

Recognising Objects on the Ground-plane[†]

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

Department of Computer Science
University of Reading, Berkshire RG6 2AY, UK
Email: T.Tan@uk.ac.reading

Abstract

Objects such as vehicles are often constrained to lie on a known plane. The ground-plane constraint reduces the problem of localisation and recognition from 6 dof to 3 dof. A novel algorithm is presented which makes effective use of the ground-plane constraint to derive pose estimates. A form of the generalised Hough transform is used to group evidence from line features, and to identify approximate poses. The single orientation parameter is decoupled from the two location parameters, and treated separately. The method is fast and robust. It copes well with complex outdoor scenes including multiple occluded objects, and image clutter from irrelevant structures.

1 Introduction

In many practical applications of computer vision, the objects to be recognised are constrained to be in contact with a known plane. 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 position on the GP (X, Y) and the orientation about the vertical axis (θ).

Object recognition depends crucially on establishing matches between 2D features (e.g. line segments) discovered by initial analysis of the image data and 3D features encoded in *a priori* object models. Even under the GPC, a given image line can arise from many different model lines, as the object takes different poses. Each match between a 2D image line (with known, accurate end points) and a 3D model line (called here, a line match) provides in principle sufficient information to identify the pose uniquely. However determination of the end points of an image line is very unreliable. In practice, a line match identifies the orientation (θ) of the model, but is imprecise along a confusion line in the ground-plane (X, Y), which is easy to determine geometrically.

The simplicity of the confusion geometry makes object identification under the GPC well suited to treatment by the generalised Hough transform. All possible

[†]. This work was carried out as part of the ESPRIT project P2152 (VIEWS).

line matches vote for a range of poses (in (X, Y, θ)), and high scores, where many line matches coincide, are used to identify likely poses. In fact the method is further simplified under the GPC because the orientation (θ) can be recovered independently of location (X, Y) . This means that the most likely (model) orientations can be identified first, and only subsets of line matches consistent with each possible orientation need be considered to determine location.

2 Previous related work

Since the work of Roberts [1], much effort has been devoted to object recognition [2]. Among the various recognition strategies [2], the generalized Hough transform technique [3-5] is intuitively appealing, and is inherently suitable for parallel implementation. The algorithm described here is of this type.

The use of the GPC in object recognition is not new, e.g. [5-7]. The overall algorithm presented here resembles that of Silberberg *et. al.* [5] in spirit, but there are many important differences. Unlike in [5] (and indeed in many other existing approaches), object recognition here is based on matching 2-D and 3-D line segments rather than isolated points. The use of the GPC in [5] is cumbersome and indirect, and introduces multiple solutions which substantially increase the computational cost at the recognition stage, whereas our use of the constraint is based on simple and intuitive geometric observations, and is mathematically easy to follow. Also, the algorithm described in [5] is based entirely on *ad hoc* voting schemes whereas our algorithm is rooted on a formal error analysis. Our algorithm also differs from that of [5] in pose verification and refinement. Finally, the scenes considered in [5] were well-controlled laboratory scenes of polyhedral objects with little clutter from irrelevant structures. Here we address natural, wide-area outdoor traffic scenes in an attempt to develop a real-time model-based vision system for automatic road traffic monitoring and surveillance [8].

3 Pose estimation from a line match

We first examine what pose parameters can be recovered from a single 2-D to 3-D line match under the GPC and without assuming the correspondences between the end points of the line segments.

3.1 Pose constraints from a single line match

It is shown in [9] that given the correspondence between a 3D model line M and a 2D image line S , the following two constraints can be derived on the three pose parameters θ , X and Y :

$$\begin{cases} F \cos \theta + G \sin \theta = H \\ A \cos \theta + B \sin \theta + CX + DY = E \end{cases} \quad (1)$$

where A, B, C, D, E, F, G and H are terms computable from known data such as

the model and image line parameters (see [9] for details).

3.2 Computing object orientation on the ground-plane

The first constraint in (1) does not involve the translational parameters, and can easily be solved for the rotation angle θ . By defining

$$\sin\phi = F/\sqrt{F^2 + G^2}, \quad \cos\phi = G/\sqrt{F^2 + G^2}, \quad Q = H/\sqrt{F^2 + G^2} \quad (2)$$

the first constraint of (1) can be rewritten as

$$\sin(\phi + \theta) = Q \quad (3)$$

If $|Q| \leq 1$, the two possible solutions for θ are given by

$$\begin{cases} \theta_1 = \text{asin}(Q) - \phi \\ \theta_2 = \pi - \text{asin}(Q) - \phi \end{cases} \quad (4)$$

Therefore, under the GPC, the object's orientation can be recovered from a single line match up to a single ambiguity (assuming non-degenerate configurations).

The input data has so far tacitly been assumed to be perfect. In practice, the image data is always subject to uncertainty. For the purpose of this work, it is assumed that the error in a 2D-3D line match is solely due to the inaccuracy in locating the 2D image line segment. If each endpoint of an image line segment can only be located up to a circular region of radius ρ (in pixels), then the orientation θ recovered from the first constraint of (1) can be shown to lie within one of two orientation intervals containing θ_1 and θ_2 respectively [10]. The probability density function (PDF) of θ in the interval $[\theta_{i,\min}, \theta_{i,\max}]$ associated with θ_i is given by (see [10] for details):

$$f(\theta) = \begin{cases} \frac{|p_{i1}q_{i0} - p_{i0}q_{i1}|}{2\Delta(p_{i1} + q_{i1}\theta_i - q_{i1}\theta)^2}, & \theta_{i,\min} \leq \theta \leq \theta_{i,\max} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $\theta_{i,\min}$, $\theta_{i,\max}$, p_{ij} and q_{ij} are terms computable from known parameters [10], and $\Delta = \text{asin}(2\rho/L)$ with L being the length of the image line segment. The orientation interval $[\theta_{i,\min}, \theta_{i,\max}]$ varies strongly with the geometry of the model and image lines - for example, a near vertical model line would produce a very broad orientation interval which indicates a weak constraint on θ .

From (5), the probability of θ being in a sub-interval (θ_{k-1}, θ_k) of $[\theta_{i,\min}, \theta_{i,\max}]$ is easily computed:

$$P(\theta_{k-1} \leq \theta \leq \theta_k) = \frac{Wq_{i1}(\theta_k - \theta_{k-1})}{(p_{i1} + q_{i1}\theta_i - q_{i1}\theta_k)(p_{i1} + q_{i1}\theta_i - q_{i1}\theta_{k-1})} \quad (6)$$

where $W = |p_{i1}q_{i0} - p_{i0}q_{i1}| / (2\Delta q_{i1})$. In what follows, we assume that $\rho = 0.5$ unless otherwise stated.

3.3 Computing object location on the ground-plane

Once the orientation θ is known, the object is confined to slide along a line on the ground-plane defined by the second constraint of (1):

$$L_{GP}: CX + DY = P \quad (7)$$

where $P = E - A\cos\theta - B\sin\theta$. The possible positions of the object on the ground-plane can be further restricted by requiring the projection of the model line M to overlap the corresponding image line segment S . A simple rule is that the projection of the midpoint of the model line segment should lie somewhere between the two end points (u_1, v_1) and (u_2, v_2) of S . Thus the object can only slide along L_{GP} between (X_1, Y_1) and (X_2, Y_2) , where (X_1, Y_1) and (X_2, Y_2) can easily be computed from known variables by matching the midpoint of the model line segment with (u_1, v_1) and (u_2, v_2) respectively.

In principle, similar probability calculations to those presented in the preceding subsection might be carried out for the GP location parameters X and Y . In practice, however, this has proved to be unnecessarily complicated, and is not discussed in this paper.

4 Implementation

Image line segments were first extracted in a region of attention using an algorithm based on the Canny operator. Lines of length greater than 10 pixels become the set of 2D image lines considered by the algorithm. The process is illustrated in Fig.1(a)-(c), for a fragment of an image.

Each 3D model line is considered as a match for each image line and, where solutions exist, the distribution intervals and their corresponding PDFs of the two possible orientations of the object are computed using (5). The recovered orientation distributions for each line match are histogrammed with a bin width of 1° (Fig.1(d)). During voting, the contribution a line match makes to a particular bin is proportional to the probability (calculated from (6)) of the line match producing an orientation which lies within the bin. The contribution is recorded only if the model line of the line match would be visible at the recovered orientation (i.e., it is not self-occluded). To reduce the effects of small errors in orientation, the histogram is then smoothed by a Gaussian function of standard deviation equal to 3° . Note, the probability-based voting scheme embedded in our algorithm is fundamentally different from the conventional equal or other *ad hoc* voting schemes (e.g., [3-5], [7]).

Fig.1(d) shows the results obtained for Fig.1(c). The orientation histogram is dominated by 4 peaks, at 90° intervals, corresponding to the strong rotational symmetries of the near-rectilinear vehicle. Each peak identifies the orientation (θ)

of the model on the GP where multiple matches agree. Taking each peak in turn, we discard all matches except those which contribute to it. Each of these matches constrains the location of the origin of the object coordinate system to lie on a line segment on the ground given by (7). In this way, each accepted line match defines a segment of line in the remaining configuration space (the ground-plane, (X,Y)).

The (X,Y) line segments are tallied in an accumulator array, and subsequently blurred by a Gaussian of standard deviation equivalent to 240mm on the ground-plane (Fig.1(e)). Local maxima in the accumulator array are identified, to recover the most likely locations of objects (having the known orientation). In the case of an isolated vehicle such as Fig.1(a), we usually obtain one conspicuous peak as seen in Fig.1(e) (where darker points indicate higher accumulator values). Fig.1(f) shows the object instantiated in the image in this pose; the fit is very close. The recovered pose is used to seed a process of model-based iconic pose refinement (see [8]). Fig.1(g) shows the result of pose refinement.

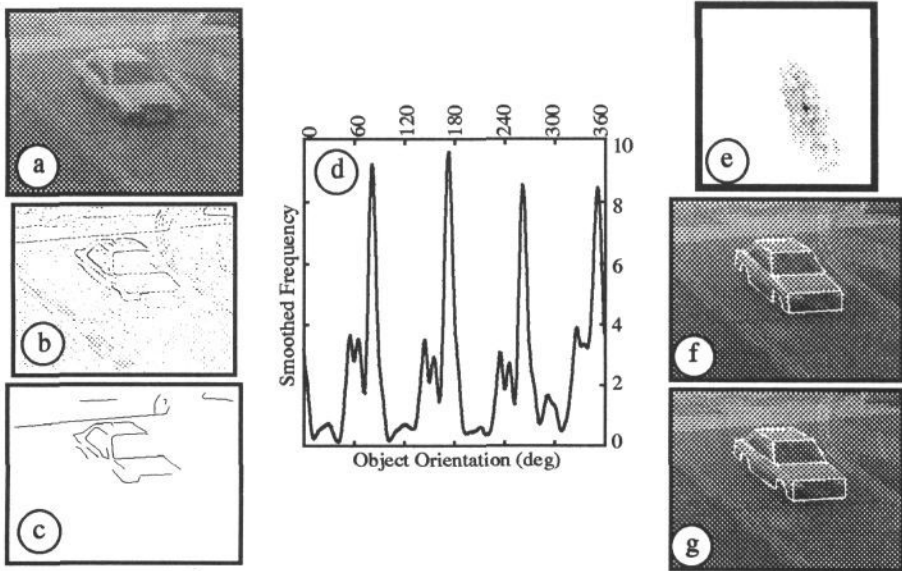


Figure 1: (a) Original image (160*200 pixels); (b) Canny edges; (c) Line description; (d) Orientation histogram; (e) X-Y accumulator at the global orientation peak ($=174^\circ$); (f) Recovered pose; (g) Refined Pose

5 Results

The procedure outlined above was applied to a sequence of images following that of Fig.1, taken at 5 Hz. In all cases the global peak of the orientation histogram corresponded to the true orientation, and the pose was recovered very accurately, with a small adjustment at the pose refinement stage. The algorithm is robust in spite of accidental alignment with the ground shadow, and distortion of the lines of the side window (due to specularities). Detailed results may be found in [12].

Fig.2 shows the (X,Y) accumulator diagrams corresponding to the 4 highest

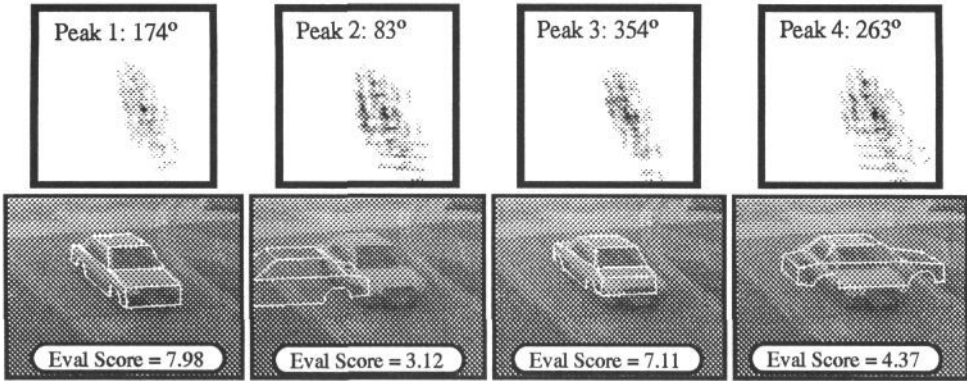


Figure 2: X-Y accumulators (top-row) and recovered initial poses (bottom-row) for the 4 main peaks in Fig.1(d).

peaks of the orientation histogram of Fig.1(d), together with the recovered poses. The 90° rotational symmetry is evident. Note that the peaks are less well defined for the cases at 90° from the correct pose. The evaluation score (=7.98) obtained at the true pose is highest, with the closely similar 180° rotation the runner-up. In practical applications the multiple symmetrical solutions may often be eliminated by using *a priori* domain knowledge, and further reduce the computational cost: only those orientation peaks falling within an expected range need be examined.

5.1 Sensitivity to Clutter

Any voting algorithm is strongly affected by irrelevant noise. Fig.3 shows a repeat

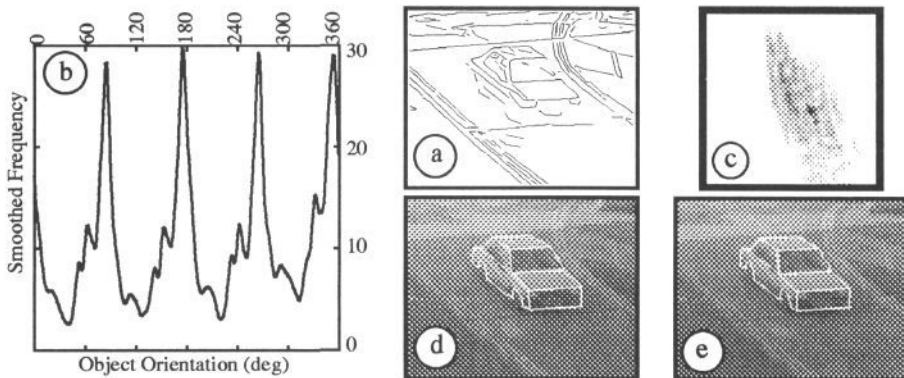


Figure 3: Fig.1 repeated, using lower thresholds in the line finding stage; (a) Line description (cf. Fig.1(c)); (b) Orientation histogram; (c) X-Y accumulator at the global orientation peak ($=174^\circ$); (d) Recovered pose; (e) Pose after iconic refinement

of Fig.1, using lower thresholds at the edge-detection stage, leading to more lines. The orientation peaks become much higher and a little less sharp, but the result is still clear.

5.2 Sensitivity to Model Geometry

Fig.4 shows the poses recovered using the identical process of Fig.1, except that

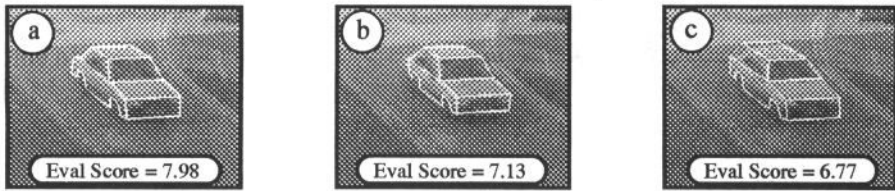


Figure 4: Recovered poses and iconic evaluation scores using (a) Saloon (correct), (b) Hatchback, and (c) Estate models.

three different vehicle models were used - a saloon, a hatchback and an estate car which differed somewhat in size as well as shape. The recovered poses are good in all cases, showing that the process is not critically dependent on an exact object model. We can therefore expect that a single model can be used to locate cars of different types. Note, however, that the evaluation scores obtained for the three models show the highest value for the correct (saloon) model, thus allowing object discrimination. 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.

5.3 Multiple Objects

The examples discussed so far are relatively simple - each image contains only one isolated object. Fig.5(a) shows an image from one of the VIEWS test sites, containing many different vehicles, in widely varying poses. The symbolic line descriptions of Fig.5(a) are shown in Fig.5(b). Fig.5(b) also shows two boxed regions such as might be identified by the analysis of image-plane movement [11]. Each box contains multiple occluding objects and a considerable amount of image clutter. The orientation histograms for each region are shown in Fig.5(c) and (d). The banded areas illustrate ranges of expected orientation which may be derived, for example, from image-plane feature-based tracking or by *a priori* knowledge of the GP and the road lay-out.

The foreground region shows 4 peaks in the orientation histogram (Fig.5(d)), within the expected orientation range. The main peak (at 256°) gives rise to the (X,Y) accumulator array shown in the top-row of Fig.6(a). The peak is poor. It corresponds to the pose shown in the middle-row of Fig.6(a), where the multiple lines on the two sides of the car are approximately aligned with image features from the street furniture. This pose corresponds to a "view catastrophe", where many nearly coplanar model features are aligned with the camera, and hence many matches with a single image line vote for the orientation. The evaluation score obtained following pose refinement (see the bottom row of Fig.6(a)) is 6.6.

The results for the second highest peak (at 291°) in the expected range are

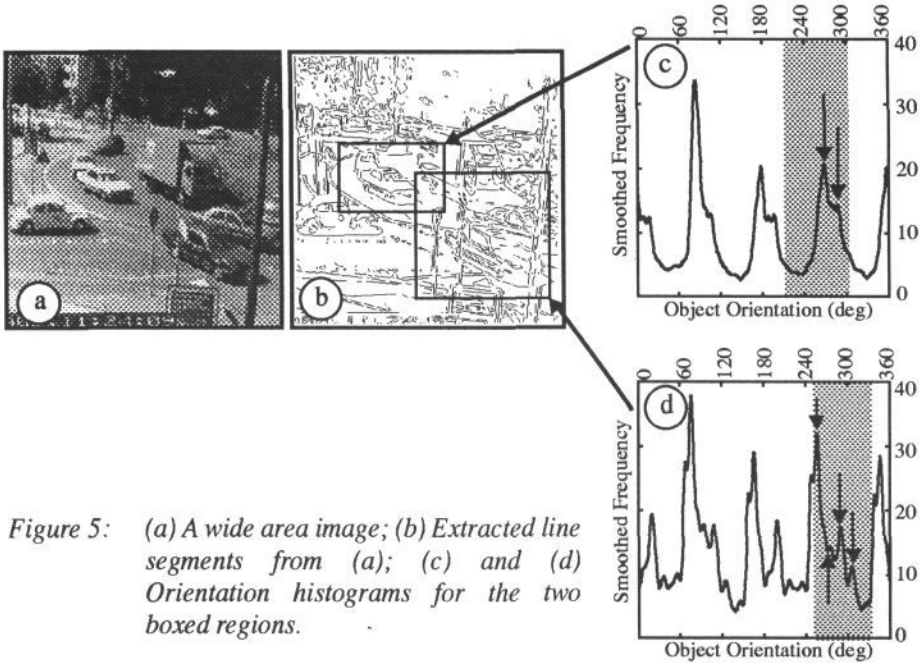


Figure 5: (a) A wide area image; (b) Extracted line segments from (a); (c) and (d) Orientation histograms for the two boxed regions.

given in Fig.6(b). In this case the pose of the nearer vehicle is recovered well, and pose refinement yields a far higher score of 8.8. The third highest peak (at 274°) also corresponds to a false alarm, shown in Fig.6(c); here the windows align with a road sign in the foreground. The evaluation score after pose refinement is 7.9. Finally the peak at 307° gives the results shown in Fig.6(d), corresponding to the second car, with an evaluation score of 9.97.

Simple geometrical reasoning shows that the first three peaks give rise to mutually exclusive object hypotheses. We therefore accept the hypothesis with the greatest evaluation score (Fig.6(b)), and identify the correct pose. It should also be noted that while the (X,Y) accumulators for the two correct poses (Fig.6(b) and (d)) exhibit conspicuous and well-defined peaks, those for the two false alarms (Fig.6(a) and (c)) are far more diffuse.

Similar processes were applied to the other boxed region in Fig.5(b), and the results are shown in Fig.7. The two main peaks of Fig.5(c) in the expected orientation range give the poses shown in Fig.7, and correctly identify the two cars, with good evaluation scores.

6 Conclusions

An algorithm has been described for object recognition under the ground-plane constraint. The algorithm is conceptually simple, fast, robust, and inherently parallel. It does not involve explicit computation of inter-feature relationships, so that its computational cost is (approximately) proportional to the number of image

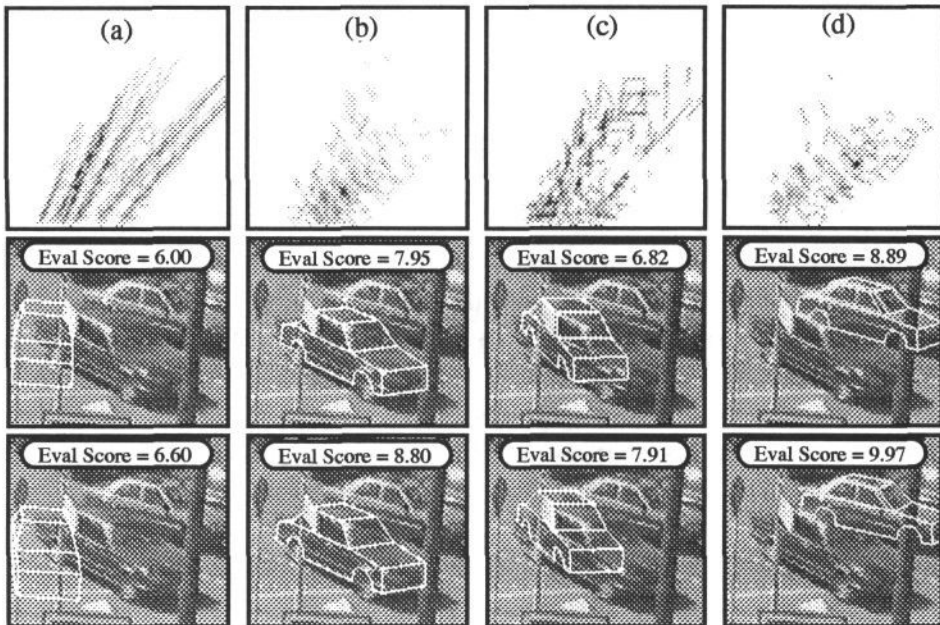


Figure 6: X-Y accumulator arrays and recovered poses for four peaks in Fig.5(d) at (a) 256° ; (b) 291° ; (c) 274° ; and (d) 307° . Top row: X-Y accumulator arrays; middle and bottom rows: initial and refined poses.

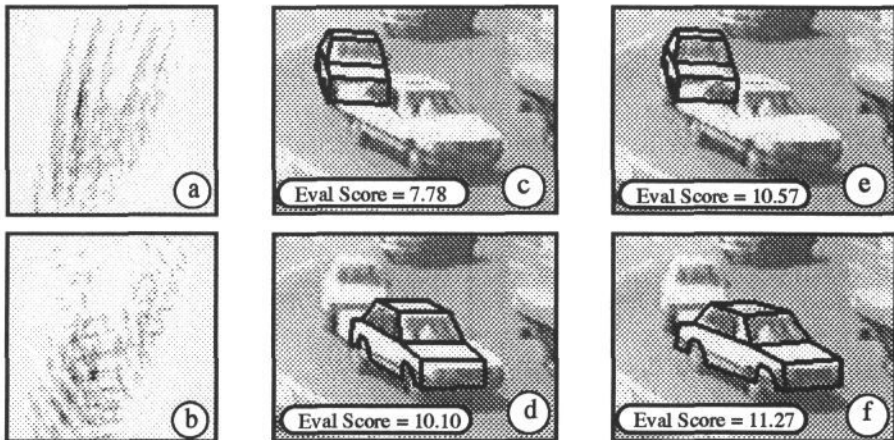


Figure 7: X-Y accumulator arrays and recovered poses for two peaks in Fig.5(c) at 269° (top row) and 284° (bottom row). (a) and (b): X-Y accumulator arrays; (c) and (d): initial poses; (e) and (f): refined poses.

features. The algorithm has been shown to work well in a wide variety of images of out-door traffic scenes.

The GPC makes model-based pose recovery fast and reliable. It eliminates the main problems in 6 dof pose recovery, since only one orientation parameter is

involved, and this can be decoupled from location. A single line match provides sufficient information to constrain the pose to a segment of a straight line in configuration space (X, Y, θ) . The simplicity of the pose constraint makes it feasible to use the Hough transform to find groups of line matches which correspond to a single pose.

The method has clear advantages over other methods for unconstrained object recognition. Interpretation tree pruning [2], for example, is grossly inefficient in 2D-3D feature matching, especially in the presence of clutter. Other methods for feature grouping and pose determination such as [5, 7] require the detection of higher order image features (e.g., vertices), entailing combinatorial searches.

The algorithm was mainly motivated by work on automatic monitoring and surveillance of road and airport ground traffic. However, it may readily be adopted to other applications where objects are in contact with a known plane.

7 References

1. L. G. Roberts, Machine Perception of Three-Dimensional Solids, in Optical and Electro-optical Information Processing, J. Tippet *et al.* (Eds.), Cambridge, MA: MIT Press, 1965, pp.159-197.
2. P. Suetens, P. Fua and A. J. Hanson, Computational Strategies for Object Recognition, ACM Computing Surveys, vol.24, no.1, 1992, pp.5-61.
3. D. H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition, vol.13, 1981, pp.111-122.
4. W. E. L. Grimson and D. P. Huttenlocher, On the sensitivity of the Hough transform for object recognition, PAMI, vol.12, 1990, pp.255-274.
5. T. M. Silberberg, D. A. Harwood and L. S. Davis, Object recognition using oriented model points, CVGIP, vol.35, 1986, pp.47-71.
6. P. C. Gaston and T. Lozano-Perez, Tactile recognition and localization using object models: The case of polyhedra on a plane, PAMI, vol.6, 1984, pp.257-265.
7. J. L. Mundy and A. J. Heller, The evolution and testing of a model-based object recognition system, Proc. of ICCV90, Osaka, Japan, 1990, pp.268-282.
8. G. D. Sullivan, Visual Interpretation of Known Objects in Constrained Scenes, Phil. Trans. Royal Soc. London, Series B: Biol. Sci., vol.337, 1992, pp.361-370.
9. T. N. Tan, G. D. Sullivan and K. D. Baker, Linear algorithms for object pose estimation, Proc. of BMVC92, 1992, pp.600-609.
10. T. N. Tan, Computing the PDF of the Object Orientation Recovered from a Single Noisy 2D-3D Line Match, Esprit Project P2152 Research Report, RU-03-WP.T3137-TNT-04, University of Reading, January 1993.
11. Z. Hussain *et al.*, Knowledge Based Image Processing: Feature Based Methods, Esprit Project P2152 Report, R122/1, 1990.
12. T. N. Tan, Object Recognition and Localisation under the Ground Plane Constraint, Esprit Project P2152 Research Report, RU-03-WP.T3137-TNT-02, University of Reading, October 1992.