Neurocomputing 166 (2015) 44–52

Contents lists available at ScienceDirect

Neurocomputing

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

Extending Laplacian sparse coding by the incorporation of the image spatial context

Mahmoud Mejdoub [a,b], Mouna Dammak [b], Chokri Ben Amar [b]

[a] Department of computer science, College of Al Ghat, P.O. BOX 445, 11914, Majmaah university, Al Majmaah, Riyadh, Kingdom of Saudi Arabia
[b] REGIM: Research Groups on Intelligent Machines, University of Sfax, National School of Engineers (ENIS), Sfax, 3038, Tunisia

ARTICLE INFO

Article history:
Received 7 August 2014
Received in revised form
14 November 2014
Accepted 31 March 2015
Communicated by Shuiwang Ji
Available online 24 April 2015

Keywords:
Spatial context
Laplacian
Encoding and pooling
Sparse Coding
Visual phrases

ABSTRACT

Diverse studies have shown the efficiency of sparse coding in feature quantization. However, its major drawback is that it neglects the relationships among features. To reach the spatial context, we proposed in this paper, a novel sparse coding method called Extended Laplacian Sparse Coding. Two successive stages are required in this method. In the first stage, the sparse visual phrases based on Laplacian sparse coding are generated from the local regions in order to represent the geometric information in the image space. The second stage aims to incorporate the spatial relationships among local features in the image space into the objective function of the Laplacian sparse coding. It takes into account the similarity among local regions in the Laplacian sparse coding process. The matching between the local regions is based on the Hungarian method as well as the histogram intersection measure between sparse visual phrases already assigned to the local regions in the first stage. Furthermore, we suggested to improve the pooling step that succeeds the encoding step by introducing the discretized max pooling method that estimates the distribution of the responses of each local feature to the dictionary of basis vectors. Our experimental results prove that our method outperforms the existing background results.

1. Introduction

Several works [1–4,37,38] have proved the performance of Bag of visual Words (BoW) in numerous computer vision applications such as image categorization and image and video retrieval [40,41]. The concept of the BoW [5,34–36] model is based on two successive steps of coding and pooling. The encoding step assigns the local descriptors to the visual words while the pooling step aggregates the visual words belonging to an image or a part of the image into a vector. The standard BoW model is based on K-means to construct the codebook of visual words. In the encoding step, each feature is assigned to the closest visual word. In the pooling step, the sum pooling is applied in order to build a histogram that counts the occurrences of each visual word. Machine learning recent studies aim to enhance and present more robust algorithms for the histogram encoding. Compared to hard assignment, sparse coding has consistently improved the results of image categorization topic [6]. Sparse coding aims to learn a dictionary of basis vectors and simultaneously find a sparse linear combination of the basis vectors to represent the image features. In the pooling step that follows the sparse coding, the max pooling is used since it is more robust to local spatial translations and more biologically plausible [7]. It has recently gained popularity due to its better performance when paired with sparse coding and simple linear classifiers. Perronnin et al. [8] show that the max pooling strategy is particularly well suited to the separation of features that are very sparse. Thus, the preciseness of the quantization process is enhanced. In recent years, several sparse coding methods have been proposed. Among them, two groups can be distinguished: supervised sparse coding and discriminative sparse coding. The methods of the first group use the class label information to learn an over-completed dictionary as well as the corresponding sparse representation for the categorization tasks. Zhang et al. [9] added the labels into the sparse coding step explicitly and suggested a Discriminative K-Singular Value Decomposition technique (D-KSVD) to enhance the separability between classes. Jiang et al. [10] extended the D-KSVD by integrating both of the classification error and class labels. The methods of the second group integrate class separability criterion into the objective function of the sparse coding. Yang et al. [11] introduced Fisher discriminative criterion in order to ensure that sparse codes will have large intra-class correlation and small inter-class correlation. Mairal et al. [12] used the softmax discriminative cost function to leverage the sparse coding. All the previous works [6,9–12] treated local features independently. The reciprocal dependency among local features is disregarded, ensuring that the sparse coding may differ widely even for close features. To overcome this drawback, different extensions of the sparse coding method [12,6,13] have been suggested recently by adding some regularization or

E-mail addresses: m.mejdoub@mu.edu.sa (M. Mejdoub), mouna.damak@ieee.org (M. Dammak), chokri.benamar@ieee.org (C.B. Amar).

http://dx.doi.org/10.1016/j.neucom.2015.03.086
0925-2312/

constraints in the sparse coding objective function. The Locality-constrained Linear Coding (LLC) technique [14] considers the locality information in the feature coding process. Contrary to the classical sparse coding, the LLC enforces locality instead of sparsity. It uses the k nearest neighbors of features as the local basis. This leads to smaller coefficients for the basis vectors far away from each local feature. Laplacian Sparse Coding (LSC) [13] learns an unsupervised dictionary as well as the sparse representation that keeps the conformity of the close local descriptors. This method has used histogram intersection similarity based on K-Nearest Neighbours (KNN) method to construct a Laplacian matrix that tries to preserve the locality consistency in the feature space. Only the K-nearest local features are selected to active the Laplacian matrix. This method obtains background results on several objects recognition.

In this paper, we propose a novel sparse coding method called Extended Laplacian Sparse Coding (ELSC) that extends the Laplacian sparse coding method by taking into account the spatial context in the image representation. Our approach is made up of two successive stages. In the first stage, the sparse visual phrases are computed for each local region sparse codes by pooling sparse codes of the patches located in the local region. In the second stage, we aim to incorporate the image space locality constraints in the encoding phase of the classical Laplacian sparse coding. We integrate into the objective function of the Laplacian sparse coding the similarity among local regions. To find the optimum matching between two local regions we consider this problem as an instance of the assignment problem [15], which can be solved by the Hungarian method [16]. We also measure the similarity between two local regions by calculating the histogram intersection between their sparse visual phrases already computed in the first stage.

Furthermore, we improved the statistical analysis of the sparse codes in the pooling step of ELSC. Instead of applying the max pooling [6,17,18] that represents an image by compacting the responses of all the sparse codes related to a basis vector into a single scalar, we propose the Discretized Max Pooling (DMP). Our purpose is to consider the locality constraints into the sparse codes space in the pooling step. For that, we construct a histogram that estimates the distribution of the sparse code weights belonging to each basis vector. The bth bin of the histogram is represented by the maximal weight of the sparse codes that falls into this bin. In comparison with the sparse BoSSA pooling presented in our previous work [19], we replaced the sum pooling on the discretized sparse codes by a max pooling. Max pooling strategy is applied because it has been shown to be considerably adapted to the sparse codes. Fig. 1 illustrates the architecture of our proposed method (ELSC-DMP): the Extended Laplacian Sparse Coding coupled with the Disretized Max Pooling. Our contributions in this paper can be summarized as follows:

1. In the encoding step, we propose a novel sparse coding method in order to enrich the image spatial information during the encoding phase. Compared to Laplacian Sparse Coding [13] that exploits the dependencies between the local features only in the feature space, we suggest exploiting the dependencies among them in both feature and image spaces. To this end, we incorporated the spatial image context information in the Laplacian sparse coding by integrating the similarity among local regions into the objective function of the Laplacian sparse coding using both the Hungarian method and the intersection histogram measure between sparse visual phrases associated to the regions.
2. In the pooling step, inspired by the BoSSA method that applies a statistical analysis on the distances between the local features and the K-means clusters, we developed a novel pooling method based on performing statistical analysis for the sparse codes.

The remainder of this paper is organized as follows: Section 2 reviewed the related works, Section 3 presented the sparse coding

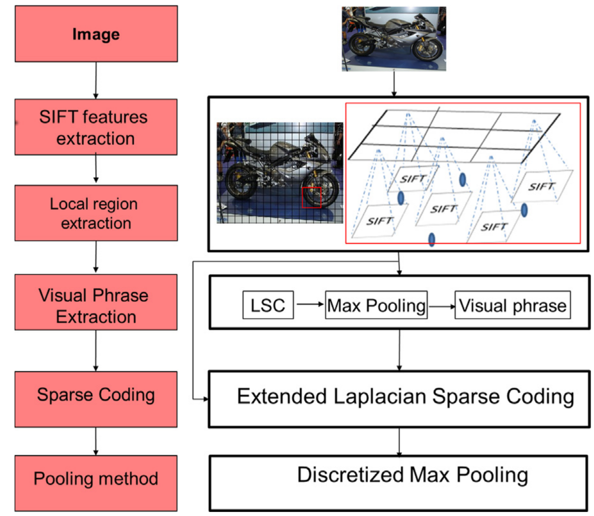

Fig. 1. Architecture of the proposed method (ELSC-DMP).

formulation and particularly the Laplacian sparse coding. Section 4 described the proposed Extended Laplacian Sparse Coding and the proposed Discretized Maximum Pooling (referred to as ELSC-DMP). Experimental results on several datasets are shown in Section 5. Finally, our conclusions were drawn in Section 6.

2. Related works

The sparse coding method solves the problem of hard quantization efficiently. But, like the BoW approach, it ignores the spatial relationship between the local features in the image space. To tackle the problem of the spatial information in the BoW and the sparse coding formalism, several approaches have been proposed. The Spatial Pyramid matching (SPM) is used in [20,6,14,13] in order to describe the global spatial information in the image. Explicitly, each image is split into progressively finer partitions and a Pyramid Match Kernel [21] is applied to match the corresponding partitions. The major inconvenient of this representation is that it neglects the local spatial information. Other approaches [22,23] group a set of visual words in terms of their proximity in the image. These approaches are called visual phrases. Using the spatial context, the visual phrases are represented by a histogram that reflects the distribution of the visual words located in a local neighborhood of the image. After that, a pairwise matching between visual phrases is applied to compute the similarity between two images. Moreover, the Bag of Graphs (BoG) model has recently been used in [23] to integrate spatial relationships among visual words. It proposes to extend the Bag of Words (BoW) model signature by computing a codebook of visual graphs. In [23], the K-means is first used to assign each feature to the closest visual word. Afterwards, to describe the spatial relationships among the visual words, the image was represented with a set of connected three sized local graphs using Delaunay triangulation. Each graph vertex is labeled by the visual word of its corresponding feature and the edge is labeled by computing the Local Binary Pattern (LBP) on the region located between two vertices. The classification graph technique based on the Hungarian method [16] was used to define the clusters of graphs. These clusters correspond to the visual graph codebook. Finally, an image can be represented by a histogram that describes the occurrences of visual graphs. The major inconvenient of the visual phrase and the BOG approaches is the hard quantization established in order to compute the visual words. Few works [24,19,39] tried to combine in the same time the distinctiveness of the local spatial information and the efficiency of the sparse