

Chapter 2

Efficient Discriminative K-SVD for Facial Expression Recognition

Weifeng Liu, Caifeng Song and Yanjiang Wang

Abstract Dictionary learning has attracted growing intention for its prominent performance in many computer vision applications including facial expression recognition (FER). Discriminative K-SVD (D-KSVD) is one of conventional dictionary learning methods, which can effectively unify dictionary learning and classifier. However, the computation is huge when applying D-KSVD directly on Gabor features which has high dimension. To tackle this problem, we employ random projection on Gabor features and then put the reduced features into D-KSVD schema to obtain sparse representation and dictionary. To evaluate the performance, we implement the proposed method for FER on JAFFE database. We also employ support vector machine (SVM) on the sparse codes for FER. Experimental results show that the computation is reduced a lot with little performance lost.

Keywords Facial expression recognition · Sparse representation · K-SVD · Discriminative K-SVD · Random projection · Gabor

2.1 Introduction

In recent years, many new technical methods have been exploited for face recognition and facial expression recognition [1–4]. Sparse representation based classification has been performed well in facial expression recognition which

W. Liu (✉) · C. Song · Y. Wang
College of Information and Control Engineering, China University of Petroleum
(East China), Qingdao 266580, People's Republic of China
e-mail: liuwf@upc.edu.cn

Y. Wang
e-mail: yjwang@upc.edu.cn

exploits sparse coding to approximate an input expression image by a sparse linear combination of samples from an over-complete dictionary. In particular, dictionary learning, which aims to learn a small dictionary including few atoms from huge amounts of original information, has achieved many successful applications including audio and vision data processing [5, 6], image analysis [7–9] and certainly facial expression recognition [10]. And K-SVD [11] is state-of-the-art dictionary learning method.

Discriminative K-SVD algorithm [12, 13] extends basic K-SVD algorithm by incorporating the linear classifier into K-SVD algorithm and finally unifies the representation power and discriminate ability to train the dictionary and classifier simultaneously. D-KSVD algorithm has been proven effective and efficiency in image classification. In [10], Liu etc. utilize D-KSVD algorithm working on Gabor feature in facial expression recognition which significantly boosts the performance. However, the high dimension of facial features will cost a lot of learning time.

On the other hand, random projection (RP) [14, 15] can project original high-dimensional data onto a low-dimensional subspace using a random matrix. And Johnson–Lindenstrauss (JL) lemma [16] identifies that RP can preserve the distance between two points. In this paper, we introduce random projection as a preprocessing for feature selection and then incorporate the reduced dimensional feature into D-KSVD framework. As a result, the proposed method can effectively reduce the computation and then save the training time significantly with only a little lost of performance. Finally, we carefully construct the experiments on JAFFE dataset [25]. Experimental results demonstrate the superiority of the proposed method.

We also employ SVM on the sparse codes for FER. The verified experiments achieved the approximate results with D-KSVD algorithm.

The rest of this paper is arranged as follows. Section 2.2 introduces discriminative K-SVD algorithm in detail. Section 2.3 describes the method of random projection. Section 2.4 presents the experiment result and analysis. Finally, we conclude with discussion in Sect. 2.5.

2.2 Discriminative K-SVD for Facial Expression Recognition

2.2.1 Sparse Representation of Facial Expression Images

The basic idea of sparse representation [2–4] is using the over-complete dictionary to replace the traditional orthogonal basis and then finding the best linear combination of several atoms to represent a signal. Figure 2.1 shows the decomposition of a facial expression image using sparse representation.

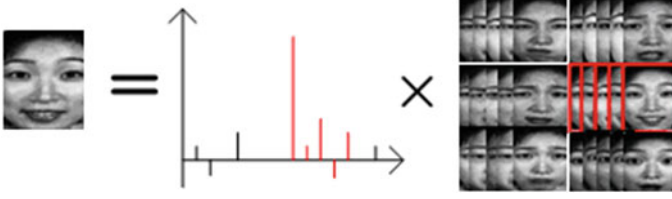


Fig. 2.1 The decomposition schematic of facial expression

As showed in Fig. 2.1, an expression can be decomposed as the linear combination of the expressions in dictionary. The coefficient corresponds to the weight of each facial expression images in dictionary.

The mathematical model can be represented as:

$$y = DX. \quad (2.1)$$

X indicates the sparse coefficients. According to compressive sensing theory [17], the problem can be transformed to solve the minimum l_1 -norm problem as.

$$\min \|X\|_1 \quad s.t. y = AX \in \mathbb{R}^m. \quad (2.2)$$

Various algorithm including L1-magic, OMP algorithm [18] etc. can be used to solve (2). Define $\delta_i(X)$ as the coefficients of the i th class. The minimum error can be as the criteria to judge the belongings of this expression.

$$\min \xi = \|y - D\delta_i(X)\|_2 \quad (2.3)$$

2.2.2 Discriminative K-SVD Algorithm

Discriminative K-SVD [7, 12, 13] is the extension of K-SVD combining the representation power of K-SVD and discriminate ability of linear classifier.

K-SVD algorithm [11] tackles the drawbacks of sparse representation through learning a small-scale dictionary from the given training samples and preserving the representation power of the original dictionary. The algorithm can be achieved by the followed problem.

$$\min_{D, X} \|Y - DX\|_2^2 \quad s.t. \|X\|_0 \leq T \quad (2.4)$$

where Y denote the training samples and T is a fixed sparsity factor. The algorithm works well in image denosing and reconstruction. But it is restricted in image classification without considering discrimination ability.

A linear predictive classifier will introduced to make the dictionary optimal for classification. The linear classifier $Q(X, W, a) = W^T X + a$ can be replaced by the followed optimal problem:

$$[W, a] = \arg \min_{W, a} \|Q - W^T X - a\|^2 + r\|W\|^2 \quad (2.5)$$

Set $a = 0$. An objective function for learning a dictionary with representation power and classification ability can be defined as the following optimal problem:

$$\min_{D, W, X} \|Y - DX\|_2 + \beta \|Q - W^T X\|_2 + r\|W\|_2 \quad s.t. \|X\|_0 \leq T \quad (2.6)$$

In order to learn the dictionary D and W simultaneously, drop the regularization penalty term $r\|W\|_2$ and convert the problem into the following equation:

$$\min_{D, W, X} \left\| \begin{pmatrix} Y \\ \sqrt{\beta} Q \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\beta} W \end{pmatrix} X \right\|_2 \quad s.t. \|X\|_0 \leq T \quad (2.7)$$

where Q is label information. Set $\beta = 1$, (2.7) can be converted to:

$$\min \|\tilde{Y} - \sum \tilde{d}_i \tilde{x}_i\| = \min_{\tilde{d}_k, \tilde{x}_k} \left\| \left(\tilde{Y} - \sum_{k \neq i} \tilde{d}_i \tilde{x}_i \right) - \tilde{d}_k \tilde{x}_k \right\|_F = \min_{\tilde{d}_k, \tilde{x}_k} \|\tilde{E}_k - \tilde{d}_k \tilde{x}_k\|_F \quad (2.8)$$

where $\tilde{Y} = \begin{pmatrix} Y \\ Q \end{pmatrix}$, $\tilde{D} = \begin{pmatrix} D \\ W \end{pmatrix}$, \tilde{d}_k indicate the k th atom of the dictionary \tilde{D} and \tilde{x}_k is the corresponded coefficient.

At testing phrase, the label of the test image can be obtained through the product between linear classifier W and the sparse coefficient α .

$$label = W * \alpha \quad (2.9)$$

The maximum of *label* can be viewed as the class of the test image.

2.3 Random Projection

There exist many dimensionality reduction algorithms [19–22] which project the data into a reduced subspace to reinforce the discriminate capability. RP (random projection) [14, 15] has been applied widely in dimension reduction. The principle of RP is very simple and easy to implement. The central idea is aroused by Johnson-Linderstrauss lemma [16]: Given an image matrix Γ which contains N points in d -dimensional vector space, the matrix can be projected to a lower-dimensional space while the distance between two points is preserved.

$$\tilde{\Gamma}_{k \times N} = R_{k \times d} \cdot \Gamma_{d \times N} \quad (2.10)$$

The transformed matrix R has many formats. In this paper, we select the sparse projection matrix proposed by Achlioptas [23].

$$R(r_{ij} = z) = \begin{cases} 1/6 & z = \sqrt{3} \\ 2/3 & z = 0 \\ 1/6 & z = -\sqrt{3} \end{cases} \quad (2.11)$$

This distribution reduces the computational time to calculate $R \cdot \Gamma$. Random vectors are sufficiently approximate to orthogonal [24], so we can use the sparse random matrix directly.

The proposed efficient D-KSVD based on random projection for facial expression recognition can be described in table:

The proposed efficient D-KSVD algorithm for FER	
Input: The training facial images and test facial image	
Output: The label of the test facial images	
Step 1	Extract the feature of facial expression images
Step 2	Reduce the dimension of the feature using random projection
Step 3	Learn the dictionary D and classifier W adopting D-KSVD algorithm
Step 4	Find the sparse coefficients α of the test sample y exploit OMP algorithm
Step 5	The label of test sample can be obtained finally.
Step 6	The coefficient X of step 3 and α of step 5 as the training sample and testing samples separately would be sent to SVM for classifying

2.4 Experimental Results and Analysis

A series of experiments of discriminative K-SVD with RP are performed on the JAFFE database [25]. JAFFE database contains 213 expression images in total including seven classes of expressions (Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral) of ten Japanese women, which each expression has two to four images.

We use the cropped and normalized face expression images of 120*96 pixels. The images are split into two groups which one contains 70 images as test sample, and the others as training sample.

2.4.1 Experiments Analysis of Efficient D-KSVD Algorithm

The role of feature [26] is not neglected and various features have been applied in facial expression recognition. Three features including gray-scale feature, LBP feature [27] and Gabor feature [28] have been selected. The original dimensions are 11520, 512, 11520*4 for gray feature, LBP and Gabor feature respectively.

Then the experiments based on discriminative K-SVD perform on the features respectively in different reduced dimensions using random projection. Each image can be separately projected into eighty-dimensions, sixty-dimensions, fifty-dimensions, thirty-dimensions, twenty-dimensions, ten-dimensions and five-dimensions. At last, the contrast of the recognition results between the original data and dimensionality reduction data is given as shown in Fig. 2.2a–c. While the comparison of training time is presented in Fig. 2.2d.

From Fig 2.2a–c, it can be seen that the recognition rates remain comparable until the dimension reduced to ten dimensional. On the other hand, in (d), the training time is decreased significantly after dimensional reduction. In addition, the performance is not only determined by the classifier, but also the extracted feature. The recognition results of D-KSVD algorithm with Gabor feature performs best in the three features.

As showed in (c), On the whole, fixing the number of atoms, the results keep steady until the dimension drops to a lower number such as five dimensions. On the other hand, fixing the dimensions, the results show a rising trend with the number of atoms increasing, but it will remain unchanged basically when the atoms reaches a certain number.

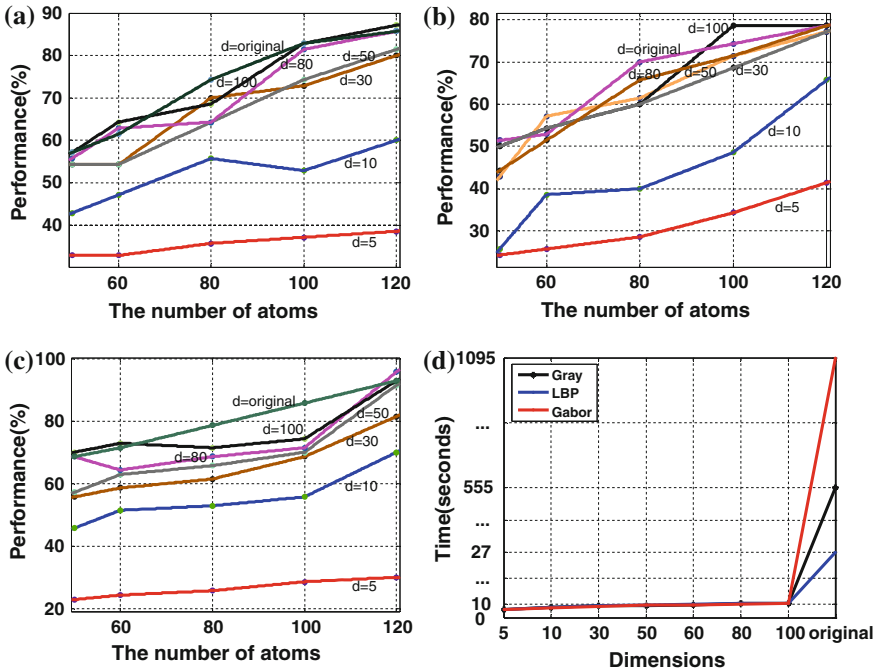
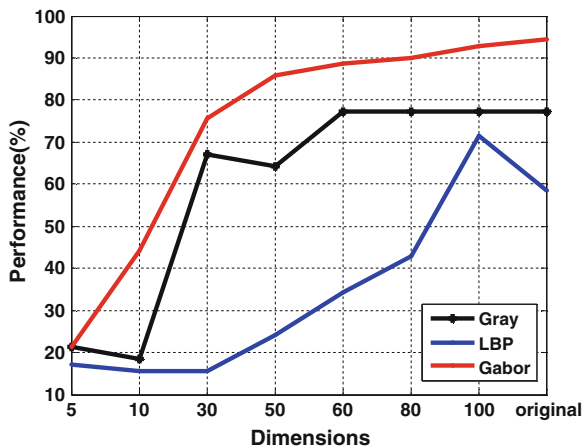


Fig. 2.2 Contrasts of performance and training time between original data and the data after dimensional reduction **a** Gray feature, **b** LBP feature, **c** Gabor feature, **d** Contrasts of training time

Fig. 2.3 The performance of three methods using SVM in different dimensions



2.4.2 Verification Experiments Using SVM

The training coefficient X and the testing coefficient α as the training feature and the testing feature separately would be classified using SVM algorithm. The classified accuracy would be presented in Fig. 2.3.

As shown from the table, the performance using SVM is in line with that using D-KSVD algorithm. The overall trend is declined with the reduction of dimensions. However, the accuracy after dimension reduction still remains unchanged basically if the reasonable and appropriate dimension is selected.

2.5 Conclusion and Future Work

An efficient discriminative K-SVD algorithm for dictionary learning is proposed for facial expression recognition. Particularly, dimensionality reduction uses random projection acted on a series of features (gray, LBP, Gabor) which are extracted from the facial expression images. Then the features after dimensionality reduction are used to train the small-size dictionary and classification simultaneously. Finally, the dictionary and classification are implemented for facial expression recognition system. Experimental results demonstrate that the training time can be greatly reduced with little performance lost after dimension reduction using RP.

Acknowledgments This paper is supported by the National Nature Science Foundation of China (No. 61271407), the Nature Science Foundation of Shandong Province (No. ZR2011FQ016) and the Fundamental Research Funds for the Central Universities (No. 13CX02096A).

References

1. Song M, Tao D, Huang X, Chen C, Bu J (2012) Three-dimensional face reconstruction from a single image by a coupled RBF network. *IEEE Trans Image Process* 21(5):2887–2897
2. Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intel* 31:210–227
3. Cotter SF (2010) Sparse representation for accurate classification of corrupted and occluded facial expressions. *ICASSP* 838–841
4. Huang M, Wang Z, Ying Z (2010) A new method for facial expression recognition based on sparse representation plus LBP. *CISP* 4:1750–1754
5. Monaci G, Vanderghelynst P, Sommer PT (2009) Learning bimodal structure in audio-visual data. *IEEE Trans Neural Netw* 20(12):1898–1910
6. Tian X, Tao D, Rui Y (2012) Sparse transfer learning for interactive video search reranking. *ACM TOMCCAP* 7665:269–276
7. Mairal J, Bach F, Ponce J et al. (2008) Discriminative learned dictionaries for local image analysis. *CVPR* 14(12):2091–2106
8. Protter M, Elad M (2009) Image sequence denoising via sparse and redundant representation. *IEEE Trans Image Process* 18(1):27–35
9. Zepeda J, Guillemot C, Kijak E (2011) Image compression using the iteration-tuned and aligned dictionary. *ICASSP* 1:739–796
10. Liu W, Song C, Wang Y (2012) Facial expression recognition based on discriminative dictionary learning. *ICPR* 1:1839–1842
11. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311–4322
12. Pham D, Venkatesh S (2008) Joint learning and dictionary construction for pattern recognition. *CVPR* 1–8
13. Zhang Q, Li B (2010) Discriminative k-svd for dictionary learning in face recognition. *CVPR* 2691–2698
14. Majumdar A, Ward RK (2010) Robust classifier for data reduced via random projections. *IEEE Trans Syst Man, Cybern, Part B: Cybernet* 40:1359–1371
15. Sulić V, Perš J, Kristan M, Kovačič S (2010) Dimensionality reduction for distributed vision system using random projection. *ICPR* 31:380–383
16. Johnson WB, Lindenstrauss J (1984) Extensions of lipshitz mapping into hilbert space. In: *Contemporary mathematics, conference in modern analysis and probability*, vol 26, pp 189–206
17. Donoho DL (2006) Compressed sensing. *IEEE Trans Inf Theory* 52(4):1289–1306
18. Cai TT, Wang L (2011) Orthogonal matching pursuit sparse signal recovery with noise. *IEEE Trans Inf Theory* 57:4680–4688
19. Yu H, Bannamoun M (2006) 1D-PCA, 2D-PCA to n D-PCA. *ICPR* 4:181–184
20. Guan N, Tao D, Luo Z, Yuan B (2012) Online nonnegative matrix factorization with robust stochastic approximation. *IEEE Trans Neural Netw Learn Syst* 23(7):1087–1099
21. Guan N, Tao D, Luo Z, Yuan B (2012) NeNMF: An optimal gradient method for nonnegative matrix factorization. *IEEE Trans Signal Process* 60(6):2882–2898
22. Guan N, Tao D, Luo Z, Yuan B (2011) Non-negative patch alignment framework. *IEEE Trans Neural Netw* 22(8):1218–1230
23. Achlioptas D (2003) Database-friendly random projections. *J Comput Syst Sci* 66(4):671–687
24. Lin J, Gunopulos D (2003) Dimensionality reduction by random projection and latent semantic indexing. *SDM*
25. Lyons M, Akamatsu S (1998) Coding facial expressions with gabor wavelets. *AFGR* 200–205
26. Song M, Tao D, Liu Z, Li X, Zhou M (2010) Image ratio features for facial expression recognition application. *IEEE Trans Syst, Man, Cybernet, Part B* 40(3):779–788

27. Tan N, Huang L, Liu C (2008) Face recognition based on LBP and orthogonal rank-one tensor projections. ICPR
28. Tao D, Li X, Wu X, Stephen, Maybank J (2007) General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans Pattern Anal Mach Intell* 29(10):1700–1715

Proceedings of 2013 Chinese Intelligent Automation
Conference

Intelligent Information Processing

Sun, Z.; Deng, Z. (Eds.)

2013, XVII, 856 p. 401 illus., Hardcover

ISBN: 978-3-642-38465-3