

Modelling Data Complexity for Model-based Vision

L. Du¹, G D Sullivan and K D Baker
Department of Computer Science, Reading University
Reading, RG6 2AY

Abstract

This paper discusses an issue of wide-ranging importance for computer vision - the systematic consideration of data complexity in assessment of computer vision systems.

We investigate 3D object recognition from 2D features as a typical problem of computer vision. We identify 3 factors which contribute to the complexity of image data: feature truncation, noise and clutter. We propose a modelling scheme for these factors, which allows us both to measure and to simulate each factor.

Using the scheme, a systematic comparison is made between two existing strategies for model-matching as a function of clutter and truncation factors. Post-model-matching object discrimination is then examined as a function of the noise factor. These two examples serve to illustrate the data complexity model, and demonstrate its use for formal assessment of model-based algorithms.

1 Introduction

This work was motivated by the desire to compare two methods for matching 3D models to 2D image features, under conditions of data complexity likely to be met in practice. Models of the factors which confound model-based vision were needed. Similar problems arise in many circumstances in model-matching and the analysis may be used generally.

The model-based approach to 3D recognition from 2D image features has inspired a number of systems and studies, including those by Roberts [13], Brooks [3], Goad [8], Ikeuchi [10], Lowe [11], Sullivan [15], Bodington et al [1], Bray [2], Zhang et al [17], Du et al [4]. These systems have adopted many different strategies for interpreting image features by matching model features to them. It is relatively easy to compare different strategies in terms of their computational costs, but they prove very difficult to compare with respect to their robustness in the presence of errors of the data. A second problem in model-based vision is that of assessing the totality of the evidence for a recognised object. Robust criteria for evidence assessment underpin the decision processes used to detect or reject an object, as well as those used to discriminate between similar objects.

The lack of any accepted basis for testing algorithms is an important obstacle towards any systematic treatment of these issues. In practice, image features suffer from

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)

3 main errors: (i) truncation, possibly due to occlusion, (ii) clutter, due to features from irrelevant objects and (iii) noise due to digitisation and measurement errors. Such errors act as confounding factors which contribute to the image complexity. We propose a modelling scheme providing measures for the factors and allowing controlled simulation of them. Different model-matching algorithms can thereby be compared objectively with respect to data complexity on a formal statistical basis.

We illustrate the use of the truncation model and clutter model by briefly comparing Lowe's incremental model-matching strategy and the VCA strategy proposed by Du et al [4]. We also illustrate the use of the noise model by an example of object discrimination in model-based vision.

2 Ideal and observed image features

In the domain of 3D recognition from 2D features, ideal image features may be defined as the result of noise-free feature extraction from images containing nothing but the target object. Assuming that the pose of a target object in the scene is known, the ideal image features are in exact agreement with the instantiation of the object model.

Observed image features are the product of an imperfect feature extraction process applied to a complex image. Any difference between the observed and ideal features can be thought of as a form of data complexity. We decompose the data complexity into 3 factors, (i) truncation of ideal features, (ii) noise contamination of ideal features, (iii) clutter caused by extraneous features.

These observations are relevant both to edge and region features, but in the remainder of this paper we consider only the more commonly used edge features. A fourth type of error, due to false extension of an image feature because of accidental alignment with other structures is relatively rare in edge-based methods (although its equivalence for regions, false merging, is more troublesome). False extension of edge features is not considered here.

3 A model of image feature truncation

Truncation of image features is common in feature extraction. It occurs when part of the object is occluded so that only a part of an ideal feature is observed, where the edge detector loses edge points near a junction (as is frequently the case with the Canny detector), when an edge has uneven contrast, or when an occluding edge is seen against a chequered background.

3.1 Truncation and its measure

Let an ideal feature be represented as $F_I = \overline{(x_1, y_1) (x_2, y_2)}$ and its corresponding observed feature by $F_O = \overline{(x'_1, y'_1) (x'_2, y'_2)}$. Truncation can be represented as perturbation of the end points along the line of the ideal feature, with only one degree of freedom (see Figure 1).

Let L_I and L_O be the lengths of F_I and F_O respectively. A measure of truncation for each fragmented ideal feature is defined as,

$$k_t = \frac{L_I - L_O}{L_I} \quad (1)$$

We define a measure of overall truncation of a set of simulated or observed features as the

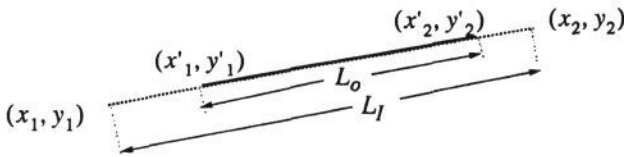


Figure 1 Feature truncation

RMS of all truncations,

$$K_t = \sqrt{\frac{\sum (k_i^t)^2}{n}} \quad (i=1, 2, \dots, n) \quad (2)$$

where n is the total number of features.

3.2 Simulation of truncation

To simulate truncated ideal features, the coordinates of each end point of an ideal feature are randomly shrunk along the feature, with a bound on the maximum of truncation. Figure 2 gives three examples of simulated image feature truncations. The examples are displayed in increasing order for the truncation measure K_t .

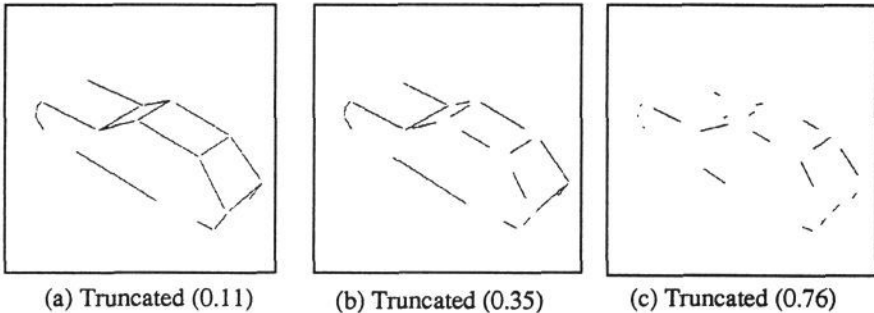


Figure 2 Examples of simulated truncation

4 A model of noise contamination for image features

Noise contamination arises from the perturbation of observed features caused by imperfect feature extraction and from the use of object models of limited accuracy.

4.1 Noise contamination and its measure

Noise contamination of an image feature is described as a random perturbation of the end points of observed feature F_o , with two degrees of freedom (see Figure 3). For a pair of ideal (dotted) and observed (continuous) features we use k_n as the measure of noise, where

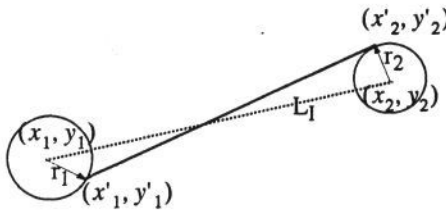


Figure 3 Noise contamination to image features

$$k_n = \frac{\sqrt{r_1^2 + r_2^2}}{L_1} \quad (3)$$

The overall measure of noise contamination is again defined as the RMS

$$K_n = \sqrt{\frac{\sum_i^n (k_{nc}^i)^2}{n}} \quad (4)$$

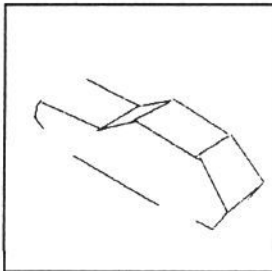
where n is the total number of features.

4.2 Simulation of noise contaminated features

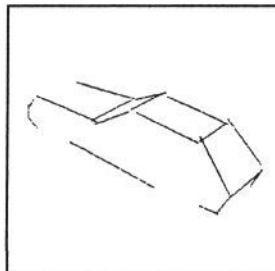
Feature noise contamination can be simulated as described in equation (5), where k is a bound on the perturbation. Examples are shown in Figure 4 for a range of the perturbation measure K_n .

$$x'_1 = x_1 + \text{random}(\text{sign}) \times \text{random}(k) \times L_1$$

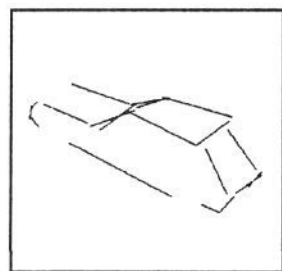
(x'_2, y'_1, y'_2 , likewise) (5)



(a) Perturbation (0.02)



(b) Perturbation (0.04)



(c) Perturbation (0.08)

Figure 4 Examples of simulated noise contamination

5 A model of clutter

The third confounding influence is that of extraneous features which arise from other objects in the scene, or from parts of the target object not described by the model.

5.1 Clutter and its measure

Extraneous features of arbitrary orientation, scattered over the image (as used for example by Bray [2]) may be distracting to human perception, but cause little distraction to most machine matching processes. Object recognition usually relies on a hypothesised initial pose, and this enables immediate exclusion of any image features falling outside a tight neighbourhood of the expected feature positions. On the other hand, extraneous features which are located very close to a ideal feature make little difference to either recognition or final pose determination. The main difficulty arises from extraneous features which are close enough to be included in the focus neighbourhood but far enough away to cause substantial error if mis-matched. Extraneous features caused by shadows, patterned backgrounds or adjacent objects are common examples.

As a measure of clutter, we therefore need to take into account only those extraneous features which have the potential to distract model matching and lead to error of pose determination. The following measure of clutter is designed to reflect the potential of an extraneous feature to mislead the model matching. It is a non-monotonic with the geometrical displacement of an extraneous image feature from a corresponding ideal feature.

Consider an ideal feature F_i ($i = 1, \dots, n$) and clutter features O_{ij} ($j=1, \dots, n_{ei}$) all of which are associated with F_i by being in its focus neighbourhood. We define the measure of clutter to be (b is explained further below)

$$k_c^{ij} = \begin{cases} (g - bg^5) & \text{if } g - bg^5 > 0 \\ 0 & \text{if } g - bg^5 \leq 0 \end{cases} \quad (6)$$

where g is defined as

$$g = \sqrt{(1.15d)^2 + \theta^2} \quad (7)$$

and combines the two geometrical displacements shown in Figure 5 (the factor 1.15 was introduced to ensure equivalent end points shift for numerically equivalent translation and rotation, as for a feature of typical length 50 pixels).

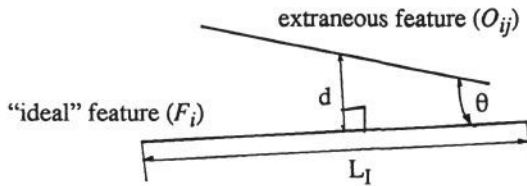


Figure 5 Measure of interference

The measure (6) is essentially an aggregate measure of the displacement subject to clipping. It falls sharply beyond the maximum, reflecting the fact that extraneous features outside the focus neighbourhood have little distracting potential. Parameter b allows its spatial extent to be tuned to suit the size of the focus neighbourhood used in the matching algorithm.

It is often the case that there are multiple extraneous features, all associated with a single ideal feature, in this case we take the maximum of their individual measures, as the measure of their collective potential for distraction

$$k_c^i = \text{MAX} \{k_c^{ij}\} \quad (\text{for } j=1, \dots, n_{ei}) \quad (8)$$

where n_{ei} is the number of extraneous features, each associated with the i^{th} feature.

A measure of the total distraction caused by all extraneous features is defined as the RMS, as before

$$K_c = \sqrt{\frac{\sum_i^n k_c^i}{n}} \quad (9)$$

where n is the total number of visible ideal features.

5.2 The simulation of extraneous features

Observed features are simulated by applying random truncation and noise contamination to instantiated model features (see Figure 6 (a)). Clutter features are obtained by simulating a truncated, noise-perturbed feature (as in section 4), and then further displacing this randomly (in a manner similar to Figure 5) with d and θ drawn from a flat bounded distribution. Note that in simulation experiments the total set of observed features comprise the perturbed ideal features plus the clutter features. The number of associated clutter features n_{ei} is randomly chosen between 1 and 5.

Figure 6 (b), (c) gives 2 examples of a set of truncated features with simulated clutter. The two images show examples of simulated clutter generated using different bounds on d and θ . In the two cases the measure of the clutter (equation 9) was computed using $b=0.047$ so that k_c maximises at $g=30.5$ pixels, (a typical size of the focus neighbourhood).

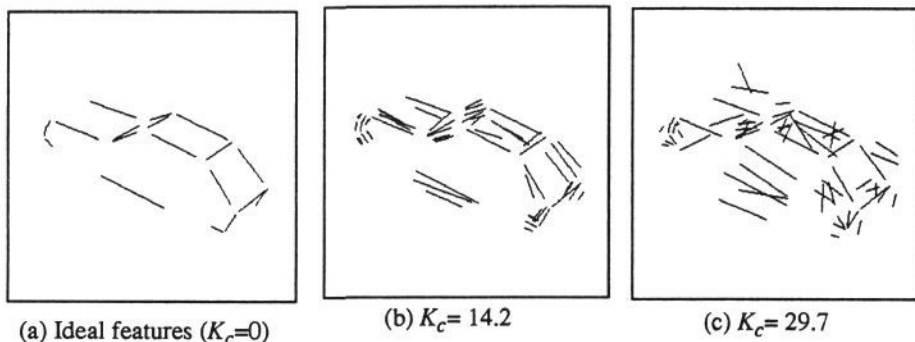


Figure 6 Examples of simulated clutter

6 Using the model of data complexity

This section describes two investigations of model-based vision, one for model-matching algorithms and the other for object discrimination, using the 3 measures of confounding factors.

6.1 Robustness of 2D-3D model matching strategies

We have recently proposed a model matching strategy which relies on viewpoint consistency ascent (VCA) [4]. This makes explicit and precise use of the viewpoint consistency constraint. In previous work it was shown to be superior to Lowe's incremental model-matching strategy [11] (which only weakly exploits viewpoint

consistency), for a few example cases. This superiority is explored here formally, using the data complexity model.

Using the truncation model and the clutter model, we conducted two Monte-Carlo experiments, each of 500 trials, to compare the sensitivity of both strategies to data complexity. Both experiments used a cube model to simulate the target object. A measure for the inaccuracy of resultant matches, D_f , is defined as the RMS of the geometrical differences between the matched features and the ideal features, as defined in equation (7).

In the first experiment, each trial used a fixed pattern of clutter and generated randomly perturbed ideal features. The measure of truncation K_t and D_f were then collected after each trial. Figure 7(a) shows the result of 500 trials and Fig. 7(b) shows histograms of inaccuracy.

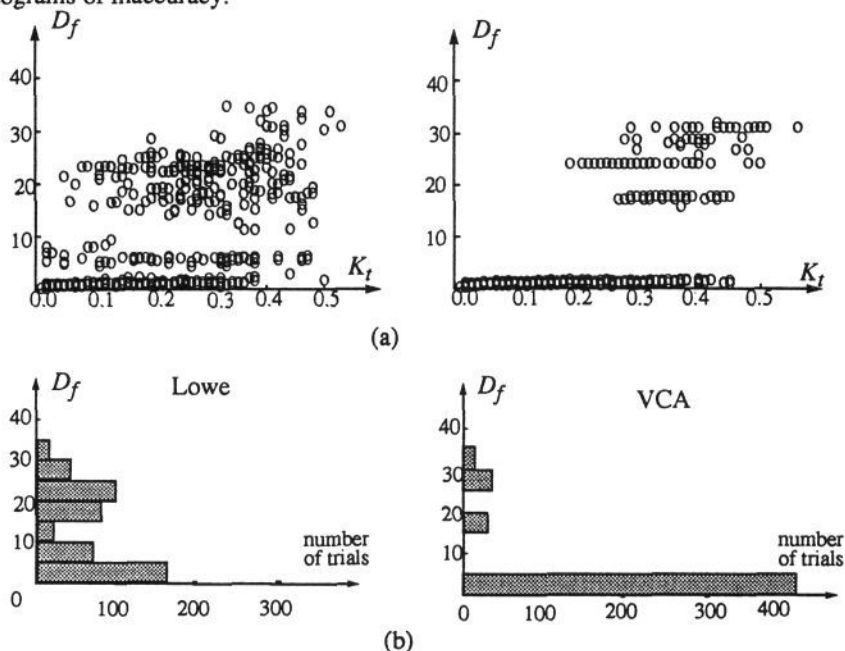


Figure 7 Model Matching Error as a function of the truncation measure

The second experiment fixed the pattern of perturbed ideal features, but allowed the measure of clutter K_t to vary and again recorded the output inaccuracy D_f . The result of two sets of 500 trials are plotted in Figure 8(a), and (b) shows the histograms.

This study reveals significant instability in Lowe's method even at very low levels of data complexity. The VCA shows a significant improvement of robustness in model-matching at all levels of data complexity examined. The study demonstrates systematically and quantitatively the benefit of a rigid enforcement of the viewpoint consistency constraint.

6.2 Object Discrimination

In model-based vision, after a set of feature matches is first established, it must be tested to decide whether the matches provide sufficient evidence for recognition or not. The

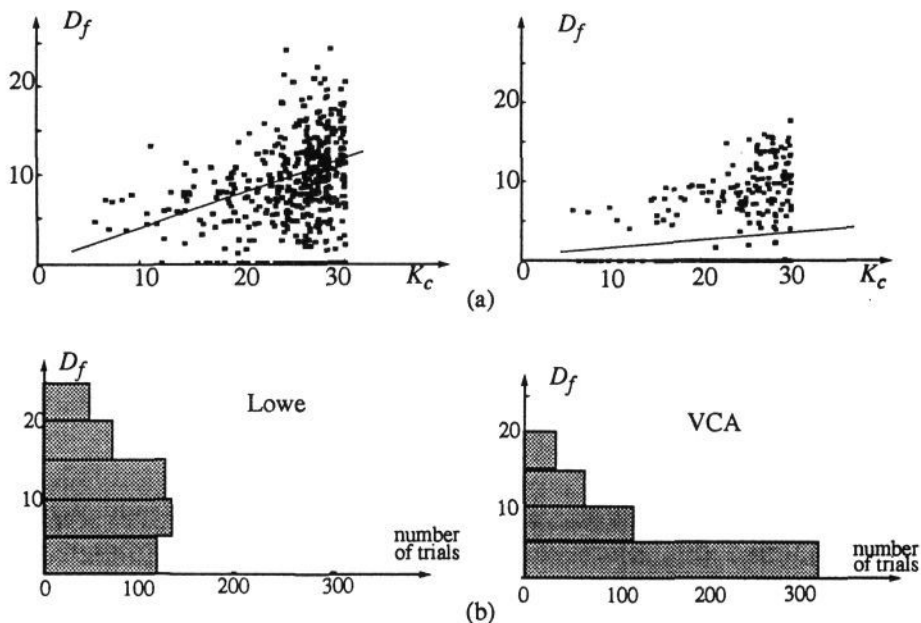


Figure 8 Matching error as a function of the clutter measure

problem is particularly acute if there exist other objects which must be discriminated from the target object.

We have investigated the object discrimination task by using the viewpoint consistency measure (VCE as defined in [4]) and adopting the using signal detection paradigm. Here we use this example to show the usefulness of the data complexity model. A companion paper in this volume contains more detailed discussion [5].

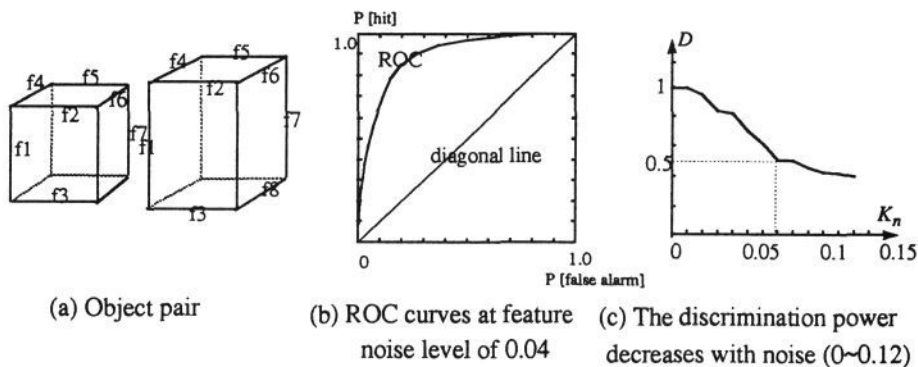


Figure 9 Discrimination of similar objects

It turns out that feature noise is the main confounding factor for this task. The theory of Signal Detection [9] and the noise model provides a general method to relate the discrimination power of the VCE with the noise level for any given pair of similar objects.

Figure 9 demonstrates the discrimination of an example pair of objects, a cube and a cuboid, (a) shows the models, (b) shows a typical ROC curve at the noise level of 0.04 (c) shows how the discrimination power decreases with noise level. [NB: D is defined as twice the area between the ROC curve and the diagonal line]. The result shows that the viewpoint-consistency-based method is able to discriminate ($D=0.5$) a cube and a cuboid up to the noise level of 0.06.

7 Conclusion

A modelling scheme has been described for the data errors encountered by algorithms for 3D object recognition based on edge-based features. Use of this model has enabled us systematically to demonstrate that a method that enforces viewpoint consistency rigidly is superior to systems such as Lowe's [11], which enforces it weakly. The data complexity model also made it possible to consider the impact of data noise on object discrimination with VCC-based object recognition algorithms. The studies illustrate the benefits of adopting better methods for testing computer vision, by modelling the confounding factors explicitly. We put this scheme forward as a standard method for model-based vision using edge based image features.

8 References

- [1] Bodington, Sullivan & Baker, Experiments on the use of the ATMS to label features for object recognition, Computer Vision-ECCV'90, Springer-Verlag, 1990.
- [2] Bray, A., Recognising and Tracking Polyhedral Objects, Ph.D Thesis, Sussex University, UK, 1991.
- [3] Brooks, A. B., Model Based Computer Vision, UMI Research Press, 1984.
- [4] 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
- [5] Du, L, G D Sullivan and K D Baker, On evidence assessment for model-based vision, Proc. BMVC'92
- [6] Grimson, et al, The combinatorics of object recognition in cluttered environments using constrained search, ICCV'88
- [7] Grimson, et al, Model-based Recognition and Localization from Sparse Range or Tactile Data, Int. J. of Robotics Research.
- [8] Goad, C., Special Purpose Automatic Programming for 3D model-based vision, Proceeding of the ARPA image understanding Workshop, Arlington, Virginia, 1983.
- [9] Green, D and Swets, J., Signal Detection Theory and Psychophysics, Robert E Krieger Publishing Co. 1974 (Reprint of Wiley & Sons 1966)
- [10] Ikeuchi, K and Takeo, K Automatic Generation of Object recognition Programs, IEEE Proceeding, Aug. 1988
- [11] Lowe, D., The viewpoint consistency constraint, International Journal of Computer Vision, 1987
- [12] Lowe, D., Fitting Parameterised 3D Models to Images, IEEE PAMI, No 5, 1991
- [13] Roberts, L Machine Perception of Three-Dimensional Solids, Chapter 9, Optical and Electro-Optical Information Processing, MIT Press, 1965

- [14] Rydz, A., et al, Model based Vision Using a Planar Representation of the Viewsphere, Alvey Vision Conference'88, Manchester, 1988.
- [15] Sullivan, G D, Alvey MMI-007 Vehicle Exemplar: Performance and Limitations, Proc. Alvey Vision Conference'97, Cambridge, 1987, pp39-45
- [16] Worrall, A., et al, Model Based Tracking, BMVC'91, Glasgow, 1991
- [17] Zhang, S, Sullivan, G D, Baker, K D, Relational Model Construction and 3D Object Recognition from Single 2D Monochromatic Image, Proceeding of the British Machine Vision Conference'91, Glasgow, 1991, pp233-239