A generic deformable model for vehicle recognition

J M Ferryman, A D Worrall, G D Sullivan and K D Baker Department of Computer Science The University of Reading

e-mail: J.M.Ferryman@reading.ac.uk

Abstract

This paper reports the development of a highly parameterised 3-D model able to adopt the shapes of a wide variety of different classes of vehicles (cars, vans, buses, etc), and its subsequent specialisation to a generic car class which accounts for most commonly encountered types of car (including saloon, hatchback and estate cars). An interactive tool has been developed to obtain sample data for vehicles from video images. A PCA description of the manually sampled data provides a deformable model in which a single instance is described as a 6 parameter vector. Both the pose and the structure of a car can be recovered by fitting the PCA model to an image. The recovered description is sufficiently accurate to discriminate between vehicle sub-classes.

1 Introduction

We have previously demonstrated the use of model-based vision for the location, recognition and tracking of vehicles [5][10]. Our techniques have used specially constructed models of 3-D objects, represented as points, lines, and surfaces, together with some topological structure [9]. The main purpose of the models is to support reasoning about the visibility of the lines as the models translate, rotate and mutually occlude each other before a calibrated camera system. This allows both bottom-up [8] and top-down [6] strategies to visual object recognition.

A major benefit of model-based vision is that it becomes easy to aggregate all image evidence that is believed to be relevant to a given recognition hypothesis. The usual technique relies on using the hypothesis (of object class and pose) to back-project an instantiated model into an iconic representation of image data. All evidence that falls near to the features of the model is then pooled. In our case, we back-project all visible model lines, and seek evidence of local image derivatives, perpendicular in the image to the lines. Such evidence can then either be aggregated to form a "goodness of fit" evaluation score, which can be searched for local maxima, or it can be used to create an error term, which can be linearised to determine the best local solution by least squares techniques. A companion paper [12] gives more details, and reports experiments to compare the relative advantages of the two methods.

A second advantage of model-based vision is that it becomes easy to impose external constraints on the possible poses of objects. In the case of road traffic this is particularly useful, since vehicles are normally confined to be upright in contact with the

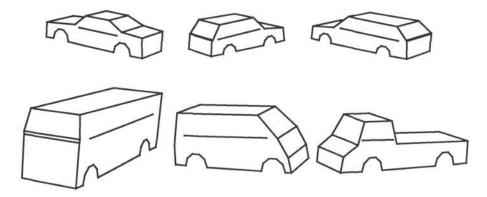


Figure 1 Examples of the 29-parameter model.

Top: Saloon, hatchback and estate sub-classes of the general car class.

Bottom: Other classes of common vehicle.

road, and this ground-plane constraint (GPC) reduces the pose recovery problem from 6 to 3 degrees of freedom: translations (X,Y) on the ground-plane and rotation (θ) about the vertical axis [7].

However, a major deficiency in our previously reported techniques is that the vehicle models are rigid. Success depends critically on how accurately the model captures the geometry of the object in view. In the limit, this might mean that a model for every different possible type of vehicle must be treated separately. Since processing time is approximately linear in the number of models, this is infeasible.

In practice, the need to assess multiple models can be greatly reduced by two factors:

- Low level cues (i.e. image attributes derived by data-driven processing) can be used
 to prescribe a subset of plausible vehicle types. We have demonstrated methods
 based on the detection of moving regions by image differencing. A region's area,
 eccentricity, direction of movement, and location in the image (which under the GPC
 can easily be related to distance from the camera) are used to index into pre-compiled
 tables of likely vehicles [6].
- 2. The search for image evidence includes a deliberate degree of "fuzziness", which reduces the sensitivity of the system to the precise dimensions of the vehicle model. This is effected by searching for derivatives within a finite interval along the normals to the projected model lines.

Any reliable way to use context-free reasoning (1, above) to reduce context-dependent processing time is extremely valuable. It just happens to be very difficult to do, and very wasteful of processing resources; this is why we seek to use context-dependent methods as soon as possible.

Reducing sensitivity to geometry using the fuzzy model (2, above) is also less than satisfactory since the object-selectivity of a model degrades rapidly as the imprecision of the evaluation is increased. Even within one class of vehicle (e.g. a saloon car) a single rigid model is only poorly able to cope with the range of vehicles encountered in typical road scenes. We have previously extended the rigid model by allowing it to

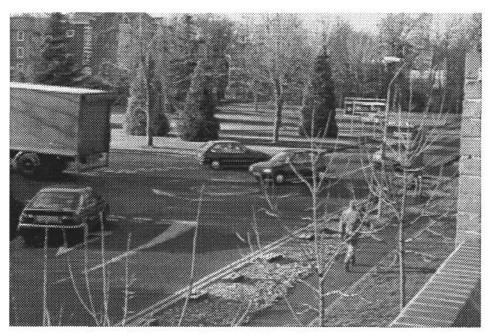


Figure 2 Test scene for sample data collection.

scale independently in its 3 dimensions, which then become 3 further parameters to search. Such anisotropic scale changes have proved useful, but they have no special relevance to the task and do not allow for style changes between different marques and types of saloon cars. For example, the wheel-base of a large car is usually a smaller fraction of overall length than that of a small car, and this consistency is not captured by scale changes alone.

The approach reported here uses a parameterised vehicle model, able to represent different classes of vehicle as separate instances of the single model (see Figure 1). The model contains 29 nominally independent parameters. These define a configuration space that is far too large to search automatically. We therefore simplify the model, based on statistical information, as introduced by[1][2][4]. To collect the required training data, we have developed a way to fit the model to images under partial interactive control, to sample the variation in vehicle shape encountered. We have then used Principal Component Analysis (PCA) to capture most of the variability of structure encountered for a single class of vehicle within a few "modes of variation".

2 The parameterised model

The underlying geometrical model is based on a bilaterally symmetrical, extruded 8-sided polygon. A few of the more meaningful instances of the model are illustrated in Figure 1. The model can be conveniently parameterised with respect to a model-centred coordinate system (x,y,z), whose yz-plane is the symmetry plane, and whose xy-plane is constrained to lie on the ground plane in the scene coordinate system (XY). The 8 vertices in the positive x direction are free to move in (x,y,z), though will be strongly constrained

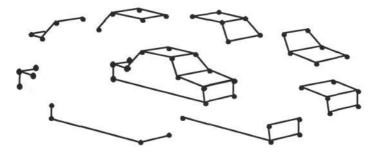


Figure 3 The 8 sub-models (all seen from one view, with hidden lines removed). as explained below. The vertices in the negative x direction are at corresponding points (-x,y,z). An additional line, corresponding to the bottom of the side windows, is also modelled using 2 further parameters, and two semi-decagonal wheel arches are added to the bottom of the car, with variable (equal) radius and longitudinal position (see below).

The model therefore has 8*3 + 2 + 3 = 29 structural parameters, as well as the 3 pose parameters (X,Y,θ) in the world coordinate system. This is far too large a configuration space to search naively. However, it is evident that there are very strong constraints between the structural parameters if the model is to be remotely like a vehicle. These constraints can be expressed by carrying out a PCA of a statistical sample of model parameters, which conform to the shape of actual vehicles. The sample data were obtained by using a specially designed interactive tool.

3 Data collection

Figure 2 shows a test scene within the University grounds, which was video-taped over a 4 hour period. The camera was calibrated using an interactive tool [11].

The parameterised model was fitted to individual images of vehicles in the following way. Firstly, the class of vehicle (saloon, hatchback or estate) was identified by eye, and a standard instance of this class was projected onto the ground plane in the image. The pose parameters (X,Y,θ) were then set by eye to provide an approximate initial fit to the image. The pose of this fixed shape model was then refined using the passive method described in [12]. This provided a first estimate of the pose and shape of the vehicle.

The model was then decomposed into 8 separate sub-models, illustrated in Figure 3^1 . Each sub-model comprises two adjacent quadrilaterals, parallel to the axis of extrusion (x) of the model, and is therefore defined by three adjacent vertices of the polygon. Each sub-model was deformed in turn, by allowing the middle vertex (and its mirror image) to vary in (x,y,z), while keeping the four other vertices fixed. Thus the fold line between the two quadrilaterals moved in 3D, while the bilateral symmetry of the sub-model was maintained.

^{1.} This approach is equivalent to successively modifying each point, but was a more natural way to utilise existing model-based vision software.

The position of the central vertex having the highest evaluation score was found by a constrained 3D separated search (as in [6]). Note that the outer parallel lines of a submodel were unchanged in the search, so did not contribute to the evaluation of the submodel. Each sub-model was treated in this way, using seed values defined by the initial pose and structure estimate. The resulting set of vertex positions (in the model-centred frame) was then used to define a new vehicle structure, still at the initial pose. This became a new structure for pose refinement, and the whole process was repeated. The fitting procedure was carried out under interactive control, so that the occasional bizarre mistake could be set right, and the stability of the recovered structure of the vehicle could be monitored.

Having recovered the outline geometry of the model, the side window line was then created in 3D between the bottoms of the front and rear windscreens, and the z-coordinates of its two ends were varied independently to obtain the best evaluation score. Finally, two equal-radius wheel arches were centred on the bottom sill line, and fitted similarly by varying position and radius.

The final 29 parameter instance is expressed with respect to the coordinate frame of the original class exemplar. Since the model deformation technique is incremental, it is possible that the model has moved in an uncontrolled way with respect to this coordinate frame. This would then equate to a change of pose and would be an irrelevant contamination of the structure data. In fact we need only worry about the position of the model along the y (front-to-back) axis, since the fitting procedure always maintained the coordinate frame on the ground-plane, symmetrically within the model. The redidual shift was therefore eliminated by repositioning the object coordinate frame to the mean y-position.

A particular merit of our supervised model-fitting approach to collecting representative sample data is that the final instances are obtained using image-fitting criteria, which are very similar to those that will later be used automatically. This avoids any bias which might be introduced by methods that attempts to recover the 3-D structure of the vehicles by other means (e.g. solving structure equations after identifying, by hand, sets of corresponding points in multiple images). Our data represent the best fitting instance of the 29 df model (given the constraints implicit in the process, and the supervisor's ability to detect and intervene in anomalous circumstances).

In the initial study reported here, a total of 90 examples of cars were processed in this way, 30 each of the three sub-classes illustrated at the top of Figure 1 (saloon, hatchback and estate cars). We deliberately chose diverse examples of vehicle makes and colours within each of the three sub-classes.

4 Stability of the recovered structure

We would like the structure recovery technique to be independent of irrelevant viewing factors. These include the colour of otherwise identical vehicles, the lighting and background conditions, and the pose of the vehicle in the particular image used. Only the last factor proves easy to investigate given our experimental facilities. A video sequence of a single vehicle coming towards the camera, and then turning right around the roundabout was selected, and sampled 30 times using every second frame. Each sample was subjected to a model-fitting similar to that outlined above, but considerably



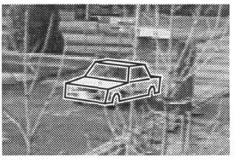


Figure 4 First (right) and last (left) frames of sequence used to test the stability of the structure recovery technique.

automated: the typical exemplar of the saloon car sub-class was first positioned approximately on each image, then the 29 df model-fitting process was carried out entirely unsupervised. The first and last frames of the sequence are shown in Figure 4, with the recovered structural models superimposed.

There are a number of measures that can be used to estimate the stability of the 29 parameter structure obtained. Given that our purpose is to generate a PCA model which captures the variation between different objects (see next section) our greatest concern is to minimise the variation introduced by our measurement of a single object.

Population sampled	Sample variance (m ²)	
30 different Saloon cars	1.24	
30 different Hatchback cars	1.55	
30 different Estate cars	1.84	
90 different cars (all the above)	2.11	
30 samples of the same Saloon car	0.238	

Table 1: Sample variance for sub-classes of vehicle.

Table 1 gives the amount of variance of the sampled populations of vehicles. It can be seen that the variability of the structure recovered from the single saloon car is considerably less than that of the set of saloon cars originally sampled, which is in turn smaller than the variance of the entire car population. This indicates that the 29-parameter model fitting process is reasonably stable (w.r.t the sub-class of the vehicle).

5 Principal component analysis

The structural data provided 90 independent samples of the 29-D configuration space for the general car class. This was subjected to a principal component analysis using MATLABTM, according to the method outlined in [3]. The covariance matrix was

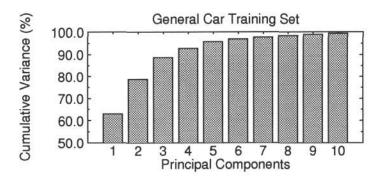


Figure 5 Percentage of the variance in the 29 df parameter model explained by the first 10 principal components in the general car vehicle class.

computed, and its eigenvectors were obtained and ranked according to magnitude. Figure 5 shows the cumulative percentage of the overall variance (for all 90 vehicles) explained by the largest n eigenvectors. More than 97% of the variation observed can be represented by just the 6 strongest eigenvectors.

The physical meaning of these 6 eigenvectors is illustrated in Figure 6. The top illustration shows the mean vehicle, and below are shown examples of the PCA model displaced from the mean by +/-2 standard deviations along each of the 6 main eigenvectors. The main modes of variation seem to conform well to simple plausible changes in types of cars: the first mode mainly codes overall length, the second accounts for much of the change of the rear between saloon and other cars, the third is dominated by width, the fourth by the relative sizes of the bonnet and roof, while the fifth and sixth also affect the shape of the rear.

It is of interest to consider how well the 6 main PCA parameters are able to discriminate between the three sub-classes of car (hatchback, saloon and estate) which made up the entire training set. The mean positions in the (normalised) 6 df PCA space of each sub-class were computed to determine prototypes of the 3 sub-classes. and all instances were classified by a simple nearest-prototype algorithm.

Classed	True class			
as	Saloon	H'back	Estate	
Saloon	27	0	1	
H'back	2	29	3	
Estate	1	1	26	

Table 2: Confusion matrix for training sets, using nearest-prototype classifier in the 6 df PCA space.

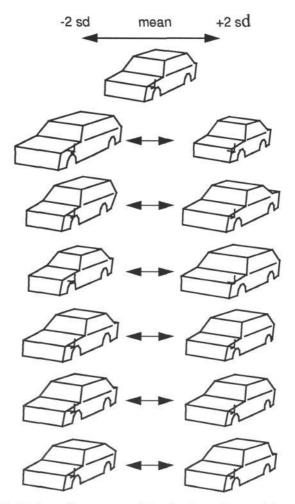


Figure 6 Deformable car model, showing the model coordinate frame, on the ground, in the symmetry plane.

Top: Mean of distribution.

Rows 2-7: +/- 2sd along the 6 principal axes.

Table 2 shows the confusion matrix obtained. It can be seen that more than 90% of the vehicle instances were "correctly" classified by the shape parameters. This is particularly interesting since the vehicle sub-classes were identified by eye, mainly according to the investigator's knowledge of the specific vehicle type, and not strictly according to the shape of the vehicle. Thus a saloon car is distinguished by the fact that the rear window is not part of the boot, whereas there is a fairly continuous variation in apparent shape between the three sub-classes. Section 6 describes the use of the PCA model automatically in this way to classify vehicles which were not part of the original training set

6 Classification using recovered structure

Our primary objective in this work is to develop a generic car model whose structure could be refined in response to the image, and thus enhance the performance of model-based vision. The companion paper [12] reports experiments on recovering the structure parameters of vehicle models, and of the benefits which accrue to pose refinement.

A second important visual task is that of classification. We have investigated how well the generic car PCA model is able to classify vehicles into the three sub-classes (hatchback, saloon and estate), simply on the basis of the deformation parameters recovered automatically. The mean generic car model was first placed by hand in images of vehicles, which did not form part of the original sample data. This identified an approximate pose for the generic model. The pose was first refined, using the generic model as a rigid object, and then the structure was refined (as in [12]).

The resulting PCA coefficients were classified by a simple nearest prototype rule as in section 5. The results are shown in Table 3.

Classed	True class			
as	Saloon	H'back	Estate	
Saloon	13	5	2	
H'back	8	15	10	
Estate	11	9	13	

Table 3: Confusion matrix for test sets, using automatically recovered structure parameters.

Unlike Table 2, none of these test cases were members of the sample data which led to the prototypes, and the structure parameters were recovered automatically. Furthermore, the test examples were selected less carefully than for the training examples, with more cases in the distance and partly obscured by the trees and other cars (Figure 2). In view of this, performance seems reasonable. We are currently exploring techniques for filtering the PCA parameters over image sequences to improve recognition.

7 Conclusion

We have described a highly parameterised vehicle model, and shown how it can be specialised by Principal Component Analysis to become a deformable generic car model. This successfully accounts for the variation between different instances of three car subclasses (saloons, estates and hatchbacks). The PCA model defines a 6 df configuration space, which can successfully be searched automatically, using the methods reported in [12]. The structure parameters recovered allow vehicle instances to be classified as members of the 3 sub-classes.

References

- [1] Cootes, T. F., Cooper, D. H., Taylor, C. J. and Graham, J. A Trainable Method of Parametric Shape Description, Proc. 2rd British Machine Vision Conference, pp 54-61, 1991.
- [2] Cootes, T. F., Cooper, D. H., Taylor, C. J. and Graham, J. Training Models of Shape from Sets of Examples, Proc. 3rd British Machine Vision Conference, pp 9-18, 1992.
- [3] Jolliffe, I. T., Principal Component Analysis, Springer-Verlag, New York, 1986.
- [4] Lanitis, A., Taylor, C. J. and Cootes, T. F. An Automatic Face Identification System Using Flexible Appearance Models. Proc. 5th British Machine Vision Conference, 13-16 September 1994, University of York, York, pp 65-74, 1994.
- [5] Sullivan, G. D. Model-based Vision for Traffic Scenes using the Ground-plane Constraint, Real-time Computer Vision, Eds: C Brown and D Terzopoulos, CUP, 1995.
- [6] Sullivan, G. D. Visual interpretation of known objects in constrained scenes, Phil. Trans. R.Soc. Lond., B, 337, pp 361-370, 1992.
- [7] Tan, T. N. Sullivan, G. D. and Baker, K. D. Recognising Objects on the Ground-Plane, Proc. 4th British Machine Vision Conference, 1993.
- [8] Tan, T. N. Sullivan, G. D. and Baker, K. D. Fast Vehicle Localisation and Recognition Without Line Extraction and Matching, Proc. 5th British Machine Vision Conference, pp 85-94, 1994.
- [9] Worrall, A. D., Baker, K. D. and Sullivan, G. D. Model-based perspective inversion, Image and Vision Computing Journal, 7(1), pp 17-23, 1989.
- [10] Worrall, A.D., Sullivan, G. D. and Baker, K. D. Advances in Model-based Traffic Vision, Proc. 4th British Machine Vision Conference, pp 559-568, 1993.
- [11] Worrall, A. D., Sullivan, G. D. and Baker, K. D. A simple, intuitive camera calibration tool for natural images, Proc. 5th British Machine Vision Conference, 13-16 September, University of York, York, pp 781-790, 1994.
- [12] Worrall, A. D., Ferryman, J. M., Sullivan, G. D. and Baker, K. D. Pose and structure recovery using active models, This Volume.