Computer Vision and Image Understanding 118 (2014) 50–60

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Multiview Hessian discriminative sparse coding for image annotation

Weifeng Liu [a], Dacheng Tao [b,*], Jun Cheng [c,d], Yuanyan Tang [e]

[a] College of Information and Control Engineering, China University of Petroleum (East China), Qingdao, Shandong, China
[b] Centre for Quantum Computation & Intelligent Systems and Faculty of Engineering & Information Technology, University of Technology, Sydney, Ultimo, NSW 2007, Australia
[c] Shenzhen Key Lab for CVPR, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
[d] The Chinese University of Hong Kong, Sha Tin, Hong Kong, China
[e] Faculty of Science and Technology, University of Macau, Taipa, Macau, China

ARTICLE INFO

Article history:
Available online 30 August 2013

Keywords:
Image annotation
Hessian
Multiview
Sparse coding

ABSTRACT

Sparse coding represents a signal sparsely by using an overcomplete dictionary, and obtains promising performance in practical computer vision applications, especially for signal restoration tasks such as image denoising and image inpainting. In recent years, many discriminative sparse coding algorithms have been developed for classification problems, but they cannot naturally handle visual data represented by multiview features. In addition, existing sparse coding algorithms use graph Laplacian to model the local geometry of the data distribution. It has been identified that Laplacian regularization biases the solution towards a constant function which possibly leads to poor extrapolating power. In this paper, we present multiview Hessian discriminative sparse coding (mHDSC) which seamlessly integrates Hessian regularization with discriminative sparse coding for multiview learning problems. In particular, mHDSC exploits Hessian regularization to steer the solution which varies smoothly along geodesics in the manifold, and treats the label information as an additional view of feature for incorporating the discriminative power for image annotation. We conduct extensive experiments on PASCAL VOC'07 dataset and demonstrate the effectiveness of mHDSC for image annotation.

1. Introduction

Due to the prodigious development of sensors such as cameras and microphones, people can exploit huge amounts of high dimensional data carrying particular kinds of information. Considering the redundancy of these high dimensional data for a particular intelligent task, such as object categorization and human behaviour analytics, it is essential to properly represent the relevant information to reveal the underlying process of these observations.

Sparse coding aims to learn a dictionary and simultaneously find a sparse linear combination of atoms from this dictionary to represent the observations (*e.g.* images and image features). It has received growing attentions because of its flexibility and promising performance for many computer vision applications, such as image denosing [1] and inpainting [4].

In recent years, dozens of sparse coding algorithms have been developed and these algorithms can be grouped into the following five categories: reconstructive sparse coding, supervised sparse coding, discriminative sparse coding, structured sparse coding and graph regularized sparse coding.

(1) Reconstructive sparse coding: Reconstructive sparse coding methods learn the optimal dictionary and find the corresponding sparse representation by minimizing the data reconstruction error. The representative optimization methods for sparse representation include matching pursuit [16], orthogonal matching pursuit [19] and basis pursuit [3].
(2) Supervised sparse coding: Supervised sparse coding methods exploit the label information to learn an over-completed dictionary and the corresponding sparse representation for classification tasks. Pham and Venkatesh [20] considered the class label and the linear predictive classification error and proposed a joint framework of dictionary construction and classification. Zhang and Li [29] incorporated the labels directly into the sparse coding stage and proposed a discriminative K-SVD (D-KSVD) method to retain the separability. Jiang et al. [13] extended D-KSVD by integrating both labels and classification error.
(3) Discriminative sparse coding: Discriminant analysis [38,39] plays an important role for classification problems. In contrast to supervised sparse coding which straightforwardly exploit the class label information, discriminative sparse coding methods incorporate class separability criterion into the objective function. Popular class separability criteria include softmax function [15], Fisher discrimination criterion [24], and hinge loss [18]. Mairal et al. [15] used the classical softmax discriminative cost function to leverage the sparse coding. Yang et al. [24] introduced Fisher's discriminative criterion to sparse coding to ensure the sparse

* Corresponding author. Fax: +61 2 95144517.
E-mail address: dacheng.tao@uts.edu.au (D. Tao).

1077-3142/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2013.03.007

representations have large between-class scatter but small within-class scatter. Lian et al. [18] proposed a max-margin sparse coding method which combined the hinge loss function with sparse coding.

(4) Structured sparse coding: Structured sparse coding methods naturally extend reconstructive sparse coding by exploiting the structure sparsity such as group sparsity [28] and hierarchical sparsity [11]. Yuan and Lin [28] extended Lasso to group Lasso which considered group/block structured dependencies among the sparse coefficients. Jenatton et al. [11] employed hierarchical sparsity-inducing norms to learn a hierarchical dictionary which solved tree-structured sparse decomposition problems. Jia and Salzmann [12] exploited structured sparsity to learning a latent space of multiview data.

(5) Graph regularized sparse coding: Graph regularized sparse coding methods use graph regularization to exploit the local geometry of the data distribution. Graph Laplacian is a representative graph regularization. Zheng et al. [30] used graph Laplacian to exploit the local geometry of the data distribution by adding a Laplacian regularization (LR) to the sparse coding framework. Gao et al. [8] proposed hypergraph Laplacian regularized sparse coding to preserve the local consistence in the feature space.

Although the aforementioned sparse coding algorithms have obtained promising performance for various applications such as clustering, classification, and dimensional reduction, they share some of the following two major problems for image annotation:

(1) Since it is expensive to label a large number of images for training a robust model, manifold assumption based semi-supervised learning (SSL) has been introduced to integrate both a small number of labelled images and a large number of unlabelled images to improve the performance of image annotation. LR is one of the most representative works in which the geometry of the underlying manifold is determined by the graph Laplacian. Although LR achieved top level performance for image annotation, it suffers from lacking of extrapolating power. It has been identified that LR biases the solution towards a constant function due to its constant null space, which possibly leads to poor extrapolation capability [14].

(2) The aforementioned sparse coding methods are only applicable to images that are represented by single view features. However, in image annotation, images are often described by multiview features. Different views (or equivalently visual features), such as colour histogram, edge sketch and local binary patterns (LBP), characterize different properties of an image [7,17,21]. Each view of a feature describes a specific property of the image, and the weaknesses of a particular view can be reduced by the strengths of others. Although we can concatenate different features into a long vector, this concatenation strategy cannot efficiently explore the complementary of different features because it improperly treats different features carrying different physical characteristics. Therefore, compared to single view feature, multiview features provide more characteristics of images and can significantly leverage the performance especially when features for different views are complementary to one another [37,40–43].

To address these problems, we present multiview Hessian discriminative sparse coding (mHDSC) in this paper. Particularly, mHDSC can well leverage multiview sparse coding by seamless integrating Hessian regularization with discrimination. According to proposition 1 in [14], the geodesic function in null space of Laplacian is no other than a const, which implicates that LR biases the solution towards a constant function and then leads to poor extrapolation capability. In contrast to Laplacian, Hessian has richer null space and drives the solution varying smoothly along the manifold. Hessian regularization (HR) is more preferable for exploiting the local geometry than LR. Kim et al. [14,36,37] has demonstrated the excellent performance of HR in regression and classification problems. The proposed mHDSC has the following advantages: (1) mHDSC incorporates multiview features into sparse coding, which effectively explores the complementation of different features from different views; (2) mHDSC treats the label information as an additional view of feature, which well boosts the discrimination without adding more computing complexity; and (3) mHDSC exploits Hessian regularization to preserve local similarity, which steers the solution varying smoothly along geodesics in the manifold.

We carefully implement mHDSC for image annotation and conduct experiments on the PASCAL VOC'07 dataset [6]. To evaluate the performance of mHDSC, we also compare mHDSC with several baseline algorithms including discriminative sparse coding (DSC), Laplacian discriminative sparse coding (LDSC), Hessian discriminative sparse coding (HDSC), multiview sparse coding (mSC), multiview discriminative sparse coding (mDSC) and multiview Laplacian discriminative sparse coding (mLDSC). The experimental results demonstrate the effectiveness of mHDSC by comparison with the baseline algorithms.

The rest of this paper is arranged as follows. Section 2 presents the proposed mHDSC framework. Section 3 details the implementation of mHDSC. Section 4 discusses some related work. And Section 5 demonstrates experimental results followed by the conclusion in Section 6.

2. Multiview Hessian discriminative sparse coding

In multiview sparse coding (mSC), we are given a multiview dataset of N observations from V views including l labelled data *i.e.* $S_L = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(V)}, y_i\}_{i=1}^{l}$ and u unlabelled data *i.e.* $S_U = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(V)}\}_{i=l+1}^{N}$, where $y_i \in R^{P_c}$ is the class labels of the i^{th} example (P_c is the number of class). In the following section of this paper, we use $X_L^{(v)} \in R^{P_v \times l}$ to denote the v^{th} view feature vectors of labelled data (P_v is the dimension of the v^{th} view feature), $Y \in R^{P_c \times l}$ to denote the label vectors, and $X_U^{(v)} \in R^{P_v \times (N-l)}$ to denote the v^{th} view feature vectors of unlabelled data.

By incorporating an additional regularization term to control the sparsity and exploit the local geometry, mSC aims to find an integrated sparse representation (code) $W \in R^{N_d \times N}$ of the multiview data and a multiview dictionary $D = \{D^{(1)}, D^{(2)}, \ldots, D^{(V)}\}$, where $D^{(v)} \in R^{P_v \times N_d}$ contains N_d dictionary atoms for the view v. Thus, mSC is written as follows

$$\min_{D,W} \frac{1}{2N} \sum_{v=1}^{V} \|X^{(v)} - D^{(v)} W\|_F^2 + \varphi(W), \tag{1}$$
$$\text{s.t. } \|D_i^{(v)}\| \leqslant 1, \quad 1 \leqslant i \leqslant N_d, \ X^{(v)} = \{X_L^{(v)}, X_U^{(v)}\}$$

where $\varphi(W) = \gamma_1 \varphi_1(W) + \gamma_2 \varphi_2(W) + \gamma_3 \varphi_3(W)$, $\varphi_1(W) = ||W||_{1,\infty}$ is a regularizer that controls the sparsity over W, $\varphi_2(W) = \sum_{v=1}^{V} \|(D^{(v)})^T\|_{1,\infty}$ is a regularizer that controls the structure of dictionary, $\varphi_3(W)$ is a regularizer to preserve the local similarity, and γ_1, γ_2 and γ_3 are parameters that balance the loss function and regularizations $\varphi_1(W)$, $\varphi_2(W)$ and $\varphi_3(W)$, respectively.

Although there are different choices for $\varphi_2(W)$ to exploit the local geometry, Laplacian regularization (LR) [30,8] is promising to