A Suspect Vehicle Tracking System Based on Video

Yadi Chen [1], Tuo Wang

Abstract. Video surveillance systems are widely used in security field. The large amounts of video data make it time-consuming for criminal investigation. Traditional researches conducted on visual analysis are extremely challenging, due to the complexity of scenes, such as large variation in viewing angle and multiple moving objects. This paper approaches these problems by proposing a video based suspect vehicle tracking system. We used a road modelling based on spatial clustering to obtain colour road information. In order to improve the accuracy of target matching, we fit the 3-D model with region of interest (ROI) to the 2-D target images, and compute the histogram in ROI as the feature for target matching. Experiment shows that system can provide accurate result for suspect vehicle tracking.

Keywords: Surveillance video; Video analysis; 3-D model; ROI

1 Introduction

With the falling cost of mounting cameras to capture video imagery, a wide area of urban road intersections has been covered with video surveillance for statistical and security purpose [1，2]. Although video surveillance is a great help for police to locate the suspect on the run and find evidences, it also causes another problem for manual search. After a crime occurs, investigator has to go through the continuous 24-hour monitoring video captured by multiple cameras to search for a particular target, which can be very time-consuming and labor-intensive. What is needed is a video-based system to search for the target automated and provide the investigator with the precise result. Although over the last decades, numerous researches conducted on visual analysis

Y. Chen (✉)
The School of Electronic and Information Engineering, Xi'an Jiaotong University,710049,Xi'an, Shaanxi, P.R.China
e-mail: chenmodidi@gmail.com

T. Wang
The School of Electronic and Information Engineering, Xi'an Jiaotong University,710049,Xi'an, Shaanxi, P.R.China
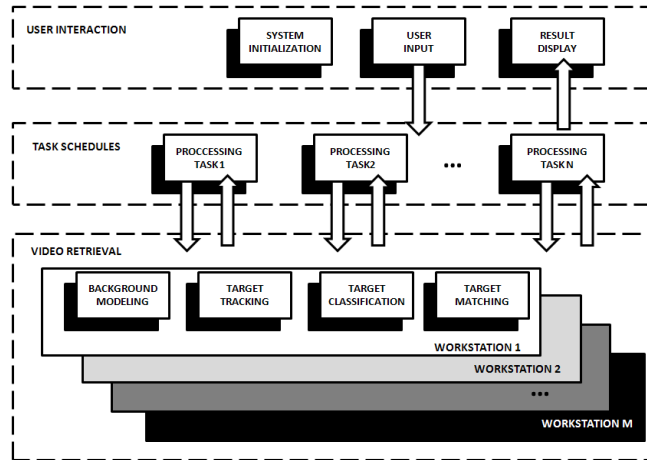e-mail: twang@sei.xjtu.edu.cn

of traffic scenes [3，4] concerns about advanced algorithms related to background modeling, target tracking and classification, few of them pay much attention to the target matching with different posture in multiple cameras.

On the basis of the above consideration, we propose a video-based computer vision system that enables a single operator to search for a specific target through a distributed network of active video sensors. It is flexible with complex area such as different camera positions and multiple moving objects. The main contribution of this paper is listed as follows: In order to obtain colored road information, a road modelling method based on spatial cluster is used for background modeling. A 3D model based strategy is employed to locate our defined region of interest (ROI), such as the front and the back of the vehicle.

The paper is organized as follows: In Sec 2, we present the brief introduction about the three-tier hierarchy architecture designed for parallel processing on multiple servers, with the description of key technologies in Sec 3. The experimental results are listed in Sec 4 and conclusions with plan for future research are drawn in Sec 5.

2 System Architecture

As a three-tier hierarchy architecture, our system from bottom to top can be divided into three level: the underlying, the middle and the top, as shown in 0.

Fig.2.1. System Diagram.

The top of the system is called user interaction, which consists of system initialization, user input and result display. System initialization module is used to provide configuration information for computing the size and speed of object. For each camera, parameters including optical axis deflection angle, focal length, and CCD area, the height and location of the camera are usually available, or if unknown, they can be calibrated by the interactive tools [5，6]. For user input module, the sample image of the target to search can be submitted from a special player with the screenshot function for user to

select the target from a piece of video. Then a method called grab cut will be employed to the sample image, to segment the target from the background for further processing. As for result display module, a report contains every detail of the matched target will be displayed to the user on the digital map based on camera location, along with the size, speed and evidence video pieces of the target.
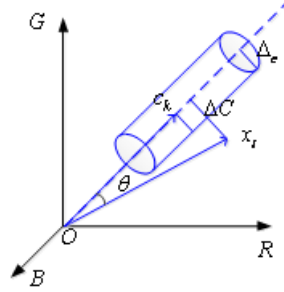
The middle level is called the task schedules, which is able to handle multiple parallel operations running simultaneously. For each user input from the top, new task is created and scheduled for further processing by servers in the underlying. After all the results are collected from the underlying, the middle level will generate a report which will be submitted to the top for result display.

The underlying level represents the parallel applications for suspect vehicle tracking on clusters of workstations, including background modeling, target tracking, target classification and target matching. The background modelling gets the background information by obtaining the colour road information from the video sequences. In the target tracking module, after the noise removal, moving subjects are marked using the connected domain algorithm and tracked using a global nearest neighbor optimization that consists of both appearance and dynamic information. As a result, each moving target in the background-subtracted image will be detected and traced until leaving the scene. Meanwhile, the target classification module will classify each vehicle detected into six categories: hatchback, car, SUV, van, bus and truck. Only the vehicles that match the category of the user input will be delivered to the target matching module. In this module, the feature of the ROI is calculated using a 3D model based strategy to measure the distance between the input sample and the detected vehicle. The result will be reported to the middle layer when the task is finished.

3 Key Technologies

3.1 Roads Background Modelling

The most commonly used methods for background modeling are frame difference, the mean background modeling, optical flow method, Gaussian background modeling and Bayesian background modeling. Because of the complexity of traffic scene, the methods above all show a great deficiency for target matching. Frame difference, the mean background modeling and optical flow method have poor adaptability to environmental change. Optical flow modeling is more accurate, but the computational complexity can hardly meet the requirement of the system. And Gaussian mixture modeling requires the assumption that the background pixels follow a known distribution, which does not meet the fact that the background of the transport network follows non-parametric distribution. Although Bayesian modeling can be performed in color space, the accuracy is not acceptable when the background color is similar to the foreground, which often occurs in traffic scene. What is needed is methods that can accurately obtain the colored road information.

Fig.3.1. The schematic diagram of color distortion.

Statistical analysis has shown that the pixels in a specific position of the road are always within a range of fluctuation, which follows the rules of color distortion, as is shown in Fig.3.1. Kyungnam Kim confirmed by experiment that, in the RGB color space, with the illumination change, most of the pixels values on the color template are changing along with the axial direction of the origin point [7]. Based on the consideration above, we employed the rules of color distortion and brightness distortion as the criteria for cluster analysis, and each cluster represents a path model possibility. Thus, we can estimate whether the pixels are background in color space and avoiding the splitting case that occurs when the color of the subject is similar to the background.

3.2 ROI Strategy Based on 3D Model

To search for the suspect vehicle is essentially a matching problem. The traditional feature-based vehicle matching method, such as the headlights matching which requires the frontal view of the vehicle, is found to be low and flexible to camera viewpoints. Other vehicle matching method using shape and appearance techniques can only make distinction between different categories other than the subjects of the same category. In order to provide an accurate result for suspect vehicle tracking, we proposed the ROI strategy based on 3D models for feature extraction.

The experiment result shows that matching accuracy can be increased by using 3D vehicle model fitting method [8], which aims at classifying target into different categories of vehicle. Based on the related works, we build six models represent for hatchback, sedan, SUV, van, bus and truck. After the system is initialized, a list of the possible 3D models is complete with the location and orientation on the road plane. In target classification module，the detected subject will be classified into the six categories, using the appearance vector to measure the similarity between the subject and each model.

As for target matching, the experimental results show that the features extracted from the informative region such as the front are more representative than a whole vehicle. The ROI was the feature-rich regions defined on the 3D vehicle models, see Fig 4. For each vehicle detected in the video, given initial pose and position parameters,

the ROI on the 3D model representing the category of the target is projected into 2D image, to generate a hypothetic location of the informative region. We mark each region with a certain color for further processing. Features of the same ROI will be measured in target matching module.

3.3 Target Matching

In target matching module, the similarity between the detected vehicle and the sample image of the suspect vehicle is calculated. Although the camera angles vary, the same region of the vehicles is detected using the ROI strategy based on 3D models. This allows us to reduce the false match caused by diversity of viewpoints.

The sample image input by user has been processed in the top layer, which include the segment of the target from the background, the ROI determined by user, the category that the vehicle belongs to and the workstations that need to track the suspect vehicle from surveillance videos. According to the user operation, the task schedules are generated to the specific workstations to search for the vehicle of specific category. Thus, the result of classification will determine whether the detected vehicle will be processed in target matching module, and discard the unqualified ones.
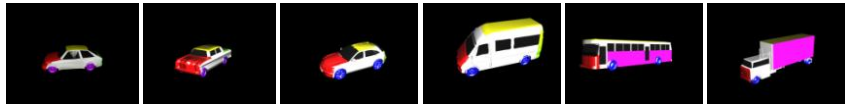
Due to the variety of the camera angles and vehicle poses, the ROI will be different in the sample image and detected video. It is important to find corresponding ROI, which can be based on the color of the region to distinguish. We define the similarity between the detected vehicle and the sample image by the summation of the distance of the color histogram d_i of each corresponding ROI, multiplying with weight coefficient w_i.

$$\frac{w_i}{w_j} = \frac{S_i}{S_j} \tag{1.1}$$

$$d_i(H_1', H_2') = \frac{\sum H_1'(j)^2 \cdot H_2'(j)}{\sqrt{(\sum H_1'(j)^2) \cdot (\sum H_2'(j)^2)}} \tag{1.2}$$
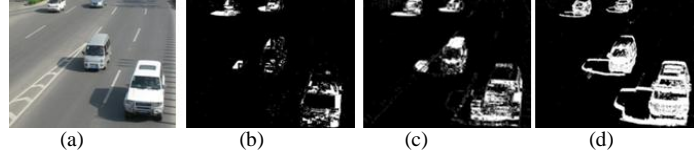
$$d = \sum_i d_i w_i \tag{1.3}$$

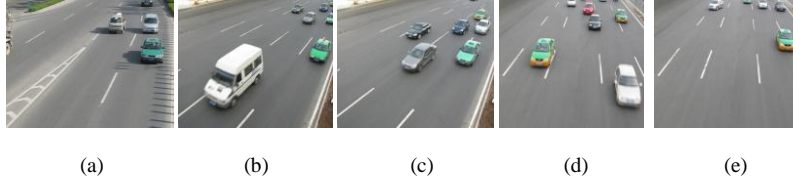S_i represents the area and $H1$ is the histogram of the ROI.

Fig.3.2. The3D models with region of interest (ROI) used for matching
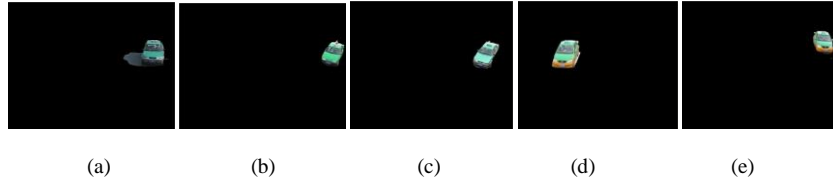
4 Result and Analysis of Experiment

In this section, we demonstrate the proposed system. The analysis of the rest modules are shown as follows. We compared the effectiveness and efficiency of the methods of road modeling and target matching with the traditional methods.
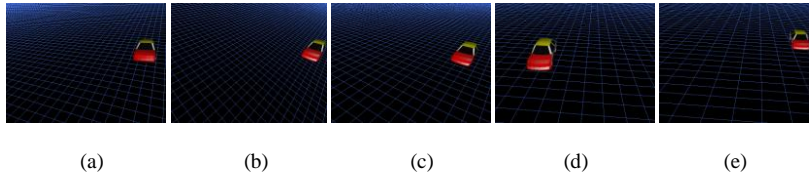

(a)   (b)   (c)   (d)

Fig.4.1.(a) The original image, (b) A result of background modeling using Bayes, (c) A result of background modeling using Gaussian, (d) A result of background modeling using the method of our system.

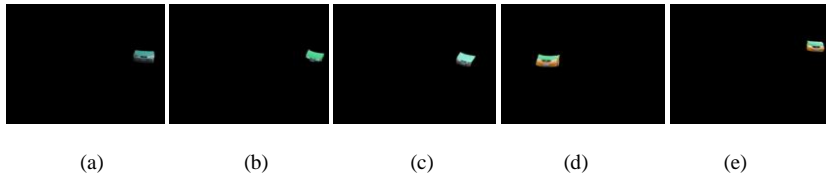(a)   (b)   (c)   (d)   (e)

Fig.4.2. The five original images captured in 3 different intersections. Fig (a), Fig (b) and Fig(c) belongs to the same type of taxi, while Fig(d) and Fig(e) blong to the other type of taxi.
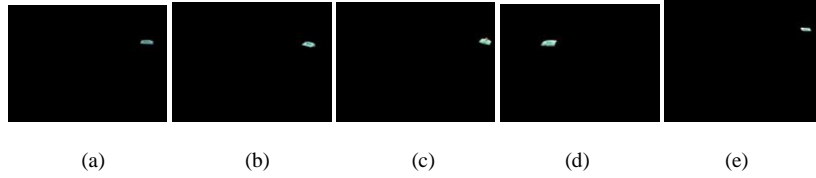

(a)   (b)   (c)   (d)   (e)

Fig.4.3. The segment results of the five pictures.

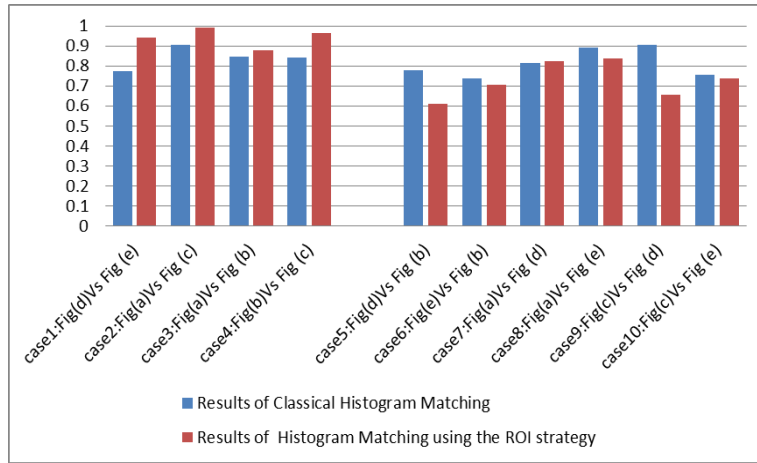(a)   (b)   (c)   (d)   (e)

Fig.4.4. Projections of the 3D model on a virtul road plane that the target belongs to.

(a)   (b)   (c)   (d)   (e)

Fig.4.5. Front face segment results, each corresponds to the red region of interest on the 3D model.

<div align="center">(a)   (b)   (c)   (d)   (e)</div>

Fig.4.6. Car roof segment results, each corresponds to the yellow region of interest on the 3D model.

Fig.4.7. The target matching results of the five images in Fig 7. The first four cases represents the matching results of the same vehicle, while the last six cases represents the different vehicles. The proposed matching strategy increases the matching rate of the same objects under different perspectives, and reduces the rate of the different. By using the ROI strategy based on 3-D models, a threshold of 0.85 can be selected to distinguish the same objects from the different ones, while the classical histogram matching method makes it difficult to select a threshold.

We compared the images processed by Gaussian background modeling, Bayesian background modeling and the method based on spatial cluster, shown in Fig.4.3. The result shows that a foreground object is lost when employing the Bayes, due to the background color is similar to the foreground object. The result of using Gaussian is also not acceptable for the same reason. 0shows the result of comparing the traditional matching method with the ROI strategy based on 3-D models. The histogram indicates that the color distribution of the ROI can describe the feature of the target better than the whole region.

5 Conclusion

In this paper, with respect to the problem of tracking suspect vehicle in videos captured by multiple cameras, we proposed a suspect vehicle tracking system. According to the experiments, the results for road modeling are much better than that of Bayes and Gaussian back ground modeling. And the 3D model-based method is

adaptive to different cameral position. The system can provide accurate result for suspect vehicle tracking.

6 Acknowledgement

7 References

1. Weiming Hu ; Xie, D. ; Zhouyu Fu ; Wenrong Zeng ; Maybank, S. " Semantic-Based Surveillance Video Retrieval [J]". Image Processing, IEEE Transactions on, 2010.

2. Räty, T.D. "Survey on Contemporary Remote Surveillance Systems for Public Safety". Systems, Man, and Cybernetics[J]. in IEEE Transactions on, 2010.

3. A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis. "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proc IEEE 90(7), 2012,pp. 1151-1163.

4. Jie Shao, Zhen Jia, Zhipeng Li, Jianwei Zhao, Pei-Yuan Peng. "Feedback strategy on real-time multiple target tracking in cognitive vision system[J]". in Optical Engineerig ,2011, vol. 50

5. Messelodi, Stefano, Carla Maria Modena, and Michele Zanin. "A computer vision system for the detection and classification of vehicles at urban road intersections." Pattern analysis and applications 8.1-2 ,2005, pp. 17-31.

6. Worrall, A. D., G. D. Sullivan, and K. D. Baker. "A simple, intuitive camera calibration tool for natural images." Proc. 5th British Machine Vision Conference. Vol. 13. No. 16. 1994.

7. Kim, Kyungnam, et al. "Background modeling and subtraction by codebook construction." Image Processing, 2004. ICIP'04. 2004 International Conference on. Vol. 5. IEEE, 2004.

8. Leotta, Matthew J., and Joseph L. Mundy. "Vehicle surveillance with a generic, adaptive, 3d vehicle model." Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.7 (2011): 1457-1469.