BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **16**, No 4
Sofia • 2016 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.1515/cait-2016-0067

Learning a Class-Specific Dictionary for Facial Expression Recognition

Shiqing Zhang[1], *Gang Zhang*[2], *Yueli Cui*[1], *Xiaoming Zhao*[1]

[1]*Institute of Intelligent Information Processing, Taizhou University, Taizhou, China*
[2]*Institute of Guangzhou Quality Supervision and Testing, Guangzhou, China*
Emails: tzczsq@163.com tzxyzz@126.com cuiyueli@tzc.edu.cn tzxyzxm@163.com

Abstract: *Sparse coding is currently an active topic in signal processing and pattern recognition. MetaFace Learning* (*MFL*) *is a typical sparse coding method and exhibits promising performance for classification. Unfortunately, due to using the l_1-norm minimization, MFL is expensive to compute and is not robust enough. To address these issues, this paper proposes a faster and more robust version of MFL with the l_2-norm regularization constraint on coding coefficients. The proposed method is used to learn a class-specific dictionary for facial expression recognition. Extensive experiments on two popular facial expression databases, i.e., the JAFFE database and the Cohn-Kanade database, demonstrate that our method shows promising computational efficiency and robustness on facial expression recognition tasks.*

Keywords: *Sparse coding, metaface learning, sparse representation, facial expression recognition, robustness.*

1. Introduction

Facial expression is the main manner of expressing and interpreting the affective states of human beings. Facial expression recognition focuses on distinguishing the human affective states by using facial expression. During the last two decades, facial expression recognition has become a hot research topic in pattern recognition, artificial intelligence and computer vision, owing to its important applications in human-computer interaction, artificial intelligence, security monitoring, social entertainment [1-3].

As far as the classification task of facial expression is concerned, a variety of conventional classification methods have been applied so far for facial expression recognition, such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and so on. In recent years, a new type of classification method called Sparse Representation based Classification (SRC） [4] has been used successfully for face recognition. However, SRC directly employs all the training samples to construct the dictionary,

resulting in lots of redundancy, noise, and trivial information in the pre-defined dictionary. In addition, when the training samples grow, SRC will suffer from the computation bottleneck since it uses the l_1-norm sparsity constraint on coding coefficients. In recent years, Yang et al. [5] developed a MetaFace Learning (MFL) method to learn a class-specific dictionary, in which metafaces are learned from the original images and then used as the dictionary to represent the input query image. It has been found in [5] that MFL is more effective than SRC and it also brings performance improvement over SRC.

However, it is noted that MFL has two shortcomings. First, similar to SRC, MFL also uses the l_1-norm sparsity constraint on coding coefficients, which needs to be solved by a time-consuming iteration process. Second, in MFL the coding fidelity is measured by the l_1-norm of coding residual under the assumption that the coding residual follows the Gaussian distribution. However, in practice this assumption may not hold well in noisy environment, where the coding residual may not conform to the Gaussian distribution. To overcome these drawbacks of MFL, we modify the objection function of MFL and derive its analytical solution with the l_2-norm regularization constraint on coding coefficients. This solution gives rise to a faster and more robust version of MFL, which is called FR-MFL. The effectiveness of the proposed FR-MFL method is verified on facial expression recognition tasks.

The remainder of this paper is organized as follows. In Section 2, the original MFL is reviewed briefly. The proposed FR-MFL is described in detail in Section 3. Section 4 gives the experiment results and analysis. Finally, conclusions are drawn in Section 5.

2. Metaface learning

MetaFace Learning (MFL) [5] aims to learn a class-specific dictionary for each object from the original training samples and then uses the learned dictionary to represent the input query image.

Suppose the training samples are denoted by $\mathbf{D} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_c] \in R^{d\times N}$ where $\mathbf{X}_c \in R^{d\times N_c}$ is the subset of all the N_c vector-represented training samples from class c, and d is the feature dimension, $N = \sum_{i=1}^{c} n_i$ is the total number of samples.

In SRC, a new test sample $x \in R^d$ can be sparsely coded by the following l_1-minimization optimization problem:

$$\alpha = \arg\min_{\alpha} \left\| x - \mathbf{D}\alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_1, \tag{1}$$

where λ is a positive scalar number which is used as a tradeoff between the reconstructed error and the coefficients' sparsity.

In MFL, a class-specific dictionary $\mathbf{D}_i$ is learned by

$$\left(\mathbf{D}_i, \mathbf{A}_i\right) = \arg\min_{\mathbf{D}_i, \mathbf{A}_i} \left\| \mathbf{X}_i - \mathbf{D}_i\mathbf{A}_i \right\|_2^2 + \lambda \left\| \mathbf{A}_i \right\|_1, \tag{2}$$

$$\text{s.t. } \left\| d_j^i \right\|_2 = 1, \quad j = 1, 2, ..., K,$$

where $\mathbf{X}_i \in R^{d\times N_i}$ denotes all the training samples from the i-th class, d_j^i represents the j-th column of the i-th class-specific sub-dictionary $\mathbf{D}_i = [d_1^i, \ldots, d_K^i] \in R^{d\times K}$, and $\|\mathbf{A}_i\|_1$ is the summation of l_1-norm of all the columns of $\mathbf{A}_i = [\alpha_1^i, \ldots, \alpha_{N_i}^i] \in R^{d\times N_i}$, i.e., $\|\mathbf{A}_i\|_1 = \sum_j^{N_i} \|\alpha_j^i\|_1$.

In Equation (2), a joint optimization problem of the metafaces **D** and the representation coefficient matrix **A** needed to be solved. As a multi-variable optimization problem, Equation (2) can be solved by optimizing **D** and **A** alternatively, as described below.

When fixing **D**, the objective function in Equation (2) can be reduced to

$$\mathbf{A}_i = \arg\min_{\mathbf{A}_i} \|\mathbf{X}_i - \mathbf{D}_i\mathbf{A}_i\|_2^2 + \lambda\|\mathbf{A}_i\|_1. \tag{3}$$

That equation is a convex optimization methods and can be obtained by quadratic programming, such as the iterative l_1-regularized least squares (l_1-ls) [6] algorithm. Fixing $\mathbf{A}$, $\mathbf{D}$ can be updated by solving the following objection function

$$\mathbf{D}_i = \arg\min_{\mathbf{D}_i} \|\mathbf{X}_i - \mathbf{D}_i\mathbf{A}_i\|_2^2, \tag{4}$$

$$\text{s.t. } \|d_j^i\|_2 = 1, \quad j = 1, 2, ..., K.$$

That equation can be solved by using the Langrage multiplier, the final analytical solution is

$$d_j = Y\alpha_j^{\mathrm{T}} / \|Y\alpha_j^{\mathrm{T}}\|_2, \tag{5}$$

where $Y = X - \sum_{l\neq j} d_l\alpha_l$. When updating d_j, all the other columns of **D**, i.e., d_l, $l \neq j$, are fixed.

3. The proposed FR-MFL method

3.1. Motivation

There are two drawbacks of the original MFL, as described below.

First, like SRC, MFL [5] imposes the l_1-sparsity constraint in Equation (2) on the representation coefficients to regularize the solution. Nevertheless, the l_1-minimization takes a time-consuming iterative process due to its large computation cost. Therefore, it is desirable to decrease the computation cost in MFL.

Second, In MFL the coding fidelity is measured by the l_1-norm of coding residual under the assumption that the coding residual follows the Gaussian distribution. However, images usually contain some additive noise, so this assumption may not hold well in noisy environment. Recently, it has been proved in [7] the l_2-norm could be used to characterize the data fidelity for an optimal maximum a posterior estimation when the observed image contains some additive Gaussian noise.

3.2. Our method

Based on the abovementioned two points, by modifying the objection function of MFL, we can get a Faster and more Robust MFL (FR-MFL) algorithm. In detail, the objection function in Equation (2) can be rewritten as

$$(\mathbf{D}_i, \mathbf{A}_i) = \arg\min_{\mathbf{D}_i, \mathbf{A}_i} \left\| \mathbf{X}_i - \mathbf{D}_i \mathbf{A}_i \right\|_2^2 + \lambda \left\| \mathbf{A}_i \right\|_2^2, \quad \text{s.t.,} \ \left\| d_j^i \right\|_2 = 1, \quad j = 1, 2, ..., K, \tag{6}$$

where $\| \ \|_2^2$ is the l_2-norm.

Like MFL, the proposed FR-MFL also needs to solve a joint optimization for the metafaces **D** and the representation coefficient matrix **A**. In other words, **D** and **A** in Equation (6) can be obtained by optimizing **D** and **A** alternatively, as described below.

When fixing **D**, based on the l_2-norm regularization constraint, **A** can be solved by using the Langrage multiplier, and the final obtained analytical solution is

$$\alpha = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda I)^{-1}\mathbf{X}^{\mathrm{T}}. \tag{7}$$

Fixing **A**, **D** can be updated by solving Equation (4), as done in MFL.

It's worth pointing out that this step of optimizing **A** in the proposed FR-MFL method is different from MFL. First, with the aid of the l_2-norm regularization constraint we can directly derive the analytical solution, and hence avoid the expensive computation in original MFL. Second, the l_2-norm is more effective in characterizing the data fidelity in noisy environment. It makes the proposed FR-MFL yield more robust feature representations than MFL. The advantages of FR-MFL are verified in the following experiments.

4. Experiments

4.1. Experiment setup

To validate the proposed FR-MFL, we performed facial expression recognition experiments on two popular facial expression databases, i.e., the JAFFE database [8] and the Cohn-Kanade database [9].

The JAFFE database has 213 images of female facial expression. Each image has a resolution of 256×256 pixels. The Cohn-Kanade database contains 100 university students. Each image has a resolution of 640×490 pixels. As done in [10], on the Cohn-Kanade database we selected 320 image sequences from 96 subjects, with 1 to 6 emotions per subject. For every sequence, the neutral face and one peak frames were employed for prototypic expression recognition, giving in total 470 images (32 anger, 100 joy, 55 sadness, 75 surprise, 47 fear, 45 disgust and 116 neutral). Figs 1 and 2 separately show some sample images from the JAFFE database and the Cohn-Kanade database.

Two experiments are performed with different configurations. The first one classifies facial expressions with the popular Local Binary Patterns (LBP) [10-13] features extracted from original clean images without any corruption. In this case,

according to the normalized value of the eye distance, a resized image of 110×150 pixels was cropped from original images before performing LBP operators, as done in [10]. The second one recognizes facial expressions on corrupted images to verify the robustness of the proposed method. In this case, all images are resized into 32×32 pixels and then the random pixel corruption is implemented to generate corrupted images.

The proposed FR-MFL is compared with linear SVM, SRC, and MFL, respectively. For SRC and MFL, the l_1-norm minimization is solved by using the iterative l_1-regularized least squares (l_1-ls) [6] algorithm. A five-fold cross validation scheme is implemented in all facial expression recognition experiments, and the average recognition results are reported. The experiment platform is Intel CPU 2.10 GHz, 1G RAM memory, MATLAB 2012a.

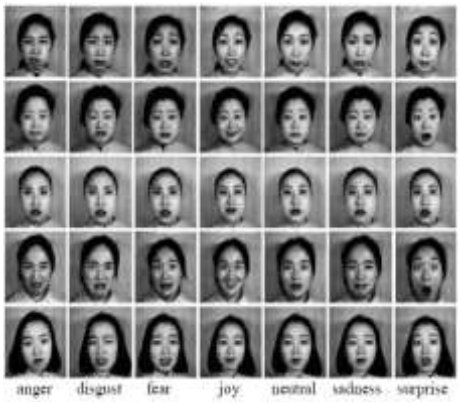

Fig. 1. Examples of facial expression images from the JAFFE database

Fig. 2. Examples of facial expression images from the Cohn-Kanade database

4.2. Experiments without corruption

As used in [10], we employed the 59-bin operator $LBP_{P,R}^{u2}$, and divided the cropped images of 110×150 pixels into 18×21 pixels regions, yielding a feature vector

length of 2478 (59×42) represented by the LBP histograms. Tables 1 and 2 give the recognition performance of all used methods, including SRC, MFL, and FR-MFL on the JAFFE database and the Cohn-Kanade database, respectively. The results are summarized in Tables 1 and 2. From the tables it is easy to observe that FR-MFL obtains an accuracy of 85.71% and 98.09% on the two databases, respectively. It performs better than MFL and SRC, and significantly outperforms the baseline SVM. This shows that l_2-norm regularization constraint is more effective for classification than l_1-norm sparsity constraint since l_2-norm can give higher ability to avoid overfitting than l_1-norm.

Table 3 presents a comparison of computation time between FR-MFL and MFL to evaluate their computation efficiency. Computation time is represented by the whole operation time for training and testing when performing classification for all face images in the corresponding face database. From Table 3, it can be observed that FR-MFL is about 2.39 times faster than MFL on the JAFFE database and about 2.46 times faster on the Cohn-Kanade database, respectively. This validates the advantages of inducing an analytical solution in FR-MFL with the l_2-norm regularization constraint.

Table 1. Recognition accuracy (%) of different methods on the JAFFE database

Method	FR-MFL	MFL	SRC	SVM
Accuracy	**85.71**	84.76	84.76	79.88

Table 2. Recognition accuracy (%) of different methods on the Cohn-Kanade database

Method	FR-MFL	MFL	SRC	SVM
Accuracy	**98.09**	97.57	97.14	95.24

Table 3. Comparison of computation time between FR-MFL and MFL

Dababase	JAFFE		Cohn-Kanade	
Method	FR-MFL	MFL	FR-MFL	MFL
Computation time (s)	167.862	402.341	250.777	617.993
Speed-up	**2.39 times**		**2.46 times**	

4.3. Experiments with corruption

In this section, we evaluate the robustness of the proposed FR-MFL method on facial expression recognition. The percentage of image pixels are randomly selected from each testing image and then replaced by random values in the range of [0, p_j], where p_j denotes the maximum value of pixels in the j-th test image. In our experiments, we change the percentage of corrupted pixels from 0 up to 90%. Fig. 3 gives an example of corrupted image on the Cohn-Kanade database. In the figure, the original image is resized to 32×32 pixels, and then is performed with 50% random pixel corruption.

Figs 4 and 5 show the recognition results of different methods under different percentage of pixel corruption. From the results, we can observe that recognition accuracies of all methods are dropped with the increase of pixel corruption. Nevertheless, the proposed FR-MFL constantly outperforms the other methods such

as SRC, SVM, and MFL. The advantage of FR-MFL over MFL clearly validates that the introduced l_2-norm regularization constraint produces better robustness to image noise.

(a) (b) (c)

Fig. 3. An example of corrupted image in the Cohn-Kanade database: Original image of 640×490 pixels (a); resized image of 32×32 pixels (b); 50% corrupted image (c)

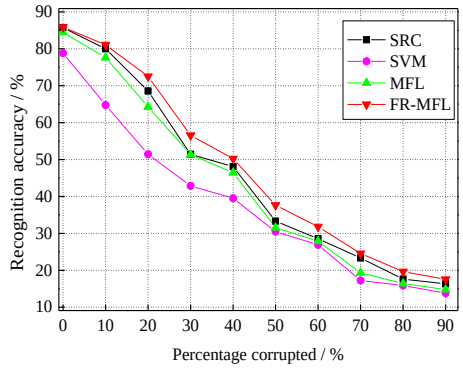

Fig. 4. Recognition accuracy under different percentage corrupted on the JAFFE database

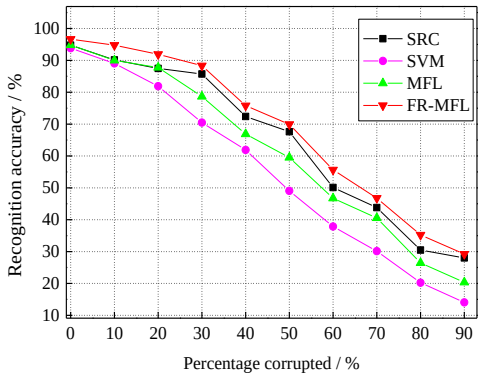

Fig. 5. Recognition accuracy under different percentage corrupted on the Cohn-Kanade database

5. Conclusion

This paper presents a faster and more robust version of MFL (FR-MFL) for facial expression recognition via the l_2-norm minimization. The proposed FR-MFL

method outperforms MFL, SRC, and SVM on facial expression recognition. More importantly, the proposed FR-MFL is much more effective than MFL in computation. This can be attributed to two aspects. Firstly, l_2-norm regularization constraint in FR-MFL is more effective for classification than l_1-norm sparsity constraint since l_2-norm can give higher ability to avoid overfitting than l_1-norm. Secondly, l_2-norm regularization constraint in FR-MFL presents an analytical solution of the objective function.

***Acknowledgements*:** This work is supported by Zhejiang Provincial Natural Science Foundation of China under Grant No LY16F020011 and No LQ15F020001, National Natural Science Foundation of China under Grant No 61203257 and No 61272261, and Taizhou Science Plan No 1501KY64.

References

1. Zheng, W., X. Zhou, M. Xin. Color Facial Expression Recognition Based on Color Local Features. – In: Proc. of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, 2015, pp. 1528-1532.
2. Eleftheriadis, S., O. Rudovic, M. Pantic. Discriminative Shared Gaussian Processes for Multiview and View-Invariant Facial Expression Recognition. – IEEE Transactions on Image Processing, Vol. **24**, 2015, No 1, pp. 189-204.
3. Wang, X., A. Liu, S. Zhang. New Facial Expression Recognition Based on FSVM and KNN. – Optik International Journal for Light and Electron Optics, Vol. **126**, 2015, No 21, pp. 3132-3134.
4. Wright, J., A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma. Robust Face Recognition via Sparse Representation. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **31**, 2009, No 2, pp. 210-227.
5. Yang, M., L. Zhang, J. Yang, D. Zhang. Metaface Learning for Sparse Representation Based Face Recognition. – In: Proc. of 2010 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, 2010, pp. 1601-1604.
6. Kim, S. J., K. Koh, M. Lustig, S. Boyd, D. Gorinevsky. An Interior-Point Method for Large-Scale l1-Regularized Least Squares. – IEEE Journal of Selected Topics in Signal Processing, Vol. **1**, 2007, No 4, pp. 606-617.
7. Dong, W., L. Zhang, G. Shi, X. Wu. Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. – IEEE Transactions on Image Processing, Vol. **20**, 2011, No 7, pp. 1838-1857.
8. Lyons, M. J., J. Budynek, S. Akamatsu. Automatic Classification of Single Facial Images. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **21**, 1999, No 12, pp. 1357-1362.
9. Kanade, T., Y. Tian, J. Cohn. Comprehensive Database for Facial Expression Analysis. – In: Proc. of International Conference on Face and Gesture Recognition, Grenoble, France, 2000, pp. 46-53.
10. Shan, C., S. Gong, P. Mc Owan. Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study. – Image and Vision Computing, Vol. **27**, 2009, No 6, pp. 803-816.
11. Zhao, X., S. Zhang. Facial Expression Recognition Using Local Binary Patterns and Discriminant Kernel Locally Linear Embedding. – EURASIP Journal on Advances in Signal Processing, Vol. **2012**, 2012, No 1, pp. 20.
12. Li, X., Q. Ruan, Y. Jin, G. An, R. Zhao. Fully Automatic 3D Facial Expression Recognition Using Polytypic Multi-Block Local Binary Patterns. – Signal Processing, Vol. **108**, 2015, pp. 297-308.
13. Luo, Y., C.-M. Wu, Y. Zhang. Facial Expression Recognition Based on Fusion Feature of PCA and LBP with SVM. – Optik-International Journal for Light and Electron Optics, Vol. **124**, 2013, No 17, pp. 2767-2770.