Noname manuscript No.
(will be inserted by the editor)

Measuring Linearity of Connected Configurations of a Finite Number of $2D$ and $3D$ Curves

Paul L. Rosin · Jovanka Pantović · Joviša Žunić

Received: date / Accepted: date

Abstract We define a new linearity measure for a wide class of objects consisting of a set of of curves, in both $2D$ and $3D$. After initially observing closed curves, which can be represented in a parametric form, we extended the method to connected compound curves – i.e. to connected configurations of a number of curves representable in a parametric form. 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. We prove that the linearity is equal to 1 if and only if the measured curve consists of two straight line overlapping segments. The new linearity measure is theoretically well founded and all related statements are supported with rigorous mathematical proofs.

The behavior and applicability of the new linearity measure are explained and illustrated by a number of experiments.

Keywords Shape · shape descriptors · $2D$ curves · $3D$ curves · compound curves · linearity measure · image processing

1 Introduction

There are many ways to quantitatively characterize 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 characterization. A certain method

Paul L. Rosin
Cardiff University, School of Computer Science, Cardiff CF24 3AA, Wales, U.K.
E-mail: Paul.Rosin@cs.cf.ac.uk

Jovanka Pantović
University of Novi Sad, Faculty of Technical Sciences, 21000 Novi Sad, Serbia
E-mail: pantovic@uns.ac.rs second address

Joviša Žunić
University of Exeter, Computer Science, Exeter EX4 4QF, U.K.
E-mail: J.Zunic@ex.ac.uk

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 measure of a wide class of configurations consisting of curves, in both $2D$ and $3D$. 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 had developed a method for the computation of such a quantity, it turned out that the method can be applied successfully to a wider class of objects/shapes – not just to the simple closed curves which represent the boundaries/frontiers of planar regions. In particular, unlike most standard shape descriptors, the method can be applied to configurations consisting of a finite number of closed curve segments which are connected in the sense that any two points, from the configuration, can be connected by a path (a continuous curve) which consists of subarcs of the curves belonging to the configuration considered. Such configurations could be termed *path-connected compound curves*, although in this paper we will use the simpler term *connected compound curves*. Such an extension of the method, from a single curve measurement to the measurement of a property of much more complex and much more generic configurations, is particularly useful, as many applications exist in which the morphology of connected configurations of curves needs to be quantified. Just two examples are networks (e.g. road networks [Chen and Chen(2011)], drainage networks [Black et al(2012)Black, Perron, Burr, and Drummond], neuronal networks [Schmitz et al(2011)Schmitz, Hjorth, Joemai, Wijntjes et al]) and cracks (e.g. in brake disks [Goo and Lim(2012)], pots [Lahlil et al(2013)Lahlil, Li, and Xu] pavements [DeCarlo and Shokri(2014)]).

Several linearity measures are already considered in the literature [Gautama et al(2004)Gautama, Mandić, and Van Hulle Gautama et al(2003)Gautama, Mandić, and Van Hulle,Stojmenović et al(2008)Stojmenović, Nayak, and Žunić, Žunić and Martinez-Ortiz(2009),Žunić and Rosin(2011)]. But they are mainly related to open curve segments. I.e. they measure how much 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* [Benhamou(2004)], denoted here 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 [Žunić and Rosin(2011)], 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 parametrization $x = x(s)$, $y = y(s)$, $s \in [0, 1]$, 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 [Žunić and Rosin(2011)], 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

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

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.
Also, any circle $\mathcal{C}$ has a higher linearity measure than rectangles $\mathcal{R}(t)$. It can be easily verified $\mathcal{S}(\mathcal{C}) = 1/4 \leq \mathcal{S}(\mathcal{C}) = 3/\pi^2$. This is also a bad property for $\mathcal{S}(\mathcal{C})$.

It is worth mentioning that there is a simple and easy way to measure the linearity 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 [Klette and Rosenfeld(2004)]) and the curve perimeter:

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

It could be said that $\mathcal{I}_{closed}(\mathcal{C})$ 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 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.

In this paper, first 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 pairwise distance) since the shape centroid is used for the measure computation. An extension to a very general class of curve configurations is obtained as well.

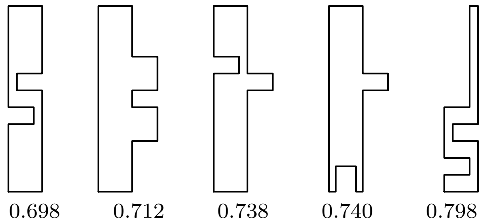

Fig. 1 The value of $\mathcal{I}_{closed}(\mathcal{C})$ is 0.699 for all the shapes, but the proposed measure $\mathcal{L}_{cl}(\mathcal{C})$ produces different linearities (shown below).

Such a generalized 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).

The measures derived are theoretically well founded. All statements are given with strict mathematical proofs. This is always an advantage because the theoretical considerations lead to a better understanding of the behavior of the measure considered. All statements are proved considering the appearing curves as $3D$ objects. The restriction to the $2D$ case is straightforward and a separate consideration and discussion is not necessary.

The paper is organized as follows. Section 2 gives the basic definitions and notations. The new linearity measure for closed curves is introduced in Section 3. Rigorous proofs of all properties of the new measure are in the same section. Extensions of the method to the connected compound curves and open curves are in Section 4. Experiments are in Section 5, and concluding remarks in Section 6.

2 Definitions and Denotations

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

- As usual, $\mathbf{d}_2(A, B) = \mathbf{d}_2((x, y, z), (u, v, w)) = \sqrt{(x-u)^2 + (y-v)^2 + (z-w)^2}$ denotes the Euclidean distance between the points $A = (x, y, z)$ and $B = (u, v, w)$.
- The Euclidean perimeter of a given curve $\mathcal{C}$ is denoted as $\mathbf{l}_2(\mathcal{C})$.
- $diam(\mathcal{C})$ 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 parametrization.

$$x = x(s), \quad y = y(s), \quad z = z(s) \quad \text{where} \quad s \in [0, \mathbf{l}_2(\mathcal{C})]. \tag{4}$$

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

Of course, the Euclidean perimeter of a given curve $\mathcal{C}$ is computed as

$$\mathbf{l}_2(\mathcal{C}) \;=\; \int_{\mathcal{C}} ds \tag{5}$$

assuming that $\mathcal{C}$ is given in an arc-length parameterization (as in (4)).

The centroid $(x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}})$ of a given curve $\mathcal{C}$ is computed as usual

$$(x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}}) \;=\; \left(\frac{\int_{\mathcal{C}} x(s)ds}{\mathbf{l}_2(\mathcal{C})}, \; \frac{\int_{\mathcal{C}} y(s)ds}{\mathbf{l}_2(\mathcal{C})}, \; \frac{\int_{\mathcal{C}} z(s)ds}{\mathbf{l}_2(\mathcal{C})} \right) \tag{6}$$

again, assuming that $\mathcal{C}$ is given as in (4). Notice that the coordinates of the centroid $(x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}})$ are the average values of the respective coordinates of all the curve points.

In several situations, we will assume that a given curve $\mathcal{C}_a$ is scaled to be of unit length. Formally, it means that $\mathcal{C}_a$ is replaced with the curve $\mathcal{C}$ defined by

$$\mathcal{C} \;=\; \frac{1}{\mathbf{l}_2(\mathcal{C})} \cdot \mathcal{C}_a \;=\; \left\{ \left(\frac{x}{\mathbf{l}_2(\mathcal{C})}, \; \frac{y}{\mathbf{l}_2(\mathcal{C})}, \; \frac{z}{\mathbf{l}_2(\mathcal{C})} \right) \;\mid\; (x, y, z) \in \mathcal{C}_a \right\} \tag{7}$$

Obviously, $\mathcal{C}$ has length equal to 1 and the centroids of $\mathcal{C}_a$ and $\mathcal{C}$ are related by

$$(x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}}) = \frac{1}{\mathbf{l}_2(\mathcal{C})} \cdot (x_{\mathcal{C}_a}, y_{\mathcal{C}_a}, z_{\mathcal{C}_a}) \,. \tag{8}$$

Initially, we will focus on *closed curves.* They are characterized the following condition:

$$(x(0), y(0), z(0)) = (x(\mathbf{l}_2(\mathcal{C})), y(\mathbf{l}_2(\mathcal{C})), z(\mathbf{l}_2(\mathcal{C}))). \tag{9}$$

We will say that $(x(0), y(0), z(0))$ is the *curve start point*, while $(x(\mathbf{l}_2(\mathcal{C})), y(\mathbf{l}_2(\mathcal{C})), z(\mathbf{l}_2(\mathcal{C})))$ is the *curve end-point*, even if they coincide (in the case of closed curves).

In the case one coordinate is identically equal to a constant value (for example $z = 0$), $\mathcal{C}$ is a planar curve (i.e. $2D$ curve).

Notice that even being simple, the definition (4) covers a wide spectrum of curves. $\mathcal{C}$ can be a simple closed curve (Fig.2(a)) or a curve which crosses itself (Fig.2(b)(c)) or even a curve whose two halves overlap. Here we define such a curve $\mathcal{B}(p)$, for $p > 0$, consisting of two (overlapping) identical straight line segments, both of length $p/2$. $\mathcal{B}(p)$ is still representable in the form of (4). One possibility for the arc-length parametrization of $\mathcal{B}(p)$ is:

$$\mathcal{B}(p) \;:\; \begin{cases} x = x(s) = s, \qquad y = y(s) = 0, \; z = z(s) = 0, \; s \in [0, \; p/2] \\ \\ x = x(s) = p - s, \; y = y(s) = 0, \; z = z(s) = 0, \; s \in [p/2, \; p]. \end{cases} \tag{10}$$

Because $\mathcal{B}(p)$ allows a parametrization 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), the curve $\mathcal{B}(p)$ looks like a single straight line segment of length $p/2$, but the parametrization (10) makes clear that $\mathcal{B}(p)$ has length p and that it consists of two identical overlapping straight line segments, both of length $p/2$. 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, and this configuration does not have a parametric representation, as given in (4). Such configurations will be called *connected compound curves* and we will develop a linearity measure which can be applied to them, as well. 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.

3 Linearity Measure for Closed $3D$ Curves

We define a new linearity measure for closed $3D$ curves. In our proofs the appearing curves are treated as $3D$ objects. Obviously, if one of the coordinates is fixed, we get planar $2D$ curves. We start with the following theorem whose results motivate the definition of the new measure.

Theorem 1 *Let $\mathcal{C}$ be a closed curve and let A be the point of $\mathcal{C}$ furthest from the centroid C of $\mathcal{C}$. Then*

(a) *The upper bound for the Euclidean distance between A and C is given by the inequality*

$$\mathbf{d}_2(A, C) \le \frac{1}{4} \cdot \mathbf{l}_2(\mathcal{C}). \tag{11}$$

(b) *The bound in* (**a**) *is the best possible upper bound, since there is a closed curve $\mathcal{C}$ such that $\mathbf{d}_2(A, C) = \frac{1}{4} \cdot \mathbf{l}_2(\mathcal{C})$.*

Proof. (a) Since the quantity $\dfrac{\mathbf{d}_2(A, C)}{\mathbf{l}_2(\mathcal{C})}$ is invariant with respect to translation, rotation and scaling transformations, we can assume, without loss of generality, the following:

(i) $\mathbf{l}_2(\mathcal{C}) = 1$;
(ii) $A = (0, 0, 0)$ and the centroid C of $\mathcal{C}$ lies on the positive part of x-axis, i.e. $x_C > 0$;
(ii) $\mathcal{C}$ is given by the arc-length parametrization

$$x = x(s), \quad y = y(s), \quad z = z(s), \quad s \in [0, 1], \tag{12}$$

such that

$$A = (x(0), y(0), z(0)) = (x(1), y(1), z(1)) = (0, 0, 0).$$

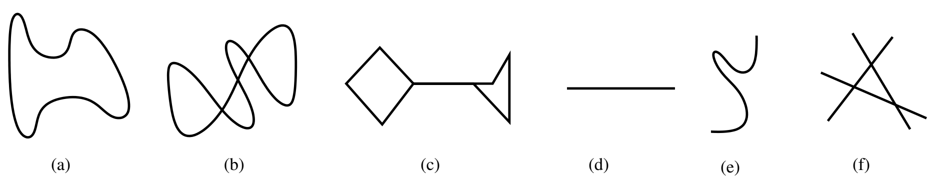

Fig. 2 Several curve examples which will be treated here as closed curves are in (a)-(e). An example of connected compound curve is in (f).

The previous parametrization provides us with

$$y_{\mathcal{C}} \;=\; \int_{\mathcal{C}} y(s) \;=\; 0, \qquad z_{\mathcal{C}} \;=\; \int_{\mathcal{C}} z(s) \;=\; 0. \tag{13}$$

Since A is furthest from the centroid C, the curve $\mathcal{C}$ lies inside the sphere with radius $\mathbf{d}_2(A, C)$ centered at C. This implies $x = x(s) \geq 0$ for any $s \in [0, 1]$. Consider the function $F(a) = \int_{s=0}^{a} x(s)\; ds,\;\; a \in [0, 1]$ where $x = x(s)$ is as in (12).

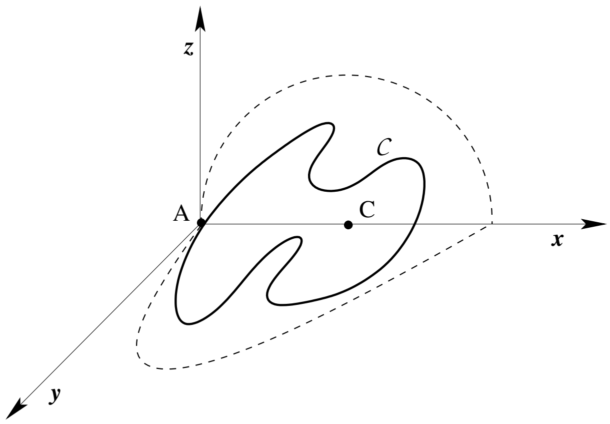

Fig. 3 A is a point from the curve $\mathcal{C}$ which is furthest from the curve centroid C.

Since $x(s) \geq 0$, $s \in [0, 1]$, $F(a)$ is a non-decreasing function. Furthermore, 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

$$\int_{s=0}^{a_0} x(s)\; ds \;=\; \int_{s=a_0}^{1} x(s)\; ds \;=\; \frac{1}{2} \cdot \int_{s=0}^{1} x(s)\; ds. \tag{14}$$

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

$$\int_{s=0}^{1} x(s)\; ds \;=\; 2 \cdot \int_{s=0}^{a_0} x(s)\; ds \;\leq\; 2 \cdot \int_{s=0}^{a_0} s\; ds = a_0^2, \tag{15}$$

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

$$\int_{s=0}^{1} x(s)\; ds \;=\; 2 \int_{s=a_0}^{1} x(s)\; ds \;\leq\; 2 \cdot \int_{s=a_0}^{1} (1 - s)\; ds = (1 - a_0)^2. \tag{16}$$

Finally, the just derived (15) and (16) give

$$\int_{s=0}^{1} x(s)\; ds \;\leq\; \min\{a_0^2,\; (1 - a_0)^2\} \;\leq\; \frac{1}{4} \tag{17}$$

(the last inequality follows because of $a_0 \in [0, 1]$)).
Taking into account (13) and the above estimate (17), we have proven (**a**).

(b) The statement follows because any closed curve $\mathcal{B}(p)$, defined as in (10), reaches the upper bound in (11). The centroid of $\mathcal{B}(p)$ is $(p/4, 0, 0)$ and the point $(0, 0, 0)$ (and also $(p/2, 0, 0)$) is furthest from the centroid of $\mathcal{B}(p)$. (Notice $\mathbf{l}_2(\mathcal{B}(p)) = p$ since $\mathcal{B}(p)$ consists of two overlapping straight line segments both having the length $p/2$). □

Note 1 The inequality in (11) is strict for simple closed curves, but still cannot be improved. This can be deduced from (17). Indeed, $\displaystyle\int_{s=0}^{1} x(s)\ ds = \min\{a_0^2,\ (1 - a_0)^2\} = \frac{1}{4}$ would imply $a_0 = 1/2$ and (see (15) and (16))

$$\int_{s=0}^{a_0=1/2} x(s)\ ds = \frac{1}{2} \cdot a_0^2 = \int_{s=1/2}^{a_0=1} x(s)\ ds = \frac{1}{2} \cdot (1 - a_0)^2 = \frac{1}{8}.$$

Since $x(s) \le s$, in order to have $\displaystyle\int_{s=0}^{a_0=1/2} x(s)\ ds = \int_{s=0}^{a_0=1} x(s)\ ds = \frac{1}{8}$, $x(s) = s$ is required 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 (11) cannot be improved for simple closed curves, even though no simple closed curves satisfy $\sqrt{(x_0 - x_{\mathcal{C}})^2 + (y_0 - y_{\mathcal{C}})^2 + (z_0 - z_{\mathcal{C}})^2} = \frac{1}{4} \cdot \mathbf{l}_2(\mathcal{C})$, 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

$$\left| \sqrt{(x_0 - x_{\mathcal{C}(\delta)})^2 + (y_0 - y_{\mathcal{C}(\delta)})^2 + (z_0 - z_{\mathcal{C}(\delta)})^2} - \frac{1}{4} \cdot \mathbf{l}_2(\mathcal{C}(\delta)) \right| < \delta.$$

Indeed, the required $\mathcal{C}(\delta)$ can be selected from a family of rectangles $\mathcal{R}(t)$ (defined as in (2)), since

$$\lim_{t \to \infty} \frac{\sqrt{(x_0 - x_{\mathcal{R}(t)})^2 + (y_0 - y_{\mathcal{R}(t)})^2 + (z_0 - z_{\mathcal{R}(t)})^2}}{\mathbf{l}_2(\mathcal{R}(t))} = \frac{1}{4}.$$

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 given in an arc-length parametrization: $x = x(s)$, $y = y(s)$, $z = z(s)$, $s \in [0, \mathbf{l}_2(\mathcal{C})]$, and let $A = (x_0, y_0, z_0)$ be the point of $\mathcal{C}$ furthest from the centroid $C = (x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}})$ of $\mathcal{C}$. The linearity measure $\mathcal{L}_{cl}(\mathcal{C})$ of $\mathcal{C}$ is defined as

$$\mathcal{L}_{cl}(\mathcal{C}) = 4 \cdot \frac{\sqrt{(x_0 - x_{\mathcal{C}})^2 + (y_0 - y_{\mathcal{C}})^2 + (z_0 - z_{\mathcal{C}})^2}}{\mathbf{l}_2(\mathcal{C})} = 4 \cdot \frac{\mathbf{d}_2(A, C)}{\mathbf{l}_2(\mathcal{C})}. \quad (18)$$

Properties of the linearity measure $\mathcal{L}_{cl}(\mathcal{C})$, defined for closed curves, are stated by the following 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) $\mathcal{L}_{cl}(\mathcal{C}) = 1 \quad \Leftrightarrow \quad \mathcal{C} = \mathcal{B}(p)$, *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 neither change the curve length nor the distance between the centroid and the curve points. Also, $\mathbf{d}_2(A,C)/\mathbf{l}_2(\mathcal{C})$ is an obvious scaling invariant. This proves **(iii)**. □

4 Connected Compound Curves

In this section we will further develop the idea applied in the previous section and define a linearity measure for a very wide class of connected configurations made by a certain number of curves – here called connected compound curves. We start with a formal definition of a connected compound curve and its centroid, and proceed with the theorem which gives the arguments for a definition of a measure for linearity of connected compound curves.

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

$$\mathcal{C}_i: \quad x = x_i(s), \;\; y = y_i(s), \;\; z = z_i(s), \quad s \in [0, l_i], \quad \text{for } \; 1 = 1, 2, \ldots, n. \tag{19}$$

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 (19)

$$\mathcal{T}(\mathcal{C}) = l_1 + l_2 + \ldots + l_n = \mathbf{l}_2(\mathcal{C}_1) + \mathbf{l}_2(\mathcal{C}_2) + \ldots + \mathbf{l}_2(\mathcal{C}_n).$$

The centroid $(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C})$ 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 $\mathcal{C}$. Formally, the centroid of connected compound curve $\mathcal{C}$ is

$$(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C}) = \frac{1}{\mathcal{T}(\mathcal{C})} \cdot \left(\sum_{i=1}^{n} \int_{\mathcal{C}_i} x_i(s)\; ds, \; \sum_{i=1}^{n} \int_{\mathcal{C}_i} y_i(s)\; ds, \; \sum_{i=1}^{n} \int_{\mathcal{C}_i} z_i(s)\; ds \right), \tag{20}$$

assuming that all curves $\mathcal{C}_i$ are given as in (19).

We give now the theorem which is a generalization of Theorem 1.

Theorem 3 *Let* $\mathcal{C} = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_n$ *be a connected compound curve whose components* $\mathcal{C}_i$ $(i = 1, 2, \ldots, n)$ *are given in arc-length parametrization:*

$$x = x_i(s), \quad y = y_i(s), \quad z = z_i(s), \quad s \in [0, l_i], \quad \textit{for each } i \in \{1, 2, \ldots, n\}.$$

Also, let $P = (x_0, y_0, z_0)$ *be an arbitrary point which belongs to* $\mathcal{C}$.

Then, the distance of $P(x_0, y_0, z_0)$ *from the centroid* $(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C})$ *of* $\mathcal{C}$ *is upper bounded by a half of the total-length* $\mathcal{T}(\mathcal{C})$ *of* $\mathcal{C}$. *More formally:*

$$\sqrt{(x_0 - x_\mathcal{C})^2 + (y_0 - y_\mathcal{C})^2 + (z_0 - z_\mathcal{C})^2} \le \frac{1}{2} \cdot (l_1 + l_2 + \ldots + l_n) = \frac{1}{2} \cdot \mathcal{T}(\mathcal{C}). \quad (21)$$

Proof. Let $\mathcal{C} = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_n$ satisfy the preconditions in the statement of the theorem. We can assume, without loss of generality, that $\mathcal{C}$ is translated such that the centroid of $\mathcal{C}$ coincides with the origin (i.e. $(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C}) = (0, 0, 0)$).

Now, let us consider another compound curve

$$\tilde{\mathcal{C}} = \left(\mathcal{C}'_1 \cup \mathcal{C}''_1\right) \cup \left(\mathcal{C}'_2 \cup \mathcal{C}''_2\right) \cup \ldots \cup \left(\mathcal{C}'_n \cup \mathcal{C}''_n\right) \quad (22)$$

where the curve segments $\mathcal{C}'_1$, $\mathcal{C}''_1$, $\mathcal{C}'_2$, $\mathcal{C}''_2$, $\ldots$, $\mathcal{C}'_n$, $\mathcal{C}''_n$, are defined as follows:

$$\mathcal{C}'_i = \mathcal{C}''_i : \begin{cases} x'_i(s) = x''_i(s) = \frac{x_i(s)}{2}, & s \in [0, l_i], \\ y'_i(s) = y''_i(s) = \frac{y_i(s)}{2}, & s \in [0, l_i], \\ z'_i(s) = z''_i(s) = \frac{z_i(s)}{2}, & s \in [0, l_i], \end{cases} \quad (23)$$

for all $i \in \{1, 2, \ldots, n\}$. In other words curves $\mathcal{C}'_i$ and $\mathcal{C}''_i$ coincide and both are, actually, identical to the curve $\mathcal{C}_i$ scaled for a factor $1/2$.

The following statements follow from the definition:

- The total-lengths of the compound curves $\tilde{\mathcal{C}}$ and $\mathcal{C}$ are the same (both equal to $l_1 + l_2 + \ldots + l_n$);
- The centroid of $\tilde{\mathcal{C}}$ coincides with the centroid of $\mathcal{C}$ (i.e. $(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C}) = (x_{\tilde{\mathcal{C}}}, y_{\tilde{\mathcal{C}}}, z_{\tilde{\mathcal{C}}}) = (0, 0, 0)$);
- If a point $P = (x, y, z)$ belongs to $\mathcal{C}$ then the point $\tilde{P} = (x/2, y/2, z/2)$ belongs to $\tilde{\mathcal{C}}$ and vice versa. Formally

$$P = (x, y, z) \in \mathcal{C} \quad \Leftrightarrow \quad \tilde{P} = (x/2, y/2, z/2) \in \tilde{\mathcal{C}}. \quad (24)$$

- The distance of a point $P = (x, y, z) \in \mathcal{C}$ to the centroid $(x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C})$ of $\mathcal{C}$, and the distance of the corresponding point $\tilde{P} = (x/2, y/2, z/2) \in \tilde{\mathcal{C}}$ to the centroid $(x_{\tilde{\mathcal{C}}}, y_{\tilde{\mathcal{C}}}, z_{\tilde{\mathcal{C}}})$ of $\tilde{\mathcal{C}}$ are related as follows:

$$\begin{aligned} \mathbf{d}_2(P, \mathcal{C}) &= \mathbf{d}_2((x, y, z), (x_\mathcal{C}, y_\mathcal{C}, z_\mathcal{C})) \\ &= 2 \cdot \mathbf{d}_2((x/2, y/2, z/2), (x_{\tilde{\mathcal{C}}}, y_{\tilde{\mathcal{C}}}, z_{\tilde{\mathcal{C}}})) = 2 \cdot \mathbf{d}_2(\tilde{P}, \tilde{\mathcal{C}}). \end{aligned} \quad (25)$$

Furthermore, if we consider the compound curve $\tilde{\mathcal{C}}$ as a graph having $2n$ edges (represented by the curves $\mathcal{C}'_i$, $\mathcal{C}''_i$, $i = 1, 2, \ldots, n$) and whose nodes are all of an even degree, from graph theory [Matoušek and Nešetril(1998)] there is an Eulerian cycle $\mathcal{E}(\tilde{\mathcal{C}})$ which includes (completely) each of curve segments $\mathcal{C}'_i$, $\mathcal{C}''_i$, $(i = 1, 2, \ldots, n)$ exactly once.

Now, let us parametrize $\mathcal{E}(\tilde{\mathcal{C}})$ by using an arc-length parametrization:

$$x = \tilde{x}(s), \quad y = \tilde{y}(s), \quad z = \tilde{z}(s), \quad s \in [0, l_1 + l_2 + \ldots + l_n]$$

Notice: $\mathbf{l}_2(\mathcal{E}(\tilde{\mathcal{C}})) = l_1 + l_2 + \ldots + l_n$, and the centroid $(x_{\mathcal{E}}, y_{\mathcal{E}}, z_{\mathcal{E}})$ of $\mathcal{E}(\tilde{\mathcal{C}})$, centroid of $\tilde{\mathcal{C}}$, centroid of $\mathcal{C}$, and the origin coincide – i.e.

$$(x_{\mathcal{E}}, y_{\mathcal{E}}, z_{\mathcal{E}}) \ = \ (x_{\tilde{\mathcal{C}}}, y_{\tilde{\mathcal{C}}}, z_{\tilde{\mathcal{C}}}) \ = \ (x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}}) \ = \ (0, 0, 0).$$

Then, by Theorem 1, the distance between the centroid of the curve $\mathcal{E}(\tilde{\mathcal{C}})$ and the point $\tilde{P} = (x_0/2, y_0/2, z_0/2)$ satisfies

$$\sqrt{\left(x_{\mathcal{E}} - \frac{x_0}{2}\right)^2 + \left(y_{\mathcal{E}} - \frac{y_0}{2}\right)^2 + \left(z_{\mathcal{E}} - \frac{z_0}{2}\right)^2} = \sqrt{\left(x_{\tilde{\mathcal{C}}} - \frac{x_0}{2}\right)^2 + \left(y_{\tilde{\mathcal{C}}} - \frac{y_0}{2}\right)^2 + \left(z_{\tilde{\mathcal{C}}} - \frac{z_0}{2}\right)^2}$$
$$\leq \frac{1}{4} \cdot \mathbf{l}_2(\mathcal{E}(\tilde{\mathcal{C}})) \ = \ \frac{l_1 + l_2 + \ldots + l_n}{4} \tag{26}$$

Finally, by using (25),

$$\sqrt{(x_{\mathcal{C}} - x_0)^2 + (y_{\mathcal{C}} - y_0)^2 + (z_{\mathcal{C}} - z_0)^2} = 2 \cdot \sqrt{\left(x_{\tilde{\mathcal{C}}} - \frac{x_0}{2}\right)^2 + \left(y_{\tilde{\mathcal{C}}} - \frac{y_0}{2}\right)^2 + \left(z_{\tilde{\mathcal{C}}} - \frac{z_0}{2}\right)^2}$$
$$\leq 2 \cdot \frac{l_1 + l_2 + \ldots + l_n}{4} = \frac{1}{2} \cdot \mathcal{T}(\mathcal{C}). \tag{27}$$

This establishes the proof. $\square$

Based on the previous observations, we have good arguments to 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 where the curves $\mathcal{C}_1$, $1 \leq i \leq n$ are given in an arc-length parametrization. Also, let the point $P = (x_0, y_0, z_0) \in \mathcal{C}$ be furthest from the centroid $(x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}})$ of $\mathcal{C}$. The linearity measure $\mathcal{L}_{comp}(\mathcal{C})$ of $\mathcal{C}$ is defined as

$$\mathcal{L}_{comp}(\mathcal{C}) = 2 \cdot \frac{\sqrt{(x_0 - x_{\mathcal{C}})^2 + (y_0 - y_{\mathcal{C}})^2 + (z_0 - z_{\mathcal{C}})^2}}{\mathcal{T}(\mathcal{C})}. \tag{28}$$

Note 3 Definition 3 can be directly applied to a single open curve. The linearity measure $(\mathcal{L}_{comp}(\mathcal{C}))$ of such an open single curve $\mathcal{C}$ is computed as a double value of the longest distance of a point from $\mathcal{C}$ to the centroid of $\mathcal{C}$ divided by the curve length (in such a situation $\mathcal{T}(\mathcal{C}) = \mathbf{l}_2(\mathcal{C})$).

Note 4 If Definition 3 is applied to a simple closed curve $\mathcal{C}_{cl}$, then (as in the previous note) $\mathcal{L}_{comp}(\mathcal{C}_{cl})$ is computed as a double value of the longest distance of a point from $\mathcal{C}_{cl}$ to the centroid of $\mathcal{C}_{cl}$ divided by the length of $\mathcal{C}$. Obviously such a linearity measure ($\mathcal{L}_{comp}(\mathcal{C}_{cl})$) differs from the linearity measure $\mathcal{L}_{cl}(\mathcal{C}_{cl}$ introduced by Definition 1. Precisely, the following relationship

$$\mathcal{L}_{comp}(\mathcal{C}_{cl}) = \frac{1}{2} \cdot \mathcal{L}_{cl}(\mathcal{C}_{cl}) \tag{29}$$

is true for all simple closed curves $\mathcal{C}_{cl}$.

Further, the linearity measure $\mathcal{L}_{comp}(\mathcal{C}_{cl})$, of any closed curve $\mathcal{C}_{cl}$, is upper bounded by 1/2 because of:

$$\mathcal{L}_{comp}(\mathcal{C}_{cl}) = \frac{1}{2} \cdot \mathcal{L}_{cl}(\mathcal{C}_{cl}) \leq \frac{1}{2}. \tag{30}$$

The upper bound of 1/2 is reached by $\mathcal{B}(p)$ (see (10)). An upper bound of $\mathcal{L}_{comp}(\mathcal{C}_{cl})$ smaller than 1 (as given in in (30)) makes sense taking into account that $\mathcal{L}_{comp}$ is designed for a much wider class of curve configurations (including open curve segments) than the measure $\mathcal{L}_{cl}$ (designed only for closed curves). Thus, it might be expected that $\mathcal{L}_{comp}$ does not reach its maximum on some particular sub-domain (in this case, set of closed curves). The linearity measure $\mathcal{L}_{comp}$ reaches its maximum (which is 1) only for straight line segments, as stated in the next theorem.

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

Theorem 4 *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.*

Proof.

(i) A direct consequence of Theorem 3.
(ii) Similar to Note 1.
(iii) Follows from the definition. □

5 Experiments

Figure 4 demonstrates the application of $\mathcal{L}_{comp}(\mathcal{C})$ to simple synthetic connected compound curves. As the curves become more elongated (lower row compared to the upper row) then the linearity values generally increase. Also, 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, $\mathcal{L}_{comp}(\mathcal{C})$ is insensitive to other factors, and so

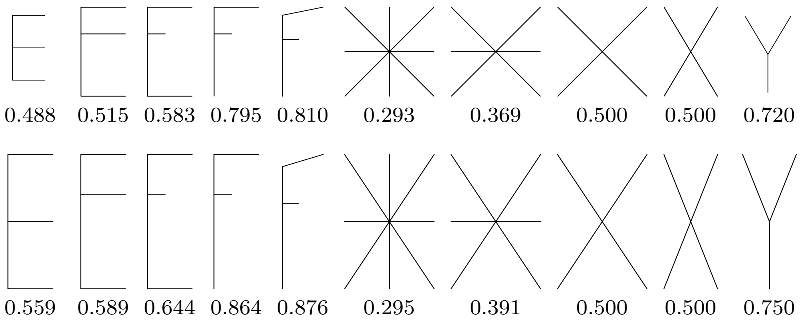

Fig. 4 Values of $\mathcal{L}_{comp}(\mathcal{C})$ for synthetic curves containing multiple components.

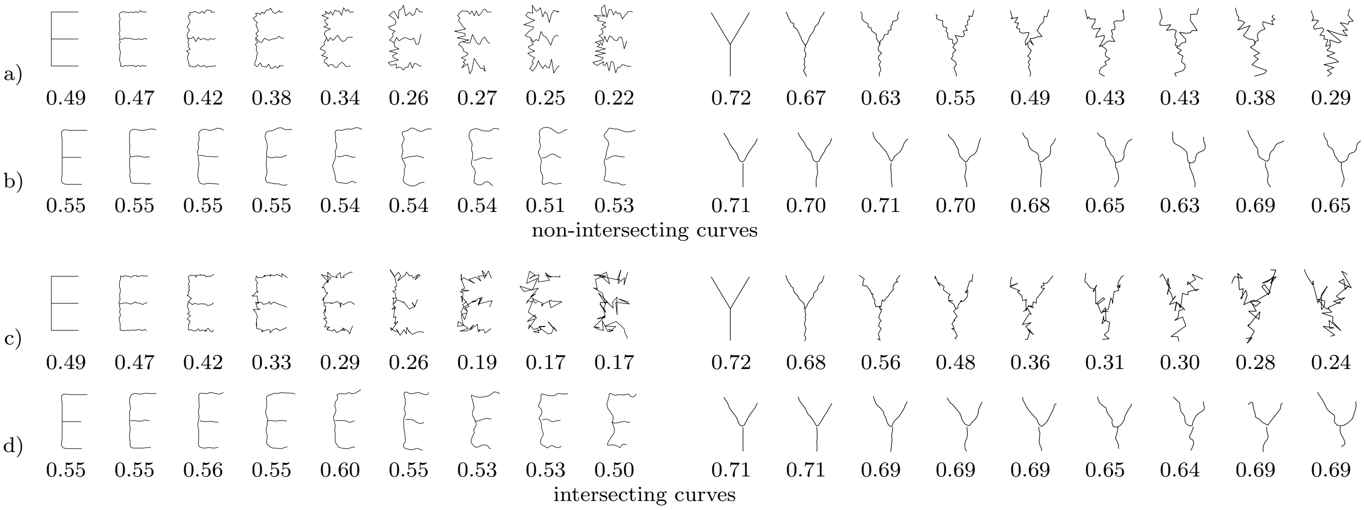

Fig. 5 a) & c) increasing amounts of noise are added to two letters; b) & d) the noise is reduced by smoothing. Values of $\mathcal{L}_{comp}(\mathcal{C})$ are provided to show the effects of noise and smoothing.

the "Xs" with different 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.

Figure 5 demonstrates how noise affects the computed linearity values. The two shapes ("E" and "Y") are uniformly sampled w.r.t. arc length, and then normally distributed noise is added to the samples. In figure 5a the noise has been added such that no line intersections are created. In contrast, in figure 5c no such constraint is applied, and in most cases line intersections occur. Increasing amounts of noise make the shapes less straight, and more convoluted, and therefore their corresponding $\mathcal{L}_{comp}(\mathcal{C})$ values decrease such that the two letters can no longer be discriminated by $\mathcal{L}_{comp}(\mathcal{C})$. However, this is as expected, since the linearity measure should be responsive to the changes in the shape. Of course, since these changes mainly affect the finer detail then their effect can be reduced by preprocessing the curves. A fixed amount of smoothing was applied to all the curves (figure 5b & d), and it can be this is sufficient to stabilize the computed values of $\mathcal{L}_{comp}(\mathcal{C})$.

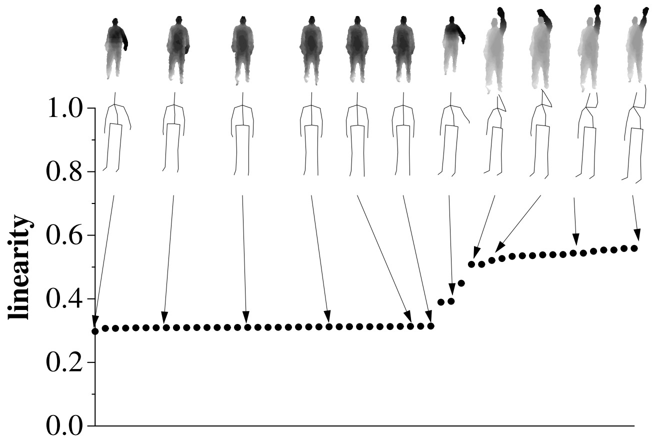

Fig. 6 2D views of 3D skeletons extracted from depth maps taken from an action sequence and plotted with their $\mathcal{L}_{comp}(\mathcal{C})$ values.

Another example of simple data is provided to demonstrate the application of the method to 3D compound curves, and consists of a human skeleton from sequence a01-s01-e01 in the MSR Action3D Dataset. The skeleton has 20 3D joint positions, and these are connected to form a 3D compound curve, from which $\mathcal{L}_{comp}(\mathcal{C})$ is measured. The sequence has 54 frames, and the linearity values are plotted in ascending order in figure 6 along with some corresponding 2D views and source depth maps of their skeletons. It can be seen that many frames have a neutral pose, which all produce a low linearity value, while raising one arm elongates the skeleton, and increases the measured linearity values. Thus, $\mathcal{L}_{comp}(\mathcal{C})$ could provide a useful feature for activity recognition. Further investigation will be carried out in a future paper.

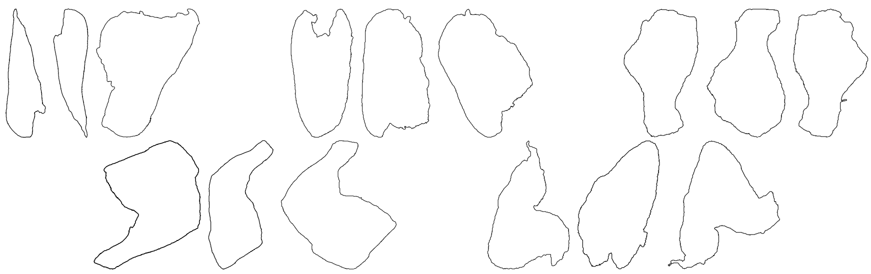

Fig. 7 Three examples of each category of chicken piece.

The remaining experiments perform a more quantitative evaluation of the linearity measure, which is used in several classification tasks. First, $\mathcal{L}_{cl}(\mathcal{C})$ is demonstrated on the dataset from Andreu *et al.* [Andreu et al(1997)Andreu, Crespo, and Valiente] which contains 446 thresholded images of chicken pieces, each of which comes from one of five categories: breast, back, drumstick, thigh and back, wing – see figure 7. 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 [Mollineda et al(2002)Mollineda, Vidal, and Casacuberta] (77.4%), edit-dist kernel [Neuhaus and Bunke(2006)] (81.1%) and contour fragments [Daliri and Torre(2009)] (84.5%).

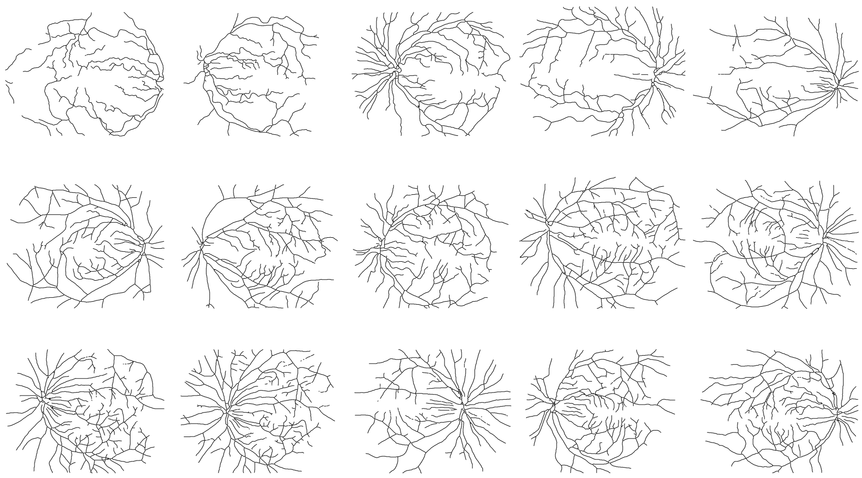

Fig. 8 Examples of retinal centrelines. Top row: AMD subjects; middle row: diabetic subjects. bottom row: normal subjects.

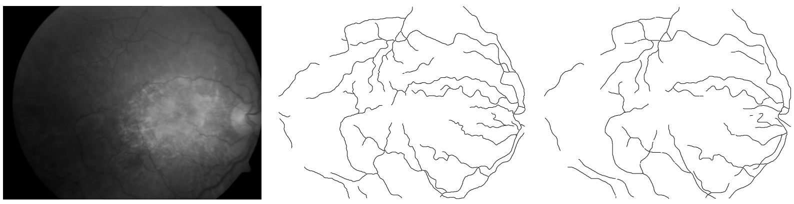

Fig. 9 Examples of retinal centrelines traced by two experts from the same source fundus image.

[1] Compactness, eccentricity, fractal dimension, roundness, Hu's moment invariants, affine moment invariants [Flusser and Suk(1993)], circularity [Haralick(1974)] and ellipticity [Proffitt(1982),Peura and Iivarinen(1997)]. Several other global shape descriptors were considered, such as elongatedness, rectangularity, and convexity, but were not found to improve classification accuracy.

Figure 8 shows examples of blood vessel centrelines traced from retinal fundus images by experts as part of the ARIA project.[2] For some subjects, their images were traced by two experts, producing slightly differing results, see figure 9. The data is organized into three categories: age-related macular degeneration (AMD) subjects, diabetic subjects and normals, and we have performed classification into these three categories using $\mathcal{L}_{comp}(\mathcal{C})$ calculated from the traced compound curves.[3] The number of centrelines (including instances of two tracings per subject) is 53 (AMD subjects), 118 (diabetic subjects) and 122 (normals). Classification was performed using an SVM [Chang and Lin(2011)] with a Radial Basis Function (RBF) kernel, and cross validation, and 57.0% accuracy was achieved. For comparison, the compound curves were also analyzed as if they were a collection of unrelated simple curves. Linearity was computed on each branch of the compound curve and then a single overall linearity score was taken as the weighted average of the individual branch linearity values, where the weights are the branch lengths. Following Note 3, linearity of the simple open curves was calculated using $\mathcal{L}_{comp}(\mathcal{C})$. Only 48.8% accuracy was achieved, showing that the compound linearity measure captures some useful global shape information unavailable to the non-compound measure.[4]

6 Conclusion

This paper has described a new approach to compute 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.

Acknowledgements J. Pantović and J. Žunić are also with the Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade. This work is partially supported by the Serbian Ministry of Science and Technology/projects OI174026/OI174008. Initial results of this paper were presented in [Rosin et al(2012)Rosin, Pantović, and Žunić]. We would like to thank the following for providing data used in this paper: Zicheng Liu (MSR Action3D Dataset), Andreu-García Gabriela (chicken pieces), St Paul's Eye Unit, Royal Liverpool University Hospital (ARIA Retinal Image Archive).

[2] We have post-processed the original vessel tracings by applying thresholding and thinning.

[3] Some of the images contain small portions of disconnected curves in addition to the main single compound connected curve. Since such curves do not satisfy the requirements of $\mathcal{L}_{comp}(\mathcal{C})$ then it is possible that $\mathcal{L}_{comp}(C) > 1$. However, in practice this did not occur for any of the retinal images (the largest computed value is 0.289). In any case, $\mathcal{L}_{comp}(C) > 1$ can still be used as a shape descriptor that is invariant with respect to similarity transformations.

[4] It is difficult to compare classification accuracy for the ARIA data set against other methods as the majority of the literature using this data set concentrates on vessel segmentation rather than disease classification. Hijazi *et al.* [Hijazi et al(2010)Hijazi, Coenen, and Zheng] did perform classification, but only considered two classes (86 AMD and 56 normal images), achieving 75% accuracy. Moreover, they treated the blood vessels as noise and eliminated them, analyzing instead image intensity histograms. In order to more directly compare our results with theirs, we removed the diabetic cases to create a two class problem, achieving 80.6% accuracy. Since traced centerlines were not provided for the full image database, we also tested the full set of source images for the two classes (92 AMD and 60 normals), and extracted blood vessel centerlines using a general purpose ridge detector [Steger(1998)]. Although the quality of the vessel detection was not high it captured some important structural information from the images, and the classification accuracy obtained using $\mathcal{L}_{comp}(\mathcal{C})$ was 82.2%.

References

[Andreu et al(1997)Andreu, Crespo, and Valiente] Andreu G, Crespo A, Valiente J (1997) Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recognition. In: Int. Conf. Neural Networks, vol 2, pp 1341–1346

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

[Black et al(2012)Black, Perron, Burr, and Drummond] Black B, Perron J, Burr D, Drummond S (2012) Estimating erosional exhumation on Titan from drainage network morphology. Journal of Geophysical Research 117(E08006)

[Chang and Lin(2011)] Chang C, Lin C (2011) LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3):27:1–27:27

[Chen and Chen(2011)] Chen Y, Chen W (2011) Morphology of Quanzhou city road network based on space syntax. Tropical Geography 6

[Daliri and Torre(2009)] Daliri M, Torre V (2009) Classification of silhouettes using contour fragments. Computer Vision and Image Understanding 113(9):1017–1025

[DeCarlo and Shokri(2014)] DeCarlo K, Shokri N (2014) Effects of substrate on cracking patterns and dynamics in desiccating clay layers. Water Resources Research 50(4):3039–3051

[Flusser and Suk(1993)] Flusser J, Suk T (1993) Pattern recognition by affine moment invariants. Pattern Recognition 26:167–174

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

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

[Goo and Lim(2012)] Goo B, Lim C (2012) Thermal fatigue of cast iron brake disk materials. Journal of Mechanical Science and Technology 26(6):1719–1724

[Haralick(1974)] Haralick R (1974) A measure for circularity of digital figures. IEEE Trans on Systems, Man and Cybernetics 4:394–396

[Hijazi et al(2010)Hijazi, Coenen, and Zheng] Hijazi M, Coenen F, Zheng Y (2010) Retinal image classification using a histogram based approach. In: Int. Joint Conf. on Neural Networks, pp 1–7

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

[Lahlil et al(2013)Lahlil, Li, and Xu] Lahlil S, Li W, Xu J (2013) Crack patterns morphology of ancient Chinese wares. The Old Potter's Almanack 18(1):1–9

[Matoušek and Nešetril(1998)] Matoušek J, Nešetril J (1998) Invitation to Discrete Mathematics. Clarendon Press, Oxford

[Mollineda et al(2002)Mollineda, Vidal, and Casacuberta] Mollineda R, Vidal E, Casacuberta F (2002) A windowed weighted approach for approximate cyclic string matching. In: Int. Conf. Pattern Recognition, vol 4, pp 188–191

[Neuhaus and Bunke(2006)] Neuhaus M, Bunke H (2006) Edit distance-based kernel functions for structural pattern classification. Pattern Recognition 39(10):1852–1863

[Peura and Iivarinen(1997)] Peura M, Iivarinen J (1997) Efficiency of simple shape descriptors. In: *et al* CA (ed) Aspects of Visual Form Processing, World Scientific, pp 443–451

[Proffitt(1982)] Proffitt D (1982) The measurement of circularity and ellipticity on a digital grid. Pattern Recognition 15(5):383–387

[Rosin et al(2012)Rosin, Pantović, and Žunić] Rosin P, Pantović J, Žunić J (2012) Measuring linearity of closed curves and connected compound curves. In: 11th Asian Conference on Computer Vision, ACCV (3), Lecture Notes in Computer Science, vol 7726, Springer, pp 310–321

[Schmitz et al(2011)Schmitz, Hjorth, Joemai, Wijntjes et al] Schmitz S, Hjorth J, Joemai R, Wijntjes R, et al (2011) Automated analysis of neuronal morphology, synapse number and synaptic recruitment. Journal of neuroscience methods 195(2):185–193

[Steger(1998)] Steger C (1998) An unbiased detector of curvilinear structures. IEEE Trans on Patt Anal and Mach Intell 20(2):113–125

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

[Žunić and Martinez-Ortiz(2009)] Žunić J, Martinez-Ortiz C (2009) Linearity measure for curve segments. Applied Mathematics and Computation 215(8):3098–3105
[Žunić and Rosin(2011)] Žunić J, Rosin PL (2011) Measuring linearity of open planar curve segments. Image Vision Computing 29(12):873–879