Computer Vision and Image Understanding 123 (2014) 14–22

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Image classification by non-negative sparse coding, correlation constrained low-rank and sparse decomposition ☆

Chunjie Zhang [a], Jing Liu [b], Chao Liang [c,*], Zhe Xue [a], Junbiao Pang [d], Qingming Huang [a,e]

[a] School of Computer and Control Engineering, University of Chinese Academy of Sciences, 100049 Beijing, China
[b] National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences, P.O. Box 2728, Beijing, China
[c] National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, 430072 Wuhan, China
[d] College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China
[e] Key Lab of Intell. Info. Process, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 6 August 2012
Accepted 28 February 2014
Available online 14 March 2014

Keywords:
Sparse coding
Image classification
Low-rank decomposition
Non-negative
Correlation constrained

ABSTRACT

We propose an image classification framework by leveraging the non-negative sparse coding, correlation constrained low rank and sparse matrix decomposition technique (CCLR-Sc+SPM). First, we propose a new non-negative sparse coding along with max pooling and spatial pyramid matching method (Sc+SPM) to extract local feature's information in order to represent images, where non-negative sparse coding is used to encode local features. Max pooling along with spatial pyramid matching (SPM) is then utilized to get the feature vectors to represent images. Second, we propose to leverage the correlation constrained low-rank and sparse matrix recovery technique to decompose the feature vectors of images into a low-rank matrix and a sparse error matrix by considering the correlations between images. To incorporate the common and specific attributes into the image representation, we still adopt the idea of sparse coding to recode the Sc+SPM representation of each image. In particular, we collect the columns of the both matrixes as the bases and use the coding parameters as the updated image representation by learning them through the locality-constrained linear coding (LLC). Finally, linear SVM classifier is trained for final classification. Experimental results show that the proposed method achieves or outperforms the state-of-the-art results on several benchmarks.

1. Introduction

As a fundamental problem in computer vision, image classification has attracted a lot of attention in recent years. Among many image representation models, the-bag-of-visual-words (BoW) model [1] has been widely used by many researchers [2–4] and shown very good performance. The BoW model contains mainly two modules: (i) codebook generation and quantization of features extracted from local image patches; (ii) histogram based image representation and prediction. Recently, it has been shown that combining the two modules with sparse representation is very effective and can achieve the state-of-the-art performance.

As to the first module of the BoW model, k-means is usually used to generate codebook and quantize visual descriptors extracted from local image patches by nearest-neighbor search. A histogram is then computed to represent each image by counting the occurrence number of each visual word within this image. Recently, Yang et al. [4] developed an extension by generalizing vector quantization to sparse coding. By using sparse coding instead of k-means, they tried to learn the optimal codebook and coding parameters for local features simultaneously. The use of sparse coding helps to reduce the quantization loss. Multi-scale max pooling is then used to get the feature representation of images. However, sparse coding has no constraints on the signs of coding coefficients. To satisfy the objective of sparse coding, negative coefficients are sometimes needed, while large numbers of zero coefficients are inevitable. Since non-zero components typically provide useful information, the encoding process with max pooling will bring the loss in terms of those negative components, and further degrade the classification performance.

Instead of learning sparse representations for local features [4], the use of sparse representation for the final classification has also been widely applied to many visual applications and achieves the state-of-the-art performances, *e.g.*, image restoration [5] and classification tasks [6–12]. These holistically sparse representations on

☆ This paper has been recommended for acceptance by Nikos Paragios.

* Corresponding author.
E-mail addresses: cjzhang@jdl.ac.cn (C. Zhang), jliu@nlpr.ia.ac.cn (J. Liu), cliang@whu.edu.cn (C. Liang), zhxue@jdl.ac.cn (Z. Xue), junbiao_pang@bjut.edu.cn (J. Pang), qmhuang@jdl.ac.cn (Q. Huang).

http://dx.doi.org/10.1016/j.cviu.2014.02.013
1077-3142/

the whole image ensure robustness to occlusions and image corruptions. Training images are often chosen as the bases for sparse representation and testing images are then classified by assigning the class with the lowest reconstruction error. Ideally, a testing image can be reconstructed by the training samples and only the coefficients of the samples within the same class may be non-zero. This means a test image can be sufficiently reconstructed by the training images of the same class. However, images are often contaminated with noise. We use noise to denote the perturbing terms for the classification task. For example, the object to be classified in an image may be occluded by other objects; the background around the target object may hinder the recognition of this object. Besides, there are often multiple objects in an image with different poses and occlusions. Sometimes using the training images as the bases is not discriminative enough to boost the final classification performance. Moreover, images with similar visual features often share a lot of similarities and correlate with each other, hence exhibit *degenerated structure* [6]. This semantic information can help make correct classification.

In this paper, we propose a new image classification framework by leveraging the non-negative sparse coding, correlation constrained low-rank and sparse matrix decomposition techniques ($\text{CCLR-Sc}^{+}\text{SPM}$). Fig. 1 shows the flowchart of the proposed method. Our proposed framework consists of two contributions. First, we extend the recent work on image classification [4] by using non-negative sparse coding along with max pooling to reduce the information loss during the encoding process of local features. The second is our main contribution. We propose a new image classification method by using the correlation constrained low-rank and sparse matrix decomposition technique. Our work is motivated by the observations that: Although images may be captured under various conditions with varied poses and occlusions, (i) the visual features of images within the same class are often similar and related. That means images of the same class exhibit some *degenerated structure* [6]. Ideally, if we stack the BoW representation of images of the same class into a matrix, this matrix will probably be low-rank. The ideal case means images are represented without any disturbance and can be well separated without any mistake; (ii) when one particular image is captured, it only contains a limited number of disturbances for classification. This results in the characteristics of noise sparsity for the stacked BoW matrix of the same class. (iii) images with similar BoW representations are more likely to have similar objects and related. This visual similarity consistency should also be preserved for their corresponding low-rank component. This correlation information should be jointly combined with the low-rank and noise components of images for better representation than directly using the BoW representation of training images.

Specially, to get more discriminative sparse coding bases with the BoW representation of images, we leverage the correlation constrained low-rank and sparse matrix decomposition technique to decompose the BoW representation of images into a low rank matrix and a sparse error matrix. We then use these bases to encode the BoW representation of images with sparsity and locality constraints. These coding parameters are used to represent images and linear SVM classifier is then utilized to predict the category labels of images. Experimental results demonstrate the effectiveness of the proposed method.

Comparing with our previous work [12], we extend the low-rank and sparse matrix decomposition per class with correlation constraints for all classes. This correlation constraints combine the relationship of images for low-rank and sparse decomposition, hence are more effective and robust to noise. Under certain conditions, [12] can be viewed as a special case of the proposed non-negative sparse coding, correlation constrained low-rank and sparse matrix decomposition technique ($\text{CCLR-Sc}^{+}\text{SPM}$). In addition, more experiments are added to clarify the effectiveness of $\text{CCLR-Sc}^{+}\text{SPM}$.

The rest of this paper is organized as follows. Section 2 introduces some related work. Section 3 presents the proposed non-negative sparse coding spatial pyramid matching method (Sc^{+}SPM). Section 4 shows the proposed image classification method by correlation constrained low-rank and sparse matrix decomposition method. Experimental results are given in Section 5. Finally we conclude in Section 6.

2. Related work

The use of the bag-of-visual-words (BoW) model [1] has been proven very useful for image classification. Over the past few years, many works have been done to improve the performance of the BoW model. Some tried to learn discriminative visual codebook for image classification [13,14]. Co-occurrence information of visual words was also modeled in a generative framework [15,16]. Others tried to learn discriminative classifiers by considering the spatial information and correlations among visual words [2–4,7,10,11]. To overcome the loss of spatial information in the BoW model, motivated by Grauman and Darrell's [3] pyramid matching in feature space, Lazebnik et al. [2] proposed the spatial pyramid matching (SPM). Since its introduction, SPM has been widely used and proven very effective.

Recently, Yang et al. [4] proposed a sparse coding based approach for soft coding and achieved the state-of-the-art performance for image classification when only one type of local feature (SIFT) is used. This method can automatically learn the optimal codebook and search for the optimal coding weights for each local feature. After this, max pooling along with SPM is used to get the feature representation of images. Inspired by this, Wang et al. [17] proposed to use locality to constrain the sparse coding process which can be computed faster and yields better performance. [11,18] also tried to jointly learn the optimal codebooks and classifiers. However, sparse coding [19] has no constraints on the sign of the coding coefficients, negative parameters are sometimes needed to satisfy the sparse coding constrains. For some particular applications [20], non-negative sparse coding [21] is needed.

Not only has sparse coding been used with local features, but also it has been widely used holistically on the entire image. Wright et al. [6] tried to view face recognition as finding a sparse representation of the test image by treating the training set as the bases and impressive results were achieved. Bradley and Bagnell [9] tried to train a compact codebook using sparse coding. Yuan and Yan [7] made visual classification with multi-task joint sparse representation by fusing different types of features. Liu et al. [20] tried to learn sparse and non-negative representations of images by solving a set of regression type non-negative matrix factorization problems. However, because images are often

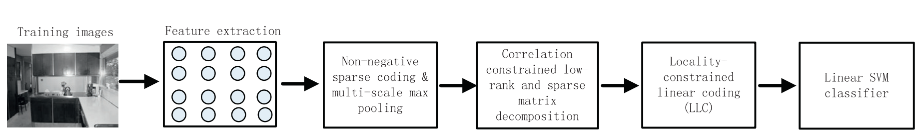

Fig. 1. The flowchart of the proposed method.