Measuring Linearity of Closed Curves and Connected Compound Curves

Paul L. Rosin[1], Jovanka Pantović[2], and Joviša Žunić[3]

[1] Cardiff University, School of Computer Science, Cardiff CF24 3AA, UK
[2] University of Novi Sad, Faculty of Technical Sciences, 21000 Novi Sad, Serbia
[3] University of Exeter, Computer Science, Exeter EX4 4QF, UK
Paul.Rosin@cs.cf.ac.uk, pantovic@uns.ac.rs, J.Zunic@ex.ac.uk

Abstract. In this paper we define a new linearity measure for closed curves. We start with simple closed curves which represent the boundaries of bounded planar regions. It turns out that the method can be extended to closed curves which self-intersect and also to certain configurations consisting of several curves, including open curve segments. In all cases, the measured linearities range over the interval $(0, 1]$, and do not change under translation, rotation and scaling transformations of the considered curve. In addition, the highest possible linearity (which is 1) is reached if and only if the measured curve consists of two overlapping (i.e. coincident) straight line segments. The new linearity measure is theoretically well founded and all related statements are supported with rigorous mathematical proofs.

1 Introduction

There are many ways to quantitatively characterise the shape of objects. Because of that, shape based object characteristics are in frequent use for object discrimination in different domains (medicine, biology, robotics, astrophysics, etc). By a shape descriptor we mean a shape-based object characteristic (e.g. compactness, elongation, etc) which allows a numerical characterisation. A certain method used for the computation of a given shape descriptor/characteristic is called here a shape measure. Several different shape measures can be assigned to a certain shape descriptor. This is because none of the shape measures is expected to outperform all the others in all applications. Measures performing well in some application could perform worse in another.

In this paper we deal with the linearity of closed curves. Initially, we were looking for a quantity, computed from the shape's boundary, which should indicate the degree to which the shape observed is linear (i.e. similar to a straight line segment). Once we developed a method for the computation of such a quantity, it has turned out that the method can be applied successfully to a wider class of curves – not just to the simple closed curves, which represent the boundaries/frontiers of planar regions.

Several linearity measures are already considered in the literature [1–3]. But they are mainly related to open curve segments. I.e. they measure how much

K.M. Lee et al. (Eds.): ACCV 2012, Part III, LNCS 7726, pp. 310–321, 2013.

an open curve segment differs from a perfect straight line segment. Generally speaking, each of these measures can be applied to closed curves, treating every closed curve as an open curve whose end points coincide. The problem is that such computed linearities might not reflect whether the structure of the observed shape is linear or not. We give two examples.

- The *straightness index* [4], denoted by $\mathcal{I}_{open}(\mathcal{C})$, is perhaps the simplest and the most natural linearity measure for open curve segments. It is defined as the ratio of the distance between the curve end points and the length of the curve. Obviously, this measure is very simple and fast to compute. Also it gets the highest possible value 1 if and only if the curve is a straight line segment. But the straightness index gives the value zero for all closed curves, independently on the choice of the start/end break point on the curve.
- A recent measure $\mathcal{S}(\mathcal{C})$, from [3], defines the linearity of open curve segments considering the distance among all the pairs of curve points (not only between the start and end point as the straight index does). Formally, for a given curve $\mathcal{C}$, given in an arc-length parametrisation $x = x(s)$, $y = y(s)$, $s \in [0, 1]$, and positioned such that the centroid of $\mathcal{C}$ coincides with the origin, the linearity measure $\mathcal{S}(\mathcal{C})$ is defined by

$$\mathcal{S}(\mathcal{C}) \;=\; 12 \cdot \int_{\mathcal{C}} \left(x(s)^2 + y(s)^2\right) \; ds. \tag{1}$$

  As it has been proven in [3], the linearity $\mathcal{S}(\mathcal{C})$ equals 1 if and only if $\mathcal{C}$ is a straight line segment, and is invariant with respect to similarity transformations. If applied to a closed curve $\mathcal{C}$, the measure $\mathcal{S}(\mathcal{C})$ has the desirable property that it does not depend (see (1)) on the choice of the breaking (start/end) point. But the problem is that $\mathcal{S}(\mathcal{C})$ does not behave as desired if applied to closed curves. Here is an illustration. Let us define a family of rectangles $\mathcal{R}(t)$ as follows

  Let $t \in (0, 0.25]$. $\mathcal{R}(t)$ *is a rectangle whose edges have length* t *and* $0.5 - t$. (2)

  The equality $\mathcal{S}(\mathcal{R}(t)) = 1/4$ easily follows from (1), for all $t \in (0, 0.25]$. Notice that $\mathcal{R}(t = 0.25)$ is a square, and as t decreases $\mathcal{R}(t)$ becomes a more and more elongated rectangle. Thus, we wish to obtain increasing linearities as $t \to 0$, but this does not happen. So, if $\mathcal{S}(\mathcal{C})$ is applied to closed curves, it would not distinguish among rectangles whose edge ratio differs, which is not a desirable property for a linearity measure.

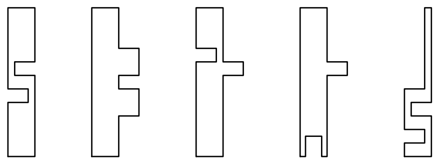
It is worth mentioning that there is a simple and easy way to measure the linearity measure for closed curves and avoid the disadvantages mentioned above. Indeed, we can define the linearity measure $\mathcal{I}_{closed}(\mathcal{C})$ based on the ratio of the *curve diameter* (the longest distance among curve points [5]) and the curve perimeter:

$$\mathcal{I}_{closed}(\mathcal{C}) \;=\; 2 \cdot \frac{diameter\_of\_\mathcal{C}}{perimeter\_of\_\mathcal{C}}. \tag{3}$$

$\mathcal{I}_{closed}(\mathcal{C})$ extends the idea of the straightness index measure to closed curves. The following desirable properties are satisfied by $\mathcal{I}_{closed}(\mathcal{C})$ measure.

(p1) $\mathcal{I}_{closed}(\mathcal{C})$ ranges over the interval $(0, 1]$.
(p2) $\mathcal{I}_{closed}(\mathcal{C})$ takes value 1 if and only if the measured curve consists of two overlapping (i.e. coincident) straight line segments.
(p3) $\mathcal{I}_{closed}(\mathcal{C})$ is invariant with respect to translations, rotations and scaling.
(p4) $\mathcal{I}_{closed}(\mathcal{C})$ is easy to compute.

An obvious drawback of $\mathcal{I}_{closed}(\mathcal{C})$ is that it depends only on the longest distance between a pair of the curve points. For example, all the closed curves in Fig.1 have the same $\mathcal{I}_{closed}(\mathcal{C})$ linearity.



0.698 0.712 0.738 0.740 0.798

Fig. 1. The value of $\mathcal{I}_{closed}(\mathcal{C})$ is 0.699 for all the shapes, but the proposed measure $\mathcal{L}_{cl}(\mathcal{C})$ produces different linearities (shown below)

In this paper we define a new measure for closed curves. The new measure satisfies the above properties **(p1), (p2), (p3)** and **(p4)** but also takes into account the distribution of all the shape points (not only these on the longest mutual distance) since the shape centroid is used for the measure computation. An extension to a very general class of curve configurations is possible as well. Such a generalised measure can be applied to open curve segments, keeping the basic requirements satisfied, and also to the configurations which are unions of certain sets of open curve segments. This is particularly suitable when estimating the linearity of an object based on the linearity of its skeleton (which usually can be represented by such configurations).

2 Definitions and Denotations

In this section we introduce the basic definitions and notation used in this paper.

– As usual, $\mathbf{d}_2(A, B) = \mathbf{d}_2((x, y), (u, v)) = \sqrt{(x-u)^2 + (y-v)^2}$ denotes the Euclidean distance between the points $A = (x, y)$ and $B = (u, v)$.
– $Per(\mathcal{C})$ denotes the Euclidean perimeter of a given curve $\mathcal{C}$.
– $diam(\mathcal{C})$ denotes the *diameter* of a given curve and equals the longest distance between two curve points. I.e.,

$$diam(\mathcal{C}) = \max_{X,Y \in \mathcal{C}} \{\mathbf{d}_2(X, Y)\}.$$

Without loss of generality, throughout the paper it will be assumed (even if not mentioned) that every curve $\mathcal{C}$ is given in an arc-length parametrisation.

$$x = x(s), \quad y = y(s), \quad \text{where} \quad s \in [0, Per(\mathcal{C})]. \tag{4}$$

The parameter s measures the distance between the points $(x(0), y(0))$ and $(x(s), y(s))$ along the curve $\mathcal{C}$.

Initially, we will focus on *closed curves*. They will also be given in an arc-length parametrisation (4) but will satisfy the following additional condition:

$$(x(0), y(0)) = (x(Per(\mathcal{C})), y(Per(\mathcal{C}))). \tag{5}$$

We will say that $(x(0), y(0))$ is the *curve start point* and that $(x(Per(\mathcal{C})), y(Per(\mathcal{C})))$ is the *curve end-point*, even if they coincide (in the case of closed curves).

Notice that even being simple, the definition (4) covers a wide spectrum of curves. The most typical situation is when $\mathcal{C}$ represents the boundary of a planar shape (see Fig.2(a)). Such curves, as usual, will be called *simple closed curves*. Curves which intersect themselves (e.g. the curve in Fig.2(b)) or even curves whose sub-sections overlap (e.g. the curve in Fig.2(c)) can also be given in the form (4). The latter also includes the curve whose two halves overlap.

Here we define such a curve $\mathcal{D}(p)$, for $p > 0$, consisting of two identical straight line segments, both of the length $p/2$, but still representable in the form of (4). So, one possibility for the arc-length parametrisation of $\mathcal{D}(p)$ is:

$$\mathcal{D}(p) \; : \; \begin{cases} x = x(s) = s, & y = y(s) = 0, \; s \in [0, \; p/2] \\ x = x(s) = p - s, & y = y(s) = 0, \; s \in [p/2, \; p]. \end{cases} \tag{6}$$

Because $\mathcal{D}(p)$ allows a parametrisation as above, it will be treated as a closed curve (but not as a simple closed curve). Of course, if displayed, see Fig.2(d), it looks like a single line straight line segment of length $p/2$, but it will not cause any confusion. If a similar reasoning is applied, we see that any open curve segment Fig.2(e) (not necessarily a straight line) can be treated as a closed curve whose halves overlap, and this will be done in this paper. Fig.2(f) displays a configuration consisting of three “connected” line segments, but the new linearity measure can also be applied to such configurations. They will be called *connected compound curves* and also can be treated as closed curves. Detailed explanation and formal definitions are in Section 4.

We note here that the straight line segment, treated as a closed curve, will have the highest linearity, measured by the new linearity measure introduced by this paper. As mentioned this might be understood as a natural preference. Also, there are simple closed curves whose measured linearities are arbitrarily close to 1, which is also a reasonable requirement.

The centroid of a given closed curve $\mathcal{C}$, with unit perimeter, will be denoted by $(x_{\mathcal{C}}, y_{\mathcal{C}})$ and is the average values of the coordinates of all the curve points. In several situations, we will assume that a given curve $\mathcal{C}_a$ is scaled to be of the unit length.

3 Linearity Measure for Closed Curves

In this section we define a new linearity measure for closed curves. We start with the following theorem whose results motivate the definition of the new linearity measure.

Theorem 1. *Let a closed curve $\mathcal{C}$ be given in an arc-length parametrisation $x = x(s)$, $y = y(s)$, $s \in [0, Per(\mathcal{C})]$, and let $A = (x_0, y_0)$ be the point of $\mathcal{C}$ furthest from the centroid $(x_\mathcal{C}, y_\mathcal{C})$ of $\mathcal{C}$. Then*

(a)

$$\sqrt{(x_0 - x_\mathcal{C})^2 + (y_0 - y_\mathcal{C})^2} \leq \frac{1}{4} \cdot Per(\mathcal{C}); \quad (7)$$

(b) *The upper bound in (7) is the best possible (i.e. it cannot be improved).*

Proof. Since the quantity $\frac{\sqrt{(x_0 - x_\mathcal{C})^2 + (y_0 - y_\mathcal{C})^2}}{Per(\mathcal{C})}$ is invariant with respect to scaling transformations, without loss of generality, we can assume $Per(\mathcal{C}) = 1$. I.e. $\mathcal{C}$ is given by

$$x = x(s), \quad y = y(s), \quad s \in [0, 1]. \quad (8)$$

(a) Let $\mathcal{C}$ and $A = (x_0, y_0) \in \mathcal{C}$ satisfy the conditions in the presumptions of the theorem. Let us choose the coordinate system such that the origin coincides with the point $A = (x_0, y_0)$ and the x-axis passes through the centroid $C = (x_\mathcal{C}, y_\mathcal{C})$ of $\mathcal{C}$ (notation is illustrated in Fig.3). In such a way

$$y_\mathcal{C} = \int_\mathcal{C} y(s) = 0 \quad (9)$$

is provided. Since A is furthest from the centroid C it implies that the curve $\mathcal{C}$ lies inside circular disc radii $\mathbf{d}_2(A, C)$ centred at C. Also, let us choose an arc-length parametrisation of $\mathcal{C}$ such that

$$x = x(s), \; y = y(s), \; \text{and} \; A = (x(0), y(0)) = (x(1), y(1)). \quad (10)$$

Further, consider the function $F(a) = \int_{s=0}^{a} x(s) \, ds, \;\; a \in [0, 1]$ where $x = x(s)$ is as in (10).

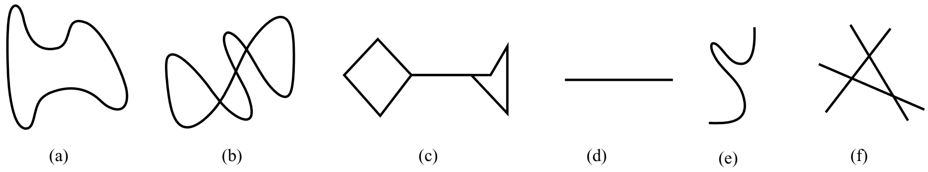


Fig. 2. Several curve examples which will be treated here as closed curves

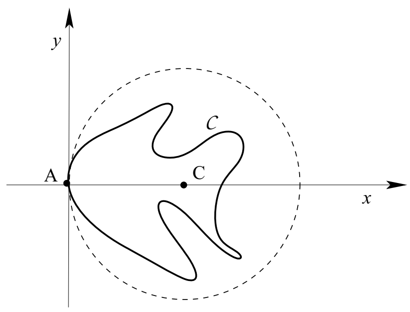


Fig. 3. A is a point from the curve $\mathcal{C}$ which is furthest from the curve centroid C

$F(a)$ is non-decreasing function because $x = x(s)$ is non-negative function (due to the choice of the coordinate system). So, there is $a_0 \in [0,1]$ such that $F(a_0) = \frac{1}{2} \cdot F(a = 1)$; e.g. there is $a_0 \in [0,1]$ such that

$$\int_{s=0}^{a_0} x(s)\ ds \ = \ \int_{s=a_0}^{1} x(s)\ ds \ = \ \frac{1}{2} \cdot \int_{s=0}^{1} x(s)\ ds. \tag{11}$$

Now, since $x(s) \leq s$, we obtain

$$\int_{s=0}^{1} x(s)\ ds \ = \ 2 \cdot \int_{s=0}^{a_0} x(s)\ ds \ \leq \ 2 \cdot \int_{s=0}^{a_0} s\ ds = a_0^2, \tag{12}$$

and similarly, since $x(s) \leq 1 - s$,

$$\int_{s=0}^{1} x(s)\ ds \ = \ 2 \int_{s=a_0}^{1} x(s)\ ds \ \leq \ 2 \cdot \int_{s=a_0}^{1} (1-s)\ ds = (1-a_0)^2. \tag{13}$$

Finally, the just derived (12) and (13) give

$$\int_{s=0}^{1} x(s)\ ds \ \leq \ \min\{a_0^2,\ (1-a_0)^2\} \ \leq \ \frac{1}{4} \tag{14}$$

(the last inequality follows because of $a_0 \in [0,1]$)).
Taking into account that the choice of the coordinate system enables (9) and the above estimate (14), we have proven (**a**).

(b) The statement follows because any closed curve $\mathcal{D}(p)$, $(p > 0)$, defined as in (6), reaches the upper bound in (7). I.e., the centroid of $\mathcal{D}(p)$ is $(p/4, 0)$ and the point $(0,0)$ (also $(p/2, 0)$) is furthest from the centroid of $\mathcal{D}(p)$. (Notice $Per(\mathcal{D}(p)) = p$ since $\mathcal{D}(p)$ consists of two overlapping straight line segments both having length $p/2$). □

Note 1. The inequality in (7) is strict for simple closed curves. This can be deduced from (14). Indeed, $\int_{s=0}^{1} x(s)\ ds \ = \ \min\{a_0^2,\ (1-a_0)^2\} \ = \ \frac{1}{4}$ would imply $a_0 = 1/2$ and (see (12) and (13))

$$\int_0^{a_0=1/2} x(s)\ ds \ = \ \frac{1}{2} \cdot a_0^2 \ = \ \int_{a_0=1/2}^{1} x(s)\ ds \ = \ \frac{1}{2} \cdot (1-a_0)^2 \ = \ \frac{1}{8}.$$

Since $x(s) \le s$, in order to have $\int_0^{a_0=1/2} x(s)\ ds \ = \ \int_{a_0=1/2}^{1} x(s)\ ds \ = \ \frac{1}{8}$, it must be $x(s) = s$ for $s \in [0, 1/2]$ and similarly $x(s) = 1 - s$ for $s \in [1/2, 1]$. I.e. both sub-arcs of $\mathcal{C}$ must be straight line segments, which implies that $\mathcal{C}$ cannot be a simple closed curve.

Note 2. The upper bound in (7) cannot be improved for simple closed curves, even though none of simple closed curves satisfies $\sqrt{(x_0 - x_{\mathcal{C}})^2 + (y_0 - y_{\mathcal{C}})^2} \ = \ \frac{1}{4} \cdot Per(\mathcal{C})$, as it has been shown in the previous note. This follows from the fact that for any $\delta > 0$ there is a simple closed curve $\mathcal{C}(\delta)$ satisfying $\left|\sqrt{(x_0 - x_{\mathcal{C}(\delta)})^2 + (y_0 - y_{\mathcal{C}(\delta)})^2} \ - \ \frac{1}{4} \cdot Per(\mathcal{C}(\delta))\right| < \delta$. Indeed, the required $\mathcal{C}(\delta)$ can be selected from a family of rectangles $\mathcal{R}(t)$ (defined as in (2)), since $\lim_{t\to\infty} \frac{\sqrt{(x_0 - x_{\mathcal{R}(t)})^2 + (y_0 - y_{\mathcal{R}(t)})^2}}{Per(\mathcal{R}(t))} \ = \ \frac{1}{4}$.

Now, by arguments of the previous theorem we give the following definition for a new linearity measure of closed curves.

Definition 1. *Let $\mathcal{C}$ be a closed curve $x = x(s)$, $y = y(s)$, $s \in [0, Per(\mathcal{C})]$, and let $A = (x_0, y_0)$ be the point of $\mathcal{C}$ furthest from the centroid $\mathbf{C}_{\mathcal{C}} = (x_{\mathcal{C}}, y_{\mathcal{C}})$ of $\mathcal{C}$. Then the linearity $\mathcal{L}_{cl}(\mathcal{C})$ of $\mathcal{C}$ is defined as*

$$\mathcal{L}_{cl}(\mathcal{C}) \ = \ 4 \cdot \frac{\sqrt{(x_0 - x_{\mathcal{C}})^2 + (y_0 - y_{\mathcal{C}})^2}}{Per(\mathcal{C})} \ = \ 4 \cdot \frac{\mathbf{d}_2(A, \mathbf{C}_{\mathcal{C}})}{Per(\mathcal{C})}. \tag{15}$$

Properties of the linearity measure $\mathcal{L}_{cl}(\mathcal{C})$ are listed in the statement of the next theorem.

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

- **(i)** $\mathcal{L}_{cl}(\mathcal{C}) \in (0, 1]$, *for all closed curves $\mathcal{C}$;*
- **(ii)** $\mathcal{L}_{cl}(\mathcal{C}) \ = \ 1 \quad \Leftrightarrow \quad \mathcal{C} = \mathcal{D}(p)$, *for some $p > 0$ (i.e., $\mathcal{C}$ consists of two overlapping straight line segments);*
- **(iii)** $\mathcal{L}_{cl}(\mathcal{C})$ *is invariant with respect to similarity transformations.*

Proof. Item **(i)** is a direct consequences of Theorem 1.

The proof of **(ii)** is actually given in Note 1.

Translations and rotations do not change either the curve length and the distance between the centroid and the curve points. Also, $\frac{\mathbf{d}_2(Q, C_{\mathcal{C}})}{Per(\mathcal{C})}$ is an obvious scaling invariant. This proves **(iii)**. □

4 Connected Compound Curves

We now demonstrate how the idea applied in the previous section can be further developed and used to define a method for measuring the linearity of connected compound curves. We give formal definitions of a connected compound curve, its centroid, and a measure for linearity of connected compound curves, followed by a theorem which summarises the basic properties of such a linearity measure for compound curves. The basic theoretical observations (analogous to those in Theorem 1) together with the proof of Theorem 3 are omitted, due to lack of space, and will be given in the authors' forthcoming papers.

Definition 2. *Let $\mathcal{C}_1,\ \mathcal{C}_2,\ \ldots\ \mathcal{C}_n$ be curve segments given in an arc-length parametrisation form*

$$\mathcal{C}_i: \quad x = x_i(s), \ \ y = y_i(s), \quad s \in [0, l_i], \quad \text{for } \ 1 = 1, 2, \ldots, n. \tag{16}$$

Also, for any two points P and Q from $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_n$ let there exist a connected path consisting of sub-arcs of curves $\mathcal{C}_1,\ \mathcal{C}_2,\ \ldots,\ \mathcal{C}_n$. Then the union

$$\mathcal{C} \ = \ \mathcal{C}_1 \ \cup \ \mathcal{C}_2 \ \cup \ldots \ \cup \ \mathcal{C}_n$$

is said to be a connected compound curve.

The total-length $\mathcal{T}(\mathcal{C})$ of the connected compound curve $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_n$ is defined as the total sum of lengths of the curves $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$. I.e., in accordance with (16)

$$\mathcal{T}(\mathcal{C}) \ = \ l_1 + l_2 + \ldots + l_n$$

The centroid $(x_\mathcal{C}, y_\mathcal{C})$ of the connected compound curve $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \ldots \cup \mathcal{C}_n$ is defined as the point whose coordinates are the average values of the coordinates of all the points which belong to $\mathcal{C}$. I.e., in accordance with (16), the centroid of $\mathcal{C}$ is

$$(x_\mathcal{C}, y_\mathcal{C}) \ = \ \left(\frac{\sum\limits_{i=1}^{n} \int_{\mathcal{C}_i} x_i(s)\ ds}{\sum\limits_{i=1}^{n} l_i}, \ \frac{\sum\limits_{i=1}^{n} \int_{\mathcal{C}_i} y_i(s)\ ds}{\sum\limits_{i=1}^{n} l_i} \right). \tag{17}$$

Now, we give the following definition for a linearity measure for connected compound curves.

Definition 3. *Let $\mathcal{C} = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_n$ be a compound connected curve and let the point $P = (x_0, y_0) \in \mathcal{C}$ be furthest from the centroid $(x_\mathcal{C}, y_\mathcal{C})$ of $\mathcal{C}$. The linearity measure $\mathcal{L}_{comp}(\mathcal{C})$ of $\mathcal{C}$ is defined as*

$$\mathcal{L}_{comp}(\mathcal{C}) = 2 \cdot \frac{\sqrt{(x_0 - x_\mathcal{C})^2 + (y_0 - y_\mathcal{C})^2}}{\mathcal{T}(\mathcal{C})}. \tag{18}$$

Note 3. Definition 3 can be directly applied to a single open curve. The linearity of such an open single curve $\mathcal{C}$ is computed as double the value of the longest distance of a point from $\mathcal{C}$ to the centroid of $\mathcal{C}$.

If Definition 3 is applied to a simple closed curve $\mathcal{C}_{cl}$, then $\mathcal{L}_{comp}(\mathcal{C}_{cl})$ is computed as double the value of the longest distance of a point from $\mathcal{C}_{cl}$ to the centroid of $\mathcal{C}_{cl}$. Indeed, however we split $\mathcal{C}_{cl}$ onto, let say, two arcs $\mathcal{C}_1$ and $\mathcal{C}_2$ and treat $\mathcal{C}_{cl} = \mathcal{C}_1 \cup \mathcal{C}_2$ as a connected compound curve (in the sense of Definition 2), the centroid, the total-length of such a connected compound curve, and the point of $\mathcal{C}_{cl} = \mathcal{C}_1 \cup \mathcal{C}_2$ furthest from the centroid of $\mathcal{C}_{cl} = \mathcal{C}_1 \cup \mathcal{C}_2$ would not vary. So, Definition 3 can be applied directly to the computation of $\mathcal{L}_{comp}(\mathcal{C}_{cl})$.

Since both Definition 3 and Definition 1 can be applied to simple closed curves $\mathcal{C}_{cl}$, it is worth mentioning that the following equality

$$\mathcal{L}_{comp}(\mathcal{C}_{cl}) = \frac{1}{2} \cdot \mathcal{L}_{cl}(\mathcal{C}_{cl}) \tag{19}$$

is true for all simple closed curves $\mathcal{C}_{cl}$. Thus, $\mathcal{L}_{comp}(\mathcal{C}_{cl})$ is upper bounded as follows:

$$\mathcal{L}_{comp}(\mathcal{C}_{cl}) = \frac{1}{2} \cdot \mathcal{L}_{cl}(\mathcal{C}_{cl}) \leq \frac{1}{2}, \tag{20}$$

with the upper bound of 1/2 reached by $\mathcal{D}(p)$ (see (6)). The estimate in (20) makes sense taking into account that $\mathcal{L}_{comp}$ is designed for a much wider class of curve configurations (including straight line segments) than the measure $\mathcal{L}_{cl}$ (designed only for simple closed curves).

Note 4. Definition 3 can be applied to connected compound curves of arbitrary topology.

Now we give the following theorem related to the properties of $\mathcal{L}_{comp}(\mathcal{C})$. The proof details are omitted because of an obvious analogy with the proof of Theorem 2.

Theorem 3. *Let* $\mathcal{C}$ *be a connected compound curve. The linearity measure* $\mathcal{L}_{comp}(\mathcal{C})$ *satisfies the following properties:*

(i) $\mathcal{L}_{comp}(\mathcal{C}) \in (0, 1]$, *for all connected compound curves* $\mathcal{C}$*;*
(ii) $\mathcal{L}_{comp}(\mathcal{C}) = 1 \quad \Leftrightarrow \quad \mathcal{C}$ *is a straight line segment;*
(iii) $\mathcal{L}_{comp}(\mathcal{C})$ *is invariant with respect to the similarity transformations.*

5 Experiments

Figure 4 demonstrates the application of $\mathcal{L}_{comp}(\mathcal{C})$ to connected compound curves. The computed value can be seen to be sensitive to the arrangement of the components, e.g. when the central horizontal line segment moves in the first two letter "E"s their linearity values change. On the other hand, $\mathcal{L}_{comp}(\mathcal{C})$ can be seen to be insensitive to other factors, and so the two "Xs" with different

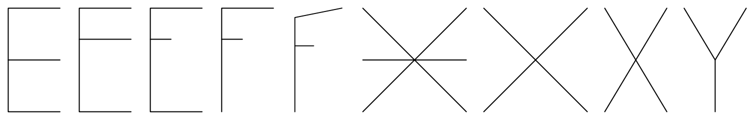


Fig. 4. Values of $\mathcal{L}_{comp}(\mathcal{C})$ for synthetic curves containing multiple components

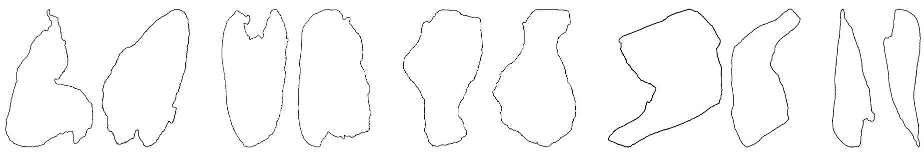


Fig. 5. Two examples of each category of chicken piece

opening angles have the same linearity value. Of course, the different "Xs" could be easily differentiated if required by using a combination of shape descriptors; in this case aspect ratio and linearity.

Now, $\mathcal{L}_{cl}(\mathcal{C})$ is demonstrated on the dataset from Andreu *et al.* [6] which contains 446 thresholded images of chicken pieces, each of which comes from one of five categories: wing, back, drumstick, thigh and back, breast – see figure 5. For classification an existing set of global shape descriptors[1] was applied to the boundaries and fed into a nearest neighbour classifier using Mahalanobis distances. Without incorporating linearity leave-one-out classification accuracy was 91.70%, while including $\mathcal{L}_{cl}(\mathcal{C})$ boosted accuracy to 93.27%. The high classification rate is a substantial improvement on previous results based on approximate cyclic string matching [11] (77.4%), edit-dist kernel [12] (81.1%) and contour fragments [13] (84.5%).

Finally, to show the versatility of the linearity measure we apply it to open curves (in accordance with Note 3, treated as compound curves) for the application to a signature verification task [14]. The data consists of pen trajectories for 2911 genuine signatures taken from 112 subjects, plus five forgers provided a total of 1061 forgeries across all the subjects. Examples of corresponding genuine and forged signatures are shown in figure 6. To provide a richer shape descriptor than a single linearity value, measurement linearity plots are created by computing

$$\mathrm{P}(\mathcal{C}, s) = \mathcal{L}_{comp}(\mathcal{C}(0, s))$$

[1] Compactness, eccentricity, fractal dimension, roundness, Hu's moment invariants, affine moment invariants [7], circularity [8] and ellipticity [9, 10]. Several other global shape descriptors were considered, such as elongatedness, rectangularity, and convexity, but were not found to improve classification accuracy.

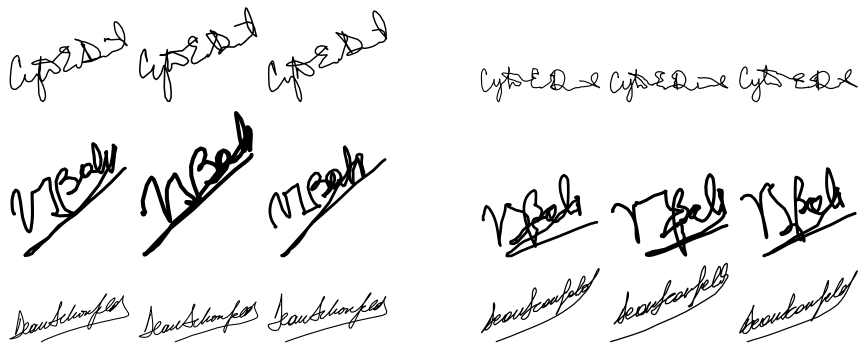


Fig. 6. Examples of genuine (leftmost three) and forged (rightmost three) signatures

where $\mathcal{C}(0, s)$ is the portion of the curve bounded by $(x(0), y(0))$ and by the point $(x(s), y(s)) \in \mathcal{C}$. Matches between two signatures are computed as

$$\min \{\text{area}\,(\mathrm{P}(\mathcal{C}_1), \mathrm{P}(\mathcal{C}_2))\,, \text{area}\,(\mathrm{P}_{rev}(\mathcal{C}_1), \mathrm{P}_{rev}(\mathcal{C}_2))\}$$

where $\mathrm{P}_{rev}(\mathcal{C}, s)$ is equivalent to applying P to the reversed curve, i.e. $\mathrm{P}_{rev}(\mathcal{C}, s) = \mathcal{L}_{cl}(\mathcal{C}(Per(\mathcal{C}), Per(\mathcal{C}) - s))$.

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. Previous results for this task using linearity [3] managed 93.1% accuracy. Application of the new linearity measure $\mathcal{L}(\mathcal{C})$ provides an improvement to achieve 95.5% accuracy.

A final example considers the set of 54 mammographic which are classified according to circumscribed/spiculated, benign/malignant, and CB/CM/SB/SM, in two group and four group classification experiments, and restricts attention to using a single descriptor per classifier. Rangayyan *et al.* [15] first achieved classification accuracies of 88.9%, 75.9% and 64.8% respectively using standard descriptors, while their later work [16] introduced the spiculation index, which improved the benign/malignant classification accuracy from 75.9% to 79%. Using a nearest neighbour classifier with Mahalanobis distances as before, we find that $\mathcal{L}_{cl}(\mathcal{C})$ is able to improve the first two (two group) classification accuracies to 94.4% and 77.8% (although only 59.3% was achieved for the third group experiment).

6 Conclusion

This paper has described a new approach to computing the linearity of shapes. It has general applicability, since it can be used for open curves, closed curves, and also connected compound curves. The new linearity measure is theoretically well founded, and experiments demonstrate its effectiveness.

Acknowledgment. This work is partially supported by the Serbian Ministry of Science and Technology/project III44006.

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

References

1. Gautama, T., Mandić, D., Van Hulle, M.: A novel method for determining the nature of time series. IEEE Transactions on Biomedical Engineering 51, 728–736 (2004)
2. Stojmenović, M., Nayak, A., Žunić, J.: Measuring linearity of planar point sets. Pattern Recognition 41, 2503–2511 (2008)
3. Žunić, J., Rosin, P.: Measuring linearity of open planar curve segments. Image Vision Computing 29, 873–879 (2011)
4. Benhamou, S.: How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension. Journal of Theoretical Biology 229, 209–220 (2004)
5. Klette, R., Rosenfeld, A.: Digital Geometry. Morgan Kaufmann, San Francisco (2004)
6. Andreu, G., Crespo, A., Valiente, J.: Selecting the toroidal self-organizing feature maps (tsofm) best organized to object recognition. In: Int. Conf. Neural Networks, vol. 2, pp. 1341–1346 (1997)
7. Flusser, J., Suk, T.: Pattern recognition by affine moment invariants. Pattern Recognition 26, 167–174 (1993)
8. Haralick, R.: A measure for circularity of digital figures. IEEE Trans. on Systems, Man and Cybernetics 4, 394–396 (1974)
9. Proffitt, D.: The measurement of circularity and ellipticity on a digital grid. Pattern Recognition 15, 383–387 (1982)
10. Peura, M., Iivarinen, J.: Efficiency of simple shape descriptors. In: Arcelli, C., et al. (eds.) Aspects of Visual Form Processing, pp. 443–451. World Scientific (1997)
11. Mollineda, R., Vidal, E., Casacuberta, F.: A windowed weighted approach for approximate cyclic string matching. In: Int. Conf. Pattern Recognition, vol. 4, pp. 188–191 (2002)
12. Neuhaus, M., Bunke, H.: Edit distance-based kernel functions for structural pattern classification. Pattern Recognition 39, 1852–1863 (2006)
13. Daliri, M.R., Torre, V.: Classification of silhouettes using contour fragments. Computer Vision and Image Understanding 113, 1017–1025 (2009)
14. Munich, M., Perona, P.: Visual identification by signature tracking. IEEE Trans. on Patt. Anal. and Mach. Intell. 25, 200–217 (2003)
15. Rangayyan, R., Elfaramawy, N., Desautels, J., Alim, O.: Measures of acutance and shape for classification of breast-tumors. IEEE Trans. on Medical Imaging 16, 799–810 (1997)
16. Rangayyan, R., Mudigonda, N., Desautels, J.: Boundary modeling and shape analysis methods for classification of mammographic masses. Medical and Biological Engineering and Computing 38, 487–496 (2000)