Neurocomputing 120 (2013) 203–213

Contents lists available at ScienceDirect

Neurocomputing

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

Superpixel level object recognition under local learning framework

Huchuan Lu *, Xuejiao Feng, Xiaohui Li, Lihe Zhang

School of Information and Communication Engineering, Dalian University of Technology, China

ARTICLE INFO

Article history:
Received 25 December 2011
Received in revised form
31 July 2012
Accepted 23 August 2012
Available online 26 March 2013

Keywords:
Object recognition
Superpixels
Local learning
Neighbor integration

ABSTRACT

In this paper, we propose a simple yet effective method for superpixel level object recognition on the bag-of-feature framework. Instead of using general classifiers for the superpixel categorization, we introduce local learning classifiers into our framework, which aims to turn a highly non-linear classification problem into multiple local linear problems within different subsets of the database, so as to tackle the intraclass variation problem brought by superpixel based representations of objects. In addition, context information is used to make better performance by combining each superpixel with its appearance-based superpixel neighbors within a certain neighborhood distance from superpixel mean color map. At last, we utilize superpixel based Graph Cuts algorithm to segment the objects from background image. We test the proposed method on Graz-02 dataset, and get results comparable to the state-of-the-art.

1. Introduction

As an important issue in computer vision and pattern recognition, object recognition has a wide range of applications in many areas, such as image matching [1–5], object classification [6–8] and image retrieval [9–12]. The recent years have witnessed significant advances in object recognition, but it remains a challenging problem.

In the past few years, sliding window [13,14] structure is acting as the most popular approach in object detection, localization, categorization and recognition, due to its simplicity in sampling and effectiveness in discrimination. By scanning the whole image with a bounding box, sliding window object recognition methods assign each image patch partitioned by the bounding box a confidence score that how much the patch contains a certain object category, and then find out the patches with the highest scores as the recognition result. However, sliding window approach also suffers from several disadvantages. On one hand, sampling of a large amount of bounding boxes at multiple locations and scales is time consuming. On the other hand, sliding window tends to obtain a coarse possible location for an entire object which will perform bad for non-rigid and non-canonical posed objects, since the "window" bounding the object may also contain much background area which may disturb the evaluation.

Pixel level categorization is another recently prevalent method in object recognition, whose aim is to directly classify individual pixels into different classifications, based on the prediction results of the patches enclosing them [15,16]. While these methods can consider more detailed information and provide more precise boundaries of objects, they have to deal with the computation and redundancy problems because of their rather dense nature.

To compromise these two kind of approaches, many researchers consider "superpixels" [17–19], small regions obtained by over segmentation, as the elementary unit for object categorization and recognition. Superpixels are qualified to be the categorization unit, for they usually preserve boundaries information, which makes them part of the object or part of the background, rather than stretch over both of the two. However, the choice of the superpixels tends to be a trouble. Russell et al. [20] and Galleguillos et al. [21] produce their work based on the assumption that there exists at least one segment from all their segmentations which contains the entire object. This kind of idea can be considered as segment based recognition, which do a variety of segmentations first, and then use classification methods to find segments that most probably contain the objects. Hence, these approaches depend too much on the reliability of the segmentation algorithms. If the segmentation process fails to provide "good" segments, the following recognition methods will consequently be affected. Practically, none of the existing segmentation methods could make sure to segment all objects from their background due to both the various appearance of the objects and the clustering context of the backgrounds.

There are mainly two ways to tackle this problem. For one thing, Pantofaru et al. [22] and Li et al. [23] relax the "entire object segment" requirement by intersecting multiple segmentations and classifying these intersections by averaging the classification results of the segments enclosing them, but multiple segmentations are computationally expensive in practice. For another, in

* Corresponding author. Tel.: +86 13 940 860 809.
E-mail address: lhchuan@dlut.edu.cn (H. Lu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.08.058

contrast to all the approaches mentioned above, Fulkerson et al. [24] generate fairly small superpixels as the elementary units of the object recognition method by using only one segmentation algorithm.

In our framework, we utilize this small superpixel idea, to make sure each superpixel segment belongs to the area of only one category (either the object area or the background area), not stretches over both of them. Then, we can assign each superpixel a clear class label, and integrate all the superpixel area with the same label of a certain object category to implement the object recognition task. However, since the object in an image might be very small, the size of superpixel should be set small enough (in our case, about 150 pixels per superpixel in average). Therefore, using these small superpixels to be the elementary unit for classification suffers many disadvantages, including intraclass variation problem and weak feature representation problem.

Intraclass variation is a challenging and difficult problem in image classification and object recognition, since objects from the same categories may be very different in appearance, shape, color, scale, etc. Local learning [10,11,25,26] is an effective way of tackling intraclass variations, for it creates multiple local models, each of which is trained on a subset of data. Hence, local models usually possess more discriminative powers for a subset of data than a global one for the whole data, thus can achieve better performance when the data is complicatedly distributed. Usually being parts of an object, superpixels with the same class labels are doom to suffer large variations, because even parts from the same object may be completely different from each other, as illustrated in Fig. 1. Therefore, the classification issue will be much too difficult if we only use one global classifier, while local learning can give a better performance if we treat the diversity of superpixels in the same class as intraclass variations.

How to build up the local models is a key problem for local learning. Kim and Kittler [27] divide the dataset into subsets by k-means clustering first, and then train a classifier with each subset respectively. Dai et al. [6] partition the dataset by considering the prior distribution of the data, and assign more local models around the data points which are more likely to be misclassified. However, since it is generally a difficult task to determine the number of the local model and estimate the distribution of the dataset, many researchers seek to design local models specifically for each training sample [10,11,28,29]. In our work, we consider superpixel as the elementary unit. However building a local model for each superpixel is redundant and computationally expensive. Therefore, in our experiments, we randomly select a number of superpixels from the training data set as center superpixels, and create local models for each of these center superpixels. By doing this, we reasonably reduce the amount of the local models with slight decrease of performance.

Another problem of superpixel classification is that superpixels are often too small to contain enough features to be correctly classified. Even a pair of neighboring superpixels who are similar and belonging to the same object might be classified very differently. Therefore, researchers seek to use neighborhood information to enhance the decision. Fulkerson et al. [24] use the context information from the neighbor superpixels to increase the discriminative power of each small superpixels, and then refine their results by using conditional random field. Actually, Markov Random Field (MRF) [7] and Conditional Random Field (CRF) [16] or some other graph based algorithms have been widely used in pixel level recognition and segmentation methods to pursue spatial consistency. In our paper, we propose a new scheme to integrate superpixel neighbors, and obtain the final object segment by Graph Cuts [30]. The flowchart of our framework is shown in Fig. 2.

The main contributions of this paper include:

- In this paper, we introduce local learning classifiers into our superpixel categorization framework, so as to tackle the intraclass variation problem brought by superpixel based representations of objects.
- Context information is used to make better performance by combining each superpixel with its appearance-based superpixel neighbors within a certain neighborhood distance from superpixel mean color map, which is different from the spatially adjacent superpixel neighbors in [24].
- At last, we utilize superpixel based Graph Cuts algorithm to segment the objects from background image and achieve good segmentation results.

The rest of this paper is organized as follows. Section 2 introduces our local learning framework for superpixel classification, while in Section 3, we represent superpixel neighbor integration and the segmentation of the object. Experimental results with comparisons are demonstrated in Section 4. Finally, Section 5 makes a conclusion of the paper.

2. Superpixel classification by local learning

Like [24], our framework works on superpixels generated by a single unsupervised segmentation method, and aims to give each superpixel a class label so as to implement both object recognition and localization at the same time. To create a more discriminative classification framework for these small superpixels, we treat the phenomenon that various segments may share the same label as an intraclass variation problem, and propose a simple yet efficient local learning method.

2.1. Superpixel generation

Utilizing the small superpixel as the elementary unit of the classification task benefits from many aspects. First, the construction

Fig. 1. (a), (b) and (c) are examples of different objects to be partitioned into superpixels. It is obvious that treating superpixels as classification units will bring in intraclass variation since even superpixels sampled from the same object appear quite different to each other. (a) Car. (b) Bike. (c) Person.