J. Vis. Commun. Image R. xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Feature discovering for image classification via wavelet-like pattern decomposition ☆

Yurui Xie *, Hongliang Li, Chao Huang, Bo Wu, Linfeng Xu

School of Electronic Engineering, University of Electronic Science and Technology of China, China

ARTICLE INFO

Article history:
Received 1 March 2016
Revised 22 June 2016
Accepted 5 August 2016
Available online xxxx

Keywords:
Dictionary learning
Image classification
Sparse representation
Hierarchical architecture

ABSTRACT

In this paper, we propose a feature discovering method incorporated with a wavelet-like pattern decomposition strategy to address the image classification problem. In each level, we design a discriminative feature discovering dictionary learning (DFDDL) model to exploit the representative visual samples from each class and further decompose the commonality and individuality visual patterns simultaneously. The representative samples reflect the discriminative visual cues per class, which are beneficial for the classification task. Furthermore, the commonality visual elements capture the communal visual patterns across all classes. Meanwhile, the class-specific discriminative information can be collected by the learned individuality visual elements. To further discover the more discriminative feature information from each class, we then integrate the DFDDL into a wavelet-like hierarchical architecture. Due to the designed hierarchical strategy, the discriminative power of feature representation can be promoted. In the experiment, the effectiveness of proposed method is verified on the challenging public datasets.

1. Introduction

Image classification is a fundamental problem in computer vision area. How to construct the discriminative feature for image representation has great impact to the performance of classification task. Among different methods, the bag-of-word (BOW) is one of the most successful methods for image representation. The standard BOW model includes four steps: feature extraction, codebook generation (dictionary learning), feature coding and feature pooling.

Codebook generation has significant influence in BOW model. Intrinsically, the codebook contains a set of visual elements. These visual elements are then used to represent each local feature within an image. Initially, the codebook is constructed using the K-means [1], and each local feature is assigned to its closest visual element in the codebook. However, this hard assignment strategy may cause information loss since the local feature is approximated by only one visual element in codebook. To address the problem, the soft-assignment method [2] is proposed to represent local feature using more than one visual element. This strategy can incorporate more visual information into the encoded feature and improve the classification performance. Sparse representation is one of the most successful soft-assignment methods and it can encode local feature using only a few number of visual elements in the codebook. To alleviate the spatial information loss, spatial pyramid matching (SPM) is proposed to incorporate the spatial layout information into the feature coding process. Recently, SPM based sparse representation (ScSPM) is presented [3] and achieves the state of the art performance in many classification tasks. Following the idea of [3], other research works [4,5] also propose a new coding property, called locality, to represent the local feature. The main advantage of [4,5] is that it incorporates the distance constrain between the local feature and the visual element of codebook into the feature coding process and obtains the good performance.

In order to exploit the more discriminative visual elements, many research works [6–17] recently focus on designing the discriminative dictionary learning model with the aid of labeled data. However, the above image classification methods based on sparse coding treat all the input samples as a whole during dictionary learning procedure. These methods neglect that the "representative visual samples" from each class tend to reflect the class-specific discriminative information. Specifically, the representative visual samples per class are defined as the feature data which are near the centroid of all the class-specific feature samples associated with the corresponding class by distance measure. Since representative data can capture the main characteristic of class-specific feature space, it is beneficial for the supervised learning

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: gloriousxyr@163.com (Y. Xie).

http://dx.doi.org/10.1016/j.jvcir.2016.08.002
1047-3203/

task that employs the representative data sufficiently. Besides, the above classification methods only use the flat dictionary learning architecture to generate the codebook for feature coding. Therefore, the discriminative visual cues may not be exploited effectively.

During the past decades, the Discrete Wavelet Transform (DWT) [18,19] is one of the most successful methods for signal analysis. The main advantage of DWT is that it offers a way to decompose the input signal into the fine component and coarse component. In particular, the fine component corresponds to the high frequency part of DWT, which characterizes the information of significant change in input signal. In contrast, the coarse component is the low frequency part in DWT and represents some information with gradual change of input signal. Besides, DWT has the great advantage of being able to analyze the signal using multiple resolutions effectively. Inspired by the main characteristics of DWT, we design a DFDDL model incorporated with a wavelet-like hierarchical architecture to alleviate the above two problems. The goal of DFDDL is to discover the representative visual samples per class and decompose the commonality and individuality visual patterns in a unified objective function. These representative samples reflect some discriminative visual patterns from each class. Furthermore, the commonality visual patterns of DFDDL are equivalent to the low frequency components in wavelet transform, which are used to exploit the communal visual information for all classes. Meanwhile, the individuality visual patterns of each class are regarded as the high frequency components of DWT, which are also collected in DFDDL model for capturing the class-specific visual information. To further explore the discriminative visual patterns, the multi-scale characteristic of DWT is also introduced to our method. Therefore, we design a wavelet-like hierarchical architecture and incorporate the DFDDL model into it. With the increase of layer in the hierarchical architecture, the more discriminative visual information can be discovered step by step. Fig. 1 demonstrates the flowchart of proposed method for image classification via feature discovering integrated with a wavelet-like hierarchical architecture.

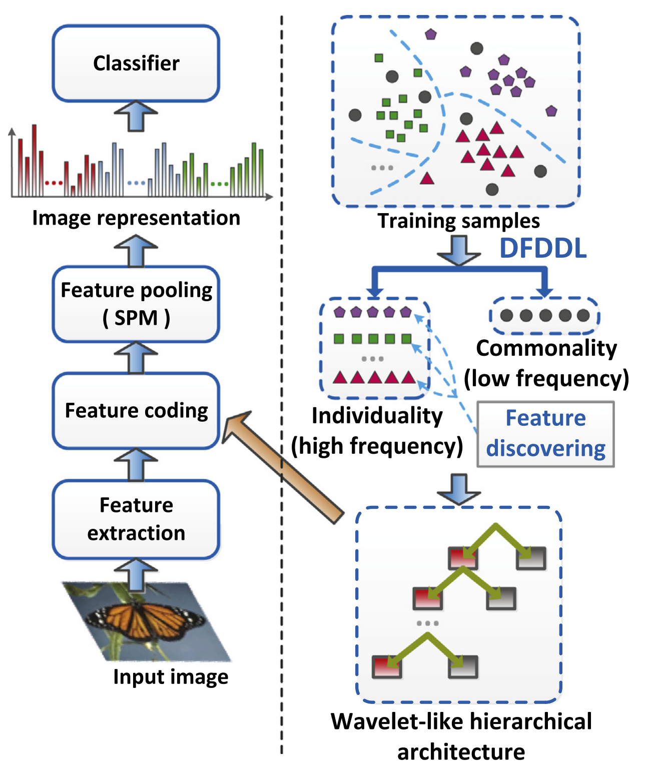

Fig. 1. The flowchart of proposed feature discovering method for image classification incorporated with a wavelet-like pattern decomposition strategy.

2. Related work

Recently, the bag-of-word (BOW) model has received great success for image classification task. Many research works have focused on developing this model. One of the main research directions aims to learn the discriminative codebook (dictionary) by the unsupervised or supervised manner. Using the learned dictionary, we can generate the feature representation with more discriminative information.

The first class of dictionary learning methods [20,21] aims to learn the dictionary in an unsupervised manner. Finally, the learned dictionary can fit the training data in terms of the reconstruction property. Specifically, the K-SVD [20] algorithm develops the k-means clustering to obtain the optimal dictionary. Then the learned dictionary can represent the input data with the minimization reconstruction error. In order to reduce the computational complexity, Lee et al. [21] cast the sparse representation into the least squares problem. Thus this optimization formulation can be solved using the Lagrange dual algorithm. Based on the above unsupervised dictionary learning algorithms, many research works [1,3–5,22–25] propose the effective framework to address the different computer vision problems.

The second class of dictionary learning methods [6–17,26–29] learns the dictionary using the label information of training data. By this manner, the more discriminative information can be captured for classification problem. Recently, there are two ways to arrive the goal. The first way is to make the representation coefficients of input feature discriminative. Following this way, some literatures [7,9,13,14,16,29] integrate the classification error term into the K-SVD [20] model to promote the classification performance. In addition, Jiang et al. [26] integrate both a label consistent constraint and the linear classification error into the K-SVD model to further improve the discriminative power of feature representation. The authors in [8] add a patch quality regularization term with supervised manner to optimize the standard K-SVD method. Another way aims to learn multiple dictionaries in a unified objective function. Each dictionary is associated with one of the classes from the training dataset. Then all the class-specific dictionaries are combined together to encode the local feature within an image. According to this idea, Wright et al. [12] use the easiest manner to construct all the class-specific dictionaries using the training samples themselves. In the test step, the label of input image is obtained by computing the reconstruction error of each class. Though it is the simple manner to generate the class-specific dictionaries, it always needs the high computational cost if the amount of training samples is too large.

To address this problem, other research works [6,11,15,17,27,28] focus on constructing the class-specific dictionaries by the learning manner. Among these methods, J. Mairal et al. [28] propose to learn multiple sub-dictionaries corresponding to each class for improving the texture segmentation problem. Ramirez et al. [11] incorporate an incoherent penalty term into the traditional dictionary learning model. This incoherent term makes the learned multiple dictionaries to be as independence as possible. Yang et al. [15] combine the fisher discrimination criterion into the dictionary learning model to make the set of class-specific dictionaries discriminative. The works of [17,27] seek to learn a set of class-specific dictionaries and a shared dictionary across all classes for enhancing the discrimination of feature representation. In addition, Gu et al. [6] learn a structured synthesis dictionary and a structured analysis dictionary to improve the discriminative power of feature representation.

Please cite this article in press as: Y. Xie et al., Feature discovering for image classification via wavelet-like pattern decomposition, J. Vis. Commun. (2016), http://dx.doi.org/10.1016/j.jvcir.2016.08.002