Expert Systems With Applications 63 (2016) 375–385

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Evolutionary optimization of a multiscale descriptor for leaf shape analysis

Marcelo M.S. de Souza[a,*], Fátima N.S. Medeiros[a], Geraldo L.B. Ramalho[b], Iális C. de Paula Jr.[c], Isaura N.S. Oliveira[d]

[a] Departamento de Engenharia de Teleinformática, Universidade Federal do Ceará, Campus do Pici s/n, Bloco 725, 60455-970, Fortaleza, CE, Brazil
[b] Curso de Tecnologia em Manutenção Industrial, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Av. Parque Central s/n, Distrito Industrial I, 61939-140, Maracanaú, CE, Brazil
[c] Curso de Engenharia da Computação, Universidade Federal do Ceará, Campus de Sobral, Bloco I - Engenharias, Rua Estanislau Frota s/n, Mucambinho, 62010-560, Sobral, CE, Brazil
[d] Curso de Engenharia Elétrica, Universidade Federal do Ceará, Campus de Sobral, Bloco I - Engenharias, Rua Estanislau Frota s/n, Mucambinho, 62010-560, Sobral, CE, Brazil

ARTICLE INFO

Article history:
Received 23 January 2016
Revised 10 July 2016
Accepted 11 July 2016
Available online 12 July 2016

Keywords:
Shape analysis
Image processing
Data visualization
Evolutionary optimization
Leaf taxonomy

ABSTRACT

Shape analysis and recognition play an important role in the design of robust and reliable computer vision systems. Such systems rely on feature extraction to provide meaningful information and representation of shapes and images. However, accurate feature extraction is not a trivial task since it may depend on parameter adjustment, application domain and the shape data set itself. Indeed, there is a demand for computational tools to understand and support parameter adjustment and therefore unfold shape description representation, since manual parameter choices may not be suitable for real applications. Our major contribution is the definition of an evolutionary optimization methodology that fully supports parameter adjustment of a multiscale shape descriptor for feature extraction and representation of leaf shapes in a high dimensional space. Here, intelligent evolutionary optimization methods search for parameters that best fit the normalized multiscale bending energy descriptor for leaf shape retrieval and classification. The simulated annealing, differential evolution and particle swarm optimization methods optimize an objective function, which is based on the *silhouette* measure, to achieve the set of optimal parameters. Our methodology improves leaf shape characterization and recognition due to the intrinsic shape differences which are embedded in the set of optimized parameters. Experiments were conducted on public benchmark data sets with the normalized multiscale bending energy and inner-distance shape context descriptors. The visual exploratory data analysis techniques showed that the proposed methodology minimized the total within-cluster variance and thus, improved the leaf shape clustering. Moreover, supervised and unsupervised classification experiments with plant leaves accomplished high Precision and Recall rates as well as Bulls-eye scores with the optimized parameters.

1. Introduction

In computer vision, shape description consists of extracting feature for recognition, classification and exploratory data analysis. Feature extraction is widely applied to shape analysis and image processing and furthermore, it has been successfully used to solve a wide range of challenges in information science, biology, taxonomy, medicine and engineering (Anuar, Setchi, & Lai, 2013; Backhaus et al., 2010; Cesar Jr. & Costa, 1997; Chaki, Parekh, & Bhattacharya, 2015; Direkoglu & Nixon, 2011; Mokhtarian & Mackworth, 1986; Rossatto, Casanova, Kolb, & Bruno, 2011; Wang, Brown, Gao, & La Salle, 2015). Identification of plant species or leaf groups within a population is a challenging task since leaves of the same specie tend to subtly differ in size and shape due to the phenotypic plasticity. In plant taxonomy, leaf dimensions and shape are degrees of freedom by which plants adapt to their environment. Thus, a large range of variation can be found, even considering a given genotype and leaf position, depending on climate and resources (Dornbusch, Jillian, Baccar, Fournier, & Andrieu, 2011). Historically, sample specimens are stored in a herbarium

* Corresponding author. Fax: +55-8533669468.
E-mail addresses: marcelo@ufc.br, marcelo.mssouza@gmail.com (M.M.S. de Souza), fsombra@ufc.br (F.N.S. Medeiros), gramalho@ifce.edu.br (G.L.B. Ramalho), ialis@ufc.br (I.C. de Paula Jr.), isauso@gmail.com (I.N.S. Oliveira).

http://dx.doi.org/10.1016/j.eswa.2016.07.016
0957-4174/

for further identification, but with the advent of digital image technology it is now possible to create and publish leaf image data sets. Therefore, there is an increasing interest in deploying automated computational systems, which support taxonomists, who examine biological specimens, in the process of leaf-based plant species identification (Chaki et al., 2015; Cope, Corney, Clark, Remagnino, & Wilkin, 2012; Ghasab, Khamis, Mohammad, & Fariman, 2015; Wang et al., 2015).

A particularly important problem in shape analysis is to fit a shape descriptor to the application domain. In fact, this task comprises parameter adjustment of the shape descriptor and investigation about its structure and how it matches information knowledge, skill and perception of specialists.

Normally, parameter setting is an empirical task (Ling & Jacobs, 2007; Mokhtarian & Suomela, 1998; Wang, Bai, You, Liu, & Latecki, 2012). Paula Jr., Medeiros, Bezerra, and Ushizima (2013) have applied a brute force optimization scheme to effectively addressing parameter adjustment and achieved better results when adjusted shape descriptor parameters according to the database. In general, the design of shape descriptors or shape matching algorithms assigns parameter values regardless of the database. Alternatively, Bai, Yang, Latecki, Liu, and Tu (2010) proposed a method that computes image descriptors to regulate learned shape similarities. Thus, the learning context-sensitive shape similarity applies a supervised learning framework to search for the best learned parameters to calculate the affinity matrix (Bai et al., 2010).

An interesting approach to unfold the structure of a shape descriptor is to assess its performance in a subset of a data set due to the inherent complexity of the high dimensional representation that the descriptor provides. Data visualization techniques, which apply projection-based data, also known as manifold learning, are powerful tools to support this sort of assessment. Such tools provide a 2D representation of multidimensional data, which yields an informative panorama about its organization (Amorim et al., 2015). Moreover, these tools promote data exploration and support specialists to extract relevant data information.

Automatic parameter adjustment of a descriptor can be addressed as an optimization problem, where computational intelligence (CI) (Engelbrecht, 2007), a sub-branch of artificial intelligence, can be used. Studies in computational intelligence encompass bio-inspired algorithms, which present intelligent behavior to solve complex problems. In computer science, evolutionary computation (EC) consists of algorithms that mimic natural evolution processes of populations. The essence of an evolutionary approach to solve a problem is to equate possible solutions to individuals in a population, and to introduce a notion of fitness on the basis of solution quality (Eiben & Smith, 2015).

This paper introduces an optimization methodology to achieve parameter adjustment of the normalized multiscale bending energy (NMBE) (Cesar Jr. & Costa, 1997) tailored to leaf shape analysis. Our approach searches for the best set of descriptor parameters, which encompasses the intrinsic shape differences within a database, to perform leaf shape supervised classification and content-based image retrieval experiments. The main contribution of this work is to provide an optimization methodology for parameter adjustment of a multiscale shape descriptor to characterize plant leaves, which can also be adapted to deal with different objective functions, data sets and applications.

The remainder of this paper is organized as follows. The next section introduces materials and methods to support the proposed methodology that optimizes a multiscale shape descriptor. It also describes the evaluation methodology which comprises the self-organization map and multidimensional scaling. Section 3 presents the public Flavia leaf data set and discusses the classification and shape retrieval results as well as the visual exploratory cluster analysis. It also evaluates the computational cost of the proposed methodology. In Section 4, we draw conclusions and summarize our main contributions.

2. Materials and methods

The proposed methodology for shape descriptor optimization follows the schemes depicted in Fig. 1. The goal of the optimization procedure is to improve shape description and hence its ability to identify hidden and subtle shape features. In fact, what guides this optimization procedure is the minimization of the objective function *median* absolute deviation error (MAD) of the *silhouette* measure (Rousseeuw, 1987). We have chosen this objective function due to its robustness to outliers (Rousseeuw & Leroy, 1987) and, it is given by

$$\mathrm{MAD} = median\left(|s_i - 1|_{i=1,\,2,\,\cdots,\,L}\right), \tag{1}$$

where the set $S = \{s_1, s_2, \cdots, s_L\}$ comprises the computed *silhouette* values for L shape descriptors. The operators |.| and *median*() return the absolute value of an argument and the median of a set of values, respectively.

Silhouette (Rousseeuw, 1987) is a cluster quality measure that indicates the affinity degree of an object i with a particular class A, taking into account the average within-class and inter-class distances of this object to the others. This measure is defined as

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \in [-1, 1], \tag{2}$$

where a_i is the mean dissimilarity between the object i and the objects belonging to the same class A and b_i is the mean dissimilarity of the object i to the neighbor class closest to i, excluding its own class. *Silhouette* may assume values in the interval $[-1, 1]$, and furthermore, a negative value indicates that an object potentially belongs to its neighbor class. Differently, a positive value denotes that it belongs to the estimated class. A value close to zero indicates that the object is close to a boundary between two classes and hence there is uncertainty about which class it belongs to. The objective function (MAD) values lie within the interval [0, 2]. Accordingly, a value equal to zero for this function indicates a perfect cluster structure, while a MAD value close to 2 indicates a cluster structure with low intra-class similarity and high inter-class similarity.

The proposed methodology adjusts the parameters of the multiscale shape descriptor following the process in Fig. 1. The first step randomly selects a subset of sample leaf shapes from Flavia leaf data set and then it runs the optimization methodology to find the best set of scale parameters $\boldsymbol{\sigma}_{best} = (\sigma_1, \sigma_2, \cdots, \sigma_k)$ of the multiscale descriptor that minimizes the MAD function (Eq. 1). Finally, the set of feature vectors is evaluated qualitatively and quantitatively by applying manifold learning algorithms and performing supervised classification and content based image retrieval experiments.

The qualitative evaluation applies two manifold learning algorithms, namely the Kohonen clustering algorithm (Kohonen, Schroeder, & Huang, 2001) and the multidimensional scaling (MDS) (Cox & Cox, 2000). These algorithms generate 2D projections of the leaf description and display these projections in shape similarity maps. Actually, these maps provide graphical representations that support cluster analysis, since they illustrate how the multiscale descriptor spatially organizes leaf shapes in a 2D space.

The quantitative evaluation methodology of the cluster quality relies on the mean *silhouette* per shape class as a measure of the leaf cluster cohesion (Rousseeuw, 1987). Finally, Precision and Recall rates assess the performance of the optimized multiscale descriptor in supervised classification experiments, where the Naive Bayes (NB) (Fukunaga, 1990), K-nearest neighbors (Knn, $K = 5$) (Fukunaga, 1990; Webb, 2002), Fisher linear discriminant analysis

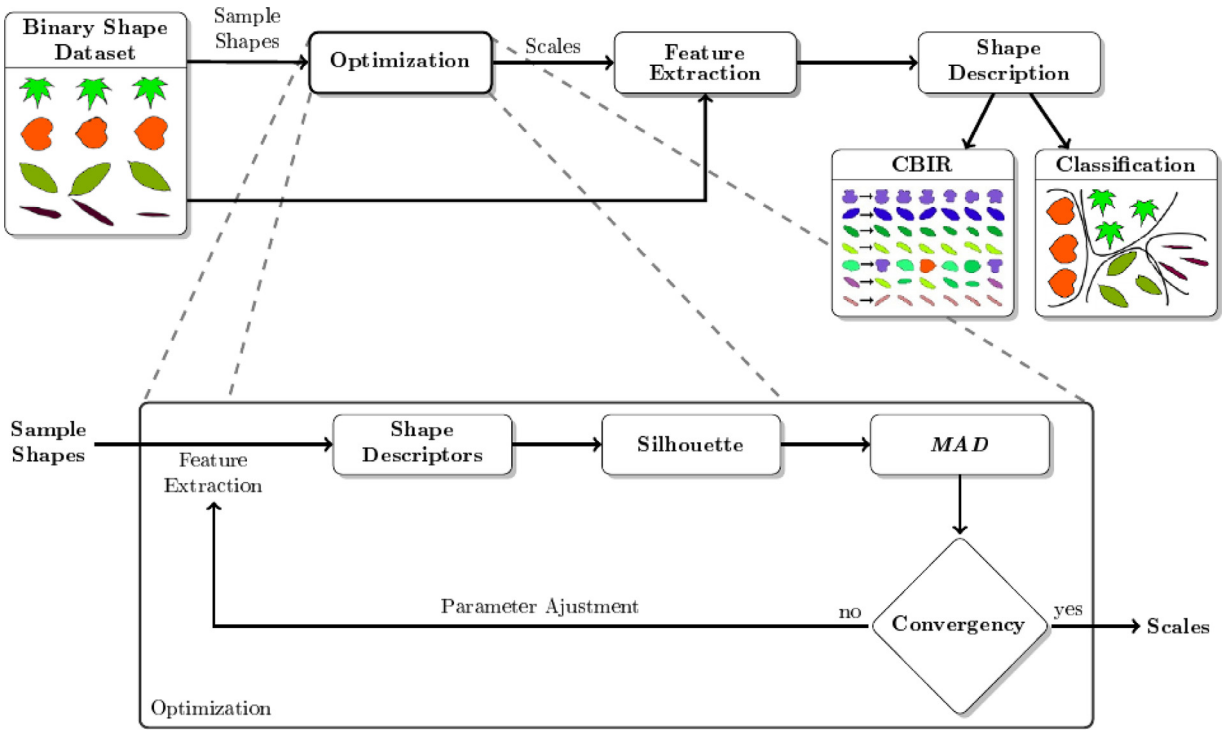

Fig. 1. The proposed methodology for evolutionary optimization of a multiscale shape descriptor.

(LDA) (Webb, 2002) and quadratic discriminant (QDA) (Fukunaga, 1990) classifiers were employed. Prior to the classification experiments, we transformed data with Fisher linear discriminant (Webb, 2002) to reduce data dimension and then, using the principal component analysis (PCA) (Webb, 2002) to reduce the correlation among feature vectors.

It is worth noting that both qualitative and quantitative evaluation methodologies rely on a similarity measure to compare shape feature vectors. Regarding NMBE, it is invariant to shape scale, translation and rotation (Cesar Jr. & Costa, 1997) and hence the similarity analysis performs element-wise correspondence between the feature vectors. Thus, the L_2 norm is appropriate to compute similarity among leaf shapes.

2.1. The normalized multiscale bending energy

The multiscale approach for feature extraction derives from the scale-space (Mokhtarian & Mackworth, 1986) and image pyramid structure (Gonzalez & Woods, 2006) theories. This approach extracts image features at various levels of details by capturing global and local features from low- to high-resolution scales. Indeed, multiscale features often produce a more reliable shape representation than monoscale features (Mokhtarian & Mackworth, 1986).

Cesar Jr. & Costa (1997) introduced the multiscale bending energy for morphometric analysis of neuron cells. Here, we revisit this shape description method and investigate whether or not it is suitable for plant leaf identification systems since leaf image classification is a relevant task for botanists.

Fig. 2a illustrates how NMBE discriminates three distinct classes of leaves from Flavia data set. Similarly, Fig. 2b also exemplifies the versatility of NMBE to discriminate shapes of two different classes from the public Kimia data set (Sebastian, Klein, & Kimia, 2004).

The bending energy (E_{be}) relies on the curvature signal (Cesar Jr. & Costa, 1997) and it measures the energy required to deform a shape to its lowest energy state, which corresponds to a circle with the same perimeter. Considering a parametric discrete contour of L points with its coordinates represented in the complex form $z[n] = x[n] + jy[n]$, $n \in [0, 1, \ldots, L-1]$, the bending energy is given by:

$$E_{be} = \frac{P^2}{L}\sum_{n=0}^{L-1} k^2[n], \tag{3}$$

where the squared shape perimeter P^2 is a normalizing constant that provides scale invariance. For curvature estimation of a discrete function, $k[n]$ is defined as:

$$k[n] = -\frac{\mathrm{Im}\left(z'[n](z''[n])^*\right)}{|z'[n]|^3}, \tag{4}$$

where $z'[n]$ and $z''[n]$ refer to the first and second derivatives of $z[n]$, respectively. Here, the discrete Fourier transform is used for estimation of the derivatives. $z^*[n]$ denotes the complex conjugate of $z[n]$ and $\mathrm{Im}(\cdot)$ stands for the imaginary part of the argument.

Although the curvature signal is sensitive to local features of shape contour, such as concavity and spatial location of salient points, its low noise immunity limits its use in shape description applications. Thus, it is recommended to smooth the contour before calculating the curvature signal in order to yield a more robust representation, despite the loss of information (Cesar Jr. & Costa, 1996). A classical smoothing strategy is the discrete convolution of $z[n]$ with a Gaussian kernel given by

$$z_\sigma[n] = \sum_{i=1}^{L} z[i]g_\sigma[n-i], \tag{5}$$

where $g_\sigma[n]$ is the Gaussian kernel and σ is the scale parameter for smoothing control. The Gaussian filter $g_\sigma[n]$ is given by

$$g_\sigma[n] = \frac{1}{\sigma\sqrt{2\pi}} e^{-n^2/2\sigma^2}. \tag{6}$$

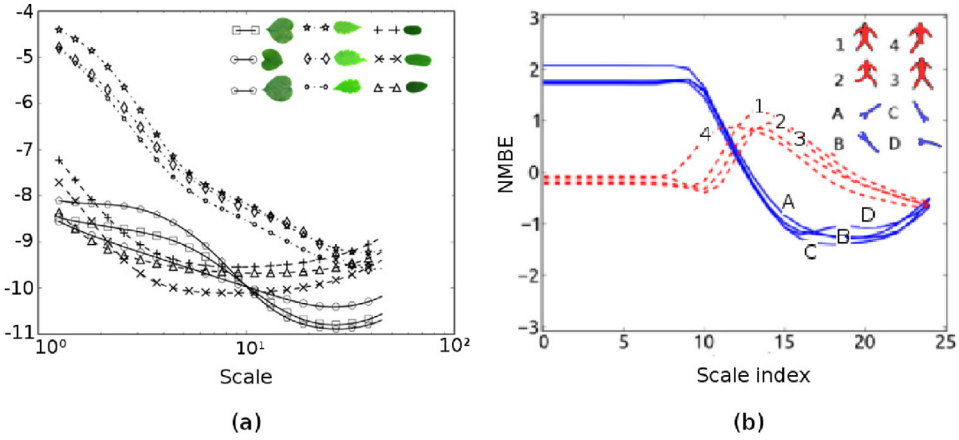

Fig. 2. Non-optimized NMBE shape descriptors for samples of (a) Flavia leaf data set and (b) Kimia99 data set.

It is well-known that the Gaussian filtering process modifies the amplitude of the spectral representation of the contour in such a manner that the contour tends to shrink as the kernel scale parameter decreases (Cesar Jr. & Costa, 1996; 1997). In order to circumvent this undesired shrinkage effect, the smoothed contour is normalized according to

$$\check{z}_\sigma[n] = \frac{P}{P_\sigma} z_\sigma[n], \tag{7}$$

where P and P_σ are the perimeters of the non-smoothed and smoothed contours, respectively.

By replacing $k[n]$ with $k_\sigma[n]$ and $z[n]$ with $\check{z}_\sigma[n]$ in Eqs. (3) and (4), and calculating them for M different smoothing scale factors $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_M)$, we obtain a multiscale representation of the normalized bending energy which is given by:

$$\text{NMBE} = (\log E_{\sigma_1}, \log E_{\sigma_2}, \ldots, \log E_{\sigma_M}). \tag{8}$$

2.2. Optimization methods

This paper investigates the solutions of three optimization algorithms to minimize the objective function MAD, namely the simulated annealing (SA), differential evolution (DE) (Storn & Price, 1997) and particle swarm optimization (PSO) (Shi & Eberhart, 1998). Here, these solutions consist of the optimized parameters of a multiscale shape descriptor.

Simulated annealing is a classical probabilistic optimization algorithm that simulates the thermodynamic process of material heating and cooling. This algorithm encompasses a temperature variable that determines the probability of accepting worse solutions as the algorithm explores the solution space. Thus, this probability decreases as well as the temperature variable exponentially diminishes over time and the algorithm converges to an optimal solution. Particularly, the implemented algorithm embodies a mechanism to disturb the current solution which consists in adding or subtracting, with a given probability $0.3 < P_r < 0.6$, a value from each coordinate of the current solution. In fact, this value is drawn from a normal distribution, with zero mean and variance, i.e. $\kappa.v_i.(1 + COST(v))$, where $\kappa = 0.6$, v_i is the i_{th} coordinate of the current solution and $COST(v)$ is the fitness. The parameters of the SA algorithm are: the initial temperature (T_0), the temperature decreasing factor (α), the number of attempts of the algorithm to replace the current solution by another (P) and the number of times the algorithm effectively replaced the current solution by another (L). The parameters of SA were empirically set to $T_0 = 80$, $\alpha = 0.95$, $P = 10$ and $L = 5$.

The differential evolution algorithm iteratively evolves a population of N vectors of candidate solutions and it produces disturbed versions of each population solution by using a disturbance vector. Here, we employ the *DE/rand/1* strategy introduced by Storn (1996), which is a modified version of DE that computes the disturbance vector according to:

$$\boldsymbol{d} = \boldsymbol{\sigma_1} + \beta.(\boldsymbol{\sigma_2} - \boldsymbol{\sigma_3}), \tag{9}$$

where $\boldsymbol{\sigma_1}$, $\boldsymbol{\sigma_2}$ and $\boldsymbol{\sigma_3}$ are three randomly selected distinct candidate solutions in the population, which are also different from the current solution to be disturbed. The β parameter is an amplification factor of the difference between $\boldsymbol{\sigma_2}$ and $\boldsymbol{\sigma_3}$. The disturbance occurs in each coordinate of the current solution, via a cross-over mechanism with probability P_r. Then, the disturbed solution replaces the current one in population if its fitness is better, otherwise the algorithm keeps the current solution. In this paper, we have tuned the algorithm parameters (N, β, P_r) by following the guidelines in (Storn, 1996) and performing several tests. Finally, we have approached the proposed optimization problem with the parameter set: $N = 65$, $\beta = 0.7$ and $P_r = 0.4$.

Particle swarm optimization is a bio-inspired algorithm that evolves a population of N solutions (particles) that move in the search space with a given velocity. This algorithm keeps track of the best positions attained by each particle (*bpp*) and the best global position obtained by all particles (*bgp*) of the swarm. At each iteration, PSO corrects the particle velocities in the search space influenced by two attraction factors c_1 and c_2. The first factor controls the trend of the particle to search around its own best position (bpp) and the second one controls the trend to move towards the best global position (bgp) found. In order to guarantee that PSO converges to a global minimum, the particles decelerate exponentially along iterations with an inertia factor ω (Shi & Eberhart, 1998). Here, we have tuned the PSO parameters to $N = 25$, $\omega = 0.7$ and $c_1 = c_2 = 1.5$. It is worth noting that these values satisfy the convergence criteria for this optimization algorithm as described by Jiang, Luo, and Yang (2007).

2.3. Self-organizing map and the U-matrix

The self-organizing map, namely SOM network (Kohonen et al., 2001), is a neural network that performs nonlinear reduction of high-dimensional data by projecting it into a low-dimensional space. This method is an exploratory data analysis tool that has been used for image visualization (Strong & Gong, 2011), cluster learning (Kuroiwa, Inawashiro, Miyake, & Aso, 2000) and texture classification (Marana, Costa, Velastin, & Lotufo, 1997).

A SOM network represents structural features of input data in a low-dimensional lattice. In fact, this tool displays data features in the lattice by using a neighborhood criterion in order

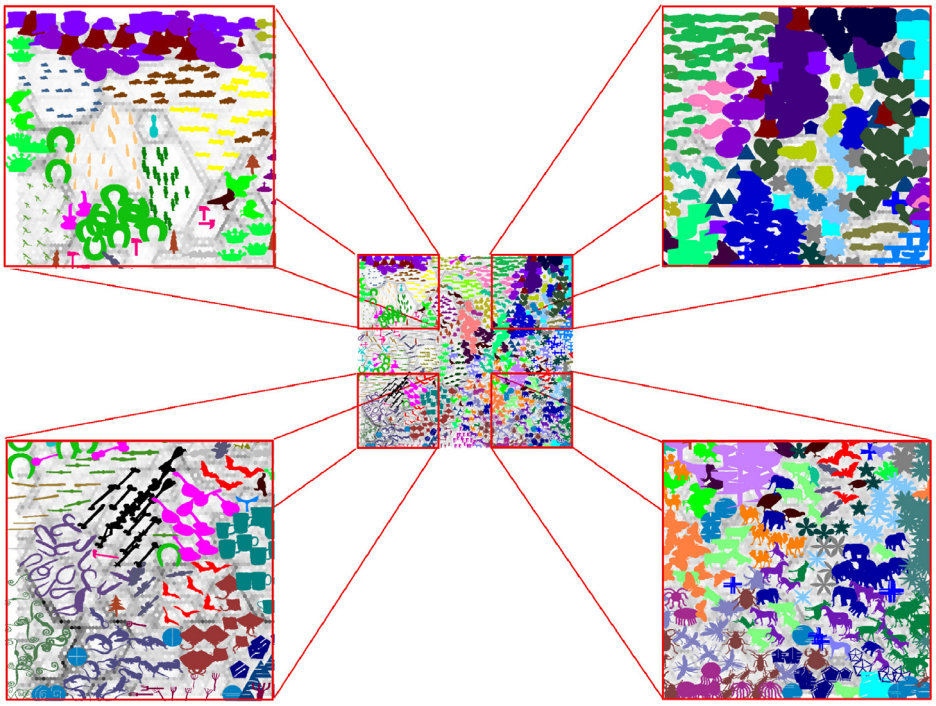

Fig. 3. U-matrices for shapes of MPEG7 CE-Shape-1 data set.

to preserve structure of the input space. According to Ultsch and Siemon (1990), the direct usage of this network output is not suitable for cluster analysis purposes. The SOM network arranges clusters in different regions, although the distances among data points are uniformly distributed. To overcome this drawback, Ultsch and Siemon (1990) developed a two-dimensional projection method, namely the unified distance matrix (U-matrix). The U-matrix displays the local distance structure of a topological preserving projection of a high-dimensional data set (Ultsch & Siemon, 1990).

Fig. 3 shows the U-Matrix for MPEG-7 CE-Shape-1 data set (Latecki, Lakamper, & Eckhardt, 2000) and the central image is the result of the whole data set described by the optimized NMBE. The four images in the corners are details of the central image. This visualization tool identifies how well NMBE describes subtle details from shapes of the same cluster and different clusters. In the detailed images, we can observe the boundaries (in gray) which separate the well-defined shape clusters (e.g. clusters in the top-left corner image). Well-defined clusters are the ones with the lowest within-class and highest inter-class distances.

2.4. *Multidimensional scaling* (MDS)

The multidimensional scaling (Cox & Cox, 2000) searches for a low-dimensional data representation which preserves the distances of the original high-dimensional space. In general, this technique analyzes data similarity or dissimilarity. Furthermore, MDS models data similarity or dissimilarity as distances in geometric spaces.

There exist two types of MDS algorithms: metric and non metric. In the former, the input similarity matrix arises from a metric where the triangle inequality holds. Thus, distances between two points are set to be as close as possible to the data similarity or dissimilarity. In the non-metric version, the algorithm attempts to preserve the order of distances and, furthermore, it seeks a monotonic relationship among the distances in the embedded space and the similarities/dissimilarities.

The R^2 coefficient measures the fitness degree of the low-dimensional representation. This coefficient indicates, in percentage, the fitness of a model to the observed data. Thus, the closer the value of R^2 is to 1, the better the model fits the observed data. Let d and $\hat{d}$ be the symmetrical distance matrices between feature vectors in the low- and high-dimensional space. The R^2 coefficient is given by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\sum_{j=i}^{n}(\hat{d}_{i,j} - d_{i,j})^2}{\sum_{i=1}^{n}\sum_{j=i}^{n}(d_{i,j} - \bar{d})^2},\ R^2 \in [0,1] \tag{10}$$

where n is the number of samples, $d_{i,j}$ the distance between samples i and j in the high-dimensional space, $\hat{d}_{i,j}$ the distance between samples i and j in the low-dimensional space and $\bar{d}$ the mean distance in the high-dimensional space.

The MDS projections in Fig. 4 illustrate how the shape clusters evolve as DE searches for the optimized parameters of the shape descriptor (NMBE). The shape samples are from Kimia data set (Sebastian et al., 2004) which comprises 99 shape images. In this paper, we have employed a manifold learning technique to produce the MDS projections of the optimized NMBE via DE. Fig. 4 also shows that, when the MAD values decreased, the inter-class distances increased and therefore the shape clusters became more evident. At the minimum value of MAD, i.e. when the optimization process converged the only clusters that were not well separated were the four-legged animals and the hand shapes.

3. Results and discussions

We performed shape classification and retrieval experiments with the proposed optimization methodology by testing two descriptors on the Flavia data set (Wu et al., 2007), which comprises 1907 leaf images with 32 different species. Fig. 5 depicts leaf images taken from this challenging data set, which presents a high between class similarity. This data set is widely used to validate automatic recognition systems of plant specimens.

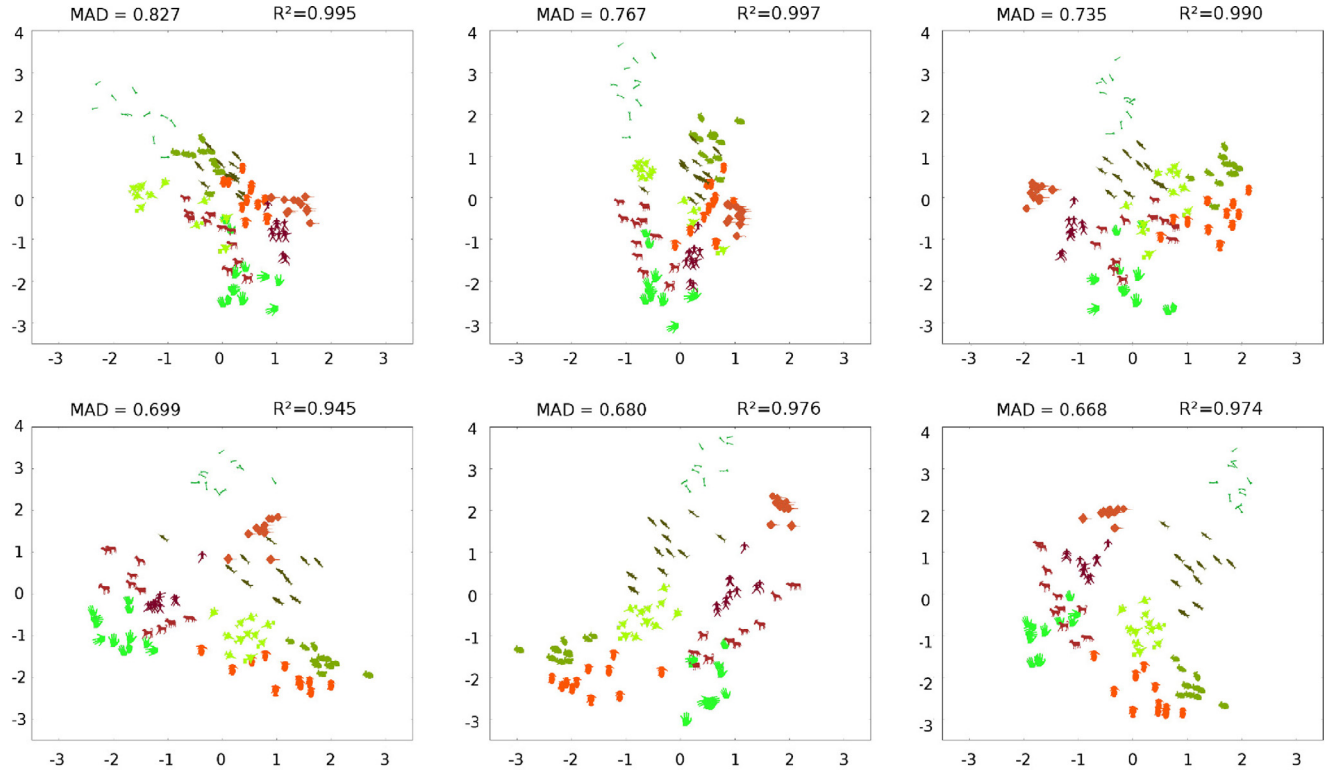

Fig. 4. The MDS projections of an experiment with 99 shapes from Kimia data set (Sebastian et al., 2004). The images display how clusters evolve during an optimization process (DE) as well as MAD and R^2 values.

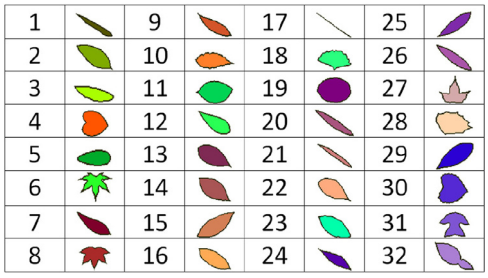

Fig. 5. Samples of Flavia leaf data set.

3.1. MAD *analysis and shape classification*

Experiments were carried out to compare the optimization algorithms regarding convergence. Fig. 6 illustrates the convergence of each optimization algorithm for 30 runs on a subset of leaf shapes. The optimization results indicate that PSO required a smaller number of iterations to converge to an optimal solution, whereas SA and DE required a larger number of iterations. Since convergence aspects are closely related to the number of iterations required to achieve an optimal solution, few iterations imply poor exploration of the search space, thus leading to premature convergence (Engelbrecht, 2007). In such case, the optimization method tends to get stuck in local minimums finding sub-optimal solutions. On the other hand, slower convergence enables the optimization method to explore more the search space, thus increasing the chance to converge to global minima, i.e. optimal solutions (Engelbrecht, 2007). In this context, PSO hit more frequently in sub-optimal solutions, reaching a higher mean MAD value of 0.805 ± 0.006. However, SA and DE achieved lower mean MAD values of 0.795 ± 0.006 and 0.798 ± 0.004, respectively.

In this context, the experimental results demonstrated that SA and DE were more efficient than PSO to find an optimal solution. However, the computational cost evaluation indicated that SA was less expensive than DE since the former evaluated the objective function less times than the latter. Section 3.4 addresses more details about the computational cost of the optimization algorithms.

Our next analysis concentrates on the assumption that the optimized scales and the minimum values of MAD, which are intertwined, lead to an increase of the correct classification rate. Table 1 presents different MAD values and their corresponding classification results. The optimized scales which attained the best classification results correspond to the minimum values of MAD. The NMBE descriptor whose scales were adjusted according to Cesar Jr. & Costa (1997) ($NMBE_{orig}$) led the classifiers to reach an intermediate performance. The optimized NMBE ($NMBE_{opt}$) improved the performance of all classifiers since they accomplished the highest Precision and Recall rates when the objective function reached the minimum values. Thus, we can state that there is an intrinsic relation between the high correct classification rates and minimum MAD values. Although the optimization algorithms achieved different MAD values and therefore different sets of optimized scales, the classification experiments showed that these scales captured leaf shape details and variations. Such results confirmed that we can take advantage of the optimization algorithms to strengthen leaf shape characterization and analysis. In spite of NMBE has been originally designed for neural cell description, this paper focus on leaf shape characterization. Moreover, the proposed objective function tailored NMBE in terms of scales to the problem and leaf shape data set under analysis, satisfactorily.

The classifiers that used the non-optimized NMBE with randomly selected scales did not perform well since they have reached the lowest values for Precision and Recall and the highest values for the objective function (MAD). Thus, we concluded that randomly selected scales were less sensitive to leaf shape variations and tended to provide more misclassifications.

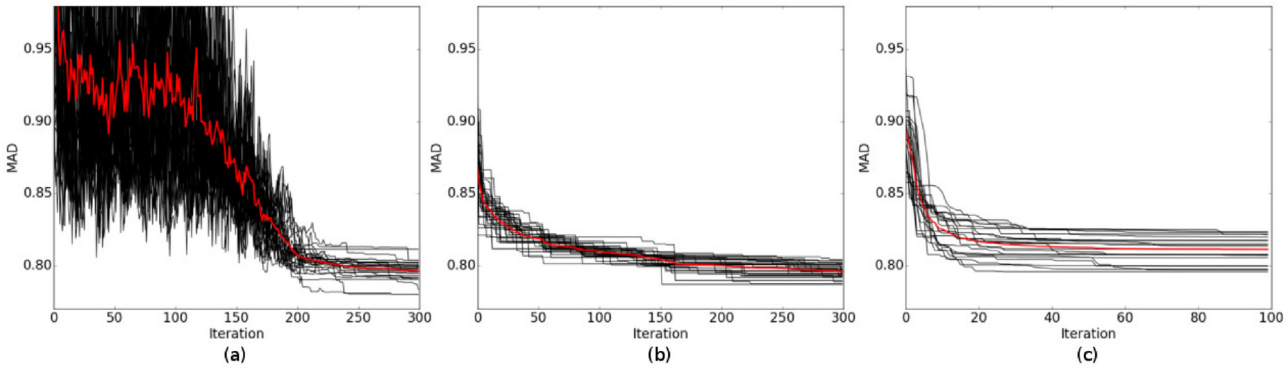

Fig. 6. Convergence of the optimization methods: (a) SA, (b) DE, (c) PSO. The red lines in each graphic refer to the mean MAD value of 30 runs (black lines) for the results of each optimization strategy within a subset of leaf shapes.

Table 1
Different MAD values and the classification results for different scale selection strategies for Flavia leaf data set.

MAD	Classifier							
	NB		Knn (n = 5)		LDA		QDA	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
0.762[a]	0.91 ± 0.02	0.89 ± 0.02	0.93 ± 0.02	0.92 ± 0.02	0.87 ± 0.02	0.85 ± 0.02	0.95 ± 0.01	0.94 ± 0.01
0.783[b]	0.88 ± 0.02	0.87 ± 0.02	0.90 ± 0.02	0.88 ± 0.02	0.85 ± 0.02	0.83 ± 0.03	0.91 ± 0.02	0.90 ± 0.02
0.829[c]	0.86 ± 0.03	0.85 ± 0.03	0.89 ± 0.03	0.88 ± 0.03	0.84 ± 0.03	0.82 ± 0.03	0.91 ± 0.02	0.89 ± 0.02
0.867[d]	0.85 ± 0.02	0.84 ± 0.02	0.89 ± 0.02	0.88 ± 0.02	0.82 ± 0.03	0.81 ± 0.03	0.89 ± 0.02	0.88 ± 0.02
0.969[e]	0.81 ± 0.03	0.79 ± 0.03	0.87 ± 0.02	0.85 ± 0.02	0.77 ± 0.03	0.77 ± 0.03	0.87 ± 0.03	0.85 ± 0.03
1.04[e]	0.69 ± 0.03	0.68 ± 0.03	0.83 ± 0.03	0.82 ± 0.03	0.74 ± 0.03	0.73 ± 0.03	0.81 ± 0.03	0.79 ± 0.03

[a] Using SA
[b] Using DE
[c] Using PSO
[d] Using scales proposed by Cesar Jr. and Costa (1996)
[e] Random selection

3.2. Visual exploratory cluster analysis

The visual exploratory cluster analysis indicates that the optimized descriptor improves the cluster arrangement of leaf shapes, which reinforces that MAD is an objective function suitable to guide parameter optimization. Fig. 7 depicts the U-matrices that support the visualization of leaf shapes. The leaf colors are in accordance with the class labels exhibited in Fig. 5. These U-matrices comprise 1907 leaf shapes and, furthermore, each leaf is represented by its descriptor.

Fig. 7a shows leaf shapes which are described by the optimized NMBE, while Fig. 7b displays, in terms of U-matrix elements, leaf shapes described by the non-optimized NMBE descriptor. The graphical representations of the *silhouette* measure in Fig. 7a and Fig. 7b demonstrated that the leaf classes, which were properly characterized and grouped, were those with positive mean *silhouette* values. Nevertheless, leaf shapes whose descriptors were not able to characterize and group them properly were those with negative *silhouette* values. Fig. 7a shows the remarkable reduction of negative mean *silhouette* values when applying $NMBE_{opt}$ to shapes from the Flavia leaf data set.

The black square in Fig. 7a highlights the performance of the optimized descriptor. It also shows how well the optimized descriptor mapped leaf shapes into groups according to their respective classes. Moreover, the optimized descriptor has remarkably improved the mean *silhouette* per class for almost all leaf classes. In contrast, the region inside the black square in Fig. 7b indicates that the non-optimized descriptor was unable to satisfactorily map leaf shapes in their corresponding classes. These results indicated that the optimized descriptor has improved the cluster arrangement of leaves due to the intrinsic shape variations within the set of the optimized scales that can not be properly embodied in the studied traditional parameter selection methods.

Fig. 8a, Fig. 8b and Fig. 8c display the U-matrices and the corresponding MAD values for experiments on Flavia leaf data set with SA, DE and PSO, respectively. An interesting aspect to examine in these results regards the convergence of the optimization algorithms to different solutions (MAD). These solutions resulted in different cluster arrangements and therefore different relations among the neighborhood structure of the clusters. For instance, the overall dispersion of the elements in Fig. 8a and Fig. 8b tend to be very similar since the corresponding MAD values are close to each other. At the same time, the result that corresponds to the highest MAD value (0.829) exhibits a relatively high dispersion within clusters, particularly at the center of the U-matrix (Fig. 8c), which reinforces the hypothesis that PSO converged to a local minimum. These findings point to the important conclusion that the lower the MAD value, the better the cluster arrangement and hence the cluster quality. Therefore, we state that the cluster quality and MAD value are negatively correlated. Our analysis also considers that there are other solutions than the minimum, in the search space, which may be suitable to the problem under study.

Fig. 9 illustrates the MDS projections for three different MAD values. Fig. 9a corresponds to the cluster arrangement after the convergence of the SA algorithm, i.e. after reaching the minimum MAD value. These projections confirmed that the minimization of the objective function provided leaf cluster arrangements with lower average within-class distances and higher inter-class distances, and thus more compact and separated clusters. The visual analysis of the three detail images in Fig. 9a discloses that the MDS projection of the optimized NMBE descriptor increased

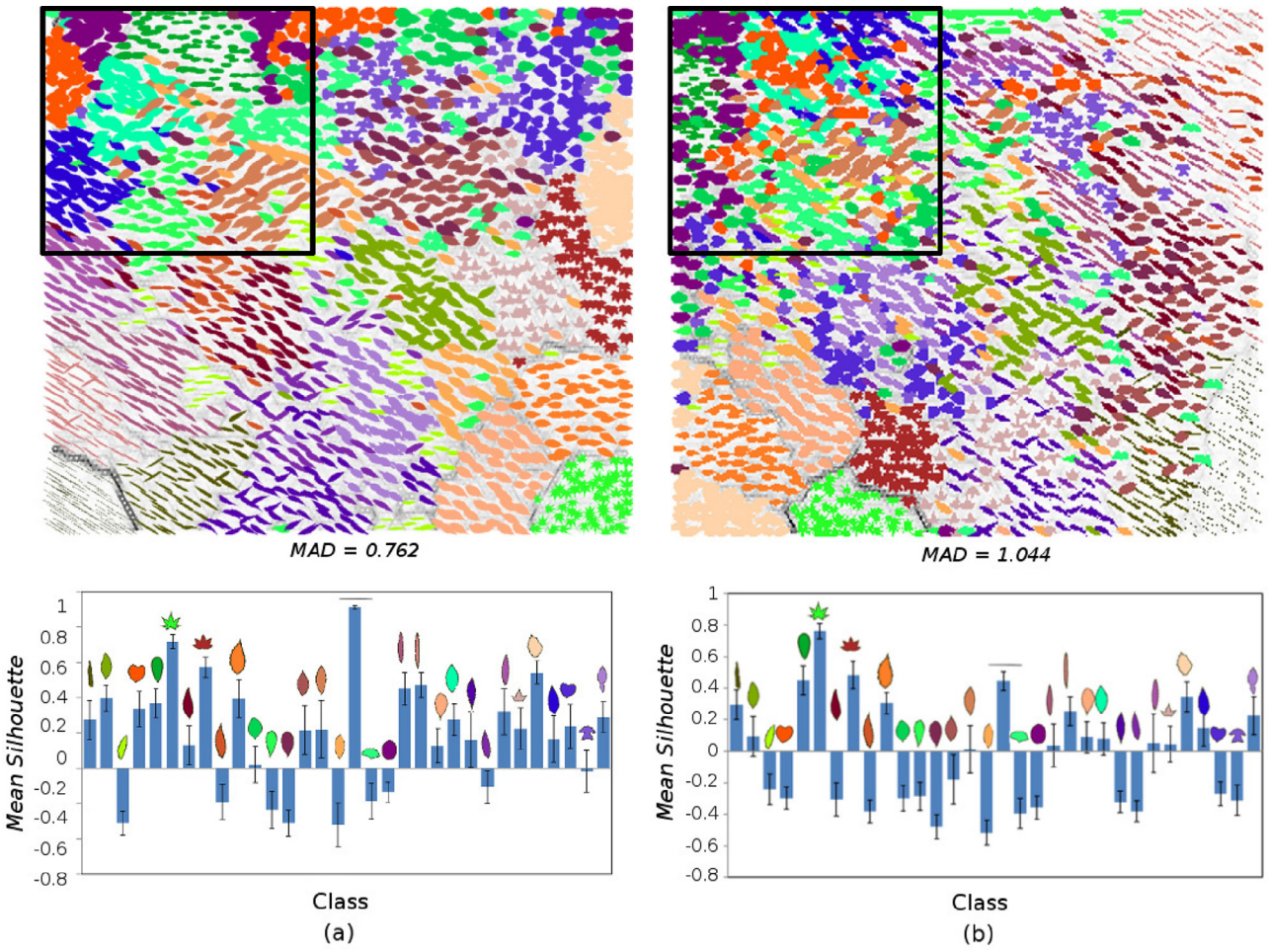

Fig. 7. U-matrices and mean *silhouette* values for Flavia leaf data set using the (a) optimized and (b) non-optimized NMBE.

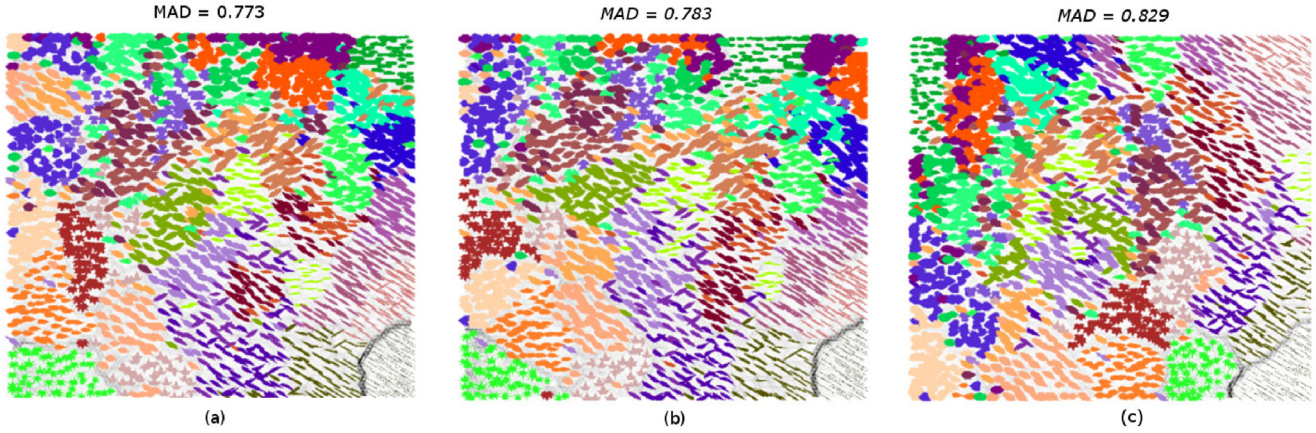

Fig. 8. U-matrices and MAD values for Flavia leaf data set achieved by using the optimized NMBE with (a) SA, (b) DE, (c) PSO.

inter-class distances, whereas it reduced the within-class distances when compared to Fig. 9b and Fig. 9c. Moreover, the R^2 coefficient values close to 1 imply that the low-dimensional representation preserved the mean distance in the original high-dimensional data space.

3.3. Shape retrieval experiment

For the sake of comparison, we have optimized the parameters of another shape descriptor namely inner distance shape context (IDSC) (Ling & Jacobs, 2007). This descriptor is mainly applied to shape retrieval, which involves unsupervised shape classification tasks. Moreover, IDSC employs dynamic programming (DP), such as Dynamic Time Warping (DTW) (Palazón-González & Marzal, 2012), to improve shape matching accuracy. In this paper, we have replaced the L_2 norm by DTW to compute the *silhouette* measure and the cost function defined in Eq. (1). Fig. 10 exhibits the comparison of shape retrieval experiments which were performed on the public Flavia leaf data set and with both non-optimized NMBE and IDSC and their optimized counterparts ($NMBE_{opt}$, $IDSC_{opt}$). The per-

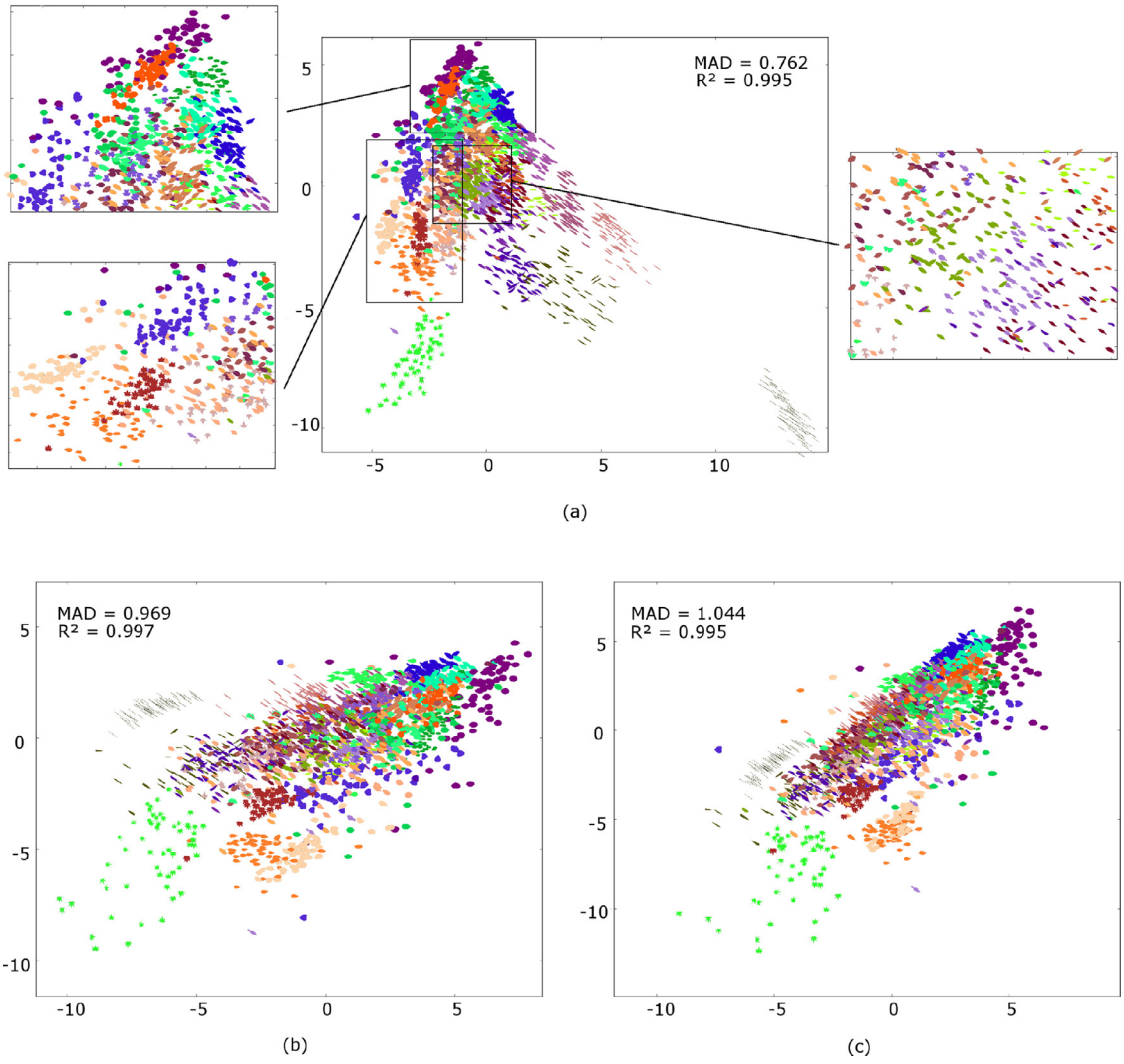

Fig. 9. MDS projections of the NMBE descriptors for leaf shapes of Flavia data set. (a) The SA-optimized NMBE descriptor with $MAD = 0.762$ (b) non-optimized NMBE with $MAD = 0.969$ and (c) non-optimized NMBE with $MAD = 1.044$.

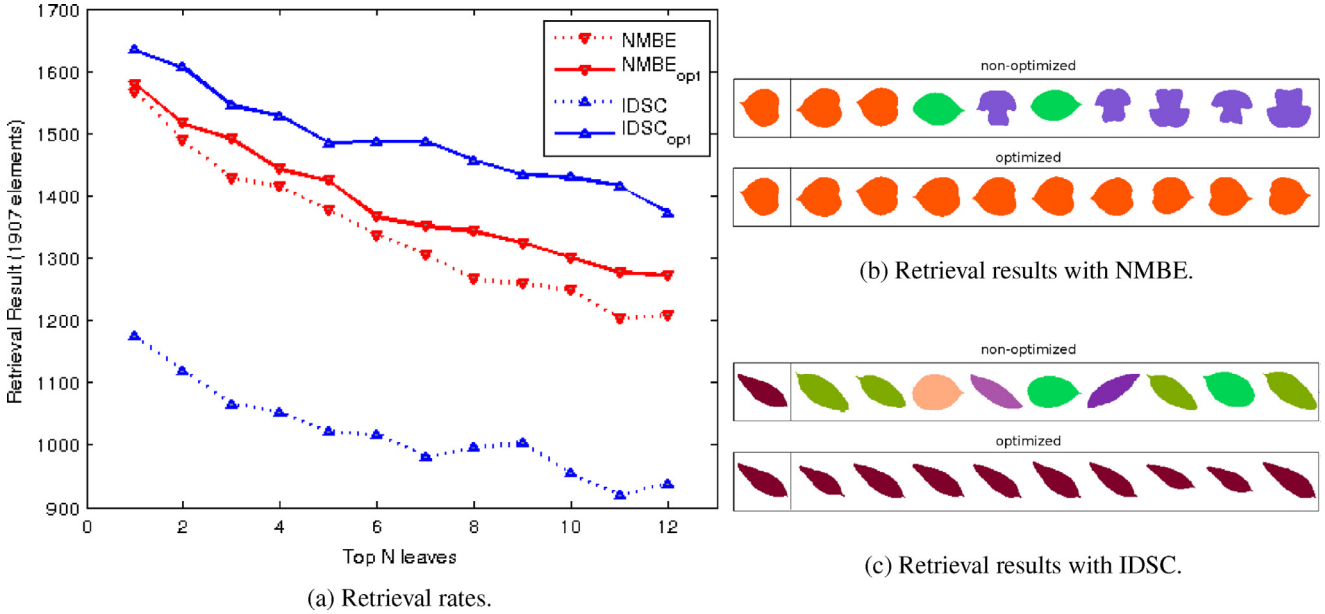

Fig. 10. Experiments conducted on Flavia leaf data set (a) retrieval rate by using both NMBE and IDSC and their optimized counterparts, (b) and (c) two leaf shape retrieval examples by using the non-optimized and optimized NMBE and IDSC descriptors, respectively.

Table 2
Bulls-eye scores for Flavia data set.

NMBE	$NMBE_{opt}$	IDSC	$IDSC_{opt}$
63.86%	71.16%	53.38%	77.50%

Table 3
Computational complexity of the optimization methods.

Method	Complexity
SA	$O(P.N_{iter}.\log N_{iter})$
DE	$O(N_{pop}.N_{iter}.D)$
PSO	$O(N_{pop}.N_{iter}.\log N_{iter})$

formance evaluation methodology has adapted the Bulls-eye measure, where the result is the overall number of shapes correctly retrieved, in each rank position, for each shape in data set taken as a query. Let N_c be the number of shapes which belong to the class c. Here, the number of retrieved shapes (N_r) was adjusted to $N_r = 2 \min_{c=1,2,\ldots 32} N_c$. Fig. 10 also demonstrates that the optimization methodology was decisive for both descriptors to achieve better retrieval rates. Fig. 10a shows that after tuning IDSC parameters for the Flavia leaf data set with the optimization methodology the retrieval rate has remarkably increased. Therefore, $IDSC_{opt}$ outperformed the optimized NMBE ($NMBE_{opt}$) and non-optimized NMBE. In this sense, the optimized parameters of $IDSC_{opt}$ may have incorporated subtle details of leaf shapes. On the other hand, we have observed that the non-optimized IDSC underperformed the non-optimized NMBE. In order to provide a non-optimized counterpart of IDSC, we have followed the parameter setting introduced by Wang et al. (2015). Fig. 10b and Fig. 10c exemplify samples of plant leaves which were retrieved according to the rank of the similarity to the two queries. These results have demonstrated that the non-optimized descriptors in Fig. 10b and Fig. 10c were unable to retrieve all samples for the two queries, whereas the optimized descriptors retrieved all of them correctly.

Table 2 presents the performance evaluation of NMBE, whose scales were computed using the scheme introduced by Cesar Jr. and Costa (1996), IDSC and their optimized counterparts. The retrieval rates show that both optimized descriptors outperformed their non-optimized counterparts on Flavia data set. The significant improvement of the Bulls-eye rates reinforces our assumption that the optimization methodology is suitable for leaf shape retrieval and analysis.

Thus, we can infer that the optimized descriptors are more likely to succeed in shape retrieval experiments because the optimized sets of parameters probably embody intrinsic and subtle information about leaf shapes. We can also assume that these optimized descriptors may reliably characterize and also discriminate shape differences within and among leaf classes.

3.4. Computational cost

Table 3 shows the computational complexity results of the three optimization algorithms. SA and PSO present similar complexity results which rely on the number of iterations to converge (N_{iter}), population size (N_{pop}) and P parameter. On the other hand, DE demands a higher complexity which relies on the dimension of the optimization problem (D), population size (N_{pop}) and number of iterations to converge (N_{iter}).

For the sake of comparison, we assumed that the computational cost is the number of times that the objective function is demanded throughout the optimization process. Its calculation takes into account the parameter setting of each method, presented in Section 2.2, and the corresponding computational complexity. The computational cost to address the optimization of the multiscale descriptor for SA, DE and PSO yielded 7, 4314, 19, 500 and 1107, respectively.

It is worth noting that there is a trade-off between the computational cost and the quality of the optimal solution found. Although the reduction of the number of shape samples, image resolution and the N_{pop} variable can lessen the computational cost, it may degrade the optimization result. Moreover, parallelism may contribute to reduce the computational cost without degrading the optimization result. However, it adds an extra computational complexity to the optimization algorithms.

4. Conclusions

This paper presents an optimization methodology for parameter adjustment of shape descriptors applied to leaf characterization and clustering. Our optimization methodology is a versatile tool for shape analysis that mainly relies on an objective function that can be adapted to different application problems or databases. Here, the minimization of the objective function accomplishes the best parameter set of a given shape descriptor to improve leaf characterization quality.

The performance evaluation of the optimization methodology led us to conclude that the optimized parameters were able to reveal subtle leaf shape features and overall they have improved cluster organization and classification of plant leaves. Actually, the optimized descriptors have reliably characterized shapes that exhibited multiscale features and they have also discriminated shape differences within and among leaf classes. Likewise, the optimized shape descriptors provided a global and robust description of the challenging Flavia data set, despite it presents a high between class similarity.

A relevant visual analysis of the proposed methodology linked the optimized and non-optimized NMBE with the classes of binary shapes that they represented in a multidimensional space by using U-matrices. This high-dimensional projection tool provided a qualitative evaluation and visual explanation on how the different classes of shapes were better spatially grouped and scattered due to the optimization methodology. Regarding the quantitative performance evaluation, we have also observed that the *silhouette* measure was suitable to numerically assess the multiscale shape descriptor and thus infer whether it was able to characterize leaf shapes or not. Our findings indicated that the optimized NMBE and IDSC were adequate to characterize leaf shapes and furthermore they may provide shape signatures to address leaf taxonomy problems. Moreover, the proposed methodology is an additional tool and source of information for plant taxonomist to discriminate and group leaf species.

Acknowledgments

The authors are grateful to Fundação de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Ceará-FUNCAP and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (401442/2014-4 and 444784/2014-4) for financial support. We would like to thank Allan Carneiro, Gerardo Lopes and Jeová Farias for insightful discussions, and for helping improve the manuscript.

Appendix A. Descriptor parameters

Table 4 reports the optimized parameters of the shape descriptors for Flavia leaf data set using the SA algorithm. We considered these parameters to carry out the supervised classification and shape retrieval experiments.

Table 4
Set of IDSC and NMBE parameters resulted from the proposed optimization methodology applied to Flavia data set.

Method	Param. names	Description	Range	Optim. Param.
IDSC	n_c	Number of samples of shape contour.	[20, 200]	91
	n_d	Number of bins in Inner distance histogram.	[3, 30]	13
	n_θ	Number of bins in angle histogram.	[3, 30]	24
	n_s	DP matching starts for multiple-path search.	[1, 200]	20
	thr	Threshold for DP matching.	[0, 2]	1.3518
NMBE	s_1	1*st* scale	[0.1, 125]	0.17
	s_2	2*nd* scale	[0.1, 125]	3.97
	s_3	3*rd* scale	[0.1, 125]	43.1
	s_4	4*th* scale	[0.1, 125]	70.5
	s_1	5*th* scale	[0.1, 125]	96.1
	s_6	6*th* scale	[0.1, 125]	118.5

References

Amorim, E., Brazil, E. V., Mena-Chalco, J., Velho, L., Nonato, L. G., Samavati, F., & Sousa, M. C. (2015). Facing the high-dimensions: Inverse projection with radial basis functions. *Computers & Graphics, 48*, 35–47.

Anuar, F. M., Setchi, R., & Lai, Y. K. (2013). Trademark image retrieval using an integrated shape descriptor. *Expert Systems with Applications, 40*(1), 105–121.

Backhaus, A., Kuwabara, A., Bauch, M., Monk, N., Sanguinetti, G., & Fleming, A. (2010). Leafprocessor: A new leaf phenotyping tool using contour bending energy and shape cluster analysis. *New Phytologist, 187*(1), 251–261.

Bai, X., Yang, X., Latecki, L., Liu, W., & Tu, Z. (2010). Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(5), 861–874.

Cesar Jr., R. M., & Costa, L. F. (1996). Towards effective planar shape representation with multiscale digital curvature analysis based on signal processing techniques. *Pattern Recognition, 29*(9), 1559–1569.

Cesar Jr., R. M., & Costa, L. F. (1997). Application and assessment of multiscale bending energy for morphometric characterization of neural cells. *Review Of Scientific Instruments, 68*(5), 2177–2186.

Chaki, J., Parekh, R., & Bhattacharya, S. (2015). Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recognition Letters, 58*, 61–68.

Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems with Applications, 39*(8), 7562–7573.

Cox, T. F., & Cox, M. A. A. (2000). *Multidimensional scaling* (2nd). Chapman and Hall/CRC.

Direkoglu, C., & Nixon, M. S. (2011). Shape classification via image-based multiscale description. *Pattern Recognition, 44*, 2134–2146.

Dornbusch, T., Jillian, W., Baccar, R., Fournier, C., & Andrieu, B. (2011). A comparative analysis of leaf shape of wheat, barley and maize using an empirical shape model. *Annals of Botany, 107*(5), 865–873. doi:10.1093/aob/mcq181.

Eiben, A. E., & Smith, J. (2015). From evolutionary computation to the evolution of things. *Nature, 521*(7553), 476–482.

Engelbrecht, A. P. (2007). *Computational intelligence* (2nd). John Wiley & Sons Ltd.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2nd). Morgan Kaufmann (Academic Press).

Ghasab, M. A. J., Khamis, S., Mohammad, F., & Fariman, H. J. (2015). Feature decision-making ant colony optimization system for an automated recognition of plant species. *Expert Systems with Applications, 42*(5), 2361–2370.

Gonzalez, R. C., & Woods, R. E. (2006). *Digital imaging processing* (3rd). Upper Saddle River, NJ, USA: Prentice Hall, Inc.

Jiang, M., Luo, Y., & Yang, S. (2007). Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm. *Information Processing Letters, 102*(1), 8–16.

Kohonen, T., Schroeder, M. R., & Huang, T. S. (Eds.). (2001). *Self-organizing maps* (3rd). Secaucus, NJ, USA: Springer-Verlag New York, Inc..

Kuroiwa, J., Inawashiro, S., Miyake, S., & Aso, H. (2000). Self-organization of orientation maps in a formal neuron model using a cluster learning rule. *Neural Networks, 13*(1), 31–40.

Latecki, L., Lakamper, R., & Eckhardt, T. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *Proceedings of cvpr 2000. ieee conference on computer vision and pattern recognition: Vol. 1* (pp. 424–429).

Ling, H., & Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence., 29*(2), 286–299. doi:10.1109/TPAMI.2007.41.

Marana, A., Costa, L. F., Velastin, S., & Lotufo, R. (1997). Oriented texture classification based on self-organizing neural network and Hough transform. In *Proceedings of ICASSP-97. ieee international conference on acoustics, speech, and signal processing: Vol. 4* (pp. 2773–2775).

Mokhtarian, F., & Mackworth, A. K. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(1), 34–43.

Mokhtarian, F., & Suomela, R. (1998). Robust image corner detection through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(12), 1376–1381.

Palazón-González, V., & Marzal, A. (2012). On the dynamic time warping of cyclic sequences for shape retrieval. *Image and Vision Computing, 30*(12), 978–990.

Paula Jr., I. C., Medeiros, F. N., Bezerra, F. N., & Ushizima, D. M. (2013). Multiscale corner detection in planar shapes. *J. Math. Imaging Vis., 45*(3), 251–263. doi:10.1007/s10851-012-0365-8.

Rossatto, D. R., Casanova, D., Kolb, R. M., & Bruno, O. M. (2011). Fractal analysis of leaf-texture properties as a tool for taxonomic and identification purposes: a case study with species from neotropical melastomataceae (miconieae tribe). *Plant Systematics and Evolution, 291*(1–2), 103–116.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*(1), 53–65.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York, NY, USA: John Wiley & Sons, Inc..

Sebastian, T., Klein, P., & Kimia, B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence., 26*(5), 550–571.

Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimizer. In *Proceedings of ieee international conference on evolutionary computation. ieee world congress on computational intelligence* (pp. 69–73).

Storn, R. (1996). On the usage of differential evolution for function optimization. In *Proceedings of biennial conference of the north american fuzzy information processing society. nafips* (pp. 519–523).

Storn, R., & Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*(4), 341–359.

Strong, G., & Gong, M. (2011). Similarity-based image organization and browsing using multi-resolution self-organizing map. *Image and Vision Computing, 29*(11), 774–786.

Ultsch, A., & Siemon, H. P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of international neural networks conference (innc)* (pp. 305–308).

Wang, B., Brown, D., Gao, Y., & La Salle, J. (2015). March: Multiscale-arch-height description for mobile retrieval of leaf images. *Information Sciences, 302*, 132–148.

Wang, J., Bai, X., You, X., Liu, W., & Latecki, L. J. (2012). Shape matching and classification using height functions. *Pattern Recognition Letters, 33*(2), 134–143.

Webb, A. R. (2002). *Statistical Pattern Recognition* (2nd). John Wiley & Sons Ltd.

Wu, S., Bao, F., Xu, E., Wang, Y.-X., Chang, Y.-F., & Xiang, Q.-L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *Proceedings of ieee international symposium on signal processing and information technology* (pp. 11–16).