Expert Systems With Applications 69 (2017) 101–109

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Predictive high-level feature representation based on dictionary learning

E. Phaisangittisagul[a,*], S. Thainimit[a], W.K. Chen[b]

[a] Department of Electrical Engineering, Faculty of Engineering, Kasetsart University, Bangkok, 10900, Thailand
[b] Division of Information Technology and Sciences, Champlain College, Burlington, Vermont, 05401, USA

ARTICLE INFO

Article history:
Received 29 March 2016
Revised 10 October 2016
Accepted 10 October 2016
Available online 14 October 2016

Keywords:
Dictionary learning
High-level feature representation
K-SVD
Sparse coding
Supervised learning

ABSTRACT

A much improved computational performance of visual recognition tasks can be achieved by representing raw input data (low-level) with high-level feature representation. In order to generate the high-level representation, a sparse coding is widely used. However, a major problem in traditional sparse coding is computational performance due to an ℓ_0/ℓ_1 optimization. Often, this process takes significant amount of time to find the corresponding coding coefficients. This paper proposed a new method to create a discriminative sparse coding that is more efficient to compute the coding coefficients with minimum computational effort. More specifically, a linear model of sparse coding prediction was introduced to estimate the coding coefficients. This is accomplished by computing the matrix-vector product. We named this proposed method as predictive sparse coding K-SVD algorithm (PSC–KSVD). The experimental results demonstrated that PSC–KSVD achieved promising classification results on well-known benchmark image databases. Furthermore, it outperformed the currently approaches in terms of computational time. Consequently, PSC–KDVD can be considered as a suitable method to apply in real-time classification problems especially with large databases.

1. Introduction

In supervised learning, one of typical objectives is to create a mapping function from input data to a target output. For example, a common task in object classification such as face classification, handwritten classification, or text classification is to build a learning model to map an input image represented by pixel intensity values to a predefined category of the object appeared in the image. Although there are many effective supervised learning algorithms, learning input data (low-level features) directly is difficult to achieve high classification performance. This is because learning mapping from pixel intensity values to object class label is a complex nonlinear function. In fact, the performance of supervised learning is highly dependent on the choice of data representation. As a result, many research groups have proposed methods that are able to capture latent and high-fidelity representation (high-level features) from the raw input data.

In signal processing, a signal or function $f(t)$ can often be described as a linear decomposition, $f(t) = \sum_{j=1}^{k} a_j \varphi_j$ where a_j are the coefficients and φ_j are the set of functions. A signal can be uniquely represented as a linear combination of those functions as long as they form a basis set. This basis set is sometimes known in the computer vision community as dictionary. In addition, the calculation of the coefficients can be done efficiently by an inner product between the input signal and the basis set. For example, a basis set of a Fourier series consists of sines and cosines (or equivalently complex exponentials) functions at different harmonic frequencies. However, this predefined set of bases is less effective in modeling complex local structure of the images. Recently, an overcomplete basis set inspired by a mechanism of human vision system was proposed. Olshausen and Field (1996); 1997) have revealed that the receptive fields can extract meaningful information from images based on sparse coding. Many studies have successfully applied sparse representation in various applications of image restoration and compressed sensing (Bradley & Bagnell, 2008; Bryt & Elad, 2008; Candes, 2006; Mairal, Elad, & Sapiro, 2008b; Wagner, Wright, Ganesh, Zhou, & Ma, 2009; Wright, Yang, Ganesh, Sastry, & Ma, 2009; Yang & Zhang, 2010; Yang, Zhang, Feng, & Zhang, 2011a). Another essential motivation of using the sparse coding is that it can be applied to unlabeled data for dictionary learning due to the limitation of labeled data. This intrigued many researchers to exploit the possibility of applying sparse coding to learn a high-level feature representation in diverse tasks

* Corresponding author. Fax: +6625797566.
E-mail addresses: fengecp@ku.ac.th (E. Phaisangittisagul), fengsyt@ku.ac.th (S. Thainimit), chen@champlain.edu (W.K. Chen).

http://dx.doi.org/10.1016/j.eswa.2016.10.019
0957-4174/

(Elad & Aharon, 2006; Li, Cichocki, & Amari, 2004; Raina, Battle, Lee, Packer, & Ng, 2007; Starck, Elad, & Donoho, 2005).

Sparse coding can be viewed as a way of constructing an approximation of an input data by a sparse linear combination of a set of overcomplete dictionary according to:

$$\langle D, A\rangle \;=\; \arg\min_{D,A} ||X - DA||_2^2 + \lambda\|A\|_p \tag{1}$$

where X is a set of n-dimensional m input data, i.e., $X = [x_1, \ldots, x_m] \in \mathbb{R}^{n\times m}$. $D = [d_1, \ldots, d_K] \in \mathbb{R}^{n\times K} (K > n)$ is an overcomplete dictionary of K atoms and $A = [a_1, \ldots, a_m] \in \mathbb{R}^{K\times m}$ is a set of coding coefficients. λ is a regularization parameter. Finally, $\|A\|_p$ is an ℓ_p-norm regularization constraint on A to control the number of nonzero elements in its coding. In (1), the dictionary (D) and the corresponding coding coefficients (A) are the parameters to be optimized. Normally, this non-convex optimization problem is solved by alternatively iterating between the dictionary and coding coefficient updates. It is important to note that all the dictionary atoms (d_i) should have a unit norm to avoid the scenario in which the dictionary elements (atoms) have arbitrary large norm so that the coding coefficients are forced to have a small value. In sparse representation, the key success depends on the choice of the dictionary. The dictionary can be obtained by either taking from off-the-shelf bases (e.g., wavelets) (Huang & Aviyente, 2006) or learning from the data. Although off-the-shelf dictionaries might be typically applied to represent all data types, these dictionaries are not well represented for highly specific applications such as text or face classifications. In order to better capture the salient features of the data, various dictionary learning methods are developed for tasks such as reconstruction and classification problems.

The current prevailing approach is to create a dictionary based on learning from data which can be divided into two categories: unsupervised dictionary learning and supervised dictionary learning. An unsupervised dictionary learning does not use class information of the data to produce the dictionary. The learned dictionary is built with the goal of minimizing the residual error between reconstructed data and original data. Aharon, Elad, and Bruckstein (2006) proposed a K-SVD algorithm which can create a learned overcomplete dictionary from a set of unlabeled data. Such unsupervised dictionary learning improved results in image denoising (Elad & Aharon, 2006; Mairal, Bach, & Ponce, 2012; Zhou et al., 2012), image compression (Bryt & Elad, 2008), and image super-resolution (Yang, Wright, Huang, & Ma, 2008a). However, without using label information from the data, the unsupervised dictionary learning is not powerful enough for classification tasks.

The other category of dictionary learning is based on supervised learning in which the class label of the data is available to exploit in the dictionary learning process. As a result, discrimination capability could be boosted by a dictionary resulting from a learning process or sparse coding, or both to enhance the classification performance (Jiang, Lin, & Davis, 2013; Mairal, Bach, Ponce, Sapiro, & Zisserman, 2009; Ramirez, Sprechmann, & Sapiro, 2010; Yang, Zhang, Feng, & Zhang, 2014). One of the main focuses of this study is to create a high-level feature representation for object classification with discriminative capability based on supervised dictionary learning strategy.

In general, an existing supervised dictionary learning can be categorized into three classes. Firstly, each dictionary atom is shared by all classes while the coding coefficients are exploited as a high-level feature to train a classifier (Zhang & Li, 2010). Some approaches proposed to jointly learn a shared dictionary with a classifier while enforcing the coding coefficients to be discriminative (Jiang et al., 2013). For example, Mairal, Bach, Ponce, Sapiro, and Zisserman (2008a) proposed an effective method to learn a shared dictionary for creating a discriminative model. Other strategies merged or chose dictionary atoms from an initial large dictionary using different criteria such as intraclass and interclass discrimination (Huang & Aviyente, 2006), mutual information of class distribution (Liu & Shah, 2008), and submodular dictionary learning (Jiang, Zhang, & Davis, 2012).

Secondly, a supervised dictionary learning is designed to improve discrimination power among classes (Castrodad & Sapiro, 2012; Mairal et al., 2008a; Mairal, Leordeanu, Bach, Hebert, & Ponce, 2008c; Perronnin, 2008; Ramirez et al., 2010; Sivalingam, Boley, Morellas, & Papanikolopoulos, 2011; Yang, Jin, Sukthankar, & Jurie, 2008b; Yang et al., 2014; Zhang, Surve, Fern, & Dietterich, 2009; Zhou et al., 2012). This is called class-specific dictionary learning. In this category, each dictionary atom is learned for only a single class. In addition, the corresponding reconstruction error can be used for class assignment. Yang, Zhang, Yang, and Zhang (2011b) applied the Fisher discrimination criteria to the objective function to encourage discriminative representation in the coding coefficients. Ramirez et al. (2010) introduced class-specific dictionary to be independent by adding an incoherence promoting term. Wang, Yang, Nasrabadi, and Huang (2013) presented a margin-based perspective to dictionary learning in order to improve the classification accuracy.

Lastly, a supervised dictionary learning that combined the shared dictionary with the class-specific dictionary is named as hybrid dictionary learning. Kong and Wang (2012) shown that the combination of class-specific dictionary and a common pattern pool led to more compact and more discriminative dictionary for classification. Zhou and Fan (2012) proposed a joint dictionary learning algorithm to leverage the correlation within a group of the visually similar object categories to enhance the discrimination of the learned dictionary.

In sparse representation for classification, it consists of two phases: coding and classification. In coding, a sparse representation or coding coefficient (A) is determined from a dictionary (D) with some sparsity constraint. Then, the coding coefficients of the entire data are used to build a classification model for class prediction. Most proposed dictionary learning algorithms have focused on designing synthesis dictionary to not only well represent the original data but also produce better classification results while adopting the ℓ_o/ℓ_1-norm sparsity constraints. Most of the existing sparse representation is based on an iterative learning process that is time-consuming on both training and testing phases, thus prohibiting real-time applications. Although an efficient sparse coding optimization algorithms have been proposed by Lee, Battle, Raina, and Ng (2006), solving sparse representation is still a challenge in computational performance for large dataset. Hence, the idea of using linear projection to predict the coding coefficients is a very attractive alternative. A novel method that rapidly computes the coding coefficients using only matrix-vector multiplication was proposed in this study. The goal of this algorithm was to make the prediction of the coding coefficients as close as possible to the optimal set of coefficients acquired from the objective function of sparse representation constraints. The main contributions of this study are described as follows.

- A modified sparse representation constraint to create a new high-level feature representation based on discriminative dictionary learning is introduced. The new proposed object representation achieved better classification results in a number of benchmark datasets.
- An implementation of linear projection model to predict a set of coding coefficients is presented. This resulted in the significant improvement of computational speed for classifying test images comparing to existing approaches.

The rest of this paper is organized as follows. Section 2 describes the related works of sparse representation based on supervised dictionary learning. A newly proposed approach to build