Shape Analysis with Trajectory Networks

Luciano da Fontoura Costa
Institute of Physics at São Carlos, University of São Paulo,
P.O. Box 369, São Carlos, São Paulo, 13560-970 Brazil
(Dated: 17th March 2008)

Image and shape analysis are amongst the most challenging abilities to be replicated artificially. One of the first important steps along these two tasks consists in obtaining comprehensive representations of the involved objects, capable not only of representing most of the original information, but also of emphasizing their less redundant portions. The current work reports an approach to shape characterization and classification which is based on trajectory networks, a special type of knitted geographical networks where the connections take into account not only the proximity between nodes, but also an associated vector field, here assumed to correspond to the electric field induced by the contours of the shapes. In this way, the original shape is mapped into a trajectory network, so that its measurements can reveal important features of the shapes under analysis. Optimal multivariate stochastic methods (namely discriminant analysis) are then applied in order to identify the topological measurements contributing most effectively for the separation between the objects to be analyzed and classified. It is shown that the weveral topological and geometrical measurements contribute differently to the separation between the considered set of shapes. The entropy of the angles defined by the edges, the number of nodes with degree 1, 4 and 5, as well as an alternative type of entropy, are found to contribute more strongly to the discrimination between the considered shapes.

PACS numbers: 89.75.Hc, 89.75.Fb, 89.75.-k

'All the rest of the city is invisible. Phyllis is a space in which routes are drawn between points suspended in the void.' (I. Calvino, Inivisible Cities

I. INTRODUCTION

Image and shape analysis remain two of the most challenging human abilities to be reproduced artificially (e.g. [1, 2]). While a substantial part of the involved difficulties are related to the need to integrate previous information and knowledge about the objects under analysis, another critical step corresponds to obtaining representations of the objects and shapes which not only encompass the majority of the original information, but also emphasizes their most important, less redundant, parts. Such a critical interplay between representation and discrimination/classification can be better appreciated from the diagram in Figure 1, which shows the basic steps normally involved in image and shape analysis. We have from this diagram that the original shape/object is first transformed into a feature vector, incorporating measurements about its geometry, which is then fed to the classification methods. Therefore, a poor choice of measurements undermines the classification stage. At the same time, the choice of features should take into account the results of the specifically adopted classification methods.

An important aspect which is often overlooked while solving image and shape analysis and classification is that the measurements and classification approach are highly dependent on the specific set of images/objects to be analyzed. In order to illustrate this important principle, consider the toy data in Figure 2. Here we have 8 shapes which are clearly recognizable by humans. Higher chances of proper identification of a given shape as one of the 8 forms in Figure 2 are achieved in case the representation of the shapes takes into account their most discriminative features. For instance, a potentially good feature to differentiate a triangle from a square is their respective number of vertices, which can be identified from curvature peaks (e.g. [2–5]). Good features typically remove the redundancy in the original object, leaving out only the most important geometrical variations (e.g. there is little profit in considering all points between the extremities of a straight line segment). It is also often expected that a good set of features be invariant to specific geometrical transformations such as rotation, translation and scaling. The unmatched performance of the human visual system is to a great extent accounted by the incorporation of such principles.

The current work presents a comprehensive approach to shape representation, analysis and classification based

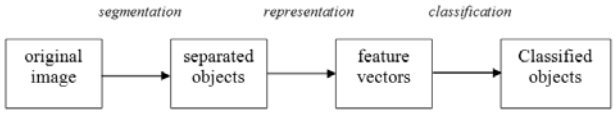


FIG. 1: The main steps along shape analysis involve the capture of the image, the separation of its constituent objects (segmentation), followed by the extraction of a meaningful representation in terms of a set of features as well as the respective identification of the category of each shape.

on complex networks concepts and methods, as well as optimal multivariate stochastic tools, namely discriminant analysis [6–8]. In particular, we suggest that the original objects and shapes be represented in terms of trajectory networks [9], so that the geometric features of the shapes under analysis can be quantified in terms of measurements of the topology and geometry of the respective trajectory networks. Then, by applying discriminant analysis we identify which of the topological measurements contribute the most for the separation between the objects to be classified as quantified by an objective optimality criterion involving the inter and intra scattering matrices [2, 6, 7].

Complex networks research (e.g. [10–14]), i.e. the study of graphs with specially intricate connectivity, are currently among the most dynamic and important scientific and technological areas because of the generality of networks for representing, characterizing, analyzing and modeling virtually any natural or human-made system composed (or decomposable) of discrete parts. Images are no exception, as they exhibit discrete organization at all scales of observation. For instance, at the more microscopic level, images are ultimately composed by picture elements, or *pixels*. At larger scales, images can be understood as being composed of several objects. At the same time as these components have well defined positions and orientations, the relationships between them are as much important as far as their identification is concerned. Complex networks can not only represent each object as a respective *node*, but also make explicit the interrelations between the objects by representing them as *edges*. In particular, networks whose nodes have well-defined positions are said to be *geographical*. Another important supercategory of complex networks, recently introduced [15, 16], corresponds to the *knitted networks*, characterized by the fact that their connections are given by paths along the nodes.

A growing number of graph and network-based approaches to image and shape analysis has been reported in the literature (e.g. [17–25]). Several approaches are based in considering pixels or points distributed along the images as corresponding nodes, while the connections are established as a consequence of gray-level similarity (e.g. [18, 19, 22, 23, 26]) or proximity/connectivity (e.g. [24, 25]). At a higher spatial scale, nodes are assigned to parts of the images or objects, while the connections are implemented by taking into account proximity and adjacent between the parts.

The representation of the visual information suggested in this article involves mapping shapes into respective trajectory networks. First, it is necessary to assign a vector field to each shape, which can be done by estimating the respective electric induced by its contour (e.g. [27]). A set of nodes are uniformly distributed (randomly) inside the object or shape of interest, and trajectories are initiated [9] from randomly chosen nodes. At each time step as the trajectories unfold, points which are no further than a maximum distance are chained into the growing path. Figure 3 illustrates trajectory networks obtained from the shapes in Figure 2 by using such a methodology. Though remindful of diffusion-limited aggregation (DLA) [28, 29], the obtained trajectory networks exhibit more aligned branches, a consequence of the locally parallel lines of force.

Once the trajectory networks are obtained for the shapes under analysis, several measurements can be extracted from them. We consider two main types of features: (i) *topological* and (ii) *geometrical*. The former category of measurements depends only on the interconnectivity between the nodes, therefore including the node degree, clustering coefficient, assortativity, amongst many other possibilities [14]. The geometrical measurements are obtained by considering the geometrical properties of the network, such as the angles of its edges and lengths of edges and involved shortest paths.

Because of the virtually infinite number of topological and geographical measurements of networks which can be considered for the respective characterization and classification, it is important to take into account sound methods which allow the choice of the most effective features. In this work we resource to scatterplots defined by pairs of measurements as well as to discriminant analysis, which ensures optimal separation of the shapes with respect to an objective criterion taking into account the inter and intra-class dispersion between the shapes.

This article starts by reviewing the related basic concepts and follows by presenting the results obtained for the characterization and classification of shapes. Several findings are reported, including the identification of the features, among those considered in this work, which most contribute to the separation between the shapes.

II. BASIC CONCEPTS

A. Complex Networks Representation and Topological Measurements

An undirected, unweighted complex network, namely a graph with specially intricate structure, can be completely represented in terms of its respective *adjacency matrix* K. The presence of an edge extending from node i to node j implies $K(j,i) = K(i,j) = 1$, while $K(j,i) = K(i,j) = 0$ expresses the absence of that connection. The *immediate neighbors* of a node i correspond to those nodes which receive an edge from i. The *degree* of node i can be defined as being equal to the number of its immediate neighbors. Two nodes are said to be *adjacent* whether they share an edge; two edges are adjacent if they share a node. A sequence of adjacent edges defines a *walk*. A *path* is a walk which never repeats either edges or nodes. The *length* of a walk or path is equal to the number of its constituent edges.

In this work we consider the number of nodes with degrees 1 to 5, henceforth abbreviated as $k1, k2, k3, k4$ and $k5$, as topological measurements of the trajectory

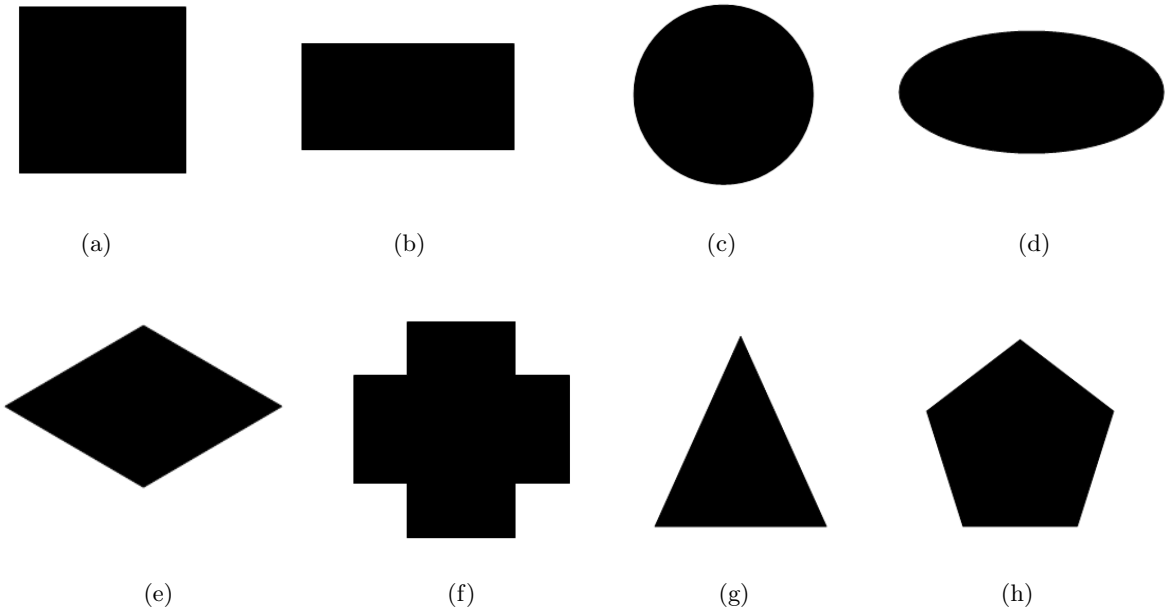


FIG. 2: The 8 shapes taken into account for the characterization and classification experiments reported in this article.

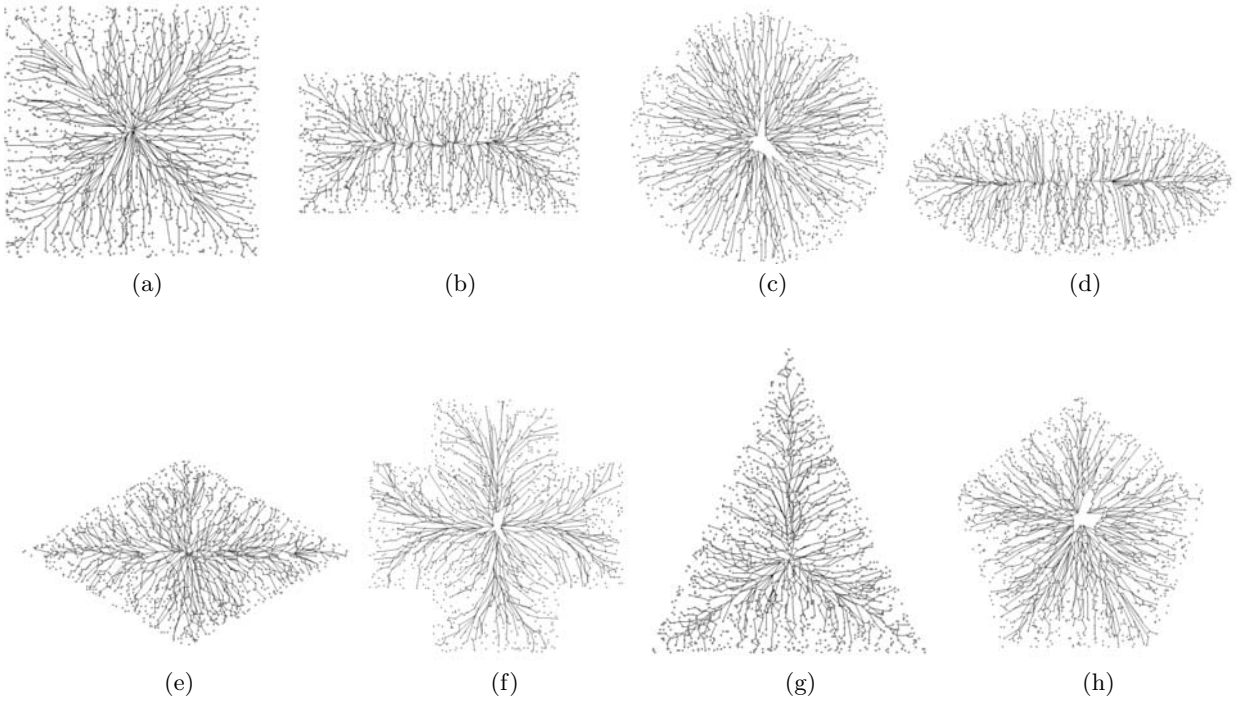
networks. Observe that $k1$ correspond to the number of extremity nodes of the trajectory networks, while the nodes with higher degree tend to be located more centrally within the shapes. These measurements can be immediately obtained at little cost from the respective adjacency matrices.

B. Trajectory Networks and Respective Geometrical Measurements

Trajectory networks are a special type of geographical (e.g. [30]) and knitted networks [16] characterized by the fact that each of their nodes possesses a well-defined spatial position while the edges follow an imposed vector field. Trajectory networks are related to gradient networks (e.g. [26, 31–33]) – where the connectivity is established by taking into account differences between fitness values assigned to nodes, field interactions [21, 34] – where the connections between nodes in geographical networks are defined in terms of imposed fields, and dynamical systems (e.g. [35, 36]) – in which attractors and specific regions of the phase space are represented as nodes while the trajectories between them are represented by the edges.

In the present work, we focus our attention on trajectory networks defined by respective shapes. Given a *shape*, henceforth understood as any set of connected points [2], it can be represented in a digital binary image by some sampling scheme, so that groups of the original points in the shape are represented by respective pixels set to 1, while the background is represented by pixels with value 0. The so-obtained *digital shapes* can also be understood as sets of connected pixels. Given a digital shape, its contour or border can be immediately extracted by marking every pixels which belongs to the shape (i.e. has value 1) and has at least one of its 4-neighbors equal to 0. Interestingly, the border of a shape captures *all* the information about its geometry, in the sense that the original shape can be perfectly reconstructed by filling with ones the interior of the contour.

Given a digital shape and its border, the method adopted in the present work to generate respective trajectory networks is as follows. In order to define the respective vector field, each of the pixels along the digital shape contour are understood as an electrical charge, so that a field of force is immediately established around the border. A particularly efficient way to calculate the vector field along the digital image lattice is to use the Fourier transform to perform the two convolutions required to calculate the $x-$ and $y-$components of the vector field [27]. Though in the present work we focus our attention on the field in the interior of the shapes, interesting results can also be obtained by considering trajectory networks defined with respect to the electrical field outside the shape. We assume repulsive force between the border of the shape and probe charges at each of the interior points, and the so-obtained vector field is normalized so that every vector has unit magnitude (this ensures constant speed magnitude during the trajectory calculations). Then, a set of N points/nodes are uniformly distributed (randomly) within the shape,



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

(e) (f) (g) (h)

FIG. 3: Examples of trajectory networks obtained for each or the shapes in Fig. 2. A total of 1500 poits were distributed uniformly and $N_p = 400$ paths were obtained by following the trajectories defined by the shape borders after starting from a randomly chones node. The maximum distance for attaching new nodes to the tip of a growing path was set as $D_p = 2$.

i.e. the region of the digital image defined by pixels with value 1. We henceforth fix the value of N in order to obtain invariance of the measurements to scaling: recall that trajectory networks obtained from circles (or other shapes) with different sizes by using the same number of spatially distributed points will tend to be invariant to the scale, represented by the size of the circles. A subset of N_p of nodes are then selected as the origins of the trajectories/paths. For each of these nodes, a trajectory is obtained by following the orientation of the respective vector field. At each step, the node which is closest to the current extremity of the growing trajectory, provided it is at a distance not exceeding D_p from the extremity, is linked into the respectively associated path. As a consequence of such a network growth scheme, the obtained trajectory networks present a predominant branching structure, with paths extending from the periphery of the network (i.e. close to the shape border) into its more central portions, which are associated to the shape skeleton (e.g. [2]). So, to a great extent the obtained trajectory networks present a tree-like structure, though several cycles are also obtained as a consequence of connections between adjacent trajectories.

Because geographical networks, such as the trajectory networks, are characterized by the fact that their nodes have well-defined spatial positions, it is possible to consider a large number of measurements defined from such an information. Such geometrical features can be used in order to complement the information about the network connectivity provided by the respective topological measurements (e.g. node degree and clustering coefficient). Two important geometrical information which can be quantified relates to the angles and distances throughout the networks. In the case of the currently assumed trajectory networks, the distances between nodes do not provide a useful measurement because they are to a great extent an immediate consequence of the uniform distribution of the N points, and such stochastic distributions of points are known to have well-defined (e.g. [37]) interpoint distances depending only on the density of points, which is assumed fixed in the present investigations. On the other hand, distances between pairs of points considering the length of each edge are computationally expensive to be calculated, so they are also not considered here. We focus attention on the angles of each of the edges, which can be immediately calculated from the coordinates of the two respectively involved nodes. More specifically, for each edges (i, j), we calculate the respective angle:

$$a(i,j) = tg^{-1}\left\{\frac{y_j - y_i}{x_j - x_i}\right\} \quad (1)$$

Observe that such angles are intrinsically related to

the vector field orientations, being almost tangent to the lines of force.

Because for each trajectory network we have several angles, we consider their respective distribution, represented by a relative frequency histogram $h(i)$, $i = 1, 2, \ldots, H$. However, such histograms still imply several measurements for each network, so that it is interesting to considering functionals capable of mapping such distributions into a single measurements. In this work we resource to the *entropy* of the angle distribution, defined as:

$$e = -\sum_{i,h(i)\neq 0} h(i) log(h(i)) \tag{2}$$

Interestingly, this functional of the angle distribution is naturally invariant to the orientation of the original shape (different orientations will shift the histograms, but do not change their relative frequency values). In order to try to obtain additional information about the geometry of the trajectory networks, we also consider the two following alternative entropies: (i) the entropy of the relative frequency histogram of the angles to the third power; and (ii) the entropy of the smallest relative frequencies. More specifically, the first of these alternative entropies is given as:

$$e1 = -3\sum_{i,h(i)\neq 0} h(i)^3 log(h(i)) \tag{3}$$

The second entropy, $e2$, is obtained by sorting the relative frequency histogram h in descending order and considering only its second half, i.e. the $H/2$ smallest relative frequency values. Therefore, we have a total of 3 geometrical measurements for our trajectory networks: the classical entropy e and two variations $e1$ and $e2$. All these three measurements can be immediately calculated from the angle distribution.

C. Inter- and Intra-Class Dispersions and Discriminant Analysis

Once the trajectory networks obtained for the shapes under analysis have been obtained and their respective topological and geometrical measurements calculated, we typically proceed with the respective characterization and classification. Among the fundamental questions which are considered at this stage, we want to know how the shapes relate one another, in the sense of being similar or distinct (characterization and analysis). We may also want to assign a category to a newly presented shape (classification [2, 6, 8]). Such tasks are often challenging because of several factors including the statistical variation of the objects, the effect of the choice of measurements, the high dimensional spaces often implied by them, as well as the lack of a large number of samples and prototypes. A particularly effective approach to characterization, analysis and classification of patterns consists in taking into account the performance of several combinations of features as far as the separation of the categories of objects are concerned. Let C be the number of involved categories of shapes (each category is represented as C_i, $i = 1, 2, \ldots, C$), each with respective $P(i)$ samples, with $i = 1, 2, \ldots, C$. In addition, assume that each of these samples j, $j = 1, 2, \ldots, T$, has been characterized by M measurements, represented by the respective feature vector $\vec{f}_j$. Observe that $T = \sum_{i=1}^{C} P(i)$. The overall separation of the C categories can be quantified in several ways (e.g. [2, 6]), for instance by taking into account the dispersion inside each category as well as the dispersion between the categories. Usually, one wants the intra-class dispersion to be small compared with the inter-class dispersion. These two types of dispersions can be properly quantified by their respective scatter matrices. The *total scatter matrix*, S, expressing the overall dispersion of the measurements is defined as:

$$S = \sum_{j=1}^{T} \left(\vec{f}_j - \vec{\mu}\right)\left(\vec{f}_j - \vec{\mu_j}\right)^T \tag{4}$$

where μ is the average feature vector considering all observations. The *scatter matrix for each class* J_i is given as:

$$S_i = \sum_{v\in C_i} \left(\vec{f}_v - \vec{\mu}_i\right)\left(\vec{f}_v - \vec{\mu}_i\right)^T \tag{5}$$

where $\vec{\mu}_i$ is the mean feature vector of the observations inside each class C_i. The *intra-class scatter matrix* quantifying the dispersion within each class is defined as:

$$S_{\text{intra}} = \sum_{i=1}^{M} S_i \tag{6}$$

The *inter-class scatter matrix*, corresponding to the dispersion between categories, is given as:

$$S_{\text{inter}} = \sum_{i=1}^{C} N_i \left(\vec{\mu}_i - \vec{\mu}\right)\left(\vec{\mu}_i - \vec{\mu}\right)^T \tag{7}$$

Therefore,

$$S = S_{\text{intra}} + S_{\text{inter}}. \tag{8}$$

The quality of the overall separation between the categories can be quantified in several ways (e.g. [6]). One of the most commonly adopted functionals expressing the separation between the classes is given as:

$$J = trace\{S_{\text{intra}}^{-1} S_{\text{inter}}\} \quad (9)$$

It can be shown that J tends to increase for large inter-class dispersion and small intra-class dispersions. Discriminant (or canonical) analysis can be used to obtain the optimal linear combination of the measurements which maximizes J. The linear transform which maximizes J is given as:

$$\vec{F_v} = \Gamma \vec{f_v} \quad (10)$$

where $\Gamma = [\vec{\gamma_1}, \vec{\gamma_2}, \ldots, \vec{\gamma_d}]^T$ corresponds to matrix composed by the eigenvectors of the matrix $S_{\text{intra}}^{-1} S_{\text{inter}}$, with the respective eigenvalues to be arranged in decreasing order. By taking only the R initial eigenvectors, the dimensionality of the measurement space can be reduced while ensuring maximum separation between the categories as quantified by J. Because J depends on the chosen measurements $\vec{f}$, we henceforth express j as a function of the adopted features, i.e. $J = J(\vec{f})$.

III. RESULTS

The experiments reported in this article involved the generation of 30 realizations of trajectories networks for each of the shapes in Figure 2. The number of points uniformly distributed inside each of the shape was always equal to 1500 nodes, ensuring scale invariance. A total of $N_p = 400$ trajectories were followed from randomly selected nodes for each realization, and nodes were attached to growing trajectories provided they were no further away than $D_p = 2$ pixels apart from the tip of each trajectory. Figure 3 illustrates realizations obtained from each of the 8 shapes in Figure 2. Observe that the trajectories are not independent one another, in the sense that one trajectory typically connects to other adjacent trajectories.

We calculated the following measurements for each of the realizations of the trajectory networks: (a) number of nodes degrees with degrees 1 to 5, henceforth represented as $k1, k2, k3, k4$ and $k5$, respectively; (b) the entropy of the distribution of the angles of the edges (e); (c) the entropy of the values of the distribution of the angles of the edges to the third power ($e1$); and (d) the entropy of the upper half of the angle distribution sorted in descending order ($e2$). Therefore, we have 5 topological features ($k1$ to $k5$) and 3 geometrical measurements (e, $e1$ and $e2$). Figure 4 shows the scatterplots obtained by considering several two-by-two combinations of such measurements.

The combinations of $k2, k3$ and $k4$ yielded relatively poor separations. More specifically, we obtained $J(k2, k3) = 1.61$, $J(k2, k4) = 2.90$ and $J(k2, k4) = 3.90$. It is clear from the scatterplots in Figures 4(a-c) that the measurement $k2$ contributes little to the separation between the classes of shapes, with $k3$ and $k4$ providing increasing levels of separation, particularly regarding the categories ellipse/rectangle and cross. The other categories are virtually undistinguishable while considering $k2, k3$ and $k4$.

On the other hand, as it is clear from Figures 4(d-f) the angle entropy e allowed a much better separation between the classes of trajectory networks. More specifically, we obtained $J(e, e1) = 13.49$, $J(e, e2) = 12.18$ and $J(e, k1) = 12.84$. Basically, the left-hand half of the scatterplots in Figures 4(d-f) corresponds to ellipses, rhombus, circles and triangles, while the right-hand half contains the squares, pentagons, crosses and circles. Observe that the latter categories corresponds to more circular shapes, which tend to produce a more uniformly varying distribution of trajectory orientations and, consequently, higher entropies. On the other hand, the less circular objects in the former categories tend to imply heterogeneous angle distributions, decreasing the respective entropies. Interestingly, the alternative entropy $e1$ allowed further separation between the left-hand group, namely grouping triangles together with rhombuses and ellipses together with rectangles, which are indeed geometrically similar as far as their elongations are concerned. The second alternative entropy, $e2$, had little effect in increasing the separation between the shape categories. Interestingly, the consideration of the measurement $k1$, shown in Figure 4(f), allowed a marked separation of the crosses from the remainder categories, but contributed little to the separation of the latter. This effect can be explained as a consequence of the fact that the number of nodes with degree 1 (i.e. extremity nodes) tend to be proportional to the perimeter of the respective shapes, and the cross had the longest perimeter (relatively high values of $k1$ are also observed for rectangles and ellipses, while the smallest perimeters have been obtained for the circles).

So far we have only considered two-by-two combinations of the 8 topological and geometrical measurements. We now apply discriminant analysis over all these measurements, obtaining the scatterplot shown in Figure 5 with respective $J(discr) = 24.58$ (considering only the two canonical variables $v1$ and $v2$). In case all 8 the measurements are taken into account, we have $J = 27.98$, so that the two-dimensional projection retains most of the maximum separability that would be allowed by considering the full original measurement space. It is clear that the consideration of all the 8 measurements allowed a substantial improvement in the separation between the trajectory networks. The weights of the linear combinations yielding the canonical variables $v1$ and $v2$ are given in Table I. In order to allow a more direct interpretation of these weights as corresponding to the contribution of each measurement for the separation between the categories, the measurements were standardize [2, 6]. Such a standardization corresponds to subtracting each measurement from its average and dividing by the respective standard deviation. The new measurements are guaranteed to have null average and unit standard deviation, with most of the values falling within the interval from

-2 to 2. Therefore, the weights provided by the eigenvector components correspond to the larges absolute value eigenvalues express to the contribution of the respective measurement to the canonical variables. We have from Table I that the angle entropy e provides the most important contribution for the separation of the considered trajectory networks, followed by the number of nodes with degree 5 ($k5$) and 4 ($k4$), and the first alternative entropy $e1$. The number of extremity nodes ($k1$) also contributed strongly for the second canonical variable $v2$. The contributions of the other measurements are relatively smaller.

The distribution of the samples of the several categories in the scatterplot in Figure 5 can be interpolated by the Parzen windows approach and used for Bayesian decision theory [2, 8, 14], allowing optimal correct classifications given the provided data. More specifically, such optimal performance is allowed also as a consequence of the proper choice of measurements identified by the discriminant analysis. In other words, it is this integration between the choice of measurements and classification stages that make the overall approach particularly effective.

IV. DISCUSSION

It is clear from the obtained results that the trajectory network representation of the original shapes does allow clear discrimination between the involved shape categories. We found that each of the considered topological and geometrical features contributed quite differently to the separation between the trajectory networks. In particular, the angle entropy allowed the overall best discrimination, followed by the number of nodes with degrees 4, 5 and 1, as well as the first alternative entropy corresponding obtained from the third power of the angle relative frequency histograms. Interestingly, except for $k4$ and $k5$, all other features are directly or indirectly related to the geometrical properties of the original shapes. For instance, all angle entropies are direct geometrical measurements, while the number of extremity nodes (i.e $k1$) is related to the perimeter of the original shapes. Among the topological features, we found that the nodes with higher degree tend to be more discriminative. We did not consider higher degrees because they become increasingly less represented and unstable in the trajectory networks. For instance, nodes with higher degrees tend to appear sporadically in trajectory networks obtained from a same original shape, implying statistical noise.

TABLE I: The values of the weights of each of the 8 standardized measurements in the linear combinations yielding the canonical variables $v1$ and $v2$.

Measurement	$v1$	$v2$
entropy e	0.86	-0.96
alt. entropy $e1$	0.22	0.22
alt. entropy $e2$	-0.11	0.14
$k1$	-0.12	-0.29
$k2$	0.10	0.01
$k3$	0.11	0.05
$k4$	0.24	0.24
$k5$	0.33	0.27

At this point in our work, it is interesting to identify the main advantages and disadvantages of the trajectory network approach to shape analysis. Among the advantages, we have:

* *Simplicity:* Contour detection and trajectory following are easy and fast to obtain, and the 8 adopted measurements can be immediately calculated at low numerical cost;

* *Invariance:* The trajectory network approach to shape analysis is naturally invariant to scaling, orientation and translation.

* *Few involved parameters:* Only three parameters are involved: the number N_p of trajectories, the number N of spatially distributed points, and the distance D_p considered while joining new nodes into the growing trajectories/paths.

* *Integration of complex networks and shape analysis:* The trajectory network approach to shape analysis natural and intrinsically integrate the two important major areas of complex networks and image analysis. This paves the way for intense cross-fertilization between these two fields, allowing not only complex networks concepts to be applied to shape analysis, but also the other way round.

The disadvantages of the approach are mainly related to the fact that distinct, though intrinsically similar, trajectory networks are obtained for the same given shape. This is a consequence of the random distribution of the N points. The consideration of fixed sets of nodes implies a certain level of orientation anysotropy.

V. CONCLUDING REMARKS

The current article has reported an integrated approach to shape analysis based on two principal concepts: (i) representation of the original shapes in terms of respective trajectory networks, and (ii) the integration of choice of measurements and discrimination/classification between the categories. It has been shown that trajectory networks do provide enough information about the original shapes so as to allow their discrimination in terms of a reduced set of topological and geometrical measurements. High quality separation (with respect to the quality functional J) can be obtained by considering the two canonical variables identified by discriminant analysis. These two variables have been found to be predominantly influenced by the angle entropy, number of nodes with degrees 1, 4 and 5, as well as the alternative entropy obtained by considering the third power of the histogram of relative

frequencies of the angles of the edges. It has been found that the number of nodes with medium degrees (except for the number of extremity nodes, which is also important) tend to contribute more intensely to the discrimination between the trajectory networks. In brief, we have shown that it is possible to discriminate between shapes represented in terms of respective trajectory networks. We have also identified among 8 topological and geometrical measurements, those which contribute the most to the separation between the categories. However, the contributions of this work extend further than classification of shapes. More specifically, wider contributions and implications of the reported concepts and methods include:

Integration of complex networks and shape analysis: The combination of complex networks and shape analysis concepts and tools allows two-direction cross-fertilization between these two areas, allowing not only concepts from complex networks research to be applied for shape analysis, and vice-versa. For instance, it would be interesting to consider the application of methods such as skeletonization [2] and watersheds on complex networks. Preliminary steps in the integration of mathematical morphology, an important area in shape and image analysis, and complex networks have been provided in [38]. The use of the transition probabilities obtained by assuming some specific dynamics, such as self-avoiding walks, has also been found to allow the identification of the borders of networks [39, 40].

Alternative entropies: The entropy of a distribution provides a particularly relevant summarization of its overall structure. In the case of the shape analysis problem addressed in the present work, the angle entropy allowed invariance to rotation of the shapes as a consequence of the fact that such rotations only shift the respective relative frequency histograms, but do not change their respective intensities, which are the only values considered during entropy calculation. In addition, entropy is naturally connected to information (e.g. [41]). Because of such special properties of the entropy of a distribution, it becomes interesting to devise means to obtain additional information from the original distribution by considering alternative definitions of entropy. Two such variants have been considered, possibly for the first time, in the present work. The first alternative entropy corresponds to the calculation of the entropy of powers of the original distribution values. Such a measurement was found to enhance the separation between some categories of the considered shapes. The second type of entropy used in this work is obtained by sorting the relative frequency values and calculating the traditional entropy only of half of the so-obtained distribution. Though this type of entropy was verified to contribute little to the separation between the trajectory networks associated to the 8 categories of shapes taken into account in this work, it is possible that such an alternative definition of entropy can provide valuable results in other occasions.

New type of geographical knitted network: Though trajectory networks have already been addressed [9], the derivation of such networks from respective shapes provide a new subcategory of networks which is interesting for several reasons. In particular, these networks are mostly tree-like and have their extremity points closely related to the borders of the original objects, so that trajectories with specific geographical structures can be easily designed from the respective shapes (e.g. a 'spiral' network can be easily derived from the shape of a spiral). This approach can also be considered for modeling real-world systems, such as the establishment of roads in an Island or growth under influence of boundary conditions.

A new way to transform a shape into a network: The use of vector fields defined by the shapes in order to obtain the respective network representation corresponds to a new way to map a shape into a graph. Previous approaches (e.g. [21]) involved the distribution of nodes along the borders of the shape. The transformation suggested in this work is simple and involves little computational efforts, especially if the spectral methodology reported in [27] is adopted.

Several are the possibilities for future work motivated by the concepts and findings reported in the current article. These possibilities include:

Other topological and geometrical measurements of networks: Though we considered 8 relevant measurements of the topology and geometry of trajectory networks in this work, it would be interesting to take into account other measurements (e.g. [14]). Of special interest would be to obtain hierarchical measurements (e.g. [14, 38]) for several of the nodes of the trajectory networks and investigate the respective potential for discrimination between the networks.

Other types of fields: Vector fields other than electrical can be associated to a given shape. Of special interest are fields defined by other powers of the distance from the border, as well as the consideration of distance fields (e.g. [42]).

Preferential trajectory networks: Given a shape, it is possible to obtain a respective network representation by considering preferential attachment. This can be done by distributing N nodes within the interior of the shape and then connecting these nodes to the points along the shape contour with probability varying inversely with the respective distances. So, the points which are closer to high curvature points tend to become more connected to the border. Once such connections between the N points and the respective shape contour are established, it is possible to apply the rich get richer attachment procedure to obtain the whole network. It would be particularly interesting to investigate whether such networks allow better separation between the respective shapes by applying the methodology reported in the current article.

Systematic investigation of alternative entropies for pattern recognition: The two types of alternative entropies proposed in this work seem to present potential for providing additional information about the respective distribution while retaining the advantages of traditional entropy, namely its invariance with permutations of the

values considered for the calculation of the relative frequencies. It would be interesting to consider such alternative entropy measurements in other types of pattern recognition problems, as well as considering other sets of shapes, in order to better evaluate the potential of such alternative features.

Extension to gray-level images: Though we have focused our attention on trajectory networks obtained by considering vector fields defined by shapes, it is immediate to extend such an approach to gray-level image analysis by considering the respective vector fields to be given by the gradient of the scalar field of gray-level intensities.

Acknowledgments

Luciano da F. Costa thanks CNPq (308231/03-1) and FAPESP (05/00587-5) for sponsorship.

[1] D. Marr, *Vision* (W. H. Freeman, 1982).
[2] L. da F. Costa and R. M. Cesar, *Shape Analysis and Classification: Theory and Practice* (CRC Press, 2001).
[3] R. M. Cesar and L. da F. Costa, Patt. Recogn. **29**, 1559 (1996).
[4] L. F. Estrozi, L. G. R. Filho, A. G. Campos, R. M. C. Junior, and L. da F. Costa, Dig. Sign. Proc. **13**, 172 (2003).
[5] R. F. Oliveira, L. da F. Costa, and A. C. Roque, Neurocomputing **65–66**, 117 (2005).
[6] K. Fukunaga, *Statistical Pattern Recognition* (Morgan Kaufmann, 1990).
[7] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition* (John Wiley and Sons, 1998).
[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley Interscience, 2001).
[9] L. da F. Costa (2008), arXiv:0803.2447.
[10] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[11] S. N. Dorogovtsev and J. F. F. Mendes, Advs. in Phys. **51**, 1079 (2002).
[12] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[13] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Phys. Rep. **424**, 175 (2006).
[14] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, Advs. in Phys. **56**, 167 (2007).
[15] L. da F. Costa (2007), arXiv:0711.1271.
[16] L. da F. Costa (2007), arXiv:0711.2736.
[17] D. W. Shattuck and R. M. Leahy, Med. Imag. **20**, 1167 (2001).
[18] R. Torres, A. Falcao, and L. da F. Costa, Patt. Recogn. **37**, 1163 (2004).
[19] L. da F. Costa (2004), arXiv:cond-mat/0403346.
[20] P. F. Felzenswalb and D. P. Huttenlocher, Intl. J. Comp. Vis. **59**, 167 (2004).
[21] L. da F. Costa (2006), arXiv:physics/0603025.
[22] T. Chalumeau, L. da F. Costa, O. Laligant, and F. Meriaudeau, in *18th Annual Symposium on Electronic Imaging, SPIE Conf. Machine Vision Applications and Industrial Inspection* (San Jose, Ca, 2006), p. 6070.
[23] L. da F. Costa (2006), arXiv:cs/0606060.
[24] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno, Th. Comp. Sci. **392**, 5 (2008).
[25] R. Manevitch, J. Berdine, B. Cook, G. Ramalingam, and M. Sagiv, in *TACAS 2007, 13th Intl. Conf. Tools and Algorithms for the Construction and Analysis of Systems* (Braga, Portugal, 2007), pp. 3–18, http://www.cs.tau.ac.il/ msagiv/tacas07.pdf.
[26] L. Diambra and L. da F. Costa, Bioinformatics **21**, 3846 (2005).
[27] L. da F. Costa, J. Real-Time Imag. **5**, 243 (1999).
[28] M. Schroeder, *Fractals, Chaos, Power Laws* (W. H. Freeman, 1991).
[29] D. ben Avraham and S. Havlin, *Diffusion and reactions in fractals and disordered systems* (Cambridge University Press, 2000).
[30] M. Kaiser and C. C. Hilgetag, Phys. Rev. E **69**, 036103 (2004).
[31] Z. Toroczkai, B. Kozma, K. E. Bassler, N. W. Hengartner, and G. Korniss (2004), cond-mat/0408262.
[32] Z. Toroczkai and K. E. Bassler, Nature **428**, 716 (2004).
[33] L. da F. Costa and G. Travieso, Intl. J. Mod. Phys. C **16**, 1097 (2005).
[34] L. da F. Costa, B. A. N. Travencolo, A. Azeredo, M. E. Beletti, G. B. Muller, D. Rasskin-Gutman, G. Sternik, M. Ibanes, and J. C. I. Belmonte, Appl. Phys. Letts. **86**, 3901 (2005).
[35] E. J. Huang and L. F. Reichardt, Europhysics News **36**, 218 (2005).
[36] E. P. Borges, D. O. Cajueiro, and R. F. S. Andrade, Eur. Phys. J. B **58**, 469 (2007).
[37] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications* (Wiley and Sons, 1996).
[38] L. da F. Costa and L. E. C. da Rocha, The Eur. Phys. J. B **50**, 237 (2005).
[39] L. da F. Costa (2008), arXiv:0801.1982.
[40] B. A. N. Travencolo and L. da F. Costa (2008), arXiv:0802.3665.
[41] T. M. Cover and J. A. Thomas, *Information Theory* (Wiley Interscience, 1991).
[42] R. Fabbri, L. da F. Costa, J. C. Torelli, and O. M. Bruno, ACM Comp. Survs. **40**, 1 (2008).

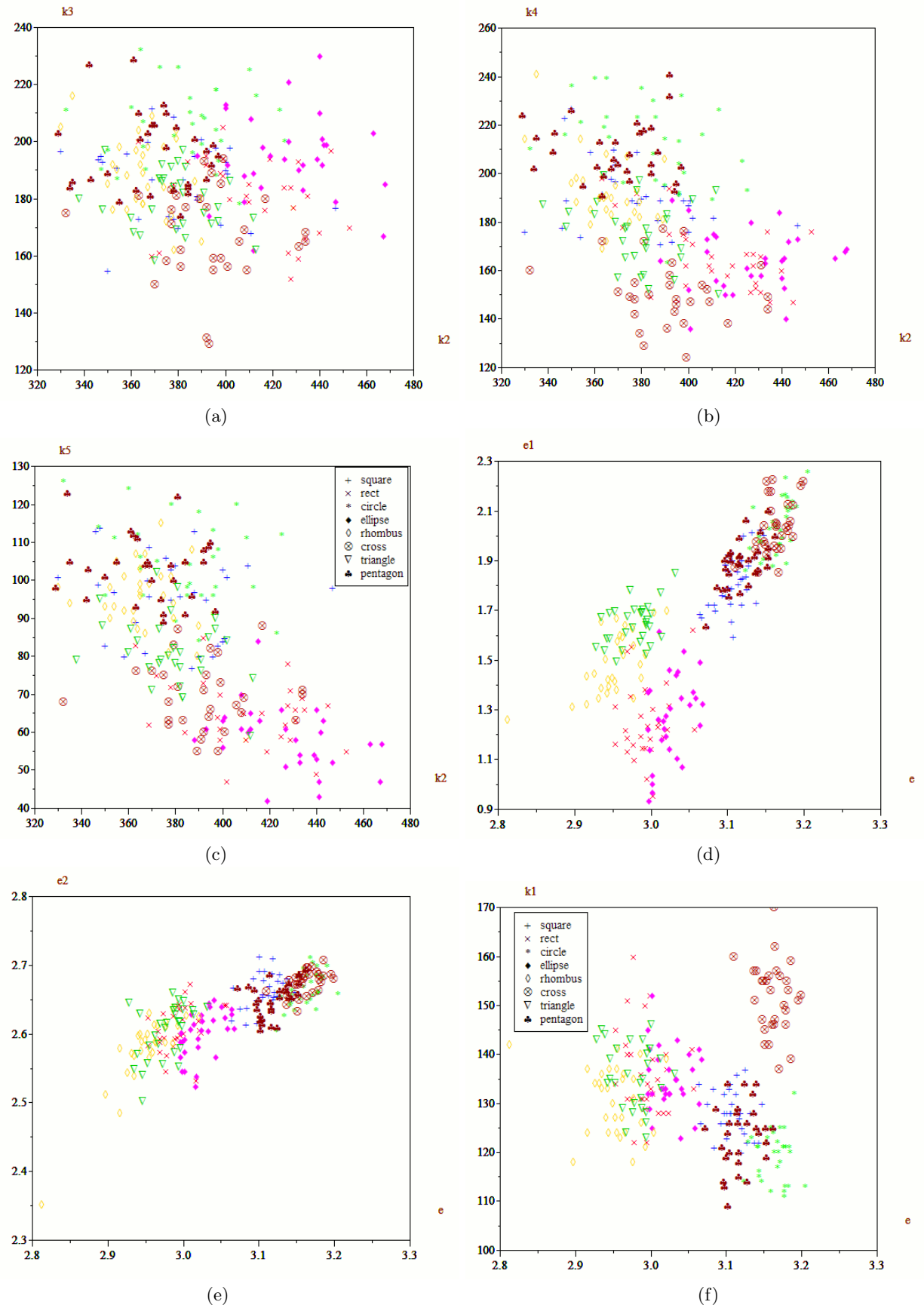


FIG. 4: The scatterplots obtained for the 30 realizations of trajectory networks obtained for each of the 8 considered shapes while considering several two-by-two combiations of the 8 adopted topological and geometrical measurements.

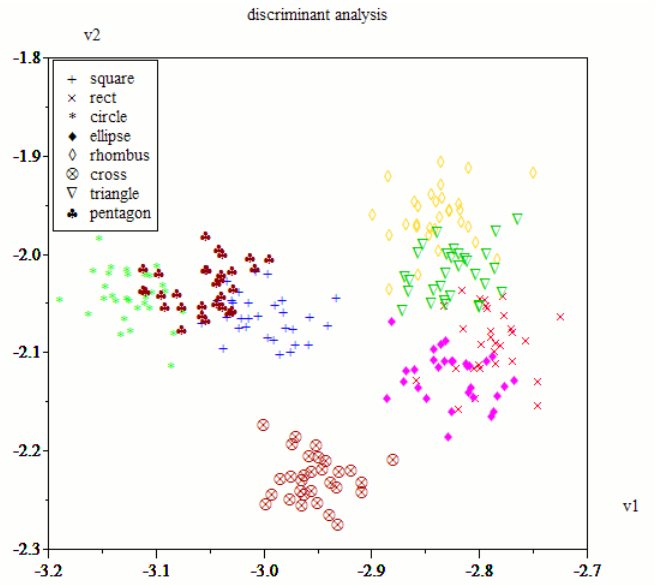


FIG. 5: The distribution of the 30 realizations of each of the 8 trajectory networks types (obtained for each of the 8 shapes in Fig. 2) as derived by discriminant (canonical) analysis considering all the measurements (i.e. the number of nodes with degree 1 ($k1$), 2 ($k2$), 3 ($k3$), 4 ($k4$) and 5 ($k5$), the angle entropy e, as well as the two alternative entropies $e1$ and $e2$.