A Shape Context Based Car Detection with Hypothesis Pruning

Sefa Kılıç, İşıl Doğa Yakut, Cansın Yıldız
Dept. of Computer Engineering
Bilkent University
Ankara, Turkey
{sefak, yakut, cansin}@cs.bilkent.edu.tr

Abstract—In this paper, a car detection methodology using shape context (SC) feature is described. The approach is adopted from a study for human detection using a generic object detection algorithm[1]. In this first part we adjusted the parameters of the introduced method in order to work with car objects. Then using the SC features each training image is assigned with several hypothesis areas. In addition, we implemented and experimented with a hypothesis verification scheme which prunes then unifies the hypothesis using empirical values as threshold for pruning and overlap ratio for merging. After the car detection phase is over model and manufacturer recognition task is studied. In particular, a method where morphological operation to detect the location of the logo was experimented. Also a dataset of clean images for supervised object recognition is provided with 750 multi-view car images.

Keywords-component; car detection, shape context, top-down object recognition.

I. INTRODUCTION

Object detection and recognition is a key topic where many research areas; robotics, navigation, surveillance benefit from. But it is a difficult task to accomplish. By specifying the object in question as car/vehicle relaxes the the problem at hand but car detection is still far from a simple task. The main reason behind this is that the object's form can vary greatly. Thus an object-detector must handle large amount of variation. For example, cars can vary by their shape, color, size, tires, headlights, etc. Similarly an object with similar characteristics should be distinguished by the detector. In addition there is occlusion and illumination conditions which makes the task even harder.

The following sections are ordered as follows: Section II gives a brief literature survey on object and car recognition. Section III, IV and V the methodology of existing object detection and introduced extension to it, is described. In section Section VI, generated dataset, results and a discussion of the results in given. A further study on logo detection is given in Section VII. Lastly, the paper is concluded in Section VIII.

II. RELATED WORK

Car detection is a very applicable area for real world needs such as surveillance. Therefore car detection has been the focus of many studies and new technologies are constantly being developed and existing technologies improved. Thus numerous approaches has been developed over the years. These methodologies can be categorized as context-based, motion-based and stereo-based[5][6]. Stereo-based approaches use stereo-vision to detect the vehicle. One study [5] that stand-out tries to sense the vehicles near-by for automatic vehicle driving. Motion-based methods detect vehicles and obstacles using optical flow [7]. But experiments show that this is a time-consuming and impractical for real-time applications. Both of stereo-based and motion-based methods are researched widely for real-world issues and application such as auto-driving, driving-assistance applications.

Context-based car detection systems use the knowledge on the objects characteristics such as shape and color. For example, the prior knowledge that vehicles are symmetric about the vertical axis has been used in vehicle detection approaches using the intensity or edge map in [8][9]. Shape context features [11] are also widely used for car recognition and detection [1]. In addition for hypothesis verification, bag of features are created for each image. Then these are learned by a training classifier or by modelling probability distribution [10][11].

III. OVERVIEW OF THE METHOD

We adopted the top-down object recognition methodology, introduced in [1]. The method is originally for recognition of generic objects. The study of the paper [1] focuses on human detection. For our purpose, we revised the parameters and adjust them accordingly.

The proposed method [1] has three parts, namely; codebook building, top-down recognition using matching and voting, and hypothesis verification. In our implementation, verification task is excluded due to time constraints. Instead, we propose a simpler hypothesis pruning method, where false positives are cancelled out based on an empirical threshold value. The remaining true positives are unified according to their overlap ratios.

IV. TOP-DOWN RECOGNITION

Top-down recognition part is divided into three tasks, which are codebook building, improved shape context and hypothesis generation. The first task is to train the top-down
recognizer, by building a codebook of shape context features. Then, given a test image, its shape context features are extracted and compared to each codebook entry. Using a voting scheme, these comparisons generate some hypotheses results for the test image.

A. Codebook Building

At this stage, using a set of images and their masks, a training is performed where each feature vector correspond to a codebook entry. The resultant codebook is a set of example features, which is used to compare with a given test image.

The theory behind codebook entries is as follows. Each entry \(cev = (u_i, \delta, m, w_i)\) represents the feature vector of a point \(p_i\) in objects train images. \(\delta\) is the position of point \(p_i\) with respect to object center. \(m\) is a binary mask for the patch centered at point \(p_i\). \(w_i\) is the weight mask computed on the binary mask. And lastly, \(u_i\) is the shape context vector of the point \(p_i\).

B. Improved Shape Context

As noted in [1], the idea of Shape Context features is first proposed by Belongie et al. [2]. Shape Context is a way to describe shapes to allow shape similarity measuring and point correspondence matching.

Basically, \(n\) points are picked on the edges of a shape. Then, for each point \(p_i\), \(n-1\) vectors are created by connecting \(p_i\) to all other points. The set of all vectors that are created by this method is a fine representation of local shape feature of that point, but it is rather too detailed. So, a coarse histogram is constructed for that \(n-1\) vectors using simple binning.

\[h_k = \# \{q \neq p_i; (q - p_i) \in bin(k)\}\] (1)

The fundamental Shape Context representation is further improved by Zhu et al. [1]. They proposed two new extensions two shape context; angular blur and mask function on shape context.

1) Angular Blur

As can be seen from Fig. 2(c), even similar edge regions can have very differing histograms, when dense bins are used. So, to overcome this problem, the angular bins can be overlapped with their neighbors, which results in more similar histograms Fig. 2(e).

2) Mask Function on Shape Context

Another important extension to shape context is to introduce masking for train images. In traditional approaches, a train image has some noise caused by background textures. To get rid of those clutters, Zhu et al. proposed using a binary mask to hide background. We further extended this mask function, to create masks from clean train images automatically.

C. Hypothesis Generation

Hypothesis generation is the process of predicting object locations in an image. The proposed generation by Zhu et al., uses a voting technique similar to [3]. In a test image, each shape context feature is compared by the each entry of the codebook, and a decision is made on possible object center. Then, the matching scores are accumulated over the whole
image. The predictions having maximum scores are selected as possible object centers.

V. HYPOTHESIS PRUNING

Top-down recognition method explained in section IV, produces high-recall/low-precision results. Therefore, a further step to increase the precision rate is needed. In Zhu et al.’s paper [1], they already proposed a false positive pruning method. But, due to time constraints, we propose a simpler hypothesis pruning method, where false positives are cancelled out based on an empirical threshold value, and remaining positive are unified according to their overlap ratio.

A. False Positive Elimination

The proposed method produces some number of hypotheses, each having an assigned score value. A higher score value means, that hypothesis is more likely to be an actual object, that is a car in our case. So, using this score values, we determine a threshold value empirically. At experiment step, we run ~40 test images, each having one or more cars. Then, we estimated a threshold value which tries to minimize number of false positives, while keeping positive hypothesis. Fig. 3(b) shows a sample output of our false positive elimination method.

B. Unification of Related Hypothesises

After false positives are pruned from the hypothesis set, we further improve the hypotheses by unifying related ones. By calculating the overlapping area ratios between each pairs of hypotheses, we select the ones having a ratio higher than a threshold, and combine these. Similar in V.A, this threshold is estimated in an empirical manner. The motivation behind this approach is that, the hypothesis having a high overlap area, are most likely correspond to same object. Fig. 3(c) shows a sample output of our final pipeline including hypothesis unification.

VI. RESULTS

In this section, some results of our implementation are demonstrated. First brief information about the data environment is given. Then, section concluded with showing some results and discussions on them.

A. Experiment Setup

In order to have a clean training dataset, we have collected over 400 images from [4]. These photos are taken from 5 different view points, and by taking horizontal flips of related images, we ended up 8 different poses, a total of 720 images. These photos are categorized in three different shape models, namely; hatchback, notchback and station-wagon. Further categorization is performed according to their brands, a total of 8 brands are collected. In authors’ knowledge, there is no such dataset which consists of clean car images of same view angles. So, we believe this dataset could be helpful for future studies on car shape and brand recognition. Fig. 4 shows a sample set of images from this clean dataset.

B. Results and Discussion

Here, we will show some results from our work. In Fig. 3, a successful run can already be seen. So, rather than showing other successful results, this section will focus on the problematic ones and discuss their reasons.

Fig. 6 shows a case where false positive elimination loses some important data. Although false positive elimination successfully eliminates all false positive almost all the time, there are some cases where a true positive is eliminated as well. In this figure, if false positive elimination is not performed, we would end up with a larger bounding box which encapsulates whole car, rather than this small one.

This clean dataset is also helped us to generate masks that are needed for train images in an automated fashion, unlike Zhu et al., which manually masked the pedestrian images. We are simply inverting the color values of train images, and using a threshold, a final binary mask is generated. Fig. 5 shows some sample masks for train images.
In Fig. 7, it can be seen that, our pruning algorithm successfully pruned lots of false positives. But, this time the unification method that we propose removed a valid data. The green box in the left figure is mistakenly unified with the larger red box. This is due to our unification assumption. As explained at V.B, we assumed if two hypotheses are overlapping an area more than a threshold, they are trying to annotate same object. But, as in this case, if two different objects are adjacent to each other in the scene, this assumption can sometimes fail.

VII. FURTHER WORK: LOGO RECOGNITION

Our next study is trying to detect and recognize the logo of a car. The detection step is based on the study in 0 which proposes a method for detection of the vehicle logo based on different morphological operations. The idea is that no matter of what color the logo is, the characters on the vehicle logo are usually bright colored.

We tried to implement an logo detector by using morphological operations. By applying histogram equalization, top/bottom hat transformations for contrast enhancement and morphological opening, we tried to extract the area which has logo from the entire image. When applied to images in the study 0, we found similar results. Although in the paper 0, it is stated that results have high accuracy rate, our results were not promising as such. We believe that the reason of the not high accuracy results is the assumption that car logos are usually bright colored. It is not always the case and also different illumination conditions of the entire image effects results.

The next step of the logo recognition. We tried SIFT for matching the logo, but we found that SIFT is not successful when it tries to match different but similar images. Even when the output of the detection step is successful, SIFT did not perform well to find the logo of the car.

VIII. CONCLUSION

In this study, we have two contributions. The first one is the clean dataset consisting of car images recorded in simple environment. The second one is the extension of pruning of hypothesis to the method mentioned. Results of the method on some images can be seen in Fig. 8. In addition to dataset and extension to the method, we also tried car logo recognition using morphological operators and SIFT descriptors. By using these methods, our results were not promising.

REFERENCES


(a) Before Hypothesis Pruning  (b) After Hypothesis Pruning  (a) Before Hypothesis Pruning  (b) After Hypothesis Pruning

**Fig. 8.** Some final results of our work.