Vehicle Re-Identification in Surveillance Environment

Kuen-Ming Lan,[1]
Jen-Hui Chuang[2] and Zhe-Wei Wu[3]
blueriver@itri.org.tw[1], jchuang@cs.nctu.edu.tw[2] and ICEwu@itri.org.tw[3]

Abstract- An intelligent transportation system (ITS) usually needs to automatically monitor the road traffic, and may also need to detect and track vehicles of interest. This paper is concerned with an extension of ITS to a campus surveillance environment to identify vehicles in images in different locations of a campus. The proposed approach uses corners of side windows as image features of a vehicle and maps them to a nominal configuration under which relative positions among them are compared. Experimental results show that the similarity measures thus established can indeed achieve the objective of vehicle identification effectively, and in an efficient way.

Keywords: intelligent transportation systems, campus traffic surveillance, vehicle identification.

1. Introduction

Recently, Intelligent Transportation Systems (ITS) has become an important research topic in the general field of transportation. A major portion of the research works on such systems is focused on automatically monitoring the traffic flow on the roads. Rather than the aggregated flow analysis, a more powerful ITS system should also provide detailed information of each vehicle, such as its position and speed in time. These systems will be very useful in reducing the workload of human operators and in improving our understanding of real traffic scenes such as congestion and collision problems on the road networks [11]. Traffic monitoring based on video sensors, which is widely adopted nowadays, has a number of advantages. First, they are easy to use and install. Second, video sensors offer the possibility of providing rich description of traffic situation, and vehicle tracking. Third, large areas can be covered with a small number of visual sensors. Fourth, the system cost is likely to decrease in terms of image acquisition devices and powerful computers. Thus, a traffic monitoring system based on video sensors is expected to collect rich information and carried out detailed analyses for real traffic surveillance.

Campus traffic surveillance is a special case of real traffic surveillance in which tracking of a vehicle of interest, such as to obtain its location and speed in time from an entrance to an exit via visual sensors of non-overlapping fields of view (FOV), is often desirable in a campus environment. In general, the vehicle speed is relatively lower in the campus environment compared to that in a highway or road network. However, different perspectives of the vehicle, and different outside factors such as light reflection and shadowing situations, will cause major variations in the appearance of the vehicle and make the tracking complicated. In order to resolve the difficulty, a hybrid approach to the identification of a vehicle under a campus surveillance situation is proposed in this paper, which uses image gradients and homographic transform to establish a vehicle appearance model.

The remainder of this paper is organized as follows. In Section 2, related works on vehicle identification are briefly described. In Section 3, the proposed hybrid approach is presented. The experimental results are given in Section 4. Finally, conclusions are drawn in Section 5.

2. Literature review

Segmentation and tracking of moving objects in an image sequences are basic tasks for various applications of computer vision, such as video monitoring, intelligent-highway system, intrusion detection, airport safety, and campus security [4]. The aim of motion detection is to identify one or more moving objects from the background image in video sequences, which is also very important for targets classification and tracking in video image processing [9]. However, because of the sensitivity to lighting and weather conditions, it is often difficult to segment moving objects from background in general. Many approaches to motion segmentation are proposed in recent years including background subtraction [13], time differencing [9], optical flow [2], and Expectation-Maximization algorithm (EM) [10].

Background subtraction is a common approach to motion segmentation, but it requires the establishment of a background model which can automatically adapt to a complex scene. A segmentation approach with adaptive background subtraction is described [7] in which Kalman filtering is adopted to predict the background image at some predetermined time intervals. The error between the prediction and the actual background image is used to update some state variables of the filter. Such an approach has the advantage that it can automatically adapt to changes in lighting and weather conditions. However, the system needs to be initialized with an image of only the background scene. Friedman and Russell [4] provided a probabilistic approach to the segmentation of an object. They used the expectation maximization (EM) method [10] to classify each image pixel as moving object, shadow, or background. This method is not only insensitive to lighting conditions, but also does not require an initialization with a background image. However, this method produces many small regions which can be difficult to be separated from real noises.

In surveillance applications, re-identification of a vehicle is often a major part of vehicle tracking. Various approaches to the vehicle tacking problem have been published. Arth *et al*. [1] presented a system to reacquire and track vehicles. They use PCA-SIFT to extract features from a large set of samples, build the feature vocabulary tree, and match the samples based on the tree. Dlagnekov [4] presented a series of approaches to compare the performance of vehicle tracking. In his thesis, he provided an Eigen system to define the feature for tracking and identification. Guo *et al*. [6] proposed a complex approach to vehicle re-identification between airial images of non-overlapping FOV. They used sequence-to-sequence matching and image alignment to track vehicles with small view angles changes. Shan [12] provided unsupervised learning approach for vehicle matching between cameras of non-overlapping FOV. They proposed a feature-based vehicle matching algorithm which will provides a probability of whether two objects are the same.

Koller *et al*. [8] presented a subtraction algorithm in which the background is dynamically estimated from incoming images, and the difference between the current and the background images was threshold to form blobs corresponding to vehicles. Beymer *et al*. [3] used corner features extracted from image of vehicles. The tracked corner features are grouped together based on the proximity of their positions and the similarity of the motion. This approach gave good results even with less favorable illumination conditions. However, it still has three limitations. First, the location and the dimension of a detected vehicle may not be accurate because they are estimated from the corner features which may not cover the whole vehicle (and some of them may belong to the shadow). Second, the position error caused by missing features (tracking failures) may introduce a significant error in the velocity estimation. Third, the feature grouping is based on only the locations and the motions of corner features. Thus, there are times when features of nearby vehicles (of the same speed) are grouped together, or the features of a large vehicle (for example, trailer trucks) are not grouped together.

Yue *et al*. [15] proposed a two-view tracking method which used homographic relation between the two views. Tomoaki *et al*. [14] presented an approach which estimated the lateral position of the vehicle based on the registration of top-view images by using homographic transform. They used homography to generate affine images for tracked vehicles; however, since the approach is based on top-view images, effective feature will be lost if cameras with lower resolutions are used. In this paper, near side-view images of vehicles obtained from cameras of non-overlapping FOV are used. The proposed re-identification algorithm, which used feature points obtained for side windows of a vehicle, will be described next.

3. The proposed approach

In this section, a hybrid approach to vehicle re-identification for campus vehicle surveillance is introduced. The overview of the proposed scheme is shown in Figure 1. In the first stage (the first block with dashed border line), stable feature points, whose correspondence can be correctly established among different surveillance images, on a vehicle are extracted by a hybrid approach. In the second stage, homographic transform is applied to some of these feature points mapping them to predetermined locations so that relative locations of other feature points can be used to re-identify the vehicle.

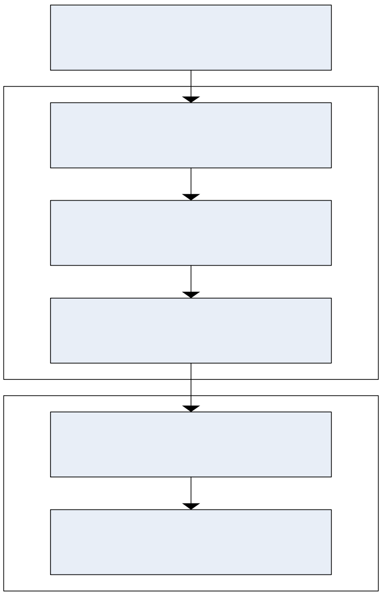


Figure 1. The overview of the proposed scheme.

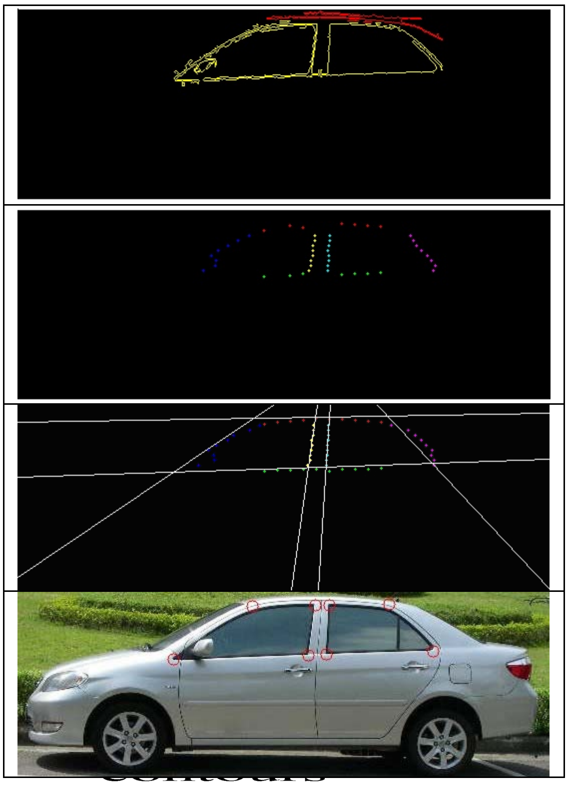
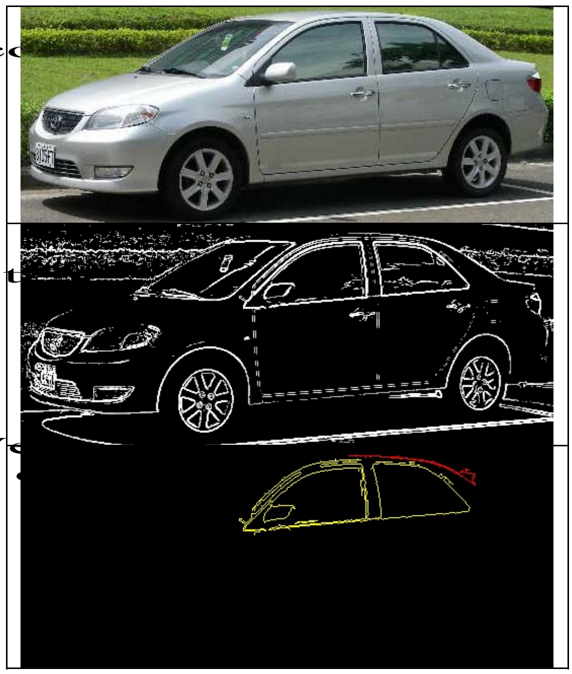


Figure. 2. Example of the proposed feature extraction scheme for side view of a vehicle.

A. The stable feature extraction process

In the stable feature extraction procedure, the four (stable) corners of the glass region of each of the two side windows are identified in a sequence of the surveillance images. The procedure is quite straightforward and can be performed in real time.

Figs. 2 and 3 show examples of the above procedure for two images (manually segmented sub-images) of the same vehicle obtained from different view angles. First, the system acquires an image of a vehicle with its side windows readily observable. Image gradients (edges) are then obtained by applying the Sobel operator. Subsequently, the two largest edge contours, for the two side windows, are identified in the upper half of the vehicle image. Finally the corners of the two windows are obtained by apply the principal axis analysis (PCA) to different sets (colors) of point samples obtained from horizontal and vertical scan-lines



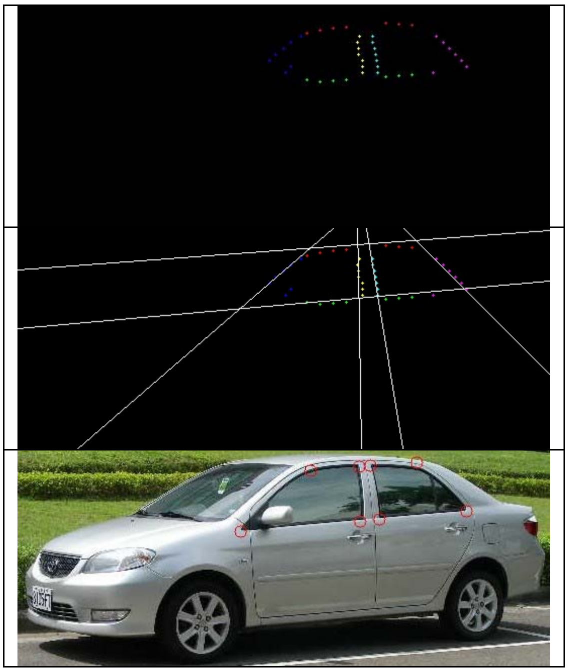


Figure. 3. Example similar to Figure 2 but with a 10˚ difference in the viewing angle.

B. The homography-based re-identification

Homography is a mathematical concept in 3D geometry. It is defined as the relationship between image locations of points on a 3D planar region such that any of such points in one image frame corresponds to one and only one point in the other, and vice versa. In the field of computer vision, a homography is defined in 2D space as a mapping between a point on a reference plane, as seen from one camera, to the same point but seen from a second camera.

Consider a corner point $\boldsymbol{x} = (x_1, x_2, 1)$ in one image and the corresponding $\boldsymbol{x}' = (x_1', x_2', 1)$ in another image. There is a 3x3 homography matrix $\boldsymbol{M}$ which relates the pixel coordinates in the two images such that $\boldsymbol{x}' = \boldsymbol{Mx}$. In homogenous coordinates, we can always introduce a scaling factor, such as $w\boldsymbol{x} = (wx_1, wx_2, w)$, without change the location of an image point. Thus, $\boldsymbol{M}$ can be found if at least four pairs of corresponding feature points are given.

In the proposed vehicle re-identification scheme, a "nominal" configuration of four outer corners ($P1 \sim P4$) of the two side windows are arbitrarily chosen, as shown in Fig. 4. While $P1 \sim P4$ are fixed in location and are used to derive the homography matrix for all vehicles, the other four (inner) corners ($P5 \sim P8$) will only have identical locations for the same type of vehicle. Thus, the locatons of the latter can be used (together with other features like color of a vehicle) for vehicle re-identification among images acquired from cameras of non-overlapped FOV.

Accordingly, some distal and angular quantities associated with the four inner corners ($R1$, $R2$ and T shown in Fig. 4) are established as quantitative measurements of the similarity between two vehicles appear in two different images. Note that the distances are measured with respect to the midpoint of $P7$ and $P8$.

One can see that the above quantities simply give the relative location and orientation of the central region between the two side windows of an vehicle (shown with the nominal configuration). In the next section, these quantities will be used to determine whether vehicles appeared in different images are of the same or different types.

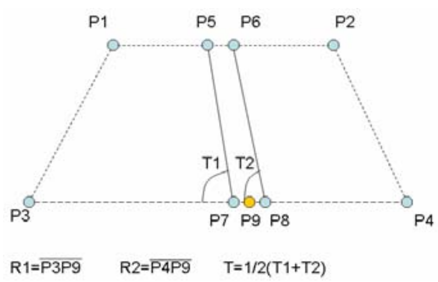


Figure 4. A "nominal" configuration for vehicle appearance comparison

4. Experimental results and discussion

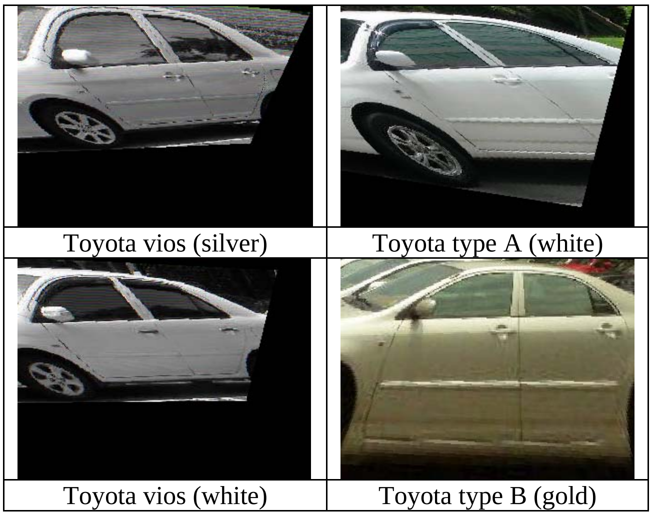


Figure 5. Toyota types of 10˚ side view with homography

Table 1. Similarity measurements

Types	T(rad)	R1	R2
Vios (silver)	0.863	132.136	90.088
Vios (white)	0.779	140.174	88.090
Type A (white)	0.583	153.160	118.067
Type B (gold)	1.526	136.014	114.004

In this section, the approach presented in the previous section is tested for the four vehicle images, two for each vehicle type, shown in Fig. 5. It is readily observable from Table 1 that the values of distal and angular quantities measured under the proposed nominal configuration can indeed provide similarity measurements among different vehicle images for re-identification, i.e., vehicles of the same type will indeed have higher similarity in these values than those of different types.

5. Conclusions

In this paper, a hybrid vehicle detection scheme is proposed. In the first stage, the stable feature extraction process is employed to find stable vehicle window corners. In the second stage, homographic transform is applied to a subset of these features to a nominal configuration. The distal and angular positions of the rest features, relative to that configuration, are then used to determine whether vehicles appeared in different images are of the same type. According to the experimental results, the proposed scheme works well for some real images. As for the computational efficiency, the proposed scheme takes almost constant time in complexity and has a system execution time suitable for real-time surveillance applications.

The proposed re-identification approach can be adopted as part of a more complete surveillance system which also takes into account other image features of a vehicle such as color and license plate data. Currently, the approach is developed for sedans with two side windows clearly separated by metallic body part. Generalization of the approach to a broader class of vehicles, as well as more complex weather condition, is under investigation.

References

[1] C. Arth, C. Leistner, and H. Bischof , “Object Reacquisition and Tracking in Large-Scale Smart Camera Networks.” In Proceedings of the First ACM/IEEE International Conference on Distributed Smart Cameras, Vienna, pp. 156-163, 2007.

[2] J. Barron, “Performance of optical flow techniques.” International Journal Computer Vision, vol. 12, pp. 42-77, 1994.

[3] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, “A real time computer vision system for measuring traffic parameters.” In Proceeding IEEE Conference Computer Vision and Pattern Recognition, pp. 495–501, 1997.

[4] L. Dlagnekov, Video-based Car Surveillance: License Plate, Make, and Model Recognition, U.C. San Diego, Masters Thesis, 2005.

[5] N. Friedman, and S. Russell, “Image segmentation in video sequences: A probabilistic approach.” In Proceeding The 13th Conference on Uncertainty in Artificial Intelligence, Rhode Island, USA, 1997.

[6] Y. Guo, S. Hsu, H.S. Sawhney, R. Kumar, and Y. Shan, “Robust Object Matching for Persistent Tracking with Heterogeneous Features.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 824-839, 2007.

[7] K.P. Karmann, and A.V. Brandt, “Moving object recognition using an adaptive background memory.” In Proc. Time-varying Image Processing and Moving Object Recognition, Amsterdam, Netherlands, pp. 289-296, 1990.

[8] D. Koller, J. Weber, and J. Malik, “Robust multiple car tracking with occlusion reasoning.” In Proceeding European Conference on Computer Vision, pp. 189-196, 1994.

[9] A. Lipton, and H. Fujiyoshi, “Moving target classification and tracking from real-time video”. In Proceeding IEEE Workshop on Applications of Computer Vision, Princeton, NJ, pp. 8-14, 1998.

[10] G. McLachlan, The EM algorithm and extensions, Wiley Interscience, 1997.

[11] N. Saunier, and T. Sayed, “Automated road safety analysis using video sensors”. Transportation Research Record: Journal of the Transportation Research Board, pp. 57-64, 2006.

[12] Y. Shan, H.S. Sawhney, and R. Kumar, “Unsupervised Learning of Discriminative Edge Measures for Vehicle Matching between Nonoverlapping Cameras.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, pp. 700-711, 2008.

[13] C. Stauffer, and W. Grimson, “Adaptive background mixture models for real-time tracking.” In Proceeding IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, vol. 2, pp. 246-252, 1999.

[14] T. Tomoaki, S. Hideo, O. Shinji, Y. Keiichi and I. Tora, “Vehicle Lateral Position Estimation Method Based on Matching of Top-View Images.” In Proceeding International Conference on Pattern Recognition, vol. 4, pp. 626-629, 2006.

[15] Z.F. Yue, K.S. Zhou, and R. Chellappa, “Robust two-camera tracking using homography”. In Proceeding IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3 pp. 17-21, 2004.