# Vehicle Type Classification from Visual-Based Dimension Estimation

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

**Abstract**—This paper presents a visual-based dimension estimation method for vehicle type classification. Our method extracts moving vehicles from traffic image sequences and fits them with a simple deformable vehicle model. Using a set of coordination mapping functions derived from a calibrated camera model and relying on a shadow removal method, vehicle's width, length and height are estimated. Our experimental tests show that the modeling method is effective and the estimation accuracy is sufficient for general vehicle type classification.

**Index Terms**-- vehicle modeling, shadow removal, camera calibration, dimension estimation

## I. INTRODUCTION

Vehicle type classification is a primary goal of most traffic surveillance tasks [1-3]. In the past, it is mainly employed in auto-toll systems to determine the charge for different types of vehicle. Traditional methods employ buried inductive loops and/or axle sensors to measure the length and the number of axles of the vehicles [1-4]. From that, more than 10 types of vehicle can be distinguished, e.g. motorcycle, car, van, bus, truck, truck+trailer and etc. On the other hand, large-scale deployment of traffic surveillance cameras and rapid development of image processing hardware and software has drawn the attention of using visual-based methods for vehicle type classification [5-7]. The major advantage of visual-based methods is that no buried inductive loops or axle sensor is required, i.e. no major civil work. Therefore, it is more flexible and can be massively deployed with the surveillance cameras.

In visual-based methods, vehicle models are usually used to describe the vehicles. Vehicle models can be as simple as a bounding box [8] or can be as complex as a detail deformable model [9-10]. In theory, once the vehicle is modeled, vehicle length can be estimated, which can then be used to classify the vehicle type. In practice, the accuracy of such approach is hindered by three factors. The first factor is how the vehicle is modeled. A more complex model tends to give more accurate estimation but demands more computation resource, whereas simple model tends to be less accurate. In general, a modeling method that can fit the vehicle in all orientation is necessary. The second factor

is that vehicle shadow is a major issue needs to be resolved. A good shadow removal algorithm is essential in this case. The third factor is that the accuracy in determining the real-world dimensions from the 2D image coordinates is critical. Therefore, a realistic camera model and accuracy coordination mapping functions are inevitable.

In this paper, we propose a vehicle dimensions estimation method for vehicle type classification. In the next section, the methodology of the proposed method is presented. Then, the camera model employed is described in section III. In section IV, the procedure to extract binary vehicle mask is described. The vehicle modeling and dimensions estimation methods are detailed in section V and section VI. Section VII evaluated a number of experimental results. Finally, we conclude our method in section VIII.

## II. METHODOLOGY

The concept of the proposed method is to fit a 2D projection of a simple 3D model on a binary representation of the vehicle extracted from the traffic image sequence, from which dimension is determined. This binary representation is called a vehicle mask. The vehicle mask is extracted by subtracting the traffic image sequence with an estimated stationary background [11]. After that, a shadow removal method is employed to remove any cast shadow extracted together with the vehicle mask [12]. Then, vehicle modeling is performed in 2D. In order to do this, a calibrated camera model is used to provide a mapping between the 3D world coordinates and the 2D image coordinates [13]. Lastly, vehicle dimension including width, length and height are calculated in 3D world coordinates. From the vehicle dimension, vehicles are classified into their respective types. Fig. 1 depicts the conceptual diagram of the proposed method.
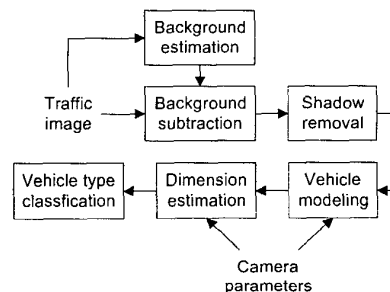


Fig. 1 Conceptual diagram of the proposed method.

The authors are with the Laboratory for Intelligent Transportation Systems Research, The University of Hong Kong, Pokfulam Road, Hong Kong. Email: hslai@eee.hku.hk.

## III. CAMERA MODEL

Fig. 2 depicts the camera model employed. This model can suit different camera setting and orientations [13]. The relationship between the image coordinates and the world coordinates can be defined in terms of pan angle $p$, tilt angle $t$, swing angle $s$, focal length $f$, camera distance $l$.
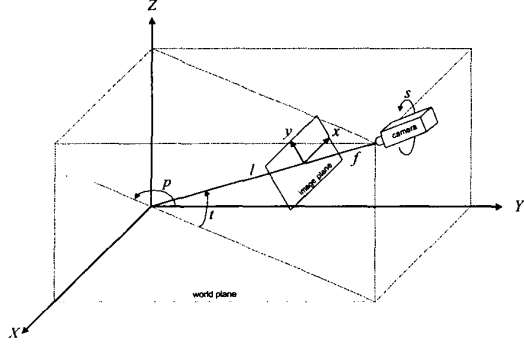


Fig. 2 Camera model.

Normally, some of these camera parameters are unknown because of possible pan-tilt-zoom actions. However, these parameters must be known in order to determine the mapping functions between the 2D image coordinates and the 3D world coordinates. Therefore, the following paragraphs describe how they are estimated.

The forward mapping, $\Phi$, of a point, $\mathbf{Q} = (X_Q, Y_Q, Z_Q)$, in the world coordinates to a point, $\mathbf{q} = (x_q, y_q)$, from the image coordinates is defined as

$$\mathbf{q} = \Phi\{\mathbf{Q}\}, \tag{1}$$

where

$$x_q = \frac{f \cdot \begin{pmatrix} X_Q(\cos p \cos s + \sin p \sin t \sin s) \\ + Y_Q(\sin p \cos s - \cos p \sin t \sin s) \\ + Z_Q \cos t \sin s \end{pmatrix}}{-X_Q \sin p \cos t + Y_Q \cos p \cos t + Z_Q \sin t + l},$$

$$y_q = \frac{f \cdot \begin{pmatrix} X_Q(-\cos p \sin s + \sin p \sin t \cos s) \\ + Y_Q(-\sin p \sin s - \cos p \sin t \cos s) \\ + Z_Q \cos t \cos s \end{pmatrix}}{-X_Q \sin p \cos t + Y_Q \cos p \cos t + Z_Q \sin t + l}. \tag{2}$$

The corresponding inverse mapping, $\Phi^{-1}$, is defined as

$$\mathbf{Q} = \Phi^{-1}\{\mathbf{q}, Z_Q\}, \tag{3}$$

where

$$X_Q = \frac{\begin{pmatrix} \sin p(l + Z_Q \sin t)(x_q \sin s + y_q \cos s) \\ + \cos p(l \sin t + Z_Q)(x_q \cos s - y_q \sin s) \\ - Z_Q f \cos t \sin p \end{pmatrix}}{x_q \cos t \sin s + y_q \cos t \cos s + f \sin t},$$

$$Y_Q = \frac{\begin{pmatrix} -\cos p(l + Z_Q \sin t)(x_q \sin s + y_q \cos s) \\ + \sin p(l \sin t + Z_Q)(x_q \cos s - y_q \sin s) \\ + Z_Q f \cos t \cos p \end{pmatrix}}{x_q \cos t \sin s + y_q \cos t \cos s + f \sin t}. \tag{4}$$

and $Z_Q$ is the assumed height of point $\mathbf{Q}$ in world coordinates. If point $\mathbf{Q}$ lies on the ground, $Z_Q$ is becomes zero.

The calibration of the camera is achieved by choosing a rectangle $ABCD$ from the traffic image as depicted in Fig. 3, which can be readily determined by the road lane [13].
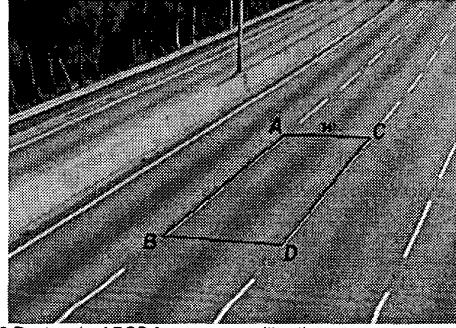


Fig. 3 Rectangle $ABCD$ for camera calibration.

In 3D, it is reasonable to assume that $\overrightarrow{AB} \mathbin{/\mkern-5mu/} \overrightarrow{CD}$ and $\overrightarrow{AC} \mathbin{/\mkern-5mu/} \overrightarrow{BD}$, where rectangle $ABCD$ satisfies Eqt. (5).

$$X_A = X_C, \tag{5a}$$

$$X_B = X_D, \tag{5b}$$

$$Y_A = Y_B, \tag{5c}$$

$$Y_C = Y_D, \tag{5d}$$

$$Y_C - Y_A = w, \tag{5e}$$

where $w$ is the width of the road and $(X_i, Y_i)$ is the world coordinate of the corners, with $i = A, B, C, D$. By substituting Eqt. (4) into Eqt. (5), we have 5 equations to solve $p, t, s, f$ and $l$.

## IV. BINARY VEHICLE MASK

We employ the background subtraction approach as given in [11] to obtain the vehicle masks of the moving vehicles in the traffic image sequence. First of all, a stationary

background image (without vehicles on the road) is estimated from the traffic image sequence. Fig. 4a depicts a typical traffic image and its corresponding estimated background image (Fig. 4b). By taking the difference between the background image and the traffic image sequence, we have a moving vehicle extracted as depicted in Fig. 4c. However, as the color and intensity of the vehicle shadow is also different from the background, it is also extracted as part of the vehicle. If the shadow is not removed, it will affect the accuracy of vehicle modeling and hence introduce huge error in the estimated vehicle dimension. On this regard, the shadow is removed by applying a shadow removal algorithm [12]. In essence, by realizing the various characteristics of shadow in luminance, chrominance, and gradient density, a shadow confidence score for each pixel is calculated. It indicates the probability of a pixel being a cast shadow. Further analysis of the neighboring edge pixel density about each pixel enables the algorithm to classify the pixels into shadow pixels or non-shadow pixels. Shadow pixels are then discarded in this case. After this, convex hull is applied to the remaining non-shadow pixels, which produces the final vehicle mask as depicted in Fig. 4d.
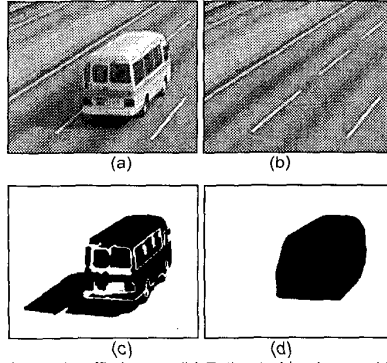


(a)    (b)

(c)    (d)

Fig. 4 (a) Typical traffic image. (b) Estimated background image. (c) Subtraction result. (d) Final vehicle mask after shadow removal and convex hull.

## V. VEHICLE MODELING

The vehicle model is a 3D cuboid wire frame projected on a 2D plane as depicted in Fig. 5 by the solid lines, where $W$, $L$ and $H$ are the width, length and height respectively. A model of this type may be used to model different shapes and sizes of vehicles under different camera parameters. As the model is a 3D cuboid, any object (including a vehicle) may be enclosed in it by varying the dimension of the model. In other words, it can be easily deformed to model the vehicles when their sizes change in view.

In principle, the model is characterized by 6 vertices, $v_i = (x_i, y_i)$, $i = 1,...,6$, which are parameterized such that their values can be adjusted to fit the vehicle mask.
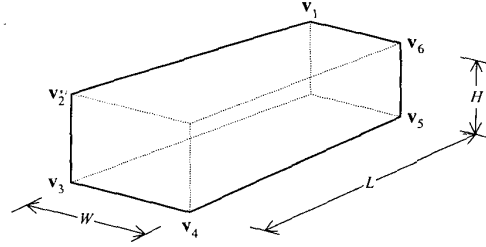


Fig. 5 Vehicle model.

Furthermore, the model consists of 6 edges defined by the 6 vertices, for example the edge $\overrightarrow{v_1v_2}$ joins $v_1$ and $v_2$. From Fig. 5, edges $\overrightarrow{v_3v_2}$ and $\overrightarrow{v_5v_6}$ are in parallel and their orientation, $\theta_Z$, is parallel to the vertical axis in 3D, which can be determined from Eqt. (6) using the forward mapping.

$$\overrightarrow{e_Z} = \Phi\{\overleftrightarrow{E_Z}\},\qquad(6a)$$

$$\theta_Z = \angle\{\overleftrightarrow{e_Z}\},\qquad(6b)$$

where $\overrightarrow{E_Z} = \{0,0,1\}^T$, which is a vertical unity vector the in world coordinates and $\overrightarrow{e_Z}$ is the corresponding vector in 2D. For edges $\overrightarrow{v_3v_4}$ and $\overrightarrow{v_1v_6}$, they are also in parallel and is more or less perpendicular to the road direction in 3D. To obtain their orientation, $\theta_X$, the road direction, $\theta_{road}$, in 2D is first mapped to 3D. Conceptually, by rotating it by $\pi/2$ and then apply the forward mapping, $\theta_X$ for the model is obtained. To do this, first we compute the pseudo center-of-gravity, $g$, of the vehicle mask and compose a point, $q$, in 2D such that the line $\overrightarrow{gq}$ is parallel to $\theta_{road}$ as depicted in Fig. 6.
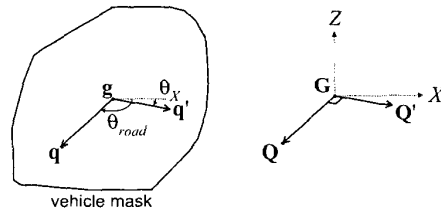


vehicle mask

Fig. 6 Obtain $\theta_X$ from road direction.

$$q = g + a \cdot \begin{pmatrix} \cos\theta_{road} \\ \sin\theta_{road} \end{pmatrix},\qquad(7)$$

where $a$ is the length of the line, which is arbitrary. The line is then mapped to the 3D coordinates using Eqt. (8).

$$G = \Phi^{-1}\{g\},\qquad(8a)$$

$$Q = \Phi^{-1}\{q\}.\qquad(8b)$$

The line $\overrightarrow{GQ}$ in 3D is rotated by $\frac{\pi}{2}$ about $G$ using Eqt. (9) and then mapped back to the 2D coordinates using Eqt. (3).

$$Q' = (Q - G)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + G, \qquad (9)$$

$$q' = \Phi\{Q'\}. \qquad (10)$$

Then,

$$\theta_X = \angle\overrightarrow{gq'}. \qquad (11)$$

Because of the perspective view, edges $\overrightarrow{v_1v_2}$ and $\overrightarrow{v_5v_4}$ are parallel and along the road direction in 3D but converge to a vanishing point, $q_v$, in 2D as depicted in Fig. 7a. The orientations of edges $\overrightarrow{v_1v_2}$ and $\overrightarrow{v_5v_4}$ are labeled as $\theta_{Y,1}$ and $\theta_{Y,2}$ respectively. To obtain them, first, we have to calculate $q_v$. Recalling the rectangle $ABCD$ used in camera calibration. As $\overrightarrow{AB} // \overrightarrow{CD}$ and they are along the road direction 3D, $q_v$ is defined as the interception point of lines $AB$ and $CD$ in the image coordinates. Second, we divide the vehicle mask into two portions along the line $\overrightarrow{q_vg}$ as depicted in Fig. 7b.
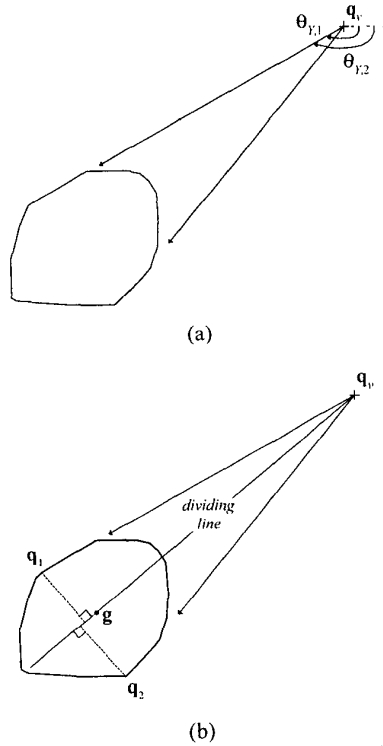
(a)

(b)

Fig. 7 (a) Vanishing point and orientations of the edges. (b) Division of vehicle mask and two extreme points.

Two extreme points, $q_1$ and $q_2$, on the vehicle mask with the largest distances from the dividing line are determined. From these two points, the required orientations can be obtained from Eqt. (12).

$$\theta_{Y,1} = \angle\overrightarrow{q_vq_1}, \qquad (12a)$$

$$\theta_{Y,2} = \angle\overrightarrow{q_vq_2}. \qquad (12b)$$

With all the orientations $\theta_X, \theta_{Y,1}, \theta_{Y,2}$ and $\theta_Z$ determined, 6 edges can be fitted to the vehicle mask as depicted in Fig. 8. The interception points of the 6 edges define the model's 6 vertices.
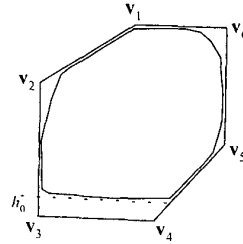
Fig. 8 Edges and vertices of the vehicle model obtained.

As seen in Fig. 4d, the bottom line of the vehicle mask corresponds to either the front or rear bumper of the vehicle, which is not on the road surface in 3D. Therefore, edge $\overrightarrow{v_3v_4}$ should not fit to that bottom line directly.

Instead, we need to shift down edge $\overrightarrow{v_3v_4}$ such that it lies on the road surface. To do this, we need to assume a bumper to road surface height, $h_0$, as depicted in Fig. 8, which varies from vehicles to vehicles. Since it is impractical to assume the knowledge of $h_0$, we need to estimate it. To do that, we measured $h_0$ of different vehicle types and observed that the measured values vary at 0.25m ± 0.10m. Therefore, a typical value of $h_0 = 0.25$m is chosen in our algorithm.

VI. DIMENSIONS ESTIMATION

From the vertices of the vehicle model, the dimensions of the vehicle can be determined by Eqt. (13). As $v_3$ to $v_5$ lie on the road, their latitudes are zero, i.e. $Z_i = 0$ for $i = 3,4,5$.

$$W = |\Phi^{-1}\{v_4,0\} - \Phi^{-1}\{v_3,0\}|, \qquad (13a)$$

$$L = |\Phi^{-1}\{v_5,0\} - \Phi^{-1}\{v_4,0\}|, \qquad (13b)$$

$$\Phi^{-1}\{v_2,H\} = \Phi^{-1}\{v_3,0\} + \begin{Bmatrix} 0 \\ 0 \\ H \end{Bmatrix}. \qquad (13c)$$

In Eqt. (13c), we assumed that vertices $v_2$ and $v_3$ have the X and Y coordinate in 3D, i.e. $X_2 = X_3$ and $Y_2 = Y_3$, while the Z coordinate of $v_2$ is $H$, i.e. $Z_2 = H$.

## VII. EXPERIMENTAL RESULTS

For evaluation purpose, three sets of traffic images of a taxi, a mini-bus and a double-decker were obtained from a traffic image sequence. The images were taken at daytime when the shadows are prominent. The first set of test images is a taxi where one of them is depicted in Fig. 9a. Fig. 9b depicts the subtraction result where both the taxi and its shadow are extracted. After applying the shadow removal algorithm and convex hull, the final vehicle mask is depicted in Fig. 9c. Based on the mask, the vehicle model is determined as depicted in Fig. 9d. From the mask, it can be seen that the shadow removal algorithm removes part of the rear of the vehicle. It is because that area is dark and the edge pixel density is low. However, it does not affect the modeling process much, which can be verified that the resulting model that fits the entire vehicle well. For the model, $\theta_Y$ and $\theta_Z$ are along the height and length orientations of the vehicle while $\theta_X$ slightly deviates from the width orientation of the vehicle, which is a source of error.
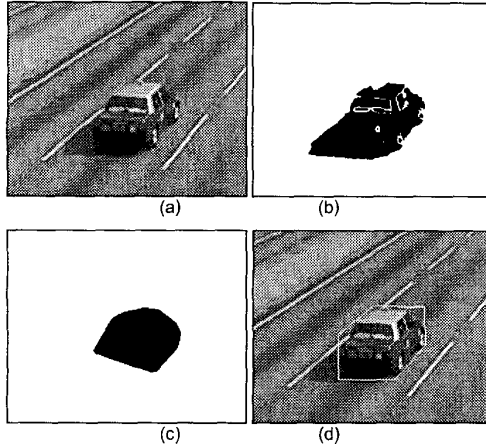

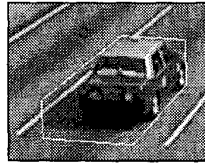Fig. 9 (a) Taxi. (b) Subtraction result. (c) Mask. (d) Model.


Fig. 10 Modeling result without removing the shadow.

Fig. 10 depicts the modeling result if we do not apply the shadow removal algorithm. In this case, the shadow is mis-interrupted as part of the vehicle body. Therefore, the resulting model is much larger than the vehicle and the shape is not a regular cuboid in 3D. This indicates that the necessity of an effective shadow removal algorithm for the

dimension estimation to be used in order for vehicle-type classification.

The second set of test images is a 16 seats mini-bus as depicted in Fig. 11a. The subtraction result and the final mask are depicted Figs. 11b and 11c respectively. The model obtained (Fig. 11d) in this case is better than the previous one visually. The shadow is removed successfully without affecting the vehicle body. The model fits the vehicle well and the orientations of the edges are close to the actual orientations of the vehicles.
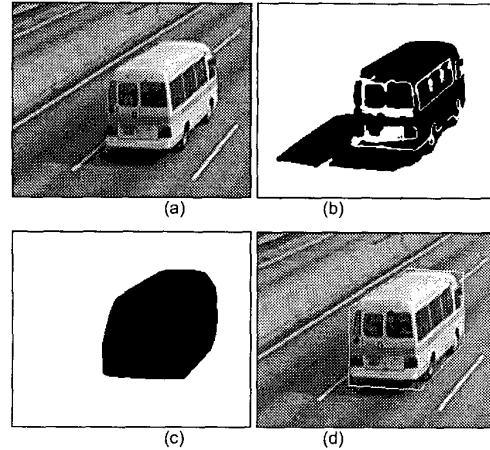

Fig. 11 (a) Mini-bus. (b) Subtraction result. (c) Mask. (d) Model.

The third set of images is a double-decker bus as depicted in Fig. 12a. The subtraction result and the final mask are depicted Figs. 12b and 12c respectively. Similar results as in the mini-bus are obtained for the double-decker. Shadow is removed successfully and the vehicle model is obtained accordingly. From the model, the leftmost edge slightly deviates from the vehicle due to the inaccuracy of the vehicle mask extracted.
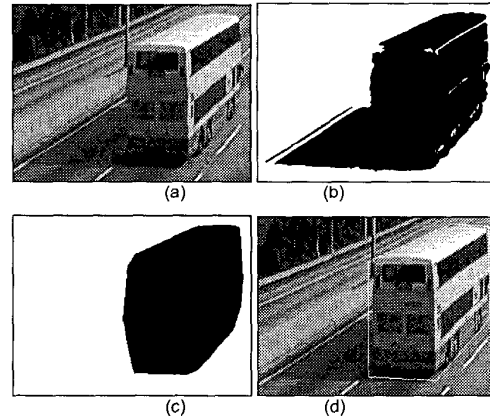

Fig. 12 (a) Bus. (b) Subtraction result. (c) Mask. (d) Model.

To evaluate the estimation accuracy, the complete test sets were evaluated and their average estimated vehicle

205

dimensions from the vehicle models are compared with the actual dimensions obtained from the manufacturers. Table 1 tabulates the values obtained for the taxi. Around 7.4% error in length estimation is encountered while the errors in width and height estimation are slightly above 10%. This is probably due to the deviation in edge orientation estimation. Tables 2 and 3 tabulate the values obtained for the mini-bus and double-decker bus. Smaller errors in length (~4%) and width (<3%) estimation are obtained. However, larger errors of >10% for the mini-bus and >7% for the double-decker bus are obtained.

| | Length | Width | Height |
|---|---|---|---|
| Actual | 4820 mm | 1765 mm | 1465 mm |
| Estimated (avg.) | 4462 mm | 1575 mm | 1306 mm |
| Error | 7.42% | 10.73% | 10.85% |

Table 1 Estimation result of Taxi (Toyota Crown).

| | Length | Width | Height |
|---|---|---|---|
| Actual | 6255 mm | 2025 mm | 2585 mm |
| Estimated (avg.) | 6507 mm | 2082 mm | 2308 mm |
| Error | 4.04% | 2.84% | 10.72% |

Table 2 Estimation result of Mini-Bus (Toyota Coaster 16seats).

| | Length | Width | Height |
|---|---|---|---|
| Actual | 11880 mm | 2500 mm | 4760 mm |
| Estimated (avg.) | 12408 mm | 2511 mm | 4396 mm |
| Error | 4.44 % | 0.47% | 7.65% |

Table 3 Estimation result of Bus (Volvo Super Olympian).

The major source of error is the inaccuracy of the vehicle mask, which affect the fitness of the vehicle model directly. Also, the model may not fit the vehicle perfectly due to small deviations in obtaining the orientations of the model edges. These can be verified from the results where the vehicle models are usually slightly larger than what they should be. It is found that the deviation is about 2 to 3 image pixels in general. Another source of error is the derivation in $h_0$. As mentioned in Section V, the derivation could be as large as 0.10m, which equals to 2.1% to 6.8% errors in height estimation. This may explain why the errors in height estimation are usually large in all cases. However, this error will be difficult to estimate.
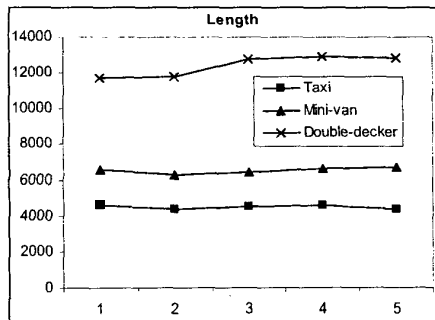


Fig. 12 Estimated lengths over 5 consective frames.

Fig. 12 shows the estimated lengths of the three vehicles over 5 consecutive frames. From the graph, the variations of individual vehicle are small. Also, there is a prominent distinction in their vehicle lengths. Thus, vehicle type can be identified from the estimated length effectively.

VIII. CONCLUSIONS

In conclusion, the visual-based dimension estimation method as presented in this paper is an effective way for obtaining vehicle length for vehicle type classification. The merits of the method are effective shadow removal algorithm is employed and using simple 3D cuboid model that can fit any types of vehicles. The drawbacks are the performance is hindered by the accuracy of the vehicle mask and the estimated bumper to road surface height. According to our evaluation, the estimation accuracy in vehicle length is above 92.5%, and the variations in estimation are very small. Therefore, major vehicle types, from private car to mini-bus to bus or truck, can be readily identified from the estimated length.

IX. REFERENCES

[1]     Ayland N., Davies P., "Automatic vehicle identification for heavy vehicle monitoring", Proc. Second International Conference on Road Traffic Monitoring, 1989, pp152-155.
[2]     Lee S.C., "Road traffic monitoring in Hong Kong", Proc. Second International Conference on Road Traffic Monitoring, 1989, pp14-18.
[3]     Harvey B.A., Champion G.H., Deaver R., "Accuracy of traffic monitoring equipment field tests", Proc. Vehicle Navigation and Information Systems Conference, 1993, pp141-144.
[4]     Pursula M., Kosonen I., "Microprocessor and PC-based vehicle classification equipments using induction loops", Proc. Second International Conference on Road Traffic Monitoring, 1989, pp24-28.
[5]     Burden M.J.J., Bell M.G.H., "Vehicle classification using stereo vision", Proc. Sixth International Conference on Image Processing and Its Applications, 1997, vol. 2, pp881-885.
[6]     Gupte S., Masoud O., Papanikolopoulos N.P., "Vision-based vehicle classification", Proc. Intelligent Transportation Systems, 2000, pp46-51.
[7]     Lai A.H.S, Yung N.H.C., "Vehicle-type identification through automated virtual loop assignment and block-based direction-biased motion estimation", IEEE Trans. Intelligent Transportation Systems, vol. 1, no. 2, 2000, pp86-97.
[8]     Yung N.H.C., Lai A.H.S., "Detection of vehicle occlusion using a generalized deformable model", Proc. IEEE International Symposium on Circuits and Systems, 1998. ISCAS '98, vol. 4, 1998, pp154-157.
[9]     Dubuisson M.P., Lakshmanan S., Jain A.K., "Vehicle segmentation and classification using deformable templates", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 3, 1996, pp293-308.
[10]     Koller D., Daniilidis K., Nagel H.H., "Model-based object tracking in monocular image sequences of road traffic scenes", Int. Journal of Computer Vision, 1995, pp75-86.
[11]     Lai A.H.S., Yung N.H.C., "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence", Proc. IEEE International Symposium on Circuits and Systems, 1998. ISCAS '98, vol. 4, 1998, pp241-244.
[12]     Fung G.S.K., Yung N.H.C, Pang G.K.H, Lai A.H.S, "Effective moving cast shadow detection for monocular color image sequences", Research report, ITS-2000G02, Laboratory for ITS Research, The University of Hong Kong.
[13]     Yung N.H.C., Pang G.K.H., Fung G.S.K., "A novel camera calibration technique for visual traffic surveillance", Proc. 7th World Congress on Intelligence Transport Systems, paper no.3024, 2000.