COLOR FACIAL EXPRESSION RECOGNITION BASED ON COLOR LOCAL FEATURES

Wenming Zheng[1] *Xiaoyan Zhou*[2] *Minghai Xin*[1]

[1]Key Laboratory of Child Development and Learning Science, Ministry of Education,
Research Center for Learning Science, Southeast University, Nanjing, 210096, P.R. China
[2]School of Electronics and Information Engineering,
Nanjing University of Information Science and Technology,Nanjing, 210044, P.R. China
wenming_zheng@seu.edu.cn

ABSTRACT

In this paper, color facial expression recognition based on color local features is investigated, in which each color facial image is decomposed into three color component images. For each color component image, we extract a set of color local features to represent the color component image, where color local features could be either color local binary patterns (LBP) or color scale-invariant feature transform (SIFT). To cope with the facial expression recognition problem, we use a group sparse least square regression (GSLSR) model to describe the relationship between the color local feature vectors and the associated emotion label vectors and then perform expression recognition based on it. Finally, experiments on the Multi-PIE color facial expression database are conducted to testify the proposed method and compare the results with state-of-the-art methods.

***Index Terms*—** Color facial expression recognition, Group sparse least square regression model, Color local features

1. INTRODUCTION

The recognition of emotion from facial images is a major research topic of affective computing and pattern recognition. During the last several decades, this research topic had been extensively explored in the literatures [1]. Basically, the recognition of facial expression mainly consists of two major steps, i.e., the facial feature extraction versus the expression classification. The target of facial feature extraction is to extract the features that contribute to the expression classification of the facial images, where the facial features can be roughly categorized into the geometric features and the appearance based features [1]. A typical geometric features is to use the landmark points that characterize that major parts of a face, such as eyes, brows, or mouth [2]. The most commonly used appearance based facial features include, local binary patterns (LBP) features [3], Gabor features [4], and scale invariant feature transform (SIFT) features [5]. Although the appearance based facial feature extraction approaches have been extensively explored, most of the methods focus on the gray-scale facial images. In contrast to gray-scale facial image, however, color facial image has been shown to be more advantageous in the facial expression recognition [6][7], where experimental results had demonstrated that using color facial image is able to obtain higher facial expression recognition accuracy than using gray-scale facial image.

In this paper, we investigate the color facial expression recognition problem based on color local features, where both color local uniform binary pattern with uniform (LBP^{u2}) [3] feature and color SIFT feature are respectively explored for this purpose. To extract the color LBP feature, we adopt the method used in [9] by dividing each color image component into a set of facial regions with multi-scale sizes and then extract the color local features from the facial regions of the color components. To extract the color SIFT features, on the other hand, we adopt the active shape model (ASM) [8] to automatically locate 49 landmark facial points from each color facial image and then use the points as the key points to extract 49 color SIFT feature vectors from each color component image.

In dealing with color facial expression recognition problem, a major issue one should be notable is how to effectively fuse the facial features associated with different color image components. In addition, it is also notable that different facial regions play different roles in discriminating facial expressions and hence contribute differently to the recognition of the expressions [9]. For this reason, Zheng proposed a novel method in [9] to deal with both facial feature extraction and facial feature selection problems. In that method, the author firstly divided each facial image into a set of facial regions using multi-level region sizes. Then, the facial feature extraction is carried out in each facial region. To differentiate the importance of the different facial regions, a group sparse regression model is developed for the simultaneous modeling training and the facial region selection.

Motivated by the work of Zheng [9], in this paper we propose a group sparse least square regression (GSLSR) model to cope with the color facial expression recognition problem, in which the GSLSR model is used for modeling the relationship between color facial image feature vectors and the associated emotion class label vectors. In addition, we also develop an efficient algorithm for the GSLSR model to simultaneously learn the model parameters and the facial region selection. Finally, we apply the proposed GSLSR method to the expression label prediction of testing color image .

2. FEATURE EXTRACTION FROM COLOR FACIAL IMAGES

Suppose that $\mathcal{I}$ is a given color facial image. Then, to extract the color local features from $\mathcal{I}$, we first decompose $\mathcal{I}$ into three kinds of color image components in terms of the color space we chose, where the color space could be either the RGB color space, the YCbCr color space, or the CIE Lab color space [6]. To demonstrate our feature extraction scheme, we take the RGB color space as an example. In this case, we divide $\mathcal{I}$ into the red (R) image component $\mathcal{I}_R$, the green (G) image component $\mathcal{I}_G$, and the blue (B) image component $\mathcal{I}_B$.

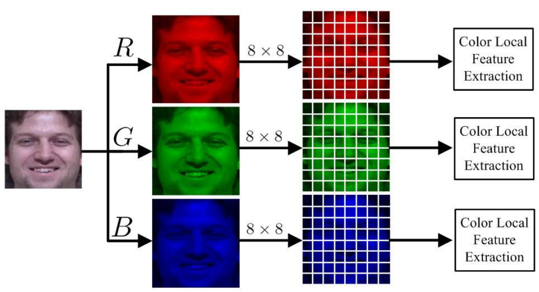

Fig. 1. Illustration of facial feature extraction from a color image, where the face image is decomposed into three components and each component is divided into 64 sub regions.

To extract the color LBP feature, for each of the three image components, we adopt the method used in [9] by dividing the image into a set of regions and then perform the color LBP feature extraction in each region. More specifically, suppose that $\mathcal{I}_R$ is the image component corresponding to the R color component of $\mathcal{I}$. Then, we divide $\mathcal{I}_R$ into K sub regions $\mathcal{R}_1, \cdots, \mathcal{R}_K$, and extract the color local features from each of the sub regions. Fig. 1 illustrates the procedures of color local feature extraction, where each component image is divided into $8 \times 8 = 64$ sub regions. Noting that the aforementioned sub regions shown in Fig.1 may only reflect the local information of a facial expression. To capture both local and global expression information, we increase the region size such that more image information are contained in the regions. As a result, we obtain the following multi-scale region division scheme shown in Fig.2, in which the component image is divided into a set of regions with different sizes. Based on the region division shown in 2, we totally obtain 85 sub regions from each color component image.

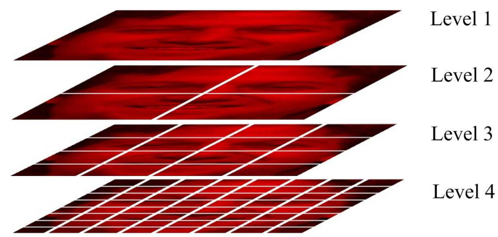

Fig. 2. The multi-scale region division framework.

To extract the color SIFT feature, we first use the ASM method to automatically locate 49 landmark facial points, and use these landmark points as the key points when dealing with the SIFT feature extraction. In this case, we are able to obtain 49 SIFT feature vectors from each facial component image. Fig.3 illustrates the 49 landmark points located by ASM.

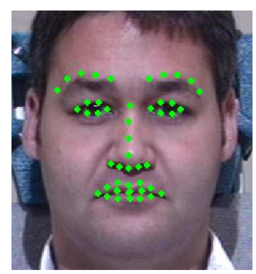

Fig. 3. The 49 landmark points automatically located by ASM for color SIFT feature extraction.

3. COLOR IMAGE FACIAL EXPRESSION RECOGNITION BASED ON GSLSR

In this section, we will address the color image facial expression recognition method based on GSLSR. Without loss of generality, the RGB color space is adopted for facial feature extraction when addressing this method.

3.1. Group Sparse Least Square Regression Model

Suppose that we are given N training face images. Let

$$\mathbf{Y}_R = \begin{bmatrix} \mathbf{Y}_{R1} \\ \vdots \\ \mathbf{Y}_{RK} \end{bmatrix}, \mathbf{Y}_G = \begin{bmatrix} \mathbf{Y}_{G1} \\ \vdots \\ \mathbf{Y}_{GK} \end{bmatrix}, \mathbf{Y}_B = \begin{bmatrix} \mathbf{Y}_{B1} \\ \vdots \\ \mathbf{Y}_{BK} \end{bmatrix}$$

denote the facial feature matrices corresponding to the three color components in the RGB color space, where $\mathbf{Y}_{Ri}$, $\mathbf{Y}_{Gi}$

and $\mathbf{Y}_{Bi}$ represent the feature matrices corresponding to the i-th sub region of the R color component, G color component, and B color component, respectively.

Let $\mathbf{X}$ denote class label matrix, where the i-th column $\mathbf{x}_i$ is a $c \times 1$ class label vector associated with the i-th facial image, where c is the number of expression classes. The entries of $\mathbf{x}_i = [x_{i1}, \cdots, x_{ic}]^T$ take the value of 0 or 1, i.e.,

$$x_{ij} = \begin{cases} 1, & \text{if } \mathbf{y}_i \text{ belongs to the } i^{th} \text{ class;} \\ 0, & \text{otherwise.} \end{cases}$$

Now we use the following least-square regression (LSR) model to describe the relationship between $\mathbf{X}$ and $\mathbf{Y}_i$, $i \in \{R, G, B\}$, i.e.,

$$\arg \max_{\mathbf{A}_R, \mathbf{A}_G, \mathbf{A}_B} \left\{ L = \left\| \mathbf{X} - \sum_{i \in \{R,G,B\}} \mathbf{A}_i \mathbf{Y}_i \right\|_F^2 \right\}, \quad (1)$$

where $\mathbf{A}_R$,$\mathbf{A}_G$,$\mathbf{A}_B$ are regression coefficient matrices.

Let $\mathbf{A}_R = [\mathbf{A}_{R1}, \cdots, \mathbf{A}_{RK}]$, $\mathbf{A}_G = [\mathbf{A}_{G1}, \cdots, \mathbf{A}_{GK}]$, $\mathbf{A}_B = [\mathbf{A}_{B1}, \cdots, \mathbf{A}_{BK}]$, such that the objective function L in (1) can be expressed as:

$$L = \left\| \mathbf{X} - \sum_{i \in \{R,G,B\}} \sum_{j=1}^{K} \mathbf{A}_{ij} \mathbf{Y}_{ij} \right\|_F^2. \quad (2)$$

Noting that different facial regions contribute differently to the expression recognition [9], we use a weight w_j to indicate the importance of the region, where w_j takes the value 1 or 0. If $w_j = 1$, the corresponding region is important and hence the features are utilized in the facial expression recognition. Otherwise, it will not be used. As a result, we can use the following feature matrices to replace the original ones:

$$\tilde{\mathbf{Y}}_R = \begin{bmatrix} w_1 \mathbf{Y}_{R1} \\ \vdots \\ w_K \mathbf{Y}_{RK} \end{bmatrix}, \ \tilde{\mathbf{Y}}_G = \begin{bmatrix} w_1 \mathbf{Y}_{G1} \\ \vdots \\ w_K \mathbf{Y}_{GK} \end{bmatrix},$$

and $\tilde{\mathbf{Y}}_B = \begin{bmatrix} w_1 \mathbf{Y}_{B1} \\ \vdots \\ w_K \mathbf{Y}_{BK} \end{bmatrix}$, where $w_j \in \{0, 1\}$. Then, the objective function L in (2) can be expressed as

$$L = \left\| \mathbf{X} - \sum_{i \in \{R,G,B\}} \sum_{j=1}^{K} w_j \mathbf{A}_{ij} \mathbf{Y}_{ij} \right\|_F^2, \quad (3)$$

where $\tilde{\mathbf{A}}_{Rj} = w_j \mathbf{A}_{Rj}$, $\tilde{\mathbf{A}}_{Gj} = w_j \mathbf{A}_{Gj}$, and $\tilde{\mathbf{A}}_{Bj} = w_j \mathbf{A}_{Bj}$.

From the expression of (3), we can see that optimizing the coefficient matrices $\mathbf{A}_{Rj}$, $\mathbf{A}_{Gj}$, $\mathbf{A}_{Bj}$ and the weight parameters w_j can be solved via jointly optimizing the new regression coefficient matrices $\tilde{\mathbf{A}}_{Rj}$, $\tilde{\mathbf{A}}_{Gj}$, $\tilde{\mathbf{A}}_{Bj}$. Consequently, the optimization problem in (1) boils down to seeking an optimal sparse matrix set $\{< \tilde{\mathbf{A}}_{Rj}, \tilde{\mathbf{A}}_{Gj}, \tilde{\mathbf{A}}_{Bj} > | j = 1, \cdots, K\}$, where the sparse matrix set means that some of the matrix pairs equal to zero matrix pair $< \mathbf{O}, \mathbf{O}, \mathbf{O} >$, where $\mathbf{O}$ denotes a zero matrix. To obtain the aforementioned sparse matrix set, we adopt the regularized approach proposed in [9], in which a regularization with respect to the Frobenius norm of the coefficient matrices is imposed on the objective function, i.e., L is replaced by.

$$\begin{aligned} L = & \left\| \mathbf{X} - \sum_{i \in \{R,G,B\}} \sum_{j=1}^{K} w_j \mathbf{A}_{ij} \mathbf{Y}_{ij} \right\|_F^2 \\ & + \lambda \sum_{i \in \{R,G,B\}} \sum_{j=1}^{K} \|\tilde{\mathbf{A}}_{ij}\|_F, \end{aligned} \quad (4)$$

where $\lambda > 0$ is a trade-off parameter.

Let $\tilde{\mathbf{A}}_j = \begin{bmatrix} \tilde{\mathbf{A}}_{Rj} \\ \tilde{\mathbf{A}}_{Gj} \\ \tilde{\mathbf{A}}_{Bj} \end{bmatrix}$ and $\mathbf{Y}_j = \begin{bmatrix} \mathbf{Y}_{Rj} \\ \mathbf{Y}_{Gj} \\ \mathbf{Y}_{Bj} \end{bmatrix}$. Then, by using the objective function in (4), we obtain the following MSLSR optimization problem:

$$\arg \min_{\tilde{\mathbf{A}}_1, \cdots, \tilde{\mathbf{A}}_K} \left\| \mathbf{X} - \sum_{j=1}^{K} \tilde{\mathbf{A}}_j \mathbf{Y}_j \right\|_F^2 + \lambda \sum_{j=1}^{K} \|\tilde{\mathbf{A}}_j\|_F, \quad (5)$$

The above optimization problem can be solved using augmented Lagrangian multiplier (ALM) approach [10].

3.2. Facial Expression Recognition Based on MSLSR

Suppose that $\mathcal{I}_t$ is a testing color facial image. Let $\mathbf{y}_R = \begin{bmatrix} \mathbf{y}_{R1} \\ \vdots \\ \mathbf{y}_{RK} \end{bmatrix}$, $\mathbf{y}_G = \begin{bmatrix} \mathbf{y}_{G1} \\ \vdots \\ \mathbf{y}_{GK} \end{bmatrix}$, and $\mathbf{y}_B = \begin{bmatrix} \mathbf{y}_{B1} \\ \vdots \\ \mathbf{y}_{BK} \end{bmatrix}$ denote the facial feature vector corresponding to the R, G, and B color component, respectively. Let $\mathbf{x}_t$ denote the expression class label vector associated with $\mathcal{I}_t$. Then, based on the MSLSR model proposed in section 3, we obtain that the relationship between the facial feature vectors $\mathbf{y}_r$, $\mathbf{y}_g$, $\mathbf{y}_b$ and the corresponding expression class label vector $\mathbf{x}_t$ can be formulated as the following optimization problem:

$$\arg \min_{\mathbf{x}_t} \left\| \mathbf{x}_t - \sum_{j=1}^{K} \tilde{\mathbf{A}}_j \begin{bmatrix} \mathbf{y}_{Rj} \\ \mathbf{y}_{Gj} \\ \mathbf{y}_{Bj} \end{bmatrix} \right\|_2^2, \text{ s.t. } \mathbf{x}_t^T \mathbf{1} = 1, \mathbf{x}_t \succeq \mathbf{0}. \quad (6)$$

The optimization of (6) can be efficiently solved using quadratic programming approach [10]. After obtaining the optimal class label vector $\mathbf{x}_t$, we can assign the class label c^* to the facial image $\mathcal{I}_t$, where c^* is calculated as follows:

$$c^* = \arg \max_j \{x_{tj}\}, \quad (7)$$

where x_{tj} denotes the j-th entry of $\mathbf{x}_t$.

Table 1. Comparisons of the facial expression recognition accuracies on the Multi-PIE color facial expression database among the various methods.

Methods	Features	Color Space	Average recognition accuracies (%)						
			Disgust	Neutral	Scream	Smile	Squint	Surprise	Average
GSLSR	LBP^{u2}	Gray	63.50	77.00	92.00	68.00	59.50	83.50	73.92
	Color LBP^{u2}	RGB	64.00	76.00	93.00	70.00	62.50	84.50	75.00
		YCrCb	73.00	76.50	92.00	71.50	60.00	88.00	76.83
		Lab	71.50	78.00	93.50	73.50	65.00	89.00	78.42
	SIFT	Gray	68.50	88.00	95.00	80.50	81.00	95.00	84.67
	Color SIFT	RGB	71.00	85.00	95.50	81.50	81.50	93.50	84.67
		YCrCb	72.00	86.00	95.00	84.50	76.50	95.50	84.92
		Lab	75.00	88.50	95.00	83.50	76.50	97.50	**86.00**
Moore et al.[14]	LBP^{ms}	Gray	/	/	/	/	/	/	73.3
Moore et al.[14]	LGBP	Gray	/	/	/	/	/	/	80.4
Zheng [9]	LBP^{u2}	Gray	73.50	81.00	91.00	73.00	80.00	89.50	81.30

4. EXPERIMENTS

In this section, we conduct experiments on the Multi-PIE facial expression database [13] to evaluate the classification performance of the proposed color facial expression method. The Mulit-PIE facial expression database consists of the color facial images covering six facial expressions (disgust, neutral, scream, smile, squint, and surprise) and seven facial views. In this experiments, only the color facial images from frontal facial view ($= 0^o$) of 100 subjects are selected for the experiments. Before the experiments, each facial color image is manually cropped and normalized into the size of 120×120. We evaluate the proposed method in three different color spaces, i.e., the RGB color space, the YCrCb color space, and the CIE-Lab color space. For each color space, each facial image is decomposed into the three color components and then the feature extraction is carried out according to the method addressed in section 2.

In the experiments, we use subject independent cross validation strategy to evaluate the recognition performance of the proposed method. Specifically, all the selected 100 subjects are randomly partitioned into a training data set with 80 subjects and a testing data set with 20 subjects. Then, color image facial expression recognition method proposed in this paper is carried out. We totally conduct 10 trials of experiments using the same experimental protocol and the recognition accuracies of all the 10 trials are averaged as the average recognition accuracy. Table 1 shows the average recognition accuracies (%) with respect to the six expressions of the proposed method, in which we can see the major differences of the experimental results with respect to the different color spaces. For comparison purpose, we also provide the state-of-the-art results previously done on the same database.

From table 1, we can see that the color feature could achieve better recognition result than the gray-scale feature. Especially, the highest recognition accuracies are achieved when the CIE-Lab color space is adopted, in which the average recognition accuracy can be as high as 86.00%. This result is much better than the state of the art result. The better recognition result of the proposed method is most likely due to the following two reasons. One is the use of color image instead of the gray-scale facia image, which make it possible to obtain more useful information than the other gray-scale based approaches. The second one may attribute to the reason of the feature selection operation, which is very advantageous for finding the most discriminative facial expression features for the expression recognition.

5. DISCUSSIONS AND CONCLUSIONS

In this paper, we have investigated the method of using color local feature for color facial expression recognition, in which both color LBP^{u2} and color SIFT are respectively utilized to evaluate the performance of the proposed method. Experiments on CMU Multi-PIE facial expression database are conducted on three color spaces. The experiments demonstrate the proposed method achieves the best recognition performance when the CIE-Lab color space is used. This coincide with the results demonstrated in [6].

6. ACKNOWLEDGEMENTS

This paper was supported in part by the National Natural Science Foundation of China under Grant 61231002 and Grant 61201444, the National Basic Research Program of China under Grant 2011CB302202 and Grant 2015CB351704, the Natural Science Foundation of Jiangsu Province under Grant BK20130020, and the Ph.D. Program Foundation of Ministry Education of China under Grant 20120092110054.

7. REFERENCES

[1] Y. Tian, T. Kanade, and J. Cohn, “Facial Expression Analysis,” In Handbook of Face Recognition, pp.247-275, Springer New York, 2005.

[2] R. Niese, A. AI-Hamadi, A. Farag, H. Neumann, B. Michaelis, “Facial expression recognition based on geometric and optical flow features in colour image sequences,” IET Computer Vision, pp.1-11, 2012.

[3] G. Zhao, M. Pietikainen, “Dyanmic texture recognition using local binary pattern with and application to facial expressions,” IEEE Transactions on Pattern Analysis Machine Intelligence, vol.29, No.6, pp.915-928, 2007.

[4] M. Lyons, J. Budynek, and S. Akamatsu, “Automatic classification of single facial images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.12, pp.1357-1362, 1999.

[5] W. Zheng, H. Tang, Z. Lin, T. S. Huang, “A Novel Approach to Expression Recognition from Non-frontal Face Images,” International Conference on Computer Vision (ICCV 2009), pp.1901-1908, 2009.

[6] S.M. Lajevardi, H.R. Wu, “Facial expression recognition in perceptual color space,” IEEE Transactions on Image Processing, Vol.21, No.8, pp.3721-3733, 2012.

[7] M. Ilbeygi, H. Shah-Hosseini, “A novel fuzzy facial expression recognition system based on facial feature extraction from color face images,” Engineering Applications of Artificial Intelligence, Vol.25, No.1, pp.130-146, 2012.

[8] X. Xiong, F. Torre, “Supervised descent method and its applications to face alignment,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.532-539, 2013.

[9] W. Zheng, “Multi-view facial expression recognition based on group sparse reduced-rank regression,” IEEE Transactions on Affective Computing, Vol.5, No.1, pp.71-85, 2014.

[10] W. Zheng, M. Xin, X. Wang, B. Wang, “A novel speech emotion recognition method via incomplete sparse least square regression,” IEEE Signal Processing Letters, vol.21, no.5, pp.569-572, 2014.

[11] W. Zheng, X. Zhou, C. Zou, L. Zhao, “Facial expression recognition using kernel canonical correlation analysis (KCCA),” IEEE Transactions on Neural Networks, Vol.17, No.1, pp.233-238, 2006.

[12] I. Kotsia, I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” IEEE Transactions on Image Processing, Vol.16, No.1, pp.172-187, 2007.

[13] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, “Multi-PIE,” Image and Vision Computng, Vol.28, pp.807-813, 2010.

[14] S. Moore, R. Bowden, “Local binary patterns for multi-view facial expression recognition,” Computer Vision and Image Understanding, Vol.115, pp.541-558, 2011.