A Vision-Based Vehicle Identification System

Chung-Lin Huang and Wen-Chieh Liao
Institute of Electrical Engineering
National Tsing Hua University
Hsin Chu, Taiwan, ROC.
e-mail: clhuang@ee.nthu.edu.tw

Abstract

This paper presents a vision-based vehicle identification system which consists of object extraction, object tracking, occlusion detection and segmentation, and vehicle classification. Since the vehicles on the freeway may occlude each other, their trajectories may merge or split. To separate the occluded objects, we develop three processed: occlusion detection, motion vector calibration, and motion field clustering. Finally, the segmented objects are classified into seven different categorized vehicles.

1. Introduction

Recently, the intelligent traffic system (ITS) has been developed to make the existing traffic infrastructure more efficient. The most common sensor for ITS is the magnetic loop detector which is used to measure the length and the number of axles of vehicles. Based on the measurement, the types of vehicles can be identified. Since, these detectors are imbedded beneath the road, they have high failure rates due to pavement failures and poor maintenance. The vision-based video monitoring systems are more cost-effective.

The vision-based ITS consists of vehicle tracking and vehicle classification [1~4]. The former can yield traffic parameters such as traffic flow, vehicle velocity, lane changes and trajectory, whereas the latter is important in the computation of the percentages of vehicle classes of the streets and highways. ITS can also lead to an accurate design of the pavements on the roads. The most difficult problem in vision-based method is that the vehicles are often occluded. The binary occluded object blobs that are obtained from background subtraction can not be effectively divided into individual objects. This paper proposes a new method to separate occluded vehicles based on their individual motion fields.

The vehicle identification is an inherently hard problem, most of the works[5~8] did not provide a total solution. Given the wide variety of shapes and sizes of vehicles within a single category alone, it is difficult to categorize vehicles using simple parameters. This task is made even more difficult when multiple categories are desired. In traffic scenes, occlusions, shadows, camera noise, changes in lighting and weather conditions make the problem complex. Other than CCD camera, range sensor [9] and automated virtual loop [10] have also been used to capture the images which are not sensitive to lighting and may be less sensitive to other environmental conditions.

There are four main steps in our system: (1) foreground object extraction; (2) object tracking, (3) occlusion determination and occluded region segregation; and (4) classification. The object tracking will find the trajectory of each moving object. If the two trajectories merged as one or one trajectory split into two, then there may be occlusion. To divide the occluded region into two individual regions, we generate the so-called "cutting region" in the middle of the occluded region. By removing the "cutting region", we may divide the occluded region into two individual parts. Having detected and tracked the vehicle, we may classify the segmented vehicle into seven different types of vehicles.

2. Foreground Object Extraction

To track the vehicles, we need to extract the foreground moving objects from the images. The object extraction consists of two processes.

1) Background Subtraction. To segment the vehicle objects form a complex background, we use the well-known background subtraction method. Here we use a self-adaptive background subtraction method. The basic principal of our method is to modify the background image (called the current background (*B*)) by using instantaneous background (*IB*) and applying an appropriate weighting α as follows

$$CB_{k+1} = (1-\alpha)CB_k + \alpha IB_k$$

where the subscript k is the frame number index, the instantaneous background is defined as $IB_k = M_k \bullet CB_k + (\sim M_k) \bullet I_k$, I_k is the current image frame, M_k is the binary vehicle mask. It is obtained by subtracting the brightness of the current image from the estimated current stationary background followed by a binary thresholding.

2) Pre-processing. After background subtraction, we may find some noise of the extracted foreground image. Here, we develop the pre-processing process to obtain a more clean-cut silhouette for the following vehicle objects tracking. There are four steps in the pre-processing: noise removal, morphological filtering, labeling, and size filtering.

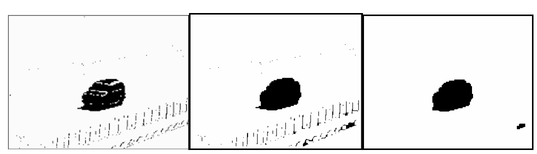


Figure 1. The images after pre-processing.

3. Object Tracking

The object tracking process traces the moving vehicles and finds the trajectory of individual vehicle. To track the regions in two consecutive frames, we may find four different phenomena: (1) the region might disappear in the next frame, (2) the region might appear in the current frame, (3) a single region in current frame might split into multiple regions in the next frame, (4) multiple regions in current frame might merge as a single region. The first two cases indicate the vehicles appear or disappear in the scene, and the latter two cases indicate when the vehicles are occluded or separated.

To track the moving objects, we rely on the correspondence between the objects in two continuous frames to creates the linkage. The linkages among the moving objects in the image sequence are their trajectories. Here we apply the path and shape coherence constraint to find the correspondence among the regions in the neighboring frames. The path coherence function ψ_p measures the smoothness of a moving component. It indicates that the deviation of the magnitude of velocity and the direction of moving object are smooth throughout the image sequence. The ψ_p is a defined as follows:

$$\psi_p(R_{i,k-1}, R_{i,k}, R_{i,k+1})=\omega_1\varphi_d+\omega_2\varphi_s.$$

where $\omega_1+\omega_2=1$, φ_d indicates the direction coherence, φ_s indicates the speed coherence.

The shape coherence function ψ_s measures the shape variation of a moving object in an image sequence. The scale of the shape of moving object may change when it is translating along the line of sight. The measurement of deformations can be described in terms of area, circumference, orientation, and so forth. The ψ_s is defined as

$$\psi_s(R_{i,k-1}, R_{i,k}, R_{i,k+1})=\omega_3\varphi_a+\omega_4\varphi_c.$$

where $\omega_3+\omega_4=1$, φ_a and φ_c, are functions of s_a and s_c , and $s_a=(a_{i,k-1}+a_{i,k-1}+a_{i,k+1})/2$, $s_c=(c_{i,k-1}+c_{i,k-1}+c_{i,k+1})/2$, $a_{i,k}$ and $c_{i,k}$ are the area and circumference of region $R_{i,k}$ respectively.

As shown in Figure 2, the trajectory of the segmented vehicle can be found for the first three frames F_{k-1}, F_k, and F_{k+1}. By using the coherence constraint functions, we can find the best trajectory for region $R_{i,k-1}$, i.e., $T_p=(R_{i,k-1}, R_{j,k}, R_{l,k+1})$, as

$$T_p = arg \min_{(i,j,l)\in I} \left[w_p\Psi_p(R_{i,k-1},R_{j,k},R_{l,k+1}) + w_s\Psi_s(R_{i,k-1},R_{j,k},R_{l,k+1}) \right]$$

where $w_p+w_s=1$. For each region $R_{i,k-1}$, we may find a trajectory based on the best correspondence between $R_{i,k-1}$, $R_{j,k}$, and $R_{l,k+1}$. However, it may not find the best correspondence due to occlusion. Suppose we apply the correspondence finding algorithm on frames F_{k-1}, F_k, and F_{k+1}. Since only two regions can be identified in F_{k+2}, we can find only one effective trajectory for three objects in the image sequence. Since the size of the merged region is much larger than the regions in the previous frame, it indicates that the two vehicles occlude each other.

If occlusion occurs, two regions merge as one, and the trajectory finding may not find the correspondence between two consecutive frames. If the measurement of all the ψ_p and ψ_s applied on any three regions in three consecutive frames are larger than certain threshold then no trajectory is found. As shown in Figure 2, the trajectory T_1 and T_2 discontinue between F_{k+1} and F_{k+2}. In such case, we link T_1 and T_2 in F_{k+1} to the same region $R_{1,k+2}$ in F_{k+2} , and two trajectories merge as a single trajectory T_{12}. Similarly, there is another possibility that two vehicles appear as one occluded region in the beginning, and then split into two regions afterward. In such case, we may link the single trajectory to two regions in the next frame to generate two new trajectories.

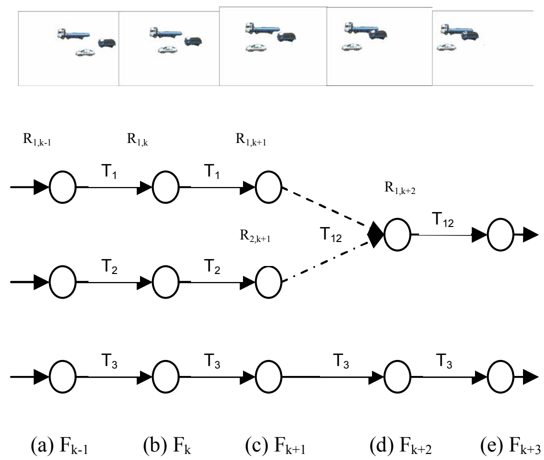


Figure 2. The three trajectories of the three vehicles.

4. Occlusion Detection and Occluded Object Separation

Here, we develop the occlusion detection and segmentation techniques. By analyzing the motion field of the merged regions and their trajectory, we may identify the occurrence of occlusion.

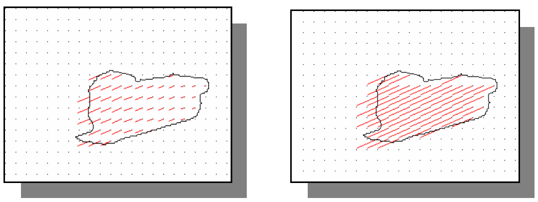
4.1 Motion Estimation

The moving vehicle is considered as a rigid body motion. We use the block matching method to search the motion vector of each pixel. Block matching does not provide accurate motion vectors especially for smooth area. We add two constrains to improve the accuracy of the motion vectors: (1) do the block matching only for the texture blocks, and (2) the block-matching searching process operates under certain fixed direction (since the vehicle is moving in a fixed direction). The fixed-direction block matching method requires less computation and generates more accurate motion vectors compared with the conventional full search method.

4.2 Motion Vector Transformation

Due to perspective projection, the motion vector filed of the same vehicle object in the same image is not homogeneous as shown in Figure 3(a). Here, we develop a motion vector transformation method to convert the non-homogeneous motion field to a homogeneous motion field as shown in Figure 3(b).

Since the direction of motion field is not changed by the projection, we only need to find a transformation function so that the motion vectors in the same moving object region will have equal magnitude. Here, we calibrate the motion field of the same vehicle to make it homogeneous. Here we take the motion vector at the centroid of the moving region as the reference, and transform all the motion vectors located in the same region to the reference velocity. Since the vehicle is moving in the fixed direction, the directions of the identified motion vectors are the same. To find the reference velocity for each path, we average the velocity as the vehicle passing through the centroid of K image frames.



(a) varying magnitudes; (b) constant magnitude.

Figure 3. Motion vector transformation.

At each location, we transform the motion vector to a constant by multiplying a scaling factor. For any two points (i, j) and (i', j'), which are on the same path with different velocity, we apply the scaling and have their motion vectors equalized. The velocities on the same path will be the same. The scaling factor for each motion vector is obtained in the off-line training process. During the training process, given a set of motion vectors and a set of constant motion vectors for each path p and each image frame. In the training process, for each location, we may obtain a set of scaling factors from the training image sequence, and then we can find the scaling factor by taking the average.

4.3 Occlusion Detection and Region Segregation

There are two ways to detect the occlusion: trajectories discontinuous detection or the motion field analysis method. After motion vector calibration, if the motion field of a binary blob is not homogeneous then we know that the occlusion occurs. The homogeneity of a region is determined by the variance of the motion field of that designated region. If the variance is larger than certain threshold, then the region is a possible occlusion region. However, if the size of the binary blob region is too small, the occlusion is not possible.

Based on the motion field homogeneity of the occluded region of two vehicles, we may separate the occluded region into two individual regions. Since the speeds of the two vehicles are different, after motion vector calibration, we may find two sets of motion vectors in the occluded silhouette. After the thresholding, we may divide the motion field into two different fields, and separate the corresponding region into two separated regions.

Here, we propose a segmentation method which removes the so-called "*Cutting Region*" to segregate the occluded region into two individual regions. To find the *cutting region*, we use the labeling method to label the occluded region as two regions. The labeling method finds two different motion fields in the silhouette. Then we use the morphological dilation method to enlarge the labeled regions. These two regions will overlap one another and generate an intersecting region called the "*cutting region*" as shown in Figure 4.

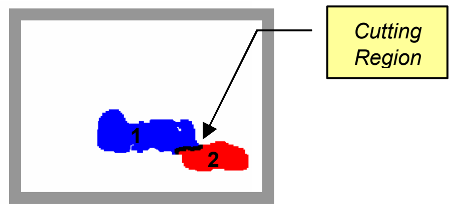


Figure 4. The cutting region extraction.

By removing the cutting region, we may cut the binary silhouette into two objects. The divided regions may not be precise. We use the opening operation to separate two divided regions as shown in Figure 5. However, the opening operation will jeopardize the divided binary silhouette. However, if the size of the operator kernel is two small, then the opening operation can not separate the two vehicles effectively. Therefore, our best trade-off is to select the opening kernel that can separate both vehicles. If both vehicles are separable, then at least one vehicle can be identified, else if both vehicles are still merged as one vehicle then no identification is possible.

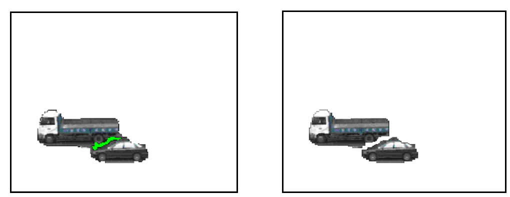


Figure 5. Occluded region separation.

5. Vehicle Classification

Our system classifies as many as seven types of vehicles. Here, we develop a so-called hierarchical vehicle classification method. The hierarchical classification consists of coarse classification and fine classification. The seven identifiable vehicle types are: (A) sedan, (B) van, (C) pickup, (D) truck, (E) van truck, (F) bus, and (G) trailer, as shown in Figure 6.

Figure 6. Seven vehicle types.

Based on the silhouette of the vehicles, we can extract three different features as: length, aspect ratio, and compact ratio. The length of the vehicle is the dominant feature which can be used to do the coarse classification to classify the vehicle in to a large or a small one. Then we may use other features to do the fine classification, which takes advantage of the second and the third feature. The classification process is shown as follows:

1. Coarse classification identifies the moving object as a large vehicle or a small vehicle.
2. If it is a small vehicle, then calculate the following

$$F_1 = \alpha_1 \bullet \text{(aspect ratio)} + \beta_1 \bullet \text{(compact ratio)}$$

where $\alpha_1=30$ and $\beta_1=1$. If $F_1>TH_{HC}$, then it is a sedan else it can be a van or pick-up.

3. To differentiate the van from the pick-up truck, we examine the height/length ratio (HLR). Since the pickup-truck with framed loading compartment will be higher, we may compare the HLR to find the pickup-truck with framed loading compartment.
4. Differentiate the pickup-truck from the van by comparing the van their compact ratio (CR).
5. Fine classification divides the larger vehicles into two groups: (a) bus and van truck; (b) truck and trailer. The first group has a larger CR than the second group.
5. To separate bus from van truck, compare their HLR.
6. Calculate the following measurement to separate the trailer from the truck as

$$F_2 = \alpha_2 \bullet \text{(length)} + \beta_2 \bullet \text{(area/length)},$$

where $\alpha_1=1$ and $\beta_2=2$. If $F_2>TH_{LC}$, then it is a trailer else it is a truck.

6. Experimental Results and Discussion

In our experiments, the frame resolution of the component color image sequence is 320×240. The image sequences are captured from the freeway traffic using digital CCD video camera. The experimental results of occluded vehicles segmentation are illustrated in Figure 7. When the silhouettes of two vehicles are merged as one, they can be divided as two individual vehicles if they travel with different speeds. The image of the segmented vehicle may not be complete and it causes incorrect classification.

Our segmentation process may fail when the speeds of the two merged vehicles are very close. The other error occurs when the opening or closing operation on the extracted foreground object creates distortion. This kind of errors is often found in the process of the small vehicles. Since the size difference of the silhouettes of the small vehicles is not obvious, if the foreground extraction process is not precise, the extracted small vehicles may not be identified accurately. The correct classification rate of the large vehicles is high, because the silhouettes of large vehicles are much dissimilar from one another. Table 1 illustrates the vehicle classification results of a video sequence lasting 463 seconds with some vehicles occluded. The numbers in bracket are the number of vehicles under occluded conditions.

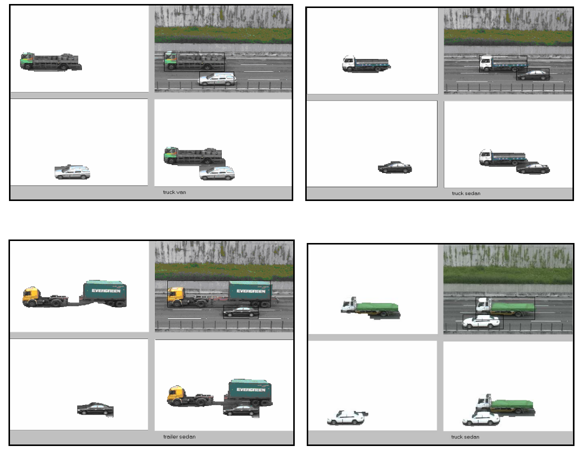


Figure 7. The occluded vehicles are divided.

Type	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*total*	*%*
A	*202 (11)*	*1*	*3*	*0*	*0*	*0*	*0*	*206 (11)*	*98*
B	*4(3)*	*25(3)*	*5(2)*	*0*	*0*	*0*	*0*	*34(8)*	*66*
C	*2*	*2*	*10*	*0*	*0*	*0*	*0*	*14*	*71*
D	*0*	*0*	*0*	*10(4)*	*(1)*	*0*	*0*	*10(5)*	*93*
E	*0*	*0*	*0*	*0*	*6*	*(1)*	*(1)*	*6(2)*	*75*
F	*0*	*0*	*0*	*(1)*	*0*	*9(1)*	*0*	*9(2)*	*91*
G	*0*	*0*	*0*	*(1)*	*0*	*0*	*3(1)*	*3(2)*	*80*

Table 1. The image sequences with occlusions and the overall classification rate is 91.35%.

References

1) B. Coifman, et al., "A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance", Transportation Research-c, December 1, 1998.
2) Y. Jung, and Y. Ho, "Traffic Parameter Extraction using Video-Based Vehicle Tracking", ITS Conference, 1999.
3) B. L. Tseng, C. Lin, and J. R. Smith, "Real-Time Video Surveillance for Traffic Monitoring using Virtual Line Analysis", Proc. of ICME'2002, Vol 2, pp.541-544.
4) I. Lee, H. Ko, and D. K. Han,"Multiple Vehicle Tracking Based on Regional Estimation in Nighttime CCD images", Proceedings of ICASSP '02, Vol. 4, 2002.
5) S. Gupte, et al., "Detection and Classification of Vehicles", IEEE Trans. on ITS, Vol. 3, No. 1, 2002, pp.37-47.
6) A. H. S. Lai, et al., "Vehicle Classification from Visual-Based Dimension Estimation", Proc. IEEE Conf. on ITS – Oakland, USA 2001, pp. 201-205.
7) S. Gupte, et al., "Vision-Based Vehicle Classification", Proc. of ITS Conf., Dearbom, USA – October 1-3, 2000, pp.46-51.
8) C. Zhang and M. Y. Siyal, "A New Segmentation Technique for Classification of Moving Vehicles", Proc. of VTC, pp. 323-326 Vol 1. Tokyo, 2000.
9) C. Harlow and S. Peng, "Automatic vehicle classification system with range sensors", Transportation Research Part C 9(2001), pp.231-247.
10) A. Lai, and N. Yung, "Vehicle-Type Identification Through Automated Virtual Loop Assignment and Block-Based Direction-Biased Motion Estimation", IEEE Trans. on ITS, No.2, June 2000, pp.86-97.