Information Sciences 281 (2014) 661–673

Contents lists available at ScienceDirect

Information Sciences

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

Learning the object location, scale and view for image categorization with adapted classifier

Shengye Yan [a,*], Xinxing Xu [b], Qingshan Liu [a]

[a] Department of Automation, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China
[b] School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 19 January 2013
Received in revised form 17 February 2014
Accepted 9 March 2014
Available online 21 March 2014

Keywords:
Image categorization
SVM
Adaptive learning
Multiple instance learning
Multiple kernel learning

ABSTRACT

In this paper, we first propose a scheme to learn with automatically aligned objects for image categorization. To fulfil this purpose, the learning is formulated into a multiple instance learning problem with bags. Each bag is a virtually generated image set from an original image. The virtually generated image set covers multiple possible object locations, scales and views. Secondly, we propose a novel adaptive learning method to take advantage of the sharing information among different categories in image categorization. The method models the sharing information between one category and the other categories by learning a classifier of the current category from perturbed pre-learned classifiers from other categories. And the adaptive learning is realized in the multiple kernel learning framework. Finally, to align the images from one category and share the information among different categories simultaneously, the proposed multiple instance learning scheme and the proposed adaptive learning method are integrated into one learning framework. A new formulation called Adaptive Multiple Instance Learning (A-MIL) is derived. The optimization method is provided. To evaluate the proposed method, comprehensive experiments are conducted on two well-known public image categorization datasets (Caltech101 and 15Scenes). The experimental results show that the proposed method outperforms the state-of-the-arts in classification accuracy.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Image categorization aims to classify an input image to a specific category (e.g., city, mountain, street). It has extensive applications such as family photo organization, content based web search engines, and automatic robot pilot, especially with the growing popularity of digital photography that led to a tremendous number of images collected in personal photo albums and online image repositories. Despite the great progresses made in the past decades, the classification performance still cannot fulfill many real-world applications and there are many unresolved problems.

One important unresolved problem in image categorization is that the objects in different images are always aligned badly. Fig. 1 shows three typical images with the 'octopus' category in object recognition. One can find that the octopuses in these images are with very different locations, scales an views. Similar problem exists in other image categorization tasks such as scene classification, where the objects are present in some loosely constrained manner [35].

* Corresponding author. Tel.: +86 15358126429; fax: +86 25 57792648.
E-mail address: shengye.yan@gmail.com (S. Yan).

http://dx.doi.org/10.1016/j.ins.2014.03.041
0020-0255/© 2014 Elsevier Inc. All rights reserved.

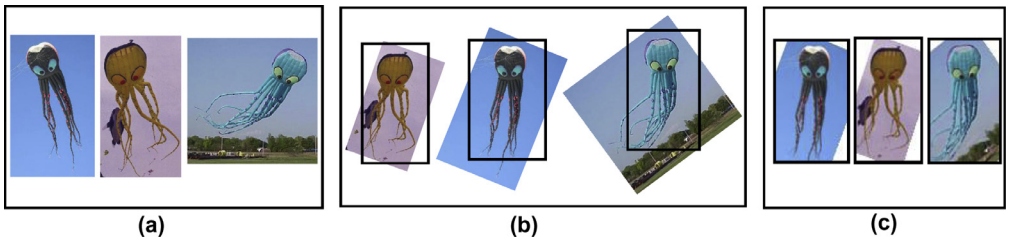

Fig. 1. The aligned training images generated from the original images. (a) The original training images; (b) The resized and rotated version of the original images; and (c) The finally generated training images.

In general understanding, badly aligned objects in the images cause large intra-class variation and always largely decrease the classification performance. This is widely acknowledged in specific object detection such as face detection [18,53] and pedestrian detection [49]. In these classification tasks, very high classification performances have been achieved through well aligning of the objects in the training images. And in contrast, it has also been proved that the classification performance will decrease largely if the objects in the training images are aligned badly.

The successes in specific object detection attributed to the well alignment motivate us to align the multiple object locations, scales and views in image categorization. In this paper, we propose a novel classification scheme to automatically learn the alignment information of the objects for image categorization. In the proposed method, to model the multiple unaligned object locations, scales and views in different training images, a virtual image set is generated for each original training image. The virtual image set is composed of the image patches cropped out from the in-plane rotated and resized images of the original image at different locations. Then the best image in the image set is learned to build the classifier. This learning procedure is naturally formulated into a multiple instance learning problem (MIL), which learns from the image set (bag) instead of a singleton [2]. In MIL, for the positive bag, only one or some samples are considered to be the positive training images. But for the negative bag, all the images are considered as negative training images. The virtual image set generated from each original image forms a bag which covers multiple possible object locations, scales and views.

Besides, another beneficial information for the multi-class classification in image categorization is the sharing information between different categories. Intuitively, in the classification of the similar concepts, if some kind of information plays an important role in one of these categories' classification, it is very likely to be beneficial for the other similar categories' classification. For example, the same background of "Water" and similar components like beaks existing in classes like "Swan" and "Duck" may be used to help the classification of the similar class like "Goose". Unfortunately, this kind of information is neglected in the traditional learning methods for image categorization.

To model the sharing information, motivated by the adaptive learning method in domain adaptation [12] where classifiers from the source domains are used to help the learning of the classifiers in the target domains by adapting pre-learned classifiers from other domains, we propose to adapt pre-learned classifiers from different categories in image categorization for information sharing. To summary, the adaptive learning is in two rounds. In the first round, the common classifiers are learned for each category independently. And these classifiers are named as pre-learned classifiers. Then in the second round, the adaptive classifier for each category is learned based on all the pre-learned classifiers. To be exact, the adaptive classifier of current category is learned by perturbing the pre-learned classifiers from other categories.

On one hand, the adaptive learning can be used to enhance the classification model. On the other hand, it can be used to boost the multiple instance learning for better alignment. This can then further boost the classification performance. Incorporating the adaptive learning into the multiple instance learning framework, we derived a new formulation called Adaptive Multiple Instance Learning (A-MIL). A-MIL learns the aligned object position, scale and view and adapt the classifiers among different categories simultaneously to build a robust classifier.

To evaluate the proposed method, extensive experiments are conducted on two widely used benchmarks for object categorization and scene categorization, i.e. Caltech101 [27] and 15Scenes [37,26,24]. The detailed experimental results and comparisons are provided. The experimental results demonstrate the effectiveness of the proposed method. And the comparisons reveal that the proposed method outperforms the state-of-the-arts in classification accuracy.

The remainder of the paper is organized as follows. We first review the previous works in Section 2. We then describe the procedure to generate virtual images in Section 3. In Section 4, we derive the proposed A-MIL. A-MIL is evaluated in Section 5. Finally, the paper concludes in Section 6.

2. Related work

From the viewpoint of *image categorization*, the most popular method is Bag of Words (BoW) [40,10] based Spatial Pyramid Matching (SPM) [24] feature extraction framework with traditional SVM for learning. BoW treats an image as a collection of unordered appearance descriptors extracted from local patches, quantizes them into discrete visual words, and then computes a compact histogram representation for semantic image classification, e.g. object recognition or scene categorization. SPM partitions an image into $2l \times 2l$ segments in different scales l = 0; 1; 2, computes the BoW histogram within each of

Download English Version:

https://daneshyari.com/en/article/393531

Download Persian Version:

https://daneshyari.com/article/393531

Daneshyari.com