J. Vis. Commun. Image R. 40 (2016) 42–50

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Extended common molecular and discriminative atom dictionary based sparse representation for face recognition ☆

Hu Zheng-ping [a,*], Bai Fan [a], Zhao Shu-huan [a], Wang Meng [a,b], Sun Zhe [a]

[a] School of Information Science and Engineering, Yanshan University, Qinhuangdao, China
[b] School of Physics and Electronic Engineering, Taishan University, Tai'an, China

ARTICLE INFO

Article history:
Received 27 October 2015
Revised 20 March 2016
Accepted 30 May 2016
Available online 31 May 2016

Keywords:
Face recognition
Sparse representation
Extended dictionary
Principal component analysis
Maximum probability

ABSTRACT

The employed dictionary plays an important role in sparse representation classification, however how to build the relationship between dictionary atoms and class labels is still an important open question. Many existing sparse representation classification dictionary models exploit only the discriminative information either in the representation coefficients or in the representation residual, which limits their performance. To address this issue, we introduce a novel dictionary building method which is constructed by two parts: the common molecular dictionary and the discriminative atom dictionary. More specifically, the discriminative atom dictionary builds its relationship to class labels and the extended molecular dictionary can reduce the representation residual for all the classes. Therefore, the new dictionary not only has correspondence to the class labels, but also has the perfect representation ability. Besides, the maximum probability representation is used for the final classification. In conclusion, the sparse coefficient of our method is sparser than the sparse representation-based classification (SRC), and our method can achieve better performance. Experiments on the AR, Extended Yale B and CMU PIE face datasets verify that our algorithm outperforms many recently proposed methods.

1. Introduction

Face recognition (FR) is a very active topic in computer vision and pattern recognition over the past two decades [1–3]. And FR makes an extensive range of applications, including video surveillance, identity authentication, safety, etc. [4]. Although numerous researchers have obtained significant progress in the field of FR, there still remains some unsolved problems [5]. In real applications, the face images are abundant in illuminations, pose variations and disguises [6,7]. Besides, the original dictionary usually only contains a few images of each individual. Thus, how to learn a more powerful representation dictionary and an excellent classification rule by the insufficient training samples is a challenging task for FR.

Among the previous classifiers for FR, the nearest neighbor (NN) is part of the most popular and common classifiers, which classifies the query face image into the class that contains the closest neighbor. However, this classifier is sensitive to outliers. The nearest subspace (NS) is the generalization of NN [8]. Instead of using a single image to perform classification, NS classifier is based on the best linear representation in terms of all the samples in each class. And NS is more robust than NN, because its classification decision is taken by all samples. Wright et al. [9] proposed a sparse representation classifier for FR based on L1-norm, which achieved a balance between these two extreme cases. SRC represents a query image by adaptively selecting a minimum number of samples from both within each class and across multiple classes. SRC shows more effective and robust than NN and NS on some common face recognition issues, such as occlusion and corruption. Encouraged by the SRC framework, various extension methods based on SRC have been proposed [10–13] and have achieved outstanding performance. However, the original dictionary cannot supply enough information, which limits the improvement of the recognition performance. So how to learn a good dictionary by insufficient training samples is a challenging task in the practical application of face recognition.

Yang et al. presented a dictionary learning method based on the Fisher Discrimination criterion [14]. In their method, the representation coefficients based on the learning dictionary had small within-class scatter and large between-class scatter. Besides, representation residual could be used to distinguish different classes. Thus, representation residual and representation coefficients contained discriminative information. To deal with the large occlusion, Ou et al. [15] proposed the occlusion dictionary learning method,

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: hzp@ysu.edu.cn (Z.-p. Hu).

http://dx.doi.org/10.1016/j.jvcir.2016.05.019
1047-3203/

which managed to find a representation that could combine the training samples and occlusion dictionary to learn a better dictionary. Thus, the recognition performance was improved. For the small sample problem (SSS), Wei and Wang presented the robust auxiliary dictionary learning approach [16]. It learned the auxiliary dictionary from external data to observe possible image variant. To better preserve the discriminative information, Wang et al. decomposed the original data matrix by singular value decomposition projection [17]. Wei et al. designed a locality-sensitive dictionary learning method which preserved local data structure and resulted in improving image classification performance [18]. To make the facial recognition more accurate, Mahoor et al. presented a PCA-based dictionary building method for sparse representation [19]. The dictionary was constructed by principal components from the expressive facial images extracted from the original samples. To represent the possible variation between the training and testing images, Deng et al. proposed an auxiliary intra-class variant dictionary [20]. It computed intra-class sample differences from either their own gallery faces or the generic faces that were outside the gallery. Then, the dictionary was consisted of the sample differences. From multi-view learning [21] aspects, Shrivastava et al. presented the multi-kernel learning sparse representation based classification model [22]. By increasing the space information and multi-kernel learning, it obtained the weight matrix and selected the best kernel matrix for classification by the linear combination of the kernel matrixes. Thus, the robustness of the classification was improved.

We can conclude from these references that the form of the dictionary plays a significant role in the performance of SRC-based method. In this paper, we construct an extended common molecular and discriminative atom dictionary to improve the recognition performance of FR. First, the principal components are learned from the whole training samples by PCA and some of the components without marks are selected to construct the common information dictionary. Simultaneously, the original training samples dictionary acts as the atom dictionary. Then, the extended dictionary is obtained by combining the molecular dictionary with the atom dictionary. The molecular dictionary which is shared by all individuals in the extended dictionary can help to find the best linear combination of test sample, which makes the representation more precise. The test samples are coded sparsely in the extended dictionary with discriminative and common information. Finally, we classify the test sample with the maximum probability.

2. Sparse representation based classification

Suppose that there exists K individuals and each sample can be represented as a column vector. Define $\boldsymbol{A_i} = [\boldsymbol{v_{i,1}}, \boldsymbol{v_{i,2}}, \ldots, \boldsymbol{v_{i,n_i}}] \in R^{m \times n_i}$ as a set of n_i training samples from ith individual, where m is the dimension of the training sample. Given sufficient training samples of the ith individual, any test sample $\boldsymbol{y} \in R^m$ from ith individual can be represented as

$$\boldsymbol{y} = a_{i,1}\boldsymbol{v}_{i,1} + \cdots + a_{i,j}\boldsymbol{v}_{i,j} + \cdots + a_{i,n_i}\boldsymbol{v}_{i,n_i} \tag{1}$$

where $a_{i,j}$ $(j = 1, 2, \ldots, n_i)$ is the weighting coefficient.

Since we can't judge the label of test sample $\boldsymbol{y}$, it defines a matrix $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_i, \ldots, \boldsymbol{A}_K] \in R^{m \times n}$, which is consisted of all the training samples, where $n = \sum_{i=1}^{K} n_i$. Then, the test sample $\boldsymbol{y}$ is a linear representation of $\boldsymbol{A}$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in R^m \tag{2}$$

where $\boldsymbol{x} = [0, \ldots, 0, a_{i,1}, \ldots, a_{i,n_i}, 0, \ldots, 0]^T \in R^n$ is the coefficient vector. The non-zero coefficients in $\boldsymbol{x}$ correspond that those associated with the ith individual.In face recognition, the equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ is typically underdetermined, that is $m < n$, so there will be many solutions. Therefore, regularization constrain is the key point for obtaining appropriate solution. To seek the solution of equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, we convert Eq. (2) into the following optimization problem:

$$\hat{\boldsymbol{x}}_0 = \arg\min \|\boldsymbol{x}\|_0 \quad s.t.\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{3}$$

However, Eq. (3) is a NP-hard problem. According to the compressed sensing, the L_0-*norm* problem (3) is equal to the L_1-*norm* problem, only if the solution $\hat{\boldsymbol{x}}_0$ is sparse enough:

$$\hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \quad s.t.\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{4}$$

Since real data may be corrupted by noise, the test sample cannot be represented exactly as the sparse superposition of the training samples. Hence, the existence of error is permitted and the limit of error tolerance is defined as ε. Thus, Eq. (4) can be revised as:

$$\hat{\boldsymbol{x}}_1 = \arg\min \|\boldsymbol{x}\|_1 \quad s.t.\ \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2 \leqslant \varepsilon \tag{5}$$

Finally, the query $\boldsymbol{y}$ is classified into the individual with the minimum residual:

$$identity(\boldsymbol{y}) = \arg\min_i \|\boldsymbol{y} - \boldsymbol{A}_i\delta_i(\hat{\boldsymbol{x}}_1)\|_2 \tag{6}$$

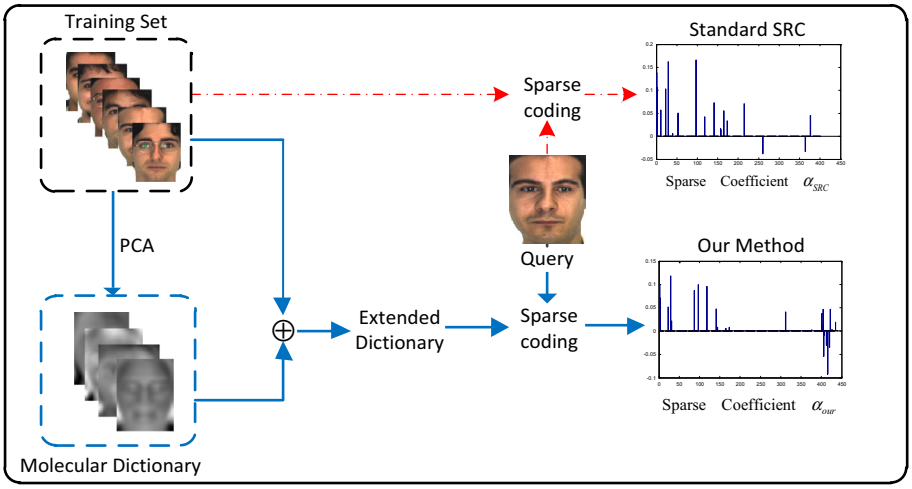

Fig. 1. Comparing the standard SRC and our method. The blue solid line describes our method, and the red dotted line describes standard SRC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)