Pattern Recognition 81 (2018) 341–356

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust, discriminative and comprehensive dictionary learning for face recognition

Guojun Lin [a,b], Meng Yang [a,e,*], Jian Yang [c], Linlin Shen [d], Weicheng Xie [d]

[a] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
[b] School of Automation and Electric Information, Sichuan University of Science and Engineer, Zigong, China
[c] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China
[d] School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[e] Key Laboratory of Machine Intelligence and Advanced Computing, Sun Yat-Sen University, Ministry of Education, China

ARTICLE INFO

Article history:
Received 31 October 2016
Revised 7 February 2018
Accepted 23 March 2018
Available online 30 March 2018

Keywords:
Dictionary learning
Face recognition
Sparse representation

ABSTRACT

For sparse representation or sparse coding based image classification, the dictionary, which is required to faithfully and robustly represent query images, plays an important role on its success. Learning dictionaries from the training data for sparse coding has shown state-of-the-art results in image classification and face recognition. However, for face recognition, conventional dictionary learning methods cannot well learn a reliable and robust dictionary due to suffering from the small-sample-size problem. The other significant issue is that current dictionary learning do not completely cover the important components of signal representation (e.g., commonality, particularity, and disturbance), which limit their performance. In order to solve the above issues, in this paper, we propose a novel robust, discriminative and comprehensive dictionary learning (RDCDL) method, in which a robust dictionary is learned from comprehensive training sample diversities generated by extracting and generating facial variations. Especially, to completely represent the commonality, particularity and disturbance, class-shared, class-specific and disturbance dictionary atoms are learned to represent the data from different classes. Discriminative regularizations on the dictionary and the representation coefficients are used to exploit discriminative information, which effectively improves the classification capability of the dictionary. The proposed RDCDL method is extensively evaluated on benchmark face image databases, and it shows superior performance to many state-of-the-art dictionary learning methods for face recognition.

1. Introduction

Inspired by sparse coding mechanism of human vision system, sparse representation represents a signal or an image vector as a sparse linear combination of representation bases which are atoms of the dictionary. Recently sparse representation technology has been successfully used in image restoration [1,2], image classification [3,4,7], and face recognition [5,6], etc. For the success of sparse representation, the dictionary is very important and should effectively represent the encoded signal or image vector [28]. As the dictionary, the analytically designed off-the-shelf bases (e.g., wavelets) might be universal to all types of images, but it will not be effective enough for specific tasks such as face recognition. Instead, many latest methods that learn properly the desired dictionary from the original training data have led to state-of-the-art results in many practical applications, which include image reconstruction [1,8], face recognition [10–12,14,15,21,36], and image classification [8,13,37,49].

Dictionary learning aims to learn the desired dictionary from the training samples. Basically the desired dictionary should well represent or code the given signal. One representative unsupervised dictionary learning model is the KSVD algorithm [16] that learns an over-complete dictionary from a set of image patches. Another unsupervised dictionary learning model is the analysis-synthesis dictionary learning method which learns a pair of dictionaries for image deblurring [47]. According to the relationship between dictionary atoms and class labels, current supervised dictionary learning can be categorized into three main types: class-shared dictionary learning, class-specific dictionary learning and hybrid dictionary learning. For class-shared dictionary learning, each dictionary atom can be used to represent all classes of data. For class-specific dictionary learning, each dictionary atom should be corresponded to only a single class. For hybrid dictionary learning, the hybrid dictionary includes class-shared dictionary and class-specific dictionary.

* Corresponding author at: School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China.
E-mail address: yangm6@mail.sysu.edu.cn (M. Yang).

https://doi.org/10.1016/j.patcog.2018.03.021

In the first category, a dictionary whose atoms are shared by all classes of data is learned while the discrimination of coding coefficients is exploited [8,10,11,19]. Marial et al. [19] proposed to learn simultaneously discriminative dictionaries with linear classifiers of coding coefficients. Based on KSVD [16], Zhang and Li [10] proposed a dictionary learning method called discriminative KSVD (DKSVD). Following DKSVD [10], Jiang et al. [11] added a label consistent term and proposed so-called label-consistent KSVD (LCKSVD). Recently, Mairal et al. [8] proposed a task-driven dictionary learning framework which minimized different risk functions of the representation coefficients for different tasks. Based on analysis dictionary learning (ADL) [40], Guo et al. [46] proposed discriminative ADL (DADL). Recently, Yang et al. [56] proposed a discriminative model of class-shared analysis and synthesis dictionary pair learning for face recognition. The class-shared dictionary that can represent all classes of data loses the relationship between dictionary atoms and class labels. Thus classifiers based on the class-shared dictionary cannot perform classification based on the class-specific representation residuals, which can weaken the classification capability.

In the second category, class-specific dictionary learning requires that each dictionary atom should correspond to a single class label, so that the class-specific reconstruction error can be used for classification [20–22,26,36]. Wright et al. [5] proposed to use the whole training set to sparsely encode a testing face image, and then classify the testing image by evaluating which class leads to the minimal class-specific reconstruction error. The sparse representation based classification (SRC) [5] framework has shown impressive face recognition results. Inspired by SRC, the class-specific dictionary is widely applied to the design of classifiers. Based on the KSVD [16] model, Mairal et al. [22] introduced a discriminative reconstruction penalty term. Yang et al. [17] and Sprechmann and Sapiro [18] learned a dictionary of sparse representation for each class. In order to encourage the dictionaries of different classes to be independent to each other, Ramirez et al. [20] proposed a model of dictionary learning with structured incoherence (DLSI) which minimized the coherence term of the dictionary to improve the discriminative capability of the dictionary. In action recognition based on images, Castrodad and Sapiro [26] learned a set of action-specific dictionaries with non-negative representation regularization. Yang et al. [21,36] proposed Fisher discrimination dictionary learning (FDDL), where both the representation residual and the representation coefficients achieved discriminative information. Inspired by FDDL, a new analysis and synthesis dictionary pair with Fisher regularized was developed in [57]. Gu et al. [41] proposed a projective class-specific dictionary pair learning algorithm for pattern classification. Although class-specific dictionary learning can achieve good performance, the coherence among the different class-specific sub-dictionaries is inevitable. The number of the dictionaries is usually large.

In the third category, the hybrid dictionary is the dictionary which combines the class-specific dictionary with the class-shared dictionary. Recently, some hybrid dictionary learning methods are proposed. Deng et al. [25] proposed extended sparse representation based classification (ESRC) which constructed an intraclass variation dictionary as a shared dictionary. ESRC achieved promising performance for face recognition with a single sample per person. Wei et al. [39] proposed undersampled face recognition via robust auxiliary dictionary learning. Zhou et al. [13] proposed joint dictionary learning (JDL) where a hybrid dictionary with a Fisher-like regularization on the coding coefficients was learned. Kong et al. [12] proposed dictionary learning with commonality and particularity (COPAR) which learned a hybrid dictionary by introducing an incoherence penalty term to the hybrid dictionary. Shen et al. [27] proposed a hybrid dictionary learning method where the desired dictionary had a hierarchical category structure. Yang et al. [48] proposed a novel dictionary learning method which was analysis-synthesis dictionary learning for universality-particularity representation based classification. Instead of predefining the relationship between dictionary atoms and class labels, Yang et al. [42] proposed a latent dictionary learning (LDL) method to learn a discriminative dictionary and build its relationship to class labels adaptively. However, these hybrid dictionary learning methods cannot well describe the disturbance such as noise, outliers and occlusion. In addition, these methods do not introduce the discriminative information to both the dictionary and the representation coefficients.

Though dictionary learning has achieved promising performance in face recognition, previous dictionary learning methods have some disadvantages. For example, for face recognition, conventional dictionary learning methods cannot well learn a reliable and robust dictionary due to suffering from the small-sample-size problem. Limited number of training samples cannot provide reliable information of face identity and variations so that the learned dictionary may not be robust in the practical application. Some methods about learning an occlusion dictionary [9,23,24] are proposed to recognize the occluded face images and achieve robust performance. However, they may not well handle the general variation in the practical face recognition. The other significant issue is that current dictionary learning do not completely cover the important components of signal representation (e.g., commonality, particularity, and disturbance), which limit their performance. In order to address the above problems, in the paper, we propose a novel robust, discriminative and comprehensive dictionary learning (RDCDL) model.

We propose RDCDL to use the training sample diversities of the same face image to get a robust dictionary. To achieve the robustness, RDCDL learns the dictionary from sample diversities by extracting real face variations and generating virtual face images that convey new possible variations, such as poses, corruption, and occlusion of the face. From original training samples, extracted face variations and virtual training samples, RDCDL learns the dictionary including class-shared dictionary, class-specific dictionary and disturbance dictionary in order to completely represent the practical data (e.g., the data of the different classes has class-shared components, class-specific components and disturbance components such as noise, outliers and occlusion). At the same time, the discriminative regularizations on the dictionary and the representation coefficients have exploited the discriminative information, which effectively improves the discriminative capability of the dictionary.

Although Xu et al. [50] proposed a dictionary learning framework which also used training sample diversities of the same face image and tried to obtain effective representations of face images and a robust dictionary, our proposed RDCDL is quite different from the framework. First, different from the class-shared dictionary learned in [50], we learn a more complete dictionary to represent the commonality, particularity and disturbance of signals. Especially the disturbance dictionary will well represent the disturbance component, with the clean part of signal represented by class-shared dictionary and class-specific dictionary. Second, our proposed model can use the powerful class-specific reconstruction error as the classification criterion which is not used in [50]. Third, apart from the sample diversities simulated by doing with the original training face images only used in [50], the practical face variations are extracted in our paper. What's more, in the proposed RDCDL, the discriminative information is introduced to the dictionary and the representation coefficients.

The rest of this paper is organized as follows. Section 2 briefly introduces related works. Section 3 presents the proposed RDCDL model. Section 4 describes the optimization procedure of RDCDL. Section 5 presents the RDCDL based classification. Section 6 con-