Relational Model Construction and 3D Object Recognition from Single 2D Monochromatic Image

S. Zhang, G. D. Sullivan, K. D. Baker

Intelligent Systems Group
University of Reading, UK, RG6 2AY
zsj@uk.ac.reading

This paper reports a method for automatically constructing a relational model of a rigid 3D object, to represent view-independent relations among its component parts, and of using such a model to recognise the object from single monochromatic images. The relational model is a graph associated with procedural constraints. It is constructed by a statistic analysis of images generated by projecting a CAD model of the object from a set of viewpoints on the Gaussian viewsphere. Object recognition is achieved by a hypothesis-and-verification process. Extended hypotheses are generated by aggregating image features which satisfy the view-independent constraints. These hypotheses are then verified by projective inversion and 3D grouping to achieve object recognition.

We illustrate the approach by means of the recognition of a hatchback model car. The method can readily be adapted to the recognition of any object defined geometrically.

1. Introduction

There has been a long history of using relational models in object recognition. Relational models capture an object, independent of its pose, so are especially suitable for recognising complex objects in unknown environments. However, this kind of model (e.g. [2,7]) has usually been constructed manually, and this severely limits the practical application of relational models in object recognition. This paper discusses the problem of automatically constructing a relational model of an object and its use in object recognition. The aim of this work is to develop a vision system which accepts as input a CAD wireframe model of an object and single monochromatic images of the object within an unknown environment, and produces as output a model instance of the recognised object superimposed on the image. No *a prior* information about the location or orientation of the object relative to the viewer is used, other than to assume that the viewpoint is within 60° above the horizontal.

Prior to the recognition process, a relational model of the object is built off-line, based on the wireframe model. The model is a graph associated with procedural constraints. The nodes of the graph are 3D component parts (model features) of the object. These can be associated with groups of image features defined by simple 2D geometrical attributes. A co-visibility constraint for model features is represented by means of an arc of the graph. Other pairwise view-independent relations are represented as constraints associated with these arcs. This relational model is then used to identify extended groups of 2D image features and to form hypotheses about the object without invoking explicit pose information in the matching process. Finally the hypotheses are verified by perspective

BMVC 1991 doi:10.5244/C.5.31

inversion to achieve object recognition. The recognition process is closely related to the work of Lowe [6], Goad [5], and Flynn & Jain [4].

2. Relational Model

Fig.1(A) shows a wireframe model of a hatchback model car, comprised of 22 model features, which in this study were identified manually, to be used for object recognition. Model features are either single isolated lines or clusters of lines forming windows. Fig.1(B) shows a small part of the graph representation of the relational model of the object (representing the six windows of the car) constructed from the wireframe model. Each arc of the graph is associated with procedural constraints representing simple relations between the corresponding model features, chosen to be largely independent of view.

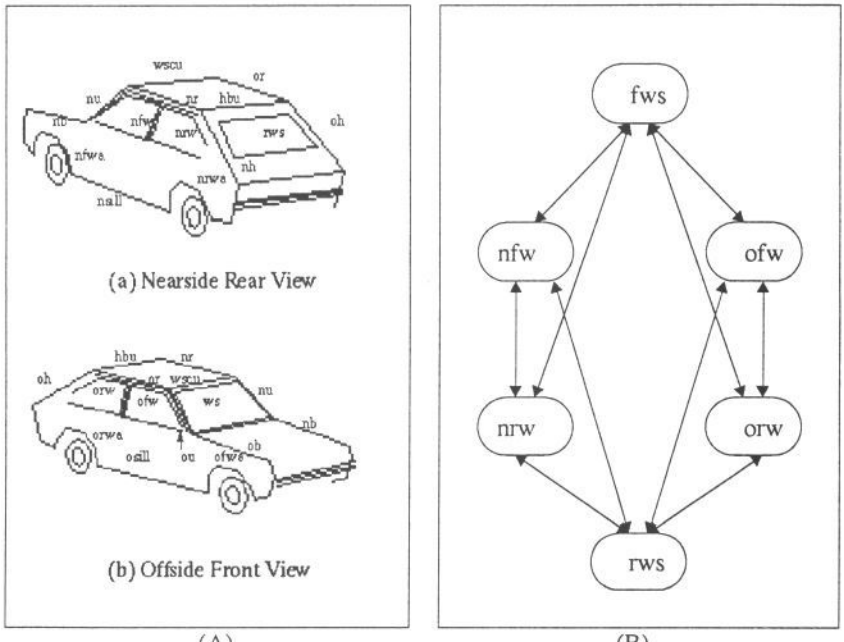


Fig. 1. Geometrical (A) and relational (B) models of a car

To build the relational model, the object is examined from many different viewpoints to collect statistical data about relations between pairs of components of the object. Five relations: co-visibility, colinearity, parallelism, side relation, and relative size are analysed and quantitative measures of these relations are defined.

As an example, we define the measure for colinearity by a co-line ratio as follows. Given two line segments *ab* and *cd* (assuming that *ab* is longer), as shown in Fig.2, we construct a minimal rectangle whose long axis is parallel to *ab* and encloses *ab* and *cd*. Let w be the length of the side of the rectangle parallel to *ab*, h be the length of the perpendicular side of the rectangle, and θ be the angle between the two line segments. The quantitative measure of colinearity between the two line segments is defined as:

$$colineratio\,(ab, cd) = \left| (1 - \frac{h}{w}) \cos\theta \right|$$

This heuristic provides an acceptable measure (between 0 and 1) for the concept of colinearity between two line segments.

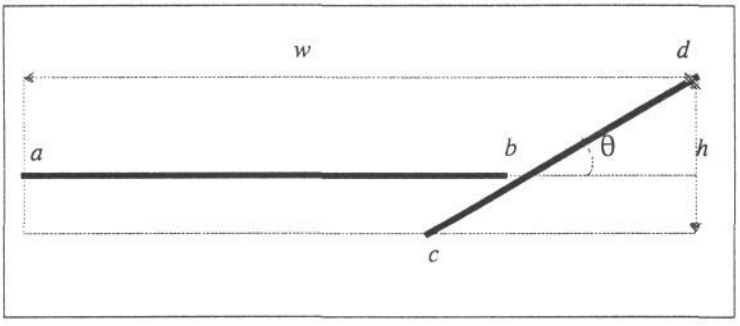


Fig. 2. Quantitative measure of colinearity between two line segments

Similarly, for any pair of line segments, we define a parallel ratio between 0 and 1 to express the degree of parallelism. The larger the parallel ratio, the more closely parallel are the two line segments. For a pair of features $\{a, b\}$, the side relation is represented by a tri-value function recording if a is to the right or left of b or they cross. Relative size is quantified by the ratio between the lengths of two features. Relations between compound features (e.g. quadrilaterals) are assessed by the relations between their component parts. Details of quantitative definition of these relations are omitted here due to space limitations.

From each viewpoint we construct an image of the wireframe model of the object and measure the selected relations (both visibility and geometrical constraints) among pairs of model features from that viewpoint. For the 22 model features, there are 231 pairs of features and therefore 1155 pairwise relations to be measured from each viewpoint. These data are subsequently examined to select those relationships which provide a view-independent description of the object.

2.1. Sampling the Viewsphere

We restrict our work to recognising cars in normal views (0°-60° from the horizontal). The surface of the Gaussian viewsphere is sampled randomly to select n (typically 500) viewpoints. The wireframe model of the object is inspected from each viewpoint to generate n sets of model features projected as 2D line segments. The measured relations between each pair of features from each view are stored to provide a statistical description of the relations between the features. This approach differs from the traditional aspect graph (e.g. [1,2]) in that the relational model involves no aspect information. This greatly simplifies the hypothesis generation process.

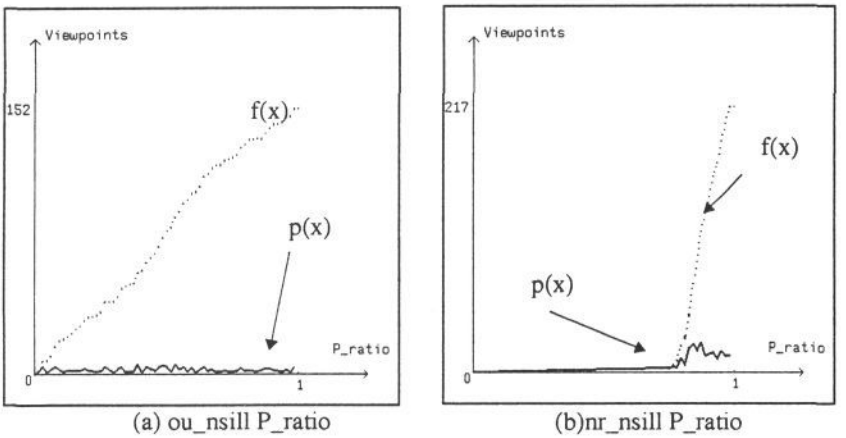
2.2. Building the Co-visibility Graph

The co-visibility of a pair of features is represented by the connectivity of the graph. Two model features $\{o_i, o_j\}$ are deemed co-visible if both the conditional probability of detecting o_j while o_i has been detected, and that of detecting o_i while o_j has been detected, are larger than a threshold σ. The conditional probabilities can be calculated by comparing the numbers of viewpoints from which o_i, or o_j or both can be seen. If $\{o_i, o_j\}$ are co-visible an arc is established between them. Each arc of the graph is weighted by the conditional probability of co-visibility, and associated with procedural constraints representing view-independent geometrical constraints on the corresponding model features. Using σ equal to 0.4 (i.e., arcs connect component that are both visible in at least 40% of the views from which either is visible), the number of arcs of the graph is 107.

2.3. Selection of Relational Constraints

The purpose of this stage is to identify those relations which would provide useful constraints on the hypothesis generation, or graph matching process. The principle is that if a measured relation is usually concentrated around a constant value, this relation is deemed to be independent of view, so that the constraint is useful in object recognition.

If two model features $\{o_i, o_j\}$are deemed co-visible, we evaluate the four quantitative measures between them. For each of the relations, if there are n_1 viewpoints from which both features can be seen, we have n_1 values of the relation between the two model features. These measurements give an estimate of the underlying probability distribution associated with this relationship between the feature pair.



(a) ou_nsill P_ratio (b)nr_nsill P_ratio

Fig. 3. Parallel ratios between feature pairs

As extreme examples, Fig. 3 shows the distributions of the parallel ratios between model features *{ou, nsill}* (the offside windscreen upright and the nearside sill) and *{nr, nsill}* (the nearside roof and the nearside sill). Black curves show the probability density functions (*pdf*) p(t) of the variable, the dotted curves show the cumulative *pdf*, that is,

$$f(x) = \int_0^{\lambda} p(t)dt$$

We can see that the parallel ratio of *{ou, nsill}* is widely distributed between [0,1] whereas that of *{nr, nsill}* is concentrated in the range from 0.8 to 1. The latter is accepted as a useful view-independent constraint on the feature pair.

More generally, let m_p be the mean value and σ_p be the standard deviation of the relational measure between s_i and s_j (the projections of model features o_i and o_j) and let e_p be the expected value when this relation is exact (for parallelism and colinearity, this is 1, for side relation this is either 1 or -1, for relative size, this value is substituted by m_p so the first expression in the following is satisfied automatically). If

$$(|m_p - e_p| < \varepsilon_1) \; and \; (\sigma_p < \sigma_1)$$

where ε_1 and σ_1 are pre-set thresholds, the relation between o_i and o_j is deemed as holding generally over all possible views.

This selection process was applied to 428 relations (4 for each of the 107 pairs of co-visible features). With the current set of thresholds used, there resulted a total of 171 useful constraints: 43 parallel constraints, 21 co-line, 72 side relation, and 35 relative size. The numbers of constraints generated in the model building process are dependent on the thresholds selected for the model generation. Such thresholds are inevitable with any recognition problem, and must be determined by experience. However, the influence of these thresholds on the final recognition are not dramatic since the relational model is designed to group 2D features into hypotheses which are then subjected to further evaluation. Minor changes in the relational model lead only to small changes in the time used for hypothesis generation and in the later assessment of the hypotheses generated. In our experiments, we have found that small changes of thresholds have little effect.

These relations provide procedural constraints which are attached to the corresponding arcs in the graph to form a relational model of the object. This model provides a set of weak constraints on model feature which can be used to aggregate 2D image features to generate a small number of reliable hypotheses about the location, orientation and size of the object. The time used for generating this relational model is significant (approximately one and half hours on a Sun4, using code written entirely in pop11) and is linearly dependent on the number of viewpoints selected. However, this model is compiled off-line and once built, it can be used to recognise any image of the object.

3. Feature Extraction

We use a heuristic approach to extract compound features from an edge map, derived using the method of Canny [1]. The features extracted from the image are line segments, U-shaped curves, and quadrilaterals. Details of the general approach has been given in [9].

4. Hypothesis Generation

A hypothesis is a group of pairings of 2D image features with 3D model features. Necessary conditions are that (1) all the model features in the group constitute a clique of the graph which is mutually view-consistent and (2) image features in the group as a whole satisfy all the procedural constraints concerned. In order to generate such hypotheses, features extracted from the image are matched against model features. To minimise the time used in hypothesis generation, we use a method employing best-first search, island-growing and early termination of the search. In our experiments, about 10 hypotheses satisfying the above criteria are typically generated by the search.

5. Hypothesis Verification

The hypotheses are subjected to further verification by model-projection followed by 3D grouping. For each hypothesis generated, we find the approximate location and orientation of the camera by a table look-up approach [9]. The CAD wireframe model of the instantiated object is then projected back into the image plane. Further image features are accepted which match the projected model line segments. If the matched image features exceed a pre-set threshold, the hypothesis is accepted and the object recognition is achieved, otherwise the hypothesis is rejected.

6. Results

Initial experience shows that the method proposed here works well. Fig. 4 shows the various stages of applying the method to a typical image. Owing to the noise in feature extraction and the resolution of the table look-up used in recovering the pose information, the model instances obtained have small displacements from the true position. This can be refined by using iconic evaluation, similar to that of Worrall [8] (to be implemented).

Image	Number of Correct Hypotheses	Number of Incorrect Hypotheses	CPU Time Used for Hypothesis Generation	CPU Time Used for Selecting Correct Hypothesis
(a)	2	8	2.23	1.95
(b)	3	10	2.42	2.10
(c)	4	12	2.59	3.01
(d)	2	10	1.90	2.05
(e)	3	7	2.08	2.12
(f)	4	7	2.45	2.59

Table 1. Summary of the Recognition Process of the Images in Fig. 4 Running on Sun4, 16 MB, times are shown in minutes

We have applied the relational model to 15 images from which reasonably reliable features are extracted. Fig.5 shows results of our approach on 6 of these images. A summary of the results is shown in Table 1. The numbers of features extracted from these images range from 120 to 285. In all the images at least one correct hypothesis is generated, together with typically fewer than 10 other incorrect hypotheses (arising either from clutter in the background or mislabelled objects). Experiment shows that this approach can generate the model instances fairly quickly (about two minutes, on a Sun 4 using pop11, for each of the two stages of (1) generating hypotheses and (2) selecting correct hypotheses from all the hypothesis generated) given that the original edge features have been extracted. No effort has been made at this stage to implement the code

efficiently. The underlying qualitative nature of this relational model makes it possible to recognise occluded objects, Fig. 5(f), or multiple objects in a scene, Fig. 5(c).

The approach discussed here can easily be used to recognise different objects. Currently we are working towards recognising other types of vehicles (Estate, Saloon, Van, etc). A drawback of this approach is that it relies closely on being able to find at least one 2D grouping of edges (e.g. quadrilaterals or U-shaped curves) and, in common with all model-based approaches, on the accuracy of features extracted. The only instances of failure we have encountered are in images where the feature extraction was obviously deficient.

Reference

[1] Canny, J. F., "Finding Edges and Lines in Images", Ph.D. Dissertation, AI-Laboratory, MIT, Cambridge, MA, 1983

[2] Dickinson, S., Pentland, A. P., Rosenfeld, A., "Qualitative 3-D Shape Reconstruction Using Distributed Aspect Graph Matching", Procs. of 3rd ICCV, pp. 257-262, Osaka, Dec., 1990

[3] Fan, T., Medioni, G., Nevatia, R., "Recognizing 3-D Objects Using Surface Descriptions", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11, No. 11, pp. 1140-1157, Nov., 1989

[4] Flynn, P. J., Jain, A. K., "CAD-Based Computer Vision: From CAD Models to Relational Graphs", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-13, No. 2, pp. 114-132, Feb., 1991

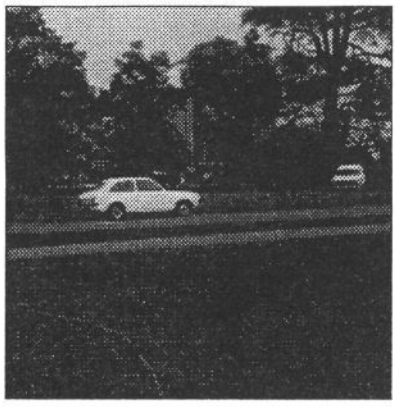
[5] Goad, C., "Special Purpose Automatic Programming for 3D Model-Based Vision", Proceedings Image Understanding Workshop, Virginia, USA, pp.94-104,1983

[6] Lowe, D. G., *Perceptual Organization and Visual Recognition*, Hingham, MA: Academic, 1985

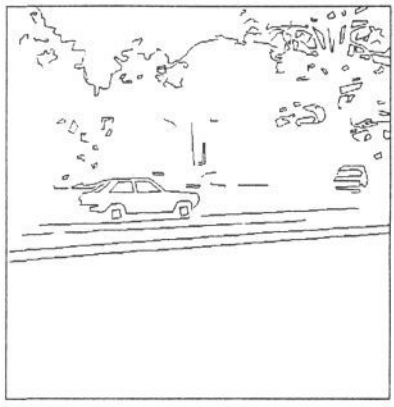
[7] Marefat, M., Kashyap, R. L., "Geometric Reasoning for Recognition of Three-Dimensional Object Features", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12, No. 10, pp. 949-965, Oct., 1990

[8] Worrall, A, Sullivan, G. D. and Baker, K. D., "Model-Based Tracking", Proc. BMVC91, Glasgow, Sept. 1991

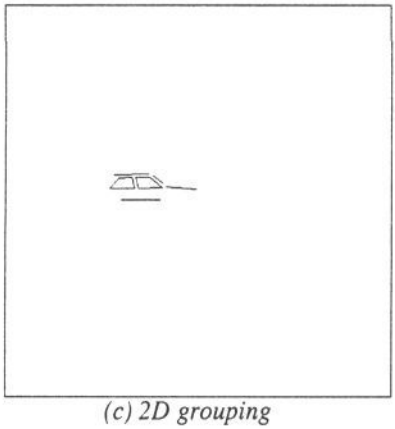
[9] Zhang, S., Du, L., Sullivan, G. D. and Baker, K. D., "Model-Based 3D Grouping by Using 2D Cues", Proc. BMVC90, Oxford, Sept., 1990



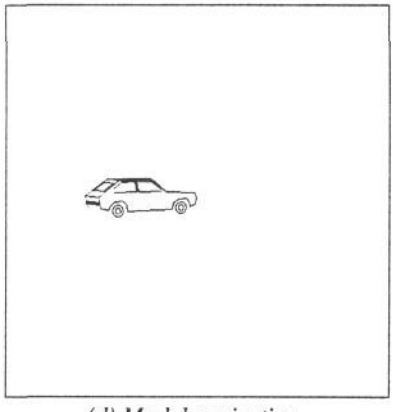
(a) Original image



(b) 2D features

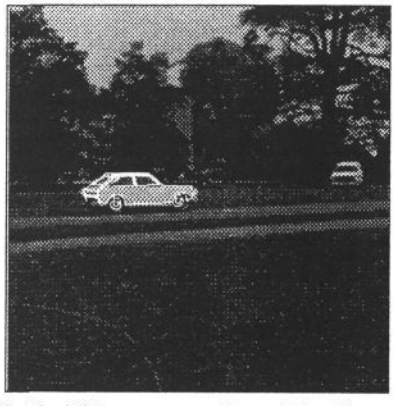


(c) 2D grouping



(d) Model projection

(e) 3D grouping



(f) Model instance on the original image.

Fig. 4. Various stages of the recognition process

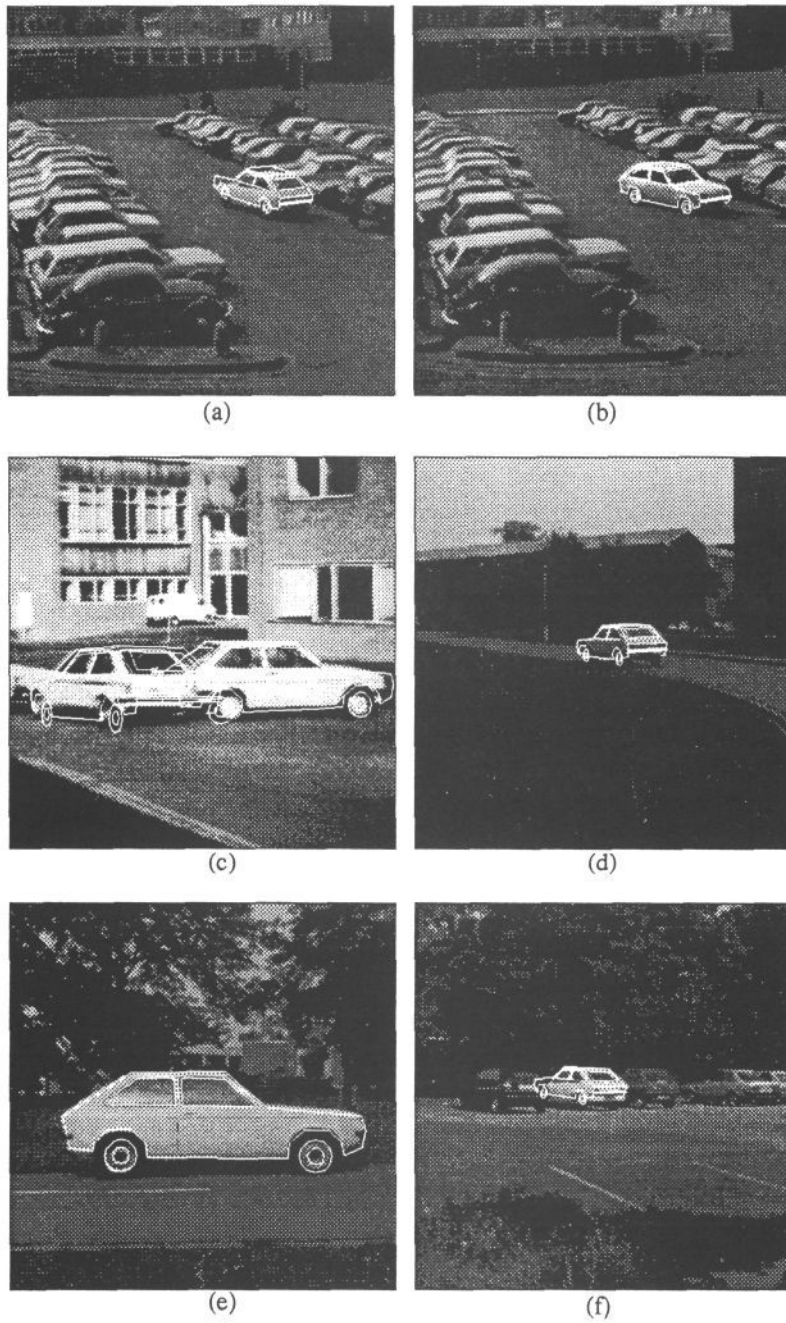


Fig. 5. Model instances superimposed on a representative set of original images