Image Anal Stereol 2020;39:281-291 doi:10.5566/ias.2183
Original Research Paper

MORPHOLOGICAL ANALYSIS OF CELLS BY MEANS OF AN ELASTIC METRIC IN THE SHAPE SPACE

Irene Epifanio✉, [1], Ximo Gual-Arnau[2] and Silena Herold-Garcia[3]

[1]Departament de Matemàtiques-IMAC-IF, Universitat Jaume I. 12071-Castelló, Spain, [2]Departament de Matemàtiques, Institute of New Imaging Technologies, Universitat Jaume I. 12071-Castelló, Spain, [3]Department of Computer Science, Universidad de Oriente. Cuba
e-mail: epifanio@uji.es, gual@uji.es, silena@uo.edu.cu
(Received May 17, 2019; revised February 26, 2020; accepted March 12, 2020)

ABSTRACT

Shape analysis is of great importance in many fields, such as computer vision, medical imaging, and computational biology. This analysis can be performed considering shapes as closed planar curves in the shape space. This approach has been used for the first time to obtain the morphological classification of erythrocytes in digital images of sickle cell disease considering the shape space $\mathscr{S}_1$, which has the property of being isometric to an infinite-dimensional Grassmann manifold of two-dimensional subspaces (Younes *et al.*, 2008), without taking advantage of all the features offered by the elastic metric related to the possibility of stretching and bending of the curves. In this paper, we study this deformation in the shape space, $\mathscr{S}_2$, which is based on the representation of closed planar curves by means of the square-root velocity function (SRVF) (Srivastava *et al.*, 2011), using the elastic metric of this space to obtain more efficient geodesics and geodesic lengths between planar curves. Supervised classification with this approach achieved an accuracy of 94.3%, classification using templates achieved 94.2% and unsupervised clustering in three groups achieved 94.7%, considering three classes of erythrocytes: normal, sickle, and with other deformations. These results are better than those previously achieved in the morphological analysis of erythrocytes and the method can be used in different applications related to the treatment of sickle cell disease, even in cases where it is necessary to study the process of evolution of the deformation, something that can not be done in a natural way in the feature space.

Keywords: elastic metric, erythrocyte deformation, geodesics, planar closed curves, shape space, SRVF.

INTRODUCTION

Shape analysis is a key step in many areas of computer vision, especially in medical imaging, since shapes are essential in understanding objects. Planar shapes can be geometrically characterized in different ways. In particular, there are different approaches to describe the continuous boundaries (contours) as curves and then analyze their shape.

In the field of medicine, the study of cell morphology using digital images offers useful results in the clinical diagnosis of certain illnesses, like sickle cell disease. This genetic disorder causes the hardening or polymerization of hemoglobin inside the cell, thus resulting in its deformation, which causes severe complications for the patient. The most frequent complication is pain crises, caused by vaso-occlusive events due to the lack of elasticity of the erythrocyte, which prevents it from circulating normally through the smaller capillaries.

Shape analysis can be performed with the use of spaces of planar shapes. A first approximation of the morphological analysis of erythrocytes with this approach was studied in Gual *et al.* (2015b), where the space of planar shapes represented by simple closed plane curves with the metric introduced in Younes *et al.* (2008) is employed in a practical way. This space has the property of being isometric to an infinite-dimensional Grassmann manifold of two-dimensional subspaces. As consequence, explicit geodesics and distances between shapes can be calculated using Jordan angles (Neretin, 2001). This previous work does not exploit all the advantages of the metric used, in relation to the possibility of considering both stretching and bending in the analysis of the curves. The algorithms to compute distances and geodesics for erythrocyte shape analysis with this approach can be found in Gual *et al.* (2015b).

This paper proposes a second approach of the use of space of planar shapes for morphological analysis of erythrocytes, based on a special representation of curves called the square-root velocity function, or SRVF, under which a specific elastic metric becomes an L^2 metric and simplifies the shape analysis (Srivastava *et al.*, 2011). This framework has not been used until now for erythrocyte shape analysis.

Related work: Automatic recognition and classification of erythrocytes considering their shapes

have been studied by the scientific community. Some techniques consider the use of features and descriptors to characterize the shape of the cell contour. A first group uses elementary descriptors, such as the circularity and elliptical factors, which do not offer the best results due to their simplicity (Wheeless *et al.*, 1994; Fernández *et al.*, 2013). Some other recent approaches propose ellipse fitting (Gonzalez-Hidalgo *et al.*, 2015), Hough transform (Mazalan *et al.*, 2013), circlet transform (Sarrafzadeh *et al.*, 2015), the computation of descriptors using Fourier series (Frejlichowski, 2010; Aziz , 2013), classification via artificial neural networks (Durant *et al.*, 2017; Rahmat *et al.*, 2018; Dalvi and Vernekar , 2016), the analysis of curvature changes (Delgado *et al.*, 2016) and the use of integral geometry-based functions to obtain efficient descriptors of the erythrocyte contour (Gual *et al.*, 2013; 2015a). All these approaches make it possible to compare and analyze the shape of erythrocytes in the feature space, but they are not suitable for computing shape statistics, such as mean shapes, or capturing the principal directions of shape variation within a specific class or between the cells in a normal and/or deformed cell class. This kind of studies can be performed in the shape space, which allows us to consider the cell contour as a point in this space and more sophisticated analyzes about shape can be made.

The first proposal that considers the space shape for analyzing the erythrocyte morphology was put forward by Gual *et al.* (2015b), where the metric introduced by Younes *et al.* (2008) is used and explicit geodesics and distances between shapes can be computed using Jordan angles (Neretin, 2001). This is the $\mathscr{S}_1$ shape space. The results obtained achieved 93.52% overall sensitivity in the supervised classification of cells and 92.39% in the classification considering distances of the cells to circle and ellipse templates, which drastically reduces the computational complexity of the process and maintains a high sensitivity value. In this proposal, the study is based completely on bending one curve into the other using linear interpolation of angle functions of the curves (with arc-length parameterization), so the advantages offered by the metric are not fully exploited. Due to this, the deformation process may not be optimal in some cases, because the behavior is similar to non-elastic metrics.

A Riemannian framework for analyzing planar shapes was developed by Srivastava *et al.* (2011). In this framework, the shape analysis is performed using the square root velocity function (SRVF) representation of curves in Euclidean spaces, under an elastic metric. This is the $\mathscr{S}_2$ space shape. With this kind of metric, geodesics and geodesic distances provide a framework for optimally matching, deforming, comparing, averaging, and inferring shapes, considering deformations of curves resulting not just from bending but from stretching as well. This approach has already been applied to various problems (Laga *et al.*, 2012; Srivastava *et al.*, 2011), but this is the first time it has been applied to the erythrocyte morphology classification problem, where intra-class and inter-class differences between certain types of erythrocytes can be seen as elastic deformations. Thus, quantifying such differences can be formulated as the problem of computing geodesics (registration) and geodesic lengths (similarity) between planar curves.

The main objective of this paper is to apply the distances and geodesics derived from the shape space and the metric introduced in Srivastava *et al.* (2011) in order to study different applications related to the morphological analysis of shape deformation of erythrocytes, for the first time, considering the elastic characteristics of the metric that takes into account both stretching and bending for shape analyzing; and to compare the results obtained with previous works introduced in Gual *et al.* (2015b). In particular, three applications are presented here: interpolation between shapes; supervised classification and unsupervised clustering.

This metric is part of a general family of elastic metrics (Mio *et al.*, 2006) which includes in particular the SRVF for the choice of parameters $a = 1$, $b = 1/2$, and the metric of Younes et al. for the choice of parameters $a = 1$, $b = 1$. More details about this can be reviewed in Bauer *et al.* (2014), Bauer *et al.* (2017), Bauer *et al.* (2018) and Kurtek *et al.* (2020). For this reason, some of the experiments were realized for the metric of Younes et al. too, in this case considering the elastic characteristics of the same.

The rest of the paper is organized as follows: section 2 presents theoretical details of the both approaches; section 3 presents the experimentation done in $\mathscr{S}_2$ for the three evaluated applications and comparison with the results obtained in Gual *et al.* (2015b); and section 4 contains a discussion of the results, the conclusions and proposals for future work. More results (Tables and Figures referenced with the letter S) are shown in Supplemental material.

MATERIALS AND METHODS

THE SPACE OF PLANE SHAPES $\mathscr{S}_1$: THE GRASSMANN MANIFOLD MODEL

Let M be the set of closed curves that are smooth boundaries of planar shapes, i.e.

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

where $\mathbb{S}^1$ denotes the unit circle and $\alpha'(t)$ stands for the parametric derivative of α. M is the space of C^∞-immersions $\alpha : [0, 2\pi] \longrightarrow \mathbb{R}^2$ with $\alpha(0) = \alpha(2\pi)$.

Let the tangent space $T_\alpha M$ at α be the set of vector fields h on α ($h : \mathbb{S}^1 \longrightarrow \mathbb{R}^2$). If $h, k \in T_\alpha M$, the Riemannian metric $G_\alpha(h,k)$ is defined as in Younes *et al.* (2008) by

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

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

The set M modulo translations, scalings, rotations, and reparameterizations of the curve ($Diff(\mathbb{S}^1)$) is the space of planar shapes. Nevertheless, let us study first the *pre-shape* space, i.e. before division by the group of diffeomorphisms $Diff(\mathbb{S}^1)$ is carried out. We refer to the group generated by translations, rotations and scalings as the group of similitudes (*sim*). Therefore, the *pre-shape* space M_d is defined by

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

and the restriction of the metric G_α is associated with M_d.

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

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

Let e, f be two functions in V and let us assume that our plane curves are curves in the complex plane $\mathbb{C}$. We define the basic mapping by

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

Let $Gr(2,V)$ be the Grassmannian of unoriented two-dimensional subspaces of V defined by an orthonormal pair $(e,f) \in V^2$ with $||e||^2 + ||f||^2 = 2$ (i.e. $l(\alpha) = 1$); then, Φ defines an isometry between M_d and a subset of $Gr(2,V)$.

Gual *et al.* (2015b) proposed detailed algorithms and pseudocodes to compute distances and geodesics in M_d from the preceding isometry.

Since our interest lies in geometric curves, i.e. we consider curves up to reparameterizations, then the shape space is defined by the following quotient space

$$\mathscr{S}_1 = \frac{M_d}{Diff(\mathbb{S}^1)}.$$

We will analyze digital contours that characterize cell boundaries with a fixed orientation and with approximately evenly spaced discrete points in these contours. The distance and the geodesic between two shapes in $\mathscr{S}_1$ are computed as in Gual *et al.* (2015b).

However, as deformations of curves respect the arc-length parameter, stretch elasticity is not incorporated to the model and resulting shape correspondences are sometimes far from optimal.

In this first proposal, only bending is considered for the evolution from one curve to another, so the metric is not used with all the advantages that it offers and the behavior is similar to that of non-elastic metrics. In the rest of the article the reference to the space $\mathscr{S}_1$ is made considering this particularity.

THE SPACE OF PLANE SHAPES $\mathscr{S}_2$: THE SRVF REPRESENTATION MODEL

In this section we review some results proposed by Srivastava *et al.* (2011). In particular, we consider the SRVF representation of closed curves in $\mathbb{R}^2$ and we summarize the main results for the shape space $\mathscr{S}_2$ with the standard elastic metric.

Let $\beta : [0,1] \longrightarrow \mathbb{R}^2$ be an absolutely continuous parameterized curve on $[0,1]$. The Square Root Velocity (SRV) of β is defined as the function $q : [0,1] \longrightarrow \mathbb{R}^n$ given by (Srivastava *et al.*, 2011):

$$q(t) = \frac{\beta'(t)}{\sqrt{|\beta'(t)|}}. \quad (2)$$

For every $q \in \mathbb{L}^2([0,1], \mathbb{R}^2)$ there is a curve β (unique up to translation) such that the given q is the SRV function of that β. In fact,

$$\beta(t) = \int_0^t q(s)|q(s)|\, ds. \quad (3)$$

To remove the scaling variability, we rescale all curves to be of unit length. If a curve β is of length one, then $\int_0^1 |q(t)|^2 dt = 1$.

Therefore, the SRV functions associated with these curves are elements of the unit hypersphere in the

Hilbert manifold $L^2([0,1],\mathbb{R}^2)$:

$$\mathscr{C}^0 = \{q : [0,1] \longrightarrow \mathbb{R}^2 \,/\, \int_0^1 |q(t)|^2 dt = 1\}, \quad (4)$$

is a Hilbert manifold.

To study the shapes of closed curves, we must impose the additional condition that the curve starts and ends at the same point, then the space of fixed length, closed curves represented by their SRV functions is

$$\mathscr{C}^c = \{q : \mathbb{S}^1 \longrightarrow \mathbb{R}^2 \,/\, \int_{\mathbb{S}^1} |q(t)|^2 dt = 1, \int_{\mathbb{S}^1} q(t)|q(t)|dt = 0\}. \quad (5)$$

To take care of the rotation and reparameterization of the curve β whose SRV function is q, we remember that a rotation is an element of $SO(2)$, the special orthogonal group of 2×2 matrices; and a re-parameterization is an element of Γ, the set of orientation-preserving diffeomorphisms of $\mathbb{S}^1$. The actions of $SO(2)$ and Γ on the SRV of β are given by:

$(O, q(t)) = q(t)$, where $O \in SO(2)$, and the SRV of the curve $\beta \circ \gamma$ is $q(\gamma(t))\sqrt{\gamma'(t)}$, where $\gamma \in \Gamma$.

The orbit of a function $q \in \mathscr{C}^c$ is

$$[q] = \{O(q,\gamma) = O(q(\gamma(t)))\sqrt{\gamma'(t)} \,/\, (\gamma, O) \in \Gamma \times SO(2)\}. \quad (6)$$

If we consider the metric in $\mathbb{L}^2$ given by the usual inner product

$$\langle v_1, v_2 \rangle_{L^2} = \int_0^1 \langle v_1(t), v_2(t) \rangle \, dt, \quad (7)$$

the feature space of interest is:

$$\mathscr{S}_2 = \{[q] \,/\, q \in \mathscr{C}^c\}, \quad (8)$$

and the distance in $\mathscr{S}_2$ is:

$$d_s([q_1],[q_2]) = \inf_{O \in SO(n), \gamma \in \Gamma} d_c(q_1, O(q_2, \gamma)), \quad (9)$$

where d_c denotes the distance in the hypersphere $\mathscr{C}^c$. The computation of the geodesics in the shape space $\mathscr{S}_2$ was detailed by Srivastava *et al.* (2011) and Joshi *et al.* (2007).

All the experiments were performed using an open source implementation of the general family of elastic metrics which include the SRVF, available from `https://github.com/h2metrics/h2metrics`. As a confirmation, the experiments were performed again using another SRVF framework toolbox available from `http://ssamg.stat.fsu.edu`, where equations (5.15), (5.28) and (5.29) from the book Srivastava *et al.* (2016) are implemented.

RESULTS

There are many interesting applications of the geometric representation of planar shapes in $\mathscr{S}_2$ proposed here. In this section, we show some of these applications and we use them to analyze digital images of blood smears in the morphological study of erythrocytes. The novelty lies in the use of the general theory in $\mathscr{S}_2$ (Sec. 2.2, Eqs. 2 and 9), which has never been used for these specific applications.

Experiments were performed using the erythrocytesIDB image database (Gonzalez-Hidalgo *et al.*, 2015) (available from `http://erythrocytesidb.uib.es/`), which consists of 202 images of normal cells, 210 images of sickle cells, and 211 images of cells with other deformations, obtained from the blood specimens of patients with sickle cell disease. Some specifications about image acquisition and characteristics are explained on the website. The other deformations class includes images with shapes that are close to circular or elongated forms, making the classification of this class a challenging task. Fig. 1 shows examples of images for all three classes (normal cells, sickle cells, and cells with other deformations). 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. The objects in this database were the data used for the experiments performed in Gual *et al.* (2015b). All the experimentation was done in MATLAB R2016a 64-bit.

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

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

INTERPOLATION BETWEEN SHAPES

Geodesic paths between shapes exposed before can be used to interpolate between shapes, which can be useful for estimating intermediate shapes of cells, whose shapes vary over time, for example those affected by sickle cell disease. Fig. 2 shows the interpolation between the same normal and the same sickle cell in $\mathscr{S}_1$ and $\mathscr{S}_2$, respectively, with the geodesic path between the two shapes. The representation of the curves is different in each space, hence the initial and final positions that are displayed are different, even when they correspond to the same cells. It can be noticed that in shape space $\mathscr{S}_2$ the deformation begins to be more pronounced than in $\mathscr{S}_1$ at a point closer to the first curve, which is because the metric is elastic and allows both stretching and bending, so the curve can deform more naturally than the case of the first attempt in $\mathscr{S}_1$, where the study only considers the bending of the curves. The distance between these two cells is 0.49 in the first study and 0.46 in $\mathscr{S}_2$, which confirms that the latter approach is more suitable for obtaining the geodesic path between those elements.

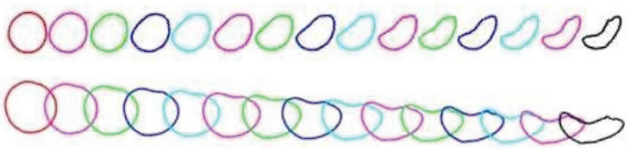

Fig. 2. *Geodesic between normal and sickle erythrocyte: in $\mathscr{S}_1$ with distance=0.49 (top); in $\mathscr{S}_2$ with distance=0.46 (bottom).*

SUPERVISED CLASSIFICATION

One of the most important issues in the study of this illness is the search for an efficient automatic classification method to quantify the number of deformed cells that a patient has and thus gauge the severity of the crisis. The distance between planar shapes was employed for the first time in Gual *et al.* (2015b) to obtain the supervised classification and unsupervised clustering of erythrocytes, achieving very good performance even when only bending was considered for the analysis of the curves. In this section, we begin by proposing the use of the distance between planar shapes (Eq. 9) in space $\mathscr{S}_2$ as the elastic metric in the supervised erythrocyte morphology classification process as normal, sickle, or those with other deformations, and given the excellent results obtained. The next section deals with unsupervised clustering of erythrocytes.

To carry out the task of supervised classification, we propose again two possibilities: using the distance in Eq. 9 to perform the normal supervised classification process and using it to perform a classification considering the cells as deformations of known templates, in this case circles and ellipses. The metrics used to evaluate the results were sensitivity or True Positive Rate (TPR), precision (P), specificity or True Negative Rate (TNR), and F1-score (F1). The overall sensitivity was given as the mean sensitivity of the classes. The measures F1-macro (F1-M) and Accuracy (Acc) were used to evaluate the whole process.

Like in the first approach, all the distances between pairs of cells were calculated and the k-nearest neighbor algorithm for supervised classification (k-NN) was used. It is not the purpose of the work to determine the most efficient classification algorithm in this case; k-NN tests are only carried out for the sake of simplicity, since our purpose is to demonstrate the feasibility of using the framework in this environment. In order to validate the results, a 5×1 cross-validation process was carried out. The confusion matrix of the process and the sensitivity, precision, specificity and F1-score measures for each class are shown in Table 1.

Table 1. *Supervised classification results in both cases. Classes N: Normal, S: Sickle, O: Other Deformations.*

$\mathscr{S}_2$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	199	0	1	1.00	0.95	0.91	0.95
S	0	193	7	0.96	0.98	0.96	0.96
O	19	7	174	0.87	0.98	0.96	0.91
$\mathscr{S}_1$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	202	0	0	1.00	0.94	0.89	0.94
S	0	202	8	0.96	0.98	0.96	0.96
O	25	8	178	0.84	0.98	0.96	0.90

In Table 1 we see that the sensitivity values obtained in $\mathscr{S}_2$ for classes normal and sickle are high, with 100% and 96% for each, respectively. No normal cell was classified as sickle or vice versa. In the case of other deformations the sensitivity decreased to 87% due to several cells with other deformations with shapes that were very close to circular and elongated, and in addition these elements differ a lot from each other.The precision values are greater than 91% for all cases, and the specificity values are greater than 95% for the three classes analyzed. The F1-score measure achieves values above 91% in all the cases and the process has an accuracy of 94.3%.

In order to assess the quality of these results, they were compared with those obtained in Gual *et al.* (2015b). This previous work were performed under the same experimental conditions and using the whole data

set in the image database, which has 202 normal cells, 210 sickle cells and 211 cells with other deformations, and the results are shown in Table 1. With the new approach, the sensitivities of normal and sickle cells were the same that in this previous work, but the sensitivity of the class of cells with other deformations and the precision of the normal class (87% and 91%) were greater than the 84% and 89% achieved in the previous work, respectively, because several objects of this class were correctly classified in $\mathscr{S}_2$. In Fig. S1 two cases of cells with other deformations that are misclassified under the previous method but are correctly classified under the proposed method are shown, the first column is the object to classify, second and third columns show the class and the distance of the object closest to the first one. Accuracy in this previous work was 93.4%, lower than 94.3% of the new approach.

The second possibility for experimentation proposes the use of the $\mathscr{S}_2$ space to make a classification considering the cells as deformations of known templates. In this particular case, normal cells have an almost circular shape and sickle cells have an almost elliptical shape and we could assume that elliptical forms appear in their geodesic path from the circle. Therefore, we can assess the possibility of considering how close the studied cells are to circle and ellipse templates in this framework, which drastically reduces the calculations needed to obtain the classification. The axes of the ellipse used as a template are obtained as the mean value of the axes of the ellipses that best fit each of the sickle cells in the sample.

The linear discriminant analysis algorithm for classification with leave-one-out cross-validation was used. The confusion matrix and measures obtained of the process in $\mathscr{S}_2$ is shown in Table 2. The results show that the sensitivity of normal and elongated classes is also high, 98% in both cases, and once again no normal cell was classified as elongated or vice versa. In the case of other deformations the sensitivity decreased to 86%, slightly lower than the 87% obtained in the supervised classification. The precision and specificity values remain as in the case of supervised classification, with more than 92% and 96% for the three classes analyzed. The F1-score measure achieves values above 91% in all cases and the process has a high accuracy of 94.2%, practically the same as the 94.3% of the supervised classification process. This is a very interesting result, because the computational cost of the process is drastically reduced with this method, maintaining the high accuracy of the process.

Table 2. *Results of classification using templates in both cases. Classes N: Normal, S: Sickle, O: Other Deformations.*

$\mathscr{S}_2$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	197	0	3	0.98	0.96	0.92	0.95
S	0	195	5	0.98	0.97	0.95	0.96
O	16	11	173	0.86	0.98	0.96	0.91
$\mathscr{S}_1$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	198	0	4	0.98	0.94	0.89	0.93
S	0	207	3	0.99	0.96	0.92	0.95
O	24	17	170	0.81	0.98	0.96	0.88

For the previous case, this experiment was performed by Gual *et al.* (2015b), under the same conditions. Table 2 shows the confusion matrix and measures of the process. In this case, the sensitivity of normal and elongated classes is also high: 98% and 99%, respectively, and once again no normal cell was classified as elongated or vice versa. In the case of other deformations the sensitivity decreased to 81%, far below the 86% obtained in the $\mathscr{S}_2$ space. The precision and specificity values for the normal and sickle classes were lower, due to the fact that more objects in the class of other deformed cells are classified as normal or sickle. The specificity values remain over 94% for all the classes, although they are below those reached in the $\mathscr{S}_2$ space, and the F1-score measure achieves lower values too, with an emphasis on cells with other deformations class, which has 88%, far below of 91% in the $\mathscr{S}_2$ space. The accuracy in this case was 92.3%, much lower than 94.2% reached with the new approach.

Both experiments, supervised classification and classification using templates, were performed using the metric of Younes et al. considering the elastic characteristics of it. The results are shown in Table S1 and are overcome those obtained previously by Gual *et al.* (2015b), which is the expected behavior due to the elastic shape analysis of curves, which has proven to outperform inelastic shape analysis, but they do not exceed those achieved with the SRVF representation of the curve. For supervised classification, accuracy was 93.8%, greater than the 93.4% of the previous work, but lower than the 94.3% achieved with SRVF, and the accuracy achieved by classification using templates was 93.7%, exceeding the results obtained by Gual *et al.* (2015b), with an accuracy of 92.3%, but not exceeding those achieved with the SRVF representation of the curve, with 94.2%.

Table S2 shows values of general performance measures for all the classification experiments performed as a whole. The accuracy and F1-Macro (94% in both cases) achieved in the $\mathscr{S}_2$ space are much greater than those in the other case (93% for supervised classification and 92% for classification using templates), and even in the case of classification using templates the performance of the process was very high.

These results allow us to conclude that the supervised classification of erythrocytes in this case is more efficient, and this is due to the possibilities offered by the $\mathscr{S}_2$ space for optimally matching, deforming, and comparing shapes, considering deformations of curves resulting not just from bending but from stretching as well, which makes it possible to obtain more representative distances of the intra-class and inter-class relationships and therefore significantly improves the results of the classification.

The experiments with the other toolbox (available from `http://ssamg.stat.fsu.edu`) also yielded excellent results in the sensitivity values: 99.5% for normal class, 97.5% for elongated class, and 88.5% for the other deformations class. The accuracy in this case was 95.1%. The results with both tools are excellent; they only differ slightly due to the alignment procedures of the shapes.

UNSUPERVISED CLUSTERING

For the study of sickle cell disease, the class of interest is sickle cells. All analyses that can be performed on their presence can be used by a specialist to issue a criterion related to the severity of a patient's crisis. For this reason, the cell differentiation into normal, sickle, or other abnormal cells that has been studied by some authors (Wheeless *et al.*, 1994; Fernández *et al.*, 2013; Gonzalez-Hidalgo *et al.*, 2015; Delgado *et al.*, 2016; Gual *et al.*, 2015a) is illustrative for this pathology. However, the analysis of other types of erythrocyte deformations present in blood samples can be of interest, because there are other diseases that can lead to this situation. Some authors have carried out research related to the morphological classification of erythrocytes in several classes (Dalvi and Vernekar , 2016; Aziz , 2013; Durant *et al.*, 2017; Frejlichowski, 2010), where homogeneous classes of deformations characterized by the morphology of the cells are defined. The use of distance between planar shapes as a dissimilarity measure in an unsupervised clustering algorithm for the $\mathscr{S}_1$ space to define homogeneous classes of deformations was first proposed by Gual *et al.* (2015b).

In this section, we present the use of this distance in an unsupervised clustering algorithm for the $\mathscr{S}_2$ space. These experiments were performed only for the $\mathscr{S}_2$ space due to the superiority of the results achieved in the supervised classification. As a 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 measurements between objects. The PAM algorithm (like k-means) is based on finding k representative objects (also known as medoids) 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), which would be a very complex task in the shape space.

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 the result with the largest average silhouette width.

It is proposed that the clustering should be performed for $k \geq 3$ because our problem requires the differentiation of at least three classes: normal, sickle, and other deformations. The results obtained in the supervised classification for $k = 3$ groups, which are shown in Table 1, will be used to validate the performance of the clustering obtained for this k. For more groups, the interest is to know how this framework can group objects that move away from the normal or sickle shape in the classes of greatest interest for the study of sickle cell disease.

The confusion matrix and measures for clustering procedure considering $k = 3$ are shown in Table 3. The sensitivity values obtained for the normal and sickle classes remain high, with 91% and 96% for each, respectively. In the other deformation class, the sensitivity increased to 97%. The values of precision are greater than 93% for all cases, and the specificity values are greater than 97% for the three classes analyzed. The F1-score measure achieves values above 92% in all the cases, and accuracy was 94.7%, similar to the supervised classification case. However, in this case there are normal elements that are classified as sickle and vice versa, even when the accuracy remains high.

Table 3. *Results of unsupervised clustering in both cases for $k = 3$. Classes N: Normal, S: Sickle, O: Other Deformations*

$\mathscr{S}_2$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	181	13	6	0.91	0.97	0.93	0.92
S	7	193	0	0.96	0.97	0.94	0.95
O	6	0	194	0.97	0.98	0.97	0.97
$\mathscr{S}_1$	Confusion Matrix			Measures			
	N	S	O	TPR	TNR	P	F1
N	200	0	4	0.98	0.91	0.84	0.91
S	0	200	10	0.95	0.94	0.90	0.92
O	37	23	151	0.72	0.97	0.91	0.80

These results were very superior to the results in the other case (Gual *et al.*, 2015b) shown in Table 3, where the sensitivity of the normal and sickle classes was very high (98% and 95% respectively), but the class of other deformations achieved only 72% of sensitivity, and the accuracy was 88%, far below 94.7% in $\mathscr{S}_2$. The average silhouette width for $k = 3$ in $\mathscr{S}_2$ space was 0.58.

Experiments for $k > 3$ were performed. Table 4 shows the results for $k = 4$ and $k = 5$, with average silhouette width (*ASW*) values for each one. For higher values of k, the average silhouette width was less than 0.56 (0.52, 0.51, 0.43 for $k = 6, 7, 8$ respectively). For the normal class the distribution remains stable, the fact that no normal cell was classified as sickle and vice versa persists. Fig. S2 shows the medoids of each class for $k = 5$.

Table 4. *Unsupervised clustering in $\mathscr{S}_2$ space: $k \geq 3$ groups generated.*

	N	S	*G1*	*G2*	*G3*
	$k = 4$ (ASW = 0.49)				
N	194	0	6	0	-
S	0	103	1	96	-
O	6	19	171	13	-
	$k = 5$ (ASW = 0.56)				
N	194	0	6	0	0
S	0	102	1	90	7
O	5	3	102	14	46

Even though the database used does not contain differentiated elements for other erythrocyte deformations, the experiment shows that it is possible to obtain clusters with good results using the $\mathscr{S}_2$ space, maintaining quality in the main classes of cells for the study of the pathology: normal and sickle cells. Nevertheless, we should stress that this is a preliminary experiment. To determine with certainty the effectiveness of using the measurement, more experiments need to be carried out with a greater amount of cell deformations, the images used are all of blood samples of sickle cells, where the principal deformation found is the characteristic of the pathology.

DISCUSSION

For an in-depth analysis of the results, considering that the method has demonstrated its efficiency in the morphological classification of erythrocytes, it is necessary to assess the feasibility of using it for possible tools to support the clinical diagnosis of patients with sickle cell disease. In the study of this disease, sickle cells are the most important class of cell, mainly due to two issues: the first is that this type of cell is the one that can cause the events that trigger vaso-occlusive crises; the second is that in a real study environment of a patient, there will always be a much smaller amount of sickle cells than other cells in the sample, so an algorithm that is not capable of performing well in this class is not recommended. The method must achieve good performance in the sensitivity and precision of the classification of normal and sickle cells because they are the most important classes, but the most complex class to process is that of other deformations, because it will be composed of elements that can have shapes close to the circular or elliptical, which therefore affects the accuracy of classification in the other two classes.

Table 5 summarizes the sensitivity and precision values achieved for each class separately, for all the experiments carried out. For the normal class, all the values of the measures obtained in $\mathscr{S}_2$ exceed those obtained in $\mathscr{S}_1$, except for the sensitivity in clustering, where 91% is obtained, lower than the 98% of $\mathscr{S}_1$. In all cases, the precision in $\mathscr{S}_2$ is greater than in $\mathscr{S}_1$. For sickle class, all the values of the measures obtained in $\mathscr{S}_2$ exceed those obtained in $\mathscr{S}_1$, except for the sensitivity in the classification from templates for sickle cells, where 98% is obtained, lower than the 99% of $\mathscr{S}_1$, although it remains very high. In all cases, the precision in $\mathscr{S}_2$ is much greater than in $\mathscr{S}_1$. Regarding the class of other deformations, the results of the sensitivities obtained in the three experiments in space $\mathscr{S}_2$ far exceed those of space $\mathscr{S}_1$, with more than 86% in all, when in $\mathscr{S}_1$ the highest sensitivity was 84%, reached in the supervised classification. Values of precision are very high in $\mathscr{S}_2$ for this class. Therefore, the proposed method achieved high sensitivity and precision in the classification of normal

and sickle erythrocytes, and in the other deformations class, it considerably improved on the previous results, reaching sensitivity values of 87% or close to this.

Based on this performance, it can be considered that the proposal is valid to be used in automated tools to support the clinical diagnosis of patients with sickle cell disease and demonstrates that this $\mathscr{S}_2$ metric is more suitable for classification in the shape space, considering both bending and stretching to obtain the smallest geodesic distance. All the experiments showed accuracy of the classification processes with greater values in $\mathscr{S}_2$ than the previous case.

Table 5. *Performance of all the experiments in both cases for each class: TPR True positive rate; P Precision; (SC) Supervised Classification; (CT) Classification using templates; (C) Clustering.*

	SC		CT		C	
	$\mathscr{S}_1$	$\mathscr{S}_2$	$\mathscr{S}_1$	$\mathscr{S}_2$	$\mathscr{S}_1$	$\mathscr{S}_2$
	Normal Class					
TPR	1.00	1.00	0.98	0.98	0.98	0.91
P	0.89	0.91	0.89	0.92	0.84	0.93
	Sickle Class					
TPR	0.96	0.98	0.99	0.98	0.95	0.96
P	0.96	0.97	0.92	0.95	0.90	0.94
	Other Deformations Class					
TPR	0.84	0.87	0.81	0.86	0.72	0.97
P	0.96	0.91	0.96	0.96	0.91	0.97

Performing the analysis of erythrocyte shapes in a different space, such as the shape space, has advantages compared to the feature space used in the proposals put forward so far. The main advantage is that it is not necessary to obtain characteristics about the shape, only its functional representation using the SRVF, which establishes an analysis framework that allows more sophisticated studies on cell deformation, not just classification according to the shape, something that can be useful, for example, in cases where it is necessary to follow the development of the deformation process itself over time. Another advantage is that the distances between the pairs of shapes are almost explicitly available.

As a comparison with previously proposed methods that are based on the use of shape characteristics, we present the results obtained by two recent methods proposed in Gual *et al.* (2013) for morphological analysis of erythrocytes, which reached high performance values. These methods propose the use of integral geometry functions $W(\phi)$ and $C_\rho(\phi)$ to describe the contour exhaustively, and they obtain excellent results in the supervised classification of erythrocytes in the three evaluated classes. Table S3 shows the confusion matrix and the measurements obtained using these functions. It can be seen that they achieve higher sensitivity in the class of other deformations; more than 93% for both descriptors with respect to the 87% obtained in $\mathscr{S}_2$. In the two most interesting classes, normal and sickle cells, supervised classification in $\mathscr{S}_2$ remains with high sensitivities. Table 6 shows the general performance measures for the two descriptor functions and the three experiments performed with the method proposed in this paper for $\mathscr{S}_2$: supervised classification, denominated $\mathscr{S}_2$-*SC*; classification using templates, denominated $\mathscr{S}_2$-*T*; and unsupervised clustering, denominated $\mathscr{S}_2$-*C*. It can be seen that the measure values in $\mathscr{S}_2$ are very high, more than 94%, and remember that in the case of $\mathscr{S}_2$-*T* it is possible to drastically reduce the amount of calculations to be made since only the distances to the two templates are needed, and for $\mathscr{S}_2$-*C* it is verified that it is possible to obtain quality clustering using this metric under the conditions of the studied data. In Gual *et al.* (2015b) the superior performance achieved by the classification in the shape space with respect to other methods is shown. Those methods are not exposed in this work given that the superiority of the performance of the classification in the space $\mathscr{S}_2$ with respect to the previous case has been proven.

Table 6. *Performance measures of supervised classification in both feature and shape spaces: $W(\phi)$ weighted generalized support function, $C_\rho(\phi)$ Crofton descriptor, $\mathscr{S}_2$-SC supervised classification, $\mathscr{S}_2$-T classification using templates, $\mathscr{S}_2$-C unsupervised clustering.*

	$W(\phi)$	$C_\rho(\phi)$	$\mathscr{S}_2$		
			$\mathscr{S}_2$-*SC*	$\mathscr{S}_2$-*T*	$\mathscr{S}_2$-*C*
F1-M	0.96	0.96	0.94	0.94	0.95
Acc	0.96	0.96	0.94	0.94	0.95

These results confirm that the elastic metric in the SRVF framework proposed by Srivastava *et al.* (2011) can be used to perform the morphological analysis of erythrocytes in sickle cell disease sample images, representing the boundaries of the erythrocytes as closed planar curves in this space and obtaining distances and geodesics between the curves for their classification. The experimentation carried out showed the superiority of the proposal in supervised classification of erythrocytes in normal, sickle, or cells with other deformations, compared to other proposals in the shape space. The performance of the classification procedure was also studied considering the distance to circle and ellipse templates, taking into account the similarity that the objects of interest have to them, and it has been concluded that this variant

can be used to reduce the computational cost of the classification process. The behavior of the metric in the process of unsupervised clustering was also studied, which achieved high accuracy in the experiment generating the three classes and maintained stability in the normal and sickle cell classes, while three new groups of other deformations were determined. Finally, the framework can be used to obtain the interpolation between curves within a geodesic path that is more representative of the process.

These excellent results are due to the characteristics of the proposed elastic metric, which allows us to obtain deformations of curves resulting not just from bending but from stretching as well. With these results it can be affirmed that the method is feasible for use in erythrocyte morphology study applications and can be used to support clinical diagnosis of the state of a patient with sickle cell disease. The methodologies introduced in this paper could be extended to other similar clinical applications. As future work, experiments must be done for images obtained from the complete visual field of view under the microscope, where the imbalance of classes is considerable, so the performance of the classification must be evaluated with appropriate measures for unbalanced classes.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and two anonymous reviewers for their very constructive suggestions, which have led to improvements in the manuscript. This work was supported by: projects PCoopUJI-18-01 and UJIB2017-13 from Universitat Jaume I, Spain, by Grant DPI2017-87333-R from the Spanish Ministerio de Ciencia, Innovación y Universidades (AEI/FEDER, UE); project 10150 within the National Basic Science Program of Universidad de Oriente, Cuba; the Belgian Development Cooperation through VLIR-UOS (Flemish Interuniversity Council-University Cooperation for Development) in the context of the Institutional University Cooperation programme with Universidad de Oriente.

REFERENCES

Aziz Toma S (2013). Using aspect ratio to classify red blood images. IOSR J Comput Eng 13:30-33.

Bauer M, Bruveris M, Marsland S, Michor PW (2014). Constructing reparametrization invariant metrics on spaces of plane curves. Differ Geom Appl, 34C:139-65.

Bauer M, Bruveris M, Harms P, Møller-Andersen J (2017). A numerical framework for Sobolev metrics on the space of curves. SIAM J. Imaging Sci. 10 (1):47–73.

Bauer M, Bruveris M, Charon N, Møller-Andersen J (2019). A relaxed approach for curve matching with elastic metrics. ESAIM:COCV. 25 (72):1-24.

Dalvi PT, Vernekar N (2016). Computer aided detection of abnormal red blood cells. In Proceedings 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore:1741-46.

Delgado-Font W, González-Hidalgo M, Herold-Garcia S, Jaume-i-Capó A, Mir A (2016). Erythrocytes morphological classification through HMM for sickle cell detection. In International Conference on Articulated Motion and Deformable Objects (AMDO), LNCS: 9756:88-97.

Durant TJS, Olson EM, Schulz WL, Torres R (2017). Very deep convolutional neural networks for morphologic classification of erythrocytes. Clin Chem 63:1847-55.

Fernández K, Herold-Garcia S, Fernández A, Escobedo M, Coello G, Marrero P (2013). Estudio Morfológico en Muestras de Sangre Periférica. In V Latin American Congress on Biomedical Engineering CLAIB 2011, Cuba. IFMBE Proceedings 33:543-46. Reading: Springer, Berlin, Heidelberg.

Frejlichowski D (2010). Pre-processing, extraction and recognition of binary erythrocyte shapes for computer-assisted diagnosis based on MGG images. In International Conference on Computer Vision and Graphics, Poland 2010, Part I. LNCS 6374:368–75.

Gonzalez-Hidalgo M, Guerrero-Pena FA, Herold-Garcia S, Jaume-i Cap A, Marrero-Fernandez P (2015). Red blood cell cluster separation from digital images for use in sickle cell disease. IEEE J Biomed Health Inform 19:1514-25.

Gual-Arnau X, Herold-Garcia S, Simó A (2013). Shape description from generalized support functions. Pattern Recogn Lett 34:619-26.

Gual-Arnau X, Herold-Garcia S, Simó A (2015a). Erythrocyte shape classification using integral-geometry-based methods. Med Biol Eng Comput 53:623-33.

Gual-Arnau X, Herold-Garcia S, Simó A (2015b). Geometric analysis of planar shapes with applications to cell deformations. Image Anal Stereol 34(3):171-82.

Joshi SH, Srivastava A, Klassen E, Jermyn I (2007). A Novel Representation for Computing Geodesics Between n-Dimensional Elastic Curves. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis.

Kurtek S, Needham T (2020). Simplifying transforms for general elastic metrics on the space of plane curves. SIAM J Imaging Sci (to appear). Arxiv:1803.10894.

Laga H, Kurtek S, Srivastava A, Golzarian M, Miklavcic SJ (2012). A Riemannian Elastic Metric for Shape-Based Plant Leaf Classification. In 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA:1-7.

Mazalan SM, Mahmood NH, Razak MAA (2013). Automated red blood cells counting in peripheral blood smear image using circular Hough transform. In 2013 IEEE 1st International Conference on Artificial Intelligence, Modelling and Simulation:320-24.

Mio W, Srivastava A, Joshi S (2007). On shape of plane elastic curves. Int J Comput Vision. 73(3):307-24.

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

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

Rahmat RF, Wulandari FS, Faza S, Muchtar MA, Siregar I (2018). The morphological classification of normal and abnormal red blood cell using self organizing map. IOP Conf Series: Mater Sci Eng 308:012015.

Sarrafzadeh O, Dehnavi AM. Rabbani H, Ghane N, Talebi A (2015). Circlet based framework for red blood cells segmentation and counting. In 2015 IEEE Workshop Signal Processing Systems:16. Reading: IEEE.

Srivastava A, Klassen E, Joshi SH, Jermyn IH (2011). Shape analysis of elastic curves in Euclidean spaces. IEEE Trans. on Pattern Analysis and Machine Intelligence. 33:1415–28.

Srivastava A, Klassen E (2016). Functional and Shape Data Analysis. Reading: Springer-Verlag, New York.

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.

Younes L, Michor PW, Shah J, Mumford D (2008). A Metric on Shape Space with Explicit Geodesics. Rend Lincei Mat Appl 39:25–57.