Fast Polynomial Segmentation of Digitized Curves

Peter Veelaert and Kristof Teelen

University College Ghent, Member of Association University Ghent, Belgium
{peter.veelaert, kristof.teelen}@hogent.be

Abstract. We propose a linear-time algorithm for curve segmentation which is based on constructive polynomial fitting. This work extends previous work on constructive fitting by taking the topological properties of a digitized curve into account. The algorithm uses uniform (or L_∞) fitting and it works for segments of arbitrary thickness. We illustrate the algorithm with the segmentation of contours into straight and parabolic segments.

1 Introduction

In this work we consider the segmentation of digitized curves into linear or parabolic segments. Curve segmentation is used for curve coding and representation, length and tangent estimation [1, 2], and shape matching. In discrete geometry one popular approach to curve segmentation is to segment a curve into digital straight segments (DSS). A rather complicated linear time-algorithm for DSS segmentation was found by Smeulders and Dorst [3]. Debled-Renesson and Réveillès gave a simple, easy to implement algorithm [4]. More recently, Buzer developed an algorithm for the recognition of straight lines of arbitrary thickness, based on convex hulls. Buzer's incremental algorithm can also be used for curve segmentation [5].

Since curve segmentation has many diverse applications in image processing it is worthwhile to look at some other approaches and their points of attention. Leclerc and Zucker [6] discuss how to segment a one-dimensional image through curve fitting, and they discuss how to avoid the misclassification of points near a discontinuity. Likewise, Dunham gives an algorithm that finds a piecewise linear approximation of a curve within uniform error and with a minimal number of vertices [7]. Some methods borrow techniques from Robust Analysis. Rosin and West propose a method that recursively subdivides a curve to form a binary tree, which is then traversed to select the best representation of the curve [8]. They use a least median of squares method to fit ellipses to the data, which is more robust than ordinary least squares. Boyer *et al* use a robust sequential estimator to parametrize and organize range images [9]. Also variable order fitting has shown its value in curve and image segmentation. An example is Besl and Jain's image segmentation algorithm which is based on variable-order surface fitting [10]. As for curve segmentation an important application of variable order fitting is feature point detection for shape matching. When a large circular arc is segmented into digital straight segments, the endpoints are unsuited as feature points, since their location is more or less arbitrary. The examples in this paper show that quadratic segmentation is more meaningful, and indicate that tangent and length estimation could also profit from variable order fitting. Also the thickness of the segments should be adjustable so that

the feature points necessary for matching are all present. One of the reasons variable order and variable thickness segmentation is not generally applied is its computational burden [11, 10, 12]. To summarize, the desirable properties for a segmentation algorithm are:

- low or even linear-time complexity;
- arbitrary segment thickness;
- more robust than least squares fitting;
- suited for variable-order fitting;
- take into account and/or benefit from the topological properties of a digitized curve;
- simple, and with predictable control and data flow for implementations on dedicated hardware.

In this paper we propose a simple, linear-time algorithm for segmentation with polynomials of arbitrary order, which we illustrate for linear and parabolic segments. Linear-time complexity, simplicity and generality do not come for free, however. The method is based on the estimation of the fitting error, not the exact computation, and therefore also the segmentation will not always be exact. The amount of errors can be reduced, however, by increasing the number of computations, and for most applications the errors are completely acceptable. Section 2 describes the mathematical basis of the segmentation algorithm. Section 3 examines the performance of the fitting cost estimators that are used in the algorithm and Section 4 shows segmentation results.

2 Constructive Fitting

In this paper we focus on one particular aspect of segmentation. Given an initial seed region, how can we find a region of maximal size where a curve or surface can be approximated sufficiently well by a polynomial of given order? And how can we find this region with minimal computational effort? In fact, although region growing and curve segmentation based on polynomial fitting is often a computationally expensive process [10, 11, 12], we propose a linear-time algorithm that solves some simple but basic segmentation problems, such as the segmentation of a curve in piecewise linear segments. This efficiency is based on a simple principle. The purpose of region growing and segmentation is to partition a curve or image into segments that satisfy a certain criterion, such as: being straight or smooth. To find such segments, a typical region growing algorithm starts from a small seed segment, and then repeatedly tries to add new points to this segment, while verifying whether the segmentation criterion is still satisfied for the enlarged segment. If not, a new segment is started, or another point is chosen. In general, the verification of the segmentation criterion requires increasingly more computation time when the segment gets larger. It is the purpose of this paper to show that this computation time can be reduced considerably by comparing the extension point with a small, but well-chosen set of reference points within the segment. Since the number of reference points is constant during the entire region growing process, the resulting algorithms have linear time complexity. This process has also been called constructive fitting.

We briefly review the basics of the particular method of fitting geometric primitives to data sets, on which the region growing processes proposed in this paper are based.

This method is called constructive fitting because it allows us to construct global fits to an entire data set from so-called elemental fits to small parts of the data [13, 14, 15]. We consider constructive fitting problems for one variable. The generalization to other dimensions is straightforward. Let $C = \{(x_0, y_0), \dots\} \subset \mathbb{Z}^2$ be the points of a finite curve. Let G be a vector space of fitting functions, for instance, the vector space of polynomial functions of the form

$$g(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_l x^l.$$

To simplify the exposition we impose the mild constraint that the curve C contains at least $n + 1$ points with distinct x -coordinates, where n is the dimension of the space of fitting functions. The uniform (or L_∞) fitting cost of fitting $g(x)$ to the curve C is defined as

$$r_g(C) = \max_{(x_i, y_i) \in C} |g(x_i) - y_i|.$$

The *best fit* is the function $g(x)$ in G for which $r_g(C)$ is minimal. We denote this minimal cost as $r(C)$, and we call it the *fitting cost*. To be precise,

$$r(C) = \min_{g \in G} r_g(C).$$

We call $\tau = 2r(C)$ also the thickness of C , because the points of C are contained in a strip of thickness τ parallel to the graph of $g(x)$. This is also related to the notion of α -thickness in [5].

The best fit and its fitting cost can be computed from elemental fits to the so-called *elemental subsets* of C . These are subsets of the curve C that contain precisely $n + 1$ points, where n denotes the dimension of the vector space of the fitting functions G . The fitting cost over an elemental subset itself can be computed in a straightforward manner. To be precise, let $D = \{(x_{i_1}, y_{i_1}), \dots, (x_{i_{n+1}}, y_{i_{n+1}})\}$ be an elemental subset of R . Let E_j denote the cofactor of the element at the intersection of the last column and the j th row of the following matrix:

$$(A_D | B_D) = \left(\begin{array}{cccc|c} 1 & x_{i_1} & \dots & x_{i_1}^l & y_{i_1} \\ \dots & & & & \\ 1 & x_{i_{n+1}} & \dots & x_{i_{n+1}}^l & y_{i_{n+1}} \end{array} \right)$$

Then one can show that the fitting cost for the elemental subset D is equal to

$$\begin{aligned} r(D) &= \det(A_D | B_D) / (|E_1| + \dots + |E_{n+1}|) \\ &= |E_1 y_{i_1} + \dots + E_{n+1} y_{i_{n+1}}| / (|E_1| + \dots + |E_{n+1}|), \end{aligned} \quad (1)$$

provided the denominator at the right side is non-vanishing [14, 16]. Furthermore, one can prove that the fitting cost over C is the maximal value of the elemental fitting costs (1) over all elemental subsets of the curve C [14]:

Proposition 1. *Let $\mathcal{M}$ be the collection of all elemental subsets D of C for which $|E_1| + \dots + |E_{n+1}| > 0$. If $\mathcal{M}$ is non-empty, then we have $r(C) = \max_{D \in \mathcal{M}} r(D)$.*

Although Prop. 1 yields a simple method to compute $r(C)$, the computational cost of evaluating $r(D)$ for all elemental subsets can be high. We can obtain a reliable estimate

of the fitting cost, with far less computations if, instead of calculating $\max_{D \in \mathcal{M}} r(D)$, we compute $\max_{D \in \widetilde{\mathcal{M}}} r(D)$ where $\widetilde{\mathcal{M}}$ is some small subcollection of $\mathcal{M}$.

One can prove that the estimate is reliable provided $\widetilde{\mathcal{M}}$ forms a so-called *rigid subcollection* of elemental subsets [14]. The rigidity of $\widetilde{\mathcal{M}}$ depends on its elemental subsets, and can be verified as follows. Assume that the points of the curve C have been ordered in a fixed but arbitrary manner, i.e., $C = \{(x_1, y_1), \dots, (x_k, y_k)\}$. Let $D = \{(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), \dots, (x_{i_{n+1}}, y_{i_{n+1}})\}$ be an elemental subset of C , with $1 \leq i_j \leq k$, and let E_j , $1 \leq j \leq n+1$, be the cofactors corresponding to the elemental subcountour D . In the vector space $\mathbf{R}^k$ let

$$v_D = \frac{(\dots, E_1, \dots, E_2, \dots, E_{n+1}, \dots)}{|E_1| + |E_2| + \dots + |E_{n+1}|} = (d_1, \dots, d_k),$$

denote a vector associated with D , where the j th cofactor E_j occurs as the i_j th element. The remaining $k - (n+1)$ entries are zero. The vectors v_D , where $D \in \mathcal{M}$, span a linear subspace L of $\mathbf{R}^k$. If we only consider vectors v_D that correspond to elemental subsets in the subcollection $\widetilde{\mathcal{M}}$, then these vectors span a linear subspace $\widetilde{L} \subset L$, where $\widetilde{L}$ can be smaller than L . Whether the elemental subsets $\widetilde{\mathcal{M}}$ can be used to accurately estimate the fitting cost depends on the question whether these two linear subspaces are equal or not, a notion formalized as follows.

Definition 1. We say that $\widetilde{\mathcal{M}}$ forms a *rigid subcollection of elemental subsets* if the linear space $\widetilde{L}$ spanned by the vectors v_D , where $D \in \widetilde{\mathcal{M}}$, is the same as the linear space L spanned by the vectors v_D , where $D \in \mathcal{M}$.

Apart from obeying the condition $r(C) \leq \tau$, a digitized curve is also an ordered set of points with topological properties, the most common being that the curve must be either 4- or 8-connected. We incorporate this property by requiring that two subsequent points (x_i, y_i) , (x_{i+1}, y_{i+1}) on the curve satisfy $|y_i - y_{i+1}| \leq \epsilon$. To obtain 4- or 8-connectivity we choose $\epsilon = 1$. With the above concept of rigid subcollections we have the following proposition which extends previous results [13, 14].

Proposition 2. Let $\widetilde{\mathcal{M}}$ be a rigid subcollection of elemental subsets chosen from the collection of all elemental subsets $\mathcal{M}$ for a curve length k . Let τ denote the maximum thickness of a curve segment, and ϵ the maximum absolute difference between y -coordinates of subsequent points on the curve. For the vectors $(z_1, \dots, z_k) \in \mathbf{R}^k$, let S be the k -dimensional polytope defined by the inequalities

$$|z_i - z_{i+1}| \leq \epsilon, \text{ for } i = 1, \dots, k-1, \quad (2)$$

and

$$\frac{|d_1 z_1 + \dots + d_k z_k|}{|d_1| + \dots + |d_k|} \leq \frac{\tau}{2}, \quad \text{for all } D \in \widetilde{\mathcal{M}}. \quad (3)$$

Then for each elemental subset $P \in \mathcal{M}$, with associated vector $(p_1, \dots, p_k)$, the values

$$\gamma_P(\widetilde{\mathcal{M}}, \tau, \epsilon) = \max_{(z_1, \dots, z_k) \in S} \frac{2}{\tau} \frac{|p_1 z_1 + \dots + p_k z_k|}{|p_1| + \dots + |p_k|}, \quad (4)$$

and

$$\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon) = \max_P \gamma_P(\widetilde{\mathcal{M}}, \tau, \epsilon) \quad (5)$$

are well defined.

Proof. It is sufficient to show that $\gamma_P()$ is well-defined for all elemental subsets P . By the rigidity assumption there is a finite set of weight vectors $v_A = (\dots, a_1, \dots, a_l, \dots)$, $v_B = (\dots, b_1, \dots, b_l, \dots)$, $\dots$, $v_D = (\dots, d_1, \dots, d_l, \dots)$ with $A, B, \dots, D \in \widetilde{\mathcal{M}}$, such that $v_P = (\dots, p_1, \dots, p_l, \dots)$ can be written as a linear combination $v_P = \alpha v_A + \beta v_B + \dots + \omega v_D$. It follows that

$$\begin{aligned} |p_1 z_1 + \dots + p_l z_l| &\leq |\alpha| |a_1 z_1 + \dots + a_l z_l| + \\ &\quad |\beta| |b_1 z_1 + \dots + b_l z_l| + \dots + \\ &\quad |\omega| |d_1 z_1 + \dots + d_l z_l|. \end{aligned}$$

As a consequence of the constraints (3), every term at the right side of this inequality has an upper bound. Hence $|p_1 z_1 + \dots + p_l z_l| / (|p_1| + \dots + |p_l|)$ will also have an upper bound, and it follows that $\gamma_P()$ is well-defined, and therefore $\gamma()$ is well defined.

The primary use of Prop. 2 is to determine the relative estimation error. Suppose we find the estimated cost $\tilde{r}(C)$ of a curve from the elemental subsets in $\widetilde{\mathcal{M}}$, and that we have $\tilde{r}(C) \leq \tau/2$. Then from Prop. 2 it follows that the outcome of any other elemental subset $D \in \mathcal{M} \setminus \widetilde{\mathcal{M}}$ satisfies $r(D) \leq \gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)\tau/2$. Thus, although the thickness of the curve is not known exactly, it must be less than $\tau\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$.

Prop. 2 yields an upper bound $\gamma()$ for the estimation error, which is found by solving linear programming problems for different cost functions (4), but over the same polytope S , defined by the inequalities (2) and (3). Each vector in S corresponds to a possible configuration of points satisfying the connectivity constraints (3) and with estimated fitting cost less than $\tau/2$. Although the fitting costs $r(C)$ and $\tilde{r}(C)$ depend on the real point positions of the curve C , the upper bound $\gamma()$ only depends on the x -coordinates of the elemental subsets in $\widetilde{\mathcal{M}}$, the thickness τ and the connectivity parameter ϵ , since these parameters completely determine the polytope S and the linear cost functions (4). Thus the upper bound $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$ can be computed once and for all, since it does not depend on the curve itself. The rigidity of $\widetilde{\mathcal{M}}$ is easy to verify [13].

As a special case, $\gamma(\widetilde{\mathcal{M}}, \tau, \infty)$ denotes the value when no topological constraint is imposed. To compute $\gamma(\widetilde{\mathcal{M}}, \tau, \infty)$ we can apply Prop. 2 with ϵ a large positive number. The common case where we fit digital straight segments to connected curves corresponds to $\gamma(\widetilde{\mathcal{M}}, 1, 1)$. By observing the expressions (4), (2), and (3) in Prop. 2 one can see that $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$ only depends on the ratio $\mu = \tau/\epsilon$. In Section 3 we will see that when ϵ is small compared to the thickness τ , then the connectivity of the curve cannot be neglected when estimating fitting costs. Previous results in [13, 14] did not take into account connectivity and only considered the case $\gamma(\widetilde{\mathcal{M}}, \tau, \infty)$.

According to Prop. 2, given $\widetilde{\mathcal{M}}$, we can compute $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$ which gives us a precise measure for the maximal error of the estimated fitting cost $\tilde{r}(R)$. Moreover, $\gamma(\widetilde{\mathcal{M}})$ depends only on the x -coordinates of the points on the subsets D in $\widetilde{\mathcal{M}}$ and not on the curve itself. In principle, Prop. 2 enables us to evaluate the maximal error for a particular estimator $\hat{\mathcal{M}}$, by performing an exhaustive search for the estimator that has the

lowest maximal error. This has been explored in previous work, but without topological constraints [13]. With topological constraints the performance depends strongly on the ratio between the allowed thickness of a curve and the connectivity.

3 Performance of Fitting Cost Estimators

When $\widetilde{\mathcal{M}}$ satisfies the conditions of Definition 1, we refer to it as a fitting cost estimator $\widetilde{\mathcal{M}}$. Using Prop. 2 we can search, in a systematic way, for the estimator $\widetilde{\mathcal{M}}$ that yields the lowest possible estimation error.

3.1 Linear Fitting Cost Estimators

We will illustrate our search with a simple fitting problem in one variable. Fig. 1 shows a curve C . Let G be a two-dimensional vector space of fitting functions of the form $g(x) = \alpha_0 + \alpha_1 x$. Then the best fit to C is the function $g(x)$ for which $\max_{(x_i, y_i) \in C} (|g(x_i) - y_i|)$ is minimal. When C is a digital arc the fitting cost $r(C)$ is a measure for the straightness of the digital arc [14]. First we compute the fitting cost exactly. According to Prop. 1, the fitting cost $r(C)$ is equal to $\max_D r(D)$, where the maximum is taken over all 3-point subsets of C . The fitting cost $r(D)$ of a subset is equal to $|E_1 y_i + E_2 y_j + E_3 y_k| / (|E_1| + |E_2| + |E_3|)$, where the E_j are the cofactors of the last column of the matrix

$$\begin{pmatrix} 1 & x_{i_1} & y_i \\ 1 & x_{i_2} & y_j \\ 1 & x_{i_3} & y_k \end{pmatrix}$$

Thus, for the function shown in Fig. 1, we find $r(C) = 8/9 \approx 0.89$. Note that the exact computation of $r(C)$ involves the evaluation of 120 elemental fitting costs $r(D)$ (a 10-point curve has 120 distinct 3-point subsets).

Estimating $r(C)$ instead of computing it exactly, considerably reduces the number of elemental fitting costs that must be evaluated. The challenge is how to choose these elemental subsets so that the estimate is as good as possible. Table 1 shows the theoretical performance of different estimators when they are applied to a curve segment consisting of 10 subsequent points $(x_1, y_1), (x_1 + 1, y_2), \dots, (x_1 + 9, y_{10})$. The first column describes the estimator, where the index i varies in each as $i = 2, \dots, 9$. For example, (p_1, p_i, p_{i+1}) denotes the estimator $\widetilde{\mathcal{M}} = \{(p_1, p_2, p_3), (p_1, p_3, p_4), \dots, (p_1, p_9, p_{10})\}$. The estimator called *iterative subdivision* is constructed by iteratively subdividing the curve into shorter curves as follows: $\widetilde{\mathcal{M}} = \{(p_1, p_5, p_{10}), (p_1, p_3, p_5), (p_5, p_7, p_{10}), (p_1, p_2, p_3), \dots, (p_7, p_9, p_{10})\}$. For $\widetilde{\mathcal{M}}_7$, the index varies from $i = 3$ to $i = 9$, and also $(1, 2, 3)$ is included.

The first two estimators are general in the sense that they assume all points are known from the start. The estimators $\widetilde{\mathcal{M}}_3, \dots, \widetilde{\mathcal{M}}_7$ can be used in incremental algorithms, where only one point at a time is added to a segment, while each time the fitting cost is evaluated. For curve segmentation the incremental algorithms are the most interesting. The estimator $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5$ use the conditions of both $\widetilde{\mathcal{M}}_4$ and $\widetilde{\mathcal{M}}_5$ to estimate the fitting cost.

Columns 2 to 5 show the worst case behavior $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$ for each estimator as computed by Prop. 2. Note that these values for $\gamma()$ do not depend on any particular curve, but that they are valid for any curve of length 10, with thickness not larger than τ and satisfying the connectivity constraint determined by ϵ . Column 6 on the other hand shows the estimated cost for one particular curve which is shown in Fig. 1. In all cases the real fitting cost of 0.89 is always less than product of the estimated cost and maximum relative error $\gamma()$. Columns 7 and 8 are discussed in Section 4.

Table 1. Performance of different estimators for varying values of the thickness and connectivity constraint, for the curve length $k = 10$

Estimators	$\gamma(\widetilde{\mathcal{M}}, 1, \infty)$	$\gamma(\widetilde{\mathcal{M}}, 1, 2)$	$\gamma(\widetilde{\mathcal{M}}, 1, 1)$	$\gamma(\widetilde{\mathcal{M}}, 2, 1)$	estim. cost	max cost	aver. cost
General estimators:							
$\widetilde{\mathcal{M}}_1: (p_1, p_i, p_{10})$	2.	2.	2	1.5	0.89		
$\widetilde{\mathcal{M}}_2$: iterative subdivision	3.1	2.94	2.65	1.71	0.72		
Incremental algorithms:							
$\widetilde{\mathcal{M}}_3: (p_{i-1}, p_i, p_{i+1})$	20.	7.11	4	2	0.5		
$\widetilde{\mathcal{M}}_4: (p_1, p_i, p_{i+1})$	4.13	3.92	3.47	2	0.5	36.	11.
$\widetilde{\mathcal{M}}_5: (p_1, p_{\lceil i/2 \rceil + 1}, p_{i+1})$	21.58	5.4	3.09	1.84	0.72	2.	1.03
$\widetilde{\mathcal{M}}_6: (p_1, p_{\lceil \frac{i+4}{3} \rceil}, p_{i+1})$	19.83	6.56	3.56	2.	0.66	3.96	1.41
$\widetilde{\mathcal{M}}_7: (p_1, p_{i-1}, p_{i+1})$	16.74	3.27	2.37	1.83	0.89	2.6	0.97
Combined incremental:							
$\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5$	3.	2.94	2.76	1.84	0.72	2.	1.02
$\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6$	2.94	2.94	2.68	1.84	0.72	2.	1.02
$\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$	2.17	2.14	1.95	1.71	0.89	2.08	0.84

Table 1 clearly shows that the performance $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$ of an estimator depends strongly on the ratio between the thickness τ and the connectivity constraint ϵ . In fact, although according to the first column $\widetilde{\mathcal{M}}_4$ performs much better ($\gamma = 4.13$) than $\widetilde{\mathcal{M}}_5$ ($\gamma = 21.58$) for arbitrary points sets, columns 4 and 5 show that when we impose the topological constraint $\epsilon = 1$, $\widetilde{\mathcal{M}}_5$ ($\gamma = 1.84$) is better than $\widetilde{\mathcal{M}}_4$ ($\gamma = 2$). The reason can be understood as follows. The estimator $\widetilde{\mathcal{M}}_4$ compares the y -coordinate of the new point p_{i+1} with the y -coordinates of the points p_1 and p_i . However, if we impose the topological condition that the difference between the y -coordinates of subsequent points is never larger than one, then for $\widetilde{\mathcal{M}}_4$ the topological condition (2) coincides almost completely with the condition (3). For $\widetilde{\mathcal{M}}_5$, however, the topological constraint makes sense, and $\gamma(\widetilde{\mathcal{M}}_5, \tau, \epsilon)$ drops when ϵ gets smaller. Also note that for all estimators in Table 1 we have $\gamma(\widetilde{\mathcal{M}}, \tau = 1, \epsilon = 1) \geq \gamma(\widetilde{\mathcal{M}}, \tau = 2, \epsilon = 1)$. That is, for a 4- or 8-connected curve, the relative error on the estimated thickness is smaller for $\tau = 2$ than for $\tau = 1$. The reason is that for a curve with thickness $\tau = 2$, connectivity is a more important additional constraint than for a curve with thickness $\tau = 1$.

The best estimator is $\widetilde{\mathcal{M}}_1$, which performs well for curves and for disconnected point sets. However, $\widetilde{\mathcal{M}}_1$ is not suited for curve segmentation. The best estimator for curve segmentation is $\widetilde{\mathcal{M}}_7$, if we assume that the curves are either 4- or 8-connected. It makes sense to combine several estimators. The overall performance of estimator $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$ is better than the performance of its parts. The drawback is that the combined estimator evaluates four times more conditions than for example $\widetilde{\mathcal{M}}_5$. Nonetheless, the time complexity of this combined estimator is still $O(N)$, where N denotes the number of points on the segment, as we use 4 elemental subsets to examine the fitting cost for each new point.

Table 1 only shows the results for curves of length 10. For lengths other than 10 we get similar results. An estimator that performs well for length 10, such as $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$, also performs well for curves of arbitrary length.

Fig. 1. Simple curve used to examine the performance of the estimators in Table 1

3.2 Quadratic Fitting Cost Estimators

Table 2 shows the results for fitting cost estimators that estimate the cost of fitting a quadratic polynomial to a curve of length 10 with points that have consecutive x -coordinates. As for line fitting, the ratio of the thickness τ and the topological constraint ϵ plays an important role. The first column describes the kind of estimator, where the index i varies in each as $i = 3, \dots, 9$. Note that the topological constraints play an even larger role for quadratic fitting than for straight fitting. The requirement that a parabolic segment is connected already imposes strong restrictions.

4 Curve Segmentation

To illustrate the fitting cost estimators we discuss the decomposition of a digitized curve into digital straight or parabolic segments. For Table 1 and 2 we made the assumption that the curve consists of subsequent points with x -coordinates that increase regularly. In the general case, a curve can be quite different and subsequent points may have the same x -coordinate, e.g., $(x_1, y_1), (x_1 + 1, y_2), (x_1 + 1, y_3), (x_1 + 2, y_4), \dots$. Without restriction, Prop. 2 can also be used to find good estimators for this case, by using an appropriate form for the curve C and its elemental subsets. The number of possibilities is unlimited, however, and for each possible configuration of x -coordinates we will find other values for $\gamma(\widetilde{\mathcal{M}}, \tau, \epsilon)$. The examples show that the best estimators of Table 1 and 2 also perform well for more general curve segments.

Table 2. Performance of different quadratic estimators for varying values of the thickness τ and connectivity constraint ϵ

Estimators	$\gamma(\widetilde{\mathcal{M}}, 1, \infty)$	$\gamma(\widetilde{\mathcal{M}}, 1, 2)$	$\gamma(\widetilde{\mathcal{M}}, 1, 1)$	$\gamma(\widetilde{\mathcal{M}}, 2, 1)$
Region growers:				
$\widetilde{\mathcal{M}}_1: (p_1, p_{\lfloor \frac{i+1}{2} \rfloor}, p_i, p_{i+1})$	3.72	3.26	2.15	1.26
$\widetilde{\mathcal{M}}_2: (p_1, p_{\lfloor \frac{i+3}{3} \rfloor}, p_{\lfloor \frac{2i+3}{3} \rfloor}, p_{i+1})$	140.	4.34	2.31	1.26
$\widetilde{\mathcal{M}}_3: (p_1, p_{\lfloor \frac{i+1}{2} \rfloor}, p_{\lfloor \frac{3i+3}{4} \rfloor}, p_{i+1})$	154.	3.8	2.23	1.26
$\widetilde{\mathcal{M}}_4: (p_1, p_{\lfloor \frac{i+5}{4} \rfloor}, p_{\lfloor \frac{i+3}{2} \rfloor}, p_{i+1})$	124.	4.58	2.51	1.26
$\widetilde{\mathcal{M}}_5: (p_{\lfloor \frac{i+5}{5} \rfloor}, p_{\lfloor \frac{i+2}{2} \rfloor}, p_{\lfloor \frac{4i+5}{5} \rfloor}, p_{i+1})$	54.	3.21	1.93	1.25
Combined growers:				
$\widetilde{\mathcal{M}}_1 \cup \widetilde{\mathcal{M}}_5$	3.01	2.72	1.91	1.25
$\widetilde{\mathcal{M}}_1 \cup \widetilde{\mathcal{M}}_2 \cup \widetilde{\mathcal{M}}_3 \cup \widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5$	2.37	2.32	1.85	1.24

Fig. 2. (a) Original contour, (b) perfect segmentation, (c) segmentation by region grower $\widetilde{\mathcal{M}}_5$, (d) segmentation by the combined region grower $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$

To find digital straight curve segments of S of maximal length, we use a simple linear-time algorithm. Each time a new point is added, we perform elemental fits to verify whether the enlarged segment is still straight. Since we estimate the fitting cost (instead of computing it exactly), the algorithm that we propose is not error-free. However, since an upper bound for the fitting cost is known instead of obtaining segments of thickness $\leq \tau$, in the worst case we have segments of thickness $\leq \gamma(\cdot)\tau$. In fact, average case behavior will be much better than worst case behavior.

Fig. 3. Segmentation of Kelloggii plant leaf: (a) Original contour, (b) piecewise linear segmentation with segment grower $\widetilde{\mathcal{M}}_4 \cup \dots \cup \widetilde{\mathcal{M}}_7$ of Table 1 with allowed thickness $\tau = 1$, (c) segmentation by quadratic segment grower $\widetilde{\mathcal{M}}_1 \cup \dots \cup \widetilde{\mathcal{M}}_5$ also with allowed thickness $\tau = 1$, of Table 2, (d) the same segments as in (c), but now also the parabolas are shown and the endpoints of the segments

Fig. 2 shows the results for a more complex 8-connected curve. Fig. 2(a) shows the original contour, (b) the result of perfect segmentation (in which all elemental subsets are used to compute the fitting cost), (c) the result of segmentation by region grower $\widetilde{\mathcal{M}}_5$, and (d) the result of segmentation by the combined region grower $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$, with $\tau = 1$ for all cases. For most applications, when extracting feature points from contours, the results in (d) are satisfactory and come with large computational gain.

Column 7 in Table 1 shows the maximal thickness of the segments when the curve in Fig. 2(a) is segmented using the corresponding estimator for $\tau = 1$. For perfect estimates the values in column 7 should not exceed 1. For curves where all x -coordinates are distinct and do not differ by more than one, the values should be less than $\gamma(\widetilde{\mathcal{M}}, 1, 1)$. This is not always the case since the topology of the curve can be different, as mentioned previously, and because the length of the curve is most of the time not equal to ten. Column 8 gives the average thickness of the segments, which shows that the average estimates are better than the worst case estimates. Again, $\widetilde{\mathcal{M}}_4 \cup \widetilde{\mathcal{M}}_5 \cup \widetilde{\mathcal{M}}_6 \cup \widetilde{\mathcal{M}}_7$ is clearly the best estimator. The average perfect fitting cost of the segments that it accepted is 0.84.

Fig. 3 shows the result of quadratic segment growers for plant leaves. One possible goal of splitting the 8-connected contour of a plant into segments is to locate the feature points that can be used for matching, i.e. the sharp corners of the leaf. Fig. 3 (a) shows

Fig. 4. Segmentation for different thicknesses τ . (a),(b),(c) segmentation by quadratic segment grower $\tilde{\mathcal{M}}_1 \cup \dots \cup \tilde{\mathcal{M}}_5$ of Table 2 with thicknesses $\tau = 2, 4, 8$, respectively; (d),(e),(f) segmentation with linear segment grower $\tilde{\mathcal{M}}_4 \cup \dots \cup \tilde{\mathcal{M}}_7$ of Table 1 with allowed thicknesses $\tau = 2, 4, 8$, respectively.

the original contour of a plant leaf, (b) the result of linear segment growing and (c) and (d) the result of quadratic segment growing, with allowed thickness $\tau = 1$. Clearly, for this kind of curves it makes sense to use quadratic segments. The number of segments in (b) is almost twice the number of segments in (d) and there are more endpoints in (b) that cannot be used as feature points. Fig. 4 shows the results of fitting with different values for the thickness τ . For this kind of curve quadratic fitting remains more accurate for all thicknesses.

5 Concluding Remarks

We have applied constructive fitting techniques to develop segment growing algorithms for curve segmentation based on the following idea. When extending a curve C_k with a new point p_k it is not necessary to compare the point with the entire curve C_k . Instead it suffices to compare the point with one or more elemental subsets. Thus the time complexity of the contour growing process is greatly reduced, and the algorithm is simple enough to be implementable on dedicated hardware. The major challenge is how to choose the subsets so that the estimation of the fitting is as reliable as possible. Theoretical results allow us to select the elemental subsets in an optimal fashion. To be specific, there is a performance parameter $\gamma()$ that allows us to quantify the quality of the subsets. In this work we have shown that the choice of the subsets depends on the allowed thickness of the contour and the connectivity constraints. The emphasis in this paper was on how to develop computationally efficient segmentation algorithms through variable order fitting algorithms. However, also the robust estimation of curve or surface parameters may benefit from fast variable order fitting [1, 17].

References

1. Lachaud, J.-O., Vialard, A., de Vieilleville, F.: Analysis and Comparative Evaluation of Discrete Tangent Estimators. *Lecture Notes in Computer Science* **3429** (2005) 240–251
2. Coeurjolly, D., Klette, R.: Comparative evaluation of length estimators of digital curves. *IEEE Trans. Pattern Anal. Machine Intell.* **26** (2004) 252–258
3. Smeulders, A., Dorst, L.: Decomposition of discrete curves into piecewise straight segments in linear time. In: *Vision Geometry* (R. A. Melter, A. Rosenfeld, and P. Bhattacharya, eds.), *Contemporary Mathematics Series*, American Mathematical Society **119** (1991) 169–195
4. Debled-Renesson, I., Revelles, J.-P.: A linear algorithm for segmentation of discrete curves. *International Journal of Pattern Recognition and Artificial Intelligence* **9** (1995) 635–662
5. Buzer, L.: An elementary algorithm for digital line recognition in the general case. *Lecture Notes in Computer Science* **3429** (2005) 299–310
6. Leclerc, Y., Zucker, S.: The local structure of image discontinuities in one dimension. *IEEE Trans. Pattern Anal. Machine Intell.* **9** (1987) 341–355
7. Dunham, J.: Optimum uniform piecewise linear approximation of planar curves. *IEEE Trans. Pattern Anal. Machine Intell.* **8** (1986) 67–75
8. Rosin, P., West, G.: Nonparametric segmentation of curves into various representations. *IEEE Trans. Pattern Anal. Machine Intell.* **17** (1995) 1140–1153
9. Boyer, K., Mirza, M., Ganguly, G.: The robust sequential estimator: A general approach and its application to surface organization in range data. *IEEE Trans. Pattern Anal. Machine Intell.* **16** (1994) 987–1001
10. Besl, P., Jain, R.: Segmentation through variable-order surface fitting. *IEEE Trans. Pattern Anal. Machine Intell.* **10** (1988) 167–192
11. Cooper, D., Yalabik, N.: On the computational cost of approximating and recognizing noise-perturbed straight lines and quadratic arcs in the plane. *IEEE Trans. Comput.* **25** (1976) 1020–1032
12. Taubin, G.: Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* **13** (1991) 1115–1138
13. Veelaert, P.: Linear-time algorithms for region growing with applications to image and curve segmentation. In: *Proceedings of the SPIE conference on Vision Geometry VI* **3168** (1997) 76–87
14. Veelaert, P.: Constructive fitting and extraction of geometric primitives. *CVGIP: Graphical Models and Image Processing* **59** (1997) 233–251
15. Stromberg, A.: Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM J. Sci. Comput.* **14** (1993) 1289–1299
16. Veelaert, P.: On the flatness of digital hyperplanes. *J. Math. Imaging and Vision* **3** (1993) 205–221
17. Teelen, K., Veelaert, P.: Improving difference operators by local feature detection. To appear in proceedings DGCi 2006.