Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Euler Label Consistent K-SVD for image classification and action recognition

Yue Song[a], Yang Liu[a,*], Quanxue Gao[a,*], Xinbo Gao[a], Feiping Nie[b], Rongmei Cui[a]

[a] State Key Laboratory of Integrated Services Networks, Xidian University, Shaanxi 710071, China
[b] Center for OPTical Imagery Analysis and Learning, Northwestern Polytechnical University, Shaanxi 710065, China

ARTICLE INFO

Article history:
Received 26 October 2017
Revised 24 April 2018
Accepted 12 May 2018
Available online xxx

Communicated by Steven Hoi

Keywords:
Similarity
ELC-KSVD
Euler space
Dictionary learning

ABSTRACT

Motivated by the fact that kernel trick can capture the nonlinear similarity of features, which may help improve the separability and enlarge the margin between nearby data points, we present an effective kernel dictionary learning approach, namely Euler Label Consistent K-SVD (ELC-KSVD), for sparse coding and image recognition. ELC-KSVD first maps the images into the complex space by Euler representation, which has a negligible effect for outliers and illumination, and then learns a discriminative dictionary in Euler space. Different from the most existing kernel dictionary learning approaches, which maps data into a hidden high-dimensional space, Euler representation not only is explicit but also does not increase the dimensionality of image space in our ELC-KSVD. This makes ELC-KSVD algorithm efficient and easy to be realized in real applications. Furthermore, an iterative method is provided to solve ELC-KSVD. This iteration algorithm is fast and has good convergence. Extensive experimental results illustrate that ELC-KSVD outperforms some representative methods and achieves impressive performance for image classification and action recognition.

1. Introduction

Sparse coding has been successfully applied to solve many problems in image processing, pattern recognition and computer vision, including image classification [1–4], image restoration [5–8] and image denoising [9–12]. Sparse coding reconstructs an input signal by a linear combination of a few atoms from an overcomplete dictionary. The most representative sparse representation classification (SRC) method uses training samples as a dictionary, called the dictionary, and obtains coefficients by solving a least squares problem (squared Euclidean distance) with L_0-norm or L_1-norm regularizer on the coefficients [13]. An obtained sparse representation is a real valued vector that has a few non-zero coefficients with respect to this dictionary. Unlike traditional methods such as Eigenface [14] and Fisherface [15], SRC does not need a feature extraction stage, and has achieved impressive performances on face recognition and image denoising.

In real applications, it is time-consuming to solve sparse code when the size of training samples is very large. Moreover, training samples usually do not contain enough discriminant information, thus, the performance of SRC is not good enough. To handle this problem, methods for learning a small-sized dictionary for sparse-coding from the training samples have been developed. For example, Aharon et al. [16] proposed a method named K-SVD to efficiently learn an over-complete dictionary from a set of training signals. This algorithm has been shown to work well in image denoising and compression. However, K-SVD only focuses on the representational power of the dictionary and ignores the label information. Thus, it cannot well encode discriminative information. To solve this problem, supervised dictionary learning methods are proposed. For example, Pham and Venkatesh [17] proposed a joint representation and classification framework and reported competitive results learning a single dictionary for all classes. Zhang and Li [18] extended this approach and proposed. Discriminative K-SVD (D-KSVD) algorithm, which aimed to learn a single dictionary with formulating a linear classifier as a joint optimization problem. Furthermore, Jiang et al. [19] proposed Label Consistent K-SVD (LC-KSVD). LC-KSVD added a label consistence term in the objective function of D-KSVD to encourage the signals from the same class to have similar sparse codes and those from different classes to have dissimilar sparse codes, thus the discrimination power of the dictionary is effectively improved.

The aforementioned dictionary learning algorithms consider a linear sparse model, which cannot effectively characterize the nonlinear properties embedded in images such as face images. To well exploit nonlinear features, which are important for classification,

* Corresponding authors.
E-mail addresses: 523240290@qq.com (Y. Liu), qxgao@xidian.edu.cn (Q. Gao).

https://doi.org/10.1016/j.neucom.2018.05.036
0925-2312/

and improve the performance of SRC, kernel sparse representation has become an active topic in image recognition and machine learning [20–25]. For example, Van et al. [23] proposed the kernel KSVD for sparse and redundant signal representations in high dimensional feature space using the kernel method. Zhang et al. [20] proposed the kernel based sparse classification. In [26], a sparse coding method was proposed by embedding Riemannian manifolds into reproducing kernel Hilbert spaces in the criterion function. Liu et al. [27] proposed Euler SRC which maps the images into a Euler space by explicit Euler representation and does not increase the dimensionality of image space. Although motivations of these methods are different, both of them need to map each image into a high-dimensional hidden space by kernel function. This makes the computational complexity significantly improved.

Motivated by the fact that kernel trick can capture the nonlinear similarity of features, we propose a novel kernel dictionary learning approach, Euler Label Consistent K-SVD (ELC-KSVD), for image classification. ELC-KSVD first maps each data such as image into a complex space by Euler transformation and then performs LC-KSVD in Euler space. Euler representation not only reveals nonlinear patterns embedded in data but also has a negligible effect for outliers and illumination. To solve ELC-KSVD, we present an effective algorithm to solve sparse code with complex valued coefficients and use classical KSVD to update dictionary. Extensive experiments illustrate that the proposed method is superior to some presentative dictionary learning methods and has good convergence. Our contributions are summarized as follows.

We use Euler transform to map each image into an explicit space that has the same dimensionality as the original image space, while most existing kernel dictionary learning methods map image into hidden and high-dimensional space. Thus, our method is simple and can be easily realized in real applications. Moreover, Euler transform not only well captures nonlinear features from data but also has negligible effect for outliers and illumination. This helps enlarge the separability of data having different labels. Finally, we propose an efficient algorithm to solve SRC with complex valued vectors.

2. Related works

2.1. Dictionary learning for reconstruction

Given an input sample matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbf{R}^{m\times n}$, learning a reconstructive dictionary for sparse representation of $\mathbf{Y}$ can be accomplished by solving the following problem:

$$< \mathbf{D}, \mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \quad s.t. \quad \|\mathbf{x}_i\|_0 < T_0 \quad (i = 1, \ldots, n) \tag{1}$$

where T_0 is the sparsity constraint, making sure each signal has fewer than T_0 items in its decomposition. $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \mathbf{R}^{m\times K}$ $(K > m)$ is the learned dictionary, and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbf{R}^{K\times n}$ is the sparse codes of training sample matrix $\mathbf{Y}$. The K-SVD algorithm [16] is an iterative method to minimize the energy in (1). It has a wide range of applications in image compression and restoration.

Given $\mathbf{D}$, sparse coding computes the sparse representation $\mathbf{x}_i$ of $\mathbf{y}_i$ by the following objective function.

$$\mathbf{x}_i = \operatorname*{argmin}_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \quad s.t. \quad \|\mathbf{x}_i\|_0 \le T_0 \tag{2}$$

Eq. (2) can be solved by the orthogonal matching pursuit algorithm (OMP) [28].

To efficient solve the model (1), an alternative formulation for (1) is to replace L_0-norm regularization with L_1-norm regularization to enforce sparsity, i.e., we solve sparse code by the model (3) in real applications.

$$< \mathbf{D}, \mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \gamma \|\mathbf{X}\|_1 \tag{3}$$

where γ is a parameter to balance the reconstruction error and sparsity.

Similarly, given $\mathbf{D}$, the sparse representation $\mathbf{x}_i$ of an input signal $\mathbf{y}_i$ can be solved by

$$\mathbf{x}_i = \arg\min_{\mathbf{x}_i} \frac{1}{2}\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \gamma \|\mathbf{x}_i\|_1 \tag{4}$$

Eq. (4) can be solved by some efficient L_1-norm optimization approaches, such as [29].

2.2. Dictionary learning for classification

Given label matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n] \in \mathbf{R}^{c\times n}$, where c denotes the number of classes. Each column $\mathbf{h}_i$ of the label matrix $\mathbf{H}$ encodes the class label of ith sample using the position of the non-zero value. For example, if the training sample $\mathbf{y}_i$ belongs to the second class, then we have $\mathbf{h}_i = [0, 1, 0, \ldots, 0]^T$. We hope to use the given label matrix $\mathbf{H}$, to learn a linear classifier $\mathbf{W} \in \mathbf{R}^{c\times K}$ taking in a sample's sparse representation $\mathbf{x}_i$, and returning the most probable class this sample belongs to. Zhang and Li [18] incorporated the classification error term into the dictionary learning formulation in (1) and proposed the discriminative K-SVD (D-KSVD). The objective function of D-KSVD is defined as follows.

$$< \mathbf{D}, \mathbf{W}, \mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{W},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \alpha\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2 \quad s.t. \quad \|\mathbf{x}_i\|_0 < T_0 \quad (i = 1, \ldots, n) \tag{5}$$

where α $(\alpha > 0)$ is a regularization parameter balancing the contribution of the classification error to the overall objective.

2.3. Label consistent K-SVD (LC-KSVD)

In order to encourage the similarity among sparse representations of signals belonging to the same class in D-KSVD, Jiang et al. [19] proposed label consistent K-SVD (LC-KSVD), which incorporated a discriminative sparse-code error term in (5). The objective function of LC-KSVD is defined as follows.

$$< \mathbf{D}, \mathbf{W}, \mathbf{A}, \mathbf{X} > = \arg\min_{\mathbf{D},\mathbf{W},\mathbf{A},\mathbf{X}} (\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \alpha\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2 + \beta\|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_F^2) \quad s.t. \quad \|\mathbf{x}_i\|_0 < T_0 \quad (i = 1, \ldots, n) \tag{6}$$

where $\mathbf{A} \in \mathbf{R}^{K\times K}$ is a linear transformation and $\mathbf{Q} \in \mathbf{R}^{K\times n}$ is the discriminative sparse codes matrix promoting label consistency. α and β are two regularization parameters balancing the discriminative sparse code errors and the classification contribution to the overall objective, respectively.

It is worthwhile to note that all of these three dictionary learning methods (K-SVD, D-KSVD and LC-KSVD) employ squared F-norm as the distance metric in the image space. However, the variations between images, which have the same class label under different illumination and time, is larger than the change of the image identity under the squared F-norm distance metric. Thus, this affects the robustness and reduces the flexibility of dictionary learning. Moreover, these methods do not well reveal nonlinear features embedded in images, while kernel trick can capture the nonlinear features of images, which encode more discriminative information [23,24].

Please cite this article as: Y. Song et al., Euler Label Consistent K-SVD for image classification and action recognition, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.05.036