Knowledge-Based Systems 90 (2015) 14–22

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Fast low rank representation based spatial pyramid matching for image classification

Xi Peng [a], Rui Yan [b], Bo Zhao [a], Huajin Tang [b,*], Zhang Yi [b]

[a] Institute for Infocomm Research, A*STAR, 138632, Singapore
[b] College of Computer Science, Sichuan University, Chengdu 610065, China

ARTICLE INFO

Article history:
Received 1 April 2015
Revised 1 October 2015
Accepted 3 October 2015
Available online 16 October 2015

Keywords:
Closed-form solution
Efficiency
Image classification
Thresholding ridge regression
ℓ_2-regularization

ABSTRACT

Spatial Pyramid Matching (SPM) and its variants have achieved a lot of success in image classification. The main difference among them is their encoding schemes. For example, ScSPM incorporates Sparse Code (SC) instead of Vector Quantization (VQ) into the framework of SPM. Although the methods achieve a higher recognition rate than the traditional SPM, they consume more time to encode the local descriptors extracted from the image. In this paper, we propose using Low Rank Representation (LRR) to encode the descriptors under the framework of SPM. Different from SC, LRR considers the group effect among data points instead of sparsity. Benefiting from this property, the proposed method (i.e., LrrSPM) can offer a better performance. To further improve the generalizability and robustness, we reformulate the rank-minimization problem as a truncated projection problem. Extensive experimental studies show that LrrSPM is more efficient than its counterparts (e.g., ScSPM) while achieving competitive recognition rates on nine image data sets.

1. Introduction

Image classification system automatically assigns an unknown image to a category according to its visual content, which has been a major research direction in computer vision and pattern recognition. Image classification has two major challenges. First, each image may contain multiple objects with similar low level features, it is thus hard to accurately categorize the image using the global statistical information such as color or texture histograms. Second, a medium-sized grayscale image (e.g., 1024×800) corresponds to a vector with dimensionality of 819, 200, this brings up the scalability issue with image classification techniques.

To address these problems, numerous impressive approaches [1–5] have been proposed in the past decade, among which one of the most popular methods is Bag-of-Features (BOF) or called Bag-of-Words (BOW). BOW originates from document analysis [6,7]. It models each document as the joint probability distribution of a collection of words. [8–10] incorporated the insights of BOW into image analysis by treating each image as a collection of unordered appearance descriptors extracted from local patches. Each descriptor is quantized into a discrete "visual words" corresponding to a given codebook (i.e., dictionary), and then the compact histogram representation is calculated for semantic image classification.

The huge success of BOF has inspired a lot of works [11,12]. In particular, Lazebnik et al. [13] proposed Spatial Pyramid Matching (SPM) which divides each image into $2^l \times 2^l$ blocks in different scales $l = 0, 1, 2$, then computes the histograms of local features inside each block, and finally concatenates all histograms to represent the image. Most state-of-the-art systems such as [14–18] are implemented under the framework of SPM and have achieved impressive performance on a range of image classification benchmarks like Columbia University Image Library-100 (COIL100) [19] and Caltech101 [20]. Moreover, SPM has been extensively studied for solving other image processing problems, e.g., image matching [21], fine-grained image categorization [22]. It has also been incorporated into deep learning to make deep convolutional neural networks (CNN) [23] handling arbitrary sized images possible. To obtain a good performance, SPM and its extensions have to pass the obtained representation to a Support Vector Machine classifier (SVM) with nonlinear Mercer kernels. This brings up the scalability issue with SPMs in practice.

Although SPM has achieved state-of-the-art recognition rates on a range of databases, its computational complexity is very high. To speed up SPM, Yang et al. [24] proposed using Sparse Code (SC) instead of Vector Quantization (VQ) to encode each Scale-Invariant Feature Transform (SIFT) descriptor [25] over a codebook. Benefiting from the good performance of sparse code, Yang's method (namely ScSPM) with linear SVM obtains a higher classification accuracy, while using less time for training and testing.

* Corresponding author. Tel.: +8615982833478.
E-mail addresses: pangsaai@gmail.com (X. Peng), ryan@scu.edu.cn (R. Yan), zhaob@i2r.a-star.edu.sg (B. Zhao), htang@scu.edu.cn, huajin.tang@gmail.com (H. Tang), zhangyi@scu.edu.cn (Z. Yi).

http://dx.doi.org/10.1016/j.knosys.2015.10.005
0950-7051/

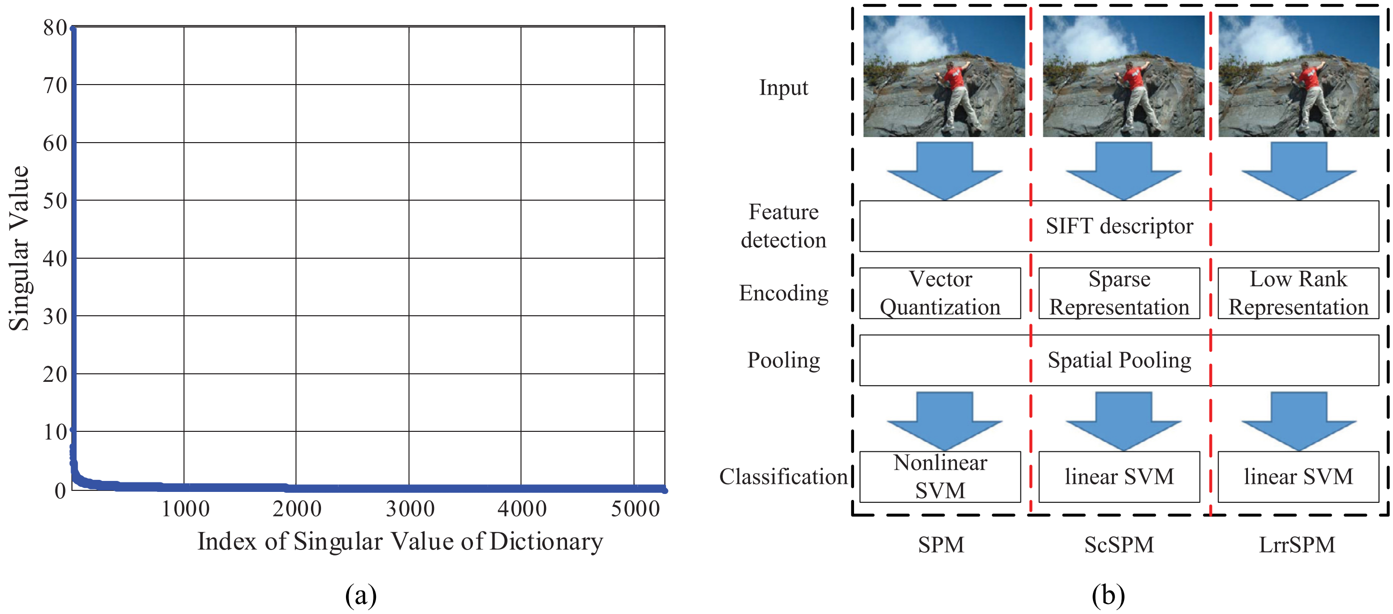

Fig. 1. (a) Singular values of a given codebook. The codebook consists of 5120 basis. It shows that most energy concentrates on the top singular values. (b) Schematic comparison of the original SPM, ScSPM and the proposed LrrSPM.

The success of ScSPM could be attributed to that SC can capture the manifold structure of data sets. However, SC encodes each data point independently without considering the grouping effect among data points. Moreover, the computational complexity of SC is proportional to the cube of the size of codebook (denoted by n). Therefore, it is a daunting task to perform ScSPM when n is larger than 10, 000. To solve these two problems, this paper proposes using Low Rank Representation (LRR) rather than SC to hierarchically encode each SIFT descriptor.

To show our motivation, i.e., the collections of descriptor and representation are low rank, we carry out k-means clustering algorithm over the SIFT descriptors of the Caltech101 database [20] and obtain a codebook $\mathbf{D}$ consisting of 5120 cluster centers. By performing Singular Value Decomposition (SVD) over the codebook shown in Fig. 1(a), one can see that most energy (over 98%) concentrates on the top 2% singular values. In other words, the data space spanned by the codebook is low rank. For a testing data set $\mathbf{X} \in span(\mathbf{D})$, its representation can be calculated by $\mathbf{X} = \mathbf{DC}$. Since $\mathbf{X}$ and $\mathbf{D}$ are low rank, then $\mathbf{C}$ must be low rank. This observation motivates us to develop an novel SPM method, namely Low Rank Representation based Spatial Pyramid Matching (LrrSPM). Fig. 1 illustrates a schematic comparison of the original SPM, ScSPM, and LrrSPM. It should be pointed out that, SPM, ScSPM, and LrrSPM are three basic models which do not incorporate the label information, kernel function learning, and multiple descriptors learning into their encoding schemes. The major difference among them is that both SPM and ScSPM perform encoding in the vector space, whereas LrrSPM calculates the representation in the matrix space.

The contributions of the paper are summarized as follows: (1) Different from the existing LRR methods [26–28], the proposed LrrSPM is a multiple-scale model which integrates more discriminative information compared to the traditional LRR. (2) Most existing LRR methods are proposed for clustering, which cannot be used for classification directly. In this paper, we fill this gap based on our new mathematical formulation. (3) Our LrrSPM has a closed form solution and can be calculated very fast. After the dictionary is learnt from the training data, LrrSPM computes the representation of testing data by simply projecting each testing datum into another space. Extensive experimental results show that LrrSPM achieves competitive results on nine image databases and is $25 - 50$ times faster than ScSPM.

Table 1
Some used mathematic notations.

Notation	Definition
n	The number of descriptors (features)
l	The scale or resolution of a given image
m	The dimensionality of the descriptors
s	The number of subjects
k	The size of codebook
r	The rank of a given matrix
$\mathbf{y}$	An image
$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$	A set of features
$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k]$	Codebook
$\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n]$	The representation of $\mathbf{X}$ over $\mathbf{D}$

The rest of the paper is organized as follows: Section 2 provides a brief review on two classic image classification methods, i.e., SPM [13] and ScSPM [24]. Section 3 presents our method (i.e., LrrSPM) which uses multiple-scale low rank representation to represent each image. Section 4 carries out some experiments using nine image data sets and several popular approaches. Finally, Section 5 concludes this work.

Notations: Lower-case bold letters represent column vectors and upper-case bold ones denote matrices. $\mathbf{A}^T$ and $\mathbf{A}^{-1}$ denote the transpose and pseudo-inverse of the matrix $\mathbf{A}$, respectively. $\mathbf{I}$ denotes the identity matrix. Table 1 summarizes some notations used throughout the paper.

2. Related works

In this section, we mainly introduce SPM and ScSPM which employ two basic encoding schemes, i.e., vector quantization and sparse code. To the best of our knowledge, most of other SPM based methods can be regarded as the extensions of them, e.g., the method proposed in [17] is a kernel version of ScSPM.

Let $\mathbf{X} \in \mathbb{R}^{m \times n}$ be a collection of the descriptors and each column vector of $\mathbf{X}$ represents a feature vector $\mathbf{x}_i \in \mathbb{R}^m$, SPM) [13] applies VQ to encode $\mathbf{x}_i$ via

$$\min_{\mathbf{C},\mathbf{D}} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{D}\mathbf{c}_i\|_2^2 \quad \text{s.t.}\ Card(\mathbf{c}_i) = 1, \tag{1}$$