Continuous Vehicle Detection and Tracking for Non-overlapping Multi-camera Surveillance System

Jinjia Peng, Tianyi Shen, Yafei Wang, Tongtong Zhao, Jun Zhang, Xianping Fu
School of Information Science and Technology, Dalian Maritime University
1 Linghai Road, Dalian, China
+86 411 84727731
fxp@dlmu.edu.cn

ABSTRACT

Vehicle detection and tracking has always been a significant research on traffic surveillance video. However, multi-camera object tracking consists of a non-overlapping video surveillance network, which makes vehicle re-identification a challenging problem. In this paper, we proposed a novel method for continuous vehicle detection and tracking in multi-camera campus surveillance videos. The method contains two main parts: One is auto vehicle detection and tracking by using background modeling combining with RCNN (Region Convolutional Neural Networks). The other one is multi-camera vehicle re-identification, which collaborates vehicle visual attributes and spatio-temporal information. The experiment results demonstrate that the proposed approach performs with high efficiency and accuracy, which can also be employed to optimize the trajectories of vehicles in multi-camera surveillance videos.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *object recognition, tracking.*

General Terms

Algorithms.

Keywords

Non-overlapping, multi-camera, vehicle detection, vehicle trajectory tracking

1. INTRODUCTION

Intelligent video surveillance for transportation has become a crucial measure to monitor the transportation situation. Multi-camera object tracking is one of the challenge and important technology of this area, which solves the difficulty of observing the complete trajectory of target vehicles by single camera with limited field of view. The intelligent video surveillance improves the monitoring efficiency effectively, and is widely used in many environments, such as traffic accident prediction, individual vehicle monitoring, crime prevention, etc.

Thus, the requirement of vehicle re-identification from large-scale surveillance image and video database in public security systems is growing. It is important to develop algorithms to track a vehicle through the road network of a city instead of human, which are also of high accuracy. To solve the problems in multi-camera object tracking, the method of recognition and tracking is divided into two main parts in this paper: single camera vehicle tracking and multi-camera vehicle tracking.

As shown in Figure 1, our task is to obtain the continuous trajectory of the target vehicle by connecting the trajectories obtained from surveillance videos. The red circles and arrows denote the locations and directions of cameras. The blue line illustrates an example trajectory of the same vehicle captured by the surveillance network. How to get the accurate trajectory in the single camera and how to search in the database of images that contain the same vehicle captured by multiple cameras are two main problems to solve.

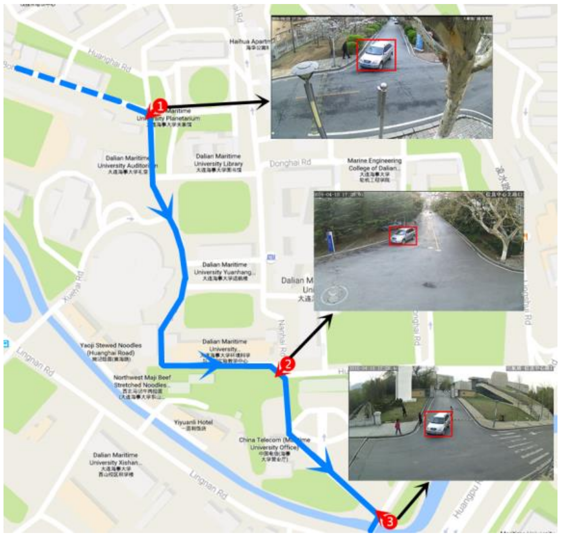

Figure 1. The task of vehicle tracking and cameras distribution in campus

In the previous research on single camera object tracking, traditional ways mostly used background modeling based on adaptive mixture Gaussian model to detect the moving object, but this will lead to coarse results because of the complicated environments. With the development of deep learning, RCNN has achieved excellent performance in many computer vision tasks. In order to achieve the result of continuous vehicle detection, we develop a new approach that integrates background modeling and RCNN together to recognize moving vehicles.

ICIMCS '16, August 19-21, 2016, Xian, China
2016 ACM. ISBN 978-1-4503-4850-8/16/08 $15.00
DOI: http://dx.doi.org/10.1145/3007669.3007705

This work is supported by National Natural Science Foundation of China (61370142, 61272368) and Teaching Reform Research Project of Liaoning, China (UPRP20140759).

Xianping Fu is the corresponding author.

However, it is challenging for vehicle detection and tracking in non-overlapping multi-camera surveillance system because the same target may have significant variations in shape and appearance across cameras. Targets may enter and exit a scene randomly with highly non-linear motions. Since vehicle has many attributes such as the diversity of brands, designs, and colors, many researches have focused on vehicle from different research interests. Feris et al. [1]proposed a system for vehicle detection and attribute-based, in which vehicles can be searched by colors, types or some other attributes in surveillance videos. Cao et al. [2] developed an approach that jointly optimize the single camera object tracking and multi-camera object tracking in an equalized global graphical model. This joint approach overcomes the disadvantages in traditional two-step tracking approaches. In [3], Bao et al. introduced a framework of object instance detection in multiple views by measuring the similarity of parts, which has more accurate detection results. Yang et al. [4]used Convolutional Neural Network (CNN) to extract visual features from images of vehicles for prediction, which inspired us to adopt CNN in multi-camera vehicle re-identification. However, the previous research of vehicle detection are usually focusing on the features of vehicle themselves. The spatio-temporal information of vehicle is also important to improve the precision of vehicle re-identification[5]. Besides, the positions of the objects could also be inferred based on a common ground assumption, which allowed the warping between the cameras' views by using a homography matrix.

In this paper, we combined background modeling and RCNN in single camera vehicles tracking, which performs with higher accuracy. For multi-camera vehicle tracking, we established a spatial topology relationship of camera networks for spatio-temporal information and matched the key frames of vehicles using CNN features, which has a great performance. Finally we measure the similarity of vehicle views by fusing attributes obtained from single camera, spatio-temporal information and the appearance of vehicles.

2. DESIGN AND ALGORITHM

2.1 Single Camera Vehicles Tracking

Considering the slow speed of vehicles in campus and the short intervals between two adjacent frames, an object tracking approach by using area overlapping ratio is proposed as the prediction model. The approach also solves the complexity of using a Kalman filter to obtain the trajectories of vehicles.

The proposed vehicle tracking approach contains two parts: detection and tracking. Moving target detection can be classified into two categories. One is to detect the moving object by using background modeling based on adaptive mixture Gaussian model[6] and then analyze the morphological filters and connectivity of this area. But in campus surveillance video, some crowd people may be detected as vehicles falsely. The other approach is to adopt RCNN[7] to recognize the vehicles, which has higher accuracy and also performs better on partially occluded vehicles. But it cannot be avoided that some tracks are missing because of not being recognized.

So we combine background modeling and RCNN together to solve the single camera vehicles tracking problem. Background modeling is used to extract the foreground blobs by background subtraction with adaptive mixture Gaussian model. After noises removed, the blobs are put into RCNN for detection to verify whether they are vehicles. This method takes advantages of both background modeling and RCNN, which can not only avoid mistaken detection of background modeling, but also accelerate the process of RCNN. If the result turns out to be a vehicle, a tracker will be distributed to the detected vehicle.

However, the real-time background updating faces great challenges in campus surveillance videos due to the slow speed and parking of vehicles. Grimson[6] used a strategy to solve this problem: if the current observation cannot be matched to all Gaussian components, sort the k-1 components in descending order, keep first k-1 components, and replace k-th Gaussian component by pixel values, low weight, high variance. When the pixel in the foreground exists for a long time in the learning process as one in the background, its weight will increase to a certain value that the model will mistake it as a component of background. Hence the learning strategy is adjusted that when the vehicle has a distributed tracker it will be marked as foreground. The pixel will be excluded in the model learning which reduces the learning rate of the pixel.

After solving the problem of recognition, we focus on the tracking of moving vehicles by using the overlapping area. The vehicle that has the largest overlapping area with the previous frame is considered as the best prediction. To deal with the occlusion of vehicles from different directions, we add a direction parameter to judge if the best predicted vehicle has the same direction with the previous one. Then we distribute new trackers to objects without past tracking states after updating the trackers. For the trackers which has not been updated, we will keep them unchanged when the object is in the center of camera. Otherwise, we need to detect whether to delete this tracker according to the update strategy.

In addition, it will be more convenient to observe the vehicle tracks by visualizing. Hence we zoom in to the map to find corresponding intersection and find the corresponding coordinate with the vehicle trajectory using homography matrix[8].

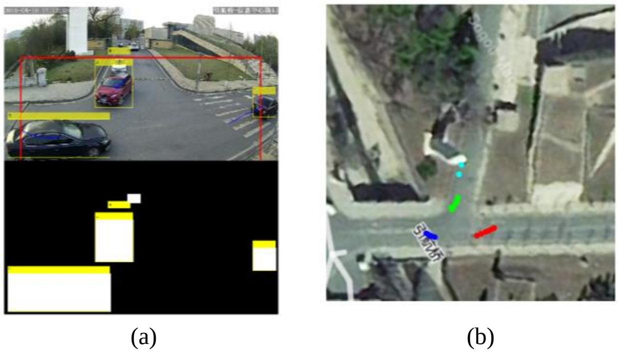
(a) (b)

Figure 2. Samples in vehicle detection and vehicle trajectories

The result is shown as Figure 2. (a) is the detection result using the proposed approach. The area in red box is the region of interest. The white box with yellow upper border is the detected vehicle with distributed tracker. The white box without yellow upper border is the detected vehicle without distributed tracker. The box only with yellow upper border is the object not being detected with distributed tracker. The blue line is the trajectory of the vehicle. And (b) shows some vehicle trajectories using homography matrix.

2.2 Multi-camera Vehicle Tracking

The non-overlapping multi-camera object tracking problem is different from traditional single camera object tracking. The blind region between cameras leads to discontinuation of the spatio-temporal information for the same vehicle, which makes it difficult to track objects with non-overlapping scenes. To solve these problems, firstly we established a topology estimation of camera networks using the single Gaussian model to obtain the spatio-temporal information.

We measure the similarity by formula (3). The features include spatio-temporal information denoted by *Sp*, attributes denoted by *A*, including types, face directions, colors, etc. and CNN features denoted by *F*.We set a weight w for each feature to control the reliability. Here, w_1, w_2, w_3 are set to: 1, 1, and 0.6 .

$$S = w_1 Sp \times (w_2 A + w_3 F) \qquad (3)$$

2.2.1 Topology estimation of camera networks using single Gaussian model

The enter and exit points of vehicles obtained from single camera object tracking are named as nodes. But the enter and exit position of different vehicles is in a range of fluctuation because of noise, vehicle size or other factors. So we merge the adjacent nodes into one, which will reduce the amount of nodes and the database of transfer time probability distribution between nodes. After obtaining the positions of nodes in camera view, we need to establish the function of transfer time between neighbor nodes. Using the single Gaussian model to train the function, we match the objects only by the appearance of objects and vehicle attributes, without the constraints of spatio-temporal information in target association. Topology model[9] will keep updating with the matching results until the relationship of topology parameters is tending stable, or learning time achieves a set value. And then this function can be used in target association.

2.2.2 Attributes Categorization

According to [10], we define a list of colors including black, gray, white, red, green, orange, yellow, golden, brown, and blue. Then for each color, each track is classified into one of seven vehicle types, i.e. sedan, SUV, MPV, van, pickup, bus and estate car. The color and type labels are assigned by three workers using a majority vote.

2.2.3 Multi-camera vehicle re-identification

Vehicle appearances are usually changed in different viewpoints, which brings the variation of visibility about attributes across cameras. To solve this problem, we can obtain the key frames of similar views of vehicles through continuously detecting and tracking in multi-cameras. We take every 5th frame as the key frames, and mark key frames of the current vehicle in this angle range. Then we will search the features of the current vehicle saved in the database of the other cameras, then match them with the key frame in which the vehicle has the nearest angle with the current one (see Figure 3).

Considering the difference of direction and brightness of vehicles in key frames of different cameras and reflective phenomenon, the traditional methods such as HOG and SURF have low matching rates. Also, the extracted points of feature are often on the ground instead of vehicles. Here we choose to use CNN features in matching the key frames of vehicles. There are many applications of using CNN, which achieves remarkable successes in many computer vision problems, such as object classification[11], detection[12], and face alignment[13]. Specifically, we used a pre-trained model[14][15] with 4096 dimensional features. It is also mentioned in the paper that this conclusion that may be applicable to many other tasks. The matching results are compared in Euclidean distance for the purpose of vehicle verification.

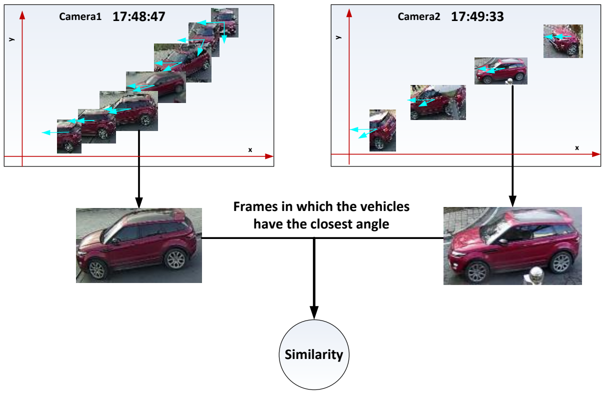

Figure 3. Illustration of multi-camera vehicle matching solution

3. EXPERIENCES AND PERFORMANCE

In order to test the performance of proposed approach, we collected vehicle monitoring videos from cameras in the real campus security monitoring system of Dalian Maritime University to analyze. All videos are in the resolution of 1920×1080 with 25 frames per second. 6 non-overlapping cameras with 60-minute durations are selected for evaluation. The complexity of experiments increased by lots of intersections, crowding of students, low speed of vehicles, etc.

A). Illustration of single camera target tracking

Evaluation results of single camera target tracking are shown in Table I. In the table, real targets indicate the number of real vehicles captured by the specified cameras. Track recall is the successful tracking rate within a camera. If the accuracy rate of a target tracked is more than 90%, it is considered as mostly tracked. When a car was wrong tracked, such as missing and lost, it is marked as wrong tracked. This approach shows a better robustness from the result.

Table 1. Evaluation of single camera tracking

Camera number	Real targets	Track recall	Mostly tracked	Wrong tracked
1	245	98%	240	5
2	218	95%	207	11
3	205	96%	197	8
4	184	96%	177	7
5	197	92%	182	15
6	206	95%	197	9
Average	209	95.5%	200	9

B). Illustration of inter-camera vehicle re-identification

Examples of inter-camera vehicle re-identification using CNN are shown as Figure 4. Table 2 shows the accuracy comparison. It is clearly to find that the re-identification results of car front, car body and car tail of vehicles using HoG and SURF are not ideal, while the results using CNN has a better robustness. This result proves that the CNN can catches more detailed features of vehicle.

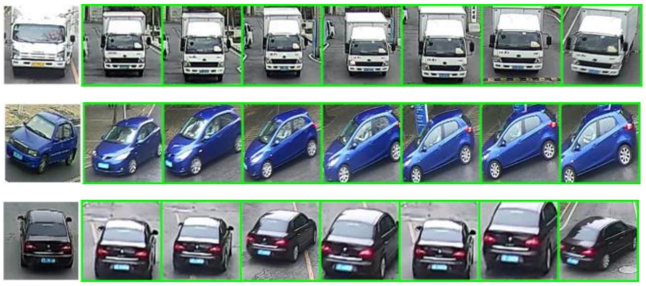

Figure 4. Examples of correct vehicle re-identification by using CNN

Table 2. Comparison of matching accuracy

Part of vehicle	CNN	HoG	SURF
Car front	96%	55%	62%
Car body	95%	43%	45%
Car tail	90%	42%	53%

C). Illustration of inter-camera vehicle tracking

Due to the key frames of vehicles in different angles from single camera, using CNN features in matching vehicles will increase the accuracy. The trajectory of tracked vehicle is shown in Figure 5. In the left satellite map, the red parts are the results from single camera target tracking, which can be seen clearly in the right satellite map. And the yellow parts are estimated trajectories calculated according to the similarity of vehicle images. In this case, improving the threshold of matching similarity can be used to determine the most similar vehicles.

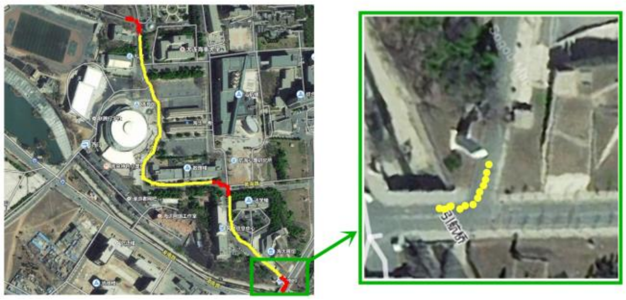

Figure 5. Example trajectory visualization for tracking and estimation on satellite map

3. CONCLUSIONS

In this paper, we proposed a method for continuous vehicle detection and tracking in multi-camera surveillance system. The trajectories of vehicles in single camera tracking are obtained by the combination of background modeling and RCNN. In addition, the vehicle trajectory was visualized using homography matrix. For multi-camera vehicle tracking, we matched the key frames of vehicles using CNN features. And after fusing attributes obtained from single camera, spatio-temporal information obtained from topology relationships and the appearance of vehicles, we can get the continuous trajectories of vehicles in multiple cameras. The method is tested by vehicle monitoring videos from cameras in the real campus security monitoring system and the results indicated its effectiveness for multi-camera vehicle tracking.

4. REFERENCES

[1]Rogerio Schmidt Feris, Behjat Siddiquie, James Petterson, Yun Zhai, Ankur Datta, Lisa M Brown, and Sharath Pankanti, “Large-scale vehicle detection, indexing, and search in urban surveillance videos”, IEEE TMM, vol. 14, no. 1, pp. 28－42, 2012.

[2]Cao, L., Chen, W., Chen, X., Zheng, S., & Huang, K. (2015). An equalised global graphical model-based approach for multi-camera object tracking.arXiv preprint arXiv:1502.03532.

[3]Bao S. Y., Xiang Y., Savarese S.. 2012. Object Co-detection In Proceedings of the 12th European Conference on Computer Vision. 7572 (Oct. 2012), 86-101.

[4]Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang, “A large-scale car dataset for fine-grained categorization and verification,” in CVPR, 2015, pp. 3973－3981.

[5]Jiang W, Xiao C, Jin H, et al. Vehicle tracking with non-overlapping views for multi-camera surveillance system, 2013 IEEE 10th International Conference on. IEEE, 2013: 1213-1220.

[6]Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. (Vol. 2). IEEE.

[7]Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[8]Ueshiba, T., & Tomita, F. (2003, October). Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 966-973). IEEE.

[9]Ellis, T. J., Makris, D., & Black, J. (2003, October). Learning a multi-camera topology. In Joint IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS) (pp. 165-171).

[10]Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang, “A large-scale car dataset for fine-grained categorization and verification,” in CVPR, 2015, pp. 3973–3981.

[11]A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[12]R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR,2014.

[13]Z. Zhang, P. Luo, C. C. Loy, and X. Tang. Facial landmark detection by deep multi-task learning. In ECCV, pages 94-108,2014.

[14]Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.Proceedings of the British Machine Vision, 1(3), 6.

[15]Vedaldi, A., & Lenc, K. (2015, October). MatConvNet: Convolutional neural networks for matlab. In Proceedings of the 23rd Annual ACM Conference on Multimedia Conference (pp. 689-692). ACM.