Facial Expression Recognition Under Varying Illumination in Perceptual Color System

Seyed Mehdi Lajevardi, Hong Ren Wu
School of Electrical & Computer Engineering, RMIT University, Melbourne, Australia
smlajevardi@ieee.org, henry.wu@rmit.edu.au

***Abstract*—This paper presents an approach to facial expression recognition based on perceptual color information (in CIELab color space). The RGB color images are converted to CIELab color images which are concatenated to create a 2-D tensor based on image analysis. The CIELab 2-D tensor based approach renders distinguished capability for improving facial expression recognition (FER) performance due to its robustness under varying illumination. The features are generated by 24 Log-Gabor filters operating on 2-D tensors to characterize facial expression textures and the optimum features are selected using the mutual information criterion. These features are classified by a linear discriminant analysis (LDA) classifier. Experimental results shows that the proposed method is more efficient and robust for facial expression recognition than other popular methods in low-resolution images under slight illumination variation.**

I. Introduction

Since 1978, there has been a significant research interest in automatic recognition of human emotion from facial expressions [3]. Facial expression is a great potential to be used as an additional input to Human Computer Interface (HCI) of various application systems. This is especially true in voice-activated control systems. This implies a facial expression recognition (FER) module can markedly improve the performance of such systems. Customer's facial expressions can also be collected by service providers as implicit user feedback to improve their services. Compared with a conventional questionnaire based method, this provides a more reliable means of acquiring customer feedback and, it incurs virtually no additional cost to that associated with the initial setup. Analyzing the emotional expression of a human face requires a number of preprocessing steps which attempt to detect and locate characteristic facial regions, extract facial expression features, and model facial gestures using anatomic information about the face. Although all these steps are equally important, current research efforts mostly concentrate on the facial expression feature extraction and classification. The feature extraction is a key importance to the whole classification process. If inadequate features are used, even the best classifier may fail to achieve accurate recognition. The two most common approaches to the facial feature extraction are the geometric feature-based methods and the appearance based methods [3]. Geometric features present the shape and locations of facial components (including mouth, eyes, eyebrows, 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.

A number of approaches have been developed for extracting features from face images. Various methods of feature extraction have been studied and compared in terms of their performance, including principal component analysis (PCA), independent component analysis (ICA), linear discriminant analysis (LDA), the Gabor filter bank, etc. Fasel reported that the Gabor filter bank has better performance than the rest [3]. However, the computational complexity and memory storage requirements of this method are very high and it's prohibiting real-time applications. To overcome these issues, the Log-Gabor filters [6] are adopted to the problem. In contrast to the entire Gabor filter bank output, the use of the Log-Gabor filters can significantly decrease the dimensionality of the characteristic features, and thus decrease the computational complexity of the whole facial expression recognition system. Most of state-of-art researches were used grayscale images for FER system. However, there were rarely studied on the color image features [7]. The term "expression" is not only a physical property of an object, but also it is a perceptual phenomenon and human subjective concept. Therefore, color representation similar to the color sensitivity of human vision system should help to obtain high performance in terms of classification rate for emotion recognition system. RGB color space is not always the most convenient space in which to process color information. So, The CIE tristimulus system proposed a new color space in terms of its three coordinates relative usually a standard illuminate to a reference color [1]. In this study, the color information in CIELab color space is investigated. The remainder of this paper describes the methods, experiments and results. Section 2 explains the image pre-processing steps. In Section 3, the feature extraction method based on the Log-Gabor filter bank and CIELab color space is explained. Section 4 contains the experimental results and in section 5 final conclusions are presented.

II. Face detection and normalization

The image pre-processing procedure is a very important step in the facial expression recognition task [2], [5]. The aim of the pre-processing phase is to obtain face images which have normalized intensity, are uniform in size and shape, and depict only the face region. The image intensity was normalized using the histogram equalization. The face area of an image was detected using the Viola-Jones method [13] based on the Haar-like features and the AdaBoost learning algorithm. The Viola and Jones method is an object detection algorithm providing competitive object detection rates in real-time. It was primarily designed for the problem of face detection. The features used by Viola and Jones are derived from pixels selected from rectangular areas imposed over the picture and show high sensitivity to the vertical and horizontal lines. After face detection stage, the face images are scaled into same size (32×32 pixels). After this stage, the color values of face images are normalized with respect to RGB values of the image. The advantage of color normalization is that it can reduce the lighting effect because the normalization process is actually a brightness elimination process. For input images represented in RGB color space, the normalized values are defined by:

$$\bar{R} = \frac{R}{\overline{RGB}}, \quad \bar{G} = \frac{G}{\overline{RGB}}, \quad \bar{B} = \frac{B}{\overline{RGB}} \tag{1}$$

where $\overline{RGB} = R+G+B$ and $\bar{R}+\bar{G}+\bar{B} = 1$. The images after normalization are used for feature extraction. Fig. 1 shows the normalized face images.

Fig. 1. Facial expression images after preprocessing stage.

III. CIELAB COLOR SPACE FEATURES

Current studies shows that the color information affect on performance of different pattern recognition systems [11], [6]. However, the color information is affected by changes of the light source (e.g., from indoor illumination to outdoor daylight), often making recognition impossible. This can be accomplished using perceptually uniform color systems [1]. The CIELab is one of colorimetric color spaces which separates a luminance variable 'L' from two perceptually uniform chromaticity variables ('a','b') [1]. The CIELab is widely used in several image processing applications include: perceptual image quality assessment, face detection, skin detection, image segmentation and etc. [1], [9]. Despite the many advantages of such color space, it's rarely used in pattern recognition. This is mainly because the transformation from RGB is more computationally expensive than other spaces. The transformation from RGB color space to CIELab color space is defined by:

$$L = \begin{cases} 116 \times (\frac{Y}{Y_n})^{\frac{1}{3}} - 16 & \frac{Y}{Y_n} > 0.008856 \\ 903 \times (\frac{Y}{Y_n}) & \frac{Y}{Y_n} \leq 0.008856 \end{cases} \tag{2}$$

$$a = 500 \times (f(\frac{X}{X_n}) - f(\frac{Y}{Y_n})) \tag{3}$$

$$b = 200 \times (f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n})) \tag{4}$$

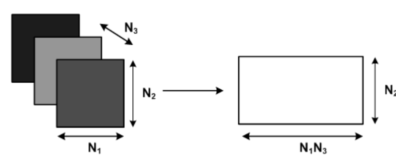
where

$$f(t) = \begin{cases} t^{\frac{1}{3}} & t > 0.008856 \\ 7.787 \times t + \frac{16}{116} & t \leq 0.008856 \end{cases} \tag{5}$$

$X_n = 0.951, Y_n = 1.000, Z_n = 1.089$ and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.431 & 0.342 & 0.178 \\ 0.222 & 0.707 & 0.071 \\ 0.020 & 0.130 & 0.939 \end{bmatrix} \begin{bmatrix} \bar{R} \\ \bar{G} \\ \bar{B} \end{bmatrix} \tag{6}$$

After conversion, the CIELab color image ($\mathbf{x}$) is unfolded to 2-D matrix tensor [12], [10]. 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. The order of tensor $A \in \mathbb{R}^{N_n}$ is n. We can unfold it along any of its dimensions to obtain a mode-n matrix version of the tensor. Given an image $\mathbf{x}$, which has three color channels, the unfolding of this image can be in three different dimensions. In this study, the image $\mathbf{x}_{N_1 \times N_2 \times N_3}$ is unfolded to $\mathbf{x}'_{N_1 N_3 \times N_2}$ which is called horizontal unfolding. Fig. 2 shows the procedure of unfolding of color and unfolded facial image in different color spaces. The features are extracted from these unfolded facial images.



(a)

(b)

Fig. 2. (a) Horizontal unfolding of facial expression image for different color spaces. (b) CIELab 2-D tensor

Gabor filters are commonly recognized as one of the best choices for obtaining localized frequency information. However, they suffer from two major 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. As an alternative to the Gabor filters, the Log-Gabor filters were proposed by Field [4]. The 2-D Log-Gabor filters can be represented in a polar form as follows:

$$H(f,\theta) = \exp\{\frac{-[\ln(\frac{f}{f_0})]^2}{2[\ln(\frac{\sigma_f}{f_0})]^2}\} \exp\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\} \qquad (7)$$

where f_0 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 bandwidth $\Delta\Omega$ in radians:

$$B = 2\sqrt{\frac{2}{\ln 2}} \times |\ln(\frac{\sigma_f}{f_0})|; \quad \Delta\Omega = 2\sigma_\theta\sqrt{\frac{2}{\ln 2}} \qquad (8)$$

In the study described here, the ratio σ_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 σ_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 lead to 24 filter transfer functions which representing different scales and orientations. The image filtering is performed in the frequency domain making the process faster compare to the space domain convolution. After the 2-D FFT transformation into the frequency domain, the image arrays $\mathbf{x}$ are changed into the spectral vectors $\mathbf{X}$ and multiplied by the Log-Gabor transfer functions $\{H_1, H_2, \cdots, H_{24}\}$, producing 24 spectral representations for each image. The spectra is then transformed back to the spatial domain via the 2-D inverse FFT. This process resulted in prohibitively large number of the feature arrays.

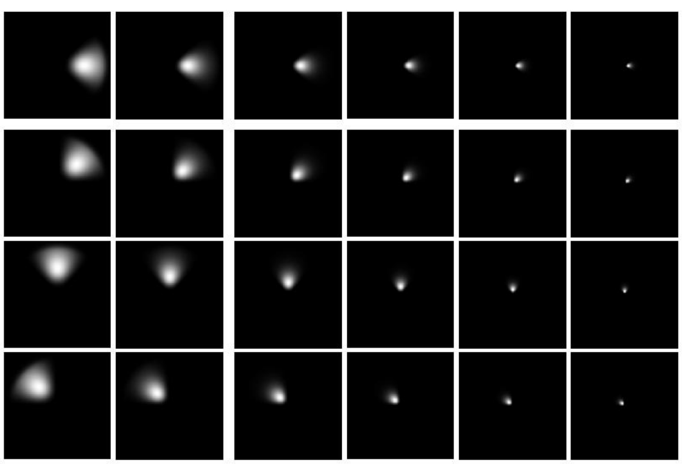


Fig. 3. Sample of Log-Gabor filters in frequency domain.

An example of Log-Gabor filters is shown in Fig. 3. For large training and testing sets, the computations are highly impractical. In order to improve the computational efficiency, it is critical to reduce the features dimensions. This is achieved through the feature selection. The mutual information quotient (MIQ) method for feature selection is adopted to select the optimum features [8]. In this methodology, if a feature vector has expressions randomly or uniformly distributed in different classes, its mutual information (MI) with these classes is zero. If a feature vector is strongly different from other features for different classes, it should have large MI. The features are selected based on solution of following problem:

$$f_m = \arg\max_{f_t}\{\frac{I(f_t, C)}{\frac{1}{|S|}\sum I(f_t, f_s)}\}, \quad f_t \in F_S, f_s \in S \qquad (9)$$

where $F_S = F_T - S$, F_T is total features, S is the desired feature subset, $I(f_t, C)$ is MI between candidate feature (f_t) and class label $(C = \{1, 2, ..., 6\})$, $|S|$ is the number of features in subset S and $I(f_t, f_s)$ is MI between candidate feature (f_t) and selected feature (f_s). So, based on above equation, the MI between selected feature and intra-class features is maximized whereas the MI between selected feature and within-class features are minimized respectively. These features are used for emotion classification.

IV. Experimental results

Facial expression image from the Binghamton University 3D Facial Expression Database (BU-3DFE) [7] are used for training and testing sets. The BU-3DFE database contains 100 subjects (56% female, 44% male), with age ranging from 18 years to 70 years old, and a variety of ethnic/racial ancestries, including White, Black, East-Asian, Middle-east Asian, Indian, and Hispanic Latino. Overall, 2300 facial expression images are selected from the database. The original resolution of the images is 256×256. The subjects represented in the training set were not included in the testing set for classification, thus ensuring a person independent FER system. Six expressions (anger, disgust, fear, happy, sad, surprise) are used for classification. Face detector is adopted to locate the face area. These faces are normalized and scaled to 32×32 pixels. Then, the RGB color face images are transformed to CIELab color space and the features are extracted from 2-D tensor CIELab color images based on 24 Log-Gabor filter bank. The optimum features are selected based on MIQ criteria. Fig. 4 shows the number of features and recognition rate for training set. 380 features are selected and classified based on multicalss LDA classifier. For comparison, the grayscale and RGB color of the same images are used under varying illumination. The structural similarity index (SSIM) is performed to find the quality of the image for slight illumination variation [14]. Table I illustrates the average recognition rate under different illumination. It can be seen from the results that the recognition rate is improved when the CIELab color space adopted in the system, whereas the grayscale has the lowest recognition rate for all expressions. Furthermore, Both gray level and RGB color images are failed under illumination variation, but, CIELab is robust to slight change of illumination. Table II shows the recognition rate of each expression and the error for CIELab color space.

V. Conclusion

Emotion recognition in perceptual color space was investigated. The RGB color images were converted to CIELab color space and they transformed to 2-D tensor matrix. The features were extracted based on 24 Log-Gabor filter bank and the optimum features selected based on MIQ algorithm. The

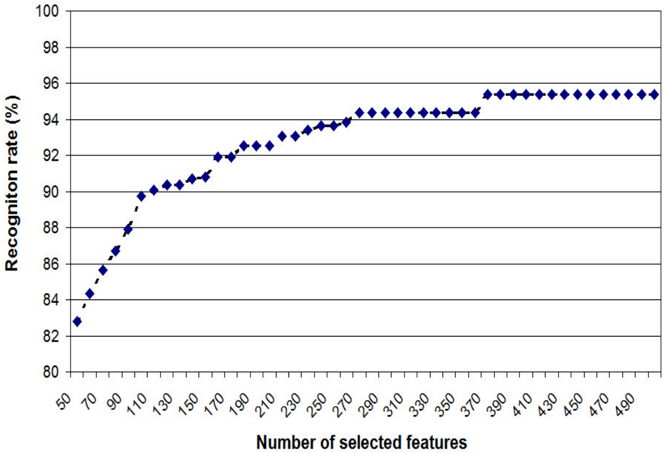


Fig. 4. 10-fold cross-validation recognition rate for training set based on number of features.

TABLE I
AVERAGE RECOGNITION RATE COMPARISON BETWEEN CIELAB, RGB AND GRAYSCALE FOR SIX EXPRESSIONS UNDER DIFFERENT ILLUMINATIONS (32×32).

SSIM index	Grayscale (%)	RGB (%)	CIELab (%)
0.35	13.1	17.5	66.3
0.84	17.9	41.7	76.3
0.93	25.4	60.1	78.8
0.96	36.5	71.1	81.2
1 (Original)	81.3	83.4	83.4

testing set were classified by multi-class LDA classifier. Excremental results showed that the perceptual color information has improved the performance of facial expression recognition under different illumination variations.

REFERENCES

[1] M. Corbaln, M. S. Milln, and M. J. Yzuel, "Color pattern recognition with CIELab coordinates," *Opt. Eng.*, vol. 41, no. 1, pp. 130-138, 2002.
[2] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification, Wiley*, New York, 2001.
[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, pp. 259-275, 2003
[4] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America,*, vol. 4, no. 12, pp. 2379-2394, 1987.
[5] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing*: Addison Wesley Publishing Company, 2008.
[6] S. M. Lajevardi and Z. M. Hussian, "Emotion Recognition from Color Facial Images Based on Multilinear Image Analysis and Log-Gabor Filters," *In Proceeding of 25th International Conference on Image and Vision Computing New Zealand, IVCNZ'10*, Queenstown, New Zealand, 2010, pp. 10-14.
[7] Y. Lijun, C. Xiaochen, S. Yi, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," *in Automatic Face and Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, Amsterdam, Netherlands, 2008, pp. 1-6.
[8] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1226-1238, 2005.
[9] P. Shih and C. Liu, "Comparative assessment of content-based face image retrieval in different color spaces," *International Journal of Pattern Recognition Artif. Intell.,* vol. 19, no. 7, pp. 873893, 2005.

TABLE II
PERCENTAGE OF CORRECT CLASSIFICATIONS IN CIELAB COLOR SPACE.

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	82.7	8.6	4.0	0.67	14.0	1.3
Disgust	3.5	78.5	8.5	1.0	1.0	0.5
Fear	1.5	6.5	75.0	9.0	7.0	2.0
Happy	1.0	3.0	6.5	89.5	1.0	1.0
Sad	7.0	4.0	5.5	0	80.0	1.0
Surprise	0	1.5	1.5	0	0.5	94.5

[10] M. Thomas, C. Kambhamettu, and S. Kumar, "Face Recognition Using a Color Subspace LDA Approach," *in Tools with Artificial Intelligence, 2008. ICTAI '08. 20th IEEE International Conference on*, Dayton, OH, USA, 2008, pp. 231-235.
[11] L. Torres, J. Y. Reutter, and L. Lorente, "The importance of the color information in face recognition," *in Proceedings International Conference on Image Processing (ICIP 99)*, 1999, vol.3, pp. 627-631.
[12] M. Vasilescu and D. Terzopoulos, "Multilinear image analysis for facial recognition," *in International Conference on Pattern Recognition (ICPR'02)*, Quebec City, Canada, 2002, pp. 511-514.
[13] P., Viola, M., Jones, "Robust real-time face detection," *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.
[14] Z. Wang, A.C., Bovik, H.R., Sheikh, E.P., Simoncelli: "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April, 2004.