# Model-based Tracking

A. D. Worrall, R. F. Marslin, G. D. Sullivan, K. D. Baker
Intelligent Systems Group
Department of Computer Science,
University of Reading, RG6 2AX, UK.
Anthony.Worrall@Reading.ac.uk

**Abstract**

Model-based vision techniques originally developed for the recognition and pose recovery of a vehicle in a single image, are used here to track a vehicle through a sequence of images. The knowledge of the position of the camera with respect to the ground plane is used to reduce the search space of possible vehicle positions from six dimensions to three.

## 1 Introduction

Model-based vision allows use of prior knowledge of the shape and appearance of specific objects to be used in the machine interpretation of a visual scene. Over the last year or two a variety of model-based methods for object tracking have been reported [1, 2, 3, 4]. In all these cases the images used were indoor laboratory images tracking simple objects. Here, model based methods are used to classify and track moving complex objects (cars and other moving vehicles) in an uncontrolled and cluttered outdoor scene. This work was carried out under ESPRIT P2152 VIEWS [5].

### 1.1 Overview of methods

The models consist of precise three dimensional geometrical representations of known objects (vehicles), which can be placed in arbitrary positions and orientations, together with a carefully constructed camera model and scene model [6]. Using the known camera and scene geometry, and given a provisional position and orientation, a three dimensional object can be accurately projected onto the two dimensional image plane and a "goodness-of-fit" score obtained by comparing the modelled features with the underlying image.

A search in position-space and orientation-space is then used to maximize this evaluation score. At each position and orientation in this search the model must be re-instantiated onto the scene and a new goodness-of-fit score evaluated. This is essentially a "gradient ascent" performed on the evaluator function. Once a maximum score is found the three dimensional position and orientation of the object is known and is used to predict a provisional position and orientation for the same object in the next frame of the

BMVC 1991 doi:10.5244/C.5.39

scene. Thus by being given an initial frame in the sequence, an object to track and an initial position it is possible to continue to track that object in subsequent frames while at all times retaining knowledge of its location and direction of travel in three dimensions.

In this paper we do not discuss the problem of how the model is selected or how the initial estimate of the position and orientation of the object is obtained.

## 1.2 Evaluation of the "goodness-of-fit"

The model comprises a set of line features specified in a three dimensional object coordinate reference frame. On instantiation the model is translated and rotated to the appropriate position in the real world and finally each line which is visible from the given camera position is projected onto the image. The evidence for each of these visible line features is then evaluated in the gaussian smoothed image (using a process called iconic evaluation) and the scores from all of the lines are aggregated to give an overall score for the model in the given position. This technique has been reported in previous BMVA conferences [7, 8, 9].

## 1.3 Pose recovery

The evaluation score defines a scalar function of six dimensions - in world coordinates these are most simply defined as the three cartesian coordinates of the object's position and the three angles needed to specify its orientation. In general, we expect that peaks in the six degree of freedom function will indicate likely matches between the model and the image. The problem is to locate the peaks, and thereby to determine the pose of the vehicles.

A considerable computational simplification can be made by limiting the object's position to the ground plane thus only permitting two dimensions of translation and one of rotation about the vertical axis. Using these simple but realistic physical assumptions, only three independent dimensions remain. Even so, an exhaustive search of three dimensions is computationally too expensive. We have therefore developed a method for decomposing the problem, by successively performing three one dimensional searches.

## 1.4 Separated gradient ascent

The separated gradient ascent algorithm is applied iteratively to find peaks in the evaluation function. Each iteration consists of three separate one dimensional searches, each search taking a number of samples either side of the initial position. This is illustrated for the x-coordinate in Figure 1.

For each sample the model is displaced by an appropriate amount (in the world coordinate frame), instantiated onto the image, and an evaluation made of its fit to the gaussian image. The results for a typical search along a single dimension are shown in Figure 2, where the abscissa represents the displacement and the ordinate represents the evaluation score obtained (the higher the score the better the fit). The best score, and its

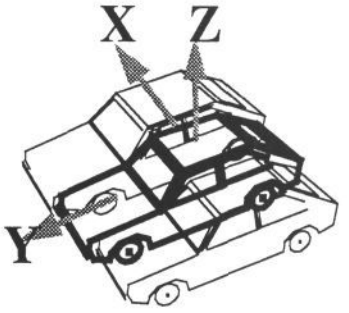



*Figure 1*: For each degree of freedom the model is displaced, instantiated and its "goodness-of-fit" evaluated

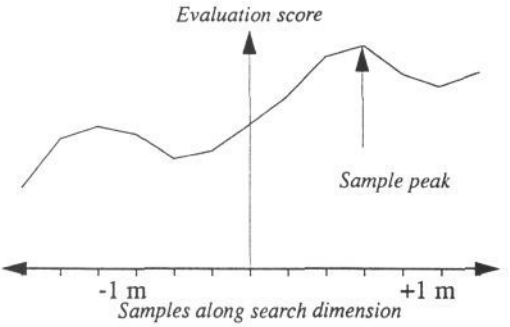



*Figure 2*: Evaluation scores obtained from a model by displacement along a single dimension

position, are noted and the process is repeated for the other two variables - each time starting from the same initial position.

When all three dimensions have been searched the position with the highest score is adopted, and becomes the initial position for the next iteration. If no higher score is found then the three ranges of the search are reduced. Since a fixed number of tests are made along each dimension this reduces the distance between the samples. Then the process repeated until all the ranges are below predetermined empirical values.

# 2 Results

A number of experiments have been carried out to explore the sensitivity and selectivity of the iconic evaluation process. Figure 3 shows a typical image from the

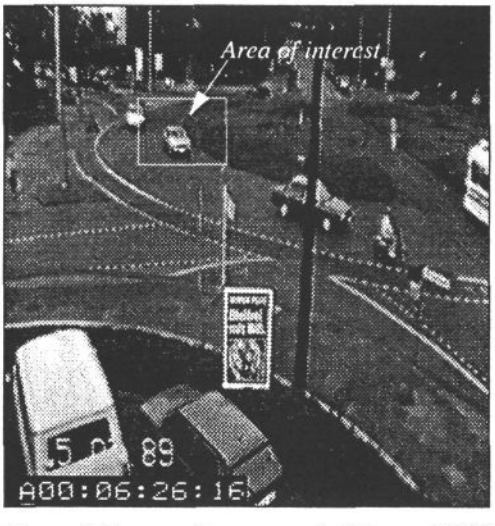
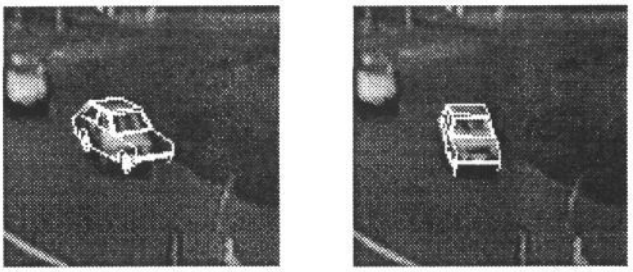



*Figure 3*: Bremer Stern scenario 3 frame 00200



A: frame 200 B: frame 200

*Figure 4*: (A) Before and (B) after gradient ascent.

Stern roundabout in Bremer, Germany. We concentrate on the results of tracking the car that is in the area of interest shown.

## 2.1 Evaluator performance

Figure 4 (A) shows a model that was very approximately instantiated on the vehicle by hand. This is fairly typical of an initial estimate of the position obtained from a simple two dimensional cueing process. The gradient ascent algorithm was allowed to run, and the result is shown in Figure 4 (B). Informally, by eye, it seems that the fit is very good.

Figure 5 shows the evaluation score for a typical two dimensional slice through the three dimensional search space. In this case the slice is in the plane of the two translations X and Y, and the range of translation of the model is ±3.2 metres about the peak.

To estimate the overall signal-to-noise ratio of the method, we have taken the best fitting position and orientation for the model on this image (as illustrated in Figure 4 (B)), and then evaluated this particular model instance on the entire sequence of images. The results are shown in Figure 6. We see one conspicuous peak, with several subsidiary

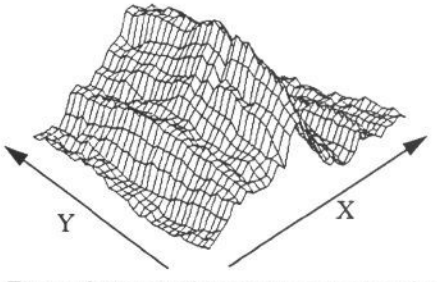



*Figure 5*: Evaluation score plotted against X and Y

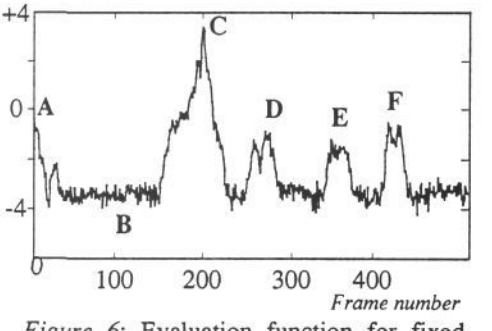
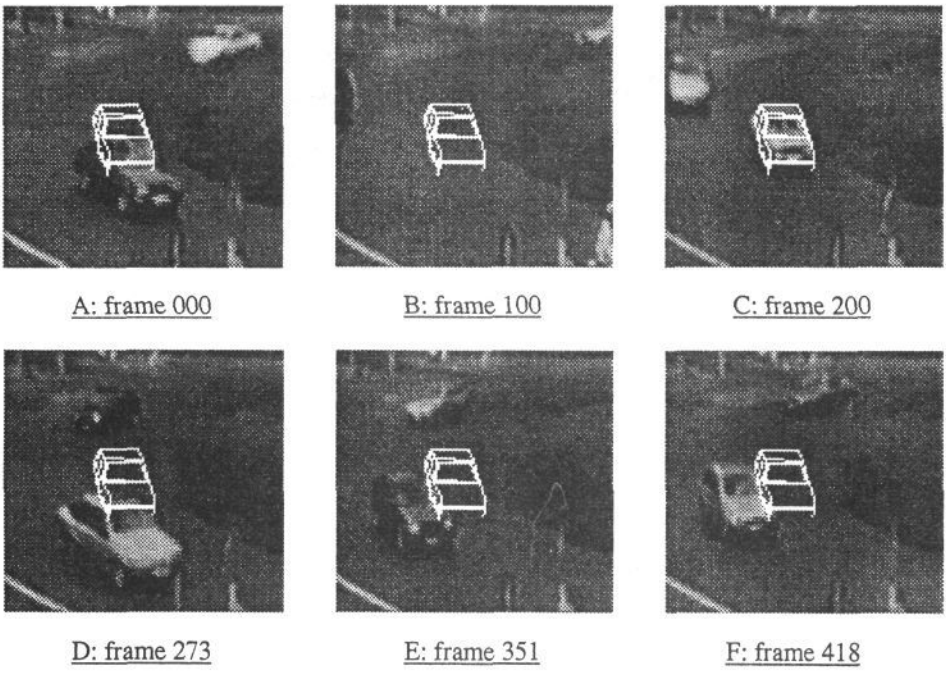



*Figure 6*: Evaluation function for fixed position on different images (A-F refers to Figure 7).



A: frame 000 B: frame 100 C: frame 200

D: frame 273 E: frame 351 F: frame 418

*Figure 7*: Fixed model in different frames (A-F refer to Figure 6).

peaks. The images corresponding to the points on the function labelled A to F are shown in Figure 7 (A-F).

The response of the evaluator is typically about -3.5 in the absence of vehicles. The minor peaks correspond to accidental alignments between the model and a "wrong" vehicle, but these scores never exceed 0.0, and the peaks are fairly flat and ragged. The score for the "correct" vehicle reaches 3.7, and is well-localised. In this simple case, where there is little distracting background image detail, the evaluator provides an intuitively acceptable measure of signal to noise ratio.

## 2.2 Results of model-based tracking

The three degree of freedom evaluation maximisation method has been used to track all the cars in the mid-ground and foreground of the sequence. The initial position of the model was first fitted very approximately by eye as each vehicle first came into view (as in Figure 4(A), but usually at a considerably further distance from the camera), and a gradient ascent performed.

Each car was then tracked through the image sequence, by using the current pose estimate to seed a 3 degree of freedom gradient ascent in the subsequent image. Each vehicle was located in the image separately with no account being taken of the vehicle's history or the other vehicles in the scene. The results are illustrated for the same vehicle

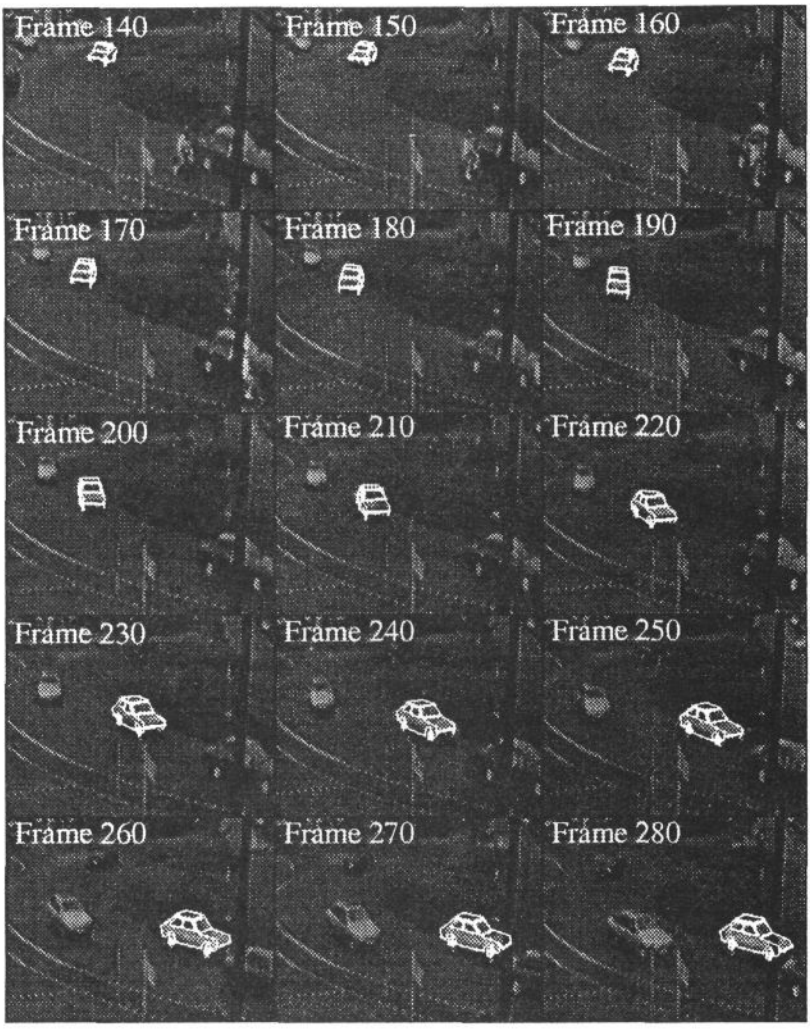



*Figure 8*: Tracking a single car from frame 140 to 280

as discussed above, in every 10th frame (i.e. every 400 msec) in Figure 8. In the main the performance seems very good, and the position and orientation of the vehicle is recovered with a fair degree of accuracy.

Figure 9 shows the score of the evaluator obtained for each of the frames in the sequence. It demonstrates that our feature pooling methods make the evaluation score at least partially independent of the pose of the vehicle before the camera, and compensate for the radical changes in the appearance of the images as the car rotates and comes closer to the camera. It should be noted that all these scores are well outside the typical subsidiary peaks shown in Figure 6.

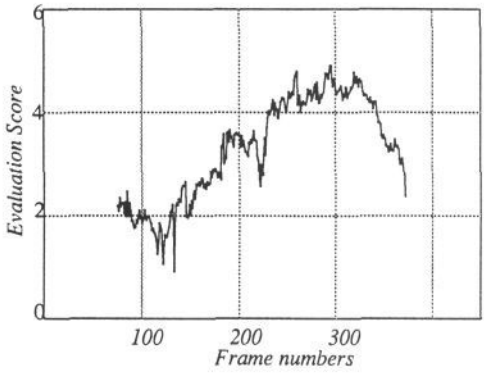



*Figure 9*: Evaluator score

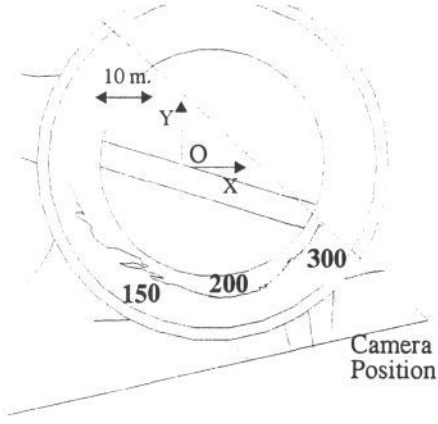



*Figure 10*: Track seen from above.

## 2.3 Model Tracks in Plan View.

The tracking algorithm recovers pose in world coordinates so we can represent the tracks in a plan view superimposed on a map of the scene, see Figure 10. Several anomalies can now be seen, where the recovered position oscillates. This is especially troublesome where the vehicle is seen head-on, around images 160-180. The ridge-like evaluation function obtained here (cf. Figure 5) makes the recovery in the model-Y direction (which in this orientation corresponds to depth) unstable.

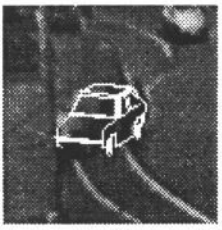
A more important problem occurs at the end of the track of another object; samples of the poses recovered are illustrated in Figure 11. The vehicle is tracked well until it approaches the lamppost (frame 145). At this point the strong edges in the image due to the lamppost cause the tracking to go awry. The model is subsequently "captured" by the white lines of the cycle track. The recovered pose gets progressively worse, and the model diverges rapidly along the cycle track, until it ends up trapped on the petrol advertisement in the foreground. Unlike the oscillatory behaviour of the first vehicle, in this case the correct pose could not be recovered. Needless to say, this behaviour involves absurd acceleration.

Further tests have been carried out to determine the robustness of this tracking algorithm. In one series of tests the position selected as the initial position for the start of tracking was varied. In another series of tests gaussian noise with a standard deviation of up to 40 grey levels was added to the image. The results of these tests were qualitatively similar to the results presented here.

Improved methods which allow physical constraints to be used during the tracking process itself are discussed in companion paper [10] and remove many of the problems associated with this naive tracking method.

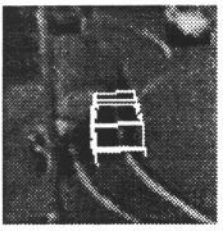
*Score = 2.446*

A. frame 00145



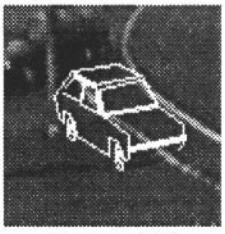
*Score = 1.293*

B. frame 00147



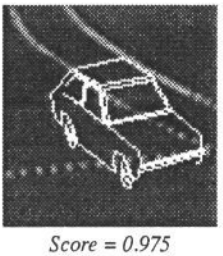
*Score = 1.305*

C. frame 00150



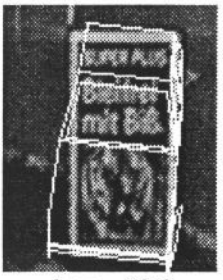
*Score = 1.453*

D. frame 00159



*Score = 0.975*

D. frame 00161



*Score = 7.218*

F. frame 00164

*Figure 11*: Unrecoverable error while tracking.

# 3 Conclusion

Model-based vision techniques, originally developed for the recognition and recovery of the pose of a vehicle in a single image, have been applied to the problem of tracking a vehicle through a sequence of images. The knowledge of the camera calibration and the ground plane were used to reduce the search space of possible vehicle positions from six dimension to three dimensions. On the whole this simple application of previous methods to the tracking problem succeeded rather well. It failed in cases when there are strong "distractor" features (such as lampposts or other vehicles). A tracking mechanism that retains more information about the vehicle's history can be used to overcome many of these problems. The model system can also be used to take into account occlusions due to known objects either fixed in the scene or other vehicles that are been tracked.

At the moment the current tracker is far from working in "real-time". During the gradient ascent around 100 evaluations are performed. Each model instantiation and evaluation is performed independently and no attempt is made to preserve useful

information about individual feature scores. Work is currently directed to developing more efficient algorithms to make the tracker's performance approach real-time.

# References


[1] Bray A.J. "*Tracking Objects Using Image Disparity*", Image and Vision Computing, Vol. 8, No. 1, Feb. 1990, pp4-9.

[2] Stevens R. S. "*Real-time 3D object tracking*", Image and Vision Computing, Vol. 8, No. 1, Feb. 1990. pp91-96.

[3] Harris C. & Stennett C. "*Rapid - A Video Rate Object Tracker*", Proc. British Machine Vision Conference, BMVC-90, Oxford, 1990 pp73-77.

[4] Lowe D. "*Fitting Parametrized 3-D Models to Images*", IEEE-TPAMI, Vol. 13, No. 5, 1991 pp 441-450.

[5] Hussain Z., Godden R., Sullivan G. D., Worrall A. D. & Marslin R. F. "*D102: Knowledge-based Image Processing*", ESPRIT P2152 deliverable PM-03-CEC.D102-01. Dec. 1990.

[6] Worrall A. D. "*Facet Based Model for Object Identification*", Alvey MMI-007 Technical Report, Reading University, 1987.

[7] Brisdon K. "*Alvey MMI-007 Vehicle Exemplar: Evaluation and Verification of Model Instances*", Proc. Alvey Vision Conference, AVC-87, Cambridge, 1987, pp33-37.

[8] Brisdon K., Sullivan G. D. & Baker K. D. "*Feature Aggregation in Iconic Model Matching*", Proc Alvey Vision Conference, AVC-88, Manchester, 1988, pp19-24.

[9] Brisdon K., "*Hypothesis Verification using Iconic Matching*" Doctoral Thesis, University of Reading, November 1990.

[10] Marslin R. F., Sullivan G. D., & Baker K. D. "*Kalman Filters in Constrained Model Based Tracking*" BMVC91, Glasgow, 1991.