Information Processing Letters 115 (2015) 677–683

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

A modular weighted sparse representation based on Fisher discriminant and sparse residual for face recognition with occlusion

Shuhuan Zhao, Zheng-ping Hu *

School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

ARTICLE INFO

Article history:
Received 15 August 2014
Received in revised form 9 March 2015
Accepted 13 April 2015
Available online 17 April 2015
Communicated by X. Wu

Keywords:
Design of algorithms
Face recognition
Sparse representation
Fisher discriminant
Modular weight

ABSTRACT

Face recognition with occlusion is one of the main problems countered in face recognition in practical application. The occlusion in the image will decline the performance of global-based methods, so most of existing methods for this problem are block-based. Our method also divides image into modules. Considering that different modules have different discriminative information, we identify a new criterion to compute modular weight. The modular weight can not only depress the effect of low discriminant module but also can detect the occlusion module to some extent. The weighting function is based on the modular Fisher rate and the modular residual. The successful application of sparse representation-based classification (SRC) in image recognition inspires us to use SRC on the weighted dictionary and test image to perform the final identification. Experiments on the AR and extended Yale B database verify the effectiveness and robustness of the method.

1. Instruction

Face recognition has extensive application prospects in national security, military security and public safety etc. So it has become a hot research focus in pattern recognition, image processing, machine vision, neural network and humanities etc. in recent years.

Nearest feature-based classifiers (NFCs) are part of the most popular face recognition methods. In general, NFCs aim to find a representation of the query image, and classify it into the class with the lowest residual. According to the mechanism of query image representation, NFCs include Nearest Neighbor (NN), Nearest Feature Line (NFL) [1], Nearest Feature Plane (NFP) [2], and Nearest Feature Subspace (NFS). Among these methods, NN is the simplest one with no parameters which classifies the query image into its nearest neighbor. The performance of NN can be easily affected by noises, for NN adopts only one sample to represent the query image. NFL classifier proposed by Li et al. forms a line by every two training samples of the same class and classifies the query image into its nearest line. Chen et al. propose the NFP classifier which uses at least three training samples from the same class to form a plane rather than a line to determine the label of the query image. Instead of using a subset of the training samples with the same label to represent the query image like NN, NFL and NFP, NFS represents the query image by all training samples of the same class. In general, the more samples are used for representation, the more stable a method is supposed to be [3]. Hence, NFS is assumed to perform better than the other NFCs. However, NFCs are not robust in real-world face recognition applications because of various occlusions.

In recent years, many new algorithms are proposed to solve this problem and the sparse representation classification (SRC), firstly applied to image processing by Wright [4], is very popular for its robustness and satisfied

* Corresponding author. Tel.: +86 13833560097.
E-mail address: hzp_ysu@163.com (Z.-p. Hu).

http://dx.doi.org/10.1016/j.ipl.2015.04.004
0020-0190/

performance [5–8]. Based on sparse representation, Qiao et al. [9] propose sparsity preserving projections (SPP) for unsupervised dimensionality reduction. It can preserve the sparse reconstructive weights and the application on the face recognition verifies the effectiveness of SPP. However, every sample is considered as an independent point in SRC, while the inner structure of the data, samples from a same class have similar coefficients in sparse representation, is ignored. So the structural sparse representation [10–12] is proposed to integrate the structural information of dictionary into the sparse representation and the identification is processed by seeking the sparsest block representation of the test sample. But the negative representation coefficients have little physical significance when the test sample is represented by the linear combination of training samples. So the nonnegative sparse representation is proposed [13]. For dictionary matrix constructed by images, there are many shared information between images. Therefore, the dictionary is low-rank. The lowest-rank sparse representation [14] is proposed which integrates the low-rank into the sparse representation. To improve the Discriminant, Chih-Fan Chem et al. propose incoherence low-rank matrix decomposition for sparse representation in paper [15] and the recognition performance is improved.

But the performance of these methods will be degraded when the face image is occluded. How to alleviate the effect of occlusion is crucial. Ran He et al. [16] propose correntropy-based sparse representation (CESR) which introduces auxiliary variables by half-quadratic optimization to detect the occlusion pixels. Some other methods divide image into modules. Paper [17] proposes WGSR (Modular Weighted Global Sparse Representation), in which the training and test samples are partitioned into some modules and the sparsity and residual of sparse representation are used to calculate the modular weight. But it needs to compute modular weight for each new test sample. In paper [18], image is partitioned into two modules, up module and down module. And the sparsity is used to estimate the occluded part. Then using the global reconstruction based on the un-occluded part and the residual, the occluded pixels are evaluated. But those two methods are not focused on the selection of discriminative information.

The election of discriminative information is an important step for face recognition with occlusion. The common used method is PCA (Principal Component Analysis) [19], which uses the feature vector corresponding to the biggest feature values of global scatter to construct the projection matrix. This method preserves the principal component, but it didn't consider the characteristic of within and between classes. However, LDA (Linear Discriminant Analysis) [20] computes the projection matrix by maximization the between class scatter and minimization the within class scatter simultaneously to improve the classification performance. One of the classical LDA algorithms is Fisher.

Considering that the identification of the sparse representation is depended on residual, we propose a new method for face recognition which uses modular Fisher rate and sparse residual to compute modular weight. Firstly, training samples are partitioned into some modules and the Fisher rate of each module is computed respectively. Secondly, the global sparse representation of each training sample on the rest of training set is calculated, so the sparse residual of each module can be obtained. Finally, combined the Fisher rate with sparse residual, the final weight of each module is obtained. The final recognition process is performed on the weighted dictionary and query image by sparse representation.

2. Sparse representation

Since our classification rule is based on SRC, for the sake of clarity, we now briefly review this algorithm. Sparse representation for classification (SRC) seeks the sparsest representation of test sample in dictionary. Suppose that there exists C subjects and the jth sample from the ith class can be represented as vector $v_{i,j}$, so all the samples from the ith class construct a matrix $A_i = [v_{i,1}, v_{i,2}, \cdots, v_{i,n_i}] \in R^{m\times n_i}$, where m means the dimension of training sample and n_i means the number of training samples from the ith class. It supposes that samples from the same class construct a linear subspace, so any test sample can be represented as linear combination of the samples from the same class, for example, test sample $y \in R^m$ from class i can be represented as:

$$y = \alpha_{i,1}\cdot v_{i,1} + \alpha_{i,2}\cdot v_{i,2} + \cdots + \alpha_{i,n_i}\cdot v_{i,n_i}$$
$$\text{where } \alpha_{i,j} \in R,\ j = 1, 2, \cdots, n_i \tag{1}$$

Since we don't know which class the test sample belongs to, it identifies a dictionary as $A = [A_1, A_2, \cdots, A_C] = [v_{1,1}, v_{1,2}, \cdots, v_{C,n_c}] \in R^{m\times N}$, where $N = \sum_{i=1}^{C} n_i$ refers to the total number of training samples. The test sample y can be represented by the linear combination of all the training samples as:

$$y = A\cdot x_0 \in R^m \tag{2}$$

where $x_0 = [0, \cdots, 0, \alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,n_i}, 0, \cdots, 0]^T \in R^N$ denotes the vector of coefficients. In x_0 the non-zero atoms correspond to the training sample from the same class with the test sample. In image recognition, Eq. (2) is usually underdetermined, that is $m < N$, so there will be many solutions. Since we know that x is sparse, we can restrain the equation by min-ℓ_0 norm:

$$(\ell_0):\quad \hat{x_0} = \arg\min \|x\|_0 \quad \text{s.t.} \quad A\cdot x = y \tag{3}$$

Because Eq. (3) is a NP-hard problem, according to sparse representation and compressive sensing, we can replace it by ℓ_1 norm only if x_0 is sparse enough, so there is

$$(\ell_1):\quad \hat{x_1} = \arg\min \|x\|_1 \quad \text{s.t.} \quad A\cdot x = y \tag{4}$$

Since there is noise in image, the linear combination of training samples cannot represent test sample accurately, so it permits the existence of error and defines the limit of error ε. Eq. (4) can be converted into the following form:

$$(\ell_1):\quad \hat{x_1} = \arg\min \|x\|_1 \quad \text{s.t.} \quad \|A\cdot x - y\|_2 \le \varepsilon \tag{5}$$

Finally, it computes the residuals of each class and classifies the test sample into the class with lowest residual:

$$identify(y) = \arg\min_i r_i(y) \tag{6}$$