Signal Processing 147 (2018) 101–109

Contents lists available at ScienceDirect

Signal Processing

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

Space–frequency domain based joint dictionary learning and collaborative representation for face recognition

Yali Peng[a,b], Liping Li[a,b,c], Shigang Liu[a,b,c,*], Tao Lei[d]

[a] Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, China
[b] Engineering Laboratory of Teaching Information Technology of Shaanxi Province, Xi'an 710119, China
[c] School of Computer Science, Shaanxi Normal University, Xi'an 710119, China
[d] College of Electronics and Information Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China

ARTICLE INFO

Article history:
Received 31 July 2017
Revised 23 December 2017
Accepted 8 January 2018

Keywords:
Face recognition
Dictionary learning
Fourier transform
Representation based classification

ABSTRACT

In this paper, we propose a novel viewpoint about dictionary learning (DL) and collaborative representation for face recognition. Different from conventional learning methods, we consider both the native spatial domain and the Fourier frequency domain of datasets for dictionary learning. Based on the Fourier spectrum of images, the proposed method provides new insights into two crucial complementations in dictionary learning: *data domain complementation* and *classification algorithm complementation*. On the one hand, we perform the dictionary learning on the original dataset and the Fourier transformed dataset respectively, which makes data complementary in both spatial and frequency domains. On the other hand, we integrate dictionary learning and collaborative representation (CRC) for classification. Specifically, CRC is conducted on frequency-domain dataset to obtain residual scores, and the residual scores are fused with the ones obtained by the previous DL algorithms as the ultimate fusion score to classify the test samples. The proposed method with two aspects of complementation promotes the discriminative ability of dictionary learning and obtains a better classification performance. The experimental results demonstrate the superior performance of our method over the original dictionary learning methods.

1. Introduction

Face recognition is a challenging computer vision task that has obtained a lot of attention in many years [1]. It is well known that conventional approaches include Eigenface [2] and Fisherface [3], and others [4–6]. These methods usually contain two stages: feature extraction and classification. Recently, as a hot topic in machine learning and pattern recognition, dictionary learning (DL) has been attracting increasing attention [7–9]. Due to its excellent learning effect and performance, dictionary learning has been widely applied in several applications, including image classification [10–12], image denoising [13–15], image deblurring [16], image super-resolution [17,18], image compression [19] and so on. For classification, the problem is to determine which class previously unlabeled data belong to. Supposed a dictionary has already been built, a new datum is classified to the class with the minimum representation residual with respect to the datum. Thus, the used dictionary plays an important role in the success of classification because learning a desired dictionary form training data allows unseen data to be faithfully and discriminatively represented as a sparse linear combination of atoms.

As for face recognition, researchers adopt the following procedures to perform recognition: image capture, feature selection or feature extraction and classification. As we all know, efficient classification algorithms play a significant role in face recognition. Numerous efficient dictionary learning based face recognition algorithms have been developed. A large class of them relies on finding an appropriate transform to represent data more efficiently. According to whether known sample with label information or not, dictionary learning methods are mainly divided into two categories: supervised learning and unsupervised learning. The former uses the labeled samples. In contrast, the latter does not exploits label information. Besides, it is called semi-supervised learning [20–22] to utilize both labeled samples and unlabeled samples to generate classification functions. The semi-supervised robust dictionary learning [23] proposed by Wang et al. designs a data adaptive dictionary by imposing structured sparsity on the data representation coefficients to automatically select prominent dictionary basis vectors. The optimal dictionary size was learned from training data in a principled way and no heuristic pre-specification was required. Besides, based on the sample diversity, such as the sym-

* Corresponding author at: School of Computer Science, Shaanxi Normal University, Xi'an 710062, China.
E-mail address: shgliu@snnu.edu.cn (S. Liu).

https://doi.org/10.1016/j.sigpro.2018.01.013
0165-1684/

metrical face [24,25] and the mirror image [26], Xu et al. proposed a representation effective and robust dictionary learning method for face recognition [27].

A popular heuristic method for sparse dictionary learning is K-SVD [28] which generalizes K-means clustering and learns an over-complete dictionary that best represents the input data sparsely. The K-SVD algorithm is generally used as a benchmark dictionary learning algorithm for face recognition. After that, researchers added a discriminative term to the dictionary learning model by incorporating the learning of the classifier parameters into the optimization of dictionary learning, such as the discriminative K-SVD for dictionary learning (D-KSVD) in face recognition [29]. Yang et al. proposed the sparse representation based Fisher discrimination dictionary learning (FDDL) [30]. FDDL presents an optimization function which forces both the learned dictionary and the resulting sparse coefficients to be discriminative. In addition, the two-phase test sample sparse representation method (TPTSR) [31] is also proposed for sparse coding. A couple of related algorithms including the l_1 -regularized least squares (L1LS) [32] can be found in the survey [33]. Moreover, Jiang et al. proposed the label consistent K-SVD method (LC-KSVD) [34], which introduces a discriminative term to optimize the regular K-SVD. In the literature [35], Cai et al. assigned different weights for each pair of sparse representation vectors to extend the work about FDDL. Recently, Li et al. proposed a locality-constrained and label embedding dictionary learning algorithm (LCLE-DL) [36], where the locality information of atoms is embedded to the framework of DL to improve the discriminative ability of dictionary.

In addition, as an important method in the field of digital signal processing, Fourier transform provides an efficient approach to analyze and study an arbitrary signal by decomposing it into multiple sinusoidal components with different values of frequency, amplitude, and phase. From a pure mathematical sense, the Fourier transform deals with transforming a function into a series of periodic functions. From the view of physical effects, Fourier transform converts images from spatial domain to frequency domain. In other words, the physical meaning of Fourier transform is to transform the gray distribution function of an image into the frequency that of an image.

The famous fast Fourier transform (FFT) computes the discrete Fourier transform (DFT) of a sequence, or its inverse (IFFT). Owning to its low requirement in arithmetic computation, FFT (including 1D-FFT and 2D-FFT) has been widely used for many applications in the areas of signal processing, data analysis, and robotics [37,38].In this paper, we take original sample and the Fourier spectrum of samples into account together during the dictionary learning process and classification process. We apply the 2D-FFT algorithm to all images (including the training samples and testing samples) to obtain the frequency features of data. Then, we perform the dictionary learning on the original dataset and the Fourier transformed dataset, respectively, such that two dictionaries, i.e., the spatial-domain dictionary and the frequency-domain dictionary, are obtained to calculate residual scores individually. Also, we use the transformed samples to represent test samples by collaborative representation classification (CRC) to obtain residual scores, and finally fuse the residual scores with that of the previous DL process as the ultimate fusion score to classify a test samples.

Our main contributions for dictionary learning based face recognition are as follows. First, we take the native spatial domain and the Fourier frequency domain of face images into account together, which makes data complementary in spatial and frequency domains each other. Second, we combine the DL with CRC during the classification process, which makes the algorithms complementary for classification. Both the two aspects of complementation promote the discriminative ability of dictionary learning and improve the classification performance.

The rest of this paper is organized as follows. In the next section, we introduce the related work about our method. In Section 3, the novel viewpoint we proposed is illustrated in detail. The experimental results are presented in Section 4. Finally, Section 5 presents the conclusions.

2. Related work

Since our proposed method is based on dictionary learning (DL) methods and the representation based classification (RBC), this section mainly reviews the typical dictionary learning method, i.e., the K-SVD algorithm, and the classical representation based classification algorithm, i.e., the collaborative representation classification (CRC) [39] algorithm.

2.1. The typical DL method: K-SVD algorithm

Fundamentally K-SVD is a generalization of K-means clustering which can be also regarded as a method of sparse representation. K-SVD algorithm focuses on learning a codebook that can best represent the training samples by solving the object function as follows.

$$\arg\min_{D,X}\left\{\|A-DX\|_F^2\right\} \text{ s.t. } \forall i,\ \|x_i\|_0 \le T_0, \tag{1}$$

where $A \in \Re^{n\times N}$ is the matrix composed of all the training examples $\{\alpha_i\}_{i=1}^N$ in the form of column vectors, and n and N are the dimension and number of them, respectively. $D = [d_1, \ldots, d_K] \in \Re^{n\times K}$ is the learned dictionary, and K is the number of atoms. $X = [x_1, \ldots, x_N] \in \Re^{K\times N}$ is the matrix of coding coefficients of training samples, T_0 is the limit of sparsity.

The Eq. (1) is a joint optimization problem with respect to D and X, and the natural method for solving it is to alternately optimize the D and X iteratively.

Step1. When fixing dictionary D, the problem in Eq. (1) can be rewritten as follows:

$$\arg\min_{X}\|A-DX\|_F^2 \text{ s.t. } \forall i,\ \|x_i\|_0 \le T_0. \tag{2}$$

For Eq. (2), each column is independent to each other, so it can be transformed into the following equivalent optimization problem:

$$\arg\min_{x_i}\|\alpha_i-Dx_i\|_2^2 \text{ s.t. } \|x_i\|_0 \le T_0,\ i=1,2,\ldots,N. \tag{3}$$

Obviously x_i is adequately be solved by pursuit algorithms that decompose signals with respect to a given dictionary such as orthogonal matching pursuit (OMP) algorithm [40].

Step2. When X is fixed, the problem in Eq. (1) can be expressed as

$$\hat{D} = \arg\min_{D}\|A-DX\|_F^2. \tag{4}$$

To obtain the optimization of Eq. (4), K-SVD's core ideal is to update the atoms of the dictionary one by one by utilizing SVD. Specifically, only one column d_g in the dictionary and it's corresponding coefficients, the gth row x_T^g in X, is putted in question. Then, Eq. (4) can be rewritten as

$$\begin{aligned}\hat{D} &= \arg\min_{D}\|A-DX\|_F^2 = \arg\min_{D}\left\|A-\sum_{j=1}^{K}d_jx_T^j\right\|_F^2\\ &= \arg\min_{D}\left\|(A-\sum_{j\neq g}d_jx_T^j)-d_gx_T^g\right\| = \arg\min_{D}\left\|E_g-d_gx_T^g\right\|_F^2,\end{aligned} \tag{5}$$

where the matrix E_g stands for the representation residual for all the X training examples when the gth atom is removed.