

Fast Vehicle Localisation and Recognition Without Line Extraction and Matching

T. N. Tan, G. D. Sullivan and K. D. Baker

Department of Computer Science

The University of Reading, Berkshire RG6 2AY, UK

Email: T.Tan@uk.ac.reading

Abstract

A novel algorithm is presented in this paper for vehicle localisation and recognition under the ground-plane constraint. Unlike the vast majority of the existing model-based object recognition schemes, the algorithm eliminates the need for explicit feature extraction and matching so that the computational cost is substantially lower. The algorithm is tested extensively with routine outdoor traffic images. Experimental results are included to illustrate the performance of the algorithm. The algorithm is developed for real-time implementation for applications in traffic scene analysis. It may readily be adapted to other industrial applications.

1 Introduction

The recognition and localisation of 3-D objects from a single intensity image using *a priori* 3-D geometric models [1-4] is a fundamental problem in computer vision. The majority of the existing algorithms address the problem of the general case of 6 degrees of freedom (DOF). Nevertheless, in many practical applications of computer vision, the objects to be recognised are constrained to be in contact with a known plane at a single stable pose. In this paper we are concerned with the localization and recognition of vehicles in traffic scenes, which under normal conditions stand on the ground-plane (GP), but similar applications such as the recognition of objects on a table, or parts on a conveyor belt, are commonplace. The ground-plane constraint (GPC) reduces the number of degrees of freedom of a rigid object from 6 to 3; these are most simply parameterized as the location on the GP (X, Y) and the orientation about the vertical axis (θ).

Existing object recognition algorithms typically involve extracting symbolic image features (e.g., lines), and matching image and model features. Both feature extraction and feature matching are error-prone and time-consuming. Even under the GPC, the matching between the image and models can be a very slow process when the number of models to be considered is large [2].

In this paper we describe a novel algorithm for model-based vehicle recognition and localisation in the context of traffic scene analysis. The algorithm takes a similar approach to [2], but does away with the early recovery of image lines. Unlike other existing object recognition algorithms, the proposed algorithm requires neither explicit line segment extraction, nor symbolic image features. The algorithm is made possible by the combined use of the GPC and the weak perspective assumption (WPA) both of which are valid in the intended application domain.

2 Coordinate Systems and Imaging Geometry

We assume a pinhole camera model with no lens distortion as sketched in Fig.1. The

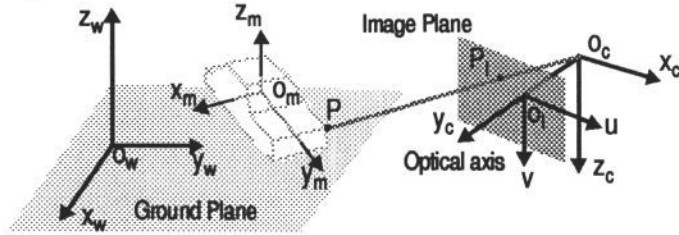


Figure 1: Coordinate systems and the imaging model.

object is located on the GP which coincides with the X_m - Y_m plane of the object coordinate system (OCS) as well as the X_w - Y_w plane of the world coordinate system (WCS). Under this imaging model and the GPC, the image coordinates (u, v) and the model or object coordinates (x_m, y_m, z_m) of point P are related to each other by:

$$\begin{bmatrix} su \\ sf \\ sv \\ s \end{bmatrix}^T = \begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix}^T \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ X & Y & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \quad (1)$$

where θ , X and Y are the three pose parameters of the object in the WCS, f the camera focal length, s a non-zero scale, and m_{ij} , $i, j = 1, 2, 3, 4$, are the elements of the 4×4 homogeneous perspective transformation matrix which maps 3D world coordinates into 2D image coordinates. The matrix is computed by off-line camera calibration (e.g., [8]) and is assumed to be known in the subsequent discussions.

It can easily be shown that when the angle subtended by the object is small and the object is viewed on axis, then the projection is scaled orthographic (also known as weak perspective), and the first three elements of the last column of the 4×4 transformation matrix are identically zero, i.e.,

$$m_{14} = m_{24} = m_{34} = 0 \quad (2)$$

We will use a similar simplification that applies locally to off-axis patches of the image.

3 Computing Orientation by Line Direction Matching

We start by discussing the computation of object orientation on the GP. Assume for the moment that a 3-D model line L is matched with a 2-D image line λ . The model line L is specified by a unit direction vector $\vec{L} = (\alpha \ \beta \ \gamma)^T$ and a point \vec{P}_0 . The perspective image λ of L is described in the Hesse normal form on the image plane:

$$(u \ v) \cdot \vec{\lambda} = \rho; \quad \vec{\lambda} = (\cos\omega \ \sin\omega)^T \quad (3)$$

where $\rho \geq 0$ is the distance from the origin of the image plane coordinate system to the image line λ , and ω is the angle between the image line and the v -axis. $\vec{\lambda}$ is therefore a

unit direction vector of the image line λ .

It is shown in [5] that a match between L and λ defines an independent constraint on the object orientation θ :

$$F \cos \theta + G \sin \theta = H \quad (4)$$

where the coefficients are given by

$$F = (\alpha \dot{\lambda}_1 + \beta \dot{\lambda}_2) \cdot \dot{n}; \quad G = (\alpha \dot{\lambda}_2 - \beta \dot{\lambda}_1) \cdot \dot{n}; \quad H = -\gamma \dot{\lambda}_3 \cdot \dot{n}; \quad \dot{n} = \begin{pmatrix} \dot{\lambda} \\ -\rho \end{pmatrix} \quad (5)$$

$$\dot{\lambda}_i = (m_{i1} \ m_{i3} \ m_{i4}), \quad i = 1, 2, 3, 4$$

Under the WPA, these coefficients become (by substituting (2) into (5)):

$$\begin{aligned} F &= (\alpha m_{11} + \beta m_{21}) \cos \omega + (\alpha m_{13} + \beta m_{23}) \sin \omega \\ G &= (\alpha m_{21} - \beta m_{11}) \cos \omega + (\alpha m_{23} - \beta m_{13}) \sin \omega \\ H &= -\gamma m_{31} \cos \omega - \gamma m_{33} \sin \omega \end{aligned} \quad (6)$$

The coefficients are independent of the position of the image line within the region of interest (ROI). In other words, under the WPA, a match between \tilde{L} and $\tilde{\lambda}$, the directions of the model and image line, defines an independent orientation constraint. An analytic error function $\varepsilon(\theta)$ of θ can easily be derived [2].

In summary, we have shown in this section that under the GPC and the WPA, the object orientation can be determined by matching the image line directions with the directions of the model lines. This paper demonstrates that the major image line directions in a region of interest can be estimated by examining the peaks of the histogram of the gradient directions of the original intensity image (see Section 5 for further details), so that no explicit line extraction is required. The scheme described in this section has the following advantages over previous ones (e.g., [2]):

- No line extraction is required and the computational cost is therefore substantially lower. In fact, the major edge directions can be identified from the video stream, “on the fly”, since only local gradient values are used.
- There are far fewer (perhaps 100 times fewer) matches of image and model lines to consider since only line directions (rather than the individual lines) are used.

4 Computing GP Location by 1-D Gradient Correlation

We next consider the location of the object on the GP by analysing the projections of intensity gradients along directions in the image determined by the orientation obtained. Under the WPA, the image-plane projection of an object model is determined once the object orientation is given. The projection is subject to an isotropic scale change depending on where the object model is located in the WCS. However the scale change is negligible within the ROI since the object subtends a small angle. The representative projection is obtained by instantiating the object model at the centre of the ROI and the known orientation.

The model projection defines a set of visible line segments on the image plane. The projection is regarded as a gradient image where the gradient vector associated with

a point on a visible line segment is defined to have a unit magnitude and a direction perpendicular to the visible line segment, and the gradient vector of a point not on any visible line segment is defined to be zero. Let $\{D_i; 1 < i \leq N\}$ be a set of 3-D directions emanating from the origin of the MCS. The projection of $\{D_i; 1 < i \leq N\}$ is a set of directions $\{d_i; 1 < i \leq N\}$ on the image plane. The 2-D model projection is represented by a set of 1-D model profiles $\{m_i(u); 1 < i \leq N\}$, where $m_i(u)$ is the model projection collapsed onto a line perpendicular to d_i (The origin of $m_i(u)$ is defined as the image-plane projection of the origin of the MCS after being collapsed onto the line perpendicular to d_i).

The 2-D gradient image of an input intensity image is similarly collapsed along $\{d_i; 1 < i \leq N\}$ to obtain a set of 1-D data profiles $\{s_i(u); 1 < i \leq N\}$. The origin of $s_i(u)$ may be defined as the centre of the ROI. If the recovered orientation is sufficiently accurate, $\{s_i(u); 1 < i \leq N\}$ and $\{m_i(u); 1 < i \leq N\}$ are expected to have similar shapes apart from a possible translation due to the unknown GP location. The similarity between $s_i(u)$ and $m_i(u)$ is measured by the cross correlation $c_i(u)$ computed as

$$c_i(u) = \sum_{u'} s_i(u') m_i(u' - u) \quad (7)$$

Peaks u_i^* of $c_i(u)$ determine the position of image lines a_i in the direction of d_i . According to the definition of $m_i(u)$, the image line a_i specifies a confusion line on which the projection of the origin of the MCS might lie.

Thus for all pairs of $\{s_i(u); 1 < i \leq N\}$ and $\{m_i(u); 1 < i \leq N\}$, we compute the 1-D correlation functions to determine a set of confusion lines $\{a_i; 1 < i \leq N\}$. The intersection of the confusion lines determines the image-plane projection (U, V) of the GP location. If 3 or more lines are perfectly consistent, then they will intersect at exactly the same point. In practice, the confusion lines do not intersect at the same point. In this case (U, V) is defined as the point which has the minimal sum of squared distances to the confusion lines. The average distance \bar{d} from (U, V) to the confusion lines may be used as a measure for the consistency of the confusion lines. Therefore the smaller \bar{d} is, the more consistent the lines are. If $c_i(u)$ possesses C_i peaks, then C_i confusion lines may be considered along direction d_i . Therefore there are a total of $C = C_1 C_2 \dots C_N$ candidates for the GP location. These candidates can be ordered in increasing \bar{d} , and only the first few need to be considered for further processing.

In practice, it has generally proved efficient and effective to use triples of lines (i.e., $N = 3$). When considering near-rectilinear objects such as vehicles, the three directions D_1, D_2 and D_3 can conveniently be chosen to be the X-, Y- and Z-axis of the MCS. In this case, the recovered confusion lines specify the possible image-plane projections of the three axes of the MCS.

In summary, we have shown in this section that if the WPA holds and the ROI is not excessively large, then the location of the object on the GP can be determined by 1-D correlations. These can be done very rapidly. In fact, the model profiles can often be reduced to a very sparse set of non-zero values (spikes), without loss of performance, thus speeding up the algorithm further. Furthermore, all calculations rely only on local gradient data and again may be performed "on the fly" once the orientation has been determined.

5 Implementation and Experimental Results

A gradient estimator such as the Sobel operator is first applied in a ROI to compute a gradient image. The gradient vector of each pixel is then accumulated as a function of the gradient direction which covers the interval $[0^\circ, 180^\circ)$ since the sign of the direction is irrelevant. The result of the accumulation process is a 1-D histogram representing the relative strength of the gradient directions. The histogram is slightly smoothed to remove noise due to the quantization of gradient directions. An example is shown in Fig.2.

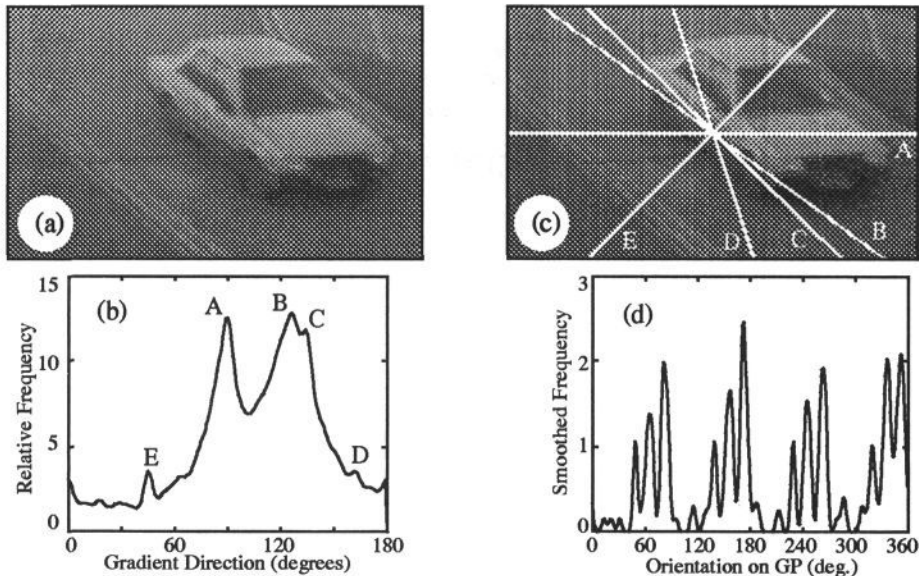


Figure 2: (a) a region of interest (322x195 pixels); (b) the corresponding histogram of the gradient directions; (c) five dominant edge directions obtained from (b); and (d) the orientation histogram obtained by matching the five directions in (c) with a saloon model.

The major peaks of the histogram are located to recover a set of dominant edge directions in the ROI (see Fig.2(c)). Each dominant edge direction is then considered as a match for each model line direction and, where solutions exist, the corresponding error functions ($\varepsilon(\theta)$) of the two possible orientations of the object are computed from (4). The recovered orientation distributions for each direction match are histogrammed. To reduce the effects of small errors in orientation, the histogram is then smoothed by a Gaussian function.

Fig.2(c) shows the five dominant edge directions detected from Fig.2(b), and Fig.2(d) is the orientation histogram obtained by matching the five edge directions with a saloon model. The orientation histogram is dominated by peaks at 90° intervals, corresponding to the strong rotational symmetry of the near-rectilinear vehicle. Each peak identifies the possible orientation (θ) of the model on the GP. When the level of clutter is not too high, the global peak of the orientation histogram usually corresponds to the true orientation of the object as is the case in Fig.2(d).

Once the orientation is determined, the object model is instantiated at the centre

of the ROI, at the known orientation. Three 1-D model profiles $m_x(u)$, $m_y(u)$ and $m_z(u)$ are then generated along directions perpendicular to the projections d_x , d_y and d_z of the three axes of the MCS. Three 1-D data profiles $s_x(u)$, $s_y(u)$ and $s_z(u)$ are similarly obtained from the gradient image of the ROI. In Fig.2 and Fig.3, the same image data is used. For “on the fly” processing, there would be little error caused by using the successive frames for the two stages. The three correlation functions $c_x(u)$, $c_y(u)$ and $c_z(u)$ are then computed by (10). The results are shown in Fig.3.

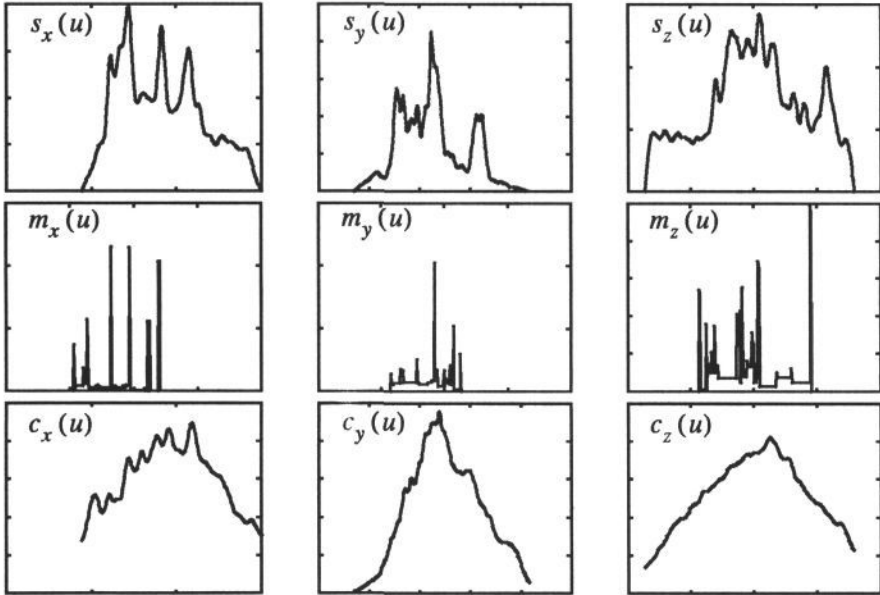


Figure 3: 1-D model and data profiles and their correlations for the ROI shown in Fig.2(a). Top row: data profiles; middle row: model profiles; bottom row: correlation functions between model and data profiles.

The global peak in each correlation function is located to determine the projections a_x , a_y and a_z of the X-, Y- and Z-axis of the MCS. The intersection of the three confusion lines is then calculated to determine the GP location of the object. This is demonstrated in Fig.4, where Fig.4(a) shows the recovered projections of the three axes

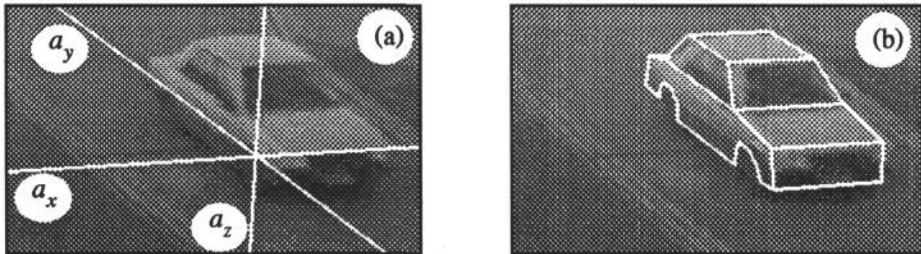


Figure 4: (a) recovered projections of the three axes of the MCS; and (b) a saloon model instantiated at the recovered pose.

of the MCS, and Fig.4(b) is the recovered pose for the vehicle. It can be seen that the three confusion lines are very consistent, and the recovered pose is very accurate.

5.1 Pose discrimination

In the examples discussed so far, only the global peak in the orientation histogram and each of the correlation functions has been considered, and in this case the global peaks identify the true object orientation and the correct projections of the axes of the MCS. In general, several orientation and correlation peaks may have to be considered since the global orientation and correlation peak does not always correspond to the correct pose due to factors such as misalignment with clutter. This leads to multiple candidate poses, and one has to identify the correct pose from these candidates. In our current work, the quality of a pose is indicated by an evaluation score (see [6] for the computation of such scores). The evaluation score measures the goodness-of-fit between the model projection and the original image. A high score indicates a good fit between the model projection and the image, and thus also signifies a good pose. The correct pose is then taken as the candidate pose which produces the highest evaluation score.

The procedure outlined above has been successfully applied in locating vehicles in multiple lanes of a motorway scene. Typical results are shown in Fig.5 for three frames of the motorway sequence. A fixed ROI of size 235x193 pixels is specified for each of the

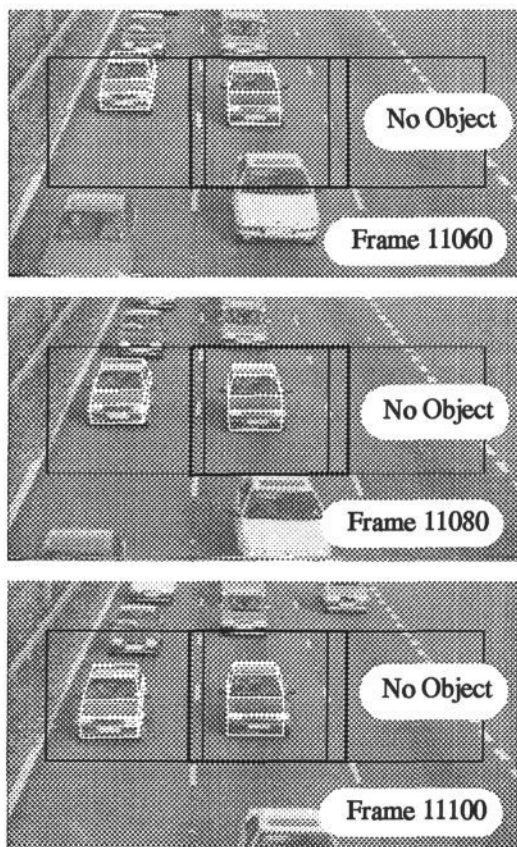


Figure 5: *Vehicle detection and localisation in motorway scenes.*

three lanes (see the boxes in Fig.5). A state of “no object” is automatically declared when the evaluation score is lower than a threshold. It can be seen that the vehicles in all lanes and all frames have been correctly detected and located. The object model used in all cases is a general saloon model.

Results for a more difficult situation from the same motorway sequence are shown in Fig.6. Because of traffic congestion, vehicles occlude each other in the image.

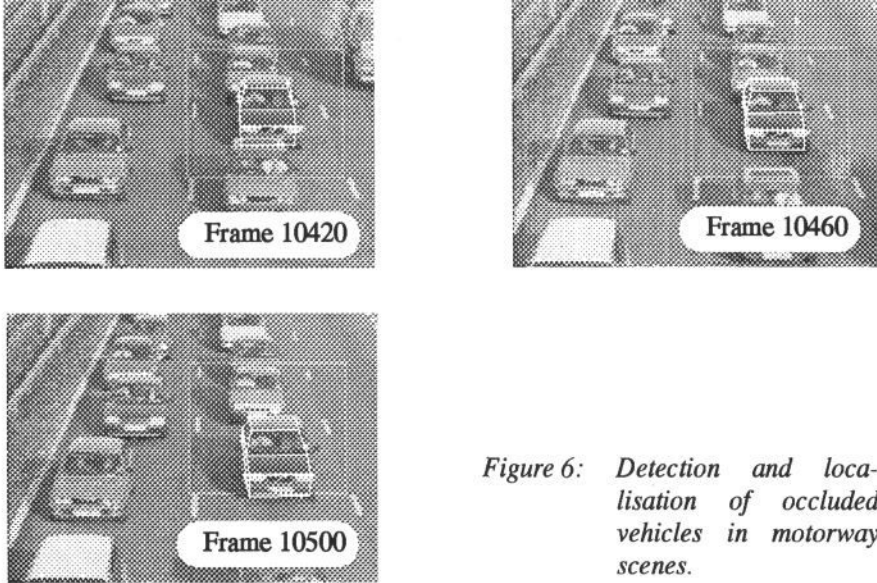


Figure 6: *Detection and localisation of occluded vehicles in motorway scenes.*

In such situations, techniques mainly based on 2-D image processing/analysis (e.g., [7]) would be unable to separate and locate the vehicles correctly. The present algorithm, on the other hand, has succeeded in accurately locating the vehicle in the ROI.

5.2 Model discrimination

When there are multiple models to be considered, the process described above is repeated with each model. It should be noted that in many practical applications, there are two further simplifications: a) the orientation is known *a priori* (e.g., in the motorway scene shown in Fig.5); and b) objects are rectilinear or near-rectilinear. Therefore, the only process of the algorithm that depends on the number of models is the 1-D correlations which can be done very rapidly. The best evaluation scores obtained with each model are then compared. The model which yields the highest score identifies the class of the object.

Fig.7 shows the results obtained with three different vehicle models - a saloon, a hatchback and an estate car which differ somewhat in size as well as shape. The ROI shown in Fig.7(a) is of size 313x217 pixels, and the four dominant edge directions detected from the gradient direction histogram (Fig.7(b)) are depicted in Fig.7(c). The four edge directions are matched, in turn, with a saloon, a hatchback, and an estate car model, and the resultant orientation histograms are shown in Fig.7(d)-(f) respectively. The three orientation histograms are similar in shape, reflecting the fact that the three car models possess a similar set of 3-D line directions. The poses recovered for each model

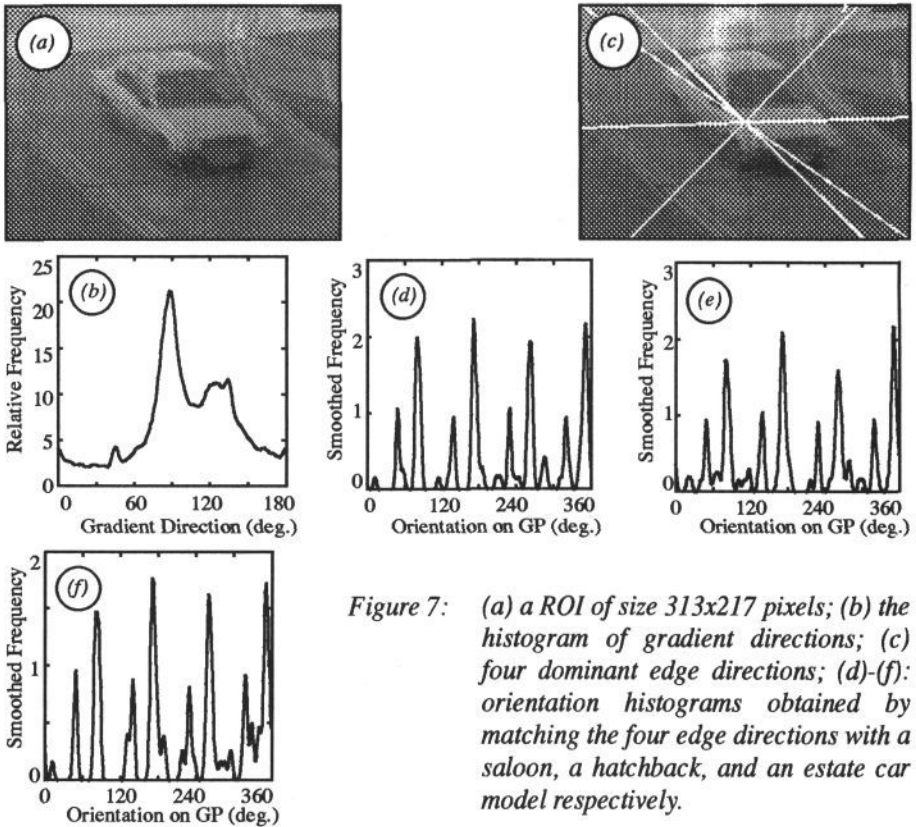


Figure 7: (a) a ROI of size 313x217 pixels; (b) the histogram of gradient directions; (c) four dominant edge directions; (d)-(f): orientation histograms obtained by matching the four edge directions with a saloon, a hatchback, and an estate car model respectively.

are shown in Fig.8. As expected, the orientation of the car is recovered well in all cases. The recovered GP location is good with the saloon and the hatchback model, but is poor with the estate model. The evaluation scores obtained for the three models show the highest value for the correct (saloon) model, thus allowing object recognition. It is also interesting to note that the intuitively more similar hatchback model produces the second highest score. Cross-over discrimination trials have also been carried out which confirm the discrimination capability of the algorithm.

6 Conclusions

We have presented in this paper a novel algorithm for the localisation and recognition of vehicles in traffic scenes which, under normal conditions, are constrained to be in contact with the ground-plane. The algorithm eliminates the need for explicit feature extraction and matching, and substantially reduces the computational cost. It has been shown that the object orientation can be found merely by matching the dominant edge directions found in a ROI using a simple gradient operator with a small set of model line directions. Once the orientation is known, the location of the object on the ground-plane is determined by simple and fast 1-D correlations.

The algorithm has been tested extensively with routine outdoor traffic images, with successful results. The algorithm has been developed for real-time implementation

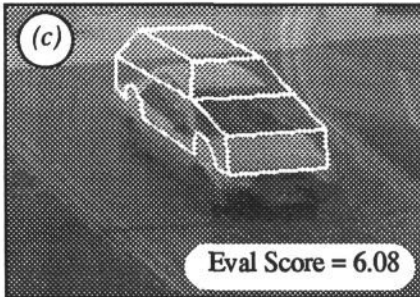
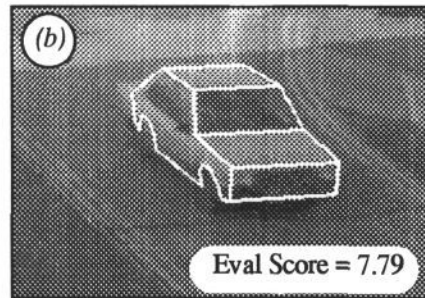
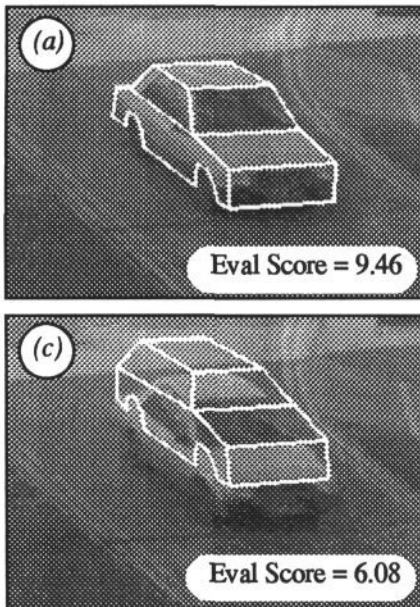


Figure 8: Recovered poses and iconic evaluation scores using (a) Saloon, (b) Hatchback, and (c) Estate models.

for applications in traffic scene analysis. It may readily be adapted to other industrial applications where objects are in contact with a known plane.

References

- [1] D. G. Lowe, Three-dimensional object recognition from two-dimensional images, *Artificial Intell.*, vol.31, 1987, pp.355-395.
- [2] T. N. Tan, G. D. Sullivan and K. D. Baker, Recognizing Objects on the Ground-plane, *Image and Vision Computing*, vol.12, no.3, 1994, pp.164-172.
- [3] T. M. Silberberg, D. A. Harwood and L. S. Davis, Object recognition using oriented model points, *CVGIP*, vol.35, 1986, pp.47-71.
- [4] D. Koller et. al., Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes, *Int. J. Computer Vision*, vol.10, June 1993, pp.257-281.
- [5] T. N. Tan, G. D. Sullivan and K. D. Baker, Linear Algorithms for Object Pose Estimation, *Proc. of BMVC92*, 1992, pp.600-609.
- [6] G. D. Sullivan, Visual Interpretation of Known Objects in Constrained Scenes, *Phil. Trans. Royal Soc. London, Series B: Biological Sciences*, vol.337, 1992, pp.361-370.
- [7] E. Sorensen and H. I. Christensen, Monitoring of Road Traffic, *Proc. of 8th Scand. Conf. Image Analysis*, May 1993, Tromso, Norway, vol2, pp.1085-1091.
- [8] A. D. Worrall, G. D. Sullivan and K. D. Baker, A Simple, Intuitive Camera Calibration Tool for Natural Images, *Proc. of BMVC94*, 13-16 September 1994, York, England.