HAL
open science

A discrete approach for decomposing noisy digital contours into arcs and segments

Phuc Ngo, Hayat Nasser, Isabelle Debled-Rennesson

To cite this version:

Phuc Ngo, Hayat Nasser, Isabelle Debled-Rennesson. A discrete approach for decomposing noisy digital contours into arcs and segments. ACCV Workshop 2016, Nov 2016, Taipei, Taiwan. 10.1007/978-3-319-54427-4_36 . hal-01375089

HAL Id: hal-01375089

https://inria.hal.science/hal-01375089v1

Submitted on 3 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A discrete approach for decomposing noisy digital contours into arcs and segments

Phuc Ngo[1,2], Hayat Nasser[1,2], Isabelle Debled-Rennesson[1,2]

[1] Universit de Lorraine, LORIA, UMR 7503, Vandoeuvre-ls-Nancy, F-54506, France
[2] CNRS, LORIA, UMR 7503, Vandoeuvre-ls-Nancy, F-54506, France
{hoai-diem-phuc.ngo,hayat.nasser,isabelle.debled-rennesson}@loria.fr

Abstract. In the paper, we present a method for decomposing a discrete noisy curve into arcs and segments which are the frequent primitives in digital images. This method is based on two tools: dominant point detection using adaptive tangential cover and tangent space representation of the polygon issued from detected dominant points. The experiments demonstrate the robustness of the method w.r.t. noise.
Keywords. Adaptive tangential cover, dominant point detection, tangent space, curve reconstruction, vectorization.

1 Introduction

The extraction of meaningful features from image contour is an important problem in computer vision. Many existing methods used critical points or straight segments as meaningful features to construct high level descriptors of images. Among the primitives (straight lines, circles, circular arcs, ellipses, parabolas, ect.), arc and straight segment are those appearing often in images, especially in graphic document images. Furthermore, a combination of circular arcs and straight line segments is an interesting solution to avoid an approximation of arcs by many straight segments or critical points. Many methods have been proposed to decompose a planar curve into arcs and segments. Surveys of the different approaches can be found in [1, 2]. Several of these methods work in a transformed domain [3, 4].

Inspired by the works about dominant point detection using adaptive tangential cover (ATC) [5] and tangent space representation of the polygon issued from dominant points detected [1, 6], we present in this paper a method for decomposing digital curves into arcs and segments. More precisely, the algorithm first detects the dominant points of the input curve. Then, it transforms the polygon of the detected dominant points into the *tangent space* [4]. In [1, 6], it is shown that a sequence of chords of an arc corresponds to a sequence of collinear points in the tangent space. Thus, the problem of digital arc detection becomes digital straight line recognition in the tangent space. In other words, one can use the tangent space to identify the segments and arcs of a curve. However, the tangent space is very sensitive to noise. By taking the advantage of the discrete structure of ATC to noisy curves, the proposed method uses the ATC to detect dominant points and thus makes it more robust to noise.

The paper is organized as follows: *Sec.* 2 consists of definitions and results used in this paper. The main results are presented in *Sec.* 3 with the proposed decomposition algorithm. *Sec.* 4 shows the experimental results and comparisons for segmenting digital curve into arcs and segments. Finally, in *Sec.* 5, we conclude and discuss the perspectives.

2 Discrete curve structure

2.1 Maximal blurred segments

Maximal blurred segment has been introduced by Debled-Rennesson *et al.* in [7] as an extension of arithmetical discrete line [8] with a width parameter for noisy or disconnected digital contours.

Definition 1. *An* ***arithmetical discrete line*** $\mathcal{D}(a, b, \mu, \omega)$*, with a direction vector* (b, a)*, a lower bound* μ *and an arithmetic thickness* ω *(with* $a, b, \mu, \omega \in \mathbb{Z}$ *and* $gcd(a, b) = 1$*) is the set of integer points* (x, y) *verifying* $\mu \leq ax - by < \mu + \omega$*.*

Definition 2. *A set* S_f *is a* ***blurred segment of width*** ν *if there exists a discrete line* $\mathcal{D}(a, b, \mu, \omega)$ *containing* S_f *has the vertical (or horizontal) distance* $d = \frac{\omega - 1}{\max(|a|, |b|)}$ *equal to the vertical (or horizontal) thickness of the convex hull of* S_f*, and* $d \leq \nu$*.*

Let C be a discrete curve and $C_{i,j}$ a sequence of points of C indexed from i to j. Let denote the predicate "$C_{i,j}$ is a blurred segment of width ν" as $BS(i, j, \nu)$.

Definition 3. $C_{i,j}$ *is called a* ***maximal blurred segment (MBS) of width*** ν *and noted* $MBS(i, j, \nu)$ *iff* $BS(i, j, \nu)$*,* $\neg BS(i, j + 1, \nu)$ *and* $\neg BS(i - 1, j, \nu)$*.*

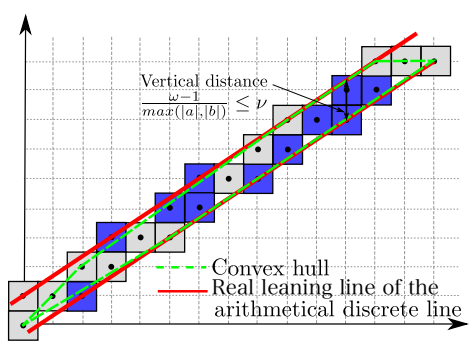


Fig. 1. Example of arithmetical discrete line $\mathcal{D}(2, 3, -3, 5)$ (grey and blue points) and a blurred segment of width $\nu = 1.4$ (grey points) bounded by D.

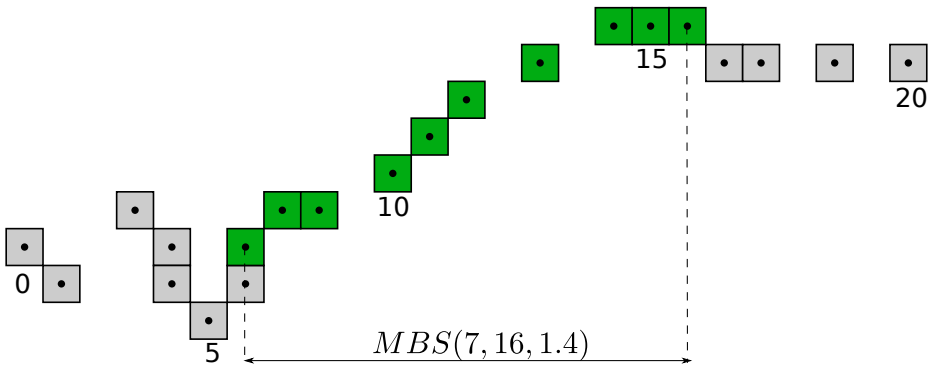


Fig. 2. Maximal blurred segment of width $\nu = 1.4$ (green points).

These notions are illustrated in *Figs.* 1–2. The sequence of maximal segments along a digital contour is called a *tangential cover* and used in numerous discrete geometric estimators (see [9] for a state of the art): length, tangent, curvature estimators, detection of convex or concave parts,*ect.* In [7], a linear algorithm is proposed for computing the tangential cover of a given curve. Inspired by this algorithm, in [1, 10] a quasi-linear algorithm is developed for tangential cover using MBS of a priori fixed width ν, namely **width ν tangential cover**.

Still in [1], the width ν tangential cover is proposed to deal with the noisy contours. However, this fixed width value ν needs to be manually adjusted and the method is inadequate to local amount of noise which can appear on real contours.

2.2 Adaptive tangential cover

Recently, a notion of **adaptive tangential cover (ATC)** was introduced in [5] which is composed of MBS with appropriated widths. More precisely, the ATC is composed of MBS of different width values varying in function of the noise perturbations presented on the contour. In particular, it uses the local noisy estimator, namely *meaningful thickness* [11], to determine the significant width locally at each point of the contour to analyze the considering curve. This meaningful thickness is used as an input parameter of the width ν tangential cover framework to compose the MBS of ATC with appropriate widths w.r.t. noise. A quasi-linear algorithm is developed in [5] to compute the ATC. Example of ATC is given in *Fig.* 3.

In the ATC, the obtained MBS decomposition of various widths transmits the noise levels and the geometrical structure of the given discrete curve. In [5], a non-parametric algorithm is proposed to build the ATC of a given discrete curve.

In the next section, we discuss about the method for decomposing noisy digital curves into arcs and segments. This method is based on: (1) the detection of key points, namely *dominant points*, on a curve using the ATC [5], and (2)

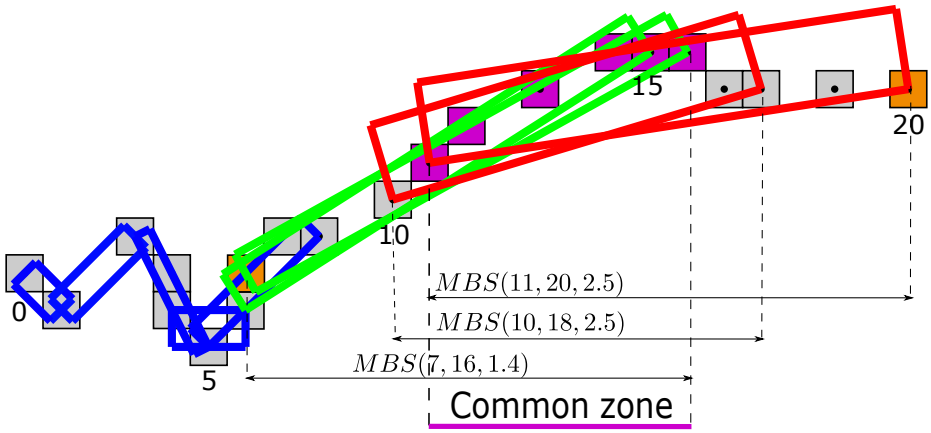


Fig. 3. Adaptive tangential cover with three width values $\nu = 1, 1.4$ and 2.5 (in blue, green and red respectively) deduced from the local noise level estimation. In pink are points in the common zone determined by the last three MBS of the ATC, and in orange are its two endpoints.

the *tangent space* [3, 4] –a tangent angle versus length segment plot– of polygon issued from dominant points detected to recognize points of arcs or segments on the considered polygon.

3 Arcs and segments decomposition method

Nguyen *et al.* in [1, 6] assumed that junctions between two different primitives are dominant points of a curve, and proposed an algorithm for decomposing curves into arcs and segments using tangential cover with a fixed width value of MBS and tangent space representation of the approximated polygon of dominant points. Such an approach is not adapted to noisy curves due to the fixed width, in particular when the noise is not uniformly distributed. Furthermore, the width value is manually adjusted. Inspired by this study, we propose a suitable approach for noisy curves by using the ATC which is a free-parameter algorithm.

3.1 Dominant point detection

Dominant points are significant points on a curve with local maximum curvature. Such points contain a rich information which allows to characterize and describe the curve.

Issued from the dominant point detection proposed in [12, 13] and the notion of ATC, a sequential algorithm is developed in [5] to determine the dominant points of a given noisy curve C. The main idea is that candidate dominant points are localized in the common zones of successive MBS of the ATC of

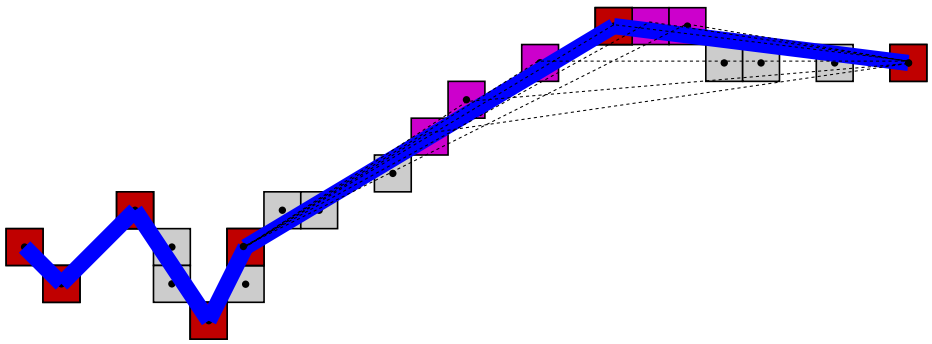


Fig. 4. Dominant points (in red) with the ATC in *Fig.* 3 are detected as point having the smallest angle measure in the common zone.

C. In particular, this step can be done by simply verifying the beginning and ending indexes of the MBS constituting the common zone (see *Fig.* 3). Once the candidates are found, an angle measure is used to determine the dominant points with local extreme curvature in the common zone. More precisely, this measure is estimated as the angle between the considered point and the two left and right endpoints of the left and right MBS involved in the studied common zone, and dominant point is identified as the point having a local minimum measure of angle. This is illustrated in *Fig.* 4.

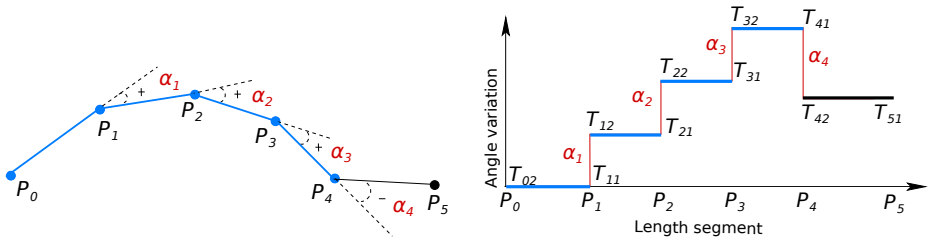
3.2 Tangent space representation (or (l,α) plane [4])

Let $P = \{P_i\}_{i=0}^{m}$ be a polygon, l_i length of segment P_iP_{i+1} and $\alpha_i = \angle(\overrightarrow{\mathrm{P_{i-1}P_i}}, \overrightarrow{\mathrm{P_iP_{i+1}}})$ such that $\alpha_i > 0$ if P_{i+1} is on the right side of $\overrightarrow{\mathrm{P_{i-1}P_i}}$ and $\alpha_i < 0$ otherwise. A tangent space representation $T(P)$ of P is a step function which is constituted of segments $T_{i2}T_{(i+1)1}$ and $T_{(i+1)1}T_{(i+1)2}$ for $0 \leq i < m$ with
$T_{02} = (0, 0)$,
$T_{i1} = (T_{(i-1)2}.x + l_{i-1}, T_{(i-1)2}.y)$ for $1 \leq i \leq m$,
$T_{i2} = (T_{i1}.x, T_{i1}.y + \alpha_i)$, $1 \leq i \leq (m-1)$.

In other words, the difference between two ordinates of an adjacent step in the tangent space represents the turn angle α_i of the corresponding pair of line segments $\overrightarrow{\mathrm{P_{i-1}P_i}}$ and $\overrightarrow{\mathrm{P_iP_{i+1}}}$. The difference between two abscissas of two endpoints of a step in the tangent space represents the length l_i of the corresponding line segments P_iP_{i+1} (see *Fig.* 5).

In [1, 6], the following result is obtained for the relationship between a set of sequential chords of an arc and the tangent space representation of the set (see *Fig.* 6).

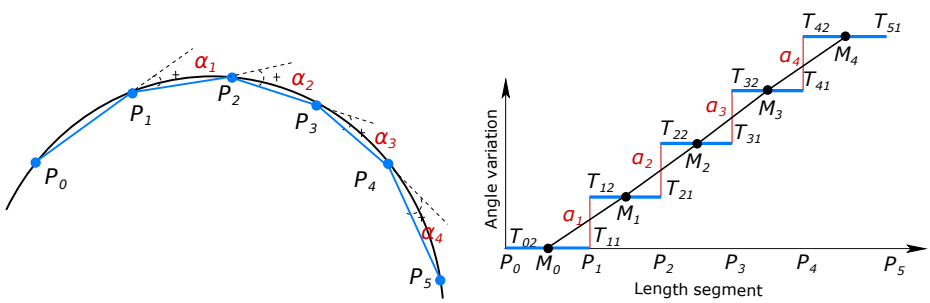
Proposition 1 ([1, 6]). *Let $P = \{P_i\}_{i=0}^{m}$ be a polygon, $l_i = \mid \overrightarrow{\mathrm{P_iP_{i+1}}} \mid$, $\alpha_i = \angle(\overrightarrow{\mathrm{P_{i-1}P_i}}, \overrightarrow{\mathrm{P_iP_{i+1}}})$ and $\alpha \leqslant \frac{\pi}{4}$ such that $\alpha_i \leqslant \alpha$ for $0 \leq i < n$. Let $T(P)$ be the tangent space representation of P and $T(P)$ constitutes of segments $T_{i2}T_{(i+1)1}$,*



(a) Input polygon (b) Tangent space of (a)

Fig. 5. Tangent space representation of a straight segment succession.

$T_{(i+1)1}T_{(i+1)2}$ *for* $0 \leq i < m$, $M = \{M_i\}_{i=0}^{m-1}$ *the midpoint set of* $\{T_{i2}T_{(i+1)1}\}_{i=0}^{m-1}$. P *is a polygon whose vertices are on an arc only if* M *is a set of quasi collinear points*[3]



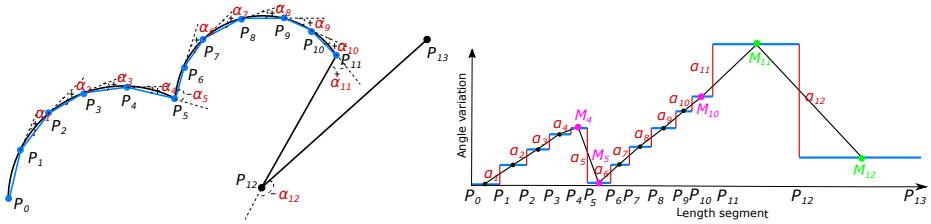
(a) Set of sequential chords of an arc (b) Midpoint curve of (a)

Fig. 6. Tangent space and the curve of midpoints of an arc. If the midpoints is quasi collinear, then the initial points belong to a circular arc [1, 6].

In other words, the arc detection of P becomes the problem of verifying the quasi collinearity of midpoints in tangent space representation of P. In particular, this quasi collinearity can be done with the algorithm of MBS of width ν with midpoints in the tangent space (see [1, 6] for the details).

Particularly, in the midpoint curve M of $T(P)$, a midpoint is said an **isolated point** if the difference of ordinate values between it and one of the two neighboring midpoints is higher than a threshold α, the point corresponds to a junction of two primitives. If this condition of isolated point is satisfied with all two neighbors, the point is called a **full isolated point** and corresponds a segment in P (see *Fig.* 7).

[3] The points are said *quasi collinear* if they belong to a small width strip bounded by two real parallel lines.



(a) Input polygon (b) Midpoint curve of (a)

Fig. 7. Classification of midpoints in tangent space. Pink (resp. green) points are isolated points (resp. full isolated points) and correspond to junctions of two primitives (resp. segments), while black points are points of arcs.

3.3 Proposed algorithm

We now describe the method to decompose a discrete curve into arcs and segments. The method is divided into three steps (see *Fig.* 8) and given in *Alg.* 1:

1. Detection of dominant points with ATC decomposition of a noisy discrete curve.
2. Tangent space representation of polygon issued from dominant points detected.
3. Computation of arcs and segments of the considered curve from the tangent space analysis and the discrete points of the curve.

In short, the decomposition process is realized by analyzing the tangent space representation of the polygon constructed from detected dominant points of the input curve. Then it outputs the sets of segments and arcs decomposed of the curve.

In *Alg.* 1, Line 9 (resp. 16) is to calculate the real segments (resp. arcs) corresponding to points in the set SEGs (resp. ARCs). For the continuity of the decomposition between the primitives, we consider the endpoints as dominant points. Thus, each segment in SEGs is simply the line segment passing through the corresponding endpoints C_{b_i} and C_{e_i} –which are also the dominant points. While, each arc in ARCs is computed as the arc passing through the two endpoints of $pARC$ and best-fitting to points in $pARC$ when the squared distance is used. Due to a high angular deviation near the endpoints of an arc, we verify the best-fitting arc with one point in the central one-third portion of $pARC$ and the two endpoints of $pARC$. Note that this is improved w.r.t. the original algorithm proposed in [1, 6] which simply pick-up a middle point without any optimization of arcs detected.

Furthermore, it is obvious that any two midpoints in the tangent space belong to a segment, thus a MBS. As a consequence, we can always determine an arc associated to these midpoints. In order to obtain an accurate segmentation and a better description of curve with arcs and segments in such cases, we consider an error criterion, namely integral sum of square errors, for the considered arc

Algorithm 1 Curve Decomposition into Arcs and Segments

Require: $C = (C_i)_{0 \leq i \leq n-1}$ a discrete curve of n points
ν width of MBS for collinear test in the tangent space
α threshold of admissible angle in the tangent space
Ensure: $ARCs$ and $SEGs$ sets of arcs and segments of C

1: **Begin**
2: $ARCs \leftarrow \emptyset, SEGs \leftarrow \emptyset, pARCs \leftarrow \emptyset, MBS_\nu \leftarrow \emptyset$
3: Detect the dominant point D of C (see *Sec.* 3.1)
4: Transform D into the tangent space $T(D)$ (see *Sec.* 3.2)
5: Construct the midpoint curve $\{M_i\}_{i=0}^{m-1}$ of $T(D)$ (see *Sec.* 3.2)
6: **for** $i \leftarrow 1$ to $m-2$ **do**
7: $C_{b_i}C_{e_i}$ is the part of C corresponding to M_i
8: **if** $(\mid M_i.y - M_{i-1}.y \mid > \alpha)\&(\mid M_i.y - M_{i+1}.y \mid > \alpha)$ **then**
9: $SEGs \leftarrow SEGs \cup \{$find a segment from $C_{b_i}C_{e_i}\}$
10: $MBS_\nu \leftarrow \emptyset$
11: **else**
12: **if** $MBS_\nu \cup \{M_i\}$ is a MBS of width ν **then**
13: $MBS_\nu \leftarrow MBS_\nu \cup \{M_i\}$
14: $pARC \leftarrow pARC \cup \{C_{b_i}C_{e_i}\}$
15: **else**
16: $ARCs \leftarrow ARCs \cup \{$find an arc from $pARC\}$
17: $pARC \leftarrow \emptyset$
18: **end if**
19: **end if**
20: **end for**
21: **End**

approximation. More precisely, we compute the error value for the part of curve approximated by an arc and this approximated by segments, if the error by the arc is smaller than this by segments then the arc is kept in the decomposition, and the segments otherwise.

Alg. 1 has a $O(n \log n + nm)$ complexity where n is the number of points of C and m is the number of midpoints. Indeed, the ATC and dominant point detection can be performed in $O(n \log n)$ [5] and the tangent space representation of dominant points as well as the recognition of MBS of the midpoint curve are done in $O(m)$ [14]. The loop iterates over the midpoints to find the corresponding segments and arcs (Lines 6–20) is performed in $O(nm)$. More precisely, Line 9, for finding a segment, is in $O(1)$ since the segment is determined by the extremities $C_{b_i}C_{e_i}$. For finding a best-fitting arc associated to $pArc$ of the curve in Line 16, we have $O(1)$ to compute an arc passing through three points and $O(\mid pArc \mid)$ to compute the fitting error, and $\mid pArc \mid = n/3$ in the worst case. Finally, the complexity of the algorithm is $O(n \log n + nm)$.

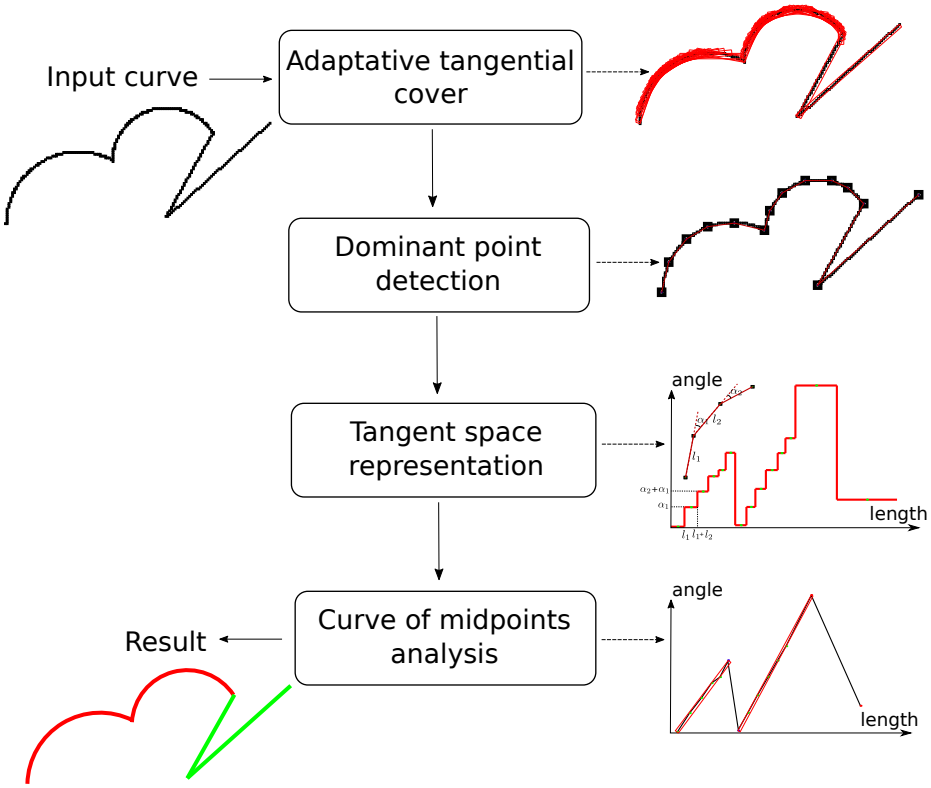


Fig. 8. Flowchart of the proposed algorithm.

4 Experimental results

We now present some experiments to illustrate the efficiency of *Alg.* 1 to decompose discrete curves into arcs and segments. To this end, we first show the decomposed results in case of noisy data. Then, we provide some comparisons to Rosin [15] and Nguyen [1] methods. Overall the experiments, the arcs and segments are colored in red and green respectively.

From *Prop.* 1, we can fix the admissible angle in the tangent space $\alpha = \frac{\pi}{4}$. While the width ν for the quasi collinear test of MBS, it is mentioned in [1] that this value relies on the approximation error $sin\alpha_i \simeq \alpha_i$ in the tangent space and it is bounded by 0.1. Therefore, we have $\nu = 0.1$.

4.1 Experimentation on noisy curves

In order to test the robustness of the proposed algorithm towards noise, we consider Gaussian noise and a statistical noise model similar to the Kanungo noise [16], in which the probability P_d of changing the pixel located at a distance

d from the shape boundary is defined as $P_d = \beta^d$ with $0 < \beta < 1$. *Fig.* 10 shows curves obtained with different noise levels for $\beta = 0.3, 0.5, 0.7$.

Comparing to the algorithm proposed in [1], tangential covers with fixed width value are applied for arcs and segments detection. Since the noise can appear randomly along the contour, this method is inadequate when contours present different noise levels. We observe in *Figs.* 9–10 that the ATC allows a better model of tangent cover and thus more relevant to noise. Still in *Figs.* 9–10, the experimental results of the proposed algorithm are shown. It can be seen that good reconstructions with arcs and segments are obtained even with an important and non-uniform noise on digital contours.

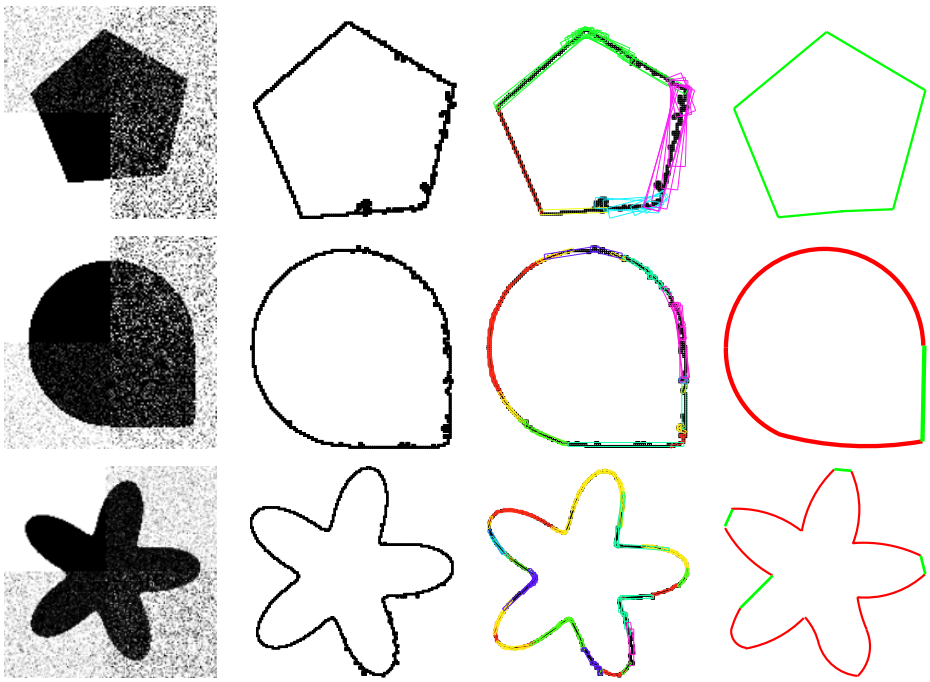


Fig. 9. Arcs and segments reconstruction of noisy curves. Left: input with Gaussian noise, middle left: extracted curves, middle right: ATC, right: decomposition results.

4.2 Comparison results with other methods

We now present some comparisons of our algorithm with two algorithms proposed by Rosin [15] and Nguyen [1]. The experiments are carried out on noisy curves, technical and real images and shown in *Figs.* 11–13. The modifications of the decomposition algorithm allows to improve the quality of the curves reconstructed with arcs and segments w.r.t. Nguyen's algorithm [1, 6] and particularly for noisy curves.

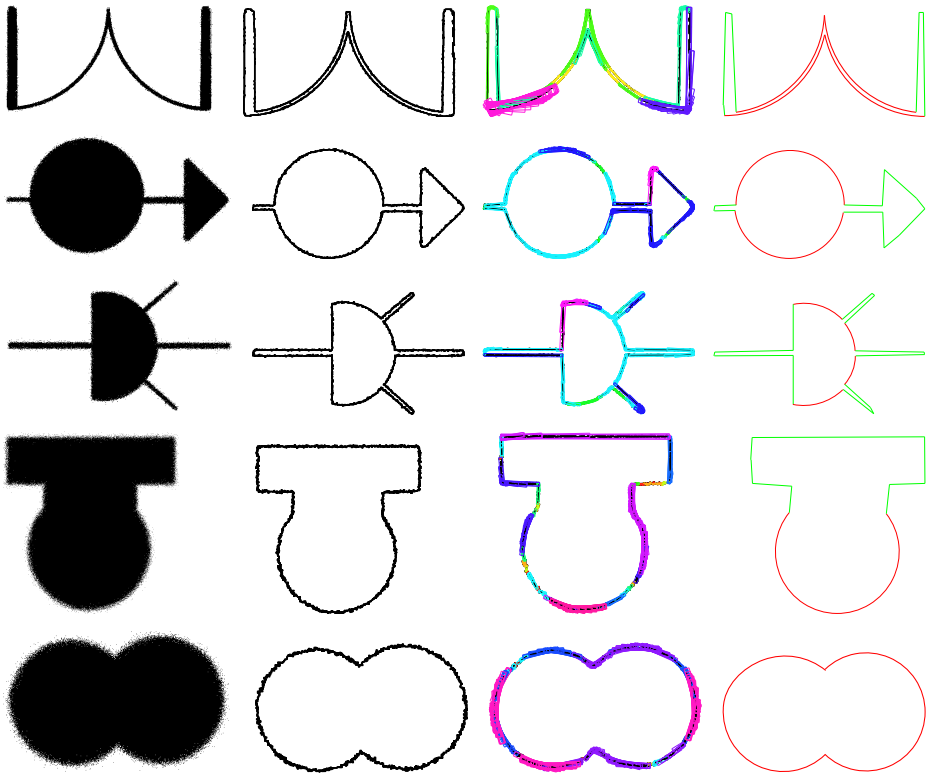


Fig. 10. Arcs and segments reconstruction of noisy curves. Left: input with Kanungo noise for $\beta = 0.3$ (first image), 0.5 (second and third images) and 0.5 (fourth and fifth images), middle left: extracted curves, middle right: ATC, right: decomposition results.

5 Conclusion

We present in this paper a discrete approach for decomposing a noisy discrete curve into arcs and segments using the notion of adaptive tangential cover deduced from the meaningful thickness. The experiments demonstrate the efficiency of the proposed approach even with an important and non-uniform noise level on input contours. An online demonstration based on the DGtal [17] and ImaGene [18] library, is available at the following website :
`http://ipol-geometry.loria.fr/~phuc/ipol_demo/ATC_ArcSegDecom_IPOLDemo`

In the paper, we consider only two primitives arcs and segments for the approximating/description of input curves. In the further work, we may use other primitives such as eclipses, parabolas, *ect.* for image processing applications. Another perspective is to extend the framework into 3D.

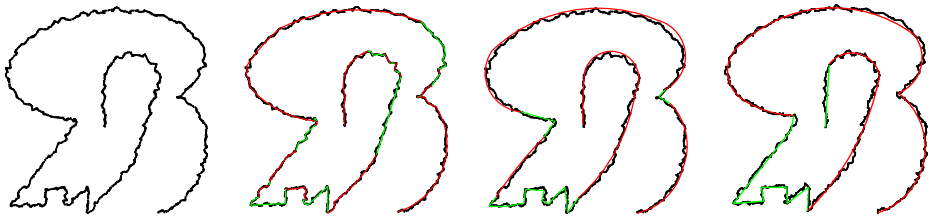


Fig. 11. Comparison result on noisy curve. Left: input curve, middle left: results with [15], middle right: results with [1], right: results with our method.

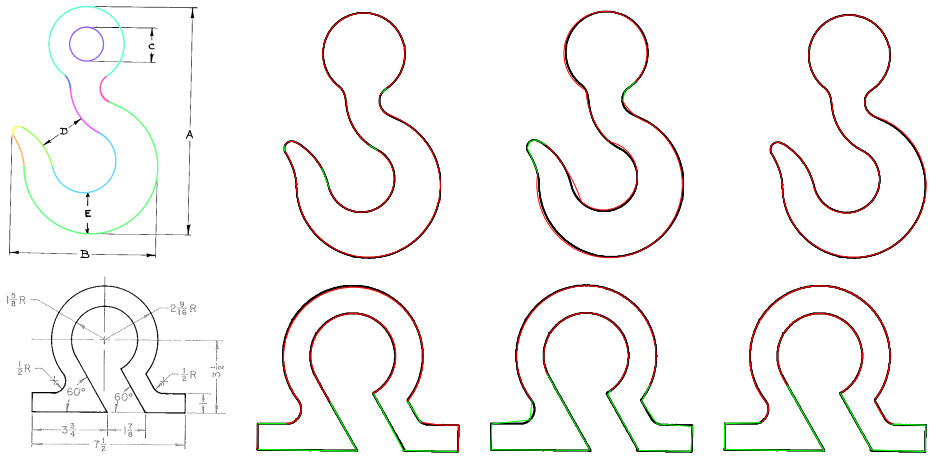


Fig. 12. Comparison result on technical images. Left: input images, middle left: results with [15], middle right: results with [1], right: results with our method.

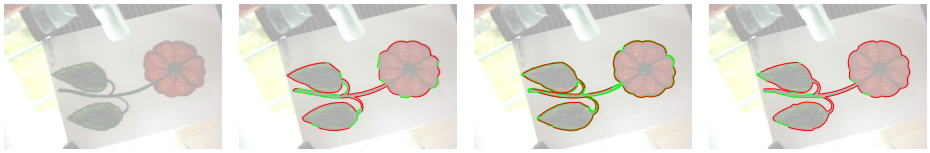


Fig. 13. Comparison result on real image. Left: input image, middle left: results with [15], middle right: results with [1], right: results with our method.

References

1. Nguyen, T.P., Debled-Rennesson, I.: Decomposition of a curve into arcs and line segments based on dominant point detection. In: SCIA '11. (2011) 794–805
2. Akinlar, C., Topal, C.: Edcircles: A real-time circle detector with a false detection control. Pattern Recognition **46** (2013) 725–740
3. Arkin, E.M., Chew, L.P., Huttenlocher, D.P., Kedem, K., Mitchell, J.S.B.: An efficiently computable metric for comparing polygonal shapes. In: Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '90 (1990) 129–137

4. Latecki, L.J., Lakämper, R.: Shape similarity measure based on correspondence of visual parts. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 1185–1190
5. Ngo, P., Nasser, H., Debled-Rennesson, I., Kerautret, B.: Adaptive tangential cover for noisy digital contours. In: DGCI '16. Volume 9647 of LNCS. (2016) 439–451
6. Nguyen, T.P., Debled-Rennesson, I.: Arc segmentation in linear time. In: CAIP '10. Volume 6854 of LNCS. (2010) 84–92
7. Debled-Rennesson, I., Feschet, F., Rouyer-Degli, J.: Optimal blurred segments decomposition of noisy shapes in linear time. Computers & Graphics **30** (2006) 30–36
8. Reveillès, J.P.: Géométrie discrète, calculs en nombre entiers et algorithmique (1991) Thèse d'état. Université Louis Pasteur, Strasbourg.
9. Lachaud, J.: Digital shape analysis with maximal segments. In: Applications of Discrete Geometry and Mathematical Morphology - First International Workshop, WADGMM 2010, Istanbul, Turkey. (2010) 14–27
10. Faure, A., Buzer, L., Feschet, F.: Tangential cover for thick digital curves. Pattern Recognition **42** (2009) 2279–2287
11. Kerautret, B., Lachaud, J.O.: Meaningful Scales Detection: an Unsupervised Noise Detection Algorithm for Digital Contours. Image Proc. On Line **4** (2014) 98–115
12. Nguyen, T.P., Debled-Rennesson, I.: A discrete geometry approach for dominant point detection. Pattern Recognition **44** (2011) 32–44
13. Ngo, P., Nasser, H., Debled-Rennesson, I.: Efficient dominant point detection based on discrete curve structure. In: IWCIA '15. Volume 9448 of LNCS. (2015) 143–156
14. Feschet, F., Tougne, L.: Optimal time computation of the tangent of a discrete curve: Application to the curvature. In: DGCI. Volume 1568 of LNCS. (1999) 31–40
15. Rosin, P.L., Wesst, G.A.W.: Segmentation of edges into lines and arcs. Image and Vision Computing **7** (1989) 109–114
16. Kanungo, T., Haralick, R.M., Stuezle, W., Baird, H.S., Madigan, D.: A statistical, nonparametric methodology for document degradation model validation. IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 1209–1223
17. : (DGtal: Digital Geometry tools and algorithms library) http://libdgtal.org.
18. : (Imagene, generic digital image library) http://gforge.liris.cnrs.frs/projects/imagene.