GEOMETRIC ANALYSIS OF PLANAR SHAPES WITH APPLICATIONS TO CELL DEFORMATIONS

XIMO GUAL-ARNAU[✉,1], SILENA HEROLD-GARCÍA[2] AND AMELIA SIMÓ[3]

[1]Departament de Matemàtiques-INIT. Universitat Jaume I. E-12071 Castellón. Spain.; [2]Department of Computing. Universidad de Oriente. Cuba.; [3]Departament de Matemàtiques-IMAC. Universitat Jaume I. E-12071 Castellón. Spain.
e-mail: gual@uji.es, simo@uji.es, silena@csd.uo.edu.cu

ABSTRACT

Shape analysis is of great importance in many fields such as computer vision, medical imaging, and computational biology. In this paper we focus on a shape space in which shapes are represented by means of planar closed curves. In this shape space a new metric was recently introduced with the result that this shape space has the property of being isometric to an infinite-dimensional Grassmann manifold of 2-dimensional subspaces. Using this isometry it is possible, from Younes *et al.* (2008), to explicitly describe geodesics, a task that previously was not at all easy. Our aim is twofold, namely: to use this general theory in order to show some applications to the study of erythrocytes, using digital images of peripheral blood smears, in the treatment of sickle cell disease; and, since normal erythrocytes are almost circular and many Sickle cells have elliptical shape, to particularize the computation of geodesics and distances between shapes using this metric to planar objects considered as deformations of a template (circle or ellipse). The applications considered include: shape interpolation, shape classification, and shape clustering.

Keywords: cell deformation, geodesics, planar closed curves, radius-vector function, shape space.

INTRODUCTION

Shapes play an important role in understanding objects and shape analysis has impact in many areas such as computer vision, medical imaging, and computational topology. Historically, there have been different geometrical characterizations of planar shapes and, in particular, different approaches to represent the continuous boundaries as curves, and then study their shapes.

In this paper we consider the space of planar shapes represented by simple closed plane curves with the metric introduced in Younes *et al.* (2008). This space has the property of being isometric to an infinite-dimensional Grassmann manifold of 2-dimensional subspaces. Therefore, explicit geodesics and distances between shapes can be computed using Jordan angles (Neretin, 2001). In fact, the representation of shapes as elements of infinite-dimensional spaces with a given metric is of interest at this time and has important applications (Klassen, 2004; Huckemann, 2011; Srivastava, 2011).

Furthermore, erythrocytes shape deformations are related to different illnesses, *e.g.*, sickle cell disease (SCD), that cause the hardening or polymerization of the hemoglobin that contains the erythrocytes. The cells are deformed and this results in a risk of various complications, for example vaso-occlusive crisis. The

study of this deformation process using digital images of peripheral blood smears offers useful results in the clinical diagnosis of these illnesses.

The first aim of this paper is to apply the distance and geodesics derived from the metric introduced in Younes *et al.* (2008) in order to study different applications related to the morphological analysis of shape deformation of erythrocytes (which are represented as planar curves). In particular, three applications are presented here: interpolation between shapes; supervised classification and unsupervised clustering.

On the other hand, curve evolution of a simple closed curve, whose points move in the direction of the normal with a prescribed velocity, has been applied to a wide variety of problems such as smoothing of shapes, shape analysis and shape recovery. The second aim of the paper consists on considering geodesics between closed-template curves and boundaries (curves) obtained as deformations of these templates. The idea of describing objects – such as cells or leaves – as deformations of a template also appears in the literature. For instance, in Granader and Manbeck (1993), the authors use an ellipse with fixed eccentricity as template in an application in the detection of defects in potatoes; and in Hobolth *et al.* (2002), the planar objects considered are deformations of a star-shaped template. Since

normal cells are almost circular and many elongated cells have elliptical shape, we will represent the shapes of cells as deformations of ellipses and circles (radius vector function). In fact, Fourier expansion of the radius-vector function has already been used in many applications; see for instance Lestrel (1997) for a review of biological applications, and Loncaric (1998) for a survey of the engineering literature. A statistical application can be found in Hobolth *et al.* (2003), and a review of several contour functions, including the radius-vector function, appears in Kindratenko (2003).

The paper is organized as follows: In the following section we give a basic résumé of the results obtained in Younes *et al.* (2008). Next, we define planar shape changes from a deformation of a template and we apply the results cited in the previous section to these particular curves. We consider the particular case where the template curve is an ellipse or a circle. Finally, we use the general results in Younes *et al.* (2008) and the results of section of the previous section, to solve different problems related to the morphological study of shape deformation of erythrocytes: interpolation between normal and sickle cells; supervised classification and unsupervised clustering. We also compare the advantages of using the general theory or this theory, particularized to deformation curves, in the study of erythrocytes.

THE SPACE OF PLANE SHAPES

In this paper we consider closed curves which are boundaries of planar shapes. That is, we consider the space of smooth planar immersed curves given by

$$M = \{\alpha \in C^\infty(\mathbb{S}^1, \mathbb{R}^2) : |\alpha'(t)| \neq 0, \forall t \in \mathbb{S}^1\},$$

where $\mathbb{S}^1$ is the unit circle and $\alpha'(t)$ is the usual parametric derivative of α. M denotes the space of C^∞-immersions $\alpha : [0, 2\pi] \longrightarrow \mathbb{R}^2$ with $\alpha(0) = \alpha(2\pi)$. The tangent space $T_\alpha M$ at α is the set of vector fields h on α; *i.e.*, $h : \mathbb{S}^1 \longrightarrow \mathbb{R}^2$.

If $h, k \in T_\alpha M$, we consider the Riemannian metric $G_\alpha(h, k)$ defined as Younes *et al.* (2008)

$$G_\alpha(h, k) = \frac{1}{l(\alpha)} \int_{\mathbb{S}^1} \dot{h}(s) \bullet \dot{k}(s) \, ds,$$

where $\dot{h}(s)$ means derivative with respect to arc length, $\dot{h}(s) \bullet \dot{k}(s)$ is the usual product in $\mathbb{R}^2$, and $l(\alpha)$ is the length of α.

The space of planar shapes will be the set M modulo translations, rotations, scalings, and reparameterizations of the curve ($Diff(\mathbb{S}^1)$). However we consider first the *pre-shape* space M_d, where the division by the group of diffeomorphisms $Diff(\mathbb{S}^1)$ has not been considered. The group generated by translations, scalings and rotations is called the group of similitudes, abbreviated as *sim*; then, the *pre-shape* space is defined as

$$M_d = \frac{M}{sim},$$

and we associate to M_d the restriction of the metric G_α.

We prefer to use this apparently complex metric rather than the much simpler metric $G_\alpha^0(h, k) = \int_{\mathbb{S}^1} h(s) \bullet k(s) \, ds$ (used, for instance, in Klassen (2004)) because the latter presents some problems – for example, a geodesic starting at a point can degenerate to a curve of vanishing length and geodesics cannot be computed explicitly. Moreover, as we will see next, M_d with this metric has the property of being isometric to a Grassmannian, and then distances and geodesics using this metric can be computed efficiently.

THE BASIC MAPPING

Let V be the vector space of all C^∞ mappings $f : \mathbb{S}^1 \longrightarrow \mathbb{R}$, with the norm

$$||f||^2 = \int_0^{2\pi} (f(x))^2 \, dx.$$

Then, given two functions $e, f \in V$ and assuming that our plane curves are curves in the complex plane $\mathbb{C}$, the basic mapping is defined as

$$\Phi : (e, f) \longrightarrow \alpha(t) = \frac{1}{2} \int_0^t (e(x) + if(x))^2 \, dx. \quad (1)$$

From this definition it follows that a curve $\alpha(t) = \Phi(e, f)$ will be a closed curve ($\alpha(0) = \alpha(2\pi)$) if and only if $||e|| = ||f||$ and $\langle e, f \rangle = \int_0^{2\pi} e(x) f(x) \, dx = 0$. Moreover the length of α is given by

$$l(\alpha) = \frac{1}{2} \int_0^{2\pi} (e(x)^2 + f(x)^2) \, dx = \frac{1}{2} \left(||e||^2 + ||f||^2 \right).$$

Let us now consider the Grassmannian $Gr(2, V)$ of unoriented 2-dimensional subspaces of V defined by an orthonormal pair $(e, f) \in V^2$ with $||e||^2 + ||f||^2 = 2$ (that is, $l(\alpha) = 1$); and let $Gr^0(2, V)$ be the subset of $Gr(2, V)$ defined by e, f with $\{t / e(t) = f(t) = 0\} = \emptyset$ (α need not be an immersion if e and f vanish simultaneously). f Although the mapping Φ has been defined on $V \times V$, it can be restricted to orthonormal pairs in $V \times V$; and then modified so as to divide out by rotations to get a new map (which is still denoted Φ) defined on $Gr^0(2, V)$ (Younes *et al.*, 2008). Then the magic of the map Φ is shown in the following result:

Theorem 1 Φ *defines an isometry*

$$\Phi : Gr^0(2,V) \longrightarrow M_d$$

when $Gr^0(2,V)$ is given with its natural metric and M_d is given with the metric G_α.

Given two functions e and f which define a point in the Grassmann manifold $Gr^0(2,V)$, from the basic mapping Φ, we obtain a closed curve that is taken as the representative of its equivalence class in M_d. On the other hand, given a curve $\alpha \in M$ we obtain the functions e and f, which correspond to the representative of α in M_d, from the following result:

Proposition 1 *Let $\alpha \in M$ and let $\theta_\alpha(t)$ be the tangent angle function of α; it follows that*

$$
\begin{aligned}
e(t) &= \sqrt{\frac{2|\alpha'(t)|}{l(\alpha)}} \cos\left(\frac{\theta_\alpha(t)}{2}\right), \\
f(t) &= \sqrt{\frac{2|\alpha'(t)|}{l(\alpha)}} \sin\left(\frac{\theta_\alpha(t)}{2}\right).
\end{aligned}
\tag{2}
$$

Proof 1 *From the definition of tangent angle function*

$$\alpha'(t) = |\alpha'(t)|e^{i\theta_\alpha(t)} = |\alpha'(t)|\left(e^{i\frac{\theta_\alpha(t)}{2}}\right)^2. \tag{3}$$

On the other hand, from Eq. 1, we have

$$\alpha'(t) = \frac{1}{2}(e(t) + if(t))^2. \tag{4}$$

Finally using the Euler formula and comparing Eqs. 3 and 4 we obtain the result. $\square$

As was said before, the curve given by $\Phi(e,f)(t)$ will be the representative curve of $\alpha \in M$ in M_d. Since $\Phi(e,f)(0) = 0$, the representative elements of M_d satisfy that $\alpha(0) = 0$ and $l(\alpha) = 1$.

COMPUTATION OF DISTANCES AND GEODESICS IN M_d

In order to compute distances and geodesic lines between any pair of shapes (closed planar curves) in the pre-shape space, we will consider the basic mapping and the distances and explicit geodesics in the Grassmannian given in Neretin (2001) and Younes *et al.* (2008).

The geodesic distance between $\alpha = \Phi(e_1,f_1)$ and $\beta = \Phi(e_2,f_2)$, is the distance between the two dimensional subspaces W_1, generated by $\{e_1,f_1\}$, and W_2, generated by $\{e_2,f_2\}$.

The singular value decomposition of the orthogonal projection p of W_1 in W_2 gives orthonormal bases $\{\hat{e}_1, \hat{f}_1\}$ of W_1 and $\{\hat{e}_2, \hat{f}_2\}$ of W_2 such that $p(\hat{e}_1) = \lambda_1 \hat{e}_2$ and $p(\hat{f}_1) = \lambda_2 \hat{f}_2$, $\hat{e}_1 \perp \hat{f}_2$, $\hat{f}_1 \perp \hat{e}_2$ where $0 \le \lambda_1, \lambda_2 \le 1$. In fact, λ_1 and λ_2 are the singular values of the $(2 \times 2)-$matrix

$$A = \begin{pmatrix} \langle e_1,e_2 \rangle & \langle e_1,f_2 \rangle \\ \langle f_1,e_2 \rangle & \langle f_1,f_2 \rangle \end{pmatrix}. \tag{5}$$

If we write $\lambda_1 = \cos\psi_1$, $\lambda_2 = \cos\psi_2$ then ψ_1, ψ_2 are the Jordan angles, $0 \le \psi_1, \psi_2 \le \pi/2$, and the geodesic distance between $\alpha = \Phi(e_1,f_1)$ and $\beta = \Phi(e_2,f_2)$ is given by Neretin (2001)

$$d(\alpha,\beta) = d(W_1,W_2) = \sqrt{\psi_1^2 + \psi_2^2}. \tag{6}$$

An upper bound of this distance is given by $d(\alpha,\beta) \le \pi/\sqrt{2}$.

The geodesic joining α and β is defined by

$$
\begin{aligned}
\gamma(t,u) &= \Phi(e(t,u),f(t,u)) \\
&= \frac{1}{2}\int_0^t \left(e(s,u) + if(s,u))^2\right) ds, \quad t \in [0,2\pi],
\end{aligned}
\tag{7}
$$

where

$$e(t,u) = \frac{\sin((1-u)\psi_1)\hat{e}_1(t) + \sin(u\psi_1)\hat{e}_2(t)}{\sin\psi_1}, \tag{8}$$

$$f(t,u) = \frac{\sin((1-u)\psi_2)\hat{f}_1(t) + \sin(u\psi_2)\hat{f}_2(t)}{\sin\psi_2}. \tag{9}$$

Therefore, if $\alpha = \Phi(e_1,f_1)$ and $\beta = \Phi(e_2,f_2)$, in order to obtain the geodesic we have to diagonalize the matrix A by rotating the curve α by a constant angle ϕ_α, *i.e.*, the basis (e_1,f_1) by the angle $\phi_\alpha/2$; and similarly the curve β by a constant angle ϕ_β. The angles ϕ_α and ϕ_β are given by the equations

$$\phi_\alpha + \phi_\beta = 2\arctan\left(\frac{\langle f_1,e_2 \rangle + \langle e_1,f_2 \rangle}{\langle f_1,f_2 \rangle - \langle e_1,e_2 \rangle}\right), \tag{10}$$

$$\phi_\alpha - \phi_\beta = 2\arctan\left(\frac{\langle f_1,e_2 \rangle - \langle e_1,f_2 \rangle}{\langle f_1,f_2 \rangle + \langle e_1,e_2 \rangle}\right). \tag{11}$$

In the newly aligned basis $\{\hat{e}_1, \hat{f}_1\} = e^{i\phi_\alpha/2}(e_1 + if_1)$, $\{\hat{e}_2, \hat{f}_2\} = e^{i\phi_\beta/2}(e_2 + if_2)$, the diagonal elements of the matrix will be the cosines $\lambda_1 = \cos\psi_1$ and $\lambda_2 = \cos\psi_2$.

THE SHAPE SPACE

We are interested in geometric curves, *i.e.*, curves considered up to reparameterizations. The shape space is then defined as the quotient space

$$S = \frac{M_d}{Diff(\mathbb{S}^1)}.$$

The map Φ can be converted in an isometry if we consider freely immersed curves and we make both quotients into Riemannian submersions in Theorem 1. However, we will consider only diffeomorphisms that are translations. We will consider digital curves which correspond to cell boundaries with a fixed orientation and approximately equally spaced discrete points have been considered in these curves, we suppose that the curves $\alpha \in M_d$ have constant speed $|\alpha'(t)|$ for all $t \in [0, 2\pi)$. Consider now that $\phi \in Diff^+(\mathbb{S}^1)$ (C^∞ orientation preserving diffeomorphisms of $\mathbb{S}^1$) with $\phi(0) = 0$, $\alpha(t)$ is a curve in M_d with $|\alpha'(t)| = K$ and $\alpha \circ \phi$ is a reparameterization of c with $|(\alpha \circ \phi)'(t)| = K$; then we have that ϕ is the identity, $\phi(t) = t$ for all $t \in \mathbb{S}^1$.

Therefore, instead of working in the shape space S we will consider distances in the pre-shape space M_d and we will define the distance between two closed immersions α and β, each of length 1, representing two shapes, as

$$d(\alpha, \beta) = \min_{\phi} d(\alpha, \widetilde{\beta \circ \phi}) , \qquad (12)$$

where $\phi \in Diff^+(\mathbb{S}^1)$ is given by $\phi(t) = t + t_0$, where $t_0 \in \mathbb{S}^1$ is a constant, and $\widetilde{\beta \circ \phi}$ is the representative element of $\beta \circ \phi$ in M_d.

With the present approach (Younes *et al.* (2008)) and the use of constant speed parameterizations, the task of finding distances and geodesics in the shape space (that is, factoring out diffeomorphisms on $\mathbb{S}^1$) is direct and does not require of sophisticated numerical methods, as in the shape space of Klassen (2004, Section 4), or in the shape space of Younes *et al.* (2008, Section 6) for general parameterized closed curves.

GEOMETRIC REPRESENTATION OF THE DEFORMATION OF A TEMPLATE

Curve evolution of a simple closed curve, whose points move in the direction of the normal with a prescribed velocity, has been applied to a wide variety of problems. In this section we will consider geodesic paths between a template curve (circle or ellipse) and a curve obtained as a normal deformation of the template. In the applications we will consider two approaches; the use of the general theory (Section "The space of plane shapes") and the results presented in this section, particularized to deformation curves.

Therefore, in this section we consider two shapes: a template curve α and a curve β obtained as a normal deformation of α. We plan to obtain suitable expressions for the functions $\{e_1(t), f_1(t)\}$ and $\{e_2(t), f_2(t)\}$ such that $\alpha = \Phi(e_1, f_1)$ and $\beta = \Phi(e_2, f_2)$.

Let α be a closed curve of length 1 parameterized by arc length and let β be a curve defined as

$$\beta(t) = \alpha(t) + f(t)\vec{n}(t) , \qquad (13)$$

where $\vec{n}(t)$ is the inner unit normal vector to $\alpha(t)$ and $f(t)$ is a differentiable function which represents the Euclidean signed distance between the points $\alpha(t)$ and $\beta(t)$.

Suppose that $\alpha(t) = (x(t), y(t))$ and $\vec{n}(t) = (n_1(t), n_2(t))$; then, identifying $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$, $\beta(t)$ can be expressed as

$$\beta(t) = (x(t) + f(t)n_1(t)) + \mathrm{i}(y(t) + f(t)n_2(t)) .$$

Proposition 2 *Let $\theta_\beta(t)$ be the tangent angle function of β; then,*

$$e_2(t) = \frac{1}{\sqrt{l(\beta)}} \sqrt{\frac{2|1 - f(t)\kappa_\alpha(t)|}{|\cos\theta(t)|}} \cos\left(\frac{\theta_\beta(t)}{2}\right) ,$$

$$f_2(t) = \frac{1}{\sqrt{l(\beta)}} \sqrt{\frac{2|1 - f(t)\kappa_\alpha(t)|}{|\cos\theta(t)|}} \sin\left(\frac{\theta_\beta(t)}{2}\right) ,$$

where κ_α denotes the curvature of α and $\theta(t) = \theta_\beta(t) - \theta_\alpha(t)$.

Proof 2 *Let $\vec{t}(t) = \alpha'(t)$ be the unit tangent vector. From the definition of β (Eq. 13), and using the Frenet-Serret formulas $\vec{t}'(t) = \kappa_\alpha(t)\vec{n}(t)$ and $\vec{n}'(t) = -\kappa_\alpha(t)\vec{t}(t)$, we obtain*

$$\beta'(t) = (1 - f(t)\kappa_\alpha(t))\vec{t}(t) + f'(t)\vec{n}(t) .$$

Since $f'(t) = \tan\theta(t)(1 - f(t)\kappa_\alpha(t))$, (Hobolth and Vedel-Jensen, 2000), we obtain

$$|\beta'(t)| = |1 - f(t)\kappa_\alpha(t)| \sqrt{1 + \tan^2\theta(t)} .$$

Finally, since

$$\beta'(t) = |\beta'(t)|\mathrm{e}^{\mathrm{i}\theta_\beta(t)} \text{ and } \beta'(t) = \frac{1}{2}(e_2(t) + \mathrm{i}f_2(t))^2 ,$$

proceeding as in Proposition 1 we obtain the result. □

Then, from Eq. 7, we obtain the geodesic in the shape space joining α to β.

174

ELLIPSES AND CIRCLES AS TEMPLATES AND THE RADIUS VECTOR FUNCTION

Since normal cells are almost circular and most sickle cells are almost elliptical, we study in this section, as a particular case, shapes of objects obtained as normal deformations of ellipses and shapes of objects obtained as normal deformations of circles.
Since ellipses and circles are boundaries of convex sets we will parameterize them using their support functions. In the second case because deformations of a circle are boundaries of star shaped objects, we will also express them in terms of their radius-vector functions.

Let $p(t)$ be the support function of a bounded convex set K; then the boundary of K, ∂K; can be parameterized as $\alpha(t) = (x(t), y(t))$ where (Santaló, 1976),

$$
\begin{aligned}
x(t) &= p(t)\cos t - p'(t)\sin t , \\
y(t) &= p(t)\sin t + p'(t)\cos t .
\end{aligned}
\tag{14}
$$

Proposition 3 *The two functions associated to the curve α by the basic mapping are:*

$$
\begin{aligned}
e_1(t) &= \sqrt{\frac{p(t)+p''(t)}{l(\alpha)}}\left(\cos\frac{t}{2} - \sin\frac{t}{2}\right), \\
f_1(t) &= \sqrt{\frac{p(t)+p''(t)}{l(\alpha)}}\left(\cos\frac{t}{2} + \sin\frac{t}{2}\right).
\end{aligned}
\tag{15}
$$

Proof 3 *Under the identification of $\mathbb{R}^2$ and $\mathbb{C}$, we have*

$$
\begin{aligned}
\alpha'(t) &= (p(t)+p''(t))(-\sin t + \mathrm{i}\cos t) \\
&= (p(t)+p''(t))\,\mathrm{i}\,(\cos t + \mathrm{i}\sin t) \\
&= (p(t)+p''(t))\,\mathrm{i}\,e^{\mathrm{i}t} \\
&= \frac{1}{2}\left(\sqrt{p(t)+p''(t)}(1+\mathrm{i})e^{\mathrm{i}t/2}\right)^2 .
\end{aligned}
$$

Then, proceeding as in Proposition 2 and, as we want the length of the curve α to be equal 1, we obtain the result. □

When the curve α is an ellipse $\alpha(t) = \{a\cos(t), b\sin(t)\}$, $t \in [0, 2\pi]$, the support function is

$$
p(t) = \sqrt{\frac{1}{2}\left[(a^2+b^2) + (a^2-b^2)\cos(2t)\right]} ,
$$

and the corresponding functions e_1, f_1 to α are known from the above proposition, and we do not have to compute them numerically. On the other hand,

$$
\kappa_\alpha(t) = \frac{ab}{(a^2\sin^2(t) + b^2\cos^2 t)^{3/2}} ,
$$

and

$$
\theta_\alpha(t) = \arctan\left(\frac{b}{a}\tan\left(t + \frac{\pi}{2}\right)\right) ,
$$

and Proposition 2 can be used to obtain the functions $\{e_2(t), f_2(t)\}$ corresponding to a curve β, obtained as a deformation of an ellipse.

For $a = b = 1$ the curve $\alpha(t) = \{\cos(t), \sin(t)\}$ is a circle and $\vec{\mathrm{n}}(t) = -\alpha(t)$; then,

$$
\beta(t) = \alpha(t) + f(t)\vec{\mathrm{n}}(t) = (1 - f(t))\alpha(t) .
$$

Therefore, $r(t) = 1 - f(t)$ is the radius vector function of the curve β with respect to the origin.

Now we present a way to obtain the functions $\{e_2(t), f_2(t)\}$ corresponding to a curve $\beta(t) = r(t)\{\cos(t), \sin(t)\}$.

Proposition 4 *Let $\theta_\beta(t)$ be the tangent angle function of β; then,*

$$
e_2(t) = \sqrt{\frac{2|r(t)|}{l(\beta)|\sin(\theta_\beta(t)-t)|}}\cos\left(\frac{\theta_\beta(t)}{2}\right) ,
$$

$$
f_2(t) = \sqrt{\frac{2|r(t)|}{l(\beta)|\sin(\theta_\beta(t)-t)|}}\sin\left(\frac{\theta_\beta(t)}{2}\right) .
$$

Proof 4 *Since $\kappa_\alpha = 2\pi$ and*

$$
\cos\theta(t) = \left\langle \alpha'(t), \frac{\beta'(t)}{|\beta'(t)|}\right\rangle ,
$$

with

$$
\alpha'(t) = \{-\sin t, \cos t\} ,
$$

and

$$
\frac{\beta'(t)}{|\beta'(t)|} = \{\cos\theta_\beta(t), \sin\theta_\beta(t)\} ,
$$

from Proposition 2 we obtain the result. □

In the second approach considered in the applications, we use two templates α: a unit circle and an ellipse, whose associated functions are given by Proposition 3. To compute the distance and geodesic from a cell to a circular template α using Proposition 4, we consider as radius vector $r(t)$ the distance from the centroid of the cell to its boundary; and to compute the distance and geodesic from a cell to an ellipse α as a template using Proposition 2, we consider that α is the ellipse that best fit the boundary of the cell. As expected, in general, the distance of a normal cell to the circle is almost zero and the distance from a sickle cell to the ellipse is almost zero.

APPLICATIONS

There are many interesting applications of the geometric representation of planar shapes proposed here. In this section we show some of these applications and we use them to analyze digital images of peripheral blood smears in the morphological study of erythrocytes.

All these applications represent a novelty for two reasons. Firstly, the novelty lies in the use of the general theory (Prop. 1, Eqs. 6 and 12) which has never been used for these specific applications. On the other hand, in some applications, additionally, we will make use of the computation of representatives for a normal deformation of an ellipse or a circle (Prop. 4).

Erythrocytes shape deformations are related to different illnesses, *e.g.*, SCD, that cause the hardening or polymerization of the hemoglobin that contains the erythrocytes. The cells are deformed and this results in a risk of various complications, for example vaso-occlusive crisis. In Fig. 1 we see images of normal and deformed cells, they will be used in the following applications. The study of this process using digital images of peripheral blood smears offers useful results in the clinical diagnosis of these illnesses.

The images were obtained of blood specimens from patients with SCD. They were of 500×375 pixels, with 480 ppi resolution. They were taken in the Clinical Laboratory of the Special Hematology Department of the "Dr. Juan Bruno Zayas Alfonso" General Hospital, Santiago de Cuba. In order to accomplish segmentation we used methods that are based on contour evolution and that appear in the literature under the name of active contours, deformable models, etc.

Three applications are presented here: interpolation between shapes; supervised classification and unsupervised clustering.

INTERPOLATION BETWEEN SHAPES

Geodesic paths between shapes cited in section "The space of plane shapes" can be used to interpolate between shapes. Interpolation between shapes can be useful to estimate intermediate shapes of cells which vary their shapes as time goes by, for example those affected by SCD. For example, given a normal and a sickle cell like those shown in Fig. 2, one can build the geodesic path between both shapes, as it is shown in this figure. In Fig. 3 some of the intermediate shapes in this process are shown.

On the other hand, taking into account that a normal cell should be 'circular', we can make profit of

Propositions 3 and 4 (where $r(t)$ is the distance from the centroid of the cell to its boundary). Using this option, calculations are considerably reduced. In Fig. 4, intermediate shapes between a "completely circular" shape and the same sickle cell of Fig. 3 are shown.

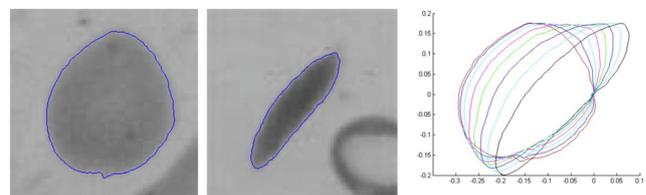Fig. 1. *Examples of normal, sickle and other deformed erythrocytes.*

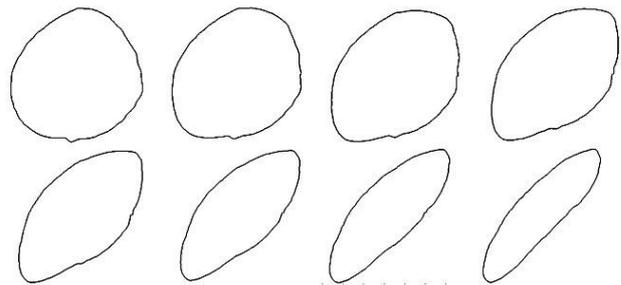Fig. 2. *Geodesics between normal and sickle erythrocyte.*

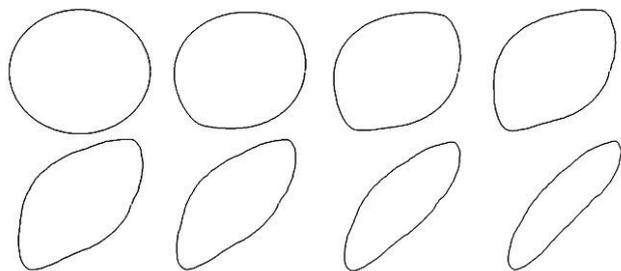Fig. 3. *Evolution from normal to sickle erythrocyte.*

Fig. 4. *Evolution from a circle to a sickle erythrocyte.*

SUPERVISED CLASSIFICATION

One of the most important issues in the study of SCD is the search for an efficient automatic classification method to quantify the number of deformed cells that a patient has and so gauge the severity of the illness. Nowadays, automatic

classification of erythrocytes using digital images is a very active field of research (Horiuchi *et al.*, 1990; Wheeless *et al.*, 1994; Asakura *et al.*, 1996; Kavitha and Ramakrishnan, 2005; Jayavanth *et al.*, 2010; Frejlichowski, 2010). All these studies are based on the extraction of shape descriptors (either boundary-based or region-based) and the use of Euclidean distance between these descriptors.

In this section, we begin by proposing the use of the distance between planar shapes cited previously as the metric in a supervised classification algorithm and demonstrated the excellent results obtained in the automatic classification of normal erythrocytes, those with sickle or those with other deformations.

Blood specimens were obtained from patients with SCD and 45 images of different fields were taken for this study. A specialist selected the cells to be studied, differentiating between normal ones, elongated ones and those with other deformations present in the images (Fig. 1).

To carry out the task of the supervised classification, we propose again two possibilities, the first one is based on the general approximation to distances between shapes (Proposition 1 and Eqs. 6 and 12). The second approach is based on considering the cells as deformations of known templates.

In the first approximation all the distances between pairs of cells were calculated and the k-nearest neighbor algorithm for supervised classification (k-NN) was used (Cover and Hart, 1967). In order to validate the results, a 5×1 cross-validation process was carried out. The confusion matrix of the process is showed in Table 1, and the sensibility, precision and specificity measures in Table 2.

Table 1. *Confusion matrix of supervised classification.*

Classes	Normal	Sickle	Other Deformed
Normal	202	0	0
Sickle	0	202	8
Other Deformed	25	8	178

Table 2. *Measures of supervised classification using the proposed distance.*

Classes	Sensibility	Precision	Specificity
Normal	100.00	88.99	90,26
Sickle	96,19	96,19	92,01
Other Deformed	84,36	95,70	98,06
General Efficiency	93,52	93,63	93,44

In Table 2 we see that the elongated class had high values of sensibility and specificity, with 96.19 % and 92.01 % respectively, and the normal class had 100.0 % of sensibility. No normal cell was classified as elongated nor vice versa. In the case of other deformations the sensibility decreased to 84.36 % due to several cells with other deformations having shapes very close to circular ones and in addition these elements differ a lot among themselves. This drop in accuracy is paralleled – though to a lesser extent – in the values given for precision and specificity in the normal class. Nevertheless, a very few elements from the "other deformations" class were classified as elongated, the majority of misclassifications being to consider them as normal. The process had a general performance of 93.53 %.

In order to asses the goodness of our results, they were compared with those obtained using other approaches previously considered in the literature on analysis of erythrocytes images. In particular with descriptors used in Wheeless *et al.* (1994), Asakura *et al.* (1996) and Frejlichowski (2010). The first one is named the UNL-Fourier method and it is based on applying two transformations, firstly, the transformation of boundary points to polar coordinates (more precisely, UNL-transform) and secondly 2D Fourier transformation. The second one is based on two shape factors: circular-shape factor (CSF) (Wheeless *et al.*, 1994) and elliptical-shape factor (ESF) (Asakura *et al.*, 1996). Results can be found in Table 3). In all the cases the proposed descriptors have similar or higher classification sensibility, specificity and precision.

Although the results obtained using the above methodology have been excellent, and the distances between all the pairs of shapes are almost explicitly available (this is, for instance, a key advantage over the shape space of Klassen (2004)); in order to reduce computational cost, we are going to consider now that normal and sickle cells correspond to known templates.

As we said before, normal cells are almost circular and all deformed cells correspond to deformations of a circle. However sickle cells are almost elliptical and we could assume that elliptical forms appear in their geodesic path from the circle. Considering this fact we propose to consider distances from each cell to a circle and to an ellipse in the classification method. The *e* and *f* functions of a circle and of an ellipse were obtained analytically, by using Eq. 15. Results given in section "Geometric representation of the deformation of a template" were used to calculate the representative of each cell and then the distances to the circle and to the ellipse. The axes of the ellipse used as a template

Table 3. *Sensibility, precision and specificity values for UNL-Fourier function and the elementary coefficients CSF and ESF.*

	UNL-F			CSF-ESF		
Class	Sensibility	Precision	Specificity	Sensibility	Precision	Specificity
Normal	96.59	92.97	90.59	100.00	63.90	73.05
Sickle	97.14	91.66	90.24	98.10	94.9	97.35
Other abnorm.	84.19	93.50	96.87	41.31	95.70	99.03

are obtained as the mean value of the axes of the ellipses that best fit each of the sickle cells in the sample.

Linear discriminant analysis algorithm for classification with Leave-one-out cross-validation was used (Rao, 2009). The confusion matrix of the process is showed in Table 4, and the sensibility, precision and specificity measures in Table 5. In Table 5 we see that a high overall performance is also obtained (92.39 %), although it is slightly worse than that obtained in the previous case. The sensibility of normal and elongated classes is also high: 98.02 % and 98.57 %, respectively, and once again no normal cell was classified as elongated nor vice versa.

Table 4. *Confusion matrix of classification considering the distances of each cell to a circle and an ellipse.*

Classes	Normal	Sickle	Other Deformed
Normal	198	0	4
Sickle	0	207	3
Other Deformed	24	17	170

Table 5. *Measures of classification considering considering the distances of each cell to a circle and an ellipse.*

Classes	Sensibility	Precision	Specificity
Normal	98.02	89,19	89,55
Sickle	98,57	92,41	89,10
Other Deformed	80,57	96,05	98,30
General Efficiency	92,39	92,55	92,32

UNSUPERVISED CLUSTERING

Another interesting problem in the study of SCD is that – in classifying cells using a more complex system – the simplicity of SCD diagnosis is made more complicated. This more complex classification system is of interest in the diagnosis of other diseases and their own particular cell deformations, a study that merits its own importance.

Several studies consider cellular deformations produced by various diseases and have been the theme of study and interest for some years now. Morphological classification in 6 classes (Bacus *et al.*, 1977), 12 classes (Frejlichowski, 2010) and 14 classes (Bacus *et al.*, 1976) have been used. Other studies considering different shapes of cells depending on varying flow conditions in different blood vessels were carried out (Jayavanth *et al.*, 2010). In all these studies homogeneous classes of deformations characterized by the morphology of the cells are defined.

In this section we propose to use distance between planar shapes cited previously as a dissimilarity measure in an unsupervised classification algorithm in order to define homogeneous classes of deformations.

As cluster procedure we will use a partitioning method called Partitioning Around Medoids (PAM). This method is a generalization of the well-known k-means algorithm, which can be used with all types of data and dissimilarity measurement between objects. The PAM algorithm (in a similar way to k-means) is based on finding k representative objects (also known as medoids (Kaufman and Rousseeuw, 1990)) from the data set in such a way that the total of the dissimilarities within any given cluster is minimized. Medoids are representative objects in the clusters that always exist and we just have to compute the dissimilarities between cells. Unlike K-means, there is no need to calculate cluster Frechet means (Pennec, 2006) that would be a very complex task in our shape space. A gradient search algorithm in a infinite dimensional Riemannian manifold is required to find the Frechet mean (Woods, 2003).

PAM algorithm can be found in the *Cluster* package of the free software, *R* (R Development Core Team, 2009), a language and environment for statistical computing and graphics.

In most applications of clustering procedures and in particular in the problem that concerns us here, the number of groups k is not known in advance. In order to select the appropriate number of groups, we will run the clustering algorithm with different

numbers of groups and we will choose (as suggested in Kaufman and Rousseeuw, 1990) the result with the largest average silhouette width.

We have to note that we want $k \geq 3$ because at least we have three classes: normal, sickle and other deformations, but previously (in order to validate the clustering procedure) we apply the method to a supervised case where $k = 3$. In Table 6 the proportions of correctly classified objects are shown and we can see that results are excellent, with a sensibility value of 97.00 % in the detection of sickle cells, 100.00 % in the case of normal cells, and a general performance of the process of 87.67 %.

Table 6. *Unsupervised classification using the proposed distance: 3 groups generated.*

Classes	Cell Quantity	Detected	Sensibility
Normal	100	100	100.00
Sickle	100	97	97.00
Other Deformed	100	66	66.00
General Efficiency			87.67

We then used our methodology for $k > 3$. The optimal clustering has been achieved with $k = 5$. For this number of groups, the average silhouette width is 0.1592418. Results are shown in Table 7. In Fig. 5 we can see the medoids of each class. The medoids objects detected in normal and elongated classes were the same as in the process using 3 groups. The new study (with $k = 5$) verified that the detection in these classes was stable.

The cells in the generated new groups mostly would have belong to the class of other deformations in our earlier ($k = 3$) study. Thus we can infer that metrics allows additional grouping of several types of other deformations that show up in a study that is adequately accurate. No normal cell was classified as sickle and vice versa, and in the case of other deformation classes, the misclassification occurring was in great majority of cases of cells detected as normal. Only one cell in the case of five groups was erroneously detected as sickle. Thus we can say that the results were good.

Table 7. *Unsupervised classification using the proposed distance: 5 groups generated.*

	Normal	Sickle	*G1*	*G2*	*G3*
Normal	100	0	0	0	0
Sickle	0	90	9	1	0
Other	12	1	24	27	36

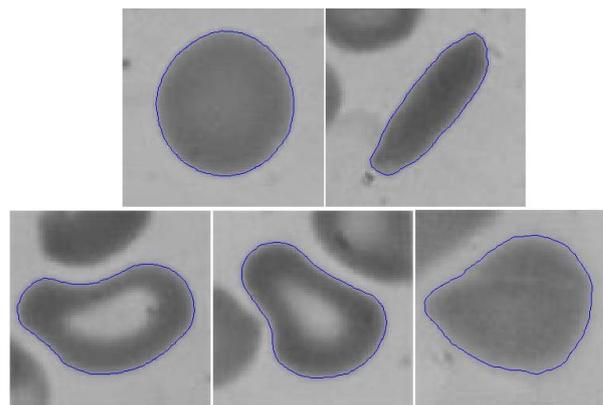

Fig. 5. *Medoids of each class obtained: Normal and Sickle classes above, the three classes of Other Deformations below.*

PSEUDOCODES AND NOTATION

PSEUDOCODES

The aim of this section is to present a description of both algorithms used to compute distances and geodesics between planar curves; from their general representation (Eq. 2), and between an ellipse (or circle) and a deformation of this template (Propositions 2-4). Since most of the points of both algorithms are identical, we will list the steps of the first algorithm, and then we discuss the differences with the second one.

Algorithm 1 (Distances and geodesics between general digital curves)

Given c_1 and c_2 two digital closed curves, each of them given by an ordered set of N points approximately equally spaced: $c_i = c_i(t) = \{x_i(t), y_i(t)\}_{t=1,\cdots,N}$, $i = 1,2$.

1. *Compute the lengths $L(c_i)$, $i = 1, 2$.*

2. *Compute the approximations of the functions given in Eq. 2, for $t = 1, \cdots, N$. The tangent angle functions $\theta_i(t)$ are derived from the cumulative angular function (Zhan et al., 1972) and $|c_i'(t)|$, although it should be constant, is approached from a second grade polynomial approximation using five points (Sauer, 2011).*

3. *Calculate the numerical integration of $\langle e_1, e_2 \rangle$, $\langle f_1, f_2 \rangle$, $\langle e_1, f_2 \rangle$ and $\langle f_1, e_2 \rangle$ from the corresponding sums of type $\sum_{t=1}^{N} e_1(t)e_2(t)$.*

4. *Compute the singular values λ_1 and λ_2 of the matrix A defined in Eq. 5 (λ_1 and λ_2 are the eigenvalues of the matrix $A^T A$, where A^T stands for the transpose matrix).*

5. *The distance between the parameterized digital curves c_1 and c_2 is $d(c_1, c_2) = \sqrt{\psi_1^2 + \psi_2^2}$, where $\lambda_1 = \cos \psi_1$, $\lambda_2 = \cos \psi_2$.*

6. *Consider all the reparameterizations of c_2 of the form $c_2(t + t_0)$, when $t_0 \in \{1, \cdots, N\}$ and go to point 3 (now $\langle e_1, f_2 \rangle \approx \sum_{t=1}^{N} e_1(t) f_2(t + t_0)$). Compute the N distances $d(c_1(t), c_2(t + t_0))$ and the minimum distance will be the distance between the geometric curves c_1 and c_2.*

7. *To compute the geodesic joining them, we consider the parameterization of c_2 which gives the minimum distance.*

   Solve Eqs.10 to 11 and, from the angles ϕ_α, ϕ_β, obtain the aligned basis $\{\hat{e}_1, \hat{f}_1\}$.

8. *From Eq. 7 we obtain the geodesic $\gamma(s, u)$ between c_1 and c_2. For $u = 0$ we obtain the curve c_1, for $u = 1$ we obtain the curve c_2, and for $0 < u < 1$ we obtain intermediate curves between c_1 and c_2. For $u > 1$ we obtain an extension of the geodesic beyond c_2. For each fixed value of the parameter u, the integral to obtain $\gamma(s, u)$ is approximated numerically.*

Algorithm 2 (Distances and geodesics using a circle or an ellipse as a template)

   In this case we consider c_1 either a circle or an ellipse (if the axes are unknown they can be estimated from the axes of the ellipses that best fit each curve in the sample). To compute the distance of a curve c_2 to the template c_1, we suppose the condition in Eq. 13. All steps are identical to the preceding algorithm except:

2. *If c_1 is an ellipse: compute the approximation of the functions e_2, f_2 from proposition 2.*

   If c_1 is a circle: compute the approximation of the functions e_2, f_2 from proposition 4.

   The corresponding functions e_1, f_1 are given from proposition 3.

6. *This step is not necessary in this case.*

7. *This step is not necessary when the template c_1 is a circle.*

NOTATION

   We present here a summary of some mathematical concepts and symbols that appear in the paper.

M	Space of smooth planar immersed curves
$\mathbb{S}^1$	Unit circle
$T_\alpha M$	Tangent space of M at α
G_α	Riemannian metric on M
$\bullet$	Usual product in $\mathbb{R}^2$
sim	Group generated by translations, scalings and rotations
M_d	Pre-shape space, $\frac{M}{sim}$
S	Space of planar shapes, $\frac{M_d}{Diff(\mathbb{S}^1)}$
V	Vector space of all C_∞ mappings $f : \mathbb{S}^1 \longrightarrow \mathbb{R}$
$\| \cdot \|^2$	Usual norm in V
$\langle \cdot, \cdot \rangle$	Usual product in V
Φ	Basic mapping, $\Phi(e, f)(t) = \frac{1}{2} \int_0^t (e(x) + i f(x))^2 dx$
$Gr(2, V)$	Grassmannian of 2-dimensional subspaces of V
$Gr^0(2, V)$	Subset of $Gr(2, V)$ with non vanishing simultaneously (e, f) elements
$\theta_\alpha(t)$	Tangent angle function of the curve α

CONCLUSIONS

In this paper we consider the metric introduced in Younes *et al.* (2008) and we apply it to calculate distances and geodesics between all pairs of a sample of closed planar curves, each one representing the boundary of an erythrocyte. From these distances and geodesics we show some practical applications in the morphological study of erythrocytes using digital images of peripheral blood smears of SCD:

1. Shape interpolation between erythrocytes.

2. Supervised classification of erythrocytes in normal cells, sickle cells, and cells with other deformations, with a general effectiveness of the process of 93.53 %

3. Unsupervised classification of erythrocytes where the detection of the groups of normal and sickle cells in relation to the supervised classification remained stable while three new class of deformations were differentiated.

On the other hand, since normal erythrocytes are almost circular and many sickle cells have elliptical shape, we have adapted the general theory of Younes *et al.* (2008) to the particular case where we have a known convex template (in particular, circle or ellipse) and a normal deformation of it. Now, we only have to calculate the distance and geodesic of each shape in the sample, to a unit circle and to an ellipse, which is obtained from the mean value of the axes of the ellipses that best fit each of the sickle cells in the sample.

For the same sample of cells as in the preceding case, we have performed a second experiment, using templates (a circle and an ellipse) with the goal of reducing computational cost. In this case, the general effectiveness of the supervised classification process of erythrocytes in normal cells, sickle cells, and cells with other deformations was of 92.39 %.

With these results it can be affirmed that the method is feasible for morphologic study applications and can be used as support to the clinical diagnosis of the state of a patient with SCD.

The methodologies introduced in this paper could be also extended to other similar clinical applications.

ACKNOWLEDGEMENTS

REFERENCES

Asakura T, Hirota T, Nelson AT, Reilly MP, Ohene-Frempong K (1996). Percentage of reversibly and irreversibly sickled cells are altered by the method of blood drawing and storage conditions. Blood Cell Mol Dis 22:297-306.

Bacus JW, Belanger MG, Aggarwal RK, Trobaugh FG (1976). Image processing for automated erythrocytes classification. J Hystochem Cytochem 24:195-201.

Bacus JW, Yasnoff WA, Belanger MG (1977). Computer techniques for cell analysis in hematology. Proc 1st Annu Symp Comput Appl Med Care (SCAMC), 1977 Oct 5, 24-35.

Cover TM, Hart PE (1967). Nearest neighbor pattern classification. IEEE Trans Inform Theory 13:21-7.

Frejlichowski D (2010). Pre-processing, extraction and recognition of binary erythrocyte shapes for computer-assisted diagnosis based on MGG images. International Conference on Computer Vision and Graphics, Poland 2010, Part I. Lect Not Comput Sci 6374:368-75.

Grenander U, Manbeck KM (1993). A stochastic shape and color model for defect detection in potatoes. J Comp Graph Statist 2:131-51.

Hobolth A, Vedel-Jensen EB (2000). Modelling stochastic changes in curve shape, with an application to cancer diagnostics. Adv Appl Prob (SGSA) 32:344-62.

Hobolth A, Pedersen J, Vedel-Jensen EB (2002). A deformable template model, with special reference to elliptical templates. J Math Imaging Vis 17:131-7.

Hobolth A, Pedersen J, Vedel-Jensen EB (2003). A continuous parametric shape model. Ann Inst Statist Math 55:227-42.

Horiuchi K, Ohata J, Hirano Y, Asakura T (1990). Morphologic studies of sickle erythrocytes by image analysis. J Lab Clin Med 115:613-20.

Huckemann SF (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. Ann Statist 39:1098-124.

Jayavanth S, Lee DH, Pak BC (2010). Multi-shape erythrocyte deformability analysis by imaging technique. J Mech Science Technol 24:931-5.

Kaufman L, Rousseeuw PJ (1990). Finding groups in data: an introduction to cluster analysis. New York: John Wiley.

Kavitha A, Ramakrishnan S (2005). Analysis on the erythrocyte shape changes using wavelet transforms. Clin Hemorheol Micro 33:327-35.

Kindratenko V (2003). On using functions to describe shapes. J Math Imaging Vis 18:225-45.

Klassen E, Srivastava A, Mio W, Joshi SH (2004). Analysis of planar shapes using geodesic paths on shape spaces.

IEEE T Pattern Anal 26:372-83.

Lestrel PE (1997). Fourier descriptors and their applications in biology. New York: Cambridge University Press.

Loncaric S (1998). A survey of shape analysis techniques. Pattern Recogn 31:983-1001.

Neretin YA (2001). On Jordan angles and the triangle inequality in Grassmann manifold. Geometriae Dedicata 86:81-92.

Osher S, Sethian JA (1998). Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation. J Comput Phys 79:12-49.

Pennec X (2006). Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. J Math Imaging Vis 25:127-54.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www. R-project.org/.

Rao CR (2009). Linear statistical inference and its applications. New York: John Wiley & Sons.

Sauer T (2011). Numerical analysis. 2nd Edition. Harlow: Pearson.

Santaló LA (1976). Integral geometry and geometric probability. London: Addison-Wesley.

Srivastava A, Klassen E, Joshi SH, Jermyn H (2011). Shape analysis of elastic curves in Euclidean spaces. IEEE T Pattern Anal 33:1415-28.

Wheeless L, Robinson R, Lapets OP, Cox C, Rubio A, Weintraub M, Benjamin LJ (1994). Classification of red-blood-cells as normal, sickle, or other abnormal, using a single image-analysis feature. Cytometry 17:159-66.

Woods RP (2003). Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. NeuroImage 18:769-88.

Younes L, Michor PW, Shah J, Mumford D (2008). A metric on shape space with explicit geodesics. Rend Lincei Mat Appl 39:25-57.

Zhan CT, Roskies RZ (1972). Fourier descriptors for plane closed curves. IEEE T Comput 21:269-81.