Vehicle Detection based on the Deformable Hybrid Image Template

Ye Li*†, Bo Li*†, Bin Tian*†, Fenghua Zhu*‡, Gang Xiong*‡, Kunfeng Wang*‡
*The State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
†Beijing Engineering Research Center of Intelligent Systems and Technology,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
‡Dongguan Research Institute of CASIA,
Cloud Computing Center, Chinese Academy of Sciences,
Dongguan, China

***Abstract*—In complex urban traffic conditions, the accurate detection of vehicles is challenging to current vehicle detection methods. To achieve the precise vehicle detection in complex urban traffic conditions, we have proposed a vehicle detection method based on a deformable hybrid image template in this paper. Our method contains two steps: constructing our hybrid image template and its probability model, and detecting vehicles from traffic images by using a three-stage SUM-MAX procedure. In the template construction step, small image patches constituting the hybrid image template are automatically learned from training images. These image patches have various features and the feature types include sketch, texture, flatness, and color. Furthermore, a probability model is proposed to assist the vehicle detection step. After the template construction, a three-stage SUM-MAX procedure is applied to achieve vehicle detection with local deformations in locations and orientations. There are two innovations in our method which are the applications of the hybrid image template and the three-stage SUM-MAX procedure in vehicle detection under complex urban traffic conditions. To evaluate our method, we have done a quantitative and contrast experiment and the experiment on complex urban traffic conditions. The experimental results show that our method can effectively cope with various vehicle poses, vehicle shapes, time-of-day and weather conditions. In particular, our method has good performance in complex urban traffic conditions.**

***Index Terms*—Vehicle detection, hybrid image template, probability model**

I. INTRODUCTION

Vehicle detection techniques play an important role in intelligent transportation systems and provide vehicle information for the design of artificial transportation systems [1]. Many vehicle detection methods have been studied [2][3][4][5][6]. Ref. [2] and [3] applied motion information to detect vehicles. But these approaches are not suitable for dealing with the slow-moving traffic. Cheon *et al.* [4] used shadow regions under vehicles to detect vehicle objects. But the shadow regions are easily occluded by other vehicles in complex urban traffic conditions, so this method is also not suitable for complex urban traffic conditions. Ref. [5] proposed a deformable 3D vehicle model for describing the shapes and appearances of vehicles. The experimental results showed that this method was powerful when detecting or tracking one vehicle in the traffic images. However, when it comes to the complex urban traffic conditions, the authors haven't discussed whether or not their method is still valid. Wu *et al.* [6] provided a deformable model to describe the shape of the vehicle object. This model consists of multiple Gabor wavelet elements with various locations and orientations and these elements are allowed to perturb their locations and orientations for achieving vehicle deformation.

Currently, a new method of object representation and detection is proposed to extract the intrinsic characteristics of object categories by integrating various types of features and it is called hybrid image template (HIT) [7]. The template consists of multiple image patches with various types of features including the sketch, texture, flatness, and color. Moreover, each image patch in the template is allowed to perturb locally its location and orientation, which makes the template adapt in a local deformation. Therefore, it will be a good idea to apply it to detect vehicles in complex urban traffic conditions.

In this paper, we have constructed our HIT to represent vehicles in complex urban traffic conditions. Based on the constructed HIT, we use the three-stage SUM-MAX procedure to detect vehicles from traffic images. Furthermore, a series of experiments are done at various vehicle poses, vehicle shapes, weather, time-of-day conditions and complex urban traffic conditions. The experimental results show that our method can effectively deal with complex urban traffic conditions. There are two main innovations in this paper which include the application of HIT in vehicle representation under complex urban traffic conditions and the usage of a three-stage SUM-MAX procedure in vehicle detection with the local deformation. Notably, in this paper we mainly focus on the rear-view vehicle which is one of the most common vehicle pose in the urban traffic video surveillance systems [8][9].

This paper is organized as follows: Section II presents the basic theory of HIT. In Section III, the implementation of our method is described in detail. Section IV presents two experiments including a quantitative and contrast experiment and a experiment on complex urban traffic conditions. Finally, Section V makes a conclusion and introduces an outlook of the future work.

978-1-4799-0380-1/13/$31.00 ©2013 IEEE

Fig. 1. A part of training images used in our method.

II. BASIC THEORY

The hybrid image template proposed by Si *et al.* is composed of, typically, 50-500 image patches with various features [7]. Let Λ be the image lattice corresponding to a hybrid image template. The template is denoted as:

$$HIT = (\{\Lambda_i, t_i, \{B_i \ or \ H_i\}, \alpha_i, i = 1, 2, ..., N\}, P) \tag{1}$$

where $\Lambda_i \subset \Lambda$ represents the k-th patch lattice, $t_i \in \{sketch, texture, flatness, color\}$ shows the type of the patch, α_i denotes the latent variables for the local perturbations in locations and orientations, P is the probability model which will be introduced in the next subsection, and N is the number of the patches. $B_i \ or \ H_i$ presents the feature prototype of the patch.

In the equation 1, if $t_i = sketch$, the patch describes the image sketch and is modeled by a Gabor wavelet element B_i. If $t_i = texture$, the patch illustrates the local image texture by using a histogram H_i of Gabor filter responses with various orientations over the patch. If $t_i = flatness$, the patch represents an image region with void of structures by H_i summing the responses of Gabor filters with various orientations over the patch. If $t_i = color$, the patch illustrates an image region with color consistency among vehicles by pooling a histogram H_i on the color space (HSV space used in this paper) of the patch.

A probability model of the hybrid image template is defined by Si *et al.* for assisting vehicle detection as:

$$p(I|HIT) = q(I) \prod_{i=1}^{N} \frac{\exp\{\lambda_i r(I_{\Lambda_i})\}}{Z_i}, \tag{2}$$

in which I is a testing image, I_{Λ_i} denotes a local image region matching Λ_i in I, $q(I)$ is a reference distribution, λ_i is the coefficient corresponding to Λ_i, $r(I_{\Lambda_i})$ measures the distance between I_{Λ_i} and Λ_i, and Z_i is a normalization constant. The detailed introduction for the probability model and the hybrid image template can be found in [7]. In the following sections, we will describe our method in detail.

III. METHOD

Our method includes two steps: the construction of our hybrid image template for vehicle object and vehicle detection using a three-stage SUM-MAX procedure. We will introduce the two steps sequentially in the following subsections.

A. Construction of Our Hybrid Image Template

The hybrid image template consists of multiple image patches with various features and these image patches are learned from training examples sequentially. As the learning algorithm proposed by Si requires a small number of training examples, 27 training images are selected and manually aligned which are partly shown in Fig. 1. These images show the rear-view vehicles which are mainly studied in this paper. Note that in this paper the bus and truck are not considered due to the large deformation comparing to other vehicle types.

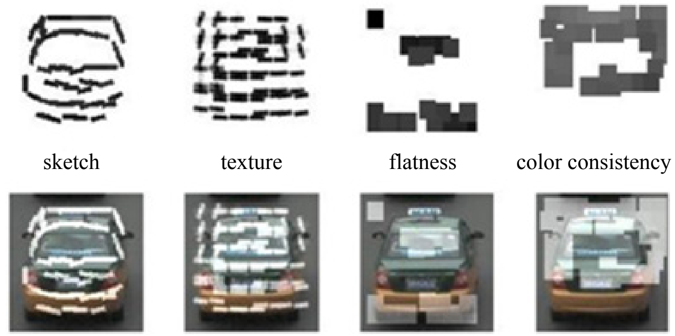

Fig. 2. Our hybrid image template for representing the vehicle object.

Based on the training images, we have learned the hybrid image template for representing the vehicle object. The learned image patches with four types of features are shown in Fig. 2 respectively. The learning algorithm is described in detail in Ref. [7]. As shown in Fig. 2, the learned sketch patches describe well the edges of the vehicle object, the flatness patches represent perfectly the features of the vehicle window, and the color patches also locate correctly the regions with consistent colors among vehicles. As each image patch corresponds to a weight which is λ_i in the equation 2, the intensity of the color shown in each patch of Fig. 2 shows the weight representing the significance of each patch in the template. The learned template in Fig. 2 shows that the learning algorithm extracts meaningful features for the rear-view vehicle object, which is helpful for the vehicle detection in complex urban traffic conditions. During the learning process, we also obtain a probability model of the template which is denoted as the Eq. 2. In order to apply the HIT on vehicle detection, we use the log likelihood ratio in the probability model calculated from Eq. 2 as the template matching score:

$$score(I|HIT) = \log \frac{p(I|HIT)}{q(I)} = \sum_{i=1}^{N} \lambda_i r(I_{\Lambda_i}) - \log Z_i \tag{3}$$

By using Eq. 3, we compute the score of each training example so as to obtain the statistical value of the real vehicle's score. This statistical value will serve as the detection threshold for vehicle detection on testing images. At this point the construction of the hybrid image template of the rear-view vehicle object is finished.

B. Vehicle Detection Using a Three-stage SUM-MAX Procedure

In this paper, our goal is to apply the constructed hybrid image template to pursue vehicle objects in an input traffic

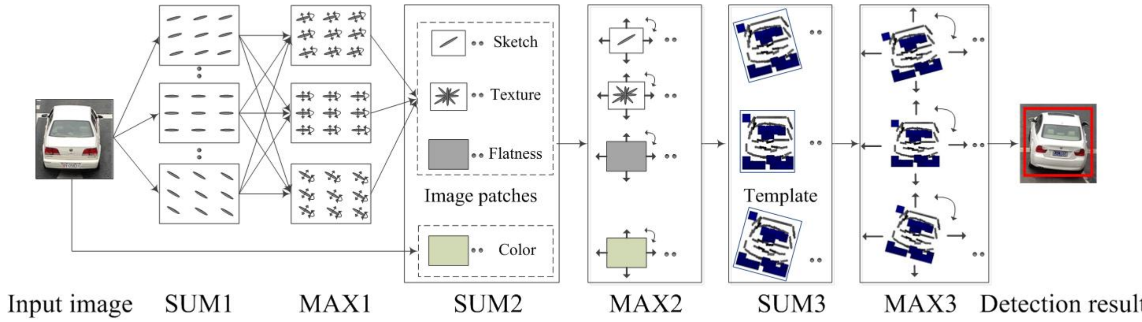

Fig. 3. The three-stage SUM-MAX procedure for detecting vehicles from the traffic image.

image. In consideration of multiple vehicles in an image, we achieve the detection of these vehicles by utilizing an iteration process. In each step, the local image region with the maximal score computed from the equation 3 is selected as a vehicle candidate. If this score is more than the learned threshold introduced above, this candidate is viewed as a detected vehicle object. After finding a vehicle, the corresponding image region will be removed from the traffic image for pursuing other vehicles. The iteration process stops until the score computed at the current step is less than the threshold. Notably, the size of a vehicle in a traffic image varies from the distance between the vehicle and the camera. Thus, we scale the input image for adapting in the deformations of vehicle sizes. The experimental results will show that the multi-scale approach is valid when dealing with different sized vehicles.

In the implementation, each iteration applies a three-stage SUM-MAX procedure to pursue a vehicle with the maximal score from the input image. In the following paragraphs, we will describe the SUM-MAX procedure in detail.

Fig. 3 illustrates the SUM-MAX procedure for the detection of a rear-view vehicle. The SUM-MAX procedure includes six steps: SUM1, MAX1, SUM2, MAX2, SUM3 and MAX3. In SUM1, Gabor filters with various orientations are applied on the input image for detecting edges with different directions. In Fig. 3, each rectangle of SUM1 represents a Gabor filter operation in one orientation. The MAX1 step applies a local maximum operation (LMO1) on the edge images outputted in SUM1 for preferably matching the edge segments in the input image. The LMO1 perturbs locally the Gabor wavelet elements at locations and orientations to pursue the local maximal filter response. The perturbation shown by arrows in MAX1 will achieve vehicle deformation. The MAX1 step outputs the modified edge images in all orientations. In SUM2, each image patch in our HIT and its tilted ones are used to filter the modified edge images or the input image for pursuing the patch. According to feature types of image patches, the pursuit of sketch, texture, and flatness patches applies the modified edge images, and that of the color patches applies color input image which is transformed into HSV color space. In MAX2 step, we also perform the local maximum operation (LMO2) on the output of the SUM2 step for adapting in more vehicle deformation. The LMO2 locally perturbs the

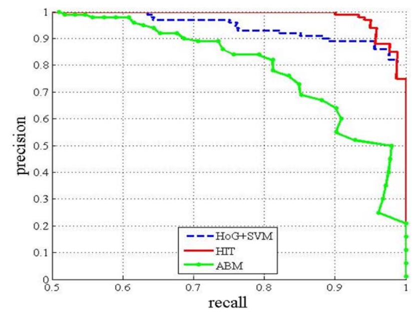

Fig. 4. The precision-recall curves for our method, HoG and ABM.

location and orientation of the detected patch to achieve a good match between the patch and the corresponding filtered image. In SUM3 step, we combine the detected patches by using their relative locations and orientations. Then the score is computed according to Eq. 3. In Fig. 3, for simplicity the hybrid image template is showed only with sketch and flatness patches. The SUM3 outputs multiple detected vehicle candidates. After the SUM3 step, the MAX3 step performs the local maximum operation (LMO3) that achieves the local perturbations of the detected vehicle candidates in locations and orientations for pursuing the local maximal score. Finally, the vehicle candidate with the maximal score in the input image is selected. If this score is more than the threshold, the vehicle candidate is viewed as a vehicle object.

IV. EXPERIMENTS

In this section we make experiments under various traffic conditions to verify the effectiveness of our method. These experiments contain a quantitative and contrast experiment, and the experiment on complex urban traffic conditions.

A. Quantitative and Contrast Experiment

We have tested the performance of our method on a set of testing images including various vehicle shapes, vehicle poses, time-of-day and weather conditions. The image set contains 100 positive and 100 negative examples which are collected from the real traffic conditions and the random background images respectively. In this experiment, a vehicle is detected correctly only when its bounding box overlaps 65 percent of the ground-truth. We apply our HIT to identify vehicles with corresponding scores from all training images and then threshold these scores at different points for computing the precision-recall curve of our method. For a particular threshold, the precision and recall are (true positive)/(true positive + false positive) and (true positive)/(true positive + false negative) respectively. This curve is shown in Fig. 4. Meanwhile, Fig. 5-7 show the vehicle detection results at various vehicle poses, vehicle shapes, time-of-day, and weather conditions. For clarity, the detected hybrid image templates are showed only with sketch, texture, and flatness patches in Fig. 5-7. In Fig. 5-7, the gray or black lines represent the sketch patches, the red lines indicate the texture patches, the rectangles with various colors show the flatness patches with different response values. The experimental results show that our method can effectively deal with various vehicle poses, vehicle shapes, time-of-day, and weather conditions.

For further verifying the performance of our method, we compare our HIT with two other methods including Histograms of Oriented Gradients (HOG) based on SVM and Active Basis Model (ABM). We learn ABM and HOG models on the same training images as ours. Furthermore, we use the learned ABM and HOG models to process the same testing set, and the corresponding precision-recall curves are showed in Fig. 4. The higher curve indicates higher precision rate and better performance. Therefore, in situation of the small size of the training image set, our method performs better than HOG and ABM.

B. Experiment on Complex Urban Traffic Conditions

In this paper, the complex urban traffic condition is mainly concerned because this condition is challenging to current vehicle detection methods. Therefore, we have done the experiment on complex urban traffic conditions. Notably, our method only studies the rear-view vehicles and utilizes the rear part of a vehicle to detect the vehicle. As shown in Fig. 8, our method detects correctly the rear parts of vehicles with various shapes and red bounding boxes show the detection results. Moreover, due to the capacity of adapting in vehicle deformation, our method locates correctly the vehicles with various poses and identifies roughly the poses shown by rotated bounding boxes in Fig. 8. The experimental results show that our method can effectively deal with the complex traffic conditions with various vehicle poses and shapes.

V. CONCLUSION

In this paper, we present a vehicle detection method based on the deformable hybrid image template for complex urban traffic conditions. The proposed method includes the construction of our HIT and vehicle detection based on a three-stage SUM-MAX procedure. In this paper, there are two main innovations. Firstly, we apply HIT to vehicle detection in complex urban traffic conditions. The experimental results show that the HIT preferably extracts meaningful features including sketch, texture, flatness, and color for vehicle object and improves the vehicle detection accuracy in complex urban traffic conditions. Secondly, the application of the three-stage

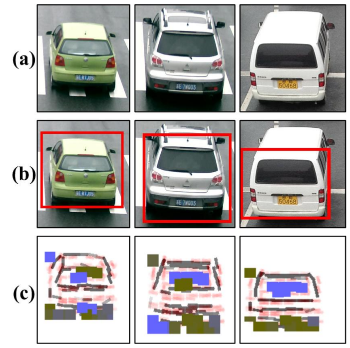

Fig. 5. The experimental results of vehicle detection with different vehicle shapes. (a) The input images. (b) The detection results. (c) The detected hybrid image templates.

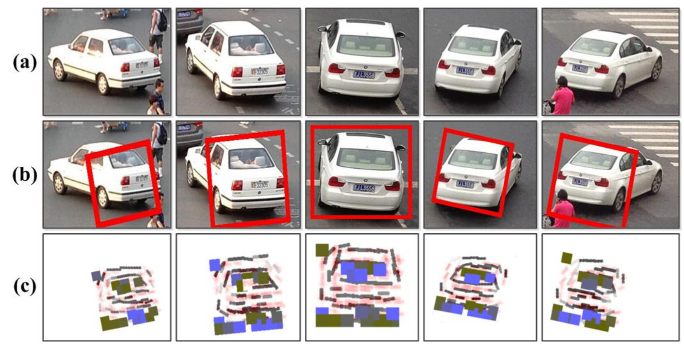

Fig. 6. The experimental results on vehicles with various vehicle poses. (a) The input images. (b) The detection results. (c) The detected hybrid image templates.

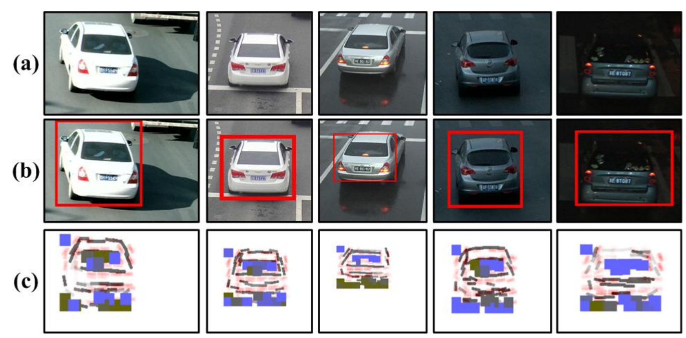

Fig. 7. The experimental results for different weather and time-of-day conditions. (a) The input images. (b) The detection results. (c) The detected hybrid image templates.

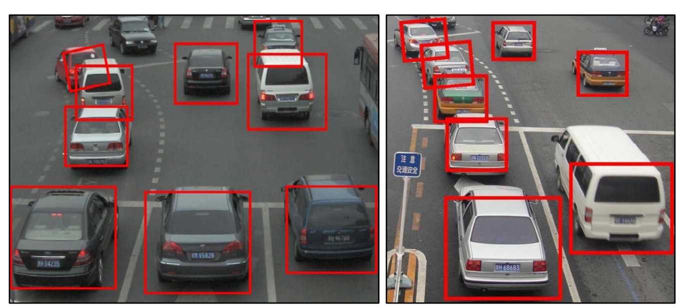

Fig. 8. The experimental results on complex urban traffic conditions.

SUM-MAX procedure makes our method adapt in slight vehicle deformation. A comprehensive test is performed in this paper and the results indicate that our method can deal with slight variance in vehicle pose and shape. Meanwhile, our method can also adapt in various weather conditions, and time-of-day conditions. In this paper, our method mainly concentrates on the rear-view vehicles. In the future, the front-view and side-view vehicle will be considered. Moreover, our method does not consider the truck and bus. It is necessary to expand our method to more vehicle types in the future work.

ACKNOWLEDGMENT

This work is supported in part by NSFC 71232006, 61233001, 91024030, and 90920305; and CAS 2F12E05 and 2F11N07. We would like to express our sincere appreciation to professor Fei-Yue Wang for his instruction and encouragement.

REFERENCES

[1] F.-Y. Wang, "Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.

[2] A. Faro, D. Giordano, and C. Spampinato, "Adaptive Background Modeling Integrated With Luminosity Sensors and Occlusion Processing for Reliable Vehicle Detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1398–1412, Dec. 2011.

[3] W. Zhang, Q. Wu, X. Yang, and X. Fang, "Multilevel Framework to Detect and Handle Vehicle Occlusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 161–174, Mar. 2008.

[4] M. Cheon, W. Lee, C. Yoon, and M. Park, "Vision-Based Vehicle Detection System With Consideration of the Detecting Location," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1243–1252, Sep. 2012.

[5] M. Leotta and J. Mundy, "Vehicle Surveillance with a Generic, Adaptive, 3D Vehicle Model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1457–1469, Jul. 2011.

[6] Y. N. Wu, Z. Si, H. Gong, and S.-C. Zhu, "Learning Active Basis Model for Object Detection and Recognition," *Int'l J. Computer Vision*, vol. 90, pp. 198–235, 2010.

[7] Z. Si and S.-C. Zhu, "Learning hybrid image templates (hit) by information projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1354–1367, 2012.

[8] J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu, "Automatic Traffic Surveillance System for Vehicle Tracking and Classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 2, pp. 175–187, Jun. 2006.

[9] N. Buch, S. Velastin, and J. Orwell, "A Review of Computer Vision Techniques for the Analysis of Urban Traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920–939, Sep. 2011.