**Research Highlights (Required)**

- A new approach for generation of polygonal approximations based on the convex hull of contour is proposed.
- The proposed algorithm takes into account the symmetry of the contour.
- A final improvement process is applied to increase the quality of the polygonal approximation.
- The new algorithm is non-optimal but unsupervised (automatic), because no parameters have to be set or tuned.
- Experiments using a public available dataset show that the new proposal outperforms other unsupervised algorithms.

ELSEVIER

Pattern Recognition Letters
journal homepage: www.elsevier.com

# Unsupervised generation of polygonal approximations based on the convex hull

Nicolás Luis Fernández García[a,b,**], Luis Del-Moral Martínez[a], Ángel Carmona Poyato[a,b], Francisco José Madrid Cuevas[a,b], Rafael Medina Carnicer[a,b]

[a]*Computing and Numerical Analysis Department, Edificio Albert Einstein, Córdoba University, Campus de Rabanales, Córdoba 14071, Spain*
[b]*Maimonides Institute for Research in Biomedicine (IMIBIC), Avenida Menéndez Pidal s/n, Córdoba 14004, Spain*

ABSTRACT

The present paper proposes a new non-optimal but unsupervised algorithm, called *ICT-RDP*, for generation of polygonal approximations based on the convex hull. Firstly, the new algorithm takes into account the convex hull of the 2D closed curves or contours to select a set of initial points; secondly, the significance levels of the contour points are computed using a symmetric version of the well-known Ramer, Douglas-Peucker algorithm; and, finally, a thresholding process is applied to obtain the vertices or dominant points of the polygonal approximation. Since the convex hull can select many initial points in rounded parts of the contour, an additional deletion process is required to remove quasi-collinear dominant points. Furthermore, an additional improvement process is applied to shift the dominant points in order to increase the quality of the polygonal approximation. Experiments performed on a public available dataset show that the new proposal outperforms other unsupervised algorithms for generation of polygonal approximations.



## 1. Introduction

Shape representation by polygonal approximation is extensively used for constructing a characteristic description of a boundary in the form of a series of straight lines. This representation is very popular due to its simplicity, locality, generality and compactness (Loncaric, 1998; Melkman and O'Rourke, 1998; Zhang and Lu, 2004). In a closed digital planar curve, most of the information is located at points of high curvature (Attneave, 1954), which are used to obtain polygonal approximations. These points are known as *dominant points* and are an important target in many machine vision applications (Wu, 2003a).

The present work proposes a new approach, called *ICT-RDP*, for generation of polygonal approximations of 2D closed curves or contours. The new proposal is based on the convex hull and it is a new version of a previous algorithm (Fernández-García et al., 2016), which uses a *symmetric* version of the well-known *RDP* algorithm proposed by Ramer (1972), Douglas and Peucker (1973), and applies an adaptive thresholding method to obtain the dominant points. The new proposal also applies a deletion process, to remove *quasi-collinear* dominant points, and a final improvement step, based on the optimization method proposed by Masood (2008b), to improve the quality of the polygonal approximation.

The present paper is arranged as follows. Section 2 describes the related work. Sections 3 and 4 explain the background and the new proposal, respectively. The experiments and results are detailed in section 5. Finally, the main conclussions and future work are summarized in Section 6.

## 2. Related work

### 2.1. Generation of polygonal approximations

Many techniques for generation of polygonal approximations have been proposed and can be classified according to different criteria, such as: (1) optimal or non-optimal and (2) supervised or unsupervised algorithms.

The optimal algorithms are based on an optimization criterion (Aguilera-Aguilera et al., 2015; Kolesnikov and Fränti, 2007; Perez and Vidal, 1994; Pikaz and Dinstein, 1995; Salotti, 2001, 2002) but have two main drawbacks (Horng and Li, 2002): the optimum depends on the applied criterion and

**Corresponding author: Tel.: +34-957-21-83-47; fax: +34 958 -21-83-60; *e-mail:* `ma1fegan@uco.es` (Nicolás Luis Fernández García)

requires a very high time complexity, which is not suitable for real time applications. Otherwise, the non-optimal (or suboptimal) algorithms do not guarantee any kind of optimum, but can find reasonable polygonal approximations for real time applications.

The supervised algorithms take into account one or more parameters to generate the polygonal approximations and, therefore, this is its main drawback, because these parameters must be tuned (Ataer-Cansizoglu et al., 2013; Carmona-Poyato et al., 2010; Kalaivani and Ray, 2019; Lowe, 1987; Masood and Haq, 2007; Masood, 2008a,b). On the other hand, the unsupervised (or automatic) algorithms generate the polygonal approximations without using any kind of parameters (Fernández-García et al., 2016; Madrid-Cuevas et al., 2016; Marji and Siy, 2004; Prasad et al., 2012; Wu, 2003a).

Combining the optimal and supervised approaches, the generation problem of polygonal approximations of 2D closed curves or *contours* can be formulated in two ways (Kolesnikov and Fränti, 2007): *minimum-distortion problem* or *minimum-rate problem*. The algorithms based on the minimum-distortion problem or $Min-\epsilon$ *problem* consider a predefined number $d$ of vertices and try to generate the *optimal* polygonal approximation with $d$ vertices or dominant points (*DP*) so that the adjustment error from the contour is minimal among all the approximations with $d$ vertices (Aguilera-Aguilera et al., 2015; Perez and Vidal, 1994; Salotti, 2001). Other algorithms are focused on the minimum-rate problem or $Min-\#$ *problem*, where a predefined error measure $\epsilon$ is set and try to generate the polygonal approximation, with the minimal number of vertices, so that its adjustment error from the contour is less than $\epsilon$ (Salotti, 2002).

In this paper, a non-optimal but unsupervised algorithm is proposed for generation of polygonal approximations suitable for real-time applications.

### 2.2. Quality measures of polygonal approximations

The quality of a polygonal approximation can be assessed by two approaches: subjective or objective. In the subjective approach, a human observer visually compares the original contour with the polygonal approximation. This approach is easy to apply, but cannot be automated and depends on the criterion of the observer. On the other hand, two criteria must be taken into account in the objective approach: (1) the *number of points* of the polygonal approximation and (2) its *adjustment error* to the contour. The objective approach can be automated but has a main drawback: the criteria on which it is based are opposed to each other. In general, if the number of points is reduced, then the adjustment error increases; otherwise, if the number of points is increased, then the adjustment error decreases. Because of this, the aim of the objective evaluation must be to achieve the best *tradeoff* between the number of points and the adjustment error of the polygonal approximation.

The measurement to evaluate the number of points of the polygonal approximation is *Compression ratio*, defined as $CR = \frac{n}{d}$, where $n$ is the number of contour points and $d$ is the number of points of the polygonal approximation or *dominant points* (*DP*). If the number of dominant points decreases, the *compression ratio* increases, and vice versa. On the other hand, different measurements have been proposed to evaluate the adjustment error, such as: (a) *Integral square error*: $ISE = \sum_{i=1}^{n} e_i^2$, where $e_i$ is the distance of the contour point $P_i$ to the polygonal approximation; or (b) *Maximum error*: $E_\infty = \max_{i\in\{1,\cdots,n\}}\{e_i\}$. *ISE* may not take into account some relevant information of the contour when hides large deviation at particular point due to closeness of approximating polygon at other parts of the curve. $E_\infty$ is proposed to solve this lack of accuracy of *ISE* (Masood and Haq, 2007).

The *Figure of Merit* $FOM = \frac{CR}{ISE}$ was proposed by Sarkar (1993) to make the *tradeoff* between the compression ratio (*CR*) and the total distortion (*ISE*) caused (Masood, 2008a). However, the terms *CR* and *ISE* used by $FOM$ are not balanced (Rosin, 1997), causing the measure to be biased towards approximations with lower *ISE* and many dominant points. Hence, $FOM$ is not the best measure for comparing contours with different numbers of dominant points. The *weighted sum of squared error* $WE = \frac{ISE}{CR}$ is defined as the inverse of $FOM$ (Wu, 2003a,b). The *weighted maximum error* $WE_\infty = \frac{E_\infty}{CR}$ has also been proposed (Lowe, 1987; Masood and Haq, 2007; Wu, 2003a). Technically, $WE$ and $WE_\infty$ are similar to $FOM$ and suffer similar problems (Masood and Haq, 2007). A parameterized version $WE_n = \frac{ISE}{(CR)^n}$ has also been used (Carmona-Poyato et al., 2010; Marji and Siy, 2004) to *tradeoff* the contribution of *ISE* and *CR*, where $n = 1, 2, 3$. From our point of view, $WE_2$ is the fairest measurement that makes the best *tradeoff* between compression ratio and adjustment error, because $WE_1$ and $WE_3$ favor polygonal approximation with many or few dominant points, respectively (Carmona-Poyato et al., 2011);

A new measure was proposed by Rosin (1997) to avoid the drawback of $FOM$: *Merit* = $\sqrt{Fidelity \times Efficiency}$, where *Fidelity* = $\frac{E_{opt}}{E_{appr}} \times 100$ and *Efficiency* = $\frac{N_{opt}}{N_{appr}} \times 100$ where $E_{appr}$ and $N_{appr}$ are the error and the number of dominant points of the suboptimal polygonal approximation, $E_{opt}$ is the error produced by the optimal algorithm with the same number of dominant points and $N_{opt}$ represents the number of dominant points that would require an optimal algorithm to produce the same error (Perez and Vidal, 1994; Salotti, 2001). The *Fidelity* measures how well the polygon obtained by the algorithm to be tested fits the curve relative to the optimum polygon, in terms of the approximation error. The *Efficiency* measures the compactness of the polygon obtained by the algorithm to be tested, relative to the optimum polygon that incurs the same error (Rosin, 1997).

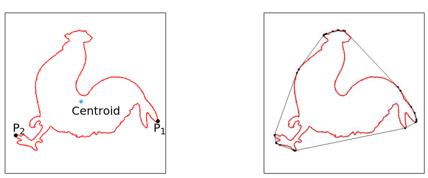
The Rosin's measurement can compare results of different algorithms with different number of dominant points. Nevertheless, Masood (2008a) pointed out that this measurement also suffers a few weaknesses. The polygon that consists of just *break points* will produce *Fidelity* = 100, *Efficiency* = 100 and *Merit* = 100. It means that the set of break points taken as dominant points will produce a perfect approximation, whereas this type of approximation is of no practical use since its compression ratio is very low. Carmona-Poyato et al. (2011) proposed a new measure for assessing polygonal approximation of curves that uses the optimal algorithm of Perez and Vidal (1994) and the optimization of an objective function based on $WE_2$. Both measurements have a main drawback: they need optimal solutions which are computationally very expensive.

## 3. Background

### 3.1. Automatic generation of polygonal approximations based on a thresholding approach

The present paper describes a new version of a previous non-optimal algorithm for unsupervised or automatic generation of polygonal approximations based on a thresholding approach, which will be referred to as $T-RDP$ for short (Fernández-García et al., 2016). $T-RDP$ consists of six steps:

1. *Selection of the initial points*: some special contour points are chosen to be considered as *initial points* (Figure 1-a): the farthest point(s) from the centroid is (are) chosen as *initial point(s)*; besides, the farthest point(s) from the previous one(s) is (are) also considered as *initial point(s)*. This method is independent of the starting point, invariant to rotations and scales and takes into account the symmetry of the contour.




(a) 2 initial points (b) 27 initial points

**Fig. 1. First step.- Selection of initial points for the contour *Chicken-5*: (a) original method: $P_1$ is the farthest point from the centroid and $P_2$ is the farthest point from $P_1$ ; (b) new method: convex hull.**

2. *Computation of the significance values of the non-initial points*: a new version of the well-known RDP algorithm proposed by Ramer (1972) and Douglas and Peucker (1973), which takes into account the symmetry of the contour, is used to compute the significance values of the non-initial points. The significance value of every *non-initial* point is its deviation error $d$ computed using a recursive process. At the beginning, if $P$ and $P'$ are two contiguous *initial points*, then the farthest *non-initial* point $Q$ located between $P$ and $P'$ is chosen and considered as *candidate point*. The significance value of the point $Q$ is its deviation error $d$, that is, the distance from $Q$ to the segment defined by the points $P$ and $P'$. This process continues recursively with the points $P$ and $Q$, and the points $Q$ and $P'$, respectively.
3. *Computation of the significance values of the initial points*: the significance values of the *initial points* must be greater than the significance values of the other contour points. Initially, the *maximum* of the significance value of the *non-initial* points is computed: $Sig_{max}$. If $Sig_{max}$ is equal to 0.0, then the significance value of the *initial* points is 1.0. This situation occurs with artificial contours in which the initial points define a perfect polygonal approximation. Otherwise, if $Sig_{max}$ is not equal to 0.0, then the significance value of the *initial* points is the maximum between $Sig_{max}$ and $d$, where $d$ is the maximum distance from the centroid to the contour points.

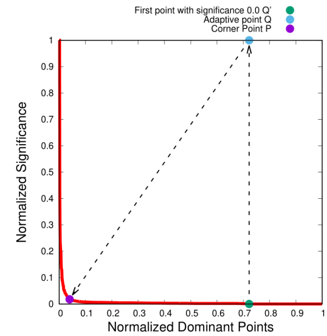



**Fig. 2. Fifth step: adaptive thresholding method applied to the normalized significance curve; the threshold $h$ is the ordinate $y$ of the farthest point $P$ from the Adaptive point $Q$.**

4. *Generation of the normalized significance curve*: this curve is a plot of polygonal approximation error defined as a function of the number of dominant points and must be normalized in order to facilitate the search for the dominant points using a thresholding method (Figure 2).
5. *Search for the threshold of the normalized significance values*: the normalized significance curve is used by a *thresholding* method to search for the threshold of the normalized significance values. Four thresholding algorithms were proposed and the *adaptive* method obtained the best results. The *adaptive* method searchs for the *farthest* point $P(x,y)$, in the region of interest of the significance curve, to the *adaptive point* $Q(x_0, 1.0)$, where the project point $Q'(x_0, 0.0)$ is the first point in ascending order with normalized significance value equal to 0.0 ( Figure 2).
6. *Thresholding of the normalized significance values to obtain the dominant points*: the normalized significance values of the contour points are thresholded to obtain the dominant points of the contour. The points of the contour with a normalized significance value greater than the threshold, or equal to 1.0 for artificial curves, are chosen as *dominant points*.

The experiments demonstrated that this previous method has a good performance for generating polygonal approximations of 2D closed contours, without requiring any parameter to be tuned. In addition, the time complexity is $O(n \log(n))$ in the best case and “quadratic in $n$ in the worst case, where $n$ is the number of points of the given polygonal contour” (Hershberger and Snoeyink, 1992).

### 3.2. Convex hull

The new proposal modifies the first step of the previous method $T-RDP$ (Fernández-García et al., 2016) and makes use of the *convex hull* of the contour to obtain the *initial points* of the polygonal approximation. The convex hull of a set X of points in the Euclidean plane (or space) is the smallest convex set that contains X (see Figure 1-b). There are many implementations of the convex hull (Chadnov and Skvortsov, 2004). The convex hull is not used to represent the whole contour, but it is used to provide a set of *initial points* that improves the performance of the algorithm to obtain better polygonal approximations.

*3.3. Masood's optimization*

The new approach includes a final step to improve the adjustment of the polygonal approximation to the contour, which is based on the optimization process proposed by Masood (2008b). Masood took into account that some dominant points may be moved to a new position in order to decrease the overall error (*ISE*) of the resulting approximation, as it is shared among all the dominant points ($DPs$).

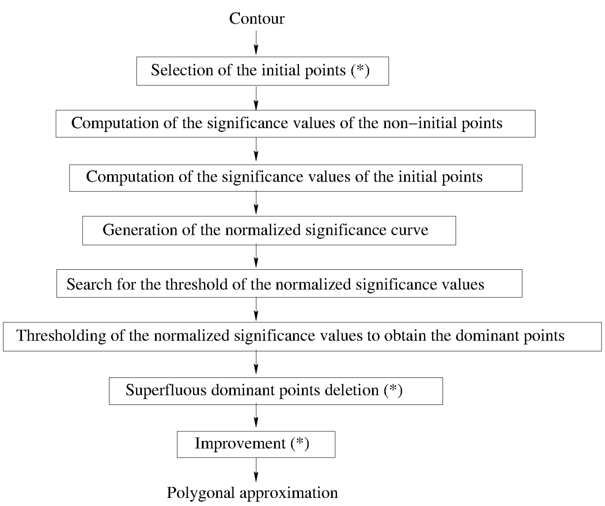



**Fig. 3. New proposal: $ICT - RDP$. The steps marked with (*) were modified or added regarding the original method.**

## 4. New proposal

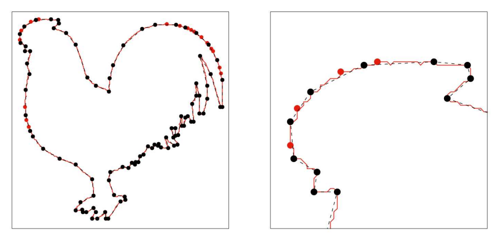
The new proposal is called $ICT - RDP$ because improves the previous algorithm $T - RDP$ (see Section 3.1 and Fernández-García et al. (2016)) by modifying the first step to select the initial points, using the convex hull (see Section 4.1), and including two final steps to eliminate the superfluous dominant points (see Section 4.2) and to improve the adjustment of the polygonal approximation to the contour (see Section 4.3), respectively (Figure 3).

*4.1. First step. Selection of the initial points*

The first step calculates the *convex hull* of the contour in order to use its vertices as the initial points for the rest of steps of the new proposal (see Figure 1-b). The aim is to compute a superset of points that can be considered as containing dominant points; that is, points that are more likely to belong to the polygonal approximation. The selection of the initial points using the convex hull is invariant to rotations or scale changes, does not depend on the parametrization of the contour points, and also takes into account the symmetry of the contour. The *Quickhull* implementation was used, being its expected time complexity of $O(n\log(n))$ (Eddy, 1977; Bykat, 1978). However, if the points are ordered by any of the coordinates or by an angle with respect a fixed vector, then its time complexity is $O(n)$ (Andrew, 1979; Graham, 1983).

The aim of the convex hull is not intended to represent the entire contour, but to provide a more reliable and complete superset of *initial points* than the previous selection method used in the original algorithm $T - RDP$ (see Figure 1). This new superset of *initial points* allows to improve the performance of the rest of the steps of the new proposal $ICT - RPD$ to generate better polygonal approximations as will be shown in the experiments (see Section 5 and Table 1).

Unlike the selection method of initial points of the original algorithm $T - RDP$ (see Section 3.1 and Figure 1-a), the new proposed selection method of initial points, based on the convex hull, is likely to generate many more points than the previous one; for instance: if the contour has rounded parts, the convex hull provides too many initial points, which could lead to generating a large amount of dominant points in the final polygonal approximation. To mitigate this effect, an aditional seventh step must be included in order to eliminate the *quasi-collinear* points introduced due to this application (see Section 4.2 and Figure 4).



(a) (b)

**Fig. 4. Seventh step. Superflous dominant points deletion on the contour *Chicken-2*: (a) polygonal approximation with superfluous dominant points in red; (b) zoom on (a). The quasi-collinear points, marked in red, will be deleted.**

*4.2. Seventh step. Superflous dominant point deletion*

The seventh step deletes the *superfluous or quasi-collinear* dominant points that may appear in rounded parts of the contour due to the computation of the *convex hull* in the first step (see Section 4.1). Many of these points are almost aligned and, therefore, can be eliminated (Figure 4). The deletion process is shown in Algorithm 1. The dominant point $P$ will be deleted if the deviation error ($Emax$ or $ISE$) of the side defined by the points $P'$ and $P''$ is lower than the deviation errors of the other two sides defined by the points $P'$ and $P$, and by $P$ and $P''$. The supression of the point $P$ improves the polygonal adjustment to the contour. This process is repeated until it is impossible to eliminate any point or until the final approximation has only three vertices (triangle). The remaining points will be the vertices of the polygonal approximation. The expected time complexity of this step is linear in $n$, because the number of dominant points ($d$) is much smaller than the number of the contour points ($n$).

**Algorithm 1** Seventh step. Superflous dominant point deletion

{Let $L$ be the set of dominant points}
$L \leftarrow \{P_1, \ldots, P_d\}$
{Let $P$ be the first point of $L$}
$P \leftarrow P_1$
**repeat**
  Let $P'$ and $P''$ be the current contiguous dominant points of $P$
  {Minimization criterion for deletion of $P$}
  **if** $((Emax(P', P'') < Emax(P', P))$
  $\wedge (Emax(P', P'') < Emax(P, P''))$
  or $(ISE(P', P'') < ISE(P', P))$
  $\wedge (ISE(P', P'') < ISE(P, P''))$ **then**
    $L \leftarrow L - \{P\}$
  **end if**
  Let $P$ be the next current dominant point of $L$
**until** (No point is deleted or $L$ has only three points.)

where $d$ is the number of dominant points, $Emax$ is the *maximum deviation error* and $ISE$ is the *integral square error*, defined as $Emax(P_1, P_2) = max\{d(Q, r) \mid Q \in (P_1, P_2)\}$, and $ISE(P_1, P_2) = \sum_{Q \in (P_1, P_2)} d^2(Q, r)$, respectively, where $Q$ is a contour point between $P_1$ and $P_2$, and $d(Q, r)$ is the distance from $Q$ to the approximated line segment $r$ defined by the points $P_1$ and $P_2$.

### 4.3. Eighth step. Improvement

The final step of the new proposal is based on the algorithm proposed by Masood (2008b) and improves the adjustment of the polygonal approximation to the contour. This improvement can be considered as a process of displacement of the dominant points to obtain a better adjustment. In addition, this process does not modify the number of dominant points of the polygonal approximation (see Algorithm 2 and Figure 5).

The time complexity of this step is $O(d \cdot n)$, where $d$ is the number of dominant points of the polygonal approximation and $n$ is the number of points of the contour, respectively. Generally, $d$ is very small in comparison to $n$, and an average of $n/d$ points of the contour are analysed for each dominant point.

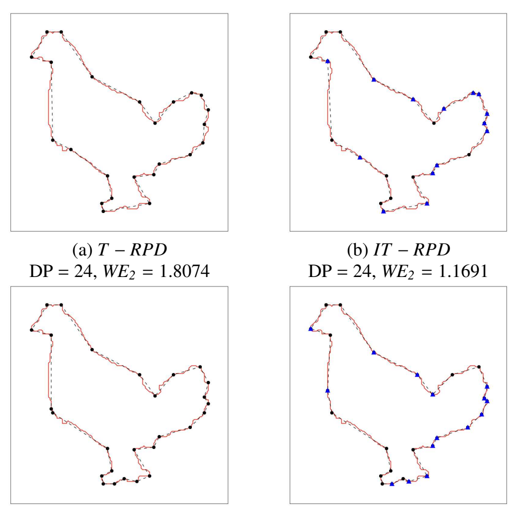
## 5. Experiments and results

### 5.1. Introduction

The quality measurements used in the experiments to compare the performance of the algorithms were (see Section 2.2): the number of dominant points ($DP$), compression ratio ($CR$) and integral square error ($ISE$) and the parameterized weighted sum of the squared error ($WE_n$), where $n \in \{1, 2, 3\}$. In addition, for every measurement $M$, the average value ($\mu(M)$) and the standard deviation ($\sigma(M)$) were also computed. Furthermore, the index $I_M = MinIndex(\mu(M))$ (or $I_M = MaxIndex(\mu(M))$) was also computed to compare the value of each algorithm with the best. For instance, if a measurement $M$ orders the algorithms in increasing way (the smallest is the best), then $I_M = MinIndex(\mu(M))$ is computed dividing the value of $\mu(M)$ by the smallest; otherwise, if a measurement $M$ orders the algorithms in decreasing way (the greatest is the best), then

**Algorithm 2** Eighth step. Improvement

{Let $L$ be the set of dominant points}
$L \leftarrow \{P_1, \ldots, P_d\}$
{Mark the points of $L$ as not revised}
**for all** $i \in \{1 \ldots d\}$ **do**
  $revised(P_i) \leftarrow false$
**end for**
**while** (there are points not revised in $L$) **do**
  Let $P$ be the first not revised point in $L$
  Let $P'$ and $P''$ be the current contiguous dominant points of $P$
  Find the *optimal point* $Q$ that minimizes the *ISE* error in the contour part bounded by $P'$ and $P''$.
  **if** $(P = Q)$ **then**
    $revised(P) \leftarrow true$
  **else**
    $L \leftarrow L - \{P\}$
    $L \leftarrow L \cup \{Q\}$
    $revised(Q) \leftarrow true$
    $revised(P') \leftarrow false$
    $revised(P'') \leftarrow false$
  **end if**
**end while**



(a) $T - RPD$
DP = 24, $WE_2 = 1.8074$

(b) $IT - RPD$
DP = 24, $WE_2 = 1.1691$

(c) Proposal: $CT - RPD$
DP = 27, $WE_2 = 1.7421$

(d) Improved proposal: $ICT - RPD$
DP = 27, $WE_2 = \mathbf{1.0452}$

**Fig. 5. Comparison with the original algorithm using the contour *chicken-10_0* (N = 468 points). The dominant points moved in the improvement step are marked with a blue triangle.**

$I_M = MaxIndex(\mu(M))$ is computed dividing the greatest one by the value of $\mu(M)$. The best value for $I_M = MinIndex(\mu(M)$ and $I_M = MaxIndex(\mu(M))$ is 1.00, and the worst value is $\infty$.

As it was pointed out in section 2.2, $WE_2$ is the fairest measurement to evaluate the quality of the polygonal aproximations, because $WE_1$ and $WE_3$ favor polygonal approximation with many or few dominant points, respectively. Due to these differences in the evaluation criteria, a new aggregate voting system was considered to compute the index $I_{averageWE_{1-3}}$, defined as the average of $I_{WE_1}$, $I_{WE_2}$, and $I_{WE_3}$ (see Table 1).

The database of shapes "MPEG-7 Core Experiment CE-Shape-1 Test Set (Part B)" was used in the experiments (Jeannin and Bober, 1999). This database is available in (MPEG-7, 1999) and contains 1,400 images, classified into 70 categories, and each category includes 20 samples, with different rotation, size and position, and even image resolution (Latecki et al., 2000). This database was also used previously (Parvez, 2015; Fernández-García et al., 2016; Madrid-Cuevas et al., 2016).

### 5.2. First experiment

The new proposal, $ICT-RDP$ was compared with other seven algorithms: $CT-RDP$, a version of the new proposal that does not include the eight step of improvement (see Section 4.3); the original algorithm $T-RDP$ (Fernández-García et al., 2016) and a new version called $IT-RDP$ which includes the improvement process (see Section 4.3); the three versions of $RDP$, Carmona and Masood algorithms proposed by Prasad et al. (2012), which will be referred to as $P-RDP$, $P-CAR$ and $P-MAS$, respectively, for short; and the algorithm proposed by Madrid-Cuevas et al. (2016), referred to as $MAD$. All these algorithms were selected because they are non-optimal and unsupervised algorithms. These algorithms were applied to all 1,400 contours of the public MPEG-7 database of shapes (Jeannin and Bober, 1999; MPEG-7, 1999). The polygonal approximations generated were evaluated using the indexes $I_{WE_1}$, $I_{WE_2}$, $I_{WE_3}$ and $I_{averageWE1-3}$, although the Table 1 also shows the values of $DP$, $CR$, $ISE$, $WE_1$, $WE_2$, and $WE_3$, for completeness. The best results are highlighted in **bold**.

The new proposal $ICT-RDP$ obtained the best result (1.000) for $I_{WE_2}$, the second best result (1.0951) for $I_{WE_3}$, but very close to the best one ($IT-RDP$), and the third best result (1.4297) for $I_{WE_1}$. The combined index $I_{averageWE1-3}$ showed that the new proposal obtained the overal best result (1.1749). Among the unimproved algorithms, the proposed $CT-RDP$ obtained the best result for the combined index $I_{averageWE1-3}$ (1.9662).

This experiment also allowed to compare the number of the *initial points*. The centroid criterion (see Section 3.1, step 1) provided a mean value of 2.0243 *initial points* to the algorithms $T-RDP$ and $IT-RDP$, which subsequently generated polygonal approximations with a mean value of 45.0693 *dominant points*; instead, the convex hull criterion criterion (see Section 4.1) provided a mean value of 31.7792 *initial points* to subsequently generate polygonal approximations with a mean value of 53.9514 *dominant points*. On average, the algorithms $CT-RDP$ and $ICT-RDP$, based on the convex hull criterion, deleted 14.6329 quasi-collinear points in the *Superflous dominant point deletion* step (see Section 4.2 and Figure 4).

### 5.3. Second experiment

The second experiment was developed to analyse the computing time of the eight non-optimal and unsupervised algorithms: (1) each algorithm was run 10 times with each of the 1,400 contours of the public MPEG-7 database (Jeannin and Bober, 1999; MPEG-7, 1999); (2) the mean value of the run time of every algorithm for contours with the same number of points was computed; the 1,400 contours from MPEG-7 database were grouped into 979 subsets, so that each subset was composed of contours with the same number of points (see Figure 6); and, finally, (3) the average of the mean values obtained in the step 2 was computed (see Table 2). The experiment was run on a compatible computer with an Intel(R) Core(TM) i5 661 @ 3.33GHz x 4 and 3,7 GB of memory. The operating system Ubuntu 16.04.6 LTS xenial 64 bits was used. The times were measured in micro seconds ($\mu s$).

In order to analyse the performance of the algorithms, the indexes $I_{Time_1}$ and $I_{Time_2}$ of the Table 2 were used. These indexes compared the value of each algorithm with the best one. The best value for these indexes is 1.0000. $I_{Time_1}$ was computed using the average of the mean values of the computing time of the algorithms, whereas $I_{Time_2}$ considered the addition of the average and standard deviation. The new proposal $CT-RDP$, without the improvement step (see Section 4.3), obtained an index $I_{Time_1}$ of 1.2648, ranked in fourth place after $P-CAR$, $MAD$ and $T-RDP$. The value of $CT-RDP$ for the index $I_{Time_2}$ was 1.1630, ranked in third place after $T-RDP$ and $MAD$. On the other hand, the new proposal $ICT-RDP$, with the improvement step, required more computing time than $CT-RDP$, but the analysis of the indexes showed that $ICT-RDP$ obtained better results ($I_{Time_1} = 11.4879$, $I_{Time_2} = 12.7916$) than the improved version $IT-RDP$ of the original algorithm ($I_{Time_1} = 12.0290$, $I_{Time_2} = 13.5994$). Therefore, if the improvement step is applied then the convex hull criterion must be used to select the *initial points*, because the computing time of the $ICT-RPD$ is better than the computing time of $IT-RDP$, and, above all, generates better polygonal approximations (see Tables 1 and 2).

Finally, an additional regression model was developed to analyse the complexity of the computing time of the algorithms. The regression model $nlog(n)$ can be accepted for the original algorithm $T-RDP$ and the new proposal $CT-RDP$, both without the improvement step, as show their coefficients of determination $R^2 = 0.9472$ and $R^2 = 0.9498$, respectively, which are greater than 0.9 (see Figure 6 (e) and (f)). In the case of the algorithms with the improvement step, $IT-RDP$ and $ICT-RDP$, their coefficients of determination $R^2 = 0.6107$ and $R^2 = 0.6865$, respectively, do not allow to confirm the regression model $nlog(n)$, because are less than 0.9 (see Figures 6 (g) and (h)). As a consequence of these results, the time complexity of $IT-RDP$ and $ICT-RDP$ is $O(n\log(n))$ in the best case and $O(n^2)$ in the worst case, where $n$ is the number of contour points.

## 6. Conclussions and future work

The present paper proposes a new algorithm, called $ICT-RDP$, for generation of polygonal approximations which improves the previous algorithm $T-RDP$ (Fernández-García et al., 2016). The main contributions of the new approach are: (1) the convex hull of 2D closed digital planar curves can be considered to select the initial points of the polygonal aproximation; (2) the superfluous or *quasi-collinear* dominant points, that may appear in rounded parts of the contour due to the application of the convex hull, are eliminated; and (3) a final improvement step is used to increase the quality of the polygonal approximation without changing its number of dominant points.

**Table 1. First experiment: comparative results of the polygonal approximations algorithms using the quality measurements.**

| | Unimproved algorithms | | | | | | Improved algorithms | |
|---|---|---|---|---|---|---|---|---|
| **Measurement** | $MAD$ | $P-CAR$ | $P-MAS$ | $P-RDP$ | $T-RPD$ | $CT-RDP$ | $IT-RD$ | $ICT-RDP$ |
| $I_{DP}$ = MinIndex($\mu(DP)$) | 2.8590 | 1.5021 | 2.9322 | 2.4433 | **1.0000** | 1.1971 | **1.0000** | 1.1971 |
| $\mu(DP)$ | 128.8529 | 67.6979 | 132.1507 | 110.1186 | 45.0693 | 53.9514 | 45.0693 | 53.9514 |
| $\sigma(DP)$ | 166.0244 | 80.2357 | 169.9686 | 116.3092 | 22.7598 | 25.0818 | 22.7598 | 25.0818 |
| $I_{CR}$ = MaxIndex($\mu(CR)$) | 2.1919 | 1.2128 | **1.0000** | 2.2232 | 1.2760 | 1.4954 | 1.2760 | 1.4954 |
| $\mu(CR)$ | 18.0295 | 32.5860 | **39.5194** | 17.7757 | 30.9705 | 26.4266 | 30.9705 | 26.4266 |
| $\sigma(CR)$ | 19.3437 | 29.7438 | 76.7221 | 20.7151 | 25.2543 | 23.4654 | 25.2543 | 23.4654 |
| $I_{DP}$ = MinIndex($\mu(ISE)$) | **1.0000** | 13.4855 | 568.3419 | 1.3643 | 12.5930 | 8.6871 | 6.2084 | 5.0507 |
| $\mu(ISE)$ | **160.7779** | 2,168.1645 | 91,376.7954 | 219.3423 | 2,024.6730 | 1,396.6875 | 998.1759 | 812.0354 |
| $\sigma(ISE)$ | 126.1560 | 15,842.7718 | 646,465.0711 | 168.2474 | 5,023.2454 | 3,686.2365 | 2,124.5349 | 2,104.4019 |
| $I_{WE_1}$ = MinIndex($\mu(WE_1)$) | **1.0000** | 3.8007 | 17.1285 | 1.1995 | 3.3110 | 2.4417 | 1.7024 | 1.4297 |
| $\mu(WE_1)$ | **16.3020** | 61.9593 | 279.2299 | 19.5540 | 53.9760 | 39.8051 | 27.7525 | 23.3071 |
| $\sigma(WE_1)$ | 22.6226 | 139.7476 | 1,466.7902 | 23.3915 | 100.5091 | 73.9782 | 44.8601 | 42.0023 |
| $I_{WE_2}$ = MinIndex($\mu(WE_2)$) | 3.0345 | 3.6343 | 3.9639 | 2.6027 | 2.0110 | 1.6762 | 1.0770 | **1.0000** |
| $\mu(WE_2)$ | 2.6419 | 3.1640 | 3.4509 | 2.2659 | 1.7508 | 1.4593 | 0.9376 | **0.8706** |
| $\sigma(WE_2)$ | 6.3755 | 4.5131 | 5.9653 | 4.1113 | 2.5403 | 1.9799 | 1.2015 | 1.1219 |
| $I_{WE_3}$ = MinIndex($\mu(WE_3)$) | 15.3190 | 6.2744 | 13.1176 | 8.1420 | 1.7907 | 1.7806 | **1.0000** | 1.0951 |
| $\mu(WE_3)$ | 0.5897 | 0.2415 | 0.5049 | 0.3134 | 0.0689 | 0.0685 | **0.0385** | 0.0422 |
| $\sigma(WE_3)$ | 2.0075 | 0.5429 | 1.3942 | 0.7565 | 0.0853 | 0.0756 | 0.0445 | 0.0446 |
| $I_{averageWE_{1-3}} = \frac{I_{WE_1}+I_{WE_2}+I_{WE_3}}{3}$ | 6.4512 | 4.5698 | 11.4033 | 3.9814 | 2.3709 | 1.9662 | 1.2598 | **1.1749** |

Notes: number of contours = 1,400; average number of contour points = 1,271.0379 points.

**Table 2. Second experiment: comparative of the mean values of the computing time of the algorithms measured in micro seconds ($\mu s$).**

| | Unimproved algorithms | | | | | | Improved algorithms | |
|---|---|---|---|---|---|---|---|---|
| **Time ($\mu s$)** | $MAD$ | $P-CAR$ | $P-MAS$ | $P-RDP$ | $T-RPD$ | $CT-RDP$ | $IT-RD$ | $ICT-RDP$ |
| $I_{Time_1}$ = MinIndex($\mu(Time)$) | 1.1326 | **1.0000** | 105.9447 | 1.4496 | 1.0872 | 1.2648 | 12.0290 | 11.4879 |
| $\mu(Time)$ | 4,626.4434 | **4,084.7578** | 432,758.2626 | 5,921.1067 | 4,440.9898 | 5,166.5681 | 49,135.7432 | 46,925.4858 |
| $\sigma(Time)$ | 3,818.4409 | 7,929.5973 | 672,547.1353 | 13,109.0317 | 3,314.4382 | 3,853.2963 | 56,333.4583 | 52,278.9179 |
| $I_{Time_2}$ = MinIndex($\mu(Time)+\sigma(Time)$ ) | 1.0889 | 1.5492 | 142.5202 | 2.4538 | **1.0000** | 1.1630 | 13.5994 | 12.7916 |
| $\mu(Time)+\sigma(Time)$ | 8,444.8843 | 12,014.3550 | 1,105,305.3979 | 19,030.1383 | **7,755.4280** | 9,019.8644 | 105,469.2015 | 99,204.4037 |

Note: the 1,400 contours from MPEG-7 database were grouped into 979 subsets, so that each subset was composed of contours with the same number of points.

The proposed algorithm $ICT-RDP$ obtained the best results in the experiments and it was faster than the improved version $IT-RDP$ of the previous algorithm $T-RDP$. $ICT-RPD$ is an unsupervised or automatic algorithm (no parameters have to be set or tuned), and non-optimal, but it can generate reasonable polygonal approximations. Its expected time complexity is $O(n\log(n))$ in the best case and $O(n^2)$ in the worst case, where $n$ is the number of contour points. Among the unimproved algorithms, the proposed $CT-RDP$ obtained the best results and its expected time complexity is $O(n\log(n))$.

Finally, future work should be aimed at designing a new quality measurement for polygonal approximations, due to the drawbacks explained in section 2.2. In relation to the new proposal $ICT-RDP$, the *Superflous dominant point deletion* step (see Section 4.2) and the *Improvement* step (see Section 4.3) could be applied taking into account the dominant point with lowest or greatest error (Emax or ISE) in every iteration, instead of being applied using the current *greedy way*.

## Acknowledgement

This work has been developed with the support of the Research Project TIN2016-75279-P of Spain Ministry of Economy, Industry, and Competitiveness, and FEDER.

## References

Aguilera-Aguilera, E., Carmona-Poyato, A., Madrid-Cuevas, F., Muñoz-Salinas, R., 2015. Novel method to obtain the optimal polygonal approximation of digitalplanar curves based on mixed integer programming. J. Vis. Commun. Image R. 30, 106–116. doi:http://dx.doi.org/10.1016/j.jvcir.2015.03.007.

Andrew, A.M., 1979. Another efficient algorithm for convex hull in two dimensions. Information Processing Letters 9, 216–219. doi:http://dx.doi.org/10.1016/0020-0190(79)90072-3.

Ataer-Cansizoglu, E., Bas, E., Kalpathy-Cramer, J., Sharp, G.C., Erdogmus, D., 2013. Somenovel method to obtain the optimal polygonal approximation of digitalplanar curves based on mixed integer programming. Pattern Recognition 46, 106–116. doi:http://dx.doi.org/10.1016/j.patcog.2012.10.014.

Attneave, F., 1954. Some informational aspects of visual perception. Phychol. Rev. 61, 189–193. doi:http://dx.doi.org/10.1037/h0054663.

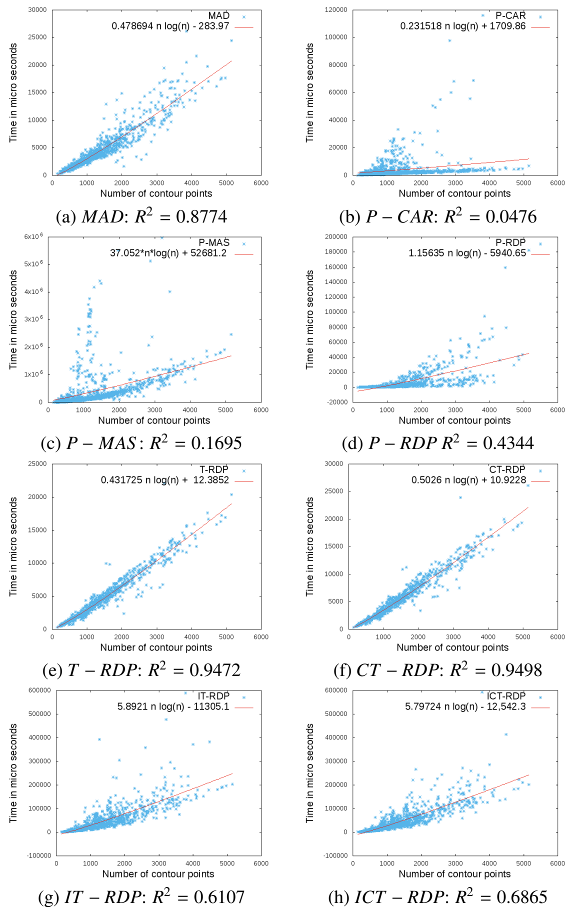
Bykat, A., 1978. Convex hull of a finite set of points in two dimensions. Information Processing Letters 7, 296 – 298. doi:http://dx.doi.org/10.1016/0020-0190(78)90021-2.

Carmona-Poyato, A., Madrid-Cuevas, F., Medina-Carnicer, R., Muñoz-Salinas, R., 2010. Polygonal approximation of digital planar curves through break point suppression. Pattern Recognition 43, 14–25. doi:http://dx.doi.org/10.1016/j.patcog.2009.06.010.

Carmona-Poyato, A., Medina-Carnicer, R., Madrid-Cuevas, F.J., Muñoz-Salinas, R., Fernández-García, N.L., 2011. A new measurement for assesing polygonal approximation of curvers. Pattern Recognition 4, 44–54. doi:http://dx.doi.org/10.1016/j.patcog.2010.07.029.

Chadnov, R.V., Skvortsov, A.V., 2004. Convex hull algorithms review, in: Proceedings. The 8th Russian-Korean International Symposium on Science and Technology, pp. 112 – 115. doi:DOI:10.1109/KORUS.2004.1555560.

Douglas, D.H., Peucker, T.K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: The International Journal for Geographic Information and




(a) *MAD*: $R^2 = 0.8774$

(b) $P-CAR$: $R^2 = 0.0476$

(c) $P-MAS$: $R^2 = 0.1695$

(d) $P-RDP$ $R^2 = 0.4344$

(e) $T-RDP$: $R^2 = 0.9472$

(f) $CT-RDP$: $R^2 = 0.9498$

(g) $IT-RDP$: $R^2 = 0.6107$

(h) $ICT-RDP$: $R^2 = 0.6865$

**Fig. 6. Regression model $n\ log(n)$ for the mean values of the computing time of the algorithms. $R^2$ is the coefficient of determination.**

Geovisualization 10, 112 – 122. doi:http://dx.doi.org/10.3138/FM57-6770-U75U-7727.

Eddy, W.F., 1977. A new convex hull algorithm for planar sets. ACM Transactions on Mathematical Software 3, 398 – 403. doi:http://dx.doi.org/10.1145/355759.355766.

Fernández-García, N.L., Del-Moral-Martínez, L., Carmona-Poyato, A., Madrid-Cuevas, F., Medina-Carnicer, R., 2016. A new thresholding approach for automatic generation of polygonal approximations. Journal of Visual Communication and Image Representation 35, 155 – 168. doi:http://dx.doi.org/10.1016/j.jvcir.2015.12.013.

Graham, R.L., 1983. Finding the convex hull of a simple polygon. Journal of Algorithms 4, 324 – 331. doi:http://dx.doi.org/10.1007/BF00993195.

Hershberger, J., Snoeyink, J., 1992. Speeding up the Douglas-Peucker line simplification algorithm, in: Proceedings of the Fifth International Symposium on Spatial Data Handling, pp. 134–143. doi:http://dx.doi.org/10.1.1.17.6932.

Horng, J.H., Li, J., 2002. An automatic and efficient dynamic programming algorithm for polygonal approximation of digital curves. Pattern Recognition Letters 23, 171–182. doi:http://dx.doi.org/10.1016/S0167-8655(01)00098-8.

Jeannin, S., Bober, M., 1999. Description of core experiments for mpeg-7 motion/shape. mpeg-7. ISO/IEC JTC1/SC29/WG11/MPEG99/N2690 .

Kalaivani, S., Ray, B.K., 2019. A heuristic method for initial dominant point detection for polygonal approximations. Soft Computing 23, 8435–8452. doi:https://doi.org/10.1007/s00500-019-03936-1.

Kolesnikov, A., Fränti, P., 2007. Polygonal approximation of closed discrete curves. Pattern Recognition 40, 1282–1293. doi:http://dx.doi.org/10.1016/j.patcog.2006.09.002.

Latecki, L., Lakamper, R., Eckhardt, U., 2000. Shape descriptors for non-rigid shapes with a single closed contour. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) , 424–429.

Loncaric, S., 1998. A survey of shape analysis techniques. Pattern Recognition 31, 983–1001. doi:http://dx.doi.org/10.1016/S0031-2023(97)00122-2.

Lowe, D.G., 1987. Three dimensional object recognition from single two dimensional images,. Artificial Intelligence 31, 355–395. doi:http://dx.doi.org/10.1016/0004-3702(87)90070-1.

Madrid-Cuevas, F., Aguilera-Aguilera, E., A.Carmona-Poyato, R.Muñoz-Salinas, R.Medina-Carnicer, N.L.Fernández-García, 2016. An efficient unsupervised method for obtaining polygonal approximations of closed digital planar curves. Journal of Visual Communication and Image Representation 39, 152 – 163. doi:http://dx.doi.org/10.1016/j.patcog.2007.05.021.

Marji, M., Siy, P., 2004. Polygonal representation of digital planar curves through dominant point detection - a nonparametric algorithm. Pattern Recognition 37, 2113 – 2130. doi:http://dx.doi.org/10.1016/j.patcog.2004.03.004.

Masood, A., 2008a. Dominant point detection by reverse polygonization of digital curves. Image and Vision Computing 26, 703 – 715. doi:http://dx.doi.org/10.1016/j.imavis.2007.08.006.

Masood, A., 2008b. Optimized polygonal approximation by dominant point deletion. Pattern Recognition 41, 227 – 239. doi:http://dx.doi.org/10.1016/j.patcog.2007.05.021.

Masood, A., Haq, A., 2007. A novel approach to polygonal approximation of digital curves. Journal of Visual Communication and Image Representation 18, 264 – 274. doi:http://dx.doi.org/10.1016/j.jvcir.2006.12.002.

Melkman, A., O'Rourke, J., 1998. Computational Morphology. G.T. Toussaint (Ed.), North-Holland, Amsterdam. chapter On polygonal chain approximation. pp. 87–95.

MPEG-7, 1999. Core experiment ce-shape-1 test set (part b) benchmarking image database for shape recognition techniques. Available at http://www.cis.temple.edu/\~latecki/TestData/mpeg7shapeB.tar.gz.

Parvez, M.T., 2015. Optimized polygonal approximations through vertex relocations in contour neighborhoods. Image and Vision Computing 34, 1 – 10. doi:http://dx.doi.org/10.1016/j.imavis.2014.10.012.

Perez, J.C., Vidal, E., 1994. Optimum polygonal approximation of digitized curves. Pattern Recognition Letters 15, 743–750. doi:http://dx.doi.org/10.1016/0167-8655(94)90002-7.

Pikaz, A., Dinstein, I., 1995. Optimal polygonal approximation of digital curves. Pattern Recognition 28, 373–379. doi:http://dx.doi.org/10.1016/0031-3203(94)00108-X.

Prasad, D.K., Leung, M.K.H., Quek, C., Cho, S.Y., 2012. A novel framework for making dominant point detection methods non-parametric. Image and Vision Computing 30, 843–859. doi:http://dx.doi.org/10.1016/j.imavis.2012.06.010.

Ramer, U., 1972. An iterative procedure for the polygonal approximation of plane curves. Computer Graphics and Image Processing 1, 244 – 256. doi:http://dx.doi.org/10.1016/S0146-664X(72)80017-0.

Rosin, P.L., 1997. Techniques for assessing polygonal approximations of curves. IEEE Trans. Pattern Analysis and Machine Intelligence 19, 659–666. doi:http://dx.doi.org/10.1109/34.601253.

Salotti, M., 2001. An efficient algorithm for the optimal polygonal approximation of digitized curves. Pattern Recognition Letters 22, 215–221. doi:http://dx.doi.org/10.1016/S0167-8655(00)00088-X.

Salotti, M., 2002. Optimal polygonal approximation of digitized curves using the sum of square deviations criterion. Pattern Recognition 35, 435–443. doi:http://dx.doi.org/10.1016/S0031-3203(01)00051-6.

Sarkar, D., 1993. A simple algorithm for detection of significant vertices for polygonal approximation of chain-coded curves. Pattern Recognition 14, 959 – 964. doi:http://dx.doi.org/10.1016/0167-8655(93)90004-W.

Wu, W.Y., 2003a. An adaptive method for detecting dominant points. Pattern Recognition 36, 2231–2237. doi:http://dx.doi.org/10.1016/S0031-3203(03)00087-6.

Wu, W.Y., 2003b. Dominant point detection using bending value. Image and Vision Computing 21, 517–525. doi:http://dx.doi.org/10.1016/S0262-8856(03)00031-3.

Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognition 37, 1 – 19. doi:http://dx.doi.org/10.1016/j.patcog.2003.07.008.