Automated B-Spline Curve Representation Incorporating MDL and Error-Minimizing Control Point Insertion Strategies

December 5, 1997

Abstract

The main issues of developing an automatic and reliable scheme for spline-fitting are discussed and addressed in this paper, which are not fully covered in previous papers or algorithms. The proposed method incorporates B-spline active contours, the minimum description length (MDL) principle, and a novel control point insertion strategy based on maximizing the Potential for Energy-Reduction Maximisation (PERM). A comparison of test results show that it outperforms one of the best existing methods.

1 Introduction

Representing curves by analytic functions instead of sets of data points allows the geometry of curves to be exploited in various ways [1], and may also be used for data smoothing [2] or for storing data efficiently. However, it is difficult to find a *fully automated spline-fitting method* which consistently performs as well as methods based on human assistance.

A number of schemes have been proposed for fitting analytic functions to image curves. Duda and Hart [3] suggested a polygonal representation scheme based on the subdivision of joined line segments, while other methods [4, 1, 5, 6, 7, 8, 9] have been suggested for fitting B-splines. These methods were found to be unsatisfactory for curved-based applications such as those discussed in [10, 11] primarily because as a result of poor control-point distribution an unduly large number of control points must be used in the fitted B-spline to reduce estimation error – this causes large instability in the computation of contour derivatives. In some cases the control points must be initialised manually. Furthermore, most of these methods terminate the spline fitting based on a predefined error threshold.

The main contributions in our paper are: (i) highlighting and discussing the main issues to spline-fitting, describing where previous methods fail – past papers have not done this; (ii) introducing a novel control-point insertion strategy which attempts to maximize error-reduction through the iterative insertion of control-points – this addresses the issue of control-point distribution more accurately compared to previous heuristic schemes; (iii) combining different techniques addressing *all* the issues into a highly functional and *automated* spline-fitting algorithm, which have not been achieved with previous algorithms.

2 Main Issues in Fitting B-Splines

The important issues to consider when fitting image contours with B-splines are:

Parameterisation When handling image curves, the B-spline parameterisation along the fitted curve does not need to unique and the B-spline model may be simplified by having *control points uniformly spaced at unit intervals along the spline parameter* (it does *not* follow that the control points are uniformly spaced in the image domain). A measure typically used for the quality of a fit is the sum-of-squared-errors (the 'error energy') given by $E = \sum \|\boldsymbol{x}_i - \boldsymbol{S}(t_i)\|^2$ where $\boldsymbol{x}_i$ are the data samples with associated spline parameter values t_i, and $\boldsymbol{S}(t)$ is the spline with parameter t[1]. Since E may be minimised through reparameterisation of $\boldsymbol{S}$ such that the $(\boldsymbol{x}_i - \boldsymbol{S}(t_i))$'s are perpendicular to the tangents $\boldsymbol{S}'(t_i)$'s of the spline, *the problem may be stated as that of assigning the optimal spline parameter value t_i to each of the sampled points x_i*. Bartels *et al.* [4] proposed the use of normalised arc-length distances between data points for the parameterisation, while Guéziec and Ayache [5] suggested using arc-length distances along a polygonal representative curve. However in both cases, the assigned parameter values are only heuristic estimates. If iterative fitting strategies [5, 6] are further used, the parameter values may be further refined numerically but these are computationally expensive and may produce erroneous results if the initial estimates are weak. There are also spline-fitting algorithms which do not take this into account (eg. [1]), and are likely to employ the common solution of assigning consecutive data points to a uniformly increasing sequence of parameter values which unsurprisingly gives poor results.

Number of Control Points The number of control points in a spline model determines the number of degrees of freedom available in the fitting process, and an optimal choice for this is needed to maintain both smoothness and closeness of fit. A number of statistical tests have been proposed including tests for the constancy of an unbiased error variance estimate [7] and Powell's test for trends [12]. However in most of the other spline-fitting methods a empirical error threshold is used, which

[1] For clarity, we call the points $\boldsymbol{S}(t_i)$ on the spline the *sampling points*

does not take into account the amount of noise present nor the scale of the curves.

Distribution of Control Points Regions which are poorly modelled by polynomials of the given spline order require a greater concentration of control points to reflect the greater information density. This is analogous to the *Shannon-Nyquist sampling theorem* [13] applied to local regions, and also with the exception that instead of frequencies we consider the amount of information per unit length. *Hence it is incorrect to assume that the optimal distributions of control points are similar between splines of different order, or that control points cluster around regions of higher curvature.* Although the initial positions of the control points may be numerically improved, it is likely that only a local minimum for E will be found (Jupp's 'lethargy' theorem [14]), and hence good initial estimates for the control points have to be obtained. However in [5], initial control points estimates are based on Duda and Hart's polygonal representation. Figure 1(c,d) show how the polygonal representation may be used to produce a control-point distribution which will give rise to control points clustering around higher curvature regions. This is clearly weaker than what is possible as shown in figure 1(b). Others [6] assume that some initial distribution of control points is available or may be obtained heuristically. Some control point insertion strategies, elaborated in section 3.2, have also been proposed but which are not satisfactory in our experience.

Data Sampling Regions of curves which require a higher proportion of control points to model should have the same higher proportion of sampled points in the vicinity in order to achieve the same rate of oversampling, by further drawing on the analogy with the *Shannon-Nyquist sampling theorem.* Many algorithms however treat

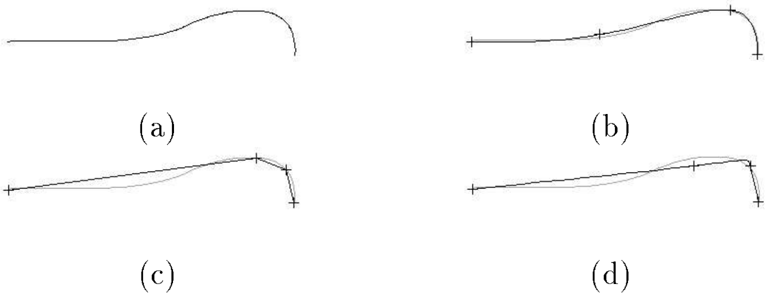

Figure 1: Control point distributions using different strategies. (a) the original curve to be fitted. (b) a spline with two control points which is fitted manually and deemed optimal to the eye; the '+'s mark the boundaries between piecewise polynomials. (c) shows the best polygonal representation which is used to produce the distribution of control points in (d). It is clear that clustering control points around high curvature regions is not necessarily optimal.

all available data points equally or weigh them according to their accuracies [5, 7]. Doing this unfortunately results in simple regions of the curve (eg. long straight sections) being modelled with too much precision, *at the cost* of poorly representing the complex (high information) regions of the curve. Lu and Milios [6] attempted to overcome this problem using curvature-dependent weighing of data points, which fails to take into account the source of the problem.

3 Theoretical Approach to B-Spline Fitting

The top-level paradigm for fitting splines proposed here is based on the continual *evolution* of a spline such that the spline deforms to approximate the curve to be represented. A sequence of images depicting the evolution of the spline in our final fitting algorithm is shown in figure 2.

The features of our paradigm are:

1. **B-Spline Active Contours**. The optimal positions of a fixed set of control points are recovered by treating the B-spline as an active contour.

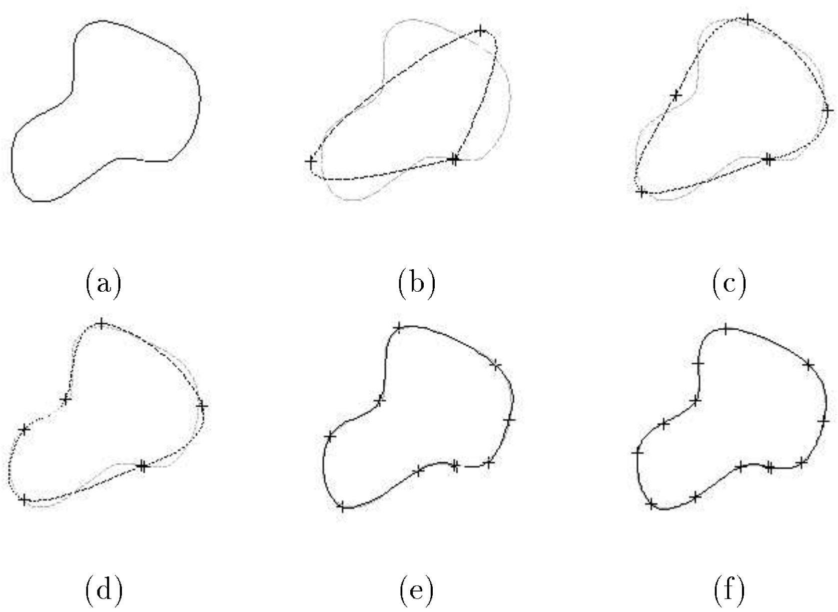

Figure 2: (a) shows the shape to be fitted, which at the moment is treated as an open contour with the break near the lower right region. (b)-(f) shows the spline with increasing number of control points.

2. **Optimal Control Point Insertion**. A control point insertion strategy is derived which maximises energy-reduction potential.

3. **Minimum Description Length (MDL) Framework**. The minimum description length theory [15] caters for a parameter-free estimation of the optimal number of control points.

These are discussed in the next few sections.

3.1 B-Spline Active Contours

The crucial difference between point-based and B-spline active contours is that instead of including an internal energy term, regularisation is *intrinsic* through the fewer degrees of freedom available in deformation. The characteristics of B-spline active contours allow two fitting issues to be explicitly addressed:

Data Sampling Instead of using all available data points, *the data is resampled according to some preferred distribution of sampling points on the active contour.* This distribution of sampling points may be changed for each iteration. In the proposed model, the number of sampling points per piecewise polynomial is fixed, and is given by $R = \frac{N}{M-1} > 1$ where R is the *oversampling ratio*, N is the number of sampling points and M is the number of control points. Hence if during the minimisation process two control points are drawn closer, the sampling also becomes proportionally denser in that region. Since the control points of the B-spline are uniformly distributed along the spline parameter, the ratio may be achieved by sampling at fixed intervals of the spline parameter.

Parameterisation Given a distribution of sampling points, data samples are obtained by searching *perpendicularly* to the spline tangent at each sampling point. These data samples therefore are associated with the spline parameter values at the corresponding sampling points.

Since the original curve is originally represented as densely sampled but discrete data points, it is convenient to model the curve as an initial polygonal curve obtained by joining all data points with line segments. The data point associated with a sampling point may then be found by solving for the intersection of the tangent perpendicular and the polygonal curve.

3.2 Control Point Insertion

Iterative fitting schemes which involve gradually increasing the number of control points during the fitting process [7, 5, 6] usually require a control point insertion strategy. In this paper, the insertion of control points is considered in its dual form of inserting *hinges*, which are the boundaries between polynomial pieces.

The shortcomings of two current strategies are discussed here:

1. **Interval Midpoint Strategy**. Dierckx [7] proposed that a control point be introduced in the middle of the piecewise-polynomial which has large displacements from the data points. However if these displacements are skewed to one side of the piecewise-polynomial, the optimisation routine may be trapped in a weak local minimum. See figure 3.

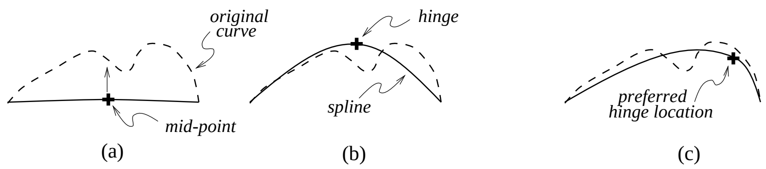

Figure 3: *Interval Midpoint Strategy*. In (a), a hinge is introduced in the midpoint of a piecewise-polynomial interval, and by using the active contour optimisation the configuration in (b) is reached. A preferred solution is shown in (c).

2. **Largest Displacement Strategy**. Lu and Milios [6] suggested that the control point be introduced such that a hinge is formed at the point on the spline which has the maximum displacement from a data point. However it is possible that this maximum displacement may correspond to a position on or close to a hinge. Introducing a hinge at the position gives very unstable results. See figure 4.

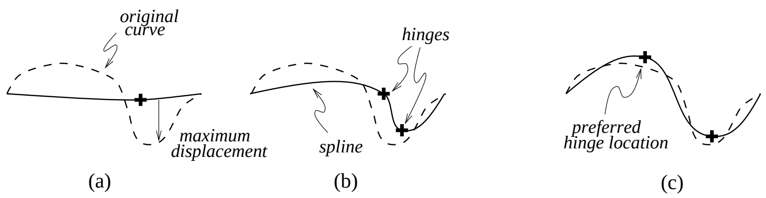

Figure 4: *Largest Displacement Strategy*.In (a), the initial hinge has little effect since a *collapse mechanism* has not been set up. If another hinge is introduced at the maximum displacement, the configuration in (b) is arrived at after using the active contour optimisation. A preferred solution involving a good collapse mechanism is shown in (c).

Observation of the failure of the largest displacement strategy reveals that control points must be introduced such that *compatible collapse mechanisms* [16] (related to the plastic deformation of structural beams) must be formed. More details may be found in [17].

3.2.1 The Potential for Energy-Reduction Maximisation (PERM) Strategy

Ideally, a control point should be inserted at a location such that the maximum error reduction occurs. This would correspond to maximizing the information represented by the new control point, and consequently regions with higher information density would contain more control points. In this strategy we attempt to estimate the position on the spline to introduce a hinge such that the *potential for reducing the error energy E is maximised.*

Consider a polynomial piece of the spline. By mapping the polynomial piece onto a horizontal line segment, the residual displacements between the sampling points and the data points are also mapped such that they become perpendicular to the line segment. The new vertical displacements d_i may be written as

$$d_i = \left| \; \boldsymbol{x}_i - \boldsymbol{S}(t_i) \quad \frac{\boldsymbol{S}'(t_i)}{||\boldsymbol{S}'(t_i)||} \; \right| \tag{1}$$

See figure 5. The hinges on either side of the polynomial piece are also noted as either fixed (for end-points) or free.

By allowing the line segment to be completely broken at some point, we proceed to calculate the maximum error energy reduction possible on both the smaller line segments which can be achieved by moving the break point and free hinges vertically, as in figure 5. For a given break point, the error-energies are quadratic. If the error energy to the left of the break point at t_i is E_l, then

$$E_{li}(\alpha, \beta) = \sum_{j=m}^{i} (\alpha + j\frac{\beta - \alpha}{i - m} - d_j)^2 \tag{2}$$

where α and β are the (vertical) displacements of the left hinge and break point respectively. The maximum energy reduction possible may be easily computed by

$$\Delta_m E_{li} = \Delta E_{li}(\hat{\alpha}, \hat{\beta}) = \frac{1}{2} (\boldsymbol{\nabla} E_{li})^T \boldsymbol{H_l}^{-1} \boldsymbol{\nabla} E_{li} \mid_{\alpha=\beta=0} \tag{3}$$

where $(\boldsymbol{\nabla} E)$ is the gradient of E and $\boldsymbol{H_l}$ is the Hessian which is constant with respect to α and β.

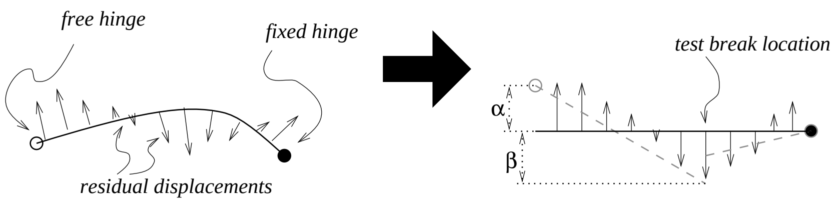

Figure 5: The PERM strategy. See text for details.

The *Energy-Reduction Potential* for a particular point is defined as the sum of the left and right maximum energy reductions if a break is introduced at that point

$$\mathcal{P}_i = \Delta_m E_{li} + \Delta_m E_{ri} \tag{4}$$

Intuitively, this *potential* does not necessarily refer to the immediate potential for energy minimisation. Instead *it measures the energy-reduction potential for a hinge at the point to be part of a compatible collapse mechanism likely to be formed after an additional number of control point (hinge) insertions.*

The PERM strategy is therefore to insert a hinge at the point on the spline corresponding the maximum energy-reduction potential $\mathcal{P}$. Results based on the

PERM strategy have shown that the control points are introduced in near-optimal regions, and compatible collapse mechanisms are formed.

3.3 Minimum Description Length (MDL)

The minimum description length (MDL) criterion by Rissanen [15] provides a generic method for comparing the optimality of different models fitted to a particular set of data, where models may have different dimensionality and structure.

Model-fitting may be expressed in a Bayesian framework as *maximising* the aposteriori probability $P(\boldsymbol{\theta}|\boldsymbol{X})$ where $\boldsymbol{X}$ is the matrix of data points given by $\boldsymbol{X} = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \dots\ \boldsymbol{x}_N]$, and $\boldsymbol{\theta}$ is a variable-length vector containing all model parameters. Then with the application of Bayes' rule and the assumption that $P(\boldsymbol{X}|\boldsymbol{\theta})$ is a normal distribution, we obtain an expression to be minimised:

$$L(\boldsymbol{X},\boldsymbol{\theta}) = \frac{\log e}{2}(\boldsymbol{y} - \bar{\boldsymbol{y}}_{\boldsymbol{\theta}})^T \boldsymbol{K}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{y} - \bar{\boldsymbol{y}}_{\boldsymbol{\theta}}) + M \log N \tag{5}$$

where $L(\boldsymbol{X},\boldsymbol{\theta}) = -\log P(\boldsymbol{X}|\boldsymbol{\theta}) - \log P(\boldsymbol{\theta})$ is the ideal code length to describe both $\boldsymbol{X}$ and $\boldsymbol{\theta}$, $\boldsymbol{y} = [\boldsymbol{x}_1^T\ \boldsymbol{x}_2^T\ \dots\ \boldsymbol{x}_N^T]$ is a vector containing the elements of $\boldsymbol{X}$, $\boldsymbol{K}$ is the covariance matrix of $\boldsymbol{y}$, and M is the number of control points.

In circumstances when the covariance matrix $\boldsymbol{K}$ is unknown, it is possible to estimate this from the residual errors. Using the usually satisfactory simplification $\boldsymbol{K} = \sigma^2 I$ in our analysis, an unbiased estimate for the uniform variance σ^2 is $\hat{\sigma^2} = \frac{\boldsymbol{y}^T\boldsymbol{y}}{N-M-1}$. As new control points are gradually introduced, each new estimate L_{j+1} is calculated based on $\hat{\sigma^2}_j$ for the current estimate. If $L_{j+1} < L_j$, previous estimates L_j may then be revised by *backpropagating* the new $\hat{\sigma^2}_{j+1}$. This scheme is similar to that used in [18]. The minimum L is usually considered found once L remains higher than the lowest estimate over a number of iterations.

4 Implementation and Experimental Results

The specific problem considered here is that of fitting B-splines to chains of edgels obtained from any edge-detection and linking algorithm. For the purpose of comparing the performance of our PERM-based algorithm to the latest spline-fitting methods, we implemented Guéziec and Ayache's algorithm [5] which in our opinion addresses the main spline-fitting issues most thoroughly among the other published methods. The algorithms were tested on the contours shown in figures 6(a) and 7(a,e). The results are compared in terms of errors for the same number of control points, as shown in figures 6(f) and 7(d,h). For the first curve, the fit obtained by Guéziec and Ayache's algorithm when errors are similar to our fit for 9 control points is also shown in figure 6(d).

The results indicate that the PERM control-point insertion strategy is superior compared to using Duda and Hart's polygonal representation scheme, since our algorithm performs better for any number of control points. However as would be expected, this advantage reduces when linear splines are used. Additionally, Guéziec and Ayache's algorithm requires less computation time. The evolution of our active contours take considerably longer since at each time-step the data needs to be resampled, and convergence is slow in some situations for a gradient-descent based optimisation.

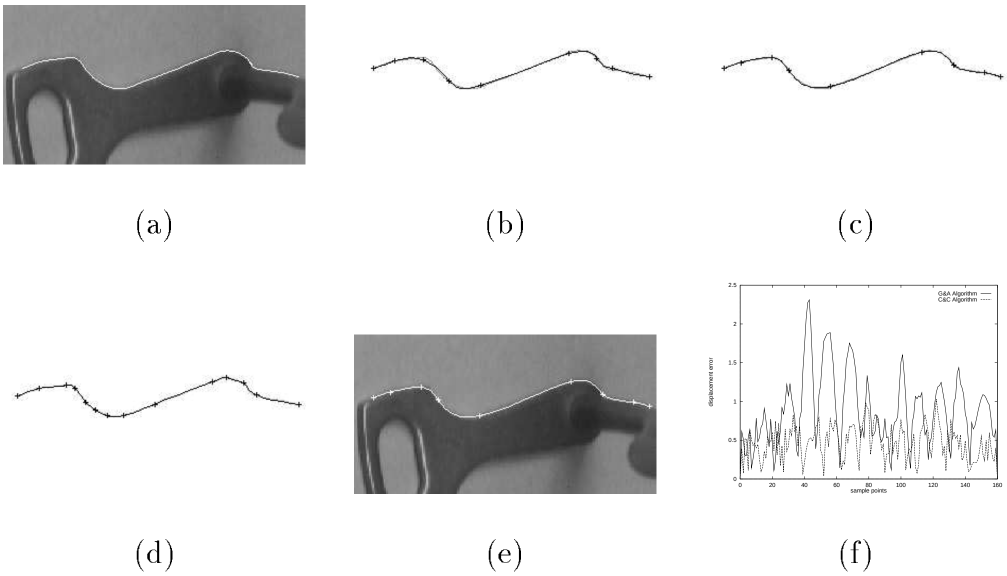

Figure 6: (a) The contour on which the two compared algorithms are tested. (b) The spline with 9 control points using Guéziec and Ayache's (G&A) algorithm, and (c) The spline with 9 control points using our (C&C) algorithm. (d) The spline from the G&A algorithm for which the average squared errors is similar to the errors for the spline in (d). (e) The spline from our algorithm superimposed on the original image. (f) is a graph comparing squared errors when using different numbers of control points with the two algorithms.

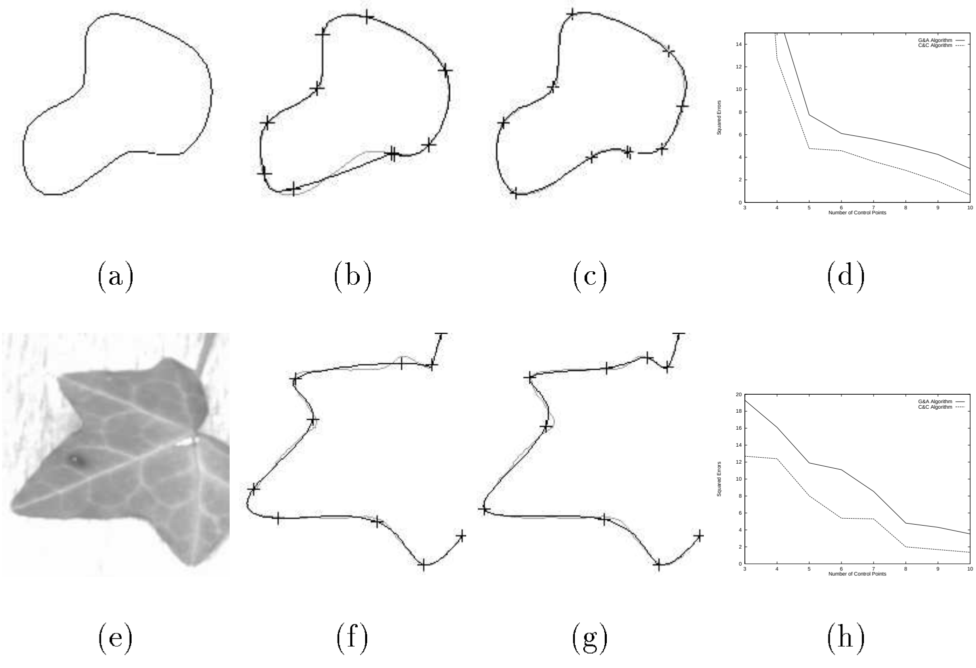

Figure 7: (a,e) show the initial curves to be fitted. In (b,f), the splines obtained by using the G&A algorithm up to 10 control points. (c,g) show the same curves fitted using our algorithm at 10 control points. A graph providing comparison of squared errors to number of control points for both algorithms is shown in (d,h).

5 Conclusions

We have shown that by explicitly describing the main issues in B-spline fitting, a model designed to resolve these issues may be developed. B-spline active contours and the minimum description length principle have been used in various ways but have rarely been applied to spline-fitting, while the PERM control-point insertion strategy is inspired by structural mechanics analysis. By combining these features, we show that it is possible to produce a *fully automatic* spline-fitting algorithm which is capable of outperforming existing methods used in computer vision. The caveat however lies in the amount of time active contours require to converge on the optimal curve-representation solutions. If fitting is to be carried out on polygonal image curves, or if a polygonal approximation suffices, a faster regression based Duda and Hart's method is often accurate enough. However when a good estimation of contour derivatives is required, the proposed algorithm is considerablt more reliable. Nevetheless, this may be improved by using more complex methods (eg. conjugate-gradient methods) in the optimisation process, and will form part of our plans for future work. As is, the algorithm has been applied in [10, 11]. It is also currently being ported to the IUE environment by the Oxford Robotics Research Group and will be publicly available in due course.

References

[1] P. Saint-Marc, H. Rom, and G. Medioni, "B-spline contour representation and symmetry detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1191–1197, Nov 1993.

[2] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by method of generalized cross validation," *Numerische Mathematik*, vol. 31, pp. 377–403, 1979.

[3] R. Duda and P. Hart, *Pattern classification and scene analysis*. John Wiley, 1973.

[4] R. Bartels, J. Beatty, and B. Barsky, *An Introduction to Splines for use in Computer Graphics and Geometric Modeling.* Morgan Kaufmann, 1987.

[5] A. Guéziec and N. Ayache, "Smoothing and matching of 3-D space curves," *Int. Journal of Computer Vision*, vol. 12, no. 1, pp. 79–104, 1994.

[6] F. Lu and E. Milios, "Optimal spline fitting to planar shape," *Signal Processing*, vol. 37, pp. 129–140, 1994.

[7] P. Dierckx, *Curve and Surface Fitting with Splines.* Clarendon Press, 1993.

[8] T. Cataldi and T. Rotunno, "Kalman filter optimized cubic spline functions for digital smoothing," *Analytical Methods and Instrumentation*, vol. 2, no. 1, pp. 27–34, 1995.

[9] I. Qamar, "Method to determine optimum number of knots for cubic splines," *Comm. Numerical Methods in Engineering*, vol. 9, pp. 483–488, 1993.

[10] T. Cham and R. Cipolla, "Geometric saliency of curve correspondences and grouping of symmetric contours," in *Proc. 4th Euro. Conf. on Computer Vision, Cambridge (England)* (B. Buxton and R. Cipolla, eds.), vol. 1064 of *Lecture Notes in Computer Science*, pp. 385–398, Springer-Verlag, 1996.

[11] J. Sato and R. Cipolla, "Quasi-invariant parameterisation and matching of curves in images." To appear in IJCV, 1997.

[12] M. Powell, "Curve fitting by splines in one variable," in *Numerical Approximation to Functions and Data*, pp. 65–83, Athlone Press, 1970.

[13] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley, 1991.

[14] D. Jupp, "The 'lethargy' theorem – a property of approximation by γ-polynomials," *Journal of Approximation Theory*, vol. 14, pp. 204–217, 1975.

[15] J. Rissanen, "Minimum-description-length principle," *Encyclopedia of Statistical Sciences*, vol. 5, pp. 523–527, 1987.

[16] S. Crandall, N. Dahl, and T. Lardner, *An Introduction to the Mechanics of Solids.* Engineering Mechanics Series, McGraw-Hill, 2nd ed., 1978.

[17] T. Cham, *Geometric Representation and Grouping of Image Curves.* PhD thesis, Department of Engineering, University of Cambridge, Aug 1996.

[18] Y. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. Journal of Computer Vision*, vol. 3, no. 1, pp. 73–102, 1989.