Face Recognition based on Discriminative Group Structured Dictionary

Yuhua Li

College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

lyhresearch@njau.edu.cn

Abstract. Given a group structured dictionary, sparse representation vector is more likely to take on a particular structure. Intuitively, better results can be achieved if the structural characteristics are reasonably utilized. Inspired by this, we propose to construct a new discriminative group structured dictionary and impose a group sparse structure on the desired sparse representation vector. Thus, each testing sample can be decomposed as the product of the discriminative group structured dictionary and a group sparse representation vector. The identity of testing sample can be found by evaluating which class could induce the minimum reconstruction residual. Numerous experimental results on benchmark face databases demonstrate the effectiveness of the proposed method.

Keywords: Face recognition, discriminative group structured dictionary, group sparsity, linear discriminant analysis, low rank matrix recovery.

1. Introduction

In last two decades, numerous algorithms have been developed for face recognition since it plays a more and more important role in security related applications [1]. Face recognition is very challenging due to the high dimensionality, bad illumination, noisy background, disguises, exaggerated expressions, etc. To overcome these limitations, manifold learning has been successfully utilized, which indicates high dimensional face images usually lie on a lower dimensional manifold, implying that face image can be sparsely represented by representative samples on manifold [2]. As a typical technique of manifold learning, sparse representation has been successfully used in face recognition. Sparse representation based face recognition algorithms usually performs good because sparse representation model fits human perception that human neurons are selective for a mount of stimuli in both low-level and mid-level vision, and the response of neurons to an input image is highly sparse. Typically, Wright et al. [3] proposed a sparse representation based classification algorithm (SRC), which outperforms many existing algorithms and runs fast. However, there are still some defects of SRC. For example, dictionary used in SRC is constructed by all training samples which are deposited as columns class by class, naturally leading to a group structure. And studies in [4] indicate the structures of dictionary and sparse representation vector are of great importance. Nevertheless, the structural property of dictionary is completely ignored, resulting in a waste of useful prior information. Besides, SRC cannot be expected to perform well if training images are not taken under well-controlled conditions or if there are not enough training images per class.

SRC has largely boosted the research of sparsity based face recognition. Many modified SRC methods achieve better results than SRC. For instance, Yang *et al.* [2] propose a gabor occlusion dictionary based SRC (GSRC) method, which greatly improves recognition accuracy nevertheless runs slower than SRC. Assuming training samples of the same individual are linearly correlated, some low rank matrix recovery based methods [5] are proposed. However, most of the above methods require training images are carefully controlled which can't be met in practical situations. Moreover, we discover that most of the sparse representation based face recognition methods even accidentally bring about a structured dictionary. But they regretfully didn't pay much attention to the structural property of dictionary, which may cause a waste of prior information. Based on the observation that nonzero elements of some sparse data are not random but often tend to be clustered, Huang et al. [4] proposed a dynamic group sparsity scheme by utilizing priors of sparsity and clustering.

In this paper, we propose to construct a discriminative group structured dictionary and apply it to frontal face recognition problem. Firstly, for each class, pure representative training samples are estimated from the corresponding raw training samples which may be contaminated by exaggerated

expressions, disguises and bad illumination. To some extent, pure representative training samples can get rid of sparse errors and possibly can better model samples from the same class. Subsequently, dimensionality reduction is carried out by PCA. An initial group structured dictionary can be constructed by concatenating pure representative training samples class by class. To make dictionary more discriminative, linear discriminant analysis is then performed on all atoms. Thus, each atom of the initial group structured dictionary is linearly projected to an atom of the discriminative group structured dictionary, where one sub-dictionary still corresponds to one class. Finally, we impose a group sparse structure on the desired sparse representation vector. Thus, each testing sample is expressed as a group sparse linear combination of atoms from the proposed discriminative group structured dictionary. The obtained group sparse representation vector is used for classification. Experimental results on benchmark face databases verified the effectiveness of the proposed method.

2. Proposed Method

Standard sparse representation mainly concentrates on sparsity while other important and useful prior information are almost ignored, such as the structural property of dictionary and the distribution of nonzero coefficients. Some theories about sparsity are aware of these problems and achieve better results by reasonably making use of these less considered but important prior information [4, 7]. It is observed that the sparse representation vector is more likely to be group sparse structured if the dictionary is group structured, implying that the nonzero elements of the sparse representation vector are not random but often tend to be clustered into a small number of groups. The sparse representation vector can be better recovered using less measurements and computations if these structural properties are reasonably utilized.

In this paper, we focus on the sparsity as well as the structural properties. Firstly, we construct a discriminative group structured dictionary which is the concatenation of several sub-dictionaries, one sub-dictionary corresponding to one class. In practical scenarios, raw training face images are often not collected under a well-controlled condition. Due to variations of illumination and expression, raw training face images couldn't correctly model face images from the same class. To alleviate this problem, for each class, we construct its corresponding initial sub-dictionary by learning pure representative training samples which are more linearly correlated and can better model face images from the same individual [8]. By concatenating initial sub-dictionaries of all classes, we obtain an initial group structured dictionary. Since the discriminative capability of dictionary is essential for sparsity based classification, we enhance the discriminative power of the initial group structured dictionary by performing linear discriminant analysis on all atoms. Hence, atoms of the aforementioned dictionary are mapped to new atoms and the initial group structured dictionary is mapped to the discriminative group structured dictionary. As the projection is class specific, group structure can be still preserved and the discriminative power of dictionary is enhanced. Secondly, based on the observation that the sparse representation vector is more likely to be group sparse structured if the dictionary is group structured, we impose a group sparse structure on the desired sparse representation vector. Each testing sample is therefore decomposed into a product of the discriminative group structured dictionary and a group sparse representation vector. The obtained group sparse representation vector is used for classification. For better invariance to illumination and expression, histogram of oriented gradients (hog) feature is extracted to substitute raw image.

2.1 Discriminative Group Structured Dictionary Preparation

Face images are often of contaminations due to illumination, expression, shadowing, etc. As hog feature is performed on small cells of image, it can tolerate mild variations of expression. Moreover, since local response is contrast normalized before using, it has a certain degree of invariance to illumination and shadowing. Therefore, we extract hog feature to substitute raw face image to preliminarily relieve the influences of illumination, shadowing and expression.

In many situations, face images are not taken under well controlled settings. Excessive variations of illumination, disguises and occlusions may exist, which can't be well handled by extracting the

hog feature. We make efforts to further diminish the influences of these contaminations. For each class, we build a raw training matrix by putting the hog features of all training face images from this class as its columns. Here, we refer to each column of this raw training matrix as a raw training sample. Because of the contaminations, new samples form this same class can't be well modeled by the raw training samples. This impels us to learn pure representative training samples from the raw training samples to better model new samples from this class. A useful assumption is that pure representative training samples of the same person are linearly correlated [2, 5], implying the matrix constructed by the pure representative training samples is of a low rank. We relieve the influences of these contaminations by learning linearly correlated pure representative training samples via low rank matrix recovery. As in formula (1), the raw training matrix of each class is expected to be decomposed into a set of pure representative training samples plus corresponding sparse errors if the rank of H_i is not too large and E_i is sparse enough [5]:

$$\min_{\mathbf{H}_i, \mathbf{E}_i} \| \mathbf{H}_i \|_* + \lambda \| \mathbf{E}_i \|_1 \quad s.t. \quad Y_i = H_i + E_i \tag{1}$$

where $Y_i = [y_1^i, ..., y_{n_i}^i] \in \mathbf{R}^{l \times n_i}$ denotes the raw training matrix of the ith class with each column $y_j^i \in \mathbf{R}^{l \times 1}$ $\left(j = 1, ..., n_i \right)$ a raw training sample. n_i is the number of samples of the ith class. $H_i = [h_1^i, ..., h_{n_i}^i] \in \mathbf{R}^{l \times n_i}$ is a low rank matrix with each column $h_j^i \in \mathbf{R}^{l \times 1}$ $\left(j = 1, ..., n_i \right)$ a pure representative training sample. $E_i \in \mathbf{R}^{l \times n_i}$ is the associated sparse error matrix. $\| H_i \|_*$ represents the nuclear norm, estimating the rank of H_i. $\| E_i \|_1$ represents the l_1-norm, estimating the sparsity of E_i. Formula (1) can be efficiently solved by the technique of inexact augmented lagrange multipliers [10]. By concatenating the low rank matrices class by class, an initial group structured dictionary $H = [H_1, ..., H_k] = [h_1, ..., h_n] \in \mathbf{R}^{l \times n}$ can be obtained, where h_i is the ith atom, $n = n_1 + ... + n_k$ is the number of atoms, and H_i is the initial sub-dictionary for the ith class. H has some advantages over the dictionary constructed by raw training samples. Firstly, initial sub-dictionary is more compact and can better model training samples from the same person. Secondly, it reduces the influence of sparse errors caused by illumination and occlusions. We reduce dimensionality using PCA. Thus H, H_i and h_i turn to $Z = P_{pca}^T H$, $Z_i = P_{pca}^T H_i$ and $z_i = P_{pca}^T h_i$ respectively using PCA projection matrix $P_{pca} \in \mathbf{R}^{l \times d}$.

Group structured dictionary Z can correct sparse errors, enhance the linear correlation of atoms within each sub-dictionary, and preserve the group structure. As the discriminative capability of dictionary is important for sparse representation based classification, we boost the discriminative power of dictionary Z by projecting the atoms of Z to new atoms via linear discriminant analysis. A linear mapping matrix $P \in \mathbf{R}^{d \times d}$ is calculated to map each atom $z_i (i = 1, ..., n)$ to a transformed atom $x_i (i = 1, ..., n)$ by $x_i = P^T z_i$ $(i = 1, ..., n)$. Thus each atom of dictionary Z is linearly transformed to a new atom. $x_i \in \mathbf{R}^{d \times 1}$ is the ith atom of the discriminative group structured dictionary $X = [X_1, ..., X_k] = [x_1, ..., x_n] \in \mathbf{R}^{d \times n}$ which is expected to be discriminative while still maintains the group structure. $X_i = P^T Z_i \in \mathbf{R}^{d \times n_i}$ is a sub-dictionary corresponding to class i. The group structure can still be preserved, one sub-dictionary still corresponding to one class. Atoms within each sub-dictionary are more compactly clustered and atoms between different sub-dictionaries are much farther apart from each other. The projection is class specific, making the dictionary more discriminative which is useful for classification.

2.2 Group Sparse Representation Based Classification

Utilizing the discriminative group structured dictionary, each testing sample is supposed to be encoded as a group sparse representation vector. For a testing image from i^{th} class, we firstly extract its hog feature, then project it using mapping matrices P_{pca} and P, and finally obtain a testing sample

$x \in \mathbf{R}^{d \times 1}$. In this paper, we casted x as a group sparse linear combination of atoms from the discriminative group structured dictionary X :

$\boldsymbol{\beta}$ *is a group sparse column vector*   *s.t.*

$$\| x - X\boldsymbol{\beta} \|_2^2 \leq \sigma^2 \tag{2}$$

where σ is a noise term with bounded energy. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1;...;\boldsymbol{\beta}_k] = [\beta(1);...;\beta(n)] \in \mathbf{R}^{n \times 1}$ is a group sparse representation vector with $\boldsymbol{\beta}_j \in R^{n_j \times 1}$ a column vector consisted of all the coefficients corresponding to the jth class and $\beta(j)$ the jth element of $\boldsymbol{\beta}$. k is the number of classes. n_j is the number of training samples of the jth class. n is the number of training samples of all the k classes. $\boldsymbol{\beta}$ is supposed to be group sparse structured, which means that the nonzero elements are clustered in a small number of groups. Ideally, the number of nonzero group is one, meaning that only the elements in $\boldsymbol{\beta}_i$ are nonzeros and the rest are zeros. A comparison of sparse vector and group sparse vector is shown in fig.1. To solve formula (2), we adopt the strategy of DGS recovery [4]. First, we calculate a column vector $v = [v(1),...,v(n)]^T$ by formula (3):

$$v(i) = \sum_{j=-\tau, j \neq 0}^{j=\tau} \beta^2(i) + \gamma\beta^2(i+j) \tag{3}$$

where $v(i)$ is the ith element of v. In this paper, γ is a positive scalar which is set as 0.5, 2τ is the neighboring elements of $\beta(i)$ which is set as 2. Second, we find out the s largest elements of v and the corresponding indices constitute a support column vector $\Gamma \in \mathbf{R}^{s \times 1}$. Third, $\boldsymbol{\beta}$ is updated via $\boldsymbol{\beta} = X_\Gamma^\dagger x$, where X_Γ represents a matrix only preserving the columns of X corresponding to Γ and setting the rest to zeros. $X_\Gamma^\dagger$ is the pseudo-inverse matrix of X_Γ. The above steps are repeated until representation error is no longer decreasing. That making use of group sparse representation instead of standard sparse representation has some advantages. Firstly, the constraint of group sparsity forces the sparse representation vector close to the actual representation. Secondly, utilizing both priors of group clustering and sparsity reduces the degree of freedom of $\boldsymbol{\beta}$, leading to an acceleration of the coding stage and an enhancement of the robustness to noises. Thirdly, group sparse representation can reduce the minimal number of necessary measurements [4]. Here, the number of measurements means the dimensionality d of the testing sample.

Group sparse representation vector $\boldsymbol{\beta}$ is used for final classification. The reconstruction residual corresponding to the jth class is defined as $r_j = \| x - X\delta_j(\boldsymbol{\beta}) \|_2^2$, where $\delta_j(\boldsymbol{\beta})$ is obtained by preserving the elements corresponding to the jth class and setting the rest to zeros. Finally, the identity of the testing image is deemed as $j^* = \underset{j}{\arg\min}(r_j)$. In the situations of group sparse occlusions, identity matrix $I \in \mathbf{R}^{d \times d}$ is used as an occlusion dictionary. Testing sample $x' = x + e$ can be expressed as:

W *is a group sparse column vector*   *s.t.*

$$\| x' - BW \|_2^2 \leq \sigma^2 \tag{4}$$

where e is a group sparse occlusion, $B = [X,I] \in \mathbf{R}^{d \times (d+n)}$ is an expanded dictionary, and $W = [\boldsymbol{\beta}^T, e^T]^T \in \mathbf{R}^{(d+n) \times 1}$ is a group sparse representation vector. $r_j = \| x' - X\delta_j(\boldsymbol{\beta}) - e \|_2^2$ is the relevant reconstruction residual of the jth class.
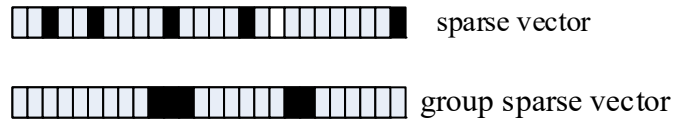
Fig. 1 Comparison diagram of sparse vector and group sparse vector, where each small white rectangle represents zero element and black rectangle represents nonzero element.

3. Experimental Results

To demonstrate the performance of the proposed method, it is compared with the methods in [2] and [3] on public face databases: AR database [11] and Extended Yale B database [12]. For each database, all selected face images are resized and aligned beforehand using the locations of eyes which are provided by databases. Parameters are empirically set as follow unless otherwise stated in this paper. Cell size is set as 8×8 pixels. Block size is set as 2×2 cells with 50% overlapping. b, γ and τ are set as 9, 0.5 and 1 respectively. s is chosen between 1 and $n/3$. d ranges from 100 to 600 with an interval 100. Parameters in GSRC and SRC are tuned for best results. PCA is utilized for dimensionality reduction for all methods. A few results of GSRC and SRC are from [2]. Numerous experimental results show the proposed method achieves state-of-art results.

(1) Experiments on AR database: We randomly select 100 individuals (50 males and 50 females) from AR database. Each individual consists of 26 face images which are taken in two separate sessions. For each individual, 7 unoccluded face images from session one are used for training and 7 unoccluded face images from session two are used for testing. There are a total of 700 training images and testing images respectively. Each face images is resized to 160×112 pixels. These selected face images are only with expression and illumination variations. Top recognition rates versus sample dimension d are illustrated in table 1. It can be seen that our method reaches much higher recognition rates than GSRC and SRC at all dimensions. The reasons why the proposed method performs well is threefold: First, the proposed discriminative group structured dictionary has powerful discriminative capability. Second, atoms in each sub-dictionary can better model testing samples from the same class. Third, the group sparsity constraint compels the sparse representation vector closer to its ideal pattern (ideal pattern: nonzero elements gather in one group). Besides, table1illustrates that even at a low dimension, our method can still achieve high recognition rates. For example, when d =100, top recognition rates of our method, GSRC, and SRC are 97.29%, 93.71% and 85.86% respectively. This is because the group sparsity constraint makes it possible that we recover a better group sparse representation vector using much less measurements. An average runtime of the testing stage for one testing sample is shown in table 2. Sample dimension d , cell size, s and λ are set as 500, 8×8, 40 and 0.08 respectively. Our method runs faster than GSRC and SRC because group sparsity constraint reduces the degrees of freedom of sparse representation vector which accelerates the sparse coding stage.

Table 1. Top recognition rates (%) versus dimensions d on AR database.

Dimension d	100	200	300	400	500	600
Our method	97.29	99.00	99.14	99.14	99.43	99.43
GSRC	93.71	95.57	96.86	96.71	96.04	96.57
SRC	85.86	90.57	90.43	89.57	90.86	90.00

Table 2. Average runtime of the testing stage for one sample on AR database.

Methods	Our method	GSRC	SRC
Runtime(sec)	0.25	14.2	1.53
Recognition rate (%)	99.29	96.04	90.86

(2) Experiments on Extended Yale B database: 31 individuals are randomly selected from Extended Yale B database. Each individual consists of 64 face images with illumination variations.

Images are resized to 192×160 pixels. For each individual, half of the face images are randomly selected for training and the rest half are used for testing, resulting in 992 training images and 992 testing images. When sample dimension d ranges from 100 to 600, top recognition rates are presented in table 3. We can see that our method achieves higher recognition rates than GSRC and SRC at all dimensions. The maximal recognition rates of our method, GSRC, and SRC are 99.5%, 96.27%, and 92.84% respectively. From table 1 and table 3, we can demonstrate that our method is robust to illumination variations to some extent. Besides, we carried out an experiment to evaluate the influence of enrollment size (the number of training samples per class). For each individual, half images are randomly selected for testing, and 10, 20 or 32 images are randomly selected from the rest half for training, illustrating that the enrollment size is 10, 20, and 32 respectively. In total, there are 992 testing images and 310, 620, or 992 training images respectively. Recognition rates versus sample dimension d and enrollment size are listed in table 4. In table 4, for each method, recognition rates from top to bottom separately correspond to the case that enrollment size is 10, 20, and 32 respectively. Here, s is set as 25 and λ is set as 0.07. Recognition rates decrease as the enrollment size decreases. We can see that our method achieves higher recognition rates than GSRC and SRC at all enrollment size, and the superiority is more significant when enrollment size is small.

Table 3. Top recognition rates (%) versus dimensions d on Extended Yale B database.

Dimension d	100	200	300	400	500	600
Our method	99.09	99.50	99.40	99.40	99.40	99.40
GSRC	80.44	91.23	92.14	95.06	95.77	95.97
SRC	86.09	90.42	92.14	90.04	92.44	92.34

Table 4. Recognition rates (%) versus different enrollment size.

Dimension d	100	200	300	400	500	600
	97.28	98.79	99.09			
Our method	98.59	99.19	99.09	98.39	98.59	98.59
	99.09	99.50	99.29	99.29	99.40	99.29
	64.52	69.25	70.46			
GSRC	73.59	79.13	81.45	84.17	85.18	85.88
	80.44	91.23	92.14	95.06	95.77	95.97
	72.58	77.32	77.32			
SRC	81.05	85.89	87.40	88.31	88.21	88.51
	86.09	90.42	92.14	90.04	92.44	92.34

(3) Recognition against disguises:100 individuals (50 males and 50 females are randomly selected from AR database. Pick out the 8 face images of expression variations per class for training. Thus there are 800 training images. Two separate testing sets are prepared. Testing set 1: one face image occluded by sunglass (or scarve) with natural expression and illumination is selected from session one, in total, 100 testing images. Testing set 2: separately select one image occluded by sunglass (or scarve) with natural expression and illumination from session one and session two respectively, in total, 200 testing images. Images are cropped into 160×112 pixels. Top recognition rates of our method, GSRC and SRC are illustrated in table 5, where, "GSRC-P" and "SRC-P" is the partitioned case of GSRC and SRC respectively. From table 5, it can be observed that our method achieves a top recognition rate of 100% in these four cases, higher than the other two methods. Here, both occlusions of sunglass and scarve are group sparse. Our method is more robust to deal with group sparse disguises because that group sparsity constraint can better model these disguises than sparsity constraint. The proposed method can well manage group sparse occlusions undoubtedly. However, how to handle general occlusions should be studies in further research.

Table 5. Top recognition rates (%) on AR database with disguises ("Rate": recognition rate, "-sg": sunglass, "-sc": scarve, "-100": 100 testing images, and "-200": 200 testing images).

Method	Our method	GSRC	SRC	GSRC-P	SRC-P
Rate-sg-100	100	96	46	99	91
Rate-sc-100	100	98	57	100	99
Rate-sg-200	100	96	45.5	99	89
Rate-sc-200	100	97.5	59.5	99.5	96.5

4. Conclusion

In this paper, we propose to construct a discriminative group structured dictionary for frontal face recognition. Firstly, for each class, we learn pure representative training samples from contaminated raw training samples. By doing this, the linear correlation among samples within each class can be strengthened, samples from the same individual can be better modeled, and sparse noise jamming can be alleviated to some extent. Secondly, the introduction of linear discriminant analysis increases the within-class scatter and decreases the between-class scatter, which promotes the discriminative capability of dictionary. Thirdly, we reasonably impose both priors of group clustering and sparse on the desired sparse representation vector. Thus, testing sample is decomposed as a product of the discriminative group structured dictionary and a group sparse coefficient vector, which reduces the degree of freedom of sparse representation vector, accelerates the coding process. Experimental results verify the effectiveness of the proposed method. It not only achieves higher recognition rates but also shows a certain degree of robustness to variations of expressions, illumination, and group sparse disguises.

References

[1]. He, X. and P. Niyogi, Face recognition using Laplacian faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2005) 328–340.

[2]. M. Yang and L. Zhang, Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, in: Proceedings of the 11th European Conference on Computer Vision, 2010, pp.448-461.

[3]. J. Wright, A. Ganesh, A. Yang, and Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31(2009) 210-227.

[4]. J. Z. Huang, X. L. Huang, and D. Metaxas, Learning with dynamic group sparsity, in: Proceedings of the 12th IEEE International Conference on Computer Vision, 2009, pp.64-71.

[5]. C. Chen, C. Wei, and Y. Wang, Low-rank matrix recovery with structural incoherence for robust face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp.2618-2625.

[6]. D. Donoho and Y. Tsaig. Fast solution of l1-norm minimization problems when the solution may be sparse, IEEE Transactions on Information Theory 54 (2008)4789–4812.

[7]. J. Huang, T. Zhang, d D. Metaxas, Learning with Structured Sparsity, in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp.417-424.

[8]. E. Candes, X. Li, Y. Ma, and J. Wright, Robust principal component analysis, JACM 58(2011).

[9]. W. Zuo, H. Zhang, D. Zhang, and K. Wang, Post-processed LDA for face and palmprint recognition: What is the rationale, Signal Processing 90 2344-2352.

[10]. Z. Lin, M. Chen, and Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, UIUC Tech. Rep. UILU-ENG-09-2214, 2010.

[11]. A. M. Martinez and R. Benavente, The AR Face Database, CVC Technical Report, no.24, 1998.

[12]. K. Lee, J. Ho, and D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (2005) 684-698.