Title	Vehicle-type identification through automated virtual loop assignment and block-based direction biased motion estimation
Author(s)	Yung, NHC; Chan, KC; Lai, AHS
Citation	Ieee Conference On Intelligent Transportation Systems, Proceedings, Itsc, 1999, p. 692-696
Issued Date	1999
URL	http://hdl.handle.net/10722/46156
Rights	Creative Commons: Attribution 3.0 Hong Kong License

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

N. H. C. Yung*, K. C. Chan & A. H. S. Lai
Laboratory for Intelligent Transportation Systems Research
The University of Hong Kong
Pokfulam Road, Hong Kong SAR
* nyung@eee.hku.hk

Abstract

This paper presents the concept of automated virtual loop assignment and loop-based motion estimation in vehicle-type identification. A major departure of our method from previous approaches is that the loops are automatically assigned to each lane; the size of virtual loops is much smaller for estimation accuracy; and the number of virtual loops per lane is large. 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 and fully utilize computing resources. Second, there is no failure rate associated with the virtual loops and installation and maintenance cost can be kept to a minimum. Third, virtual loops may be re-allocated anywhere on the frame, giving flexibility in detecting different parameters.

1. Introduction

One very important aspect of traffic management is the timely acquisition of travel parameters on the road network. Information such as vehicle speed and count using a particular road at a particular time provides a glimpse of its flow rate, queue length and even the degree of congestion.

Traditionally, such information is acquired through inductive loop detectors (ILD) [1]. Although the ILD concept is simple and parameters such as speed, count and type classification can be extracted, it needs to be physically installed under the surface of the road. Once it is installed, it would be difficult to alter its detection configuration, let along re-deploying 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. Due to these reasons, other type of sensors such as ultrasonic, radar, laser, infrared detectors and video cameras have also been developed and evaluated in the hope of finding a new and better solution [2]. However, most of these sensors operate in a similar manner as the ILD and acquire similar parameters, except for the video cameras.

Video or CCTV cameras have been extensively deployed for visual surveillance in many countries. These cameras are usually mounted by the roadside or overhead at strategic locations on freeways or junctions, with video links to a control center. The video outputs may be inspected by traffic officials, from which decisions are made. The attraction of this video surveillance approach is that it offers a far richer information content than those obtained from ILD and similar sensors. However, processing video information creates a new set of problems too. If these problems are resolved, then visual surveillance methods present an attractive alternative to ILD.

Since the late 80's, a number of papers have been published on detecting and/or tracking moving vehicles using image and video processing techniques [3-5]. Dubuisson & 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. Model-based here refers to those approaches that use a parameterized or polygonal model to describe and represent a vehicle. By-and-large, model-based approach is the most popular and has been adopted widely by many researchers [6-16].

For most deformable models, the vehicle shape and interior attributes are fitted with pre-defined parameterized models consisting of 20-30 vertices. It can distinguish the vehicle types in a limited sense and indicate the vehicle directions 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. 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 by [14, 17]. For non-model-based approaches, there are quite a few cases reported too [15, 17-19]. 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 [20-23]. AUTOSCOPE is one system that employs this approach and claims that it can replace ILD directly [20].

From the above, we can observe that the automated tracking and modeling methods are usually generic and flexible but complex and computation demanding. Whereas

the non-model-based approaches are simpler but require a high degree of human interaction in their operations especially when zoom, pan and tilt (ZPT) actions are involved.

To tackle these issues, we propose in this paper the concept of automated virtual loop assignment and loop-based motion estimation. A major departure of our method from previous approaches is that first, the loops are automatically assigned to each lane. This allows ZPT actions without needing further region definition manually. Second, the size of virtual loops is much smaller for estimation accuracy. This enables us to use standard block-based motion estimation techniques developed for video coding. Third, the number of virtual loops per lane is large. This 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 and fully utilize computing resources. Second, there is no failure rate associated with the virtual loops and installation and maintenance cost can be kept to a minimum. Third, virtual loops may be re-allocated anywhere on the frame, giving flexibility in detecting different parameters.

This paper is organized in the following: Section 2 presents the virtual loops concept, assignment criteria and their limitations. Section 3 discusses the assigned virtual loops can be used for type identification based on unique 'signatures'. This paper is concluded in Section 4.

2. Virtual loops

The philosophy of virtual loop (VL) is to emulate the functionality of ILD on the image sequence. In essence, 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. As different types of vehicles produce different 'signatures' in theory, therefore ILD may be used for vehicle-type identification.

Figure 1. Typical view from assumed camera mounting.

To emulate the functionality of ILD, VL may be defined as regions within an image frame over the entire image sequence such that processing may be confined to the VL instead of the whole image. The question is how this can be done automatically and what should be the size and number of VL per lane.

To begin with, let us assume the video camera is mounted by the roadside on a 3-meter post, with optical axis along the road direction and its field of view (FOV) covers all the lanes in one travelling direction. The camera settings are assumed to be fixed while the VL are automatically assigned, without ZPT actions. This assignment is expected to be repeated if there are ZPT actions. It is also assumed that the camera operates under normal day light and road conditions. The road direction and the number of road lanes are assumed known a priori. A typical view near a junction with such camera setting is depicted in Figure 1.

Let us define the ith VL in frame f to be an $N \times N$ block given by $s_i(n_x, n_y, f)$ for $i=1,...,M$, where M is the number of VL in the frame and (n_x, n_y) is the spatial coordinates of the VL center. The initial assignment of the VL starts from the line dividing the lanes or from the stop line [24]. If none of these information are available, the assignment will start at $(X/2, Y/4)$, where X and Y are the width and height of the frame respectively, and the origin is assumed at the lower left-hand corner. In fact, the initial number of virtual loops and their locations can be arbitrary to some extend as most of these VL will be removed subsequently. A typical initial assignment is depicted in Figure 2.
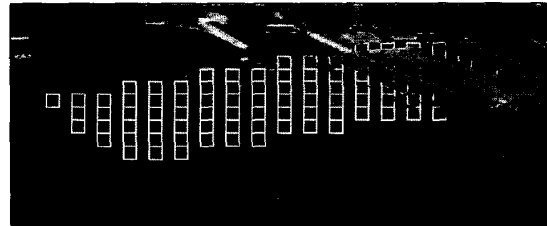

Figure 2. Initial VL assignment M =100, N =15.

For a two-lane road, an initial M is found to be between 100 and 200. As the lane number is known a priori, we can set M accordingly for roads with more lanes. This concept is significantly different from those of [20-23] where they only use one assigned region per lane for detection, apart from the fact that they are drawn manually. By using large M, we are able to reduce N such that block-based or loop-based motion estimation per virtual loop can be performed. In addition, M is inversely proportional to N, where N is upper bounded by 16. Although there is no formal criterion for the selection of N, it is observed that based on the assumed camera setting, N between 13 to 15 gives reasonable detection results. This can be explained as while the vehicle moves away from the camera, large N tends to exaggerate the panning out effect of the vehicle, giving a poor match in motion estimation. On the other hand, if N is small, the number of VL has to increase to ensure a correct detection.

This increases the overall computing delay in motion estimation as well as the latency of the overall detection.

After the initial assignment, M is reduced by the following method. Over a short image sequence of N_v frames (10-20 frames) with vehicles passing under the FOV, the MAD (Mean Absolute Difference) of the ith VL between consecutive frames, f and $f+1$ is determined by the following equation:

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

and $d_x, d_y : |d_x, d_y| < r$, $\left| \angle(d_x, d_y) - \theta_{road} \right| \leq \beta$, where r is the search range and β is a constrained search angle in the calculation. The values of r and β are dependent on the camera view, frame rate and the maximum allowable vehicle speed on that section of the road. For a maximum vehicle speed of 50kph, r is set to 50 pixels and for the assumed camera setting, β is set to 45°. Essentially, the relationship between vehicle speed and r is governed by the frame rate. In this case, r was chosen to be proportional to vehicle speed/frame rate. For 10 frames/sec and 50kph, this works out to be 1.38m/frame. From the camera setting, this is equivalent to just less than 50 pixels. Therefore, r is chosen to be 50. This equation is a modification of the standard MAD calculation. The first reason for doing this is that the direction of the vehicle should be roughly along the road direction in the normal instance. Secondly, with high object speed in traffic video compared with typical coding scenery, the search range is extended to accommodate the possible large motion. From Eqt. (1), the motion vector (MV) corresponding to the ith VL is calculated using the so-called loop-based direction biased three-step search [25]:

$$v_i(f, f+1) = (d_x, d_y)^T$$
$$= \arg\left(\min_{(d_x, d_y)} MAD(d_x, d_y) \right) \quad (2)$$

The mean and standard deviation of the MV magnitude ($|v_i|$) and orientation ($\angle v_i$) of the ith VL are computed over N_v as follows:

$$\mu_{|v_i|} = \frac{1}{N_v - 1} \sum_{f=0}^{N_v - 2} |v_i(f, f+1)| \quad (3a)$$

$$\sigma_{|v_i|} = \sqrt{\frac{1}{N_v - 1} \sum_{f=0}^{N_v - 2} \left[v_i(f, f+1) - \mu_{|v_i|} \right]^2} \quad (3b)$$

$$\mu_{\angle v_i} = \frac{1}{N_v - 1} \sum_{f=0}^{N_v - 2} \angle v_i(f, f+1) \quad (3c)$$

$$\sigma_{\angle v_i} = \sqrt{\frac{1}{N_v - 1} \sum_{f=0}^{N_v - 2} \left[\angle v_i(f, f+1) - \mu_{\angle v_i} \right]^2} \quad (3d)$$

From Eqt. (2) and (3), those VL with MV that satisfy all the following conditions are retained in the final assignment:
(1) Large mean magnitude and small standard deviation;
(2) Its mean orientation is in the direction of the road;
(3) Its orientation standard deviation is small.

Condition (1) retains those VL that has consistent large motion, i.e., significant motion due to vehicle. Condition (2) retains those VL that lie along the direction of the road direction. The VL that have motion due to other objects are eliminated. Condition (3) further eliminates those VL with large orientation deviations. Such deviation in orientation often represents motion due to other sources, for example, shadows.

From the final set of VL, they are further divided into groups according to the number of lanes the FOV covers. This is achieved by considering their MV magnitudes in each frame. In the two-lane case presented in Figure 1, vehicles motion on the outer lane (right) only cause the VL in the outer lane to record large MV magnitudes, but not the inner lane (left). On the other hand, vehicles motion in the inner lane may cause large MV magnitudes in both lanes. Over N_v, based on the above conditions, the VL can be divided accordingly. Figure 3 depicts these VL that have been divided into two groups for two lanes. In this case, there are 8 final VL in the left lane, and 10 in the right lane. It should be noted that if the camera is overhead mounted, the motion detected in each lane will be separated and therefore the division of VL in to groups will be more distinct.
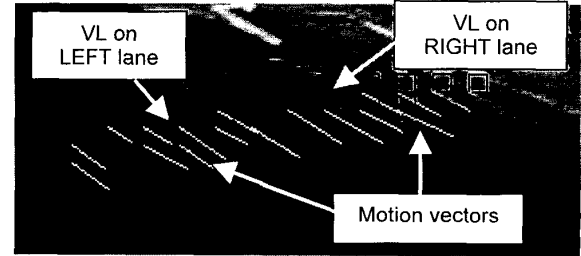

Figure 3. Final VL assignment.

To estimate vehicle motions on the road, we first compute the MV, $v_i(f, f+1)$ (Eqt. (2)) over the final set of VL. The resulting MV are also depicted as straight line segments in Figure 3.

3. Vehicle-type identification

In the case of ILD, vehicle-type identification is based on the signature of individual vehicle recognized by the ILD. The 1D signature may be current magnitudes or counts versus time depending on the type of ILD used. For example, a private car would have a 1D signature that consists of a high and sharp peak, while a truck would have a flat peak of much lower magnitude but perhaps twice as wide over time (Figure 4).
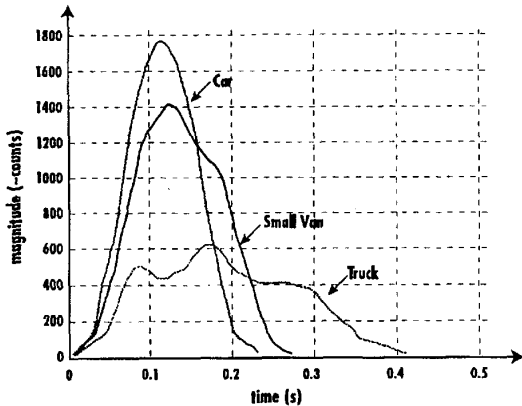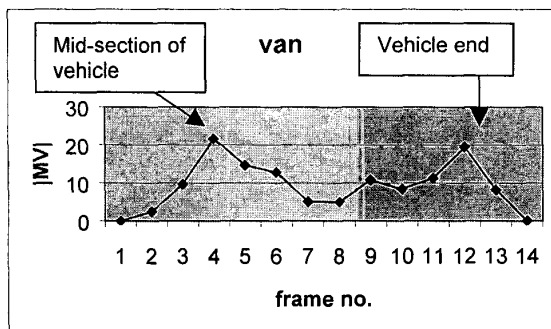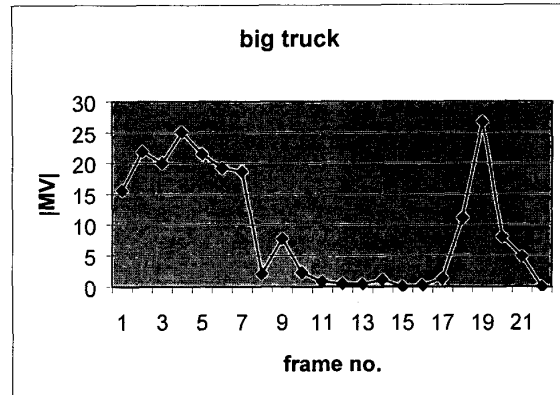
Figure 4. Vehicle signatures from ILD.

To perform the same functionality using the virtual loops, we propose to analyze the function of the averaged MV magnitude over time for each lane. In theory, when a vehicle moves across the VL, MV are estimated in the loops, and their magnitudes will only diminish when the vehicle has passed. Therefore, by taking an average of the VL in each lane, we are able to detect the presence of vehicle as in the case of ILD, as well as extracting its 'signature' by compiling and analyzing this information over time.

In order to verify this approach, we have chosen a traffic video taken from an existing camera post at the Wang Chiu Road, Kowloon Bay, Hong Kong. The post was roughly 3m high by the side of the road. During the experiment, the camera setting was fixed and the time of recording was chosen to be a sunny afternoon, where a large shadow from a nearby tree is evident on the road. Such sequence is chosen to represent a realistic and practical scenario.

The traffic video was subsequently digitized and processed by our proposed algorithm on a Pentium PC, where virtual loops were automatically assigned, reduced and divided into groups per lane. From the VL, loop-based, direction biased three step search was performed on each VL and their average per lane was calculated over time. From these results, Figure 5 depicts the MV signatures of three vehicles: car, van and truck.

(a) 'Van'

(b) 'Truck'
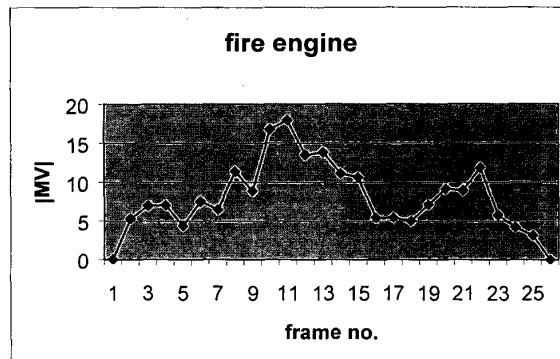
(c) 'Fire engine'
Figure 5. MV signatures of vehicles.

From Figure 5, it can be observed that all three MV signatures have common and unique characteristics. The common characteristic is the twin peak in all cases, which the left peak denotes the mid-section of the vehicle and the right peak denotes the vehicle end. For the van, its signature is 13 frames, while the big truck has 21 frames, and the even bigger fire engine has 25 frames. In this case, the difference between the truck and fire engine may be arguable, but the van can be clearly identified.

4. Conclusions

From the above results, we can conclude that the virtual loop concept is an effective mean for vehicle-type identification. The method proposed here offers an automated approach to virtual loop assignment and reduction. Based on motion estimation, the final set of virtual loops works effectively and is road independent. The selection of larger number of loops and smaller loop size compared with existing manual methods also enable us to use a loop-based direction biased three step search to construct a motion vector signature of the vehicle. The unique description of the vehicle type allows us to identify vehicle types reasonably easily. This is verified by our experiment of analyzing a real and practical traffic video. Future direction in this research will be focused on the

further sub-division of vehicle types and a more extensive analysis of the signatures.

5. References

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

[3] Ashworth, et al (1987), "Image Processing for Traffic Monitoring", Information Technology Applications in Transport, VNU Science Press, Utrecht, Netherlands, pp.65-86, 1987.

[4] Hoose N. (1991), "Computer Image Processing in Traffic Engineering", *Research Studies Press Ltd.*, London, 1991.

[5] Irani M & Anandan P (1998), "A unified approach to moving object detection in 2D and 3D scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.20, no.6, pp.577-589.

[6] Dubuisson M-P & Jain A. K, (1995), "Contour extraction of moving objects in complex outdoor scenes", International journal of computer vision, 14, 83-105.

[7] Koller, D. et al. (1993a), "Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes", *Int. Journal of Computer Vision,* vol. 10, no. 3, pp. 257-281, 1993.

[8] Dubuisson Jolly M-P, et al (1996), "Vehicle Segmentation and Classification Using Deformable Templates", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 18, no. 3, pp. 293-308, 1996.

[9] Gardner W.F. & Lawton D.T., (1996), "Interactive model-based vehicle tracking", IEEE Trans. On PAMI, Vol.18, No.11, Nov 1996, pp.1115-1121.

[10] Sullivan, G.D. et al. (1995), "Visual Object Recognition Using Deformable Models of Vehicles", *Proc. Workshop on Context-Based Vision,* pp. 75-86, 1995.

[11] Yung N. H. C. & A. H. S. Lai (1998b), "Detection of Vehicle Occlusion using a generalized deformable model", *Proceedings of the IEEE ISCAS'98,* IEEE Catalog # 98CH36187, ISBN: 0-7803-4458-8, MPA13-16, 1998.

[12] Koller D, J. Weber, and J. Malik (1993b), "Robust Multiple Car Tracking with Occlusion Reasoning", *UCB Technical Report UCB/CSD-93-780,* University of California, October, 1993.

[13] Malik, S. et al (1994), "A Machine Vision Based Surveillance System for California Roads", *PATH Project MOU-83 Final Report,* University of California, November, 1994.

[14] Gil, S. et al. (1996), "Combining Multiple Motion Estimates for Vehicle Tracking", *Computer Vision – ECCV'96,* vol. II, Springer, Germany, pp. 307-320, 1996.

[15] Gloyer, B. et al. (1995), "Video-based Freeway Monitoring System using Recursive Vehicle Tracking", *SPIE Proceedings on Image and video processing III,* vol. 2421, pp. 173-180, 1995.

[16] HIDO (1997), "Intelligent Transport Systems handbook in Japan", Highway Industry Development Organization, Japan, 1997.

[17] Bartolini, F. et al. (1996), "Motion Estimation and Tracking for Urban Traffic Monitoring", *Proceedings of International Conference on Image Processing'96,* vol. III, pp. 787-790, 1996.

[18] Malik, J & S. Russell (1997), "Traffic surveillance and detection technology development: new traffic sensor technology final report", PATH publications, UCB-ITS-PRR-97-6, Jan 1997.

[19] Rojas J. C. and J. D. Crisman (1997), "Vehicle Detection in Color Images", *Proc. IEEE Conference on Intelligent Transportation Systems,* 1997.

[20] P. Michalopoulos, "Vehicle Detection Video Through Image Processing: the Autoscope System", IEEE Transactions on Vehicular Technology, vol. 40, no. 1, 1991.

[21] B. W. Hwang, T. H. Kang and J. G. Lee, "A Study on the Real Time Measurement of Vehicle Speed Using Dynamic Image Processing", 5th World Congress on Intelligent Transport Systems, 1998.

[22] N. Hoose, "Impact: An Image Analysis Tool for Motorway Surveillance", Traffic Engineering and Control, vol. 33, no. 3, pp. 140-147, 1992.

[23] Y. Iwasaki, "An Image Processing System to Measure Vehicular Queues and an Adaptive Traffic Signal Control by Using the Information of The Queues", IEEE Conference on Intelligent Transportation Systems '97, pp. 195-200, 1997.

[24] Lai H. S. & N. H. C. Yung (1998d), "A video-based system methodology for detecting red-light runners", *Proceedings of the IAPR Workshop on Machine Vision Applications,* ISBN 4-091122-98-3, 1998, pp.23-26.

[25] Chung H. Y., P. Y. S. Cheung & N. H. C. Yung (1998), "Adaptive search center non-linear three step search", *Proceedings of the IEEE ICIP'98,* 1998.