Measuring Linearity of Open Planar Curve Segments

Joviša Žunić,[1,*] Paul L. Rosin[2]

[1] Department of Computer Science, University of Exeter
Exeter EX4 4QF, U.K.

J.Zunic@ex.ac.uk

[2] School of Computer Science, Cardiff University
Cardiff CF24 3AA, Wales, U.K.

Paul.Rosin@cs.cf.ac.uk

Abstract

In this paper we define a new linearity measure for open planar curve segments. We start with the integral of the squared distances between all the pairs of points belonging to the measured curve segment, and show that, for curves of a fixed length, such an integral reaches its maximum for straight line segments. We exploit this nice property to define a new linearity measure for open curve segments. The new measure ranges over the interval $(0, 1]$, and produces the value 1 if and only if the measured open line is a straight line segment. The new linearity measure is invariant with respect to translations, rotations and scaling transformations. Furthermore, it can be efficiently and simply computed using line moments. Several experimental results are provided in order to illustrate the behaviour of the new measure.

Key words: Shape, curves, linearity measure, image processing.

*J. Žunić is also with the Mathematical Institute, Serbian Academy of Arts and Sciences, Belgrade.

1 Introduction

Shape descriptors are a useful tool in the area of many computer vision and image processing tasks (e.g. image retrieval, object classification, object recognition, and so forth). Different mathematical tools have been used to define the shape descriptors: algebraic invariants [10], Fourier analysis [13], statistics [17], morphological operations [24], curvature [25], integral transformations [21], fuzzy approaches [32], computational geometry [36], and so forth.

Generally speaking there are two approaches to analyse shapes: boundary-based (which use the information from boundary points only) and area-based ones which use all shape points. It could be said that, in the past, more attention has been given to the area-based methods. The area-based methods are more robust (e.g. with respect to noise). Although not often mentioned, an additional reason for a higher number of methods which are based on 'interior' shape points, rather than the methods based on the boundary points, is that the area-based methods are usually simpler to compute. For example, to estimate accurately the area of a given planar shape it is sufficient to enumerate the number of pixels inside the shape [11, 15] while perimeter estimation is still a challenging task [4, 27]. Another example is geometric (area) moment invariants [10] which are easily and accurately computable from the corresponding object images, while their boundary-based analogues [3] involve computation of path integrals. However, these are not easy to estimate from discrete data, which are mainly used in image processing and computer vision tasks. Nevertheless, high resolution image data is now available, enabling sufficiently accurate computation of line features (e.g. perimeter, curvature), making the use of boundary-based methods practical for many applications. The boundary-based methods are more suitable for computer vision and image processing tasks that require high sensitivity to small details on shape boundaries. In addition, boundary-based methods, very often, have a much lower time complexity, because shape boundaries are represented by a significantly smaller number of pixels than complete shapes are. Of course, there are methods which cannot be classified as either boundary-based or area-based ones. For example, a very standard shape compactness measure [28, 33]:

$$\mathcal{C}_{st}(S) = \frac{4 \cdot \pi \cdot Area\_of\_S}{(Perimeter\_of\_S)^2}$$

(for more details about the isoperimetric inequality see Chapter 1 in [14]) which can be considered as indicating how much a given shape S differs from a perfect circle, obviously uses both boundary and interior information. Also, there are methods which use only information from specific points (corners, for example) or specific boundary parts – e.g. belonging to the convex hull of the given shape [34], or from significant object characteristics in a correlogram structure [6].

As mentioned, due to the high demand for efficient image verification, identification and recognition systems there is an ongoing interest for shape properties that could be derived from their boundaries [16, 18, 26, 31]. A particular benefit of boundary-based methods is that some objects, like human signatures for example, are linear by their nature and area-based descriptors cannot be used for their analysis.

In this paper we derive a new linearity measure that indicates the degree to which an open curve segment corresponds to a perfectly straight line segment. Several linearity measures

have already been considered. Most of them relate to ordered sets of points [1, 7, 8, 12, 30, 35], but there are also linearity measures for unordered sets of points [29, 33].

It is not possible to establish a strict ranking among shape measures. A measure dominant in one application could fail in another. All of them have their strengths and their weaknesses and their performance strongly depends on the application performed. For example, a simple and natural linearity measure defined as the ratio of the squared distance between the curve end points and the curve length (so called *straightness index* [1]) is simple and fast to compute. Such a measure assigns the same value of linearity to the two open curve segments in Figure 1(a) and Figure 1(b), which could be desirable in some applications

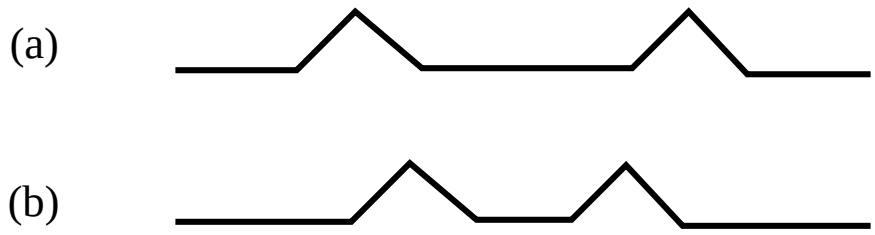


Figure 1: Straightness index [1] assigns the same linearity for the curves displayed. The new linearity measure and the measure from [35] assign different linearities to these curves.

where the curves displayed are judged as curves of the same linearity. On the other hand, if we would like to distinguish between these curves, what could be a requirement in some applications, we need another measure. One such measure which assigns different linearities for the curve segments in Figure 1 was developed in [35]. Informally speaking, the measure from [35] defines the linearity of a given curve segment as the sum of the average square distances of the curve points from the curve end points. Obviously, such a linearity measure also depends on the distance between the end points (as the straightness index does) but is also sensitive to curve deformations (the straightness index is not). Further, since the definition of the linearity measure from [35] uses the curve endpoints it cannot be applied to compound curves (like curves in Figure 4, for example) which do not have strictly defined end points. The measure introduced here overcomes such a problem and can be applied to such compound curves. For some applications this can be a big advantage because these curves appear very often in applications (e.g. shape skeletons, or letters, as displayed in Figure 4). Notice that the new linearity measure also considers the squared distance between curve points, but not only the sum of squared distances between a curve point and the curve end points.

The measure defined here is easy to compute, and is invariant with respect to translations, rotations, and scaling. It ranges over the interval $(0, 1]$ and takes the value 1 if and only if the measured curve is a perfect straight line segment. Several experiments are provided to illustrate the behaviour of the new linearity measure. Note that the linearity measure defined here is a global shape descriptor, and it can be used as a component along with others in a feature vector used for shape classification. This approach makes it easy to update the feature vectors of a shape database if new shapes are added, and consequently makes shape matching and classification relatively simple and fast. In contrast, local shape descriptors

may provide higher classification rates, but their computation and matching tends to be time consuming.

2 New Linearity Measure

In this section we define a new linearity measure for open curve segments. We will use the following denotations:

- Let $\mathbf{C}$ be an open curve given in the arc-length parametrisation $x = x(s)$, $y = y(s)$, and $s \in [0, 1]$;
- Let $F(s_0, \mathbf{C})$ denote the integral of the squared distances of a fixed point $(x(s_0), y(s_0)) \in \mathbf{C}$ to all the points of $\mathbf{C}$, i.e.

$$F(s_0, \mathbf{C}) = \int_{s=0}^{1} \left((x(s_0) - x(s))^2 + (y(s_0) - y(s))^2 \right) ds. \tag{1}$$

Now we give the following theorem.

Theorem 1 *The following is true for any planar curve* $\mathbf{C}$*:*

(a) $F(s_0, \mathbf{C}) \leq s_0^2 - s_0 + \dfrac{1}{3}, \qquad$ *for any* $s_0 \in [0, 1]$.

(b) $\displaystyle\int_{s_0=0}^{1} F(s_0, \mathbf{C})\ ds_0 \leq \frac{1}{6}.$

(c) $\displaystyle\int_{s_0=0}^{1} F(s_0, \mathbf{C})\ ds_0 = \frac{1}{6} \Leftrightarrow \mathbf{C}$ *is a straight line segment.*

Proof. Since the parameter s measures the distance along the curve $\mathbf{C}$, between the points $(x(s), y(s))$ and $(x(0), y(0))$, it means that $|s_0 - s|$ measures the distance (again, along the curve $\mathbf{C}$) between points $(x(s_0), y(s_0))$ and $(x(s), (y(s))$. This implies (see Figure 2 for an illustration):

$$(x(s_0) - x(s))^2 + (y(s_0) - y(s))^2 \leq (s_0 - s)^2, \qquad \text{for all } s \in [0, 1]. \tag{2}$$

Now, we prove **(a)**:

$$F(s_0, \mathbf{C}) = \int_{s=0}^{1} \left((x(s_0) - x(s))^2 + (y(s_0) - y(s))^2 \right) ds \leq \int_{s=0}^{1} (s_0 - s)^2\, ds = s_0^2 - s_0 + \frac{1}{3}. \tag{3}$$

Item **(b)** easily follows from **(a)**. I.e.:

$$\int_{s_0=0}^{1} F(s_0, \mathbf{C})\ ds_0 \leq \int_{s_0=0}^{1} \left(s_0^2 - s_0 + \frac{1}{3} \right)\ ds_0 = 1/6.$$

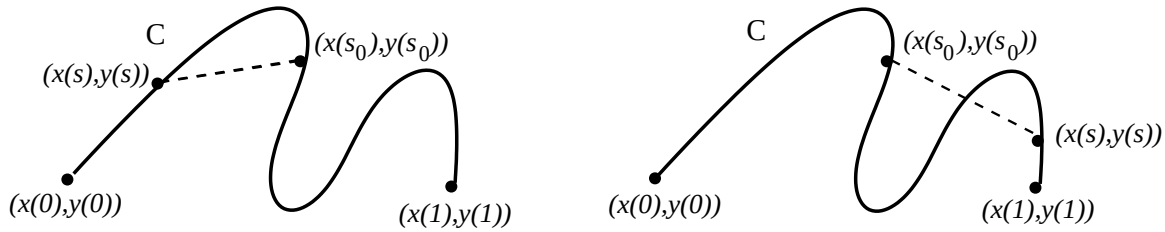


Figure 2: The Euclidean distance (length of dashed line) between the points $(x(s_0), y(s_0))$ and $x(s), y(s))$ is not bigger than $|s - s_0|$ (which equals the length of the arc of $\mathbf{C}$ between those points) in both cases: $s \in [0, s_0]$ (on the left) and $s \in [s_0, 1]$ (on the right).

Item **(c)** follows from the fact that the equalities in (2) and (3) are true if and only if $\mathbf{C}$ is a straight line segment. □

Further, by using (1) it is easy to see that $F(s_0, \mathbf{C})$ can be expressed as

$$F(s_0, \mathbf{C}) = m_{2,0}(\mathbf{C}) + m_{0,2}(\mathbf{C}) - 2 \cdot x(s_0) \cdot m_{1,0}(\mathbf{C}) - 2 \cdot y(s_0) \cdot m_{0,1}(\mathbf{C}) + x(s_0)^2 + y(s_0)^2 \quad (4)$$

where

$$m_{p,q}(C) = \int_{\mathbf{C}} x^p \cdot y^q \; ds$$

are the well-known line moments (see [3]). In addition, (4) gives

$$\int_{s_0=0}^{1} F(s_0, \mathbf{C}) \; ds_0 \;\; = \;\; 2 \cdot m_{2,0}(\mathbf{C}) + 2 \cdot m_{0,2}(\mathbf{C}) - 2 \cdot m_{1,0}(\mathbf{C})^2 - 2 \cdot m_{0,1}(\mathbf{C})^2. \quad (5)$$

Now, by using statements **(b)** and **(c)** of Theorem 1 which say that $\int_{s_0=0}^{1} F(s_0, \mathbf{C}) ds_0$ is bounded from above by $\frac{1}{6}$ and reaches this upper bound if and only if $\mathbf{C}$ is a straight line segment, we naturally come to the following definition for a new linearity measure $\mathcal{L}(\mathbf{C})$ of a given open curve segment $\mathbf{C}$.

Definition 1 *The linearity measure* $\mathcal{L}(\mathbf{C})$ *of the open planar curve* $\mathbf{C}$, *having the length* 1, *is defined as*

$$\mathcal{L}(\mathbf{C}) = \frac{\int_0^1 F(s, \mathbf{C}) ds}{1/6} = 12 \cdot m_{2,0}(\mathbf{C}) + 12 \cdot m_{0,2}(\mathbf{C}) - 12 \cdot m_{1,0}(\mathbf{C})^2 - 12 \cdot m_{0,1}(\mathbf{C})^2. \quad (6)$$

Note. For the sake of simplicity, the above note assumes $\mathbf{C}$ to have unit length. If such an assumption is not made then the linearity of $\mathbf{C}$ is computed as:

$$\frac{12 \cdot m_{2,0}(\mathbf{C}) + 12 \cdot m_{0,2}(\mathbf{C}) - 12 \cdot m_{1,0}(\mathbf{C})^2 - 12 \cdot m_{0,1}(\mathbf{C})^2}{(length\_of\_\mathbf{C})^3}$$

The following theorem summarises the properties of $\mathcal{L}(\mathbf{C})$.

Theorem 2 *The linearity measure $\mathcal{L}(\mathbf{C})$ has the following properties:*

(i) $\mathcal{L}(\mathbf{C}) \in (0, 1]$, *for all open curve segments* $\mathbf{C}$*;*

(ii) $\mathcal{L}(\mathbf{C}) = 1 \quad \Leftrightarrow \quad \mathbf{C}$ *is a straight line segment;*

(iii) $\mathcal{L}(\mathbf{C})$ *is invariant with respect to similarity transformations.*

Proof. Items **(i)** and **(ii)** are a direct consequence of Theorem 1.

To prove the item **(iii)** it is sufficient to notice that the subintegral function in (1) is invariant with respect to translation and rotation, while the invariance with respect to scaling transformations follows from the Definition 6. □

In the next section we perform several experiments in order to illustrate the behaviour and applicability of the new linearity measure. When applying (6) to discrete data the computation can be performed incrementally. This is advantageous if linearity is to be repeatedly measured from overlapping sections of data since it allows more efficient computation than recomputing all the moments from scratch each time. Remember that (6) assumes that the curve has unit length, and so to enable the incremental computation the moments need to be rescaled as

$$\frac{m_{1,0}}{m_{0,0}^2}, \quad \frac{m_{0,1}}{m_{0,0}^2}, \quad \frac{m_{2,0}}{m_{0,0}^3}, \quad \frac{m_{0,2}}{m_{0,0}^3}.$$

3 Experiments

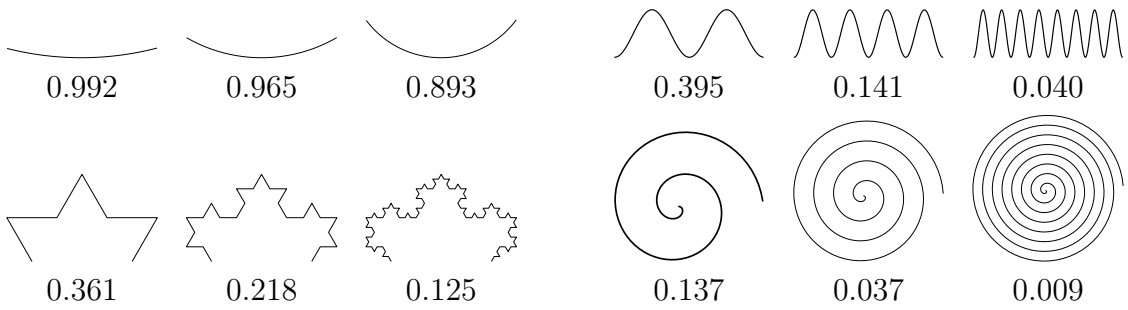


Figure 3: Values of $\mathcal{L}$ for synthetic curves.

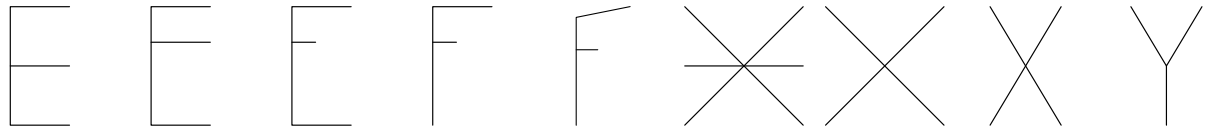


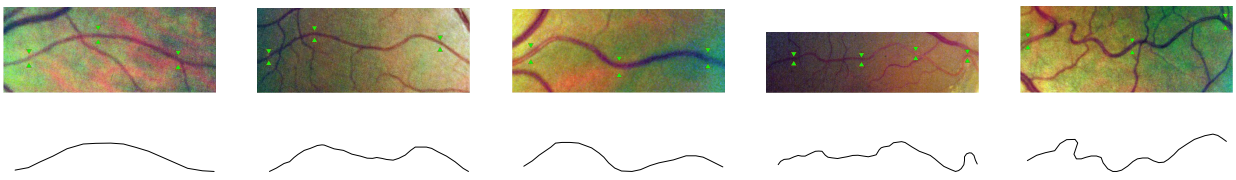
Figure 4: Values of $\mathcal{L}$ for synthetic curves containing multiple components.

In Figure 3 we show some simple synthetic curves with their calculated linearity values to illustrate the behaviour of $\mathcal{L}(\mathbf{C})$. It can be seen that the linearity measure matches our

intuition: as the curves become "more complex" (e.g. as it has more extreme points (as the last three curves in the first row in Figure 3), or as it has more corners (see the first three curves in the second row)) then their linearity value decreases.

It is worth mentioning that in addition to the fact that $\mathcal{L}(\mathbf{C}) = 1$ fully describes the shape of the curve $\mathbf{C}$ (it must be a straight line segment, as stated by Theorem 2), very small values of $\mathcal{L}(\mathbf{C})$ also give some information about the shape of $\mathbf{C}$: such curves must be inscribed into a small disc around the curve centroid. This is in accordance with the fact that $\mathcal{L}(\mathbf{C})$ actually considers the average squared values between pairs of points of $\mathbf{C}$. More than this is not known – i.e. for other values of $\mathcal{L}(\mathbf{C}) \in (\varepsilon, 1-\varepsilon)$ (ε indicates a small positive number) it is not possible to say something precise about the shape of $\mathbf{C}$. Curves very different in shape can have the same linearity measure, and the same happens for all shape measures. Notice that there are many known measures which do not characterise the shape of $\mathcal{C}$ for both extreme values (i.e. minimum and maximum). Such examples could be the orientability measures from [34] or the convexity measure from [21]. The situation is even worse for the well-known Hu moment invariants [10] where the shapes which maximise/minimise these invariants (apart from the first one [33]) are still unknown. A similar situation occurs with Fourier descriptors, and so forth.

The new measure can also applied to curves with multiple components and varying topology, as demonstrated in Figure 4. Notice that these curves are not simple curves because they cannot be presented in the form $x = x(s)$, $y = y(s)$, $s \in [0, 1]$. This is because these curves have more than 2 end points. E.g. the first curve in Figure 4 has 3 end points, while a simple open curve has exactly two such points: $(x(0), y(0))$ and $(x(1), y(1))$. Nevertheless, the new measure can be applied to compound curves. Indeed, each of these curves $\mathbf{C}$ can be represented as a union $\mathbf{C} = \mathbf{C}_1 \cup \ldots \cup \mathbf{C}_n$ of a number of simple non-overlapping curves $\mathbf{C}_1, \ldots \mathbf{C}_n$ and the moments $m_{p,q}(\mathbf{C})$ appearing in the definition of $\mathcal{L}(\mathbf{C})$ (see (6)) are computed as the sum $m_{p,q}(\mathbf{C}) = m_{p,q}(\mathbf{C}_1) + \ldots + m_{p,q}(\mathbf{C}_n)$ of the moments corresponding to these components. Notice that the decomposition of $\mathbf{C}$ into simple curve segments is not unique, but the computed moments $m_{p,q}(\mathbf{C})$ do not depend on the specific decomposition of $\mathbf{C}$.



0.935 0.872 0.842 0.781 0.642

Figure 5: Retinal images showing the blood vessels and their manually extracted centrelines [9]. The examples are shown manually ordered by increasing tortuosity. Values of $\mathcal{L}$ are displayed below each centreline.

We now demonstrate the application of the linearity measure to real data. This was taken from Grisan *et al.* [9], and consists of 60 retinal images of which 30 arteries and 30 veins of similar length and diameter were manually extracted. The vessels have also been assessed

by a clinician who ranked them according to their tortuosity (which is an indication of how twisted the curve is). Figure 5 shows five examples of arteries which cover the full range of tortuosity ratings. The goal is to automatically measure tortuosity, which can then be used to indicate retinal pathologies. This is evaluated by computing the correlation between the measured tortuosities against the clinical rankings using Spearman's rank-order correlation coefficient. Table 1 is based on [9] and shows results for a variety of methods ordered by their correlation values.[1]

arteries		veins	
measure	correlation	measure	correlation
τ	0.949	τ	0.853
tc/L_x	0.939	tc/L_x	0.842
$1-\mathcal{L}_{22}$	**0.935**	$1-\mathcal{L}_{10}$	**0.842**
tsc/L_x	0.928	tc	0.837
tsc	0.925	tsc	0.826
tc	0.922	tc/L_c	0.814
MAC	0.920	MAC	0.814
tc/L_c	0.919	$1-\mathcal{L}_{22}$	**0.814**
tsc/L_c	0.917	tsc/L_x	0.804
$1-\mathcal{L}_{10}$	**0.888**	tsc/L_c	0.773
$1-\mathcal{L}$	**0.858**	$1-\mathcal{L}$	**0.715**
TN	0.838	TN	0.695
L_c/L_x	0.792	L_c/L_x	0.656
DCI	0.787	DCI	0.589
ICM	0.684	ICM	0.575

Table 1: Spearman's rank-order correlation coefficient computed between various computed shape measures against clinical rankings of tortuosity. See [9] for full details of the shape measures.

We use the inverse of linearity as a tortuosity measure, and as Table 1 shows, this provides reasonable results. However, global non-linearity does not exactly capture the concept of tortuosity, since curved blood vessels are not considered to have high tortuosity as long as they are smooth. Therefore, we also consider a more local version of our linearity measure in which the mean is taken of (6) after it has been computed over all (overlapping) local sections of a given length.[2] More precisely, let $\mathbf{C}_{i,j}$ denote the section of discretely sampled

[1]In brief, the tortuosity measures in table 1 are defined as follows. $\mathcal{L}$: our proposed global linearity measure, $\mathcal{L}_{10}$ $\mathcal{L}_{22}$ DCI: integral of squared deviation of curvature, ICM: number of inflection points, MAC: mean angle change, tc: integral of absolute curvature, TN: number of times the local change in direction $\geq \frac{\pi}{6}$, tsc: integral of squared curvature, τ: decomposes a curve into sections at turns and measures how much each turn is different from a smooth curve. In addition, the following are used: L_c: curve length, L_x: distance between end points of curve.

[2]The average lengths of the arteries/veins were 980/890 pixels. We have applied linear interpolation to the sparse and irregular centreline points of the vessels to provide a denser and regular sampling with a 10

curve between $\mathbf{C}(i)$ and $\mathbf{C}(j)$. Given a uniformly sampled curve $\mathbf{C}(i); i = 1 \ldots n$, and using all $n - w$ local sections of length w pixels, the mean local linearity is denoted by $\mathcal{L}_w$, and calculated as

$$\mathcal{L}_w = \frac{1}{n-w} \sum_{s=1}^{n-w} \mathcal{L}(\mathbf{C}_{s,s+w}).$$

This substantially improves their correlation values, which are now ranked in the top two or three. Arteries and veins appear to have different characteristics, and so best results are obtained for each using different sub-section lengths (22 for arteries as compared to 10 for veins).

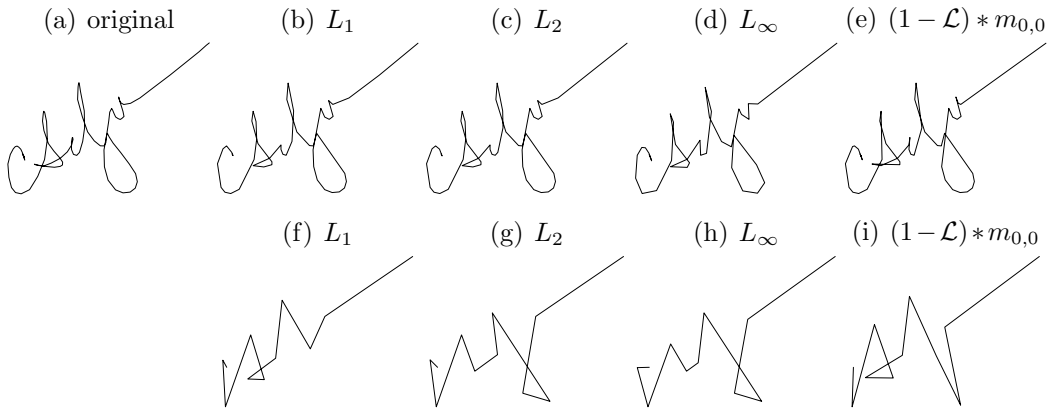


Figure 6: A signature with its optimal polygonal approximations using various error functions at two scales: (b)-(e) 50 lines, (f)-(i) 10 lines.

A second application of the linearity measure is now given, for polygonal approximation. Dynamic programming was used to produce the optimal solution of partitioning a discrete curve into a given number of segments [20].[3]

Rather than use a standard error measure for each segment such as the summed shortest distances between points along the curve segment and the straight line segment connecting its end points we shall use linearity. More specifically, to convert it into an appropriate error term to be minimised we calculate the error for each curve segment as $(1-\mathcal{L}(\mathbf{C}_{ij})) \cdot m_{0,0}(\mathbf{C}_{ij})$, where $\mathbf{C}_{ij}$ denotes the section of curve between $\mathbf{C}(i)$ and $\mathbf{C}(j)$. That is, for a section of curve $\mathbf{C}_{ij}$, $(1 - \mathcal{L}(\mathbf{C}_{ij}))$ measures its deviation from linearity which is to be minimised to create a good polygonal approximation. The total error of a polygonal approximation is computed as the sum of the linearity errors for each curve segment which are weighted by the length of the curve segment ($m_{0,0}$). This is shown in Figure 6 where the optimal polygonal approximation

pixel spacing.

[3]If the total error can be decomposed into the sum of local errors then a polygonal approximation can be made by generating a table T_{ij} containing the approximation errors for replacing each segment of curve $\mathbf{C}(i) \rightarrow \mathbf{C}(j)$ by a straight line segment through $\mathbf{C}(i)$ and $\mathbf{C}(j)$. Dynamic programming can then be used to efficiently determine the globally optimal polygonal approximation which corresponds to the minimal cost path through the table.

of a signature is computed as described; note that this data is not a simple curve, but it contains several self intersections. For comparison, approximations are also given using the standard L_1, L_2 and L_∞ errors norms on the distances between the curve segment and the corresponding straight line segment. Another example is demonstrated in Figure 7. It can be seen that there are differences between the polygonal approximations produced by the various error terms, although they are relatively small except for very coarse approximations.

4 Linearity Plot

The linearity measure $\mathcal{L}(\mathbf{C})$, as defined here, provides an extension to a richer description of open curve segments. Instead of assigning a single number (e.g. linearity measure $\mathcal{L}(\mathbf{C})$) to the given shape $\mathbf{C}$, we will assign a function, called here the *linearity plot*. The idea is to compute linearity incrementally, i.e. to compute linearity of sub-segments of $\mathbf{C}$ determined by the start-point $(x(0), y(0))$ of $\mathbf{C}$ and another point $(x(s), y(s)) \in \mathbf{C}$ which moves from the beginning $(x(0), y(0))$ of $\mathbf{C}$ to the end $(x(1), y(1))$ of $\mathbf{C}$, as s varies from 0 to 1. The linearity plot $\mathrm{P}(\mathbf{C})$, associated with the given curve $\mathbf{C}$ is formally defined as follows.

Definition 2 *Let* $\mathbf{C}$ *be a curve given in an arc-length parametrisation:* $x = x(s)$, $y = y(s)$, *and* $s \in [0, 1]$. *Let* $\mathbf{A}(s)$ *be the part of the curve* $\mathbf{C}$ *bounded by the starting point* $(x(0), y(0))$ *and by the point* $(x(s), y(s)) \in \mathbf{C}$. *Then the linearity plot* $\mathrm{P}(\mathbf{C})$ *is defined by:*

$$\mathrm{P}(\mathbf{C}) = \{(s, \mathcal{L}(\mathbf{A}(s)) \mid \ s \in [0, 1]\}. \tag{7}$$

Note. $\mathrm{P}(\mathbf{C})$ is actually a graph of the function $\mathcal{L}(\mathbf{A}(s)))$, $s \in (0, 1]$. Since in our applications we work with discrete data, real curves are represented by a certain polygonal approximation [22]. This implies that in our experiments $\mathrm{P}(\mathbf{C})$ starts at the point $(0, 1)$ and ends at the point $(1, \mathcal{L}(\mathbf{C}))$.

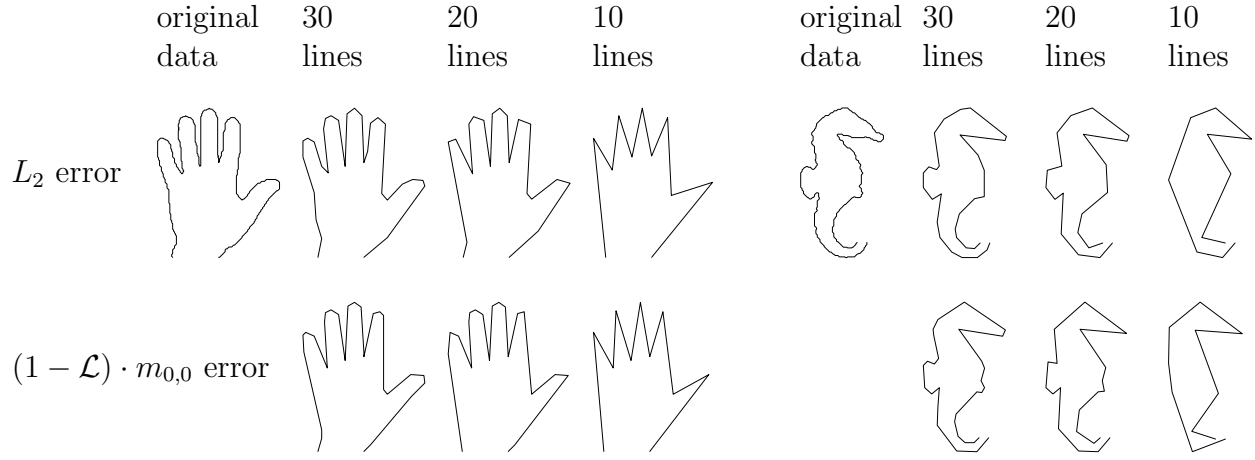


Figure 7: Two shapes with their optimal polygonal approximations using the L_2 error function (top row) and the $(1 - \mathcal{L}) \cdot m_{0,0}$ error function (bottom row). Approximations are determined using 30, 20 and 10 line segments.

We also will use the *reverse linearity plot* $\mathrm{P}_{rev}(\mathbf{C})$ defined as:

$$\mathrm{P}_{rev}(\mathbf{C}) = \{(s, \mathcal{L}(\mathbf{A}_{rev}(1-s)) \mid \ \ s \in [0,1]\}, \tag{8}$$

where $\mathbf{A}_{rev}(1-s)$ is the segment of $\mathbf{C}$ determined by the end point $(x(1), y(1))$ and the point $(x(1-s), y(1-s))$ (which moves from the point $(x(1), y(1))$ to the point $(x(0), y(0))$, along the curve $\mathbf{C}$, as $s' = 1-s$ varies from 1 to 0, i.e. as s varies from 0 to 1). In other words, $\mathrm{P}_{rev}(\mathbf{C})$ is the graph of the function which measures the linearity of $\mathbf{A}_{rev}(s' = 1-s)$.

We also note that the linearity plot differs from well-known *shape signature* descriptors, which are also 'one-dimensional' representations of planar shapes, often used in shape analysis tasks [5, 37].

Now, an additional application is demonstrated, using data from Munich and Perona [19] to perform signature verification. The data consists of pen trajectories for 2911 genuine signatures taken from 112 subjects. In addition, five forgers provided a total of 1061 forgeries across all the subjects. Some examples are provided in Figure 8.

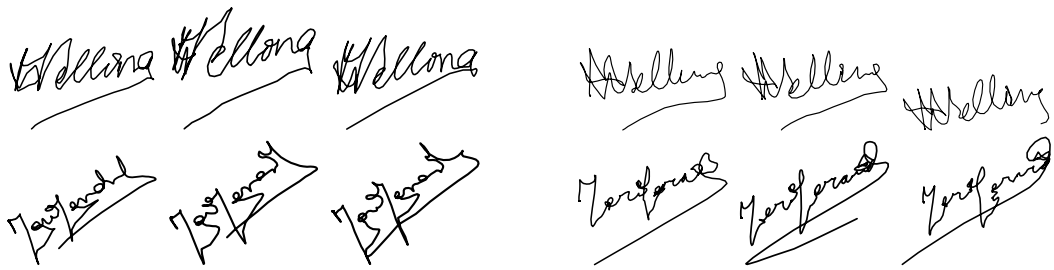
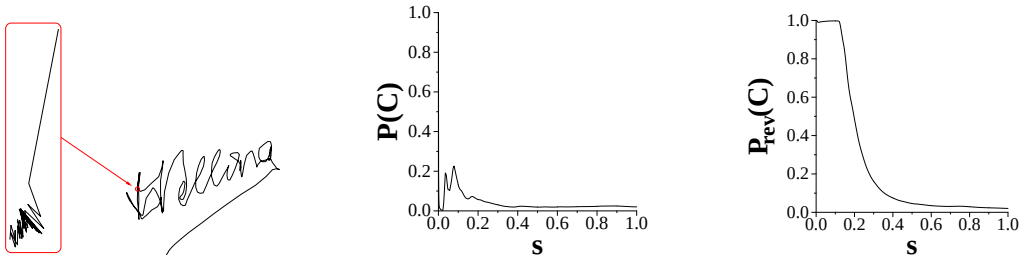


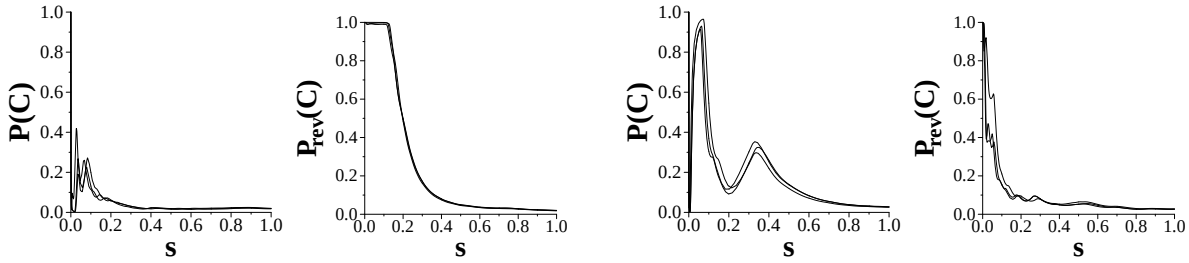
Figure 8: Examples of genuine (leftmost three) and forged (rightmost three) signatures.

In our experiment we have only considered the curve shape, and discounted temporal information (i.e. writing speed) although Munich and Perona found this useful to reject overly slow signatures as forgeries.



(a) (b) (c)

Figure 9: (a) A genuine signature from Fig.8. A magnified start of the signature is inside the rectangle displayed. (b) The linearity plot P($\mathbf{C}$) of the signature obtained by forward traversal of the curve. (c) The reverse linearity plot $\mathrm{P}_{rev}(\mathbf{C})$ obtained by reverse traversal of the curve.



(a) (b) (c) (d)

Figure 10: (a & b): Linearity and reverse linearity plots $\mathrm{P}(\mathbf{C})$ and $\mathrm{P}_{rev}(\mathbf{C})$ of three genuine signatures in Figure 8 (first row). (c & d): Linearity and reverse linearity plots $\mathrm{P}(\mathbf{C})$ and $\mathrm{P}_{rev}(\mathbf{C})$ of three forged signatures in Fig.8 (first row).

The quality of match between signatures $\mathbf{C}_1$ and $\mathbf{C}_2$ is measured by the similarity between the linearity plots $\mathrm{P}(\mathbf{C}_1)$ and $\mathrm{P}(\mathbf{C}_2)$. This similarity is measured by the area bounded by the linearity plots $\mathrm{P}(\mathbf{C}_1)$ and $\mathrm{P}(\mathbf{C}_2)$ and by the vertical lines $s = 0$ and $s = 1$. We achieve better results by also consider traversing the curve in reverse direction, i.e. by considering the reverse linearity plots $\mathrm{P}_{rev}(\mathbf{C}_1)$ and $\mathrm{P}_{rev}(\mathbf{C}_2)$. The similarity of two signatures $\mathbf{C}_1$ and $\mathbf{C}_2$ is now determined as

$$\min\left\{\text{area}\left(\mathrm{P}(\mathbf{C}_1), \mathrm{P}(\mathbf{C}_2)\right), \text{area}\left(\mathrm{P}_{rev}(\mathbf{C}_1), \mathrm{P}_{rev}(\mathbf{C}_2)\right)\right\}.$$

Of course, there are many other possibilities to measure the similarity between $\mathrm{P}(\mathbf{C}_1)$ and $\mathrm{P}(\mathbf{C}_2)$, i.e. between $\mathrm{P}_{rev}(\mathbf{C}_1)$ and $\mathrm{P}_{rev}(\mathbf{C}_2)$.

Figure 9 shows the linearity plots for one signature. It can be seen that the two plots, obtained by traversing the curve from either end, provide useful complementary information. The linearity plot that starts from the letter H generally has a low value due to the complex shape of the signature (see Figure 9(b)). In addition, as the magnified inset in Figure 9(a) shows, there is an erratic wiggle containing about fifteen low magnitude oscillations at the start of the curve, which appears to by a systematic artifact produced by the data acquisition process (probably created when the pen is initially at rest). This non-linear wiggle causes the initial sharp drop (close to 0) in linearity values in Figure 9(b). When the curve is traversed from the other end there is no starting artifact, and the long linear flourish underneath the name causes a significant high peak in the corresponding linearity plot (see Figure 9(c)). Figure 10 shows the linearity plots for the six signatures in the top row of Figure 8. The linearity plots for the three genuine signatures appear significantly more similar in Figure 10(b) than either of the two linearity plots for the three forged signatures in Figure 10(c) & (d).

Nearest neighbour matching is then performed on all the data using the leave-one-out strategy. Signature verification is a two class (genuine or fake) problem. Since the identity of the signature is already known, the nearest neighbour matching is only applied to the set of genuine and forged examples of the subject's signature. We achieved 93.1% accuracy on this task, which given the close similarities between the genuine and forged signatures, is reasonable. Munich and Perona [19] report an accuracy of around 99%; however their system was specifically designed for signature analysis, and used temporal information in addition to shape.

5 Conclusion

A new linearity measure for open curve segments has been defined.[4] We start by considering the average square distance between the points on the curve and show that such quantity is maximised by the straight line segments. By exploiting such a nice property we have defines a new linearity measure and prove several properties on it. The new measure ranges over the interval $(0, 1]$, and produces the value 1 if and only if the measured open line is a straight line segment. It is invariant with respect to similarity transformation, and can be efficiently and simply computed using line moments. The new measure has a clear geometric meaning, and because of this, the behaviour of the new measure can be well understood and predicted to some extent, prior to the application planned to be performed.

A variety of experiments demonstrate its effectiveness for measuring tortuosity of blood vessels for indicating retinal pathologies, and use as the error function for computing polygonal approximations. The linearity measure can be extended to give a linearity plot, which provides a richer descriptor of open curve segments. Matching curves based on matching linearity plots was demonstrated on the application of identifying forged signatures.

It would be straightforward to extend the current 2D linearity measure to 3D. This measure could also be applied to in a similar way to all the applications demonstrated in this paper, namely: 3D polygonal approximation [23], classification of 3D blood vessels [2], and detection of forged signatures, using temporal dynamics as an additional dimension [19].

References

[1] S. Benhamou. How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension. *Journal of Theoretical Biology*, 229(2):209–220, 2004.

[2] E. Bullitt, G. Gerig, S.M. Pizer, W. Lin, and S.R. Aylward. Measuring tortuosity of the intracerebral vasculature. *IEEE Trans. Med. Imaging*, 22(9):1163–1171, 2003.

[3] C.-C. Chen. Improved moment invariants for shape discrimination. *Pattern Recognition*, 26(5):683–686, 1993.

[4] D. Coeurjolly and R. Klette. A comparative evaluation of length estimators of digital curves. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 26(2):252–257, 2004.

[5] A. El-ghazal, O. Basir, and S. Belkasim. Farthest point distance: A new shape signature for Fourier descriptors. *Signal Processing: Image Communication*, 24(7):572–586, 2009.

[6] S. Escalera, A. Fornés, O. Pujol, J. Lladós, and P. Radeva. Circular blurred shape model for multiclass symbol recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(2):497–506, 2011.

[7] T. Gautama, D.P. Mandić, and M.M. Van Hull. A novel method for determining the nature of time series. *IEEE Transactions on Biomedical Engineering*, 51(5):728–736, 2004.

[4] Source code is available at http://users.cs.cf.ac.uk/Paul.Rosin.

[8] T. Gautama, D.P. Mandić, and M.M. Van Hulle. Signal nonlinearity in fMRI: A comparison between BOLD and MION. *IEEE Transactions on Medical Images*, 22(5):636–644, 2003.

[9] E. Grisan, M. Foracchia, and A. Ruggeri. A novel method for the automatic grading of retinal vessel tortuosity. *IEEE Trans. Med. Imaging*, 27(3):310–319, 2008.

[10] M. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory*, 8(2):179–187, 1962.

[11] M.N. Huxley. Exponential sums and lattice points III. *Proc. London Math. Soc.*, 87(3):591–609, 2003.

[12] A.R. Imre. Fractal dimension of time-indexed paths. *Applied Mathematics and Computation*, 207(1):221–229, 2009.

[13] R. Kakarala. Testing for convexity with Fourier descriptors. *Electronics Letters*, 34(14):1392–1393, 1998.

[14] R. Klette and A. Rosenfeld. *Digital Geometry*. Morgan Kaufmann, San Francisko, 2004.

[15] R. Klette and J. Žunić. Digital approximation of moments of convex regions. *Graphical Models and Image Processing*, 61(5):274–298, 1999.

[16] S. Manay, D. Cremers, B.-W. Hong, A. Yezzi, and S. Soatto. Integral invariants for shape matching. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(10):1602–1618, 2006.

[17] R. Melter, I. Stojmenović, and J. Žunić. A new characterization of digital lines by least square fits. *Pattern Recognition Letters*, 14(2):83–88, 1993.

[18] W. Mio, A. Srivastava, and S.H. Joshi. On shape of plane elastic curves. *Int. Journal of Computer Vision*, 73(3):307–324, 2007.

[19] M.E. Munich and P. Perona. Visual identification by signature tracking. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 25(2):200–217, 2003.

[20] J.C. Perez and E. Vidal. Optimum polygonal approximation of digitized curves. *Pattern Recognition Letters*, 15(3):743–750, 1994.

[21] E. Rahtu, M. Salo, and J. Heikkilä. A new convexity measure based on a probabilistic interpretation of images. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 28(9):1501–1512, 2006.

[22] P.L. Rosin. Techniques for assessing polygonal approximations of curves. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(6):659–666, 1997.

[23] P.L. Rosin and G.A.W. West. Non-parametric segmentation of curves into various representations. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 17:1140–1153, 1995.

[24] C. Di Ruberto and A. Dempster. Circularity measures based on mathematical morphology. *Electronics Letters*, 36(20):1691–1693, 2000.

[25] M.P. Segundo, L. Silva, O.R.P. Bellon, and C.C. Queirolo. Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(5):1319–1330, 2010.

[26] X. Shu and X.-J. Wu. A novel contour descriptor for 2d shape matching and its application to image retrieval. *Image and Vision Computing*, 20(4):286–294, 2011.

[27] N. Sladoje and J. Lindblad. High precision boundary length estimation by utilizing gray-level information. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 31(2):357–363, 2009.

[28] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. PWS, 1998.

[29] M. Stojmenović and A. Nayak. Measuring linearity of ordered point sets. *PSIVT 2007, Lecture Notes in Computer Science*, 4872:274–288, 2007.

[30] M. Stojmenović, A. Nayak, and J. Žunić. Measuring linearity of planar point sets. *Pattern Recognition*, 41(8):2503–2511, 2008.

[31] M. Stojmenović and J. Žunić. Measuring elongation from shape boundary. *Journal Mathematical Imaging and Vision*, 30(1):73–85, 2008.

[32] V. Venkatraghavan, M. Agarwal, and A.K. Roy. Shape orientation pattern. *IEEE Signal Processing Letters*, 16(8):711–714, 2009.

[33] J. Žunić, K. Hirota, and P.L. Rosin. A Hu moment invariant as a shape circularity measure. *Pattern Recognition*, 43(1):47–57, 2010.

[34] J. Žunić, L. Kopanja, and P.L. Rosin. On the orientability of shapes. *IEEE Transactions on Image Processing*, 15(11):3478–3487, 2006.

[35] J. Žunić and C. Martinez-Ortiz. Linearity measure for curve segments. *Applied Mathematics and Computation*, 215(8):3098–3105, 2009.

[36] J. Žunić and P.L. Rosin. Rectilinearity meaurements for polygons. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 25(9):1193–3200, 2003.

[37] D. Zhang and G. Lu. Study and evaluation of different Fourier methods for image retrieval. *Image and Vision Computing*, 23(1):3349, 2005.