Information Sciences 310 (2015) 118–129

Contents lists available at ScienceDirect

Information Sciences

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

Image classification using boosted local features with random orientation and location selection

Chunjie Zhang [a], Jian Cheng [b,*], Yifan Zhang [b], Jing Liu [b], Chao Liang [c], Junbiao Pang [d], Qingming Huang [a,e], Qi Tian [f]

[a] School of Computer and Control Engineering, University of Chinese Academy of Sciences and Key Laboratory of Big Data Mining and Knowledge Management, 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] Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, 100124, China
[e] Key Lab of Intell. Info. Process, Institute of Computing Technology, Chinese Academy of Sciences, 100190 Beijing, China
[f] Department of Computer Sciences, University of Texas at San Antonio, TX 78249, USA

ARTICLE INFO

Article history:
Received 15 August 2014
Received in revised form 4 January 2015
Accepted 9 March 2015
Available online 17 March 2015

Keywords:
Sparse coding
Image classification
Random orientation
Boosting
Local feature selection

ABSTRACT

The combination of local features with sparse technique has improved image classification performance dramatically in recent years. Although very effective, this strategy still has two shortcomings. First, local features are often extracted in a pre-defined way (e.g. SIFT with dense sampling) without considering the classification task. Second, the codebook is generated by sparse coding or its variants by minimizing the reconstruction error which has no direct relationships with the classification process. To alleviate the two problems, we propose a novel boosted local features method with random orientation and location selection. We first extract local features with random orientation and location using a weighting strategy. This randomization process makes us to extract more types of information for image representation than pre-defined methods. These extracted local features are then encoded by sparse representation. Instead of generating the codebook in a single process, we construct a series of codebooks and the corresponding encoding parameters of local features using a boosting strategy. The weights of local features are determined by the classification performances of learned classifiers. In this way, we are able to combine the local feature extraction and encoding with classifier training into a unified framework and gradually improve the image classification performance. Experiments on several public image datasets prove the effectiveness and efficiency of the proposed method.

1. Introduction

Image classification is a classic problem in computer vision. It tries to classify one image to a pre-defined class by analyzing the image's content. Recently, the use of sparse coding for image classification becomes popular. Sparse coding [6] tries to minimize the reconstruction error of one given feature by selecting a relatively small subset of basis sets. Since its introduction, the sparse coding technique and its variants have attracted more and more researchers' attention and have been proved effective for many vision applications [12,31,44,45].

* Corresponding author.
E-mail addresses: cjzhang@jdl.ac.cn (C. Zhang), jcheng@nlpr.ia.ac.cn (J. Cheng), yfzhang@nlpr.ia.ac.cn (Y. Zhang), jliu@nlpr.ia.ac.cn (J. Liu), cliang@whu.edu.cn (C. Liang), junbiao_pang@bjut.edu.cn (J. Pang), qmhuang@jdl.ac.cn (Q. Huang), qitian@cs.utsa.edu (Q. Tian).

http://dx.doi.org/10.1016/j.ins.2015.03.011

Basically, sparse coding based image classification method can be divided into three steps. First, local features are extracted either by detection or dense sampling. Second, the codebook is generated (using sparse coding or its various variants) and local features are encoded accordingly. Finally, images are represented using the encoded parameters and SVM classifiers are trained to predict the categories of images. Although this strategy has been proven very effective for image classification. It still has two shortcomings. On one hand, the local feature extraction process and the classification task are relatively independent. On the other hand, the optimal parameters for minimizing reconstruction error cannot be able to help training discriminative classifiers for prediction. For example, the classifying of tiger and cat is different from separating cat and dog. Although researchers have tried [2,9,46] to alleviate the two problems, the results are far from satisfactory with heavy computational cost. If we can combine the local feature encoding and classifier training into a unified framework, we are able to model images better for classification.

To alleviate the two problems mentioned above, we propose to classify images using boosted local features with random orientation and location selection. We first extract local features by randomly choosing the locations and orientations with re-weighting to extract more types of information than pre-defined local feature extraction strategies (e.g. dense sampling of SIFT features). For each random extraction strategy, we generate the corresponding codebook with re-weighting and encode the features accordingly. Then we train SVM classifiers to make prediction of image classes. The outputs of the trained classifiers are used to re-weight images. This process is iterated in a boosting way to combine the discriminative power of a series of classifiers for classification. In this way, we can unify the extraction of local features, the generation of codebook and the training of classifiers into a unified framework.

Our main contribution are as follows.

- We propose a random local feature extraction strategy with orientation and location selection. This helps us to extract more types of information for classification. Besides, a re-weighting scheme is also imposed to extract more information from the 'hard' images which may help the classification task.
- We iteratively generate codebooks using the randomly extracted local features with re-weighting. The weights are determined by the predictions of learned classifiers. The misclassified images gains more weights compared with the correctly classified images, making the proposed method concentrates on the 'hard' images for each round.
- We unify the local feature extraction, codebook generation and classifier training into a unified framework iteratively by using the boosting strategy. The discrepancy between the predicted classes and groundtruth is used to re-weight the training images which are then used for local feature extraction. In this way, we can gradually improve the image classification performances by combining the sparse coding based image representation with the boosting strategy.

The rest of this paper is organized as follows. Section 2 introduces some related work. Section 3 gives the details of the proposed boosted local feature with random orientation and location selection method for image classification. Experimental comparisons are given in Section 4 and finally we conclude in Section 5.

2. Related work

The bag-of-visual-words (BoW) model [38] is widely used for image classification in recent years due to its simplicity and efficiency. k-means is usually used for codebook generation and nearest neighbor assignment is leveraged to quantize local features. This hard assignment strategy causes information loss which hinders final classification performance. Gemert et al. [14] tried to softly encode local features while Yang et al. [45] used the sparse coding technique. Motivated by this, a lot of works have been done [13,42,48–51]. Wang et al. [42] added locality constraints during the sparse coding process to speed up the computation and improve the performance. Gao et al. [13] explored the relationship of local feature similarities and encoded parameter similarities with Laplacian sparse coding. Zhang et al. [48] used non-negative sparse coding instead of sparse coding to ensure consistency with the max pooling strategy.

Most image classification methods used histogram based features for local region description, such as SIFT [29] and HoG [5]. To speed up computation of local features, many works have been done [1,18]. Speeded up robust features (SURF) was proposed by Bay et al. [1] which can be computed 3–7 folds faster than SIFT. The hashing technique [18] was also proposed by Indyk and Motwani. These local features are then encoded either by sparse coding or its variants to get the image representation for classification. This is often achieved by minimizing the reconstruction error with sparsity constraints. Although very effective, most of the above mentioned methods treated local feature generation and classifier training separately for image classification. The objectives of codebook generation and local feature encoding are minimizing the reconstruction errors while the objective of classifier training is to minimize the classification error. To solve this problem, Yang et al. [46] tried to unify the codebook generation with classifier training for object recognition by modeling on the SIFT features directly while the use of nearest neighbor information for direct image classification was also proposed by Boiman et al. [2]. Lowe [30] extended it by using nearest neighbor information to speed up the computation. However, the computational cost of [2,30,46] are very high compared with the BoW model for classification.

Instead of using pre-defined local feature extraction strategy, the use of features learned from the images also becomes popular [8,16,20,37] in recent years. Deep belief network (DBN) [16] and convolutional neural network (CNN) [20] tried to learn multiple layers of nonlinear features from images. Shao et al. [37] used multi-objective genetic programming