

An Optimized Shape Descriptor Based on Structural Properties of Networks

Gisele H. B. Miranda,^{1,*} Jeaneth Machicao,^{2,†} and Odemir M. Bruno^{2,1,‡}

¹*Institute of Mathematics and Computer Science,
University of São Paulo, São Carlos - SP, 13560-970, Brazil.*

²*São Carlos Institute of Physics, University of São Paulo,
São Carlos - SP, PO Box 369, 13560-970, Brazil.*

The structural analysis of shape boundaries leads to the characterization of objects as well as to the understanding of shape properties. The literature on graphs and networks have contributed to the structural characterization of shapes with different theoretical approaches. We performed a study on the relationship between the shape architecture and the network topology constructed over the shape boundary. For that, we used a method for network modeling proposed in 2009. Firstly, together with curvature analysis, we evaluated the proposed approach for regular polygons. This way, it was possible to investigate how the network measurements vary according to some specific shape properties. Secondly, we evaluated the performance of the proposed shape descriptor in classification tasks for three datasets, accounting for both real-world and synthetic shapes. We demonstrated that not only degree related measurements are capable of distinguishing classes of objects. Yet, when using measurements that account for distinct properties of the network structure, the construction of the shape descriptor becomes more computationally efficient. Given the fact the network is dynamically constructed, the number of iterations can be reduced. The proposed approach accounts for a more robust set of structural measurements, that improved the discriminant power of the shape descriptors.

I. INTRODUCTION

In computer vision, shape boundaries are important attributes that can be used for the characterization and the classification of objects. In shape analysis there are many pattern recognition applications covering different areas, such as neuroscience [1–3], agriculture [4–9], medical imaging [10–13], remote sensing [14, 15], to mention but a few. Over the last decades many methods were proposed in the literature of pattern recognition which are based on classical approaches, such as, Fourier descriptors [16, 17], wavelets [18, 19], fractal dimension [9, 20–22], curvature scale space (CSS) [23] among others. These methods support a wide range of applications. In addition, methods based on structural properties of shape boundaries have been drawing attention for classification tasks. Such methods are strongly influenced by graph and network theory [20, 24–27]. Instead of considering the shape boundaries only as a chain of connected points, this new approach also takes advantage of the topological properties of the shape contour.

In a paper published in 2009 [24], Backes *et al.* proposed a method, named as CNDescriptor, for boundary shape analysis based on the connectivity among the contour pixels. In this one-parameter model, each pixel is modeled as a vertex of a graph and the connections between them are established according to that parameter, which represents a distance threshold. For a given threshold value, there is one graph realization and a set of measurements regarding the connectivity of this graph can be obtained. These measurements are the so-called network descriptors which characterize the graph structure, and, consequently, unveil the topological properties of the shape boundary. In the referred paper, the authors also detail some properties of the graphs obtained through the proposed method, e.g., for some thresholds the graphs present a high clustering coefficient, characterizing them as small-world graphs. In addition, the CNDescriptor is invariant to geometric transformations such as rotation and scale and robust for what concerns the presence of noise. This method was also evaluated in the context of image degradation, in which parts of the shape contour were removed. The authors demonstrated that the CNDescriptor is also very robust in such cases, showing promising results. It is also possible to find an extension of the method for texture analysis [28].

However, in spite of being well suited for shape analysis tasks, the dynamic evolution signature provided by the method CNDescriptor only accounts for degree-related measurements as feature vectors. However, a much more robust set of measurements could improve the classification performance in different shape recognition applications. Along the last two decades, Complex Networks (CN) has been established as a new research field which integrates graph theory and statistical mechanics [29–31]. Many related studies have provided new insights about the topological characterization of networks leading to a deep understanding of how the connectivity patterns of the nodes are related

* gimiranda@usp.br

† machicao@usp.br

‡ bruno@ifsc.usp.br

to the network model [32–34]. Therefore, the measurements extracted from the network structure are important features for the network characterization. In a survey of 2007, Costa *et al.* emphasize this approach [35]. The authors summarize a huge set of measurements that can be used to describe the topology of a network. They present different categories of measurements, such as, connectivity measurements which include, for instance, mean degree and degree distributions and correlations. Distances and path lengths as well as hierarchical and spectral measurements are other important categories. Centrality measurements can quantify how important is a node to the network topology for what concerns its robustness and noise tolerance, e.g., betweenness, closeness and eigenvector centrality [36, 37]. It is also possible to quantify clustering and cycles in a network through measurements like transitivity and the clustering coefficient, which can be used to characterize the small-world property [34]. In addition, Costa *et al.* also addresses the use of classical pattern recognition techniques for network analysis [38].

In this paper we present an extension of the previous work of Backes *et al.* [24] and we propose a generalized approach for shape characterization in Computer Vision based on network analysis. This approach accounts for different categories of measurements. We have demonstrated how these measurements are related with the properties of shape boundaries. For this task we used samples obtained through an interpolation process performed between regular shapes. This way, we could analyze, for instance, how the emergence of internal angles of a shape influences the connectivity of the resulting network. Besides, we also performed an analysis of the curvature signal of those shapes and its relationship with structural network measurements. We observed that the pixels of high curvature are not the ones with the highest degree, instead, they present a high clustering coefficient. In addition, we also evaluated the performance of the proposed approach in three different applications concerning shape recognition. The first image dataset contains geometric shapes of ten different classes. The second dataset contains generic shape contours belonging to nine different categories like animals, fishes and tools among others. This dataset has been used as benchmark in many other studies [39, 40]. The last dataset evaluated in this paper is the same dataset yielded by Backes *et al.*, which was used for comparing purposes as well as in order to validate the methodology in a real-world application. This dataset contains images of leaf contours from 30 different plant species.

This paper is organized as follows: Section II presents a detailed description of the previous work of Backes *et al.* as well as a basic introduction in what concerns the structural characterization of networks. In Section III, we present a study regarding structural properties of networks for different geometric shapes using an interpolation approach as well as a curvature analysis. In section IV, we evaluated the proposed approach regarding the classification task for three distinct datasets, and, finally, Section ?? presents the discussion.

II. BACKGROUND

A. Previous work

In the work of Backes *et al.* [24] is introduced a shape descriptor based on the dynamic evolution of a network built from the contour points. This method is based on a distance criterion in order to establish the connections between the pixels. The threshold parameter, T , is used for modeling shape boundaries and to obtain the corresponding feature vectors. All this process is detailed next.

Let S be the contour of an image, such that $S = \{s_1, s_2, \dots, s_N\}$, where $s_i = [x_i, y_i]$ are the coordinates of point i , represented by discrete values. Given a graph of the form $G = (V, E)$, each pixel of the contour represents a node in the graph, and, therefore, $S = V$. The set of non-directed edges E is defined for each pair of nodes and the corresponding weight is calculated by the Euclidean distance, as follows:

$$d(s_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (1)$$

The adjacency matrix W of this weighted network is represented by the $N \times N$ matrix, such that $w_{ij} = W([w_i, w_j]) = d(s_i, s_j)$, which is then normalized in the interval $[0, 1]$: $W = \frac{W}{\max_{w_{ij} \in W}}$. N is the total number of nodes in the network.

Initially, the network is regular, since each node is connected to all the others. At this step, a threshold transformation, T_l , can be applied in order to obtain a new set of edges E' . This set is composed by the edges whose weights are smaller than T_l . If T_l is small, the number of edges will be also small and the network will be composed of many connected components without intersection. Otherwise, if T_l is large, the network will be almost fully connected. For intermediate values of T_l , other properties begin appearing, such as the small-world property, characterized by the presence of many connected triples. The authors have demonstrated this property for a set of distinct shapes. Therefore, the threshold parameter controls the connectivity of the network. Notice that the same process can be im-

plemented considering the connections which are above the threshold, i.e., two pixels will be connected if the distance between them is greater than T_l .

Based on the properties that arise from the transformations defined by T_l , the shape descriptor is obtained through the connectivity measures extracted from the network topology. The dynamic evolution of the network as a function of T_l provides distinct attributes which are then combined to compose the feature vector. The transformation is formally defined by the δ operation as follows:

$$A_{T_l} = \delta_{T_l}(W) = \forall_W \in W \begin{cases} a_{ij} = 0, & \text{if } w_{ij} \geq T_l \\ a_{ij} = 1, & \text{if } w_{ij} < T_l \end{cases}, \quad (2)$$

where A is the resulting unweighted and thresholded matrix. Therefore, the shape characterization is performed through a series of δ transformations where T_l is regularly incremented by T_{inc} . Fig. 1 illustrates the $\delta_{T_l}(W)$ transformation. The left part of the figure presents the networks obtained by applying the comparison *smaller than*. Therefore, as T_l increases more connections are established. Similar analysis is shown in the right part of the figure, which illustrates the condition *greater than*, $\Delta_{T_l}(W)$, however with the opposite behavior. Given that the distance values are normalized between 0 and 1, the threshold values are defined in the same interval.

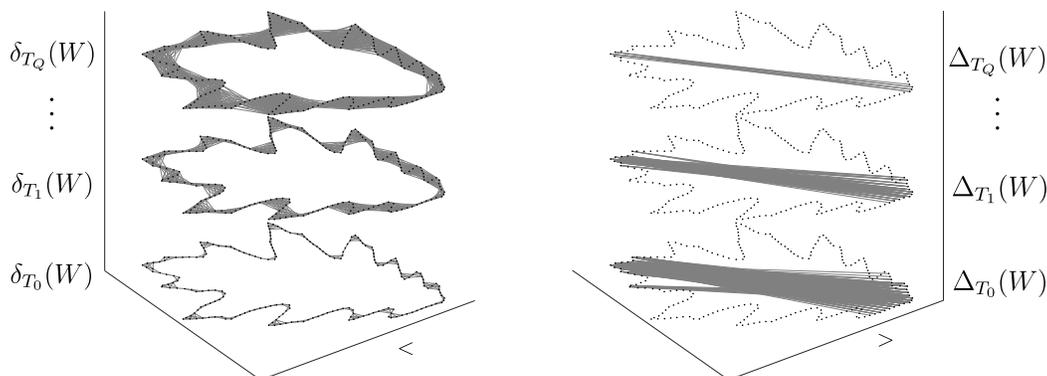


FIG. 1. Network construction from the contour pixels. The left-side depicts the connections established for different threshold values by applying the comparison *smaller than*. The right-side illustrates the comparison *greater than*.

The measurements obtained at each $\delta_{T_l}(W)$ are: the *average degree* (k_μ) and the *max degree* (k_κ), which corresponds to the average and to the maximum degree of all the network nodes, respectively. The final feature vector, φ , is represented by

$$\varphi = [k_\mu(T_0), k_\kappa(T_0), k_\mu(T_1), k_\kappa(T_1), \dots, k_\mu(T_Q), k_\kappa(T_Q)], \quad (3)$$

where the values k_μ and k_κ correspond to the *average* and *max degree* for each $\delta_{T_l}(W)$ considered. Backes *et al.* [24] showed that the proposed descriptor is invariant to rotation and scale transformations applied on images, and, it is robust to noise and image degradation.

In the present paper, in addition to degree-related measurements, we investigated a more generalized approach for shape characterization based on network analysis. Different sets of measurements were evaluated which are capable of describing important structural properties of shape boundaries. In the next subsection, we introduce the definition of some network measurements. And further in this paper we discuss how these measurements can be used as feature vectors.

B. Structural network characterization

A large variety of structural measurements can be used to characterize networks [35]. It is possible to identify different categories such as connectivity and distance related measurements. Some of these groups are described next.

1. Connectivity measurements

The number of neighbors of a given node i , k_i , is one of the simplest measures regarding the connectivity of a network. From k_i a set of measurements can be derived [35] as the average degree,

$$\langle k \rangle = \frac{1}{N} \sum_i k_i, \quad (4)$$

which is computed considering the degrees of all the nodes in the network. In contrast to $\langle k \rangle$, the hierarchical degree, k^h , accounts for the connectivity of the node neighbors that are restricted to a hierarchy level h . For instance, the hierarchical degree of level 2 of a given node i , k_i^2 , is the sum of the degrees of its neighbors. Therefore, the average hierarchical degree of a network, $\langle k^h \rangle$, is given by,

$$\langle k^h \rangle = \frac{1}{N} \sum_i k_i^h, \quad (5)$$

where h defines the hierarchy level. Another interesting measurement is the clustering coefficient [34], cc_i , which express the probability of two vertices j and k being connected to each other since both are connected to node i . It is calculated by $cc_i = 2e_i/k_i(k_i - 1)$, where e_i represents the number of edges between the neighbors of node i . Therefore, the average clustering coefficient, $\langle cc \rangle$, is given by

$$\langle cc \rangle = \frac{1}{N} \sum_i cc_i. \quad (6)$$

2. Distance-related measurements

A *path* that connects two nodes i and j in a network is defined by the sequence of nodes which must be visited to go from i to j , and, the number of links is denoted by the distance between them, d_{ij} . The geodesic path is the path that has the smallest value of d_{ij} . If there is no path between i and j , then $d_{ij} = \infty$ or only the largest connected component of the network can be considered. The average path length is defined as follows [35]:

$$\langle l \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (7)$$

where $\frac{1}{N(N-1)}$ is the normalization factor for a totally connected network.

3. Degree correlation

The degree correlation quantifies the tendency of nodes that present high k to connect with nodes that also have a high degree, or, with nodes that present lower values of k . It is also called assortativity [41] and can be calculated as follows:

$$\rho = \frac{(1/M) \sum_{j>i} k_i k_j a_{ij} - [(1/M) \sum_{j>i} (1/2)(k_i + k_j) a_{ij}]^2}{(1/M) \sum_{j>i} (1/2)(k_i^2 + k_j^2) a_{ij} - [(1/M) \sum_{j>i} (1/2)(k_i + k_j) a_{ij}]^2}, \quad (8)$$

ρ belongs to the interval $-1 \leq \rho \leq 1$. Value of ρ that are close to 1 indicate a positive correlation, whereas values close to -1 , a negative correlation. Finally, values near zero indicate no linear dependency for what concerns the degrees of the network nodes.

4. Betweenness centrality

The betweenness centrality is a measure that accounts for the importance of a node regarding the information load it is responsible in the network [36]. Given j and k , two nodes that are not adjacent, i.e., that are not directly linked, the communication between them depends on the nodes that belong to the paths that connect j and k . Therefore, the betweenness b_i of a node i is calculated as a function of the number of geodesic paths passing through i [36]:

$$b_i = \sum_{j,k,j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \quad (9)$$

on which, $n_{jk}(i)$ is the number of geodesic paths connecting j and k passing through i and n_{jk} is the number of geodesic paths connecting j and k . We adopted the normalized version of Eq. 9 in the experiments we performed in this paper: $b_i = \frac{1}{n^2} \sum_{j,k,j \neq k} \frac{n_{jk}(i)}{n_{jk}}$, where n^2 is the total number of possible links. The definition of the betweenness centrality can also be extended to the links of the network, therefore b_i is given as a function of the number of paths passing through an specific link or edge [35].

III. SHAPE PROPERTIES REVEALED BY STRUCTURAL MEASUREMENTS

The dynamic evolution of the network, defined by the $\delta_{T_l}(W)$ transformation over the shape boundary, allows the characterization of its contour as well as the intrinsic structural properties that can be associated to that shape. The connectivity of the underlying network is directly influenced by the shape contour, and, therefore, it can be used for pattern recognition.

Besides degree-related measurements, we investigated the performance of a more general set which comprises clustering, distance and centrality measurements. Given that each structural measurement captures specific properties of the network topology, we investigated a set measurements for the characterization of networks which were constructed from regular geometric shapes. The use of these specific shapes allows to relate geometric characteristics, such as angle and curvature, to the structural properties of the network generated according to the proposed approach.

The next sections illustrate how such measurements are related to shape properties. Therefore, the proposed approach generalizes the method that we presented in the previous section regarding the attributes that can be obtained from the dynamic evolution of the network. From now on, the generalized shape descriptor will be denoted by Φ in contrast to the network descriptor based only on the degree of the nodes, φ (Eq. 3).

A. Curvature analysis & Structural features

The curvature analysis explores parts of the image that exhibit high contrast and plays an important role in shape recognition. As pointed out in the famous work of Attneave [42], points that present high curvature are a rich source of information, and the human visual system can recognize a shape considering only those points.

For the network constructed from the shape boundaries, it should be noticed that, the degree of a pixel is also conditioned to the threshold parameter T_l . Therefore, the smaller the internal angles of a shape, the smaller the distance between the pixels of the shape boundary and, consequently, the higher the mean degree.

Fig. 2 illustrates how curvature C (dashed line) and the node degree are related to each other for different shapes. The values of both measurements were normalized between 0 and 1. Each row represents a specific shape and each column represents a different threshold which increases from left to right. The x -axis of each plot corresponds to the pixels of each shape, therefore, the curvature and the other measurements are presented per pixel. The method used for the estimation of the curvature signal is based on the DFT (*Discrete Fourier Transform*) [43]. The initial pixel (point), from which the curvature and is calculated, is highlighted for each polygon as a green dot. The aliasing effect is more apparent for some shapes, like the hexagon. This effect is reflected as small oscillations in the curvature, however we can easily identify the points of maximum curvature which are close to the corners of those shapes. For the square shape (second line), for instance, there are four equally spaced peaks that correspond to the corners. In general, for all shapes, it is possible to observe that, except for $T_l = 0.025$, pixels with the highest curvature are the ones with the lowest degrees. This occurs because they are distant from the other pixels when compared to the points located on the sides of the square. For $T_l = 0.025$, the most connected pixels are the ones near the corners, since at this level the network presents only a few connections. Then, as the threshold increases and the network gets more connections, the pixels between two corners are the ones with the highest degrees. For what concerns the stars, we can observe a complementary behavior between the corners with positive and negative curvature. Finally, the last

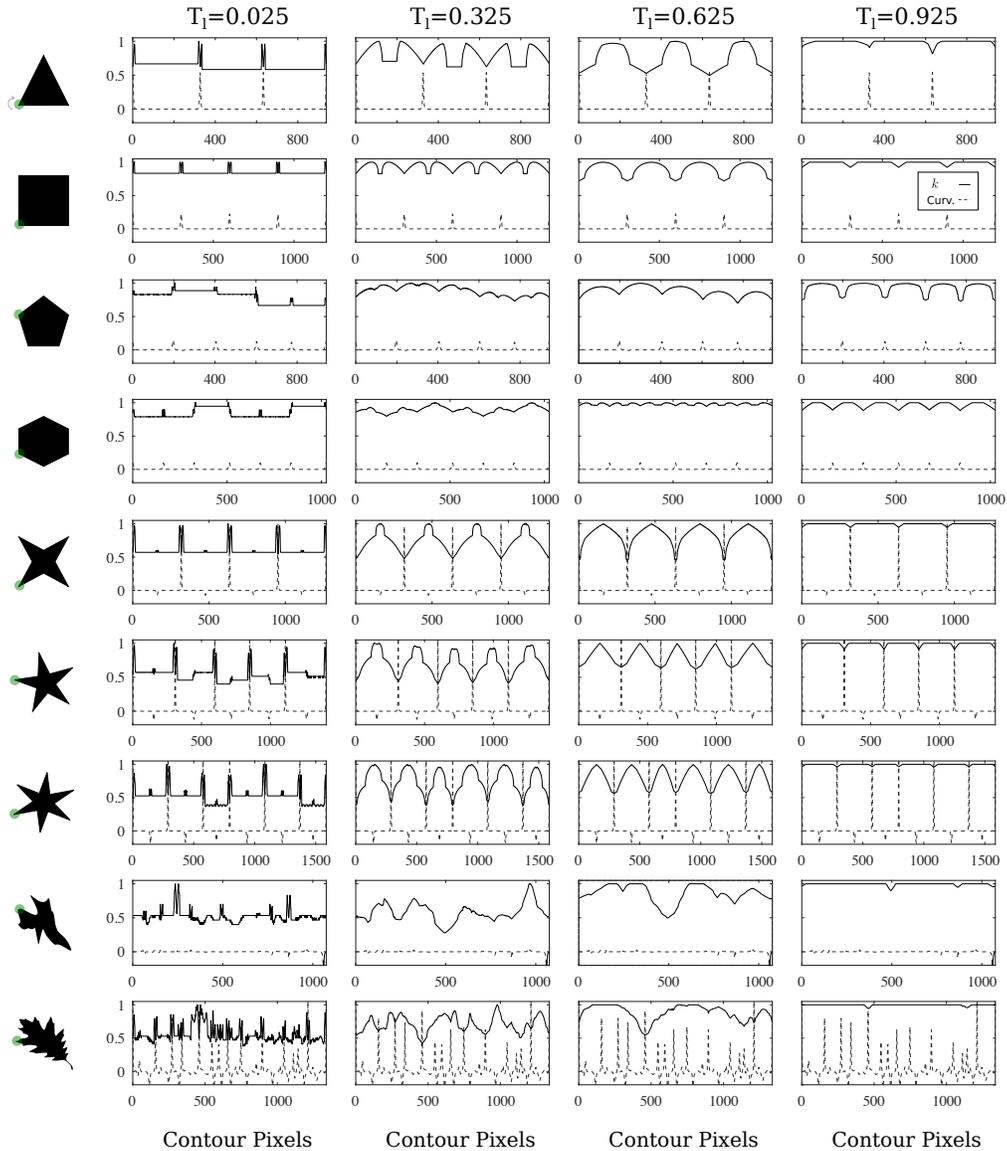


FIG. 2. Comparison of the curvature signal (dashed line) and the degree of each node as a function of the threshold parameter T_l for different geometric shapes.

two rows present the degree and curvature signals belonging to irregular shapes, for which the general assumptions are also valid.

Similar comparison was performed for other measurements which were also obtained from the network structure. The results are presented in Fig. 3 for the triangular shape. An opposite behavior was obtained for the clustering coefficient curve (first row) when compared to the degree analysis presented in Fig. 2. The pixels with the highest curvature are the ones with the highest values of cc . In the case of the triangle, the corner presents connections between the two line segments that define this corner, which increases the clustering coefficient of that node. Moreover, in general, for the great majority of the plots, all the analyzed shapes present high values of average clustering coefficient (greater than 0.5). As the threshold increases, the number of connected triples of the corresponding network also increases, but, now, also reflecting the increase for what concerns the density of the network nodes. This means that if pixel i is connected to pixel $i - 1$, given a radius r , then pixel $i + 1$ will also be connected to pixel $i - 1$, giving rise to a small-world network.

For what concerns the betweenness centrality (second row of Fig. 3), it is possible to observe that, similarly to the node degree, that measurement is also high for the nodes that are not in the corners of the shape. As both T_l and the number of connections of the network increase, new pathways are established among the pixels. Consequently, the pixels with the highest curvature values will not be in the shortest paths, which influences the betweenness of the

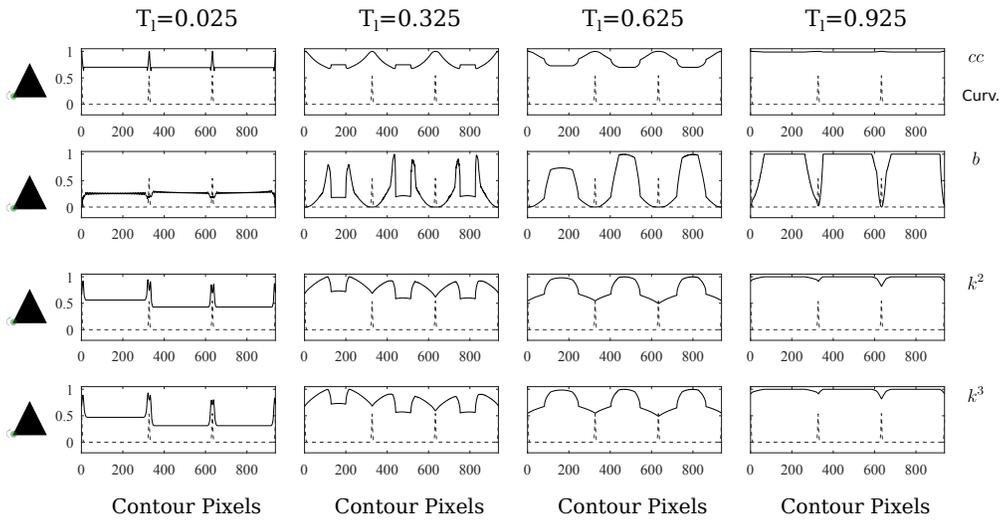


FIG. 3. Comparison of the curvature signal (dashed line) and other structural measurements: clustering coefficient, betweenness, hierarchical degree of level 2 and hierarchical degree of level 3 (from top to bottom), as a function of the threshold parameter T_l for the triangular shape.

node. Finally, the last two rows of Fig. 3 illustrate the relationship between the curvature and the hierarchical degree of levels two and three, respectively, for each pixel. Once more, we can observe that the general behavior of both measurements follows the same trend observed for the node degree, as observed in Fig. 2.

Moreover, both Figs. 2 and 3 illustrate the behavior of the dynamical properties of the networks built upon shape contours. The calibration of T_l may give rise to sparsely or highly connected networks, as well as other intermediate connectivity patterns. For all the plots presented in these figures we used the comparison *smaller than*, but the same analysis can be performed for the comparison *greater than*. Together with the curvature analysis, we demonstrated the properties of the networks evolved on the topologies defined over the contour pixels.

B. Shape Interpolation Analysis

Fig. 4 presents three interpolation processes that consist of different series of shapes representing the steps performed to fill the space between two geometric shapes. Particularly, the first and the last shapes of the sequence are the reference shapes

Fig. 4-a) presents the interpolation between a circle and a square. The two plots of this figure correspond to the average clustering coefficient $\langle cc \rangle$ and the average path length $\langle l \rangle$, as a function of the threshold parameter, T_l . Each color connects the geometric shape to its corresponding measurements which were obtained for different values of T_l . We can observe in the interpolation process the emergence of internal angles, and, consequently, the curvature values for some pixels start to increase. As largely explored in literature [34, 35], small-world networks are characterized by a high clustering coefficient and a small average path length. As expected, independently of T_l , the value of $\langle cc \rangle$ is constant for the circular shape (light gray curve). Regarding the square (dark blue curve), the values of $\langle cc \rangle$ tend to increase with the threshold T_l . This is observed due to the fact that the mean degree, $\langle k \rangle$, also increases with T_l , therefore, more connections will be established among the contour, and the clustering coefficient will increase.

Similarly, Fig. 4-b) presents the interpolation between a circle and a star. Since the star has acute angles, the values of $\langle cc \rangle$ are higher from $T_l > 0.55$ when compared to the corresponding plot of Fig. 4-a). The interpolation process also introduces some irregularities in the intermediate shapes, therefore, they are not totally uniform as the circle and the square. Consequently, the values of $\langle cc \rangle$ for these irregular shapes tend to be higher. Finally, Fig. 4-c) presents the interpolation between two irregular shapes. A rapid growth in the values of $\langle cc \rangle$ is observed also due to the irregularities of the shapes.

Meanwhile, the analysis of the average path length for the three interpolations indicates no significant differences for what concerns the different thresholds. It is possible to observe that $\langle l \rangle$ is high for thresholds $T_l < 0.1$ and present a fast decay. This can be explained by the fact that for such small thresholds the network is not totally connected and, therefore, the path between two nodes is large or that it may not exist. In order to calculate $\langle l \rangle$, we consider the missing paths as having distance one value higher than the largest possible geodesic path.

We also performed an analysis of the clustering coefficient as a function of the internal angles of different geometric

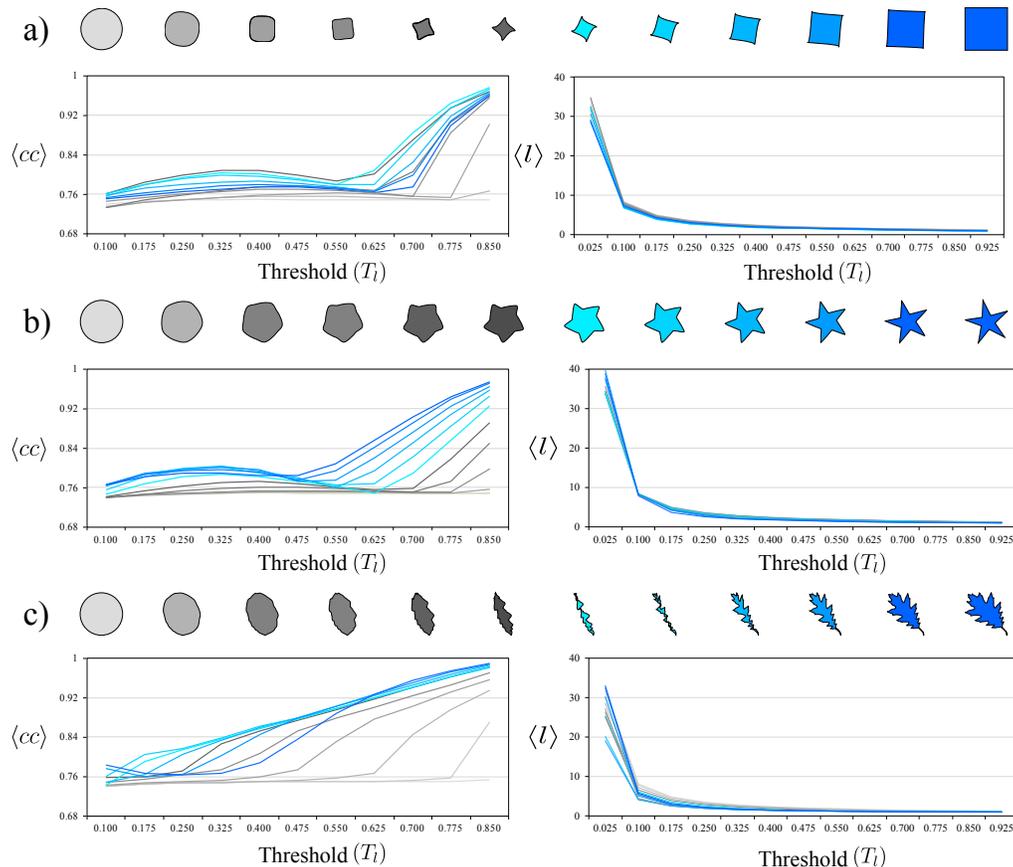


FIG. 4. Interpolation analysis. Each row represents the interpolation between two shapes. The plots in the left side show the clustering coefficient as a function of the threshold parameter, T_i , and, the plots in the right side show the average path length also as a function of T_i . The colors make the correspondence with the shapes.

shapes, which can be seen in Fig. 5. From left to right the internal angles decrease and the respective average clustering coefficients increase rapidly, since the distance between the pixels will decrease and more connections will be established between them.

IV. SHAPE CLASSIFICATION

In this section we present an analysis of the structural network measurements when used as feature vectors for shape classification. We evaluated different image datasets regarding the correct identification of the categories of interest. This evaluation was performed taking into account both the threshold variations as well as the different combinations of structural measurements chosen to compose the feature vector. In addition, we also evaluated the classification accuracy in the presence of noise for what concerns the shape structure and when applying geometric transformations as rotation and scale.

A. Datasets

Three image datasets were used in the experiments performed in this paper. The first dataset, the *geom-shapes* dataset, is composed of 250 geometric shapes of 10 distinct classes: circle, triangle, square, pentagon, hexagon, star, arrow, “pacman”, heart and moon. All these shapes were manually designed by different people. Therefore, the shapes of the same class contain intrinsic variations due to the free-hand design. Also the shapes of this dataset incorporate variations regarding the size and the structure of the geometric objects. Samples of this dataset are shown in Fig. 6-a).

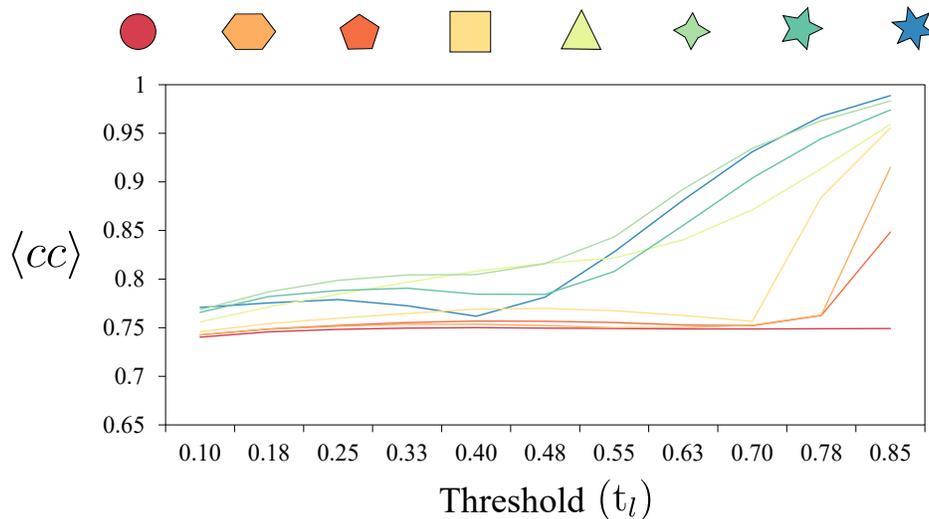


FIG. 5. Average clustering coefficient ($\langle cc \rangle$) as a function of the threshold parameter T_l for eight geometric shapes. Each color corresponds to a specific geometric shape.

The second dataset, the *generic-shapes*, contains generic shapes from different categories, which also present structural variations among instances of the same class. The *generic-shapes* dataset is composed by 99 generic shapes of 9 different classes [39, 40]. This dataset has been used as a benchmark comparison in many computer vision applications. Some examples of shapes belonging to this dataset can be seen in Fig. 6-b).

Finally, the last dataset was yielded by Backes *et al.* [24] and contains real-world leaf contours from distinct plants. The *leaves* dataset is composed by 600 images of leaf contours. There are 600 images from 30 distinct classes of plants. Therefore, there are 20 image samples for each class. The leaf images comprise distinct trees of the Brazilian vegetation. The complete description of this dataset is available in [44]. Fig. 6-c) presents some image samples belonging to the *leaves* dataset. All the three datasets presented in this section have uniform distribution of classes, i.e., all the classes of each dataset have the same number of instances. The composition of each dataset is summarized next.

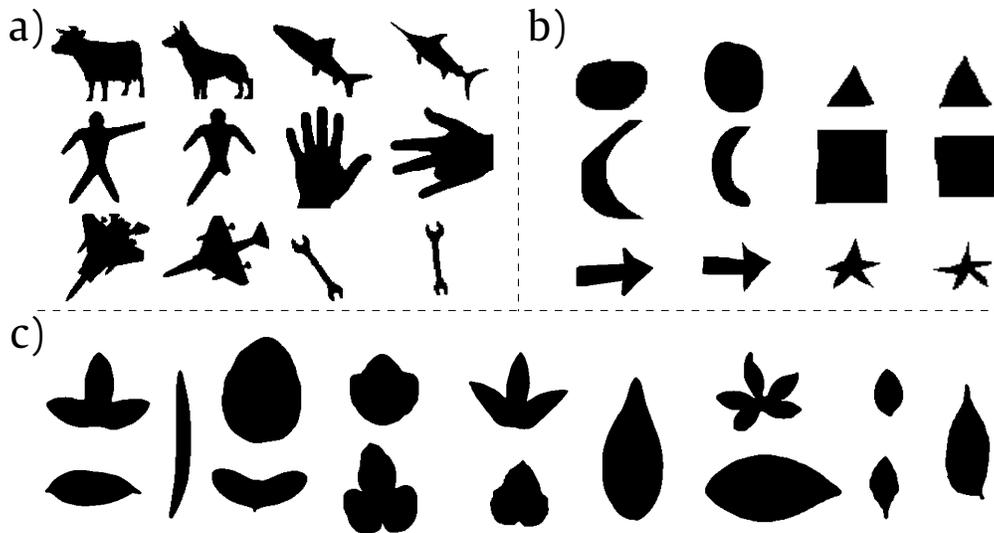


FIG. 6. Sample instances of the three datasets evaluated in this paper. a) *geom-shapes*, b) *generic-shapes* and c) *leaves*.

B. Training and validation strategies

We used n -fold cross-validation to evaluate the performance of the shape classification tasks for the different datasets. Therefore, for each validation, we have different training and validation sets. The cross-validation procedure was applied 100 times for each dataset and the standard deviation was calculated over all these runs since the instances are randomly assigned to each fold. We analyzed the classification performance for three classifiers: k -NN (k -Nearest Neighbors) for $k = 1$, SMO (an optimized version of SVM - *Support Vector Machines*) [45, 46] and Naive Bayes.

The performance of the classifier is given by the number of correctly classified instances of a dataset in relation to the total number of instances, i.e., the accuracy represents the percentage of correctly classified instances. Finally, the measurements chosen to compose the feature vector are: $\langle k \rangle$ (mean degree), $\langle k^h \rangle$ (average hierarchical degree), $\langle cc \rangle$ (clustering coefficient), $\langle l \rangle$ (average path length), ρ (degree correlation) and $\langle b \rangle$ (average betweenness).

C. Performance evaluation and comparison with CNDescriptor

The performance of the proposed shape descriptor was evaluated when applied to shape classification tasks considering the three datasets described in the previous section: *geom-shapes*, *generic-shapes* and *leaves*. In Fig. 7 we can observe the classification performance (%) for each dataset and for each classifier, with the corresponding standard deviation. The different colors, purple and green, represent the two different configuration settings of the descriptor: connections (network links) *smaller than* or *greater than* the threshold T_l . Meanwhile, the different shadings of each color represent the comparison between both network-based descriptors: the generalized (Φ) and the degree specific (φ). For each single plot, the different bar groups on the x -axis, correspond to the number of thresholds resulting from the sampling of the threshold range ($[0, 1]$), i.e., the different number of thresholds, n_T , chosen to compose the feature vector for a specific run of the experiment. For instance, the first bar group corresponds to 13 values of thresholds equally spaced in the range $[0, 1]$. The n_T times the number of network measurements correspond to the final number of attributes. In this experiment, all the network measurements were combined in the same feature vector: $\Phi = [\langle k \rangle(T_0), \langle k^2 \rangle(T_0), \langle k^3 \rangle(T_0), \langle cc \rangle(T_0), \langle l \rangle(T_0), \rho(T_0), \langle b \rangle(T_0), \dots, \langle k \rangle(T_{n_T}), \langle k^2 \rangle(T_{n_T}), \langle k^3 \rangle(T_{n_T}), \langle cc \rangle(T_{n_T}), \langle l \rangle(T_{n_T}), \rho(T_{n_T}), \langle b \rangle(T_{n_T})]$.

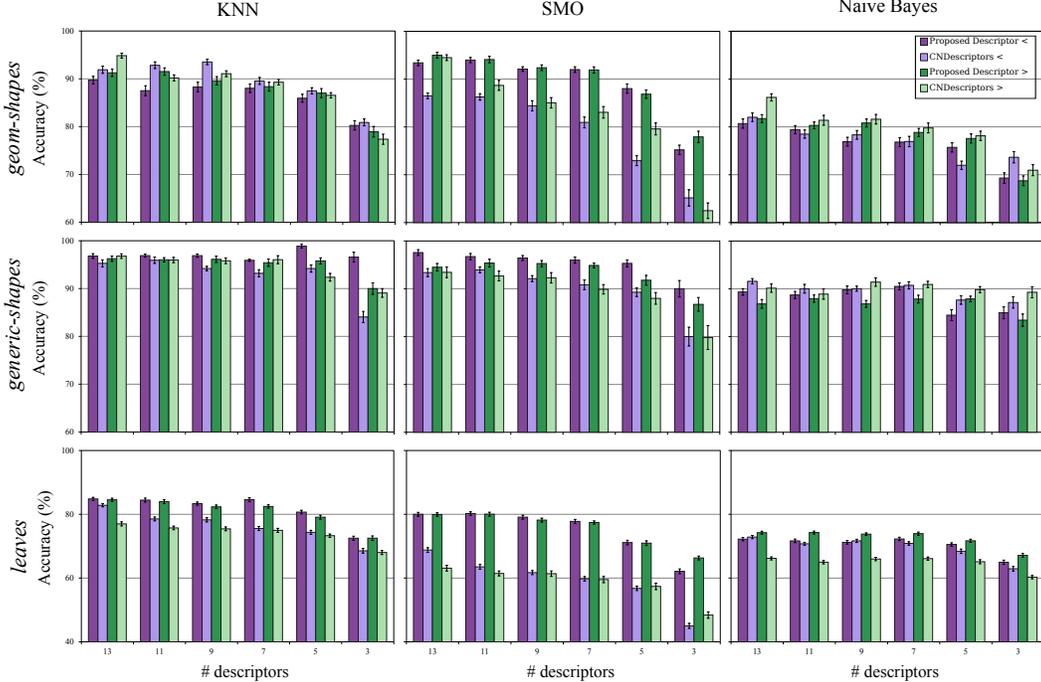


FIG. 7. Classification accuracy (%) and standard deviation obtained for the datasets a) *geom-shapes*, b) *generic-shapes* and c) *leaves*, using the proposed shape descriptor (Φ) and different classifiers.

An analysis of Fig. 7 reveals that the performances of all the classifiers decrease when using a small number of attributes. However, there is a convergence in what concerns the accuracy values when n_T is increased. This result

shows the influence of the number of attributes, and, consequently, the influence of the sampling size of the threshold interval in the final performance. Regarding the comparison between the two configuration settings of the descriptor for the threshold T_i , no significant differences in accuracy were observed. This result leads to the conclusion that both *smaller than* and *greater than* approaches provided similar performances, except in a few cases, like the *barplot* showing the *Naive Bayes* performance for the *geom-shapes* dataset. In contrast, we can observe very similar accuracy values for both settings in the SMO *barplot* for the *leaves* dataset.

For the *geom-shapes* dataset, using a larger set of measurements only improved the classification accuracy for SMO classifier. In this case, the feature vector, Φ , provided a significant increase in the final performance when compared to φ . However, this improvement was not observed for the other two classifiers and, in this case, φ provided better results for many different values of n_T . This can be observed for both comparisons, *smaller than* and *greater than*. Notwithstanding, from an overall analysis of the results for this dataset, it is possible to observe that the difference in performance between the two feature vectors is small and their use is interchangeably in many cases. The same analysis can be applied for the results obtained for the *generic-shapes* dataset, even with a smaller difference in accuracy when comparing both feature vectors. For k-NN classifier, for instance, the performance improvement when using a more robust set of measurements is more evident for smaller values of n_T . This is mainly related to the fact that when there is a greater variability of the measurements that are in the feature vector, the sampling of the threshold interval can be reduced.

For what concerns the *leaves* dataset, there was a considerable improvement in the classification performance when using the proposed feature vector, Φ . Its robustness to the threshold sampling can also be observed, specially for the SMO classifier. This result also accounts for the fact that when using a reduced n_T , a greater variability of measurements in the feature vector is more suitable and computationally more efficient, since the vector dimension is also reduced. A major difference between the *leaves* and the other two datasets is that the first represents real-world shapes. Also, it is a challenging dataset regarding the number of classes (30) and the number of samples per class (20). The *geom-shapes* dataset, for instance, was evaluated mainly as a benchmark comparison. Therefore, in this case it is expected high accuracy values for both feature vectors.

D. Single threshold analysis

A main characteristic of the proposed methodology is the dynamic evolution of the network constructed over the shape contour. This evolution is implemented as a subsequent extraction of measurements from the network topology for each incremental threshold. A small T_i will result in a poorly connected network and, in contrast, a high T_i will provide a highly connected network. Therefore, the classification accuracy is improved with the union of the attributes obtained incrementally. However, it is expected in such a method that the most extreme thresholds will contribute less to the final accuracy than the intermediate thresholds. In the experiments presented in this section, we evaluated the accuracy provided by the different threshold values individually. The result for each dataset and for each configurations of the shape descriptor (*smaller than* and *greater than*) are presented in Fig. 8.

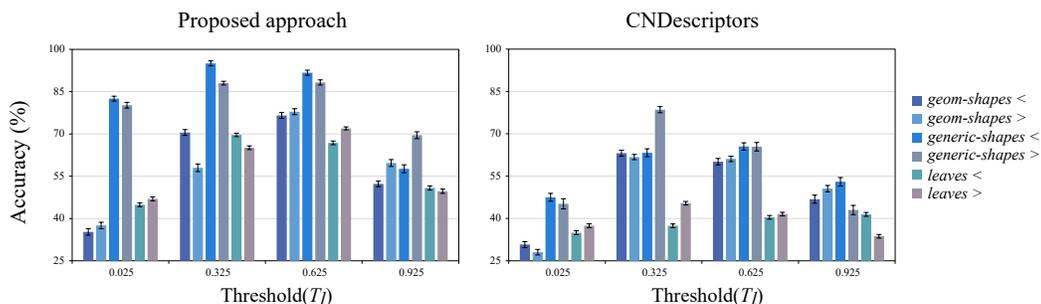


FIG. 8. Classification accuracy (%) and standard deviation obtained for the datasets a) *geom-shapes*, b) *generic-shapes* and c) *leaves*, using the proposed shape descriptor (Φ) for a single threshold value T_i without considering its dynamical evolution.

As expected the extreme thresholds are less accurate in both plots. In spite of that, the performance obtained when using Φ descriptor is much higher than the performance obtained with the φ , for all the datasets. This result corroborates the previous experiments when a small threshold sampling was used. In this case, a large variability regarding the measurements can increase the classification performance. Another result that should be highlighted in this experiment is the fact that for both descriptors the concatenation of network measurements for a sequence of thresholds, taking into account its dynamic evolution, will provide better accuracies than considering only a single

threshold.

E. Robustness & noise tolerance analysis for the *leaves-dataset*

In the work of Backes *et al.* [24], the authors have shown that the CNDescriptor is invariant to scale and rotation transformations and also that it is robust to noise and image degradation. Within this context, the same characteristics were observed using the proposed approach in this paper which generalizes the feature vector regarding the measurements extracted from the network topology (Φ). We evaluated how such measurements as well as the classification accuracy are affected when those transformations are applied to the *leaves* dataset.

The first transformations are rotation and scale. The *leaves-rotated* dataset contains the original images of the *leaves* dataset rotated by the following angles: 7° , 35° , 104° , 132° , 201° and 298° , and, the *leaves-scaled* dataset is composed by the same original images scaled by the factors: 200%, 175%, 150% and 125%. The noise tolerance and the robustness to degradation are other two properties that we investigated for the *leaves* dataset. The noise applied to the original images was uniformly generated in the range $[-n \dots n]$, being n the intensity level of the noise. Given the bi-dimensional contours of the shapes, the noise pattern is applied in the x and y coordinates. The *leaves-noise* dataset contains four different noise values applied to each original image. Finally, the robustness to degradation property was evaluated using the *leaves-degraded* dataset, which contains 17 different levels of degradation which were applied in each leaf image. This dataset presents two types of degradations, continuous and random. In the continuous degradation, only adjacent pixels of the shape contour are eliminated. Meanwhile, in the second type of degradation the pixels are randomly chosen to be eliminated. For both methods, as higher the degradation level is, more pixels will be removed from the contour. Fig. 9 presents some examples of the transformations applied to the *leaves* dataset.

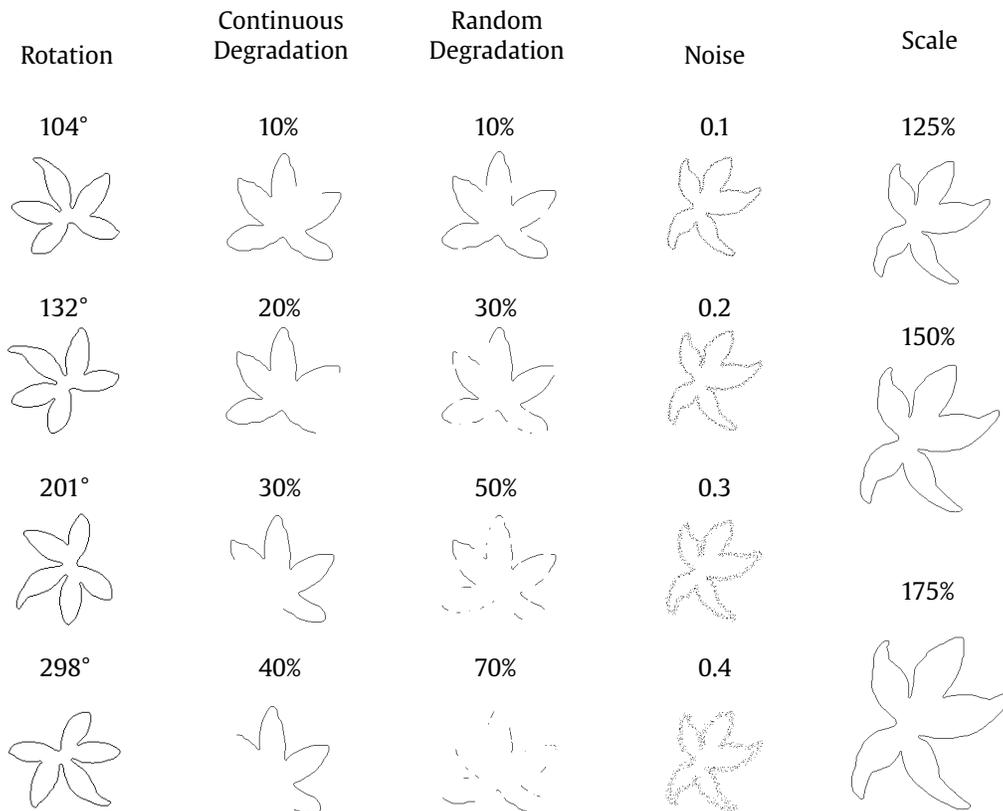


FIG. 9. Transformations applied to the images of the *leaves* dataset (per column): 1) rotation by the angles 104° , 132° , 201° and 298° ; 2) Degradation (continuous and random) on which parts of the contour are removed; 3) Noise, which was uniformly generated, and, 4) Scale by the factors: 125%, 150% and 175%.

Degree

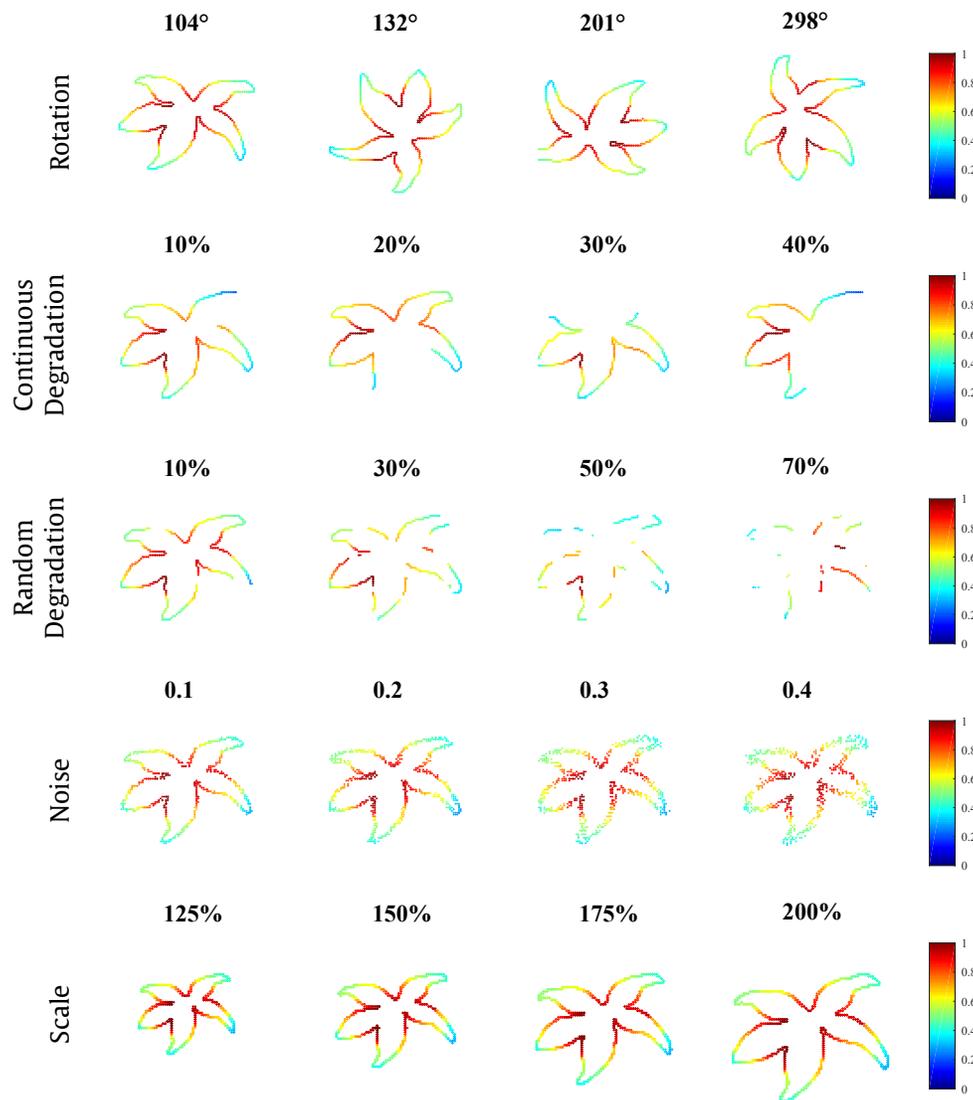


FIG. 10. Color-map representing the degree (k) of the contour pixels using threshold $T_l = 0.325$ for the network construction. Each row illustrates a transformation and its respective parameters.

Clustering

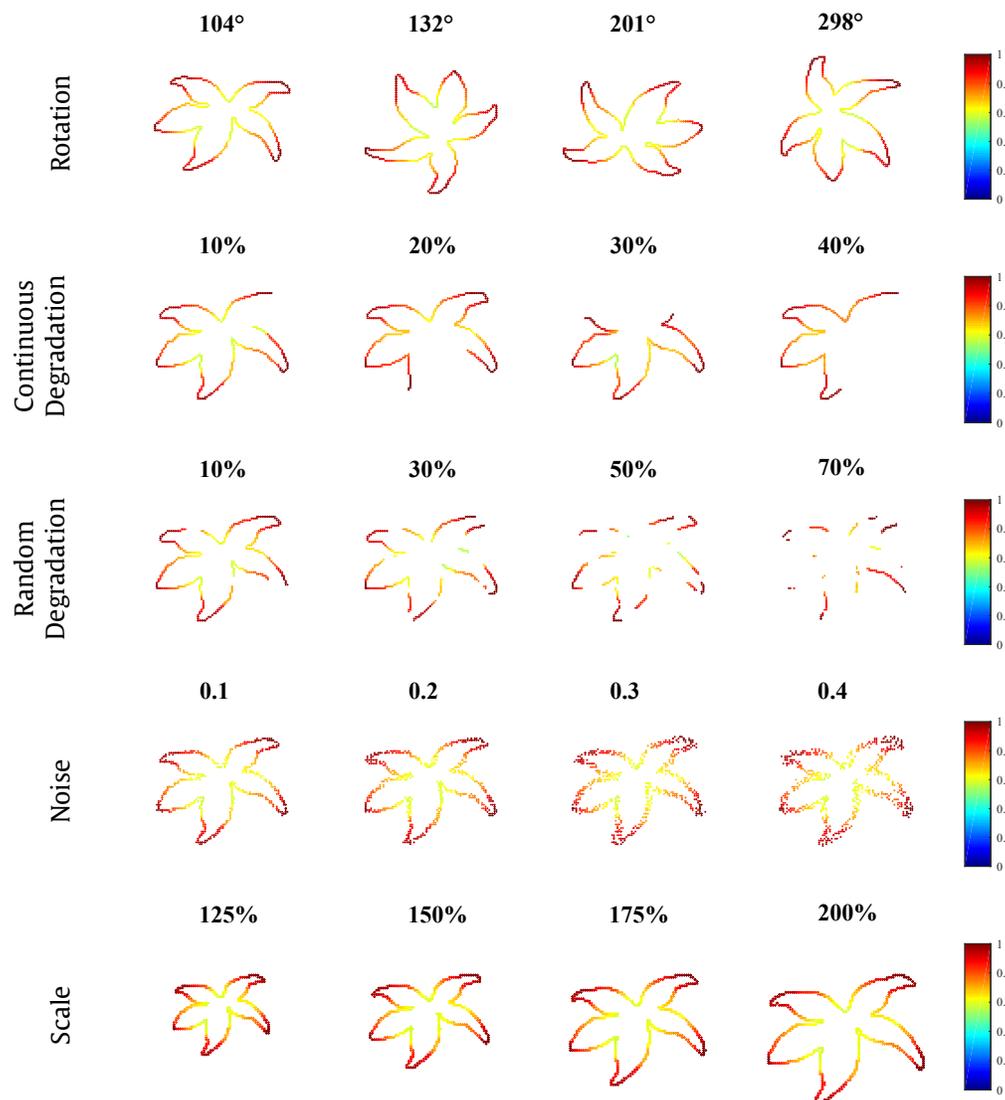


FIG. 11. Color-map representing the clustering coefficient (cc) of the contour pixels using threshold $T_l = 0.325$ for the network construction. Each row illustrates a transformation and its respective parameters.

Betweenness

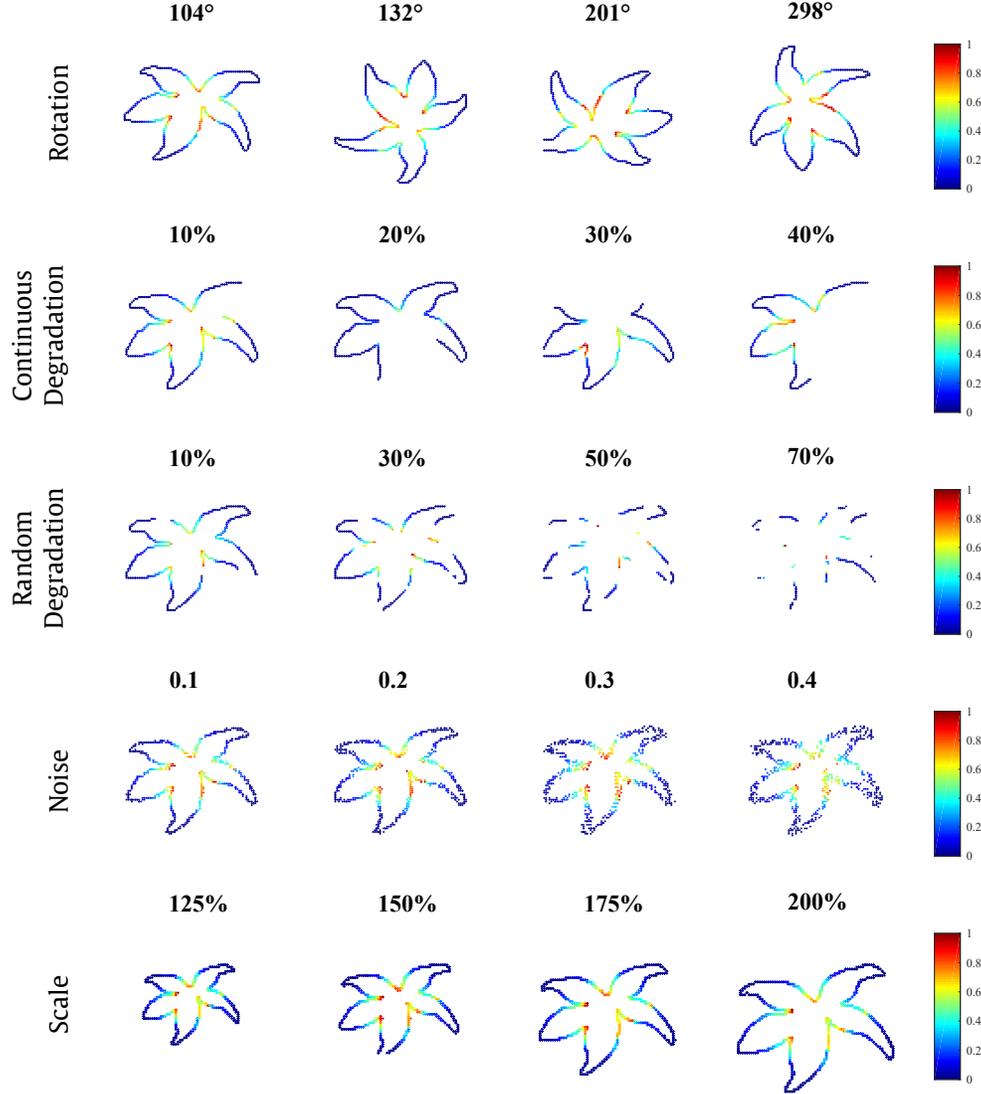


FIG. 12. Color-map representing the betweenness centrality (b) of the contour pixels using threshold $T_l = 0.325$ for the network construction. Each row illustrates a transformation and its respective parameters.

Figs. 10, 11 and 12 present the distribution per network node (per pixel) of the following measurements: degree (k), clustering coefficient (cc) and betweenness (b). In this example, we used the $\delta_{T_l}(W)$ transformation, therefore, two pixels are connected to each other if the euclidean distance between them is *smaller than* $T_l = 0.325$. The five transformations which were applied to the *leaves* dataset are illustrated in each row. For what concerns the degree (Fig. 10), we can see that the pixels that are closer to the leaf center are the most connected ones. In the case of the degraded contours, there are missing parts in the leaf folds, and, consequently, the degree of the pixels near those parts are smaller.

For the clustering coefficient (Fig. 11), we can observe that it is higher near the leaf tips. This can be explained by the fact that, at this threshold ($T_l = 0.325$), there are many connected triples among the pixels of the tips, which increases the cc value. For the pixels near the leaf center there is also a high number of connected triples, however, their degree is also higher, which increases the number of possible connections among the neighbors of a pixel, but they are not always connected as triples. Consequently, the cc value for those pixels is lower. Even for the degraded contours, a similar distribution pattern of the clustering coefficient can be observed. Yet, it should be noticed that for all the examples in Fig. 11, the clustering coefficient is already high, above 0.5.

Similarly to the degree color-map, we can observe that the betweenness centrality (Fig. 12) is also higher for the

pixels near the leaf center, which are between the leaf tips. The pixels in that region belong to many pathways of the contour network, therefore, they connect pixels of different leaf tips. Consequently, their betweenness is also higher.

In addition, we also performed an experiment to test the classification accuracy for what concerns the different variations of the *leaves* dataset. In this experiment, each transformation (rotation, scale, noise and the two degradation types) was evaluated separately. The classification accuracy was calculated taking into account the 30 classes of the *leaves* dataset. The threshold interval was sampled in 13 equally spaced values of T_l . Table I presents the results.

TABLE I. Classification accuracy (%) for the datasets *leaves-rotated*, *leaves-scaled* and *leaves-noise*.

	Scale Factor	k-NN	SMO	NB
Scale	1.25	86.55 ± 0.51	82.52 ± 0.52	73.64 ± 0.54
	1.50	87.10 ± 0.44	83.44 ± 0.56	74.50 ± 0.54
	1.75	86.93 ± 0.49	83.37 ± 0.61	75.18 ± 0.46
	2.00	87.67 ± 0.53	83.92 ± 0.58	74.64 ± 0.54
	Degree	k-NN	SMO	NB
Rotation	7°	82.84 ± 0.61	80.65 ± 0.58	72.06 ± 0.41
	35°	82.06 ± 0.60	77.90 ± 0.53	72.07 ± 0.48
	104°	80.40 ± 0.56	79.12 ± 0.66	72.58 ± 0.55
	132°	85.37 ± 0.53	78.59 ± 0.61	72.25 ± 0.48
	201°	81.21 ± 0.54	76.80 ± 0.55	71.96 ± 0.51
	298°	81.85 ± 0.59	77.04 ± 0.59	72.28 ± 0.49
	Noise level	k-NN	SMO	NB
Noise	1	77.08 ± 0.59	76.25 ± 0.68	68.30 ± 0.56
	2	75.82 ± 0.68	78.43 ± 0.47	68.35 ± 0.56
	3	75.42 ± 0.59	77.36 ± 0.59	67.00 ± 0.67
	4	76.54 ± 0.58	76.84 ± 0.64	67.87 ± 0.61
Original		k-NN	SMO	NB
		84.81 ± 0.51	79.99 ± 0.62	72.22 ± 0.52

For what concerns rotation, the resulting network for a given threshold T_l will be the same before and after the application of the transformation. The rotation does not affect the adjacency matrix of the network, and, consequently, the descriptor will be preserved. The differences observed for the accuracy values given the different rotation degrees (Table I) are due to the random sampling of the folds during the cross-validation. The descriptor is also not affected by linear transformations such as scale. Therefore, we can observe in Table I that the performance for rotation and scale transformations are similar to the performance obtained for the original dataset (*leaves*). In the case of the scale transformation, the accuracy improved for all the classifiers, using the same feature vector Φ . Therefore, the interpolation process associated with the scale transformation provided a better separation of the classes for the *leaves* dataset. Yet, the measurements extracted from the network topology were more discriminative after the application of the scale transformation. Regarding the robustness to noise, we can see that the accuracy for the *leaves-noise* dataset is lower when compared with the other two datasets presented in the this table. The decrease in accuracy was the highest when compared to the other datasets. However, the accuracy is not much affected by the different levels of noise.

Fig 13 presents the classification accuracy (%) as a function of the different degradation levels for the *leaves-degraded* dataset. As the degradation level increases, the classification performance decreases. As expected, when parts of the shape boundaries are removed, the correct identification of the classes becomes more difficult.

V. CONCLUSION

We presented in this paper an approach for shape analysis aiming at pattern recognition applications. This approach extends the method of Backes *et al.* [24] in what concerns the structural analysis performed over the network obtained through shape boundaries. Therefore, we went a step further performing a study of the relationship between the shape architecture and the network topology. This was first investigated for regular polygons and then for other irregular and generic shapes. Through the use of the curvature analysis, it was possible to investigate how the network measurements vary according to some specific shape properties, and, consequently, how the network descriptor, Φ is constructed. In addition, we demonstrated the effect of the threshold sampling in the classification accuracy regarding the dynamical construction of Φ (Fig. 7), and, also, the accuracy provided by each threshold individually (Fig. 8).

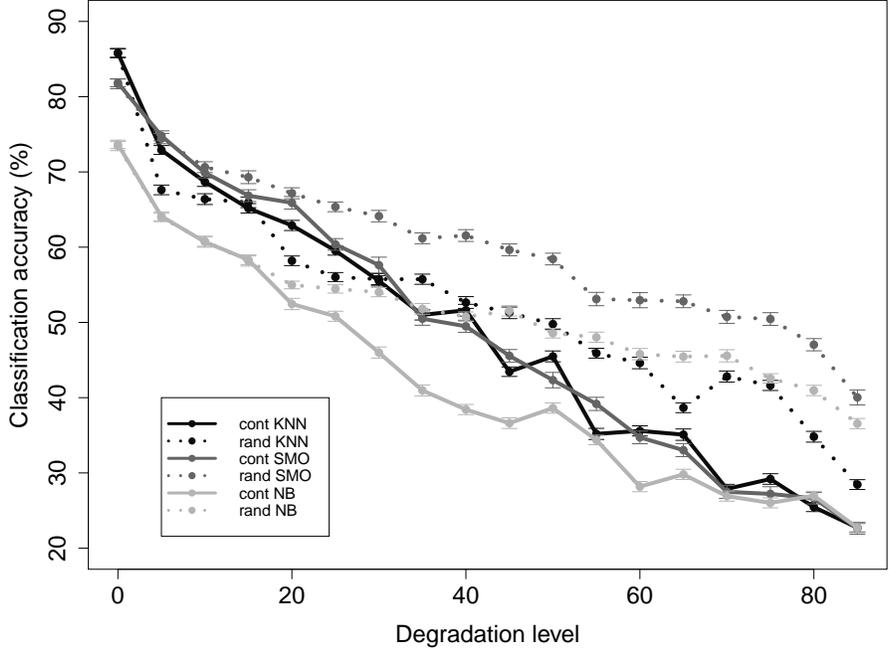


FIG. 13. Classification accuracy as a function of the degradation levels for the *leaves-degraded* dataset.

Through the use of a more robust set of structural measurements we could improve the classification performance in different pattern recognition problems. The *geom-shapes* and the *generic-shapes* datasets represent applications that can be used as a benchmark comparison. We have shown that, for both datasets, the proposed shape descriptor can improve the classification performance when compared to the CNDescriptor for the different comparisons considered in the experiments (*smaller-than* and *greater-than*), and, also for different numbers of thresholds and features. The increase observed in the classification accuracy also depends on the classifier. The SMO classifier provided the highest improvement regarding the classification performance (%). In spite of that, the NB classifier achieved a slightly better performance for those datasets when using CNDescriptor. For what concerns the *leaves* dataset, there was a considerable increase in the classification accuracy for all classifiers and for both comparisons when using the proposed feature vector, Φ . It should also be noticed that when using a smaller number of descriptors for this dataset the increase in classification performance is even higher for Φ than for the CNDescriptor. We consider *leaves* a more challenging dataset both because the number of classes (30) is higher than the number of samples per class (20), and, also because because of the intrinsic nature of the samples, which are part of a real-world application. In addition, Backes *et al.* [24] performed a comparison of the accuracy obtained with CNDescriptor with classical methods found in literature for shape analysis, e.g., Fourier descriptors, Zernike moments, curvature descriptors and others. They showed that their method could surpass these classical approaches for what concerns the classification accuracy (%). In this paper we have shown that a more robust set of network measurements could even surpass the CNDescriptor for the *leaves* dataset as well as in many comparisons performed for the other two datasets.

Finally, one important contribution of this paper is related to the construction of the feature vector from the network obtained from shape boundaries. We have shown that not only degree related measurements are capable of distinguishing categories associated with a specific problem. In addition, we also demonstrated how other classes of measurements can be used together in pattern recognition applications. Yet, when using a set of measurements that accounts for different structural properties of networks, the number of thresholds can be reduced, and, consequently, the construction of the shape descriptor becomes more computationally efficient. Therefore, there is a balance between the amount of measurements and thresholds that should be selected to compose the feature vector.

ACKNOWLEDGEMENTS

G.H.B.M. and O.M.B. are grateful for the support from São Paulo Research Foundation (FAPESP) with grant #2015/05899-7. G.H.B.M. and J.M. are grateful for the support of the Coordination for the Improvement of Higher Education Personnel (CAPES). O.M.B. gratefully acknowledges the financial support of National Council for Scientific and Technological Development (CNPq), grants #307797/2014-7 and #484312/2013-8, and, FAPESP with grant #2014/08026-1.

REFERENCES

-
- [1] G. Gerig, M. Styner, M. E. Shenton, J. A. Lieberman, Shape versus size: Improved understanding of the morphology of brain structures, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 24–32.
 - [2] R. d. S. Torres, A. X. Falcao, L. d. F. Costa, A graph-based approach for multiscale shape analysis, *Pattern Recognition* 37 (2004) 1163–1174.
 - [3] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, C. Davatzikos, Compare: classification of morphological patterns using adaptive regional elements, *IEEE transactions on medical imaging* 26 (2007) 93–105.
 - [4] M. Aitkenhead, I. Dalgetty, C. Mullins, A. J. S. McDonald, N. J. C. Strachan, Weed and crop discrimination using image analysis and artificial intelligence methods, *Computers and electronics in Agriculture* 39 (2003) 157–171.
 - [5] J. C. Neto, G. E. Meyer, D. D. Jones, A. K. Samal, Plant species identification using elliptic fourier leaf shape analysis, *Computers and electronics in agriculture* 50 (2006) 121–134.
 - [6] T. Brosnan, D.-W. Sun, Inspection and grading of agricultural and food products by computer vision systems a review, *Computers and electronics in agriculture* 36 (2002) 193–213.
 - [7] C. Costa, F. Antonucci, F. Pallottino, J. Aguzzi, D.-W. Sun, P. Menesatti, Shape analysis of agricultural products: a review of recent research advances and potential application to computer vision, *Food and Bioprocess Technology* 4 (2011) 673–692.
 - [8] J. Hemming, T. Rath, Paprecision agriculture: Computer-vision-based weed identification under field conditions using controlled lighting, *Journal of agricultural engineering research* 78 (2001) 233–243.
 - [9] R. d. O. Plotze, M. Falvo, J. G. Pádua, L. C. Bernacci, M. L. C. Vieira, G. C. X. Oliveira, O. M. Bruno, Leaf shape analysis using the multiscale minkowski fractal dimension, a new morphometric method: a study with passiflora (passifloraceae), *Canadian Journal of Botany* 83 (2005) 287–301.
 - [10] R. M. Rangayyan, N. R. Mudigonda, J. L. Desautels, Boundary modelling and shape analysis methods for classification of mammographic masses, *Medical and Biological Engineering and Computing* 38 (2000) 487–496.
 - [11] T. Heimann, H.-P. Meinzer, Statistical shape models for 3d medical image segmentation: a review, *Medical image analysis* 13 (2009) 543–563.
 - [12] D. Selle, B. Preim, A. Schenk, H.-O. Peitgen, Analysis of vasculature for liver surgical planning, *IEEE transactions on medical imaging* 21 (2002) 1344–1357.
 - [13] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, A. Willsky, A shape-based approach to the segmentation of medical imagery using level sets, *IEEE transactions on medical imaging* 22 (2003) 137–154.
 - [14] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, M. Heynen, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for gis-ready information, *ISPRS Journal of photogrammetry and remote sensing* 58 (2004) 239–258.
 - [15] T. Blaschke, Object based image analysis for remote sensing, *ISPRS journal of photogrammetry and remote sensing* 65 (2010) 2–16.
 - [16] S. Osowski, et al., Fourier and wavelet descriptors for shape recognition using neural networks a comparative study, *Pattern Recognition* 35 (2002) 1949–1957.
 - [17] D. Zhang, G. Lu, et al., A comparative study on shape retrieval using fourier descriptors with different shape signatures, in: *Proc. of international conference on intelligent multimedia and distance education (ICIMADE01)*, pp. 1–9.
 - [18] C. Davatzikos, X. Tao, D. Shen, Hierarchical active shape models, using the wavelet transform, *IEEE transactions on medical imaging* 22 (2003) 414–423.
 - [19] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, M. J. Cree, Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification, *IEEE Transactions on medical Imaging* 25 (2006) 1214–1222.
 - [20] A. R. Backes, O. M. Bruno, Shape classification using complex network and multi-scale fractal dimension, *Pattern Recognition Letters* 31 (2010) 44–51.
 - [21] A. R. Backes, J. B. Florindo, O. M. Bruno, Shape analysis using fractal dimension: A curvature based approach, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22 (2012) 043103.
 - [22] O. M. Bruno, R. de Oliveira Plotze, M. Falvo, M. de Castro, Fractal dimension applied to plant identification, *Information Sciences* 178 (2008) 2722–2733.

- [23] F. Mokhtarian, M. Bober, Curvature scale space representation: theory, applications, and MPEG-7 standardization, volume 25, Springer Science & Business Media, 2013.
- [24] A. R. Backes, D. Casanova, O. M. Bruno, A complex network-based approach for boundary shape analysis, *Pattern Recognition* 42 (2009) 54–67.
- [25] A. R. Backes, O. M. Bruno, Polygonal approximation of digital planar curves through vertex betweenness, *Information Sciences* 222 (2013) 795–804.
- [26] Z. Wu, X. Lu, Y. Deng, Image edge detection based on local dimension: A complex networks approach, *Physica A: Statistical Mechanics and its Applications* 440 (2015) 9–18.
- [27] Y. Sui, F. Shao, C. Wang, R. Sun, J. Ji, Complex network modeling of spectral remotely sensed imagery: A case study of massive green algae blooms detection based on modis data, *Physica A: Statistical Mechanics and its Applications* 464 (2016) 138–148.
- [28] A. R. Backes, D. Casanova, O. M. Bruno, Texture analysis and classification: A complex network-based approach, *Information Sciences* 219 (2013) 168–180.
- [29] M. E. Newman, The structure and function of complex networks, *SIAM review* 45 (2003) 167–256.
- [30] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, *Physics reports* 424 (2006) 175–308.
- [31] M. Newman, A.-L. Barabasi, D. J. Watts, The structure and dynamics of networks, Princeton University Press, 2011.
- [32] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, *science* 286 (1999) 509–512.
- [33] A.-L. Barabási, R. Albert, H. Jeong, Scale-free characteristics of random networks: the topology of the world-wide web, *Physica A: Statistical Mechanics and its Applications* 281 (2000) 69–77.
- [34] D. J. Watts, S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *nature* 393 (1998) 440–442.
- [35] L. d. F. Costa, F. A. Rodrigues, G. Travieso, P. R. Villas Boas, Characterization of complex networks: A survey of measurements, *Advances in physics* 56 (2007) 167–242.
- [36] M. E. Newman, A measure of betweenness centrality based on random walks, *Social networks* 27 (2005) 39–54.
- [37] S. P. Borgatti, Centrality and network flow, *Social networks* 27 (2005) 55–71.
- [38] L. da F Costa, P. V. Boas, F. Silva, F. Rodrigues, A pattern recognition approach to complex networks, *Journal of Statistical Mechanics: Theory and Experiment* 2010 (2010) P11015.
- [39] D. Sharvit, J. Chan, H. Tek, B. B. Kimia, Symmetry-based indexing of image databases, in: *Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on*, IEEE, pp. 56–62.
- [40] T. B. Sebastian, P. N. Klein, B. B. Kimia, Recognition of shapes by editing their shock graphs, *IEEE Transactions on pattern analysis and machine intelligence* 26 (2004) 550–571.
- [41] M. E. Newman, Assortative mixing in networks, *Physical review letters* 89 (2002) 208701.
- [42] F. Attneave, Some informational aspects of visual perception., *Psychological review* 61 (1954) 183.
- [43] R. M. Cesar, L. D. F. Costa, Towards effective planar shape representation with multiscale digital curvature analysis based on signal processing techniques, *Pattern Recognition* 29 (1996) 1559–1569.
- [44] Leaves dataset, <http://scg.ifsc.usp.br/dataset/ShapeCN.php>, 2009. [Online; accessed 30-October-2017].
- [45] J. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Schoelkopf, C. Burges, A. Smola (Eds.), *Advances in Kernel Methods - Support Vector Learning*, MIT Press, 1998.
- [46] S. Keerthi, S. Shevade, C. Bhattacharyya, K. Murthy, Improvements to platt’s smo algorithm for svm classifier design, *Neural Computation* 13 (2001) 637–649.