The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	Vehicle-Type Identification Through Automated Virtual Loop Assignment and Block-Based Direction-Biased Motion Estimation
Author(s)	Lai, AHS; Yung, NHC
Citation	Ieee Transactions On Intelligent Transportation Systems, 2000, v. 1 n. 2, p. 86-97
Issued Date	2000
URL	http://hdl.handle.net/10722/42869
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Vehicle-Type Identification Through Automated Virtual Loop Assignment and Block-Based Direction-Biased Motion Estimation

Andrew H. S. Lai and Nelson H. C. Yung, *Senior Member, IEEE*

Abstract—**This paper presents a method of automated virtual loop assignment and direction-based motion estimation. The unique features of our approach are that first, a number of loops are automatically assigned to each lane. The merit of doing this is that it accommodates pan–tilt–zoom (PTZ) actions without needing further human interaction. Second, the size of the virtual loops is much smaller for estimation accuracy. This enables the use of standard block-based motion estimation techniques that are well developed for video coding. Third, the number of virtual loops per lane is large. The motion content of each block may be weighted and the collective result offers a more reliable and robust approach in motion estimation. Comparing this with traditional inductive loop detectors (ILDs), there are a number of advantages. First, the size and number of virtual loops may be varied to fine-tune detection accuracy. Second, it may also be varied for an effective utilization of the computing resources. Third, there is no failure rate associated with the virtual loops or physical installation. As the loops are defined on the image sequence, changing the detection configuration or redeploying the loops to other locations on the same image sequence requires only a change of the assignment parameters. Fourth, virtual loops may be reallocated anywhere on the frame, giving flexibility in detecting different parameters. Our simulation results indicate that the proposed method is effective in type classification.**

Index Terms—**Inductive loop detector, motion estimation, vehicle type identification, virtual loops.**

I. INTRODUCTION

ONE very important aspect of traffic management is the timely acquisition of relevant travel information on the road network. Typical information, such as the average vehicle speed and volume on a particular road segment is especially useful if it is up to date. It provides a glimpse of the segment's utilization and possibly the degree of congestion. Even if it is not up to date, an estimation of the traffic parameters based on historical data is a good enough indication of the network's likely utilization in many cases. In addition to this, the detection of vehicle types in applications such as electronic road pricing is also required for charging purpose, or for the enforcement of road restrictions according to these types. As the application of both requirements is broad, the number of classes in vehicle-type classification is usually small.

Traditionally, vehicle count and even speed are acquired through inductive loop detectors (ILD) [1], [2]. The ILD concept is simple, making use of the induced electromotive force (EMF) principle in electric circuits. However, it has to be physically installed under the surface of the road, and once it is installed, it would be difficult to alter its detection configuration, let alone redeploying it to a new location. Despite the high installation cost, its application is further hindered by its high failure rate (25%–35%) and maintenance cost. Because of these reasons, other type of sensors such as ultrasonic, radar, laser, infrared detectors, and video cameras have also been developed and deployed [2], [3]. For instance, gantry- or roadside-mounted ultrasonic, radar, laser, and infrared detectors have been extensively evaluated. When vehicles pass within the footprint of a sensor, the wave emitted by the sensor hits the vehicle and bounces back. The return wave, as detected by the sensor, can be used to either signal the presence of a vehicle, or calculate the distance of the vehicle from the sensors, from which the height, width, and length of the vehicle may be estimated. Moreover, some of these sensors are able to measure the speed of the vehicles by calculating the time lag between the emitted and return waves. In summary, all these sensors, including the ILD operate on a similar principle.

On the other hand, video or closed-circuit television (CCTV) cameras operate on quite a different principle and have been extensively deployed for visual surveillance in many countries. These cameras are usually roadside- or gantry-mounted at strategic locations on freeways or junctions, with video links to a control center. The video outputs may be inspected by traffic officials, and decisions can be made based on these observations. The attraction of this video surveillance approach is that it offers a far richer information content than that obtained from ILD and similar sensors. This information content may also be analyzed by a computer algorithm at the same time, enabling a fast and effective utilization of the data for traffic management.

However, processing video information creates a new set of problems too. For example, the configuration of the camera affects the field of view, directly determines whether a vehicle can be seen or detected with reasonable accuracy. Besides, the problem of how to detect the presence of vehicles is challenging especially in outdoor environment, which is often hostile and uncontrollable. To make it worse, video processing often requires intensive computation. If real-time processing is part of the requirement, efficient algorithms with fast hardware imple-

mentation are inevitable. If most of these problems can be resolved, then visual surveillance methods truly present an attractive alternative to ILD and other similar sensors.

Since the late 1980s, a number of papers have been published on detecting and/or tracking moving vehicles using image and video-processing techniques [3]–[5]. Dubuisson and Jain [6] classified them into four categories according to how the camera is mounted or whether motion information is used. In general, they can be grouped into two broad classes: model-based approach and those that are not. The model-based approach refers to those approaches that use a parameterized or polygonal model to describe and represent a vehicle. By and large, the model-based approach is the most popular and has been adopted widely by many researchers [6]–[14]. A typical example is the deformable model, in which the vehicle shape and interior attributes are fitted with predefined parameterized models consisting of 20–30 vertices. Such a model can distinguish vehicle types in a limited sense and indicate the vehicle direction if there is no occlusion. However, computation complexity increases with the number of vertices used to define the model, and better accuracy requires more vertices. Often, one of these has to be sacrificed for the other. For polygonal models, the vehicle outline is fitted with a polygon. It does not suffer from occlusion but it does not distinguish the vehicle types, dimensions, and orientations either. For both methods, vehicle motion can be estimated using techniques proposed in [11]–[14]. For non-model-based approaches, there are quite a few cases reported as well [15]–[18]. The simplest approaches are probably those that manually define a bounding area per lane on the video sequence where vehicles can be detected and their motion estimated [19]–[22]. AUTOSCOPE is one of these systems that employs this approach and claims that it can replace ILD directly [19]. The obvious pitfalls of this approach are that the manual definition of bounding area is nontrivial and has to be repeated whenever there are pan–tilt–zoom (PTZ) actions. Inherently, there is also no concept of vehicle or object in this type of methods.

From the above, we observe that automated tracking and modeling methods are usually generic and flexible but complex and computation-intensive. Whereas the non-model-based approaches are simpler but require a high degree of human interaction in their operations. To tackle some of these issues, we propose, in this paper, a method of automated virtual loop assignment and direction-based motion estimation [23]. The unique features of our approach are that first, a number of loops are automatically assigned to each lane. The merit of doing this is that it accommodates PTZ actions without needing further human interaction. Second, the size of the virtual loops is much smaller for estimation accuracy. This enables the use of standard block-based motion estimation techniques that are well developed for video coding. Third, the number of virtual loops per lane is large. The motion content of each block may be weighted and the collective result offers a more reliable and robust approach in motion estimation. Comparing this with traditional ILD, there are a number of advantages. First, the size and number of virtual loops may be varied to fine-tune detection accuracy. Second, it may also be varied for an effective utilization of the computing resources. Third, there

is no failure rate associated with the virtual loops or physical installation. As the loops are defined on the image sequence, changing the detection configuration or redeploying the loops to other locations on the same image sequence requires only the change of the assignment parameters. Fourth, virtual loops may be reallocated anywhere on the frame, giving flexibility in detecting different parameters. Our simulation results indicate that the proposed method is effective in type classification.

This paper is organized as follows. Section II gives a brief presentation of the principle of existing ILD. Section III outlines the proposed method. Section IV presents the virtual loop concept, and how the loops are assigned and reduced automatically. Section V describes the detection of moving vehicles using a direction-based block-based motion estimation technique. Section VI formulates the vehicle identification approach, while the test results are discussed in Section VII. This paper is concluded in Section VIII.

II. PRINCIPLE OF INDUCTIVE LOOP DETECTION

Essentially, ILD are inductive loops buried under the surface of the road junction. In theory, when a metallic object passes over the loop and interferes with the loop's magnetic field, the change in loop current indicates the presence of a vehicle. The magnitude of the current in the loop is dependent on the closeness of the object and the speed of the metallic object passing over it. Therefore, if the vehicle is moving slowly or stops over the loop, very small or no current change will be induced. The meaning of this is that ILD may not be able to detect slow moving vehicles or vehicles that stop on top of it. In order to have an effective detection, the inductive loops are usually buried in grooves on the road surface that resembles the shape of loops. Because of this, ILD are prone to stress and strain when vehicles traverse over it. When ILD are exposed to this condition continuously, breakage of the loop becomes a major reason of failure for the ILD, which requires high maintenance cost as well as disruption of traffic.

ILD may be employed to estimate the speed of the vehicle. To do this, two loops are buried with a known separating distance. From the time difference between the vehicle passing over the first and the second loops, the speed of the vehicle can be estimated. As different types of vehicles have different wheel-base configuration, they will produce different "signatures" in practice. Therefore, ILD has also been used for vehicle-type identification. Vehicle-type identification is based on the signature of individual vehicle recognized by the ILD. The one–dimensional (1-D) signature may be current magnitudes or counts versus time depending on the type of ILD used. For example, a private car would have a 1-D signature that consists of a high and sharp peak, while a truck, because of its wheel-base distance from the ground, would have a flat peak of much lower magnitude but perhaps twice as wide across time [24], as depicted in Fig. 1. As can be seen, using ILD for type classification is viable, but can only differentiate a few broad classes, as in this case, small, medium, and large. Furthermore, the accuracy of the 1-D signature is not very reliable and, therefore, certain degree of misclassification is inevitable. To improve on the drawbacks of ILD, a new method is required.
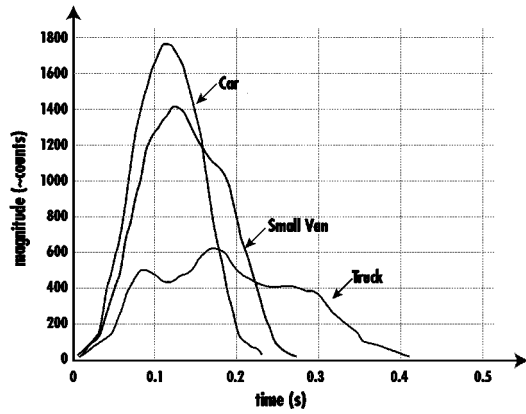
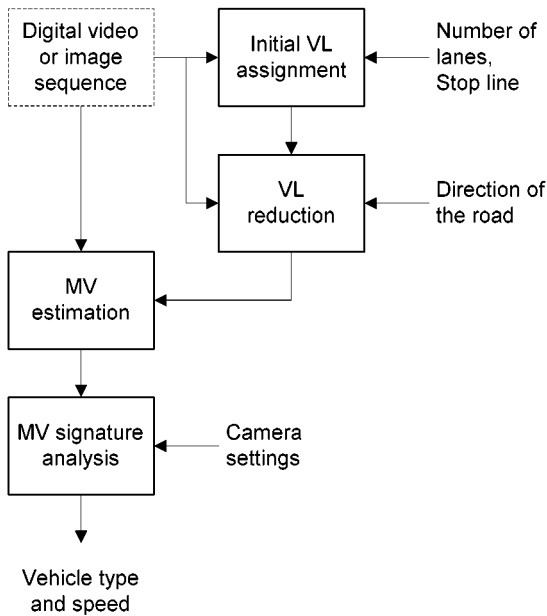Fig. 1.   Vehicle signatures from ILD.

Fig. 2.   Block diagram of the proposed method.

Fig. 3.   Typical image frame considered in this research.

III. OVERVIEW OF THE PROPOSED METHOD

The block diagram of the proposed method is depicted in Fig. 2. It is developed around four major modules: initial virtual loop (VL) assignment, VL reduction, motion estimation, and signature analysis. Assume that the number of lanes, road direction, camera parameters, and a reference line are known *a priori*. The assumption of a known number of lanes and road direction is considered reasonable as these two values would not change even if there were PTZ camera actions. For the camera parameters, if the camera is fixed in height, focal length, the distance from the focal point on the road surface, pan and tilt angles, these parameters can be used directly for the calculation of the vehicle type. However, if there are PTZ actions, the method given in [25] for automatically determining these camera parameters can be adopted. In practice, no PTZ action is allowed during the automated VL assignment, reduction, and vehicle type identification phases, else all the steps must be repeated to locate the correct VLs again.

As for the reference line, it could be the stop line at a junction or any other lines that are drawn across the road, or can be drawn using some landmarks. As for the former, it has been shown in [26] that the stop line can be detected correctly. However, the latter two cases may not be very common. In those cases, if no other lines or landmarks can be found, a worst case scenario may be assumed, that is, the initial automated assignment of VL will cover the entire frame without any spatial restrictions. As a result, the VL reduction phase would require more extensive removal of static or redundant VLs.

Given a video captured by a CCTV or video camera, it is digitized into a sequence of image frames. A typical frame at a junction is depicted in Fig. 3. Based on these image frames and the road parameters (number of lanes, direction of the road), a number of VLs is automatically assigned in an image frame without any motion. This motionless information is easily obtained by considering the dense motion field generated by block-based motion estimation on all the blocks in the frame. This assignment is classified as initial as there is the probability that the VL assignment at this stage may include those regions that would seldom have motion in the subsequent analysis, or they provide redundant motion information. Therefore, a VL reduction method is required to identify and eliminate these VLs, although leaving them in the subsequent calculations should not affect the performance of the method as such. The rationale for VL reduction is for reducing computational complexity rather than performance. It should be noted that severe reduction of the VLs may eventually affect the performance of the method, if some of the dynamic VLs (those representing major vehicle motions) are removed.

Once a set of final VLs is determined, a direction-based block-based motion estimation method is proposed to estimate the motion vectors in those VLs along the direction of the road. This is essential as it will reduce the number of search points needed to find a match. The motion vectors as a result of the motion estimation are then grouped together to form a 1-D representation (or signature) of the vehicle, which may be further analyzed, as to which group does the vehicle belong to. Details of these proposed modules are discussed in the following sections.

IV. VL ASSIGNMENT AND REDUCTION

A. Concept of VLs

The concept of VL is to emulate the functionality of ILD on an image sequence. The image sequence may be acquired from

a roadside- or gantry-mounted camera. Based purely on the sequence, a number of regions within an image frame may be defined and called virtual loops, which can be extended throughout the entire sequence, or redefined at other points in the sequence. The regions may take any size or aspect ratio, although intuitively, the number of regions and the region size should be inversely proportional to each other. Once the VLs are defined, motion estimation methods may be employed to determine motion vectors of objects that traverse over the VLs, as in the case of the relationship between the ILD and the vehicle that traverses over it. Except in this case, the object can be vehicles or other moving objects, such as human beings, since the motion estimation method would not differentiate metallic objects from nonmetallic objects. This capability of being able to detect humans crossing the road adds value to the VL approach. Furthermore, vehicles and other objects that move slowly or even stop could also be detected, whereas they would be missed in the ILD's case. From the motion vector, information regarding the speed, direction, and the length of the object becomes available, from which vehicle type may be identified.

B. Virtual Loop Assignment

The simplest way to automatically assign the VL location is to fill up the whole image frame with VLs. However, this would result in a large number of VLs. For example, a 640×480 pixels image frame would need 4800 VLs of size 8×8. Obviously, processing this large number of VLs is computationally intensive. Moreover, most of these VLs may be located on static regions with little motion throughout the sequence, e.g., sky and pavement. On the other hand, some of these VLs may provide redundant information. Therefore, it is more appropriate to select only a subset of the VLs for processing.

Let us define the ith VL in frame f to be an $N \times N$ block given by

$$s_i(n_x, n_y, f) \qquad (1)$$

for $i = 1, \cdots, M$, where M is the number of VLs in the frame and (n_x, n_y) is the spatial coordinates of the VL center. With reference to the frame shown in Fig. 3, as the stop line can be seen at the junction, the initial assignment of the VLs starts beyond the stop line [26] as it would be reasonable to assume that all the vehicles would traverse across the stop line. The assignment starts at the middle of the image $(H/2)$, where H is the height of the image frame. The VLs are placed downward until the region below $(H/2)$ and above the stop line is filled. Usually, it is unnecessary to closely pack all the VLs as neighborhood VLs are likely to give similar results. In practice, small gaps are left between the VLs. If the stop line is unavailable, the assignment will start at $(H/2)$ and the VLs are placed downward until the entire lower half of the image is filled. This is generally applicable for the image frames as given. In this case, one quarter of the image frame will be filled with VLs. In fact, the initial number of virtual loops and their locations can be arbitrary to some extend as most of these VLs would be removed subsequently. A typical initial assignment relying on the stop line is depicted in Fig. 4.

Fig. 4.   Typical initial VL assignment.

For a two-lane road and the field of view (FOV) as illustrated in Fig. 3, an initial M is found to be between 100 and 200. In fact, M is dependent on several factors:

1) the FOV of the image frame;
2) the number of road lanes;
3) the size of the VL.

For the first point, the FOV directly determines the likely trajectories of vehicles in the frame and their sizes as perceived by the camera. Therefore, if FOV is smaller, then M would increase in order to cover a larger area. If the FOV is larger, then M would be smaller for the opposite reason. For the second point, if the vehicles are going to be detected at different road lanes, different sets of VLs must be assigned to individual lanes. Therefore, more lanes simply means that a larger M is required. As the lane number is known *a priori*, we can set M accordingly. This concept is significantly different from those of [19]–[22] where they only use one assigned region per lane for detection, apart from the fact that they are drawn manually. In the case of the VL size, for a fixed detection region, M is inversely proportional to N^2, i.e.,

$$M = \frac{1}{N^2} \cdot \text{area of detection region}. \qquad (2)$$

As depicted in Fig. 5, by choosing a certain M, we are able to reduce N to suit the particular block-based methods used for motion estimation. In particular, if M is reduced to one per lane, it reverts to the cases as described in [19]–[22]. Although there is no formal criterion for the selection of N, its value should neither be too small nor too large for the following reasons. If N is small $(N < 8)$, the number of pixels within the VL is small, hence the features contained in the VL are limited. In this case, the results obtained from motion estimation may have large mean-square errors due to the mismatch of motion blocks or noise present in the image frame. On the other hand, if N is large $(N > 16)$, each VL may contain numerous features. Given the perspective view in Fig. 3, when the vehicle moves away from the camera, large VL tends to exaggerate the panning out effect of the vehicle, and again would give a poor match in motion estimation. For these reasons, N is usually chosen to be between 8 and 16.

C. Virtual Loop Reduction

After the initial assignment, M is reduced by the following step. The principle of reduction is based on the fact that those
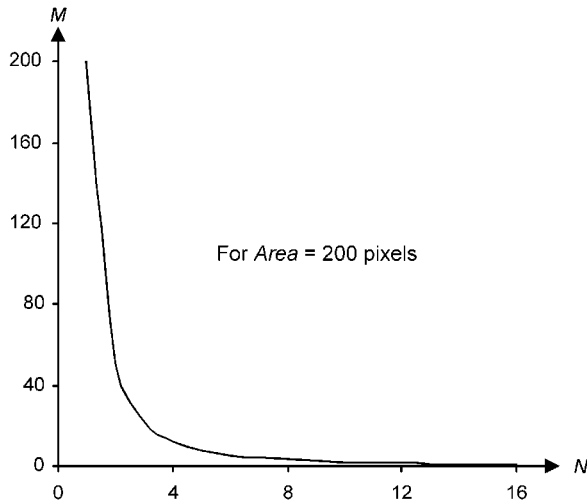
Fig. 5. Relationship between M and N.

VLs with substantial motion would be retained. This can be visualized by considering a busy (plenty of motion) segment of the image sequence, and determining the mean and standard deviation of the motion vector magnitude and direction. If the mean value is large with small standard deviation, then it must denote a detection of part of an object traversing across the FOV. If the mean motion vector direction is also consistently alone the direction of the road, then it reinforces the preceding deduction. Based on this argument, VLs that are characterized by the above properties are retained. The following paragraphs depict the details of the reduction.

Over a short image sequence of F_v frames (10–20 frames) with vehicles passing under the FOV, the MAD (mean absolute difference) [27] of the ith VL between consecutive frames, f and $f + 1$ using full search [27], [28] is determined by the following equation:

$$\text{MAD}(d_x, d_y)$$
$$= \frac{1}{N^2} \sum_{\forall(n_x, n_y)} |s_i(n_x, n_y, f) - s_i(n_x + d_x, n_y + d_y, f+1)|$$
(3)

for a set of candidate motion vectors (d_x, d_y), where $-R \le d_x$, $d_y \le R$, and R is a predetermined integer to define the size of the "search window." The value of R can be determined from the camera setting, frame rate of the image sequence, and the expected maximum speed of the vehicle traversing across the FOV. It should be noted that full search is used in this case without direction bias, which is different from the method presented in Section V.

From (3), the estimated motion vector (MV) corresponding to the ith VL is the value of (d_x, d_y) which minimizes the MAD, that is,

$$v_i(f, f+1) = \arg\left(\min_{(d_x, d_y)} \text{MAD}(d_x, d_y)\right).$$
(4)

Denoting $v_i(f, f+1)$ by v_i, the mean and standard deviation of the MV magnitude $(|v_i|)$ and orientation $(\angle v_i)$ of the ith VL are computed over F_v as follows:

Mean of $|v_i|$:

$$\mu_{|v_i|} = \frac{1}{F_v - 1} \sum_{f=0}^{F_v - 2} |v_i(f, f+1)|$$
(5)

S.D. of $|v_i|$:

$$\sigma_{|v_i|} = \sqrt{\frac{1}{F_v - 1} \sum_{f=0}^{F_v - 2} \left[|v_i(f, f+1)| - \mu_{|v_i|}\right]^2}.$$
(6)

Mean of $\angle v_i$:

$$\mu_{\angle v_i} = \frac{1}{F_v - 1} \sum_{f=0}^{F_v - 2} \angle v_i(f, f+1)$$
(7)

S.D. of $\angle v_i$:

$$\sigma_{\angle v_i} = \sqrt{\frac{1}{F_v - 1} \sum_{f=0}^{F_v - 2} \left[\angle v_i(f, f+1) - \mu_{\angle v_i}\right]^2}.$$
(8)

From (4)–(8), those VLs with MV that satisfy all the following conditions are retained in the final assignment.

Condition 1):   Large mean magnitude and small standard deviation, i.e.,

$$\mu_{|v_i|} \gg \frac{1}{M} \sum_{i=0}^{M-1} \mu_{|v_i|}$$

and

$$\sigma_{|v_i|} \ll \frac{1}{M} \sum_{i=0}^{M-1} \sigma_{|v_i|}.$$

Condition 2):   Mean orientation is in the direction of the road, i.e., $|\mu_{\angle v_i} - \theta_{\text{road}}| < \varepsilon_1$, where θ_{road} is the direction of the road in the image frame and ε_1 is predetermined from the camera setting.

Condition 3):   Small orientation standard deviation, i.e., $\sigma_{\angle v_i} \ll \overline{\sigma}_{\angle v_i}$, where

$$\overline{\sigma}_{\angle v_i} = \frac{1}{M} \sum_{i=0}^{M-1} \sigma_{\angle v_i}$$

which is the mean of the orientation standard deviation.

Condition 1) retains those VLs having consistently large motion, i.e., significant motion due to the motion of one or more vehicles. Condition 2) retains those VLs that lie along the direction of the road. Therefore, VLs having motion due to other objects (such as pedestrians cross the junction) are eliminated. This is based on the fact that pedestrians crossing the junction will produce MV at very different orientation to that of the direction of the road. Condition 3) further eliminates those VL with large orientation deviations. Such deviation in orientation often represents motion due to other sources, for example, motion of shadows.

From the final set of VLs, they are further divided into groups according to the number of lanes the FOV covers. This is achieved by further considering the VL's MV magnitudes in each frame. In the two-lane case presented in Fig. 3, vehicle motion on the outer lane (right) only causes the VLs in the outer lane to experience large MV magnitudes, but not the inner lane (left). On the other hand, vehicle motion in the inner lane may cause large MV magnitudes in both lanes. Over F_v
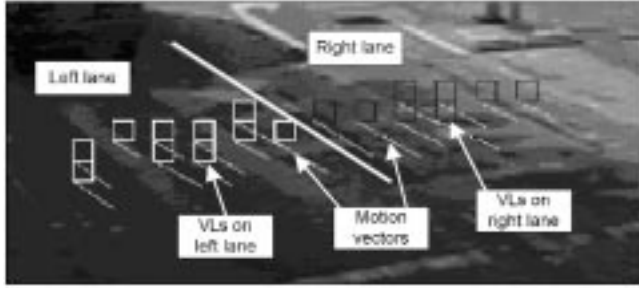
Fig. 6. Final VL assignment after reduction.

frames, based on the above criterion, the VLs may be grouped accordingly. Fig. 6 depicts the VLs that have been divided into two groups for two lanes. In this case, there are ten final VLs in the left lane, and eight in the right lane. It should be noted that if the camera is gantry-mounted, the motion detected in each lane will be separated and, therefore, the division of VLs into groups will be more distinct.

V. VEHICLE MOTION DETECTION

To estimate vehicle motions on the road, the MVs of the final set of VLs are computed using (3). However, for the purpose of computing efficiency, a three-step search is used instead of a full search [27], [29], [30]. In our case, as only the vehicles moving along the direction of the road are of interest, a direction-biased three-step search method is employed for motion estimation. The search points at each step of the direction-biased three-step search are depicted in Fig. 7, assuming that the search window is further constrained by the road direction and its tolerance, which forms the fan-like search areas centered at (n_x, n_y). Typical MVs determined by this method are also given in Fig. 6.

In this search method, the first step is to determine the mean MV magnitude and direction as given in (9) and (10)

$$\rho = \frac{1}{M_0} \sum_{i=0}^{M_0-1} \mu_{|v_i|} \qquad (9)$$

and

$$\beta = \frac{1}{M_0} \sum_{i=0}^{M_0-1} \mu_{\angle v_i} \qquad (10)$$

where M_0 is the number of final VLs. ρ and β define the two first-step search points. For ① as in Fig. 7, it is defined as ρ from (n_x, n_y) and sustained an angle of β in the fan-like window, where the numbers define the search points visited in that particular step and the circled numbers illustrate the possible search points with minimum MAD. For "1," it takes on the values of $\rho/2$ from (n_x, n_y) and sustains an angle of β. The four nearest neighbors of ① are defined as two along β at locations of $3\rho/4$ from (n_x, n_y) and $5\rho/4$ from (n_x, n_y), whereas the other two are ρ from (n_x, n_y) but with direction difference of $\beta \pm \sigma_{|\angle v_i|}$. These points are labeled as "2" and are the four points visited in the seccond step. The four nearest neighbors of ② say, are defined as $\rho' \pm (\rho/8)$ or $\beta' \pm (\sigma_{|\angle v_i|}/2)$. These points are labeled as "3" and are those being visited in the third step. Given this search pattern, the MAD values are calculated and the smallest value is located as a match. The MV of the v_i is then defined by the distance of the point from (n_x, n_y) and its direction.
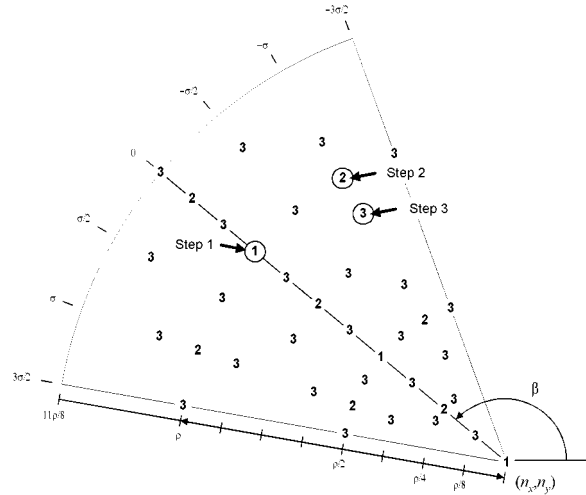

Fig. 7. Direction-biased three-step search.

As can be seen, the search is within the angular direction $\beta \pm (1.5\sigma_{|\angle v_i|})$ and a magnitude range of $11\rho/8$. This eliminates objects that are traveling in a different direction, and of MV magnitude larger than the prescribed boundary. The first restriction seems reasonable as objects moving in other directions are unlikely vehicles of interest in this paper. To combat the issue of detecting objects or vehicles moving in a direction different from the road direction, the same direction-biased search pattern may be applied with β set to a different value. The second restriction limits the speed of the vehicles to be detected. In other words, if the speed of a vehicle is much larger than ρ, it will not be detected by this search pattern. In practice, it is acceptable to set such boundary to the search pattern, even though in some rare cases the search will fail.

VI. VEHICLE-TYPE IDENTIFICATION

To perform vehicle-type identification using the reduced VL set, we propose to analyze the function of the MV magnitude averaged over the nonzero VLs and over time of each lane. In principle, when a vehicle first enters the VLs, the MVs of the VLs can be estimated, and the same applies when it leaves the VLs. The magnitude of all these MVs should give a clear indication of whether the vehicle is entering or leaving the VLs. However, the magnitude of the MVs between these two unique points is dependent on the type of vehicle and possibly other factors. For example, container trucks often have a long trunk of some homogeneous properties, the movement of which is quite difficult to detect. In other words, the MV magnitudes in this case may be small although the truck is moving. On the other hand, fire trucks for instance, often have heterogeneous structure along their trunks that would possible give large MV magnitude when detected. One point for certain is that during this period, the content of the VLs should be different from the road surface. By taking advantage of this knowledge, we can define the movement of a vehicle across the FOV by the following:

1) changes in the average MV magnitude from low to high;
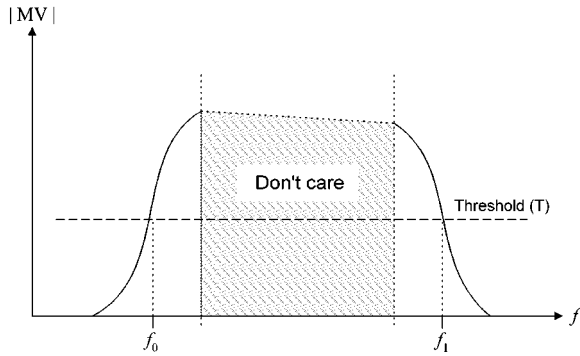2) majority of the VL contents are different from the road surface;

Fig. 8.   Vehicle signature from the VLs.

3) changes in the average MV magnitude from high to low. This can be visualized in Fig. 8. With this framework, we can proceed to define the 1D signature of a vehicle based on motion vector magnitudes.

The 1-D signature of a vehicle is defined as when the average MV magnitude rises above a threshold (T) at frame f_0, a vehicle is said to have entered the VLs. When the average MV magnitude drops below T at frame f_1, the vehicle is said to have left the VL. The threshold T in this case is computed as 50% of the maximum MV magnitude determined. For the average MV magnitude between these f_0 and f_1, as long as the majority of the VL contents are different from the road surface, the actual average MV magnitude is not of great concern. Therefore, they can be considered as "Don't care." From the 1-D signature, the vehicle's type can be identified by its pseudolength (L), which is computed by the following equations:

$$\mu_{|v_i|}(f_0,\, f_1) = \frac{1}{2}\left[\frac{1}{M_0}\sum_{i=0}^{M_0-1}|v_i(f_0,\, f_0+1)| \right.$$
$$\left. +\frac{1}{M_0}\sum_{i=0}^{M_0-1}|v_i(f_1,\, f_1+1)| \right] \quad (11)$$

which is the pseudospeed of the vehicle in the 2-D image coordinates. The pseudolength in 2-D image coordinates is given by (12) and the pseudolength in 3-D world coordinate is given by (13)

$$L_{2D} = \mu_{|v_i|}(f_0,\, f_1) \cdot \left(\frac{f_1-f_0}{f_R}\right) \quad (12)$$

$$L_{3D} = \Phi\left(\mu_{|v_i|}(f_0,\, f_1)\right) \cdot \left(\frac{f_1-f_0}{f_R}\right) \quad (13)$$

where $\Phi(\cdot)$ is the 2-D image to 3-D world coordinates mapping and f_R is the frame rate of the image sequence [11]. Therefore, if the mapping function is available, the speed of the vehicle and its pseudolength can be estimated, which would be useful for road management and law enforcement.

In general, there are a number of properties inherent in this approach. First, only rough type identification based on the vehicle length is possible. At present, this method is unable to distinguish a cargo truck from a bus. Second, only those vehicles that traverse across the VLs along the road direction could be detected. Exceptional cases such as vehicle traveling in an unknown direction will not be detected. As the automated assignment of VLs attempts to generalize the placement of VL, this

drawback is to some degree unavoidable. However, performing motion estimation in different directions such as perpendicular to the road direction (for pedestrian motion), or opposite to the road direction may prevent this to occur. Third, the frame rate of the image sequence must be high enough. Our observation is about 10 frames per second (f/s) or higher. This is because if the frame rate is not high enough, the distance traveled by the vehicle between consecutive frames would be too large. As a result, the search window for motion estimation would have to be much larger, which means the processing delay would be longer and the accuracy of the MV could be lower.

In practice, if vehicles pass over the VLs in a queue and if the camera is roadside-mounted, the occlusion of vehicles may cause the method to fail. However, this problem could be resolved by a gantry-mounted camera with a steep tilt angle of up to 90°. For night scenes, as the proposed approach views vehicles from behind, the rear vehicle lights will simply be viewed as part of the vehicle features, which will have minimal effect to the identification process. In the rare event that a vehicle changes lane while passing over the VLs, its likely the proposed approach will still hold because the final number of VLs is small and it is unlikely that the vehicle can change lane abruptly over such a short distance. On the other hand, if the vehicle stops over the VLs, due to traffic condition or other incidents, then the proposed method will not be able to determine its type.

For shadows cast by the vehicle itself, the effect can be handled as follows. For roadside-mounted camera, a vehicle on the left lane naturally appears to overlap on the right lane in the perspective view, which causes the VLs of both lanes to register motion. On the other hand, a vehicle on the right lane only causes the right lane VLs register motion. If the sun is on the left, then the shadow effect is mild and can be dealt with by considering the difference in motion in both lanes. If the sun is from the right such that both sets of VLs register motion disregarding which lane the vehicle passes over, then the homogenous nature of the shadow on the road and the motion vector magnitudes will have to be considered. From our tests, it appears that the shadow, whether it is from a tree or vehicle, is homogenous in color on the road. When motion estimation is done on the shadow, the internal region of the shadow will have close to zero motion vector magnitude, while the edge of the shadow will have larger magnitudes. The distribution of large magnitude motion vectors is therefore considerably different from those due to the vehicle. For gantry-mounted camera, the same can be applied.

VII. RESULTS

To verify the propose method, we have chosen a traffic video taken from an existing CCTV camera post currently operated by the Hong Kong Police. The post was roughly 3 m high by the side of the road. During the experiment, the camera parameters were fixed. The weather was fine with bright afternoon sun, where a nearby tree cast a large shadow on the road. With a slight breeze during the time of acquisition, the shadow moved back and forth. This sequence is chosen as it represents a realistic outdoor scenario. The traffic video was digitized and processed by our algorithm on a Pentium PC, where virtual loops were automatically assigned, reduced, and divided into groups
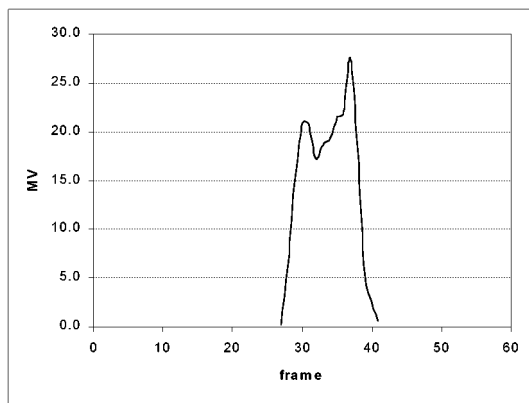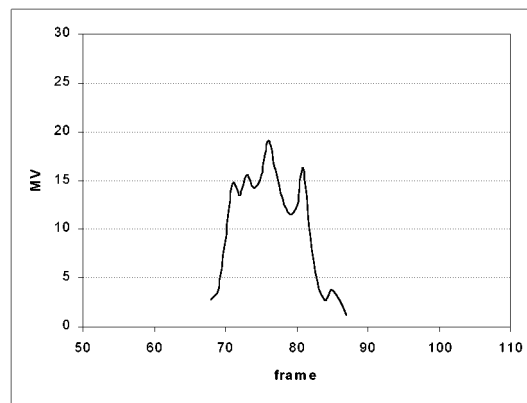
Fig. 9.   1-D signature of a seven-seat van.

Fig. 10.   1–D signature of a fire engine.
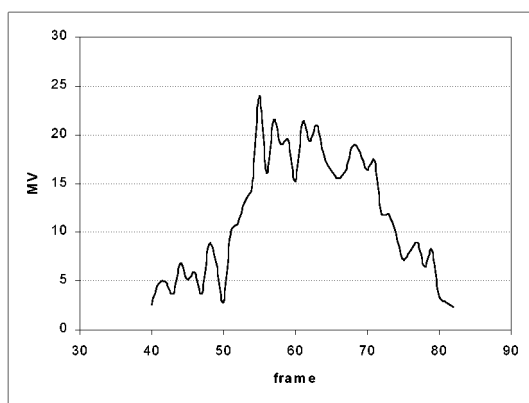

Fig. 11.   1-D signature of saloon car.

Fig. 12.   1-D signature of motocycle.

TABLE I
TYPE IDENTIFICATION RESULT: PSEUDO VEHICLE LENGTH IN 2-D

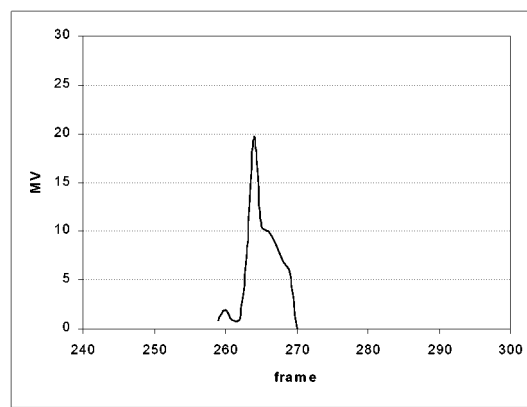|	f_0	f_1	$\mu_{|v_i|}(f_0, f_1)$	L_{2D}
7-seat van	28	37	20.20	18.18
Fire engine	53	74	17.41	36.56
Sedan	71	82	14.74	16.21
Motorcycle	264	268	13.43	5.37

per lane. From the VLs, direction-biased three-step search was performed on each VL and their average MV magnitudes were calculated over time, which are plotted as their 1-D signature. Equations (11) and (12) were used to calculate the pseudospeed and length. Figs. 9–12 depict the MV signatures of four vehicles: a seven-seat van, a fire engine, a sedan, and a motorcycle. Their respective values are given in Table I.

It should be noted that the pseudospeed and length values are given in 2-D values. The reason is that our interest is in the relativity of these values. Actual values in kilometers per hour or meters can only be obtained via a 2-D-to-3-D transformation.

From Figs. 9–12, we can observe that all four 1-D signatures share a set of common characteristics. These are first, the transition from low to high is in accordance with our definition of defining a vehicle entering the VLs. Similarly, the transition from high to low also behaves as predicted. Secondly, the frame time between the two transitions is proportional to the length of the vehicles. As depicted in Table I, for the van, its signature is nine frames wide and has the highest magnitude among the four cases. This indicates that the van is traversing across the VLs at the highest relative speed, which was verified in the image sequence visually, and is also confirmed by the pseudospeed calculation. For the fire engine, its has a wider signature of 21 frames and the magnitude between the two transitions fluctuates substantially. This can be explained as the effect of the complex structure of the fire engine's body on the

VLs. For the sedan, its signature is similar to that of the van except that it is slightly wider (11 frames) and the magnitude is smaller. This indicates that the sedan was traveling at a lower speed than the van. But what is important is that the actual length of the van and the sedan is similar, and their estimated length confirms this point. Finally, for the motorcycle, its signature is narrow and only has one peak, as the length of the motorcycle is short. It is also observed that in the case of the motorcycle, because it is much smaller than the other vehicles, the number of nonzero VLs is small (about 50% of the total VL number). Even though our detection is based on this small number of nonzero VLs, the result still illustrates a clear signature. As there is no vehicle smaller than a motorcycle, it should be reasonable to assume that the motorcycle defines the lower bound of all possible cases. Figs. 13–16 depict the image segments of the four vehicles traversing across the VLs. In these image frames, the VL centers are indicated by dots, where the motion vectors are displayed as line segments from the VL center. Their magnitudes
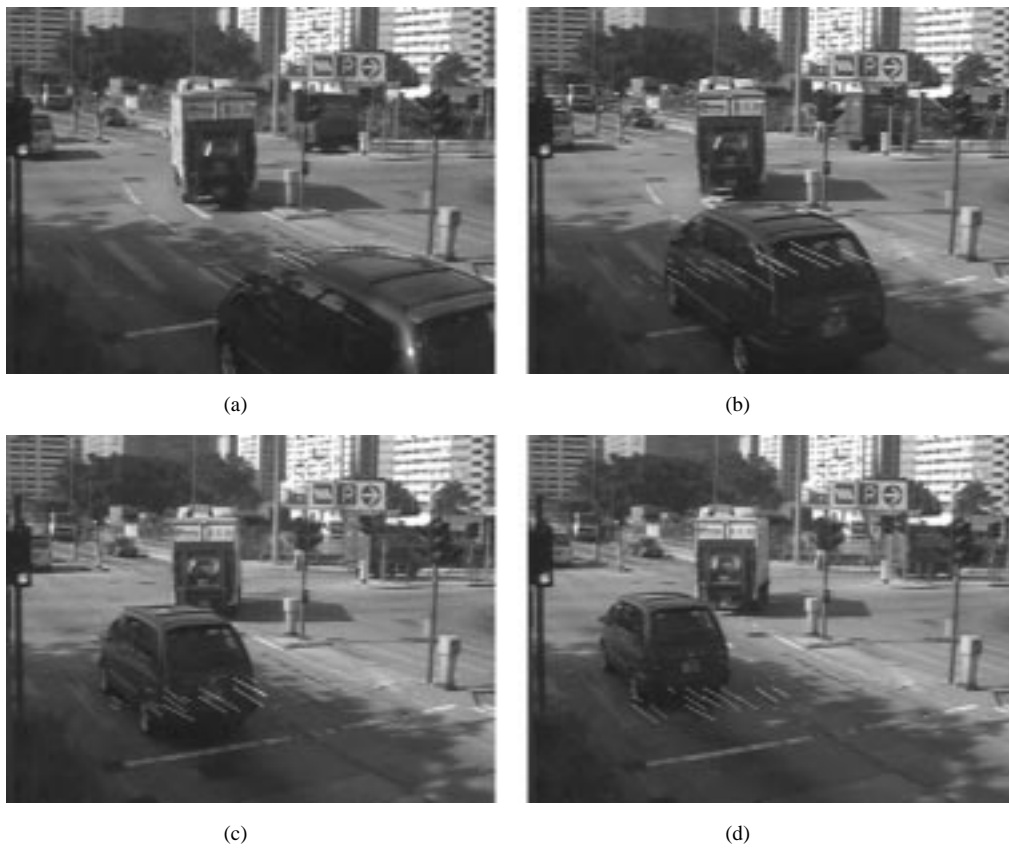
Fig. 13.   Image sequence and MV of a seven-seats van: (a) frame 29, (b) frame 32, (c) frame 35, and (d) frame 37.
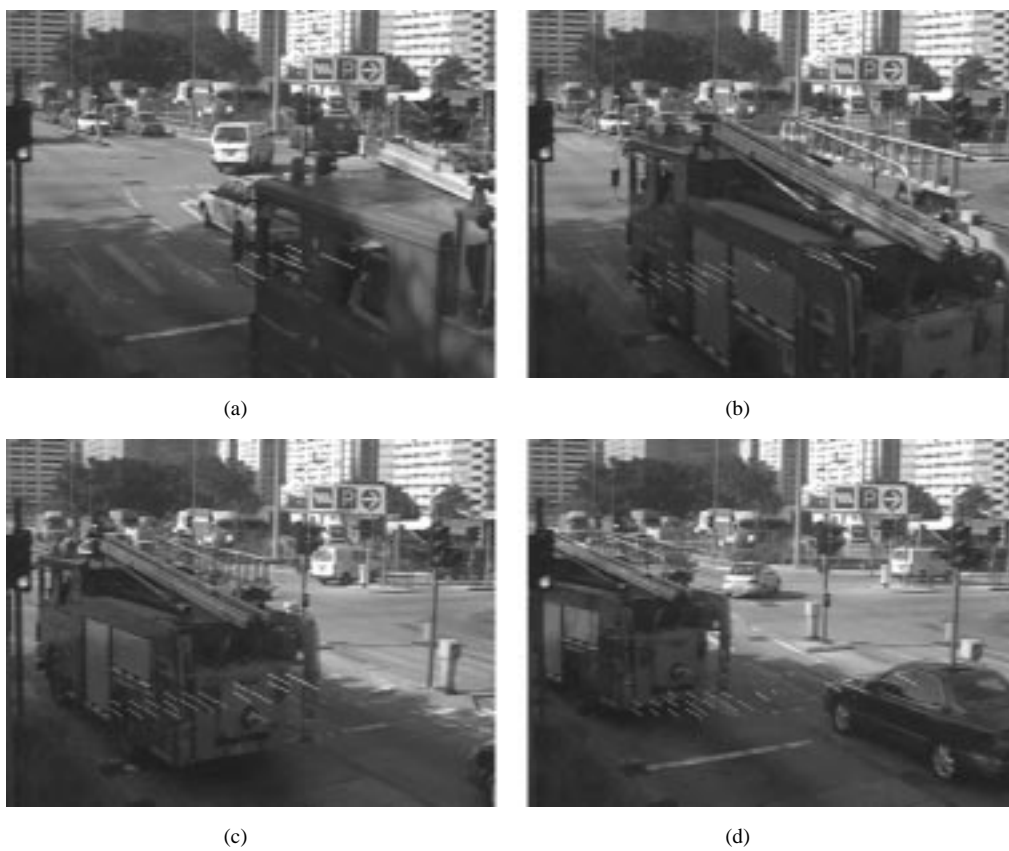

Fig. 14.   Image sequence and MV of a fire engine: (a) frame 51, (b) frame 59, (c) frame 65, and (d) frame 71.
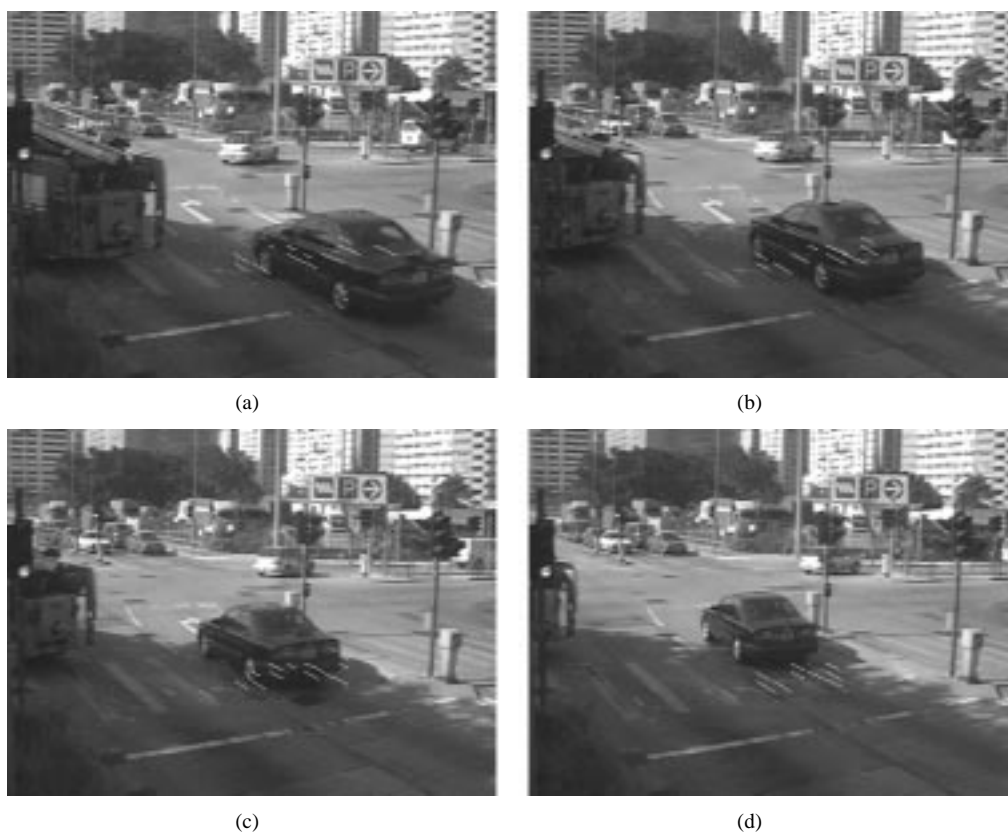
(a)

(b)

(c)

(d)

Fig. 15.   Image sequence and MV of a sedan: (a) frame 74, (b) frame 76, (c) frame 79, and (d) frame 82.
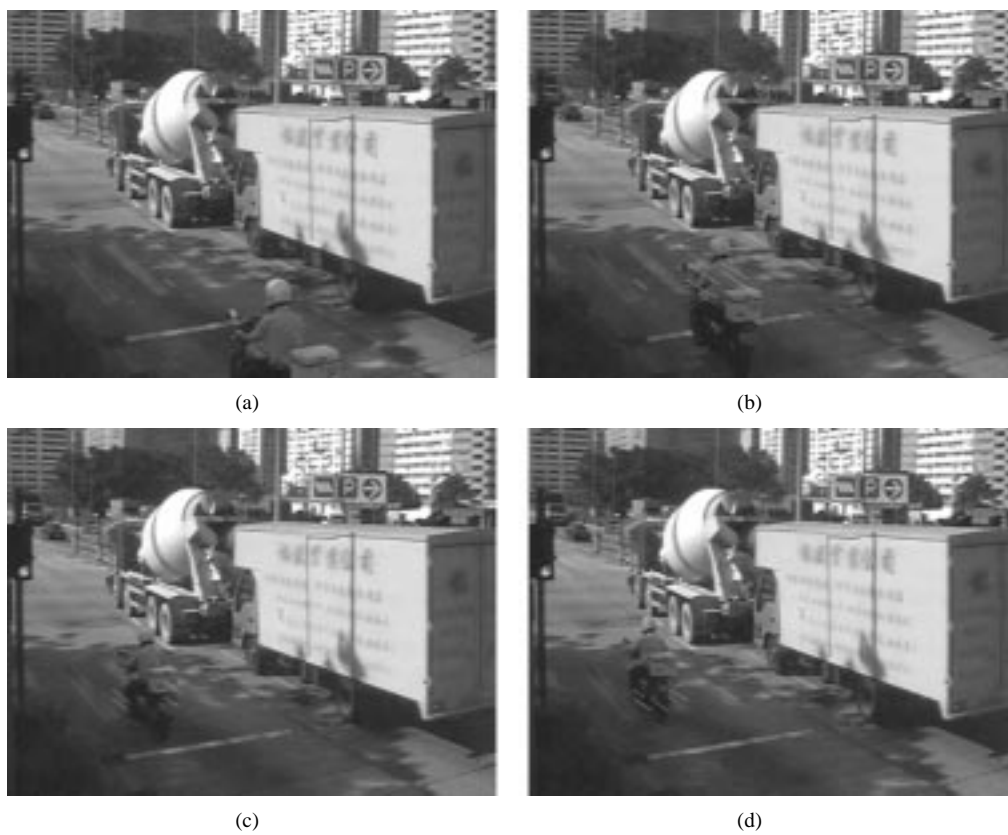

(a)

(b)

(c)

(d)

Fig. 16.   Image sequence and MV of a motorcycle: (a) frame 264, (b) frame 266, (c) frame 268, and (d) frame 269.

can be clearly identified, which correspond to the 1–D signatures shown in the previous four diagrams.

As can be seen from Table I, the pseudolengths in 2-D image coordinates of the vehicles are given. It can be observed that the fire engine has the longest pseudolength in two dimensions, which is almost twice that of the seven-seats van and the sedan. For the motorcycle, its pseudolength in two dimensions is the shortest. Assume that the sedan has an actual length of 4 m, the calculated actual length of the van is about 4.6 m, the fire truck is about 9 m, and the motorcycle is about 1.3 m. These values are considered to be fairly close to the actual lengths of these vehicles. From these values, we are able to classify the vehicles into at least three types (long, medium, and short) according to their pseudolength.

VIII. CONCLUSION

This paper has presented a vehicle-type classification method that is completely vision-based, without needing to rely on inductive-loop detectors or other kinds of sensors. It employs the idea of automated virtual loop assignment and direction-based motion estimation on the image sequence acquired via a roadside- or gantry-mounted CCTV camera. It is characterized by a number of features that are unique to this proposed method. First, the use of CCTV image sequence is flexible and cost-effective as image sequence carries an enormous amount of information good for visual inspection as well as algorithm analysis. Second, it accommodates PTZ actions without requiring manual adjustment or corrections. This is superior to the methods that use single-detection region, which requires human interactions when PTZ occurs. The added advantage of the automated virtual loop assignment process is that it is applicable to multilane cases, with or without any landmarks for the initial assignment. Third, this method offers more accurate motion estimation because the virtual loop size is much smaller than those employed by existing methods. Furthermore, this enables the use of standard block-based motion estimation techniques that are well developed for video coding. Fourth, the number of virtual loops per lane can be varied. The motion content of each block may be weighted and the collective result offers a more reliable and robust approach in motion estimation. Comparing this with traditional ILD, there are a number of advantages. First, the size and number of virtual loops may be varied to fine-tune detection accuracy. Second, it may also be varied for an effective utilization of the computing resources. Third, there is no failure rate associated with the virtual loops or physical installation. As the loops are defined on the image sequence, changing the detection configuration or redeploying the loops to other locations on the same image sequence requires only a change of the assignment parameters. Fourth, virtual loops may be reallocated anywhere on the frame, giving flexibility in detecting different parameters. As for future direction, we would consider expanding the virtual loop concept to detecting pedestrian motion and other vehicle motions such as lane changing.

ACKNOWLEDGMENT

The authors wish to thank the Hong Kong Police Department and the Transport Department of the Hong Kong SAR Government, for help in acquiring the image sequences used for this research.

REFERENCES

[1] N. Ushio and T. Shimizu, "Loop vs ultrasonic in Chicago: Ultrasonic vehicle detector field test isolating diffused reflection and enduring harsh environment," in *Proc. 5th World Congr. Intelligent Transport Systems*, 1998.

[2] S.-W. Kim *et al.*, "Performance comparison of loop/piezo and ultrasonic sensor-based traffic detection systems for collecting individual vehicle information," in *Proc. 5th World Congr. Intelligent Transport Systems*, 1998.

[3] Ashworth *et al.*, "Image processing for traffic monitoring," in *Information Technology Applications in Transport*.   Utrecht, The Netherlands: VNU Science, 1987, pp. 65–86.

[4] N. Hoose, *Computer Image Processing in Traffic Engineering*.   London, U.K.: Research Studies, 1991.

[5] M. Irani and P. Anandan, "A unified approach to moving object detection in 2D and 3D scenes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 577–589, June 1998.

[6] M.-P. Dubuisson and A. K. Jain, "Contour extraction of moving objects in complex outdoor scenes," *Int. J. Comput. Vision*, vol. 14, pp. 83–105, 1995.

[7] D. Koller, J. Weber, and J. Malik, "Robust multiple car tracking with occlusion reasoning," Univ. Calif. Berkeley, UCB Tech. Rep. UCB/CSD-93-780, Oct. 1993.

[8] J. M.-P. Dubuisson *et al.*, "Vehicle segmentation and classification using deformable templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 293–308, Mar. 1996.

[9] W. F. Gardner and D. T. Lawton, "Interactive model-based vehicle tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1115–1121, Nov. 1996.

[10] G. D. Sullivan *et al.*, "Visual object recognition using deformable models of vehicles," in *Proc. Workshop on Context-Based Vision*, 1995, pp. 75–86.

[11] N. H. C. Yung and A. H. S. Lai, "Detection of vehicle occlusion using a generalized deformable model," in *Proc. IEEE ISCAS'98*, 1998, IEEE Catalog 98CH36187, ISBN: 0-7803-4458-8, MPA13-16.

[12] D. Koller *et al.*, "Model-based object tracking in monocular image sequences of road traffic scenes," *Int. J. Comput. Vision*, vol. 10, no. 3, pp. 257–281, 1993.

[13] S. Malik *et al.*, "A machine vision based surveillance system for California roads," Univ. Calif. Berkeley, PATH Project MOU-83, Final Rep., Nov. 1994.

[14] S. Gil *et al.*, "Combining multiple motion estimates for vehicle tracking," in *Computer Vision—ECCV'96*.   Berlin, Germany: Springer-Verlag, 1996, vol. II, pp. 307–320.

[15] B. Gloyer *et al.*, "Video-based freeway monitoring system using recursive vehicle tracking," *SPIE Proc. Image and Video Processing III*, vol. 2421, pp. 173–180, 1995.

[16] F. Bartolini *et al.*, "Motion estimation and tracking for urban traffic monitoring," in *Proc. Int. Conf. Image Processing'96*, vol. III, 1996, pp. 787–790.

[17] J. Malik and S. Russell, "Traffic surveillance and detection technology development: New traffic sensor technology final report," *PATH Publications*, no. UCB-ITS-PRR-97-6, Jan. 1997.

[18] J. C. Rojas and J. D. Crisman, "Vehicle detection in color images," in *Proc. IEEE Conf. Intelligent Transportation Systems*, 1997.

[19] P. Michalopoulos, "Vehicle detection video through image processing: The autoscope system," *IEEE Trans. Veh. Technol.*, vol. 40, Feb. 1991.

[20] B. W. Hwang, T. H. Kang, and J. G. Lee, "A study on the real time measurement of vehicle speed using dynamic image processing," in *Proc. 5th World Congr. Intelligent Transport Systems*, 1998.

[21] N. Hoose, "Impact: An image analysis tool for motorway surveillance," *Traffic Eng. Contr.*, vol. 33, no. 3, pp. 140–147, 1992.

[22] Y. Iwasaki, "An image processing system to measure vehicular queues and an adaptive traffic signal control by using the information of the queues," in *IEEE Conf. Intelligent Transportation Systems '97*, 1997, pp. 195–200.

[23] N. H. C. Yung, K. C. Chan, and A. H. S. Lai, "Vehicle-type identification through automated virtual loop assignment and block-based motion estimation," in *Proc. IEEE/IEEJ/JSAI Int. Conf. Intelligent Transportation Systems*, 1999, pp. 692–696.

[24] C. Sun *et al.*, "Vehicle signature detection using standard loop detectors," *Intellimotion*, vol. 6, no. 2, pp. 4–5, 1997.

[25] A. H. S. Lai, "An effective methodology for visual traffic surveillance," Ph.D. dissertation, Univ. Hong Kong, Jan. 2000.

[26] H. S. Lai and N. H. C. Yung, "A video-based system methodology for detecting red-light runners," in *Proc. IAPR Workshop on Machine Vision Applications*, 1998, ISBN 4-091 122-98-3, pp. 23–26.

[27] A. Murat Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[28] Y. C. Lin and S. C. Tai, "Fast full-search block-matching algorithm for motion-compensated video compression," *IEEE Trans. Commun.*, vol. 45, pp. 527–531, May 1997.

[29] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 438–442, Aug. 1994.

[30] B. Liu and A. Zaccarin, "New fast algorithms for the estimation of block motion vectors," *IEEE Trans. Circuits Syst. Video Technology*, vol. 3, pp. 148–157, Apr. 1993.

Andrew H. S. Lai received the B.Eng. degree from the University of Hong Kong in 1994, the M.Sc. degree from the University of Surrey, U.K., in 1995, and the Ph.D. degree from the University of Hong Kong in 2000.

He was Honorary Assistant Professor at the University of Hong Kong from 1999 to 2000. He is currently a Senior Research Engineer at the Laboratory for Intelligent Transportation Systems Research. His research interest includes visual traffic surveillance, digital image processing, and wireless data communication.

Nelson H. C. Yung (S'82–M'85–SM'96) received the B.Sc. and Ph.D. degrees from the University of Newcastle-Upon-Tyne, U.K., in 1982 and 1985 respectively.

He was Lecturer at the University of Newcastle Upon Tyne from 1985 until 1990, where he was involved in the R&D of digital image processing and parallel processing. From 1990 to 1993, he was a Senior Research Scientist in the Department of Defence, Australia, where he headed a team on the R&D of military-grade signal analysis systems. He joined the University of Hong Kong in late 1993 as an Associate Professor. He currently leads a research group in Digital Image Processing and Intelligent Transportation Systems. He is the Founding Director of the Laboratory for Intelligent Transportation Systems Research, and has published over 90 research papers.

Dr. Yung serves as Reviewer for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *Signal Processing, Proc. IEE Pt. G, SPIE Optical Engineering, HKIE Proceedings* and *Microprocessors and Microsystems Journal*. He is a Chartered Electrical Engineer, Member of the HKIE and IEE. His biography is published in *Who's Who in the World* (Marquis, USA).