Information Sciences 281 (2014) 373–385

Contents lists available at ScienceDirect

Information Sciences

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

Spatial adjacent bag of features with multiple superpixels for object segmentation and classification

Wenbing Tao [a,e,*], Yicong Zhou [b], Liman Liu [c,d], Kunqian Li [a], Kun Sun [a], Zhiguo Zhang [a]

[a] School of Automation, National Key Laboratory of Science & Technology on Multi-Spectral Information Processing, Ministry Key Laboratory for Image Processing and Intelligence Control, Huazhong University of Science and Technology, Wuhan 430074, China

[b] Department of Computer and Information Science, University of Macau, Macau, China

[c] School of Biomedical Engineering, South-Central University for Nationalities, Wuhan 430074, China

[d] National Engineering Research Center for E-learning, Central China Normal University, Wuhan 430073, China

[e] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

ARTICLE INFO

Article history:
Received 18 September 2013
Received in revised form 9 May 2014
Accepted 19 May 2014
Available online 4 June 2014

Keywords:
Spatial adjacent bag of features
Superpixel adjacent histogram
Object recognition
Multiple segmentations

ABSTRACT

In the paper we present a new Spatial Adjacent Bag of Features (SABOF) model, in which the spatial information is effectively integrated into the traditional BOF model to enhance the scene and object recognition performance. The SABOF model chooses the frequency of each keyword and the largest frequency of its neighboring pairs to construct the feature histogram. Using the feature histogram whose dimension is only twice larger than that of the original BOF model, the SABOF model drastically enhances the discrimination performance. Combining the Superpixel Adjacent Histogram (SAH) Fulkerson et al., 2009 [12] with multiple segmentations Pantofaru et al., 2008 [33] and Russell et al., 2006 [36], the SABOF method effectively deals with the segmentation and classification of objects with different sizes. Changing the segmentation scale parameter to obtain multiple superpixel segmentations and correspondingly adjusting the neighbor parameters of the SAH method multiple classifiers are trained so that, the SABOF method can fuse multiple results of these classifiers to obtain better classification performance than any single classifier. The superpixel-based conditional random field (CRF) is used to further improve the classification performance. The experimental results of scene classification and of object recognition and localization on classical data sets demonstrate the performance of the proposed model and algorithm.

1. Introduction

Object localization and classification are both challenging tasks in the computer vision society, and are extremely important for image understanding. Recently, a number of attentions have been paid to solving these two problems in a unified framework. Sliding window approaches are of the successful object localization techniques [4,8,26,42]. Considering a sliding window around each pixel, these approaches apply a classifier function to find the best classification to fit the sliding window. They have been extensively used to detect the location of an object in an image. In [26] Blaschko et al. used the branch and bound method to search all possible subwindows in an image. In [47] Wei and Tao proposed an efficient

* Corresponding author at: School of Automation, National Key Laboratory of Science & Technology on Multi-Spectral Information Processing, Ministry Key Laboratory for Image Processing and Intelligence Control, Huazhong University of Science and Technology, Wuhan 430074, China. Tel.: +86 27 87541924.

E-mail address: wenbingtao@hust.edu.cn (W. Tao).

http://dx.doi.org/10.1016/j.ins.2014.05.032

histogram-based sliding window method that utilizes the spatial coherence of natural images and computes the objective function in an incremental manner. However in many cases, we want to perform the pixel-level object segmentation. Recently, some joint segmentation and classification methods [12,33,38] have been developed to integrate the segmentation and recognition into a unified framework and to automatically segment the image into several semantically meaningful regions where each region is labeled as a specific object class. Most of these approaches are based on the bottom-up local feature representation and often use the conditional random field [17,37] model to constrain the spatial consistency.

The classical Bag of Features (BOF) method represents an image with an orderless collection of local features and has been demonstrated to have impressive performance in object segmentation and classification [5,18,38]. However, due to the lack of information about the spatial structure of features, its descriptive ability is extremely limited. To overcome this, this paper proposes the Spatial Adjacent Bag of Features (SABOF) method to effectively integrate the spatial information for the pixel-level object segmentation and classification. To construct the feature histogram, the SABOF method considers not only the frequency of each keyword but also the frequency of every pair of keywords which are spatially neighboring. The frequency of each keyword and the largest frequency of its neighboring pairs are used to represent the feature histogram. Thus, using the feature histogram with the dimension just twice larger than that of the original BOF model, the SABOF method drastically enhances the discriminative power. Our experiments are provided to demonstrate that including more neighboring pairs to construct the feature histogram is not necessary and does not benefit the classification performance of the SABOF method.

Moreover, based on the proposed SABOF model, we integrate multiple segmentations with the Superpixel Adjacent Histogram (SAH) framework [12]. The SAH [12] was proposed to avoid the sparse features of each single superpixel and to provide context information learning. The quick shift algorithm [44] was used to extract superpixels from images with a fixed scale parameter. However, it was not explicitly stated how many adjacent superpixels would lead to the best performance in [12]. In our observation, the effect of the neighbor parameter N in SAH is closely related to the scale of superpixels. If each superpixel includes a small number of pixels, larger N would lead to better performance. On the other hand, if each superpixel includes a large number of pixels, a small value should be given to N. Therefore in this paper, considering the varied sizes of objects in images, we propose to fuse the changed parameter N with the multiple superpixel segmentations to enhance the adaptability and robustness of the SAH framework. Furthermore, more structure information of objects is obtained to enhance the performance of object segmentation and classification.

Multiple superpixel segmentations can be obtained by changing the scale parameter of the quick shift algorithm. See Fig. 1, from left to right, the scale parameter σ gradually becomes larger. For each superpixel result, a suitable neighbor parameter N is used in the SAH framework to construct a SABOF classifier. Thus, multiple SABOF classifiers are obtained combining multiple segmentation scale parameters with neighbor parameters of SAH. For testing images, multiple segmentations are provided and classified by the multiple SABOF classifiers. The multiple classifications of one image are combined to obtain the final segmentation and recognition result. Moreover, the superpixel-based conditional random field (CRF) is incorporated to further improve the segmentation performance.

The outline of the paper is as follows. In the next section we overview the related object segmentation and classification methods. In Section 3 we present the SABOF model. The SAH method with multiple segmentations is presented in Section 4. We validate our algorithm in Section 5 to show the advantages of our proposed model and algorithm, followed by a brief conclusion in Section 6.

2. Related works

Joint segmentation and classification have been studied in [6,9,24,32,38,48], where a global shape model is usually exploited and a unified framework integrates segmentation and recognition. They can efficiently classify only the highly structured objects but difficultly address the cases of severe occlusion and arbitrary viewpoints. Local features like textons [38,40] were usually applied for class segmentation algorithms to obtain pixel-level results. Moreover, a conditional random field or other spatial coherency constraint [37,43] was exploited to refine the results. Considering computational costs, some class segmentation algorithms operate on a reduced grid of the image to achieve a fast speed while sacrificing pixel accuracy. Other methods used superpixels [2,21,31,32,36] to increase the computational efficiency. The superpixels correspond to small regions obtained from an over-segmentation. Gould et al. [19] proposed a CRF to learn relative location offsets of categories based on superpixels. Fulkerson et al. [12] developed a classifier using histograms of local features based on superpixels. Multiple superpiexel segmentations are often exploited to assist classification. In [36], Russell et al. used multiple segmentations to build a BOF model to discover and label object categories automatically. Similarly, Galleguillos et al. [21] localized objects using the multiple-instance learning in the weakly labeled data. Pantofaru et al. [33] integrated multiple

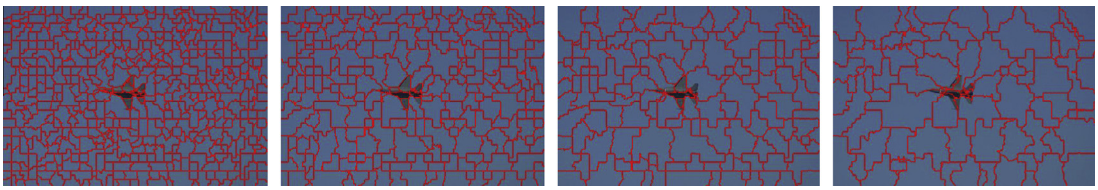

Fig. 1. Four segmentations using the quick shift algorithm with the scale parameter σ = 2, 4, 6 and 8, respectively.