DISCRIMINATIVE FILTER BASED REGRESSION LEARNING FOR FACIAL EXPRESSION RECOGNITION

*Zizhao Zhang, Yan Yan, and Hanzi Wang**

School of Information Science and Technology, Xiamen University, Xiamen, P. R. China

zizhao.zhang@yahoo.cn; {yanyan, hanzi.wang}@xmu.edu.cn

ABSTRACT

In this paper, we propose a novel discriminative filter based regression learning (DFRL) method, which can effectively remove irrelevant information while preserving useful information for facial expression recognition. DFRL integrates the filter technique and the linear analysis techniques (i.e., Linear Discriminant Analysis-LDA and Linear Ridge Regression-LRR) to obtain an effective image representation. Two steps are involved in DFRL: 1) The discriminative filters corresponding to different facial expressions are separately trained by optimizing the cost function of the two-class LDA, 2) LRR is used to extract valuable expressional information with high discriminability from the combined filtered images. Experimental results on several challenging datasets demonstrate the superior effectiveness and generalization ability of the proposed DFRL compared with other competing methods.

Index Terms— Filter design, regression learning, facial expression recognition

1. INTRODUCTION

Automatic facial expression recognition could facilitate communication and has many potential applications, such as human computer interaction and data-driven animation. However, automatic facial expression recognition still faces many challenges, such as the variations of pose and illumination.

An effective image representation plays an important role in the success of facial expression recognition. The representation of facial expression images can be roughly classified into geometric feature-based methods and appearance-based methods. Geometric feature-based methods have achieved good results on active unit recognition. However, the precise localization of facial feature points is still a challenging task. Appearance-based facial image representation methods, including Local Binary Patterns (LBP) [1] and LDA [2], have been successfully applied in facial expression analysis. Several LBP-based methods have been proposed, such as m-LBP (representing salient micro-patterns of face images) [3] and Boost-LBP [4], which have shown superior performance

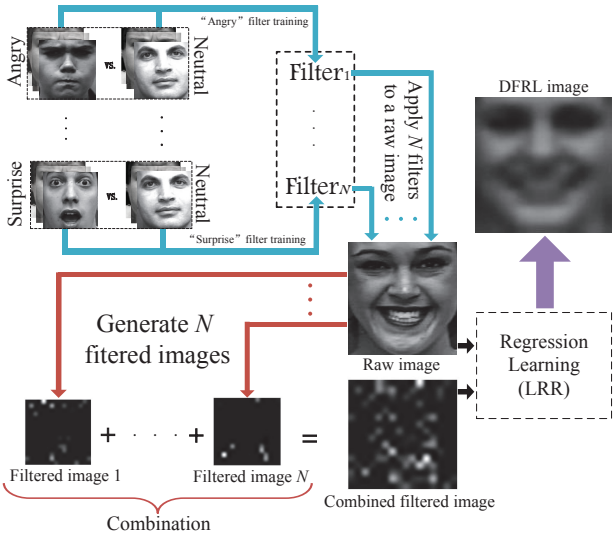


Fig. 1. The outline of the proposed DFRL method.

over the Gabor wavelet-based methods. The drawback of the LBP-based methods is their sensitiveness to the size of the chosen block. As a matter of fact, the goal of representing facial image can be considered as filtering out irrelevant information (i.e., facial feature difference) in the test images so that valuable information (e.g., wrinkled eyebrow, smiling mouth and other features which are discriminative in different expressions) is enhanced.

The above-mentioned methods consider a facial image as a whole without specifying its important parts (e.g., eye and mouth). There are some learning-based methods which try to learn part-based facial image representations. For instance, Zhong *et al.* [5] proposed a multi-task sparse learning framework to explore the discriminative information in specific facial patches of different expressions. In addition, improved nonnegative matrix factorization (NMF) algorithms have been successfully applied to facial expression recognition task, such as the graph-preserving sparse NMF (GSNMF) algorithm [6].

LDA is a supervised method which attempts to seek for a

*-Corresponding author

transformation of variables best explaining data. The goal of LDA tries to make data distributions in different classes far apart in the re-mapped subspace. Traditional LDA handles two-class problems. LDA can also be extended to multi-class cases (where the between-class matrix is the sum of the pairwise distances of any two different classes). However, multi-class LDA suffers from the problem of unbalanced pairwise distances which can reduce the recognition performance [7].

In this study, we propose a novel discriminative filter based regression learning method for effective facial image representation. To be specific, we propose to employ LDA in the design of image filters. The discriminative filters (DFs) based on the cost function of the two-class LDA are separately trained to discriminate a specific expression from the neutral expression. When the DFs are applied to a multi-class facial expression problem, we use a regression learning method (i.e., LRR) based on the filter technique to solve the multi-class classification problem. The main idea of using LRR is to extract the optimal feature information with high discriminability from the filtered images. Experimental results on several popular facial expression datasets demonstrate the effectiveness of the proposed method.

2. DISCRIMINATIVE FILTER BASED REGRESSION LEARNING METHOD

An effective image filter can increase the discriminability of facial expression images. To achieve this, a key step is to design a suitable principle for learning filters. In our work, we combine the objective function of LDA with a filter function. From Section 2.1 to 2.3, we introduce how to build the relationship between the filter function and the cost function of LDA, and give the details of the proposed filter learning procedure. Moreover, we propose to use a regression learning method to extract highly discriminative information in the filtered facial images shown in Sections 2.4 and 2.5. The outline of the proposed DFRL is shown in Fig. 1.

2.1. Linear Discriminative Filter

The linear discriminative filter function f is defined as [8]

$$f(\lambda, p) = \lambda p \tag{1}$$

where f is the dot product of a d-dimensional column vector λ and a face image p (represented as a d-dimensional column vector), and $\lambda, p, f \in \mathbb{R}^d$. Thus λ_i decides the intensity of the pixel p_i that is allowed to pass through. We also define a matrix $P = [p_1, \cdots, p_n]$ in $\mathbb{R}^{d\times n}$ (n is the number of images), and a matrix $F(\lambda, P) = [f(\lambda, p_1), \cdots, f(\lambda, p_n)]$ in $\mathbb{R}^{d\times n}$. The differentiability of the filter function is an important property for optimization.

2.2. Linear Discriminant Analysis

As a popular feature extraction algorithm, LDA attempts to seek for a linear transformation which minimizes the intra-class variation S_W while maximizing the inter-class variation S_B by maximizing the cost function of LDA. We select the neutral faces and the faces with a non-neutral expression as the inputs of LDA. We define the cost function as

$$L_{DA}(X_0, X_1) = \frac{\omega^T S_B(X_0, X_1)\omega}{\omega^T S_W(X_0, X_1)\omega} \tag{2}$$

where $X_0 = F(\lambda, P_0)$, and P_0 is the set of the neutral faces; $X_1 = F(\lambda, P_1)$, and P_1 is the set of the faces with an expression other than the neutral expression. ω is the linear transformation in $\mathbb{R}^d$. And we have

$$S_B(X_0, X_1) = (m_1 - m_0)(m_1 - m_0)^T \tag{3}$$

$$\begin{aligned} S_W(X_0, X_1) = &(X_1 - M_1)(X_1 - M_1)^T \\ &+(X_0 - M_0)(X_0 - M_0)^T \end{aligned} \tag{4}$$

where the column vector m_i is the mean of X_i ($i = \{0, 1\}$) in $\mathbb{R}^d$. The matrix M_i includes n copies of m_i. The optimal ω can be computed [8] as

$$\begin{aligned} \hat{\omega} &= \arg\max_{\omega} \frac{\omega^T S_B(X_0, X_1)\omega}{\omega^T S_W(X_0, X_1)\omega} \\ &= S_W(X_0, X_1)^{-1}(m_1 - m_0) \end{aligned} \tag{5}$$

From (1) to (5), the relationship between the cost function of LDA L_{DA} and the filter function f is built up.

2.3. Discriminative Filter Design

Based on Sections 2.1 and 2.2, the final goal is to maximize the cost function of LDA in DF. Alternatively, we treat this problem as a minimization problem which can be solved by using the traditional optimization method. We define $O(\lambda)$ as

$$O(\lambda) = \log \frac{1}{L_{DA}(F(\lambda, P_0), F(\lambda, P_1))} + Ctr(\lambda^T\lambda) \tag{6}$$

where $tr(\lambda^T\lambda)$ is a regularization term, and C is a constant. As a result, the problem can be solved by finding an optimal λ by minimizing $O(\lambda)$. Since f and L_{DA} are differentiable, we choose the gradient descent algorithm [9] to find the local minimum effectively.

2.4. Linear Ridge Regression

LDA attempts to solve the two-class classification problem (i.e., a non-neutral expression vs. a neutral expression). Thus a trained DF can discriminate only one specific expression from the neutral expression. Suppose there are N different expressions, N DFs (each corresponds to one expression) are required to be trained. When these N DFs are used to filter a test image respectively, N filtered images are generated, where only one filtered image (corresponding to the test expression) is enhanced while the other filtered images are suppressed. The enhanced filtered image keeps the valuable

information in the tested expression. The problem is how to extract the valuable information among the N filtered images. Based on the observation that the correlation between the enhanced filtered image and the test image is higher than those between the suppressed filtered images and the test image, regression learning can be used to extract the valuable information in the enhanced image.

The N filtered images are combined as follows:

$$G = \sum_{i=1}^{N} F(\lambda_i, P) \tag{7}$$

where G is the combination of one enhanced image and $(N-1)$ suppressed images. The irrelevant information in the suppressed images can be considered as noises and should be removed without impairing the valuable information in the enhanced filtered image. Hence, we propose to use the regression learning method as described below.

As an improved variant of the Least Square (LS) estimation method, LRR [8] overcomes the potential drawbacks of LS. LS is not effective if independent variables are well-correlated. The variance estimation of LS may be large due to limited samples that are used [7]. LRR tries to reduce the variance by adding a regularization term. Hence, we reserve the valuable information in G by using LRR.

LRR is used to resolve an optimization problem

$$\min_{K} \|P - K^T G\|^2 + \beta \|K^T I\|^2 \tag{8}$$

where I is a diagonal matrix in $\mathbb{R}^{d \times d}$; $\|K^T I\|^2$ is the regularization term, and β is the regularization parameter; K is a transformation which projects the samples to a new space. We can compute K as

$$\hat{K} = (GG^T + \beta I)^{-1} G P^T \tag{9}$$

Hence, the estimated Y can be obtained as

$$Y = \hat{K}^T G \tag{10}$$

The advantage of using LRR instead of LS is that the irrelevant information (i.e., noises) in G can be removed by regressing from the original image to the combined filtered image (see Fig. 2). For example, when there is an input raw image with a "smile" expression, the most valuable information is around the area of the mouth. After applying N DFs to this image, it yields $(N-1)$ suppressed images (which can be treated as irrelevant information). When this image is filtered by the "angry" filter (i.e., a filter is trained for the angry expression) which emphasizes the eyebrows information (i.e., wrinkled eyebrows), because the raw "smile" image does not have the wrinkled eyebrows appearance, this irrelevant information (i.e., wrinkled eyebrows) in the combined filtered image will be effectively removed by using LRR.

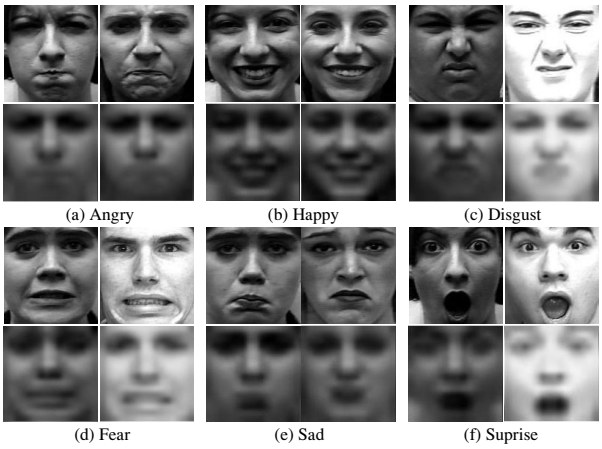


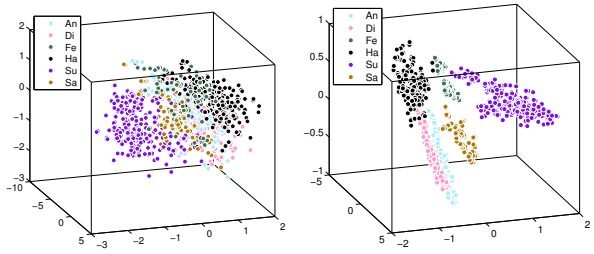
Fig. 2. The raw images (1^{st} and 3^{rd} rows) and the corresponding DFRL images (2^{nd} and 4^{th} rows) of six expressions. We can see that the similarity of the images obtained by the proposed DFRL is higher than that of the raw images. This is due to the irrelevant information (e.g., personal facial feature difference) is removed and only the valuable information around the area of mouths and eyes is preserved.

2.5. Regularization Parameter Estimation

The regularization parameter β in (8) is vital for the performance of LRR, which can be estimated by using cross-validation [10]. We use 10-fold cross-validation to estimate the optimal regularization parameter. When $\beta = 0$, m degenerates to the LS estimation. In this case, the regressed images will be similar to the raw images without losing much information. We should avoid that case because the irrelevant information is also reserved which can reduce the recognition performance. However, when β is properly chosen, the regressed images will only reserve the valuable information. To demonstrate the effectiveness of β visually, we use N sets of data with N expressions and apply our proposed method to these sets respectively. Then, PCA is used to obtain the first 3 principal components under different β values. As shown in Fig. 3, we can see that the distributions of the samples of different expressions in the subspace can be separated well if the β value is properly chosen.

3. EXPERIMENTS

In this section, we use the extended Cohn-Kanade (CK+) dataset [11] to evaluate the performance of the proposed D-FRL. We also compare DFRL with several other competing methods. Moreover, we use a hybrid dataset composed of three popular facial expression datasets (i.e., CK+, JAFFE [12], and MMI [13]) to demonstrate the generalization ability of the proposed method. In all experiments, six basic expressions (i.e., Angry (An), Disgust (Di), Fear (Fe), Happy (Ha),



(a) $\beta = 0.1$ (b) $\beta = 2$

Fig. 3. The distribution of the six expressions in the first 3 principal components obtained by DFRL under two different regularization parameter values.

Surprise (Su) and Sad (Sa)) and one neutral expression are selected. CK+ and MMI are taken from video sequences. In each video sequence, three frames with the most frequent expression intensity and one frame with the neutral expression are selected. The manually cropped face images are resized to the same size of (16×16) and converted to the gray scale.

3.1. Experiments on the Extended Cohn-Kanade Dataset

CK+ includes 593 short videos and 123 subjects. Firstly, we randomly select 50% of the samples as the training set to train the filters. The remaining 50% samples are used as the testing set. Then, we apply DFRL to the training and testing sets respectively. Finally, the recognition results of the proposed DFRL are obtained by using the SVM (the RBF kernel is used) shown in Table 1 and KNN classifiers [8], respectively.

From Table 1, DFRL can achieve good performance on most expressions. However, we can see that Sa and Di are often misclassified as An. This is because that both Di and Sa have the wrinkled eyebrow expression which is similar to An. Moreover, we observe that the subjects with Sa and Di sometimes do not have obvious motions around the mouth area, so these two expressions may be misclassified as An.

We compare the proposed DFRL with the other state-of-the-art methods which are based on different popular facial image representation algorithms, including m-LBP [3], Boosted-LBP [4], GSNMF [6] and CSPL [5], and two basic methods (i.e., multiclass-LDA and PCA [2]). As shown in Table 2, we can see that the proposed DRFL can effectively extract the discriminative information in the facial images for facial expression recognition and significantly outperforms the other competing methods under small image size.

3.2. Experiments on the Hybrid Dataset

The discrepancies, such as illumination, face color and nationality, in different datasets are large, which may have negative influence on the recognition performance. Furthermore,

	An	Di	Fe	Ha	Su	Sa
An	**97.01**	2.99	0	0	0	0
Di	2.27	**96.59**	0	1.14	0	0
Fe	0	0	**100**	0	0	0
Ha	0	0	0	**100**	0	0
Su	0	0	0	0	**100**	0
Sa	2.5	0	0	0	0	**97.5**

Table 1. The confusion matrix obtained by using SVM (with RBF kernel) on the CK+ facial expression dataset.

Methods	Recognition Rate (%)	Image Size (pixels)
m-LBP, 2005	88.4	110×150
Boosted-LBP, 2009	91.1	110×150
GSNMF, 2011	94.3	60×60
CSPL, 2012	89.9	96×96
PCA+SVM	47.3	16×16
LDA+SVM	87.1	16×16
DFRL+SVM	**98.7**	**16×16**
DFRL+KNN	**96.6**	**16×16**

Table 2. Comparisons with the competing methods on the CK+ facial expression dataset.

the small CK+ dataset can hardly verify the generalization ability of DFRL if the randomly selected training set covers all the challenging expressions. JAFFE is a female expression dataset consisting of 219 images from 10 persons. MMI includes 213 video sequences with six basic expressions, among which parts of the participants wear hats, hoods and glasses. By using this hybrid dataset, the proposed DFRL still achieves good performance. The recognition results are 95.3% for DFRL+SVM and 91.0% for DFRL+KNN, but only 80.6% for LDA+SVM and 74.2% for LDA+KNN.

4. CONCLUSION

In this paper, a novel image representation method called discriminative filter regression learning (DFRL) is proposed for effective facial expression recognition. The strategy of the discriminative filter design is to optimize the cost function of the LDA by a gradient descent procedure, which can filter out useless information in human face. Furthermore, a regression learning method is employed to explore the most discriminative information. Experimental results show the superior performance of DFRL compared to the competing methods.

5. ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China (61170179, 61201359), the Special Research Fund for the Doctoral Program of Higher Education of China (20110121110033), and the Xiamen Science & Technology Planning Project Fund (3502Z20-116005) of China, and the Natural Science Foundation of Fujian Province of China (2012J05126).

6. REFERENCES

[1] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 915–928, 2007.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 711–720, 1997.

[3] C. Shan, S. Gong, and P.W. McOwan, "Robust facial expression recognition using local binary patterns," in *Image Processing. IEEE*, 2005, vol. 2, pp. 370–373.

[4] C. Shan, S. Gong, and P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, pp. 803–816, 2009.

[5] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D.N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition. IEEE*, 2012, pp. 2562–2569.

[6] R. Zhi, M. Flierl, Q. Ruan, and W.B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE*, vol. 41, pp. 38–52, 2011.

[7] S. An, W. Liu, and S. Venkatesh, "Face recognition using kernel ridge regression," in *Computer Vision and Pattern Recognition. IEEE*, 2007, pp. 1–7.

[8] C. Bishop., "Pattern recognition and machine learning," *Springer*, 2006.

[9] Carl Edward Rasmussen and Christopher K. I. Williams, "Gaussian processes for machine learning," *The MIT Press*, 2006.

[10] M.plutowski., "Cross-validation in theory and in practice," *Research Report. Dept. of Computational Science Reserach, David Sarnoff Reserach Center, Princeton, New Jersey.*, 1996.

[11] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops. IEEE*, 2010, pp. 94–101.

[12] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, pp. 1357–1362, 1999.

[13] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo. IEEE*, 2005, pp. 317–321.