Response to Kanatani

Paul L. Rosin and Geoff A.W. West

Abstract— **We discuss the advantages and disadvantages of two approaches to model selection: the information theoretic method suggested by Kanatani [3] and others, and our heuristic sequential selection method [9].**

keywords: curve segmentation, model selection, line, ellipse

I. The Problem

Kanatani has highlighted the potential problem of fitting multiple representations to the data. Since lower-order feature (LOF) models form a subset of higher-order feature (HOF) models then the data can always be fit by a HOF with equal or less error than a LOF. In our case of fitting straight lines and elliptical arcs this is not strictly true since the conic that includes lines is the hyperbola rather than the ellipse and only ellipse fits are allowed. However, in practice, large enough elliptical arcs are indistinguishable from straight lines.

II. Information Theoretic Solutions

Due to its simplicity of application the AIC [1] provides an attractive method for model selection. But while it has received much attention is has also been criticised on several accounts (see e.g. [6], [10]). The AIC is not consistent, and has a tendency to overfit the data. This can be overcome by modifying (increasing) the penalty term related to the number of model parameters. This is the approach taken by Kanatani with the geometric AIC [3]. However, a further problem is that the penalty term needs to be data dependent in a way that is not taken into account by these versions of the AIC either [10]. For example, tightly clustered data may not constrain the model parameters in the same way as well distributed data. An additional question is how effective the AIC would be for deciding whether to represent a curve section by one instance of a model or two instances of the same model.

A related approach for model selection is to use the minimum description length (MDL) [7]. When segmenting curves Lindeberg and Li [5] represented sections by straight lines or conics according to which had the shortest description. However, in practise this is not always straightforward. In a previous report [4] Li describes how additional criteria had to be used, involving (i) the ratio of lengths of the straight line and conic, (ii) the percentage of outliers, and (iii) the area between the line and curve. These tests required five threshold values which leads to the problem of the choice of thresholds.

P.L. Rosin is with the Department of Computer Science and Information Systems, Uxbridge, Middlesex, UB8 3PH, UK, e-mail: Paul.Rosin@brunel.ac.uk G.A.W. West is with the Department of Computer Science, Curtin University of Technology, GPO Box U1987, Perth 6001, Western Australia; e-mail: geoff@cs.curtin.edu.au

Yet another recent criterion for model selection is the Bhattacharyya metric [10] which appears very promising. However, a limitation with the information theoretic solutions is that they require the noise to be explicitly modelled, which is usually done by a zero mean Gaussian distribution.

III. Our Heuristic Solution

The heuristic used in our algorithm to circumvent the model selection problem is based on the reverse relationship between LOFs and HOFs: *a sequence of LOFs provides a good piecewise approximation to a single HOF*, as shown in figure 1. Like most other model selection approaches we consider simpler (i.e. lower order) features preferable since they are more concise and probably more robust. This constraint is incorporated by fitting feature models in sequence, starting with the lowest order, and progressing to the highest order.[1] When approximating the curve, straight lines are fitted first. Thereafter, curve sections fitted by a single straight lines are not considered for approximation by elliptical arcs. The rationale is that even though an arc may provide a lower error of fit to the data the straight line fit is deemed sufficient, and therefore preferable due to its simplicity. Ellipses are only considered for curve sections that cannot not be adequately approximated by a single straight line and have therefore been approximated by a sequence of lines. Then, if an ellipse fits the data sufficiently well (according to the significance measure) it will replace the sequence of lines.

Not only does the sequential feature model testing approach avoid overfitting, it is also efficient. Unlike the information theoretic methods it does not require all the models to be fitted to each curve section. Most of the fitting only involves the LOFs while the more computationally expensive HOF fitting is performed relatively infrequently.

We also note that our algorithm does not require that the fitted features be constrained to pass through the endpoints of the curve sections. For example, in [9] line and circular arcs fits were constrained while parabolic, elliptical and superelliptical arcs were unconstrained, in [11] lines were unconstrained, and in [8] elliptical arcs were constrained. The only caution is that if the features are not constrained then some care has to be taken in the subsequent selection of breakpoints [11].

As Kanatani points out, for our algorithm features should be fitted to minimise L_∞ since the significance measure is a function of the maximum deviation. In practice, for convenience we have actually minimised other error

[1] This sequential approach has been used by others. For instance, Besl and Jain [2] implemented region growing by fitting first, second, and fourth order bivariate polynomials in turn until one was found to produce a below threshold error.

functions such as mean square error and median square error. We have not investigated the effect of this discrepancy on the segmentation.

IV. An Experimental Comparison

In addition to our original line and elliptical arc segmentation program described in [9] we have written two new programs for the purpose of comparing the different model selection techniques. Both algorithms recursively segment the curve and fit both a straight line and elliptical arc to each section of data. As the segmentation tree is built up the better of the two features is retained for each curve segment. This is determined either by choosing the feature with the better (lower) significance value, or by using Kanatani's method. With Kanatani's geometric AIC the L_∞ error norms were used, and an ellipse is selected if

$$\frac{D_{\text{ellipse}}}{D_{\text{line}}} < 1 + \frac{6}{N-4}.$$

Elliptical arcs are fitted in a least squares sense, and are constrained to pass through the end points by minimising

$$\sum_{i=1}^{N} Q(x_i, y_i) - \lambda_1 Q(x_s, y_s) - \lambda_2 Q(x_f, y_f)$$

where $Q(x_i, y_i) = Ax_i^2 + Bx_i y_i + Cy_i^2 + Dx_i + Ey_i + 1$, and the end points of the curve section are (x_s, y_s) and (x_f, y_f). The maximum deviation and significance are calculated from the curve and the selected feature as before. Unlike in our previous work, the segmentation tree can now contain a mixture of feature types throughout the tree (i.e. lines are not restricted to the leaves of the tree). The tree is collapsed into a single scale curve representation by traversing and testing significance values as described in [9].

The first example shows the synthetic curve given as figure 10a in [9]. The results using the single stage simultaneous line and arcs fitting method where models were selected by their significance values or Kanatani's geometric AIC are shown in figure 2a and figure 2b respectively.

Another example on more realistic data is given in figure 3. Figure 3a shows the original set of curves, and the results of approximating these using the two stage sequential process is shown in figure 3b. The results of the simultaneous methods are shown in figure 3c where models were selected by their significance values and figure 3d where models were selected with Kanatani's geometric AIC.

In both examples (and in others not shown) the sequential method tends to fit less small elliptical arcs than the simultaneous methods. This is as expected since small sections will be fitted by lines if possible. Although the simultaneous method with significance fared better than the simultaneous method with AIC on the synthetic example both simultaneous methods produce results of a similar quality on the real data. This can be better seen by showing the lines and arcs that differed between the two methods in figure 3e and figure 3f.

Footnotes

P.L. Rosin is with the Department of Computer Science and Information Systems, Uxbridge, Middlesex, UB8 3PH, UK, e-mail: Paul.Rosin@brunel.ac.uk G.A.W. West is with the Department of Computer Science, Curtin University of Technology, GPO Box U1987, Perth 6001, Western Australia; e-mail: geoff@cs.curtin.edu.au

This sequential approach has been used by others. For instance, Besl and Jain [2] implemented region growing by fitting first, second, and fourth order bivariate polynomials in turn until one was found to produce a below threshold error.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19(6):716–723, 1974.

[2] P.J. Besl and R.C. Jain. Segmentation through variable-order surface fitting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(2):167–192, 1988.

[3] K. Kanatani. Comments on "Non-parametric segmentation of curves into various representations". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, ??:??–??, 199?

[4] M.X. Li. Minumum description based 2-D shape description. Technical Report CVAP114, Computational Vision and Active Perception Laboratory, Royal Institute of Technology, Stockholm, Sweden, 1992.

[5] T. Lindeberg and M.X. Li. Segmentation and classification of edges using minimum description length approximation and complementary junction cues. In *Scandanavian Conference on Image Processing*, pages 767–776, 1995.

[6] Special Section on AIC. *Psychometrika*, 52(3), 1987.

[7] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, 1989.

[8] P.L. Rosin. A hierarchy of representations for curves. In Arcelli et al., editor, *Aspects of Visual Form Processing*, pages 465–474, 1994.

[9] P.L. Rosin and G.A.W. West. Non-parametric segmentation of curves into various representations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17:1140–1153, 1995.

[10] N. Thacker, D.Prendergast, and P.Rockett. B-fitting: A statistical estimation technique with automatic parameter selection. In *Proc. British Machine Vision Association*, 1996.

[11] G.A.W. West and P.L. Rosin. Techniques for segmenting image curves into meaningful descriptions. *Pattern Recognition*, 24:643–652, 1991.

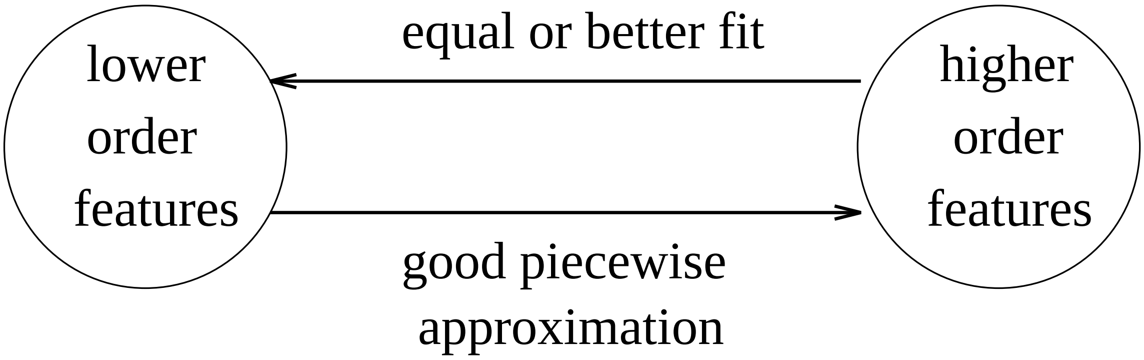


Fig. 1. Relationship between different orders of features

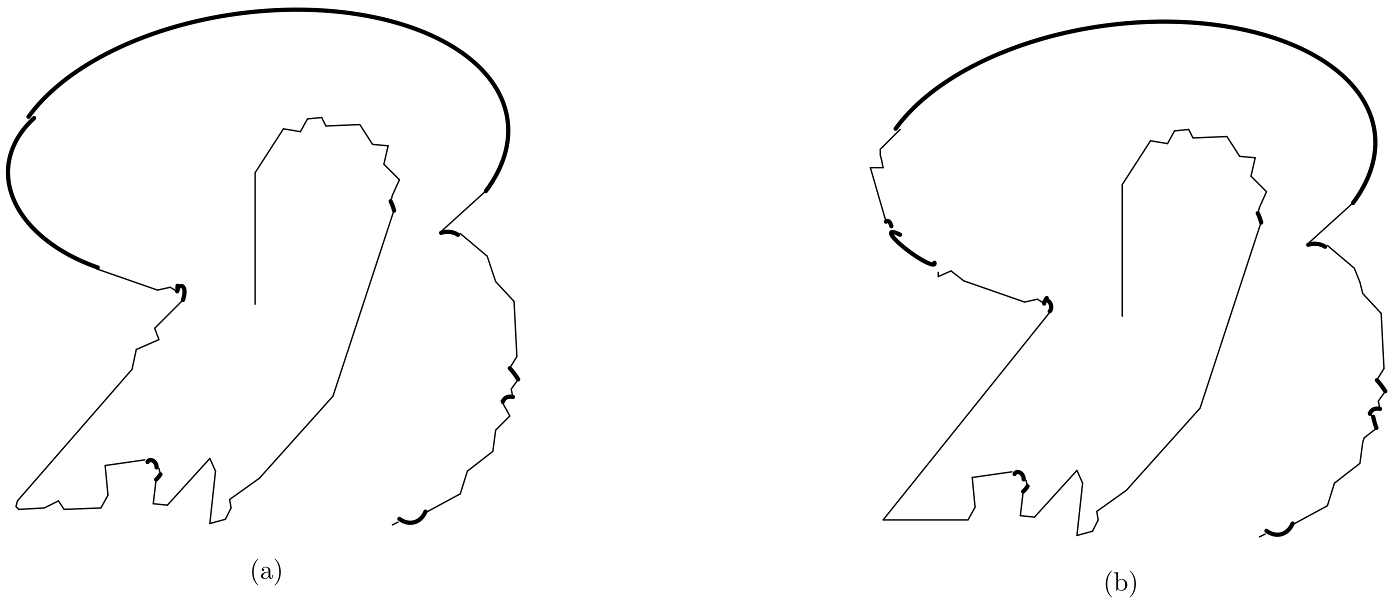
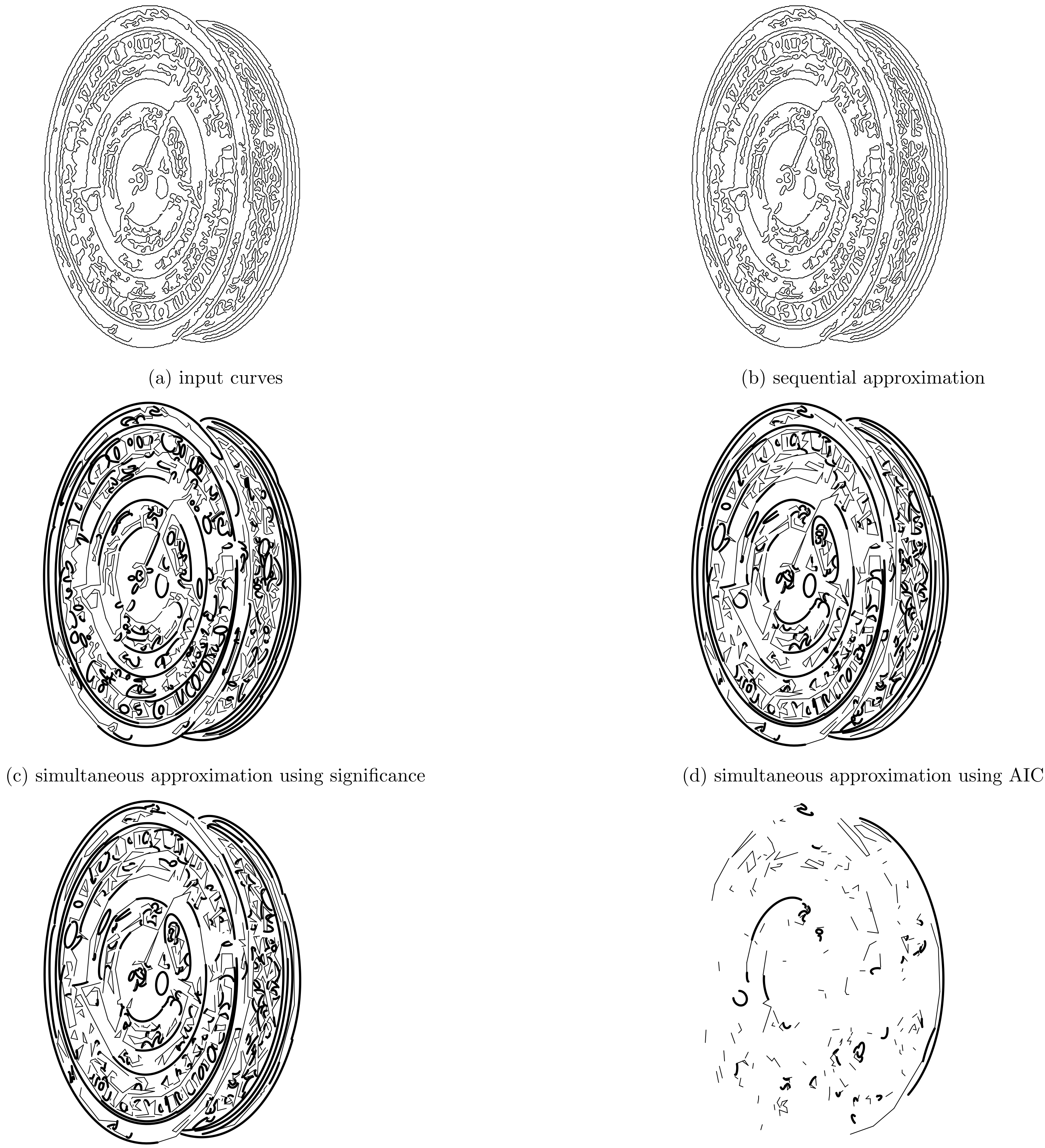


Fig. 2. Comparison between significance and AIC model selection. (a) selection by significance, (b) selection by AIC



(a) input curves

(b) sequential approximation

(c) simultaneous approximation using significance

(d) simultaneous approximation using AIC

(e) features in (c) changed in (d)

(f) features in (d) changed in (c)

Fig. 3. Comparison between different model selection techniques. (a) input curves, (b) sequential approximation, (c) simultaneous approximation using significance, (d) simultaneous approximation using AIC, (e) features in (c) changed in (d), (f) features in (d) changed in (c)