On Evidence Assessment for Model-Based Recognition

L Du[1], G D Sullivan and K D Baker
Department of Computer Science, Reading University
Reading, RG6 2AY

Abstract

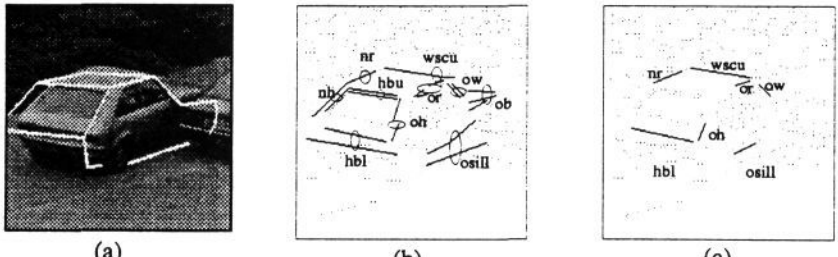
Evidence assessment for model based recognition is concerned with determining if a set of correspondences between image features and model features gives sufficient evidence for recognition. While many previous studies have addressed strategies to establish the correspondences, post-model-matching evidence assessment has remained largely primitive and *ad hoc*.

This paper presents a novel two-stage scheme of evidence assessment for model-based vision based on: (i) evidence against coincidental configuration of random image features and (ii) evidence against mis-recognition of other objects. We demonstrate this scheme for model based 3D recognition from 2D image features.

1 Introduction

Model-based vision usually comprises two phases: (i) *model matching* in which a search is carried out for evidence consistent with a pose, and (ii) *evidence assessment* to arrive at a recognition verdict. The second issue has received far less attention than model matching, and the majority of existing systems have employed *ad hoc* thresholds based on some form of goodness measure to give the final verdict.

An example drawn from car recognition [5] is illustrated in Figure 1. Figure1 (a) shows the original image and an hypothesised initial pose; (b) shows the candidate image features; (c) is the final 3D clique. The question addressed is: does this clique provide sufficient evidence to recognise the hatchback car?



(a) (b) (c)

Figure 1 Example of Object Recognition

1. Currently a SERC research fellow at the Department of Electrical and Electronic Engineering, University of Surrey, Guildford, GU2 5XH, (email: L.Du@ee.surrey.ac.uk)

BMVC 1992 doi:10.5244/C.6.16

Bolles et al [2] treated a detected feature as providing one of 3 types of evidence, (i) positive evidence (a feature found close to the predicted feature); (ii) neutral evidence (a feature between the sensor and the predicted feature); (iii) negative evidence (a feature further from the sensor than the predicted feature). Their decision algorithm was based on an *ad hoc* threshold for a simple sum of these pieces of evidence. Fan, et al [8] presented a measure of the goodness of their graph matching based on 3 factors, (i) the ratio of the number of matched model nodes to the total number of model nodes; (ii) the 3D surface area of matched model nodes comparing to that of all model nodes; (iii) the ratio of the area of matched scene nodes to that of all the scene nodes. This measure is quantitative but the choice of threshold still remains *ad hoc*. Lowe [13] used an arbitrary threshold on the minimum number of matches needed for recognition, which is typical of the majority of model based recognition systems [9] [14].

Brisdon's iconic verification [5, 6] for 2D-3D recognition and Grimson's symbolic verification for 2D-2D recognition [12] have given systematic treatments to the problem of evidence assessment. They both defined evidence assessment as a process intended to eliminate accidental configuration among random image features. Implementation of this definition for two domains has allowed informed selection of verification thresholds, on the basis of a relation between a probability of an accidental configuration (called conspiracy by Grimson) and the quality measure of a set of correspondences.

This paper proposes a novel two-stage evidence assessment for model based vision, which not only assesses accidental configurations caused by random image features but also takes object confusability into consideration.

2 A two stage approach to evidence assessment

Several definitions of evidence assessment have been put forward as a process to assess the significance of a set of matches (against accidental configuration caused by random image features), but this only captures one aspect of the problem. Mistaken matches are also very likely to happen to occur non-randomly, where other structures may be confusable with the target object. For example, a model-based recognition process for a car may well find sufficient evidence from an image of a house.

Therefore, we treat evidence assessment in model-based vision as a two-stage task. Firstly, we assess evidence against the possibility that the set of feature correspondences arises by accidental configuration of random data features; This is the task addressed by traditional evidence assessment. Secondly, we assess the evidence against the possibility of matching the model of an object by mistake to data features caused by other structures; this has not previously been reported. We use the SDT [11] approach to investigate the second issue, by focusing on discrimination between a target object and a highly confusable non-target object (as defined by the specific application). This puts a lower bound on the discrimination ability of a system, and thus gives a measure of total performance.

The above discussion applies to model-based vision using edge, region or surface features. In the remainder of this paper we concentrate on 2D-3D vision using edge features, forming cliques of 2D-3D feature correspondences (referred to as a 3D clique).

3 Significance of a set of feature correspondence

The significance of a 3D clique may be assessed by comparison with the null hypothesis that a clique is due to an accidental configuration among essentially random image features.

3.1 Significance according to cardinality and VCE

Given a 3D clique, we consider two attributes which reflect its significance: (i) the size of the set (the cardinality of the clique), (ii) a measure of viewpoint consistency error, *VCE* (defined as the disagreement between the clique and the 2D model template for the pose derived from clique [5]).

3.2 The VCE distribution of a single random clique

Probability distributions of VCE for random cliques of different cardinalities form the basis for the first stage of evidence assessment with 3D cliques. A probability distribution function is defined as:

$$p_{ca}(v) = \frac{d}{dv} P_{ca}(v) \qquad (1)$$

where the subscript *ca* denotes the particular cardinality and $P_{ca}(v)$ is the probability that a random clique of cardinality *ca* scores a VCE $\leq v$.

Grimson analytically established a similar probability for the 2D-2D recognition situation, as a function of the fraction of model feature, and sensor noise. However, in the current situation, these probability functions are extremely difficult to obtain analytically due to the inherent complexity of the 2D-3D problem. An experimental approach has been adopted, in which random 3D cliques are simulated by assigning random image features to features of a cube model. We assume that the VCE distribution obtained on the basis of those random 3D cliques is representative of random 3D cliques formed using general polyhedra.

Each experiment comprised trials of 10,000 random cliques of a fixed cardinality (*ca*). Each random 3D clique was created by assigning *ca* random image features to the same number of model features. A cube has at most 9 visible line features, so that it allows 7 experiments (from *ca* =3 to *ca*= 9), giving 7 histograms of VCE (Figure 2). These histograms are used as empirical approximations to the *VCE* distribution functions for significance testing.

3.3 Significance test for a 3D grouping problem

In order to carry out the significance test, we formulate the null hypothesis that the 3D clique is the result of matching an object to random image features. The test of significance becomes an attempt to contradict the null hypothesis.

Images containing several random features give rise to many cliques of different cardinality. Let k be the number of possible cliques of a given cardinality. The probability that at least *one* of the cliques happens to score a *VCE* as low as *v*, is

$$Q_{ca}(v) = 1 - [1 - P_{ca}(v)]^k \qquad (2)$$

We assume that a perfect model matching method has been used, which always produces the clique with the lowest VCE among the k possible cliques. Therefore, refutation of the null hypothesis depends on a sufficiently small Q.

The significance test can be arranged as the following steps:

- Establish look-up tables for $P_{ca}(v)$, off-line.
- Calculate k for different cardinalities according to the set of candidate features fed into the model-matching process (those features, which fall into the focus neighbourhood defined by the initial pose estimate before model-matching). Record the cardinality and the VCE.
- Calculate $P_{ca}(v)$.

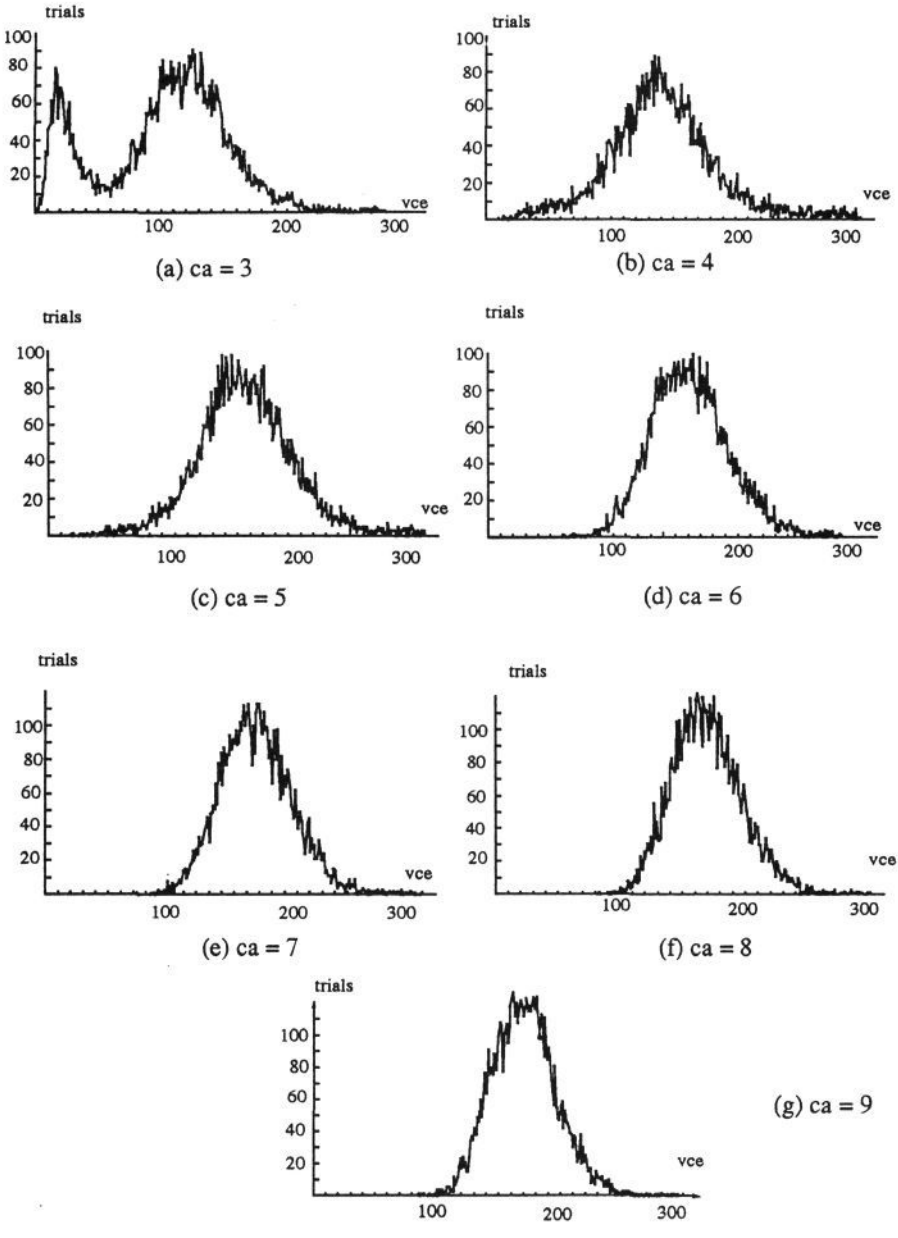


Figure 2 VCE distribution for random cliques

- Select a confidence threshold T_{cfd} to fit the particular application. If $T_{cfd} > Q$, reject the null hypothesis and accept a significant configuration. Otherwise reject the 3D clique.

The process can be illustrated for the example shown in Figure 1. Under the null hypothesis (that the candidate image features are random), the total number of possible cliques of cardinality=6 is k=18080 (obtained by combination and permutation of all

potential matches). The VCE of the clique shown in Fig. 1(c) is 4.6. Referring to Fig.2 (d), it is found that the probability for a single random clique of ca=6 to score 4.6 is $Q_6(4.6)=0$. We also find $Q = 0$, by using (2). Therefore, the clique can be taken as highly unlikely to have arisen by chance.

4 Object discrimination

A 3D clique which is significantly different from an accidental configuration of random image features, as described above, may yet be the result of incorrectly matching the target model to features due to a similar object. A second stage is needed to assess the possibility.

4.1 Signal Detection Theory (SDT)

SDT [11] considers classification as a signal detection problem, where a series of events E_i are presented consecutively to a decision process. Each event E_i causes a response in the receiver V_i. The event E_i may be of two types, (i) a signal with noise (E_s), (ii) pure noise (E_n). A rational decision process uses a threshold (Thd) to classify the event according to V_i. The rule is that If $V_i > Thd$ then, classify E_i as arising from pure noise, otherwise classify E_i as arising from signal plus noise.

The Receiver Operating Characteristic (ROC) [16] is defined as the function relating the rate of hits (correct detection of E_s) versus the rate of false alarms (false detection of E_n) over all possible choices of threshold. It provides an accepted method to assess the performance of a detector, irrespective of the selected threshold.

4.2 Object discrimination by using the viewpoint consistency measure

We use the SDT approach to investigate object discrimination using the VCE as a receiver function. Each application of the model-based vision process has one target object. The task in the second evidence assessment stage is to discriminated between the target object and other non-target objects. We try to estimate a lower bound on discrimination performance by considering the target object and a non-target object which is highly confusable with it. The choice of this non-target object depends on the intended application, and should be chosen to provide a lower bound on performance.

Once a non-target object has been chosen, the problem can be fit into the SDT framework. There are two types of cliques: (i) the cliques of features arising from the target object being matched to the target model (called G_c - analogous to E_s) (ii) the cliques with features arising from the other object being matched to the target model (called G_w - analogous to E_n). Monte-Carlo experiments were conducted to establish the two probability distributions of the VCE response to two types of cliques ($p_{G_c}(v)$ and $p_{G_w}(v)$, with random noise added to all features.

Following SDT practice, we use the Receiver Operating Characteristic (ROC) to measure discrimination power (D), which is defined here as twice the area between the positive diagonal line and the ROC curve. If D=0 then the two distributions completely overlap and no discrimination is possible, if D=1 then they are distinct and discrimination is perfect.

To study the performance of the VCE, we measured discrimination the distribution for a number of noise levels, using random perturbation to extracted image features, defined in [7].

4.3 Discrimination between a Hatch-back car and Saloon car as function of noise

To illustrate the approach we selected two highly confusable models from the library available at Reading University (Fig. 3). There are small differences in the dimensions of the lines, but the conspicuous difference is in the shape of the back of the vehicle (though this structure is not explicit in the test). The following analysis answer the question: how good is the VCE for discriminating this pair of confusable objects.

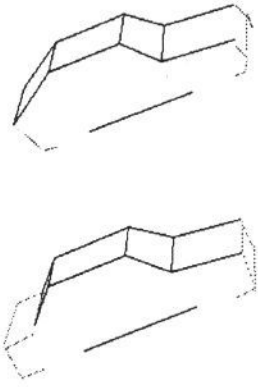
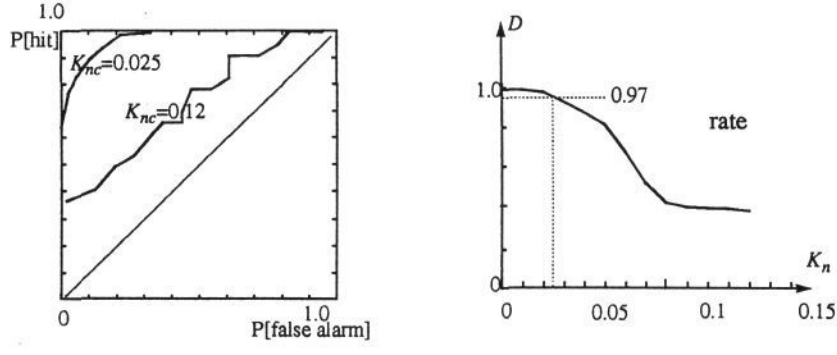


Figure 3

Simplified hatchback model and saloon model

(the broken lines indicates features removed)

Random G_w cliques were simulated by assigning instantiated features from the saloon model to the corresponding features of the hatchback model. In this case of non-identical objects, the wire-frame models for a hatchback and a saloon car were simplified (Figure 4) so that each feature in one model has exactly one corresponding feature of the same name in the other model. Therefore, the creation of G_w cliques meets the assumption of an ideal model-matching process.



(a) Two example of ROCs

(b) D vs. noise level

Figure 4 Discrimination of saloon and hatchback

Random G_c cliques were simulated by assigning instantiated model features from a hatchback model to features on the same model. Random noise was added to all template features by random perturbation.

The VCE responses to both types of cliques were collected from a large number of G_c and G_w cliques at a range of noise levels. At each level of noise, the histograms for the VCE response to each type of cliques was used to approximate the two distributions ($p_{G_c}(v)$ and $p_{G_w}(v)$. It is found that at all noise levels (0.01~0.12) two distributions overlapped significantly. Therefore, it is impossible to specify a threshold which perfectly discriminates two types of cliques, and SDT applies.

Figure 4 (a) illustrates the ROC curves at two noise levels concerning a hatchback and a saloon car. Figure 4 (b) plots the resulting D as a function of noise level against 12 noise levels (0.0~0.12)

4.4 Using D-curves

Experiments with the two types of cliques produce distributions, which lead to estimates of D for a particular pair of objects. The experiment with G_c cliques can also be plotted to produce an approximate relation between the VCE and the noise level (see Fig. 5).

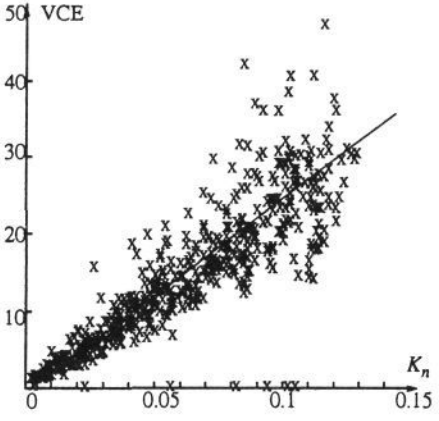


Figure 5 Corelation between noise and VCE value for correct cliques

An important point is that once the D-curve has been established for a target object (e.g. a hatchback car), D can be estimated for any future image provided we have an estimate of the image noise level. To obtain this estimate, we run the feature extraction process and manually pick out the correct features to form a number of correct cliques. We then run the VCE evaluator, and find their average VCE value. The noise level is acquired by finding an equivalent noise level on the VCE and noise relation.

For example, for the image shown in Fig. 1(a) the average VCE of manually-selected correct cliques is 9.5, which gives an equivalent noise level of 0.025 from Fig. 5. This further gives a D = 0.97 (see Fig. 4(b)).

5 Conclusion

The two-stage evidence assessment process provides an approach to answering the question raised at the beginning of this paper. The evidence assessment as a two-stage statistical process applies to model-based object recognition in general. It allows two important decisions: (1) a significance test to eliminate possible accidental configuration of random features and (2) discrimination of confusable objects by using a threshold on VCE of the set. When a clear choice of the discrimination threshold (100% hit and no

false alarm) is not possible, we can still determine the best possible compromise, as a function of the noise level.

The paper also described, in particular, a realisation of this process for 2D-3D model-based vision using edge-based features.

6 References

[1] Bodington, Sullivan & Baker, Experiments on the use of the ATMS to label features for object recognition, Computer Vision-ECCV'90, Spring-Verlag, 1990.

[2] Bolles, R, Horaud, P., 3DPO: A Three-Dimensional Part Orientation System, Int. J. of Robotics Research, Vol. 5. No. 3, 1986, pp3-26

[3] Bray, A., Recognising and Tracking Polyhedral Objects, Ph.D Thesis, Sussex University, UK, 1991.

[4] Brisdon, K, Evaluation and Verification of Model instances, Proceeding of Alvey Vision Conference'87, Cambridge, 1987, pp33-37

[5] Brisdon, K Hypothesis Verification using Iconic Matching, Ph.D. Thesis, Reading University.

[6] Du, L, G D Sullivan and K D Baker, 3D grouping by viewpoint consistency ascent, Image and Vision Computing, Special Issue on BMVC'91, 1992

[7] Du, L, G D Sullivan and K D Baker, Modelling data complexity for model-based vision, submitted to BMVC'92

[8] Fan, T, Medioni, G and Nevatia, R Recogising 3D Object Using Surface Descriptions, IEEE PAMI, Vol. 11, No. 11, 1989, pp1140-1157

[9] Fisher, R, From Surfaces to Objects: Computer Vision and 3 Dimensional Scene Analysis, John Wiley & Sons

[10] Goad, C., Special Purpose Automatic Programming for 3D model-based vision, Proceeding of the ARPA image understanding Workshop, Arlington, Virginia, 1983.

[11] Green, D and Swets, J, Signal Detection Theory and Psychophysics, Robert E Krieger Publishing Co. 1974 (Reprint of Wiley & Sons 1966)

[12] Grimson, L & Huttenlocher, D, On the Verification of Hypothesized Matches in Model-Based Recognition, Proceeding of the European Computer Vision Conference, France, 1990, pp489-498, Spring Verlag

[13] Jain, A and Hoffman, R, Evidence based Recognition of 3-D Objects IEEE PAMI, Vol. 16, No. 6, 1988, pp783-800

[14] Ikeuchi, K and Takeo, K, Automatic Generation of Object recognition Programs, IEEE Proceeding, Aug. 1988

[15] Lowe, D., The viewpoint consistency constraint, International Journal of Computer Vision, 1987

[16] Schiffmen, H, Sensation and Perception: An Integrated Approach, John Wiley and Sons, Inc, 1976

[17] Worrall, A., et al, Model Based Tracking, BMVC'91, Glasgow, 1991