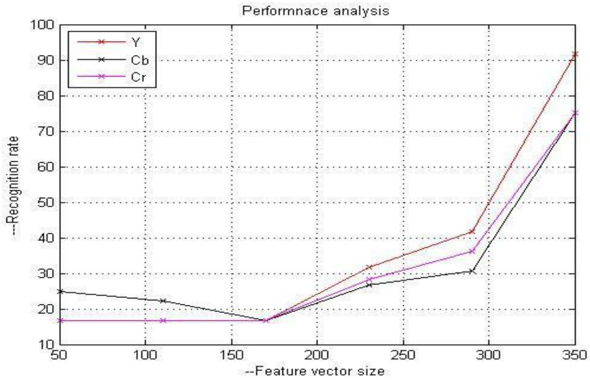
Size	Gray	RGB	YCbCr	CIELab
128X128	83.33	88.99	80.55	72.32
144X144	83.33	88.99	72.33	72.33
196x196	83.33	91.66	83.44	88.99



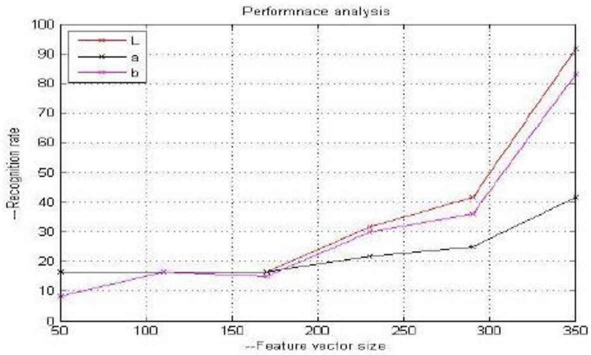
(a)



(b)



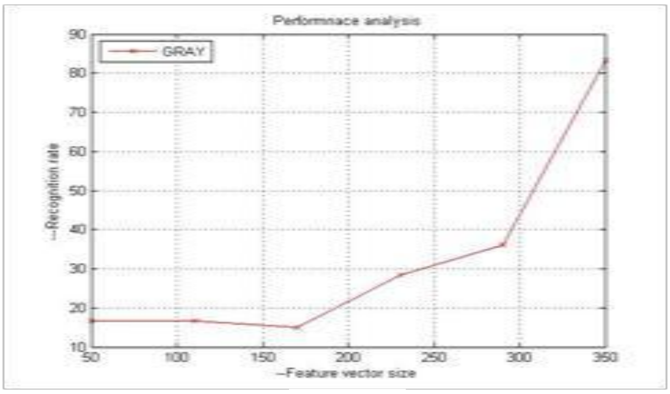
(c)



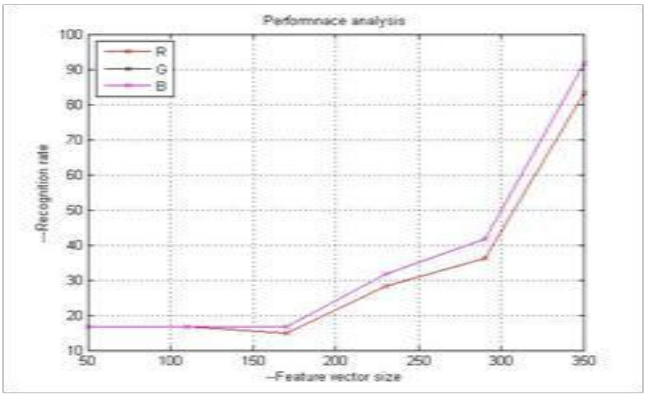
(d)

Fig 3. Comparative evaluation of performances in different color spaces from 128×128 images (a) Gray. (b) RGB. (c) YC_bC_r. (d) CIELab

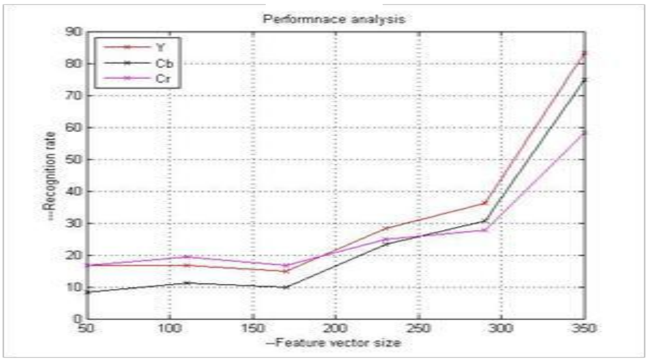
ISSN: 2321-8169
3890 – 3895



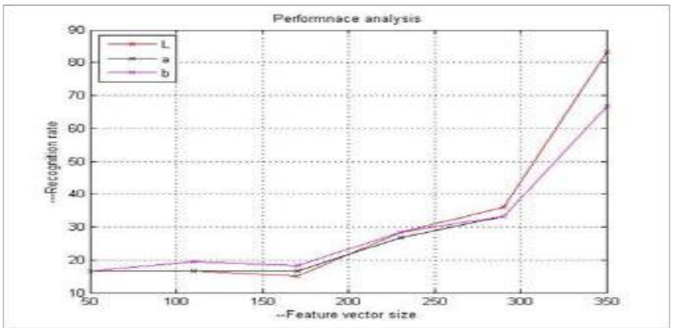
(a)



(b)

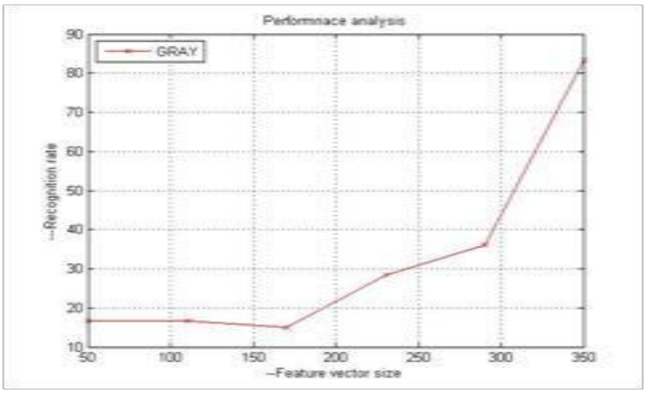


(c)

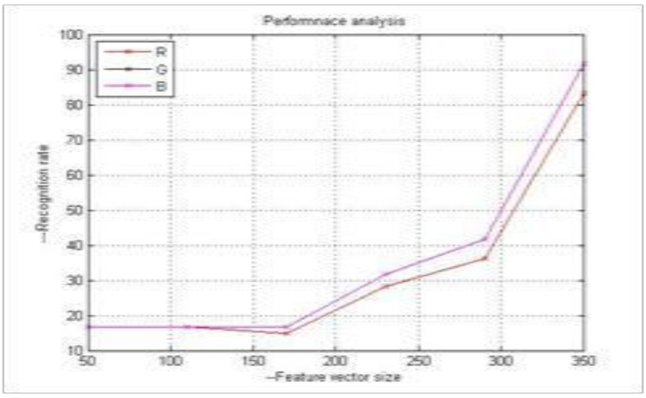


(d)

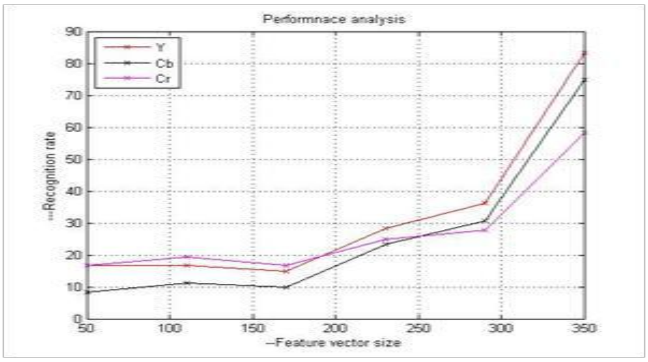
Fig 4. Comparative evaluation of performances in different color spaces from 144×144 images (a) Gray. (b) RGB. (c) YC_bC_r. (d) CIELab.



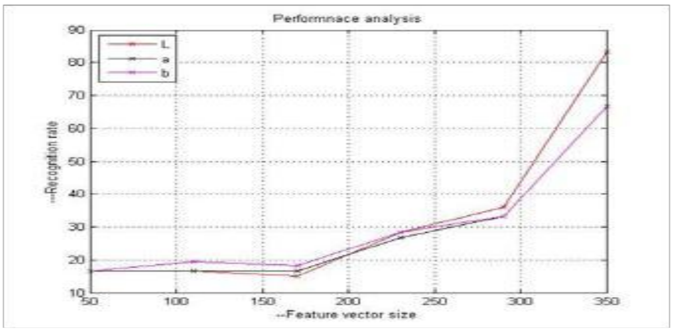
(a)



(b)



(c)



(d)

Fig 5. Comparative evaluation of performances in different color spaces from 196×196 images (a) Gray. (b) RGB. (c) YC_bC_r. (d) CIELab.

 ISSN: 2321-8169
3890 – 3895

V. CONCLUSION

A novel TPCF was proposed for FER system in perceptual color space. Based on TPCF, the RGB color images were first transformed to perceptual color spaces after which the horizontal unfolded tensor was adopted to generate the 2-D tensor for feature extraction. The 2-D tensor was normalized before the features were extracted using a bank of 24 Log-Gabor filters, and the optimum features were selected based on MIQ algorithm. 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. To the best of author's knowledge, the achieved average recognition accuracy of the TPCF for FER system is better than any other.

REFERENCES

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey,"b*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. 3–69, 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 highresolution 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.

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

[15] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust.Speech Signal Process.*, vol. 36, no. 7, pp. 1169–1179, Jul. 1988.

[16] T. S. Lee, "Image representation using 2-D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct.1996.

[17] J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, Dec. 1987.

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

[19] T. Cover and J. Thomas, *Elements of Information Theory* (Telecommunications and Signal Processing). New York: Wiley, 2006.

[20] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[21] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," in *Proc. Int. Conf. Pattern Recognit.*, Quebec City, QC, Canada, Aug. 2002, pp. 511–514.

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