

# Advanced leaf image retrieval via Multidimensional Embedding Sequence Similarity (MESS) method

F. Fotopoulou · N. Laskaris · G. Economou ·  
S. Fotopoulos

Received: 30 September 2010 / Accepted: 6 November 2011 / Published online: 4 December 2011  
© Springer-Verlag London Limited 2011

**Abstract** A novel method for shape analysis and similarity measurement is introduced based on a time series matching approach. It applies to shapes represented through one-dimensional signals and has as objectives to utilize efficiently the provided information and to optimize the shape matching process. The new technique is tested on boundaries from leaf images, after their conversion into 1D sequences using either the Centroid Contour Distance (CCD) or the Angle code (AC) measurements. In the core of the new method lies the ‘time delay’-based transformation of a given 1D sequence to an ensemble of vectors embedded in a multivariate phase space. The resulting point set is considered as representative of the leaf identity. Inter-leaf comparisons are carried out in a pairwise fashion by employing the multidimensional, Wald–Wolfowitz, statistical test for the ‘two-sample problem’, which implicitly performs shape matching and similarity quantification. The comparative experimentation shows that the complexity of our method is moderate, while the leaf retrieval performance, compared to that achieved by standard matching procedures usually employed with the CCD and AC representations, is greatly improved.

**Keywords** Shape descriptors · Shape matching · Leaf image retrieval · Time-delay embedding

## 1 Introduction

Content-Based Image Retrieval (CBIR) has become an essential issue in computer vision. According to a usual scenario, the user provides a query image and the system returns a predefined number of similar images included in the database. Automated leaf image retrieval falls in this theme, so as to alleviate the need for a botanist whenever the category of a leaf has to be identified. Dedicated methods for retrieving similar leaves (and of known classification) are becoming popular and applications in mobiles and other devices have started to appear [1, 2]. A system for plant image retrieval, that could potentially provide plant identification as a web service, has been introduced in [3]. In general, algorithms for leaf representation and comparison are an emerging topic in CBIR research. Motivated by the need for efficient image retrieval in the area, we attempted to develop a new method based on the shape representation of leaves.

Images can be recognized either by color, texture, shape, or by an appropriate combination of these characteristics. In particular for the leaf recognition task, shape descriptors are of wide use. A plethora of methods can be found in the literature for shape representation [4, 5]. Chain code, Fourier descriptors, moments and curvature scale space are only a few of them. Two-dimensional shapes can be described either by encapsulating the information provided by the object’s boundary and its features or by description of the region occupied by the object on the image plane. An appropriate shape descriptor should be invariant to several

---

F. Fotopoulou  
Department of Computer Engineering and Informatics,  
University of Patras, Patras, Greece

N. Laskaris  
Artificial Intelligence and Information Analysis Lab,  
Department of Informatics, Aristotle University of Thessaloniki,  
Thessaloniki, Greece

G. Economou · S. Fotopoulos (✉)  
Electronics Lab., Department of Physics,  
University of Patras, Patras, Greece  
e-mail: spiros@physics.upatras.gr

geometrical transformations, such as rotation, reflection, scaling, and translation.

In the case of leaf image retrieval, which is the targeted application of this paper, a great deal of work has been reported [3, 6–13]. Here, we focus on methods based on the closed contour of objects, since it is a highly informative characteristic. A common procedure is to parameterize the two-dimensional closed contours of shapes by a one-dimensional sequence (i.e., a chain of scalars), where the running index is often referred as ‘time’. Many methods using information from the leaves boundary [10, 13] have been reported in literature, representing the leaf by a one-dimensional sequence. An important issue is the matching of two sequences that represent two distinct leaves [13]. The selection of a fixed starting point serves as a solution to the problem of alignment [6], but this strategy carries the obvious disadvantages to the subsequent similarity measure stage. Centroid Contour Distance (CCD) [21] and Angle Code (AC) [14] measurements are among the most popular contour-based features that are extracted from leaf images.

The CCD sequence is considered well suited for leaf description, imitating the measurement taken by an observer which is placed in the coordinate system origin. While the difference in successive points of CCD sequence or other processed versions like CCD histogram could also be used, more often the CCD sequence is employed in the leaf-shape representation. In all cases, when the feature extraction stage is completed, the matching step follows, which in the CCD case is usually accomplished by searching for the minimum Euclidean distance between properly aligned CCD sequences.

The AC on the other hand assigns to each boundary point an angle formed by two consecutive vectors coming to and leaving from this point. The sequence of angles has not yet been utilized as a shape feature. The angle code histogram (ACH) is usually used instead. Adopting the ACH descriptor guarantees scaling and rotation invariance, due to the properties of histogram: Scaling invariance can be easily achieved by expressing the appearance of each bin (angle) as a percentage, while rotation invariance is inherent in all approaches employing aggregation. Weighted combinations of CCD sequence and ACH improve the retrieval performance, while other geometrical characteristics of leaves (like eccentricity) have been exploited as additional features [6].

Beyond the CCD and AC sequences, there are other alternative methods that intensively elaborate on shape contour points to extract useful information. Shape context (SC) descriptor [18] is one of them and had a significant impact to the feature extraction and shape matching research. For each boundary point of the object, a 2D histogram is produced, presenting the other points’ distribution. The inner distance method (IDSC) [19] is an improved version of it,

best suited for articulated objects. The complexity and the computational cost of the above method is considered very high ( $O(N^3)$ ). Probabilistic Neural Network [20] and other intelligent data analysis methods have also been used to handle a larger set of features. Shape representations that result in one-dimensional sequences call for dynamic programming techniques to express (dis)similarity between them [15]. Dynamic Time Warping (DTW) is such a technique that has several applications, including shapes [16] and leaves [17] retrieval. The increased retrieval efficiency, but at the expense of computational complexity, was noticeable in recent papers [16–19].

Alternative representations for 1D contour sequences accompanied with a suitable matching procedure are the new elements introduced in the current work. The current approach employs the embedding of successive sequence segments (which are extracted from the serial data) in an orthogonal vectorial space. Phase space techniques are used to encapsulate partial information from the closed leaf contour in terms of *embedding dimension* and *time-delay* parameters. In such a representation there is no need of a predefined starting point in the matching stage, as rotation invariance is inherent. Using the suggested representation, original sequences are transformed into distinct point swarms in a common space and, therein, their comparison can be handled efficiently as a ‘two-sample problem’ via the multidimensional Wald–Wolfowitz (WW) test [22]. Considering the mechanics (built in the WW-test) underlying the pair wise comparison of two such sequences, it becomes evident that the matching is approximately performed simultaneously in the selected scale of the sub sequence length and an aggregate score is finally derived. This is a similarity measure that can be used for subsequent ordering and retrieval of the corresponding objects. Besides its good discriminating capability the new approach is characterized by rotational invariance, a property that is difficult to achieve in shape recognition problems. In addition, its complexity is significantly reduced compared to the techniques mentioned above. The dimensionality of the phase space affects the performance: appropriate selection of embedding dimension is needed to represent the one-dimensional sequence. Although the proposed method is of general use and can be applied to all kind of one-dimensional time series data, in this work we focus on CCD, AC shape features and show that direct application of the MESS method improves results a great deal. The performance of the new technique is evaluated by means of image retrieval experiments using a leaf database.

The organization of the paper is as follows: Sect. 2 provides a review of the shape sequences to be used. The core of the MESS method will be described in Sect. 3. Details of the database used and the evaluation of retrieval performance will be given in Sect. 4.

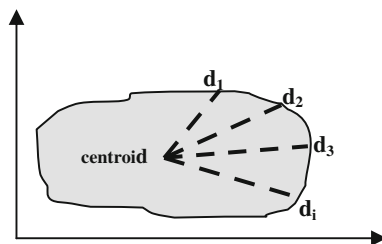
## 2 Sequences from shapes boundaries

In this section a description of two well-known shape sequences that are used to capture information from the shape boundary is provided. Appropriate formulation of CCD and AC as well as their properties respecting scaling, translation and rotation invariance is reported.

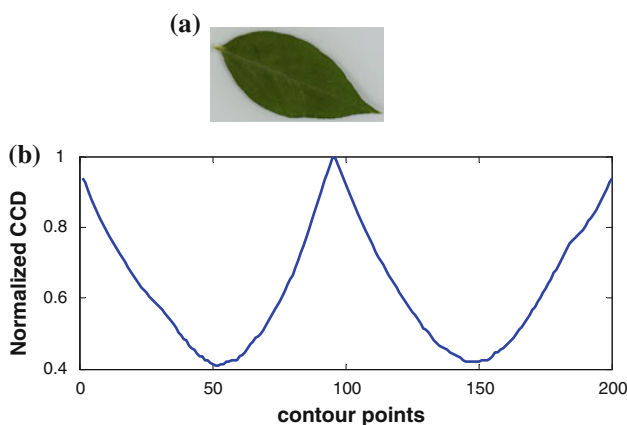
### 2.1 The CCD sequence

The CCD sequence presents the distance  $d_i$  of each boundary point  $i$  from the contour centroid. As shown in Fig. 1, each sequence value corresponds to the distance between the centroid of the shape and the contour point  $i$ . A sequence  $x(n)$  of centroid-to-contour distances is formed, where  $n = 1 \dots i \dots N$ .

It contains both global information, related to the object's structure, as well as local detailed contour information. It is a translation invariant representation and under appropriate normalization can be rendered scale invariant as well. In this paper, the normalization with respect to the max value has been adopted. Rotation invariance is not straightforward and for this reason a process of defining a certain origin is required [6]. A leaf and its CCD sequence are depicted in Fig. 2.



**Fig. 1** Definition of Centroid Contour Distance (CCD) sequence



**Fig. 2** A leaf of the *Silene stellata* species (a) and the corresponding CCD sequence (b)

Rotation invariance is important in the matching step and could be bypassed by a procedure where the minimum Euclidean distance from all the circularly shifted versions is selected [21]. However, this process is sensitive to partial distortions. As it will be clarified in Sect. 3, the leaf-shape representation built in MESS method does not require such a circular-shifting procedure during shape matching as it is rotation invariant due to the part-based sequence representation and the statistical nature of the matching procedure.

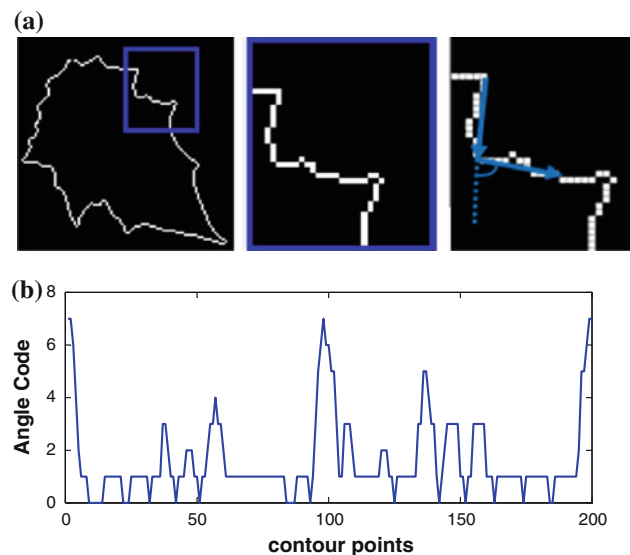
### 2.2 The AC sequence

The AC [14] is a common, very descriptive characteristic which employs a sequence of angles to represent the object shape. The geometrical characteristics of the shape in terms of the angle produced at each point  $i$  of the contour are computed in the contour coordinate system. The angle is formed by two successive line segments (vectors) connecting the reference point  $i$  with the points  $i - r$  and  $i + r$  (Fig. 3a). At each contour point  $i$  the angle is calculated as follows:

Let  $x_i, y_i$  be the coordinates of point  $i, i = 1:N$ . We define two vectors coming to and leaving from the point  $i$  as  $\vec{v}_{ic} = (x_i - x_{i-r}, y_i - y_{i-r})$  and  $\vec{v}_{il} = (x_{i+r} - x_i, y_{i+r} - y_i)$ . Then the angle between the two vectors is computed as:

$$\text{angle}_i = \cos^{-1} \left( \frac{\vec{v}_{ic} \cdot \vec{v}_{il}}{|\vec{v}_{ic}| \cdot |\vec{v}_{il}|} \right) \in [0^\circ, 180^\circ]$$

Afterward,  $\text{angle}_i$  is quantized to nine values and coded according to Table 1 [6, 14]. For each contour point  $i$ , a



**Fig. 3** Computation of the angle (a) and the AC sequence (b). The values are in the range 0–8 according to the coding scheme seen in Table 1

**Table 1** Angle code scheme

Degrees	(−5,5)	(5,40)	(40,50)
Code	0	1	2
Degrees	(50, 85)	(85, 95)	(95, 135)
Code	3	4	5
Degrees	(130, 140)	(140, 175)	(175, 185)
Code	6	7	8

value is assigned and therefore a sequence  $x(n)$ ,  $n = 1 \dots i \dots N$  is obtained.

The AC has been extensively coded by means of its Histogram [6] but significant information is lost in this way. However, the AC sequence is very informative with respect to the shape and complements the information of CCD. Especially in leaf image retrieval, AC has been used in the form of ACH which guarantees the rotation invariance that is otherwise difficult to achieve.

For our MESS method we adopt the AC sequence, since its detailed format complies better with our approach. An example of the AC sequence is given in Fig. 3. For the data used, the value of parameter  $r$  was set to ' $r = 10$ ' that was found experimentally to be the optimum.

### 3 The MESS method

The feature extraction and matching stages of the method are accomplished by the following three steps:

- boundary acquisition and sequence formulation
- phase space representation
- similarity measurement via WW multivariate statistical test

The first step was described in Sect. 2.1 using CCD and AC sequences. CCD sequences will serve in the description of the second and third steps of MESS method, which will be described in details. The implementation of MESS method in the case of any other sequences (e.g. AC sequences) proceeds in a similar manner.

#### 3.1 Phase space representation

The idea of transforming a 1D time series to a trajectory in a reconstructed state-space is borrowed from the field of nonlinear dynamics [23]. It has been successfully applied in many applications, including heart-signal classification [24] and biometrics [25]. The trajectory is considered as the phase space representation of underlying dynamics. Since in our case, the 1D sequence (related to shape boundary) is artificially considered as a signal from a dynamical process, the previous notions have only practical utility. We keep, however, the term phase space

representation to indicate the origin of the methodological aspects, which are incorporated in MESS as follows.

A CCD sequence  $x(n)$  like the one presented in Fig. 1 consists of a number of  $N$  points. In fact this sequence can be considered as periodic and be continued in both directions. The MESS method starts by transforming this sequence to a set of  $N$  vectors  $\mathbf{X}(\mathbf{n})$  defined as:

$$\mathbf{X}(\mathbf{n}) = \{x(n)x(n+1.\tau) \dots x(n+(m-1).\tau)\} \quad (2)$$

$$n = 1, 2, \dots, N$$

Under this new representation the time sequence  $x(n)$ , instead of a single  $N$ -dimensional vector, is represented by  $N$ ,  $m$ -dimensional vectors. Equation (2) formulates the *time-delay embedding* technique for various parameters' selection: *time delay*  $\tau$  and *embedding dimension*  $m$  [23]. The above process therefore could be considered as a reconstruction in phase space, where  $\mathbf{X}(\mathbf{n})$  depicts the multidimensional trajectory of the sequence  $x(n)$  [24], but now in the form of a set of vectors (and without the need of a time index). The number of points that form a single vector in phase space is defined by the  $m$  parameter in (2).

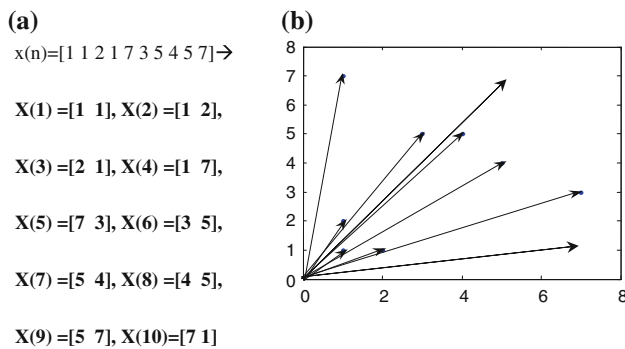
Selection of  $\tau = 1$  for the *time delay* in (2), was motivated by the aspiration to preserve a detailed description. In this way, the 1st, 2nd and  $m$ th order correlation of the contour sequence at successive index values are implicitly considered. Seen it from a different point of view, we are dealing with an overlapping, part-based representation of the original sequence in a vector space. In the case where  $m = 2$ , only the information of two consecutive points is utilized which is strongly related to the local direction of the curve, like the coded direction of Chain Code sequences. In the same way when  $m = 3$ , the angle of two consecutive directions is the coded feature of the curve. In the extreme case where the embedding dimension  $m = N$  the series of  $N$  vectors stands for the rotated versions of the original curve while, for  $m = 1$  no partial information of the contour is taken into account.

An illustrated example is given in Fig. 4, where a 10 points sequence is represented in the phase space with embedding dimension  $m = 2$ .

It should be noticed that although the complete order within a sequence is not considered by the MESS method, it is implicitly taken into account by the existing partial ordering within each  $m$ -dimensional vector.

A few properties are directly implied from the phase space representation:

- Invariance to translation is inherited from the original CCD sequence while scale invariance is easily obtained by applying an appropriate normalization step to each time series.
- The MESS method offers the advantage of producing descriptors that are insensitive to rotation. This is



**Fig. 4** The 10 point sequence  $x(n)$  (a), is represented in the 2D space with ten vectors (b). The sequence order, i.e., the trajectory in phase space, is not taken into account

deduced considering the fact that rotation does not pose any alternation to the set of  $N$ ,  $m$ -dimensional vectors that represent a given sequence.

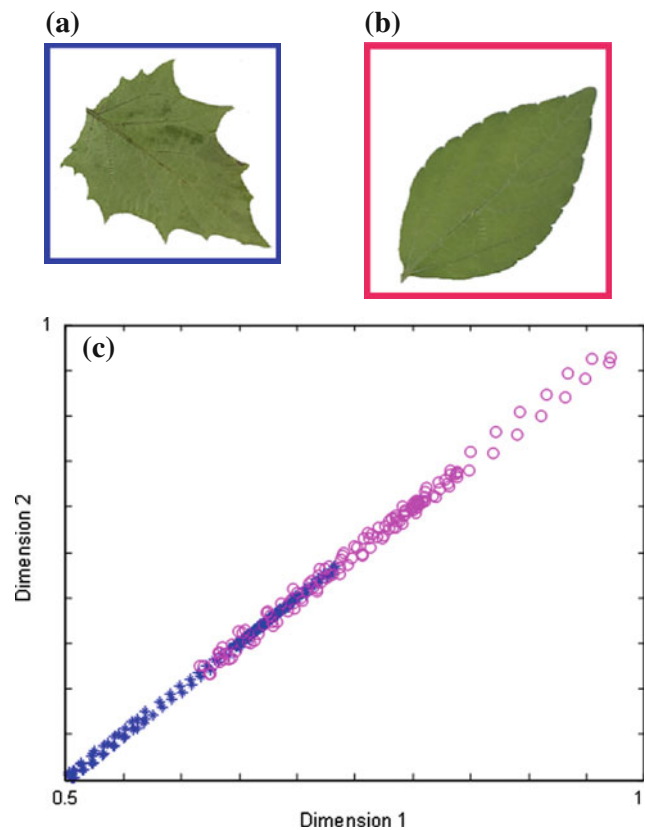
- Partial occlusion or distortion of a leaf part is tolerated due to the part-based nature of the representation, as it influences segments of the trajectory only. This means that when a limited part of the leaf contour is destroyed then a small number of points in phase space will be affected.

A realistic example for the representation of two different leaves is given in Fig. 5. Although the embedding dimension is not the optimal one, as it has been selected  $m = 2$  for visualization purposes, it is observed that the two sets of vectors are only partially overlapping and there exists a clear separation of the two distributions.

### 3.2 The optimum embedding dimension

Time-delay techniques are usually used to discover the degrees of freedom and underlying dynamics of a physical system. For this reason single or multiple time series measurements of the system are taken over time. Time-delay techniques preserve the local differential signal structure and find an embedding where its unique characteristics can be accurately mapped and used as a signature. Stated using a different terminology, based on the leaf boundary sequence, we are interested to find the minimum embedding dimension sufficient to describe the properties of the corresponding phase space.

Although finding the optimum embedding dimension is not an easy task, we can resort to manifold models for 1D signals [26]. Smooth signals are parameterized in lower dimensions. In the ideal case where the sequence values as well as their first derivatives are bounded, the reconstructed space is well defined by two parameters, i.e., two dimensions are enough to describe the curvature. However, this condition is rarely met in real sequences and the embedding dimension has to be set at a higher value. On the other



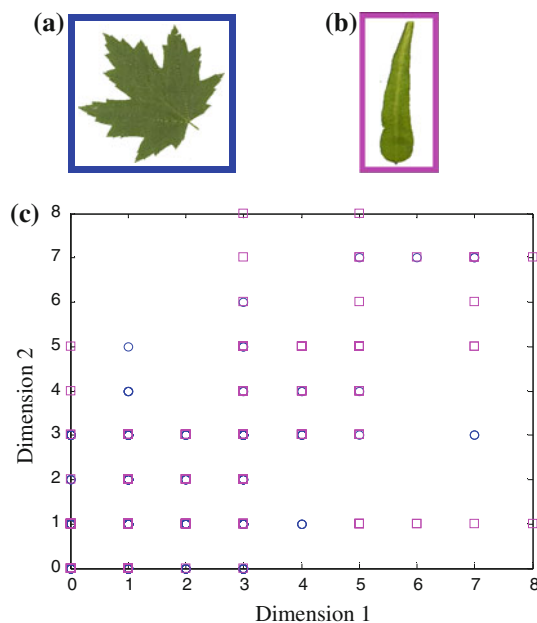
**Fig. 5** The phase space representation of CCD sequences. Two leaves **a** *Datura stramonium* and **b** *Lamium purpureum* are represented in the scatter plot **c** by the set of 2D vectors. The selected dimension  $m = 2$  is just for visualization purposes. Although the two set of vectors (stars and circles respectively) share some points, in general they are well separated, denoting two different leaves

hand, a very high dimension is not suggested due to ‘curse of dimensionality’.

Several experimental techniques for finding the correct embedding dimension from signal parameters have been proposed [27]. In general, we consider that the optimal parameter  $m_o$  depends not only on the particular leaf database but also on the image resolution. In this work this is selected, through experimentation, based on retrieval performance for the given database.

As shown in Fig. 5, we can observe that the 2D vectors form an elongated ellipsoid. It means that the reconstructed space—in case of CCD sequences—is well defined by low parameters. On the other hand, in Fig. 6 the scatter plot for two AC sequences represented by the two sets of 2D vectors, form a cloud which is spread throughout the area. This unstructured distribution reveals that the embedding dimension cannot be  $m = 2$ , and higher dimensions must be adopted to define the correct reconstructed space, capable of discriminating efficiently the two time series. Therefore, in this case, the two set of vectors are not separated.





**Fig. 6** The phase space representation of AC sequences. The leaves **a** *Acer saccharinum* and **b** *Galium aparine* are represented in the scatter plot **c** by the set of 2D vectors. The selection of two dimensions is just for visualization purposes. The two set of vectors (circles and squares respectively) are neither well described nor well separated. A higher dimension should be employed

In practice, the optimum dimension is set experimentally in terms of the best retrieval performance for a given database. Such results will be shown in Sect. 4.

### 3.3 The shape similarity measure

Matching two shapes very often involves a correspondence stage which is of high complexity. In the present work, the representation for the  $N$  points sequence provides  $N$  vectors in the  $m$ -dimensional space. The comparison of two different sequences involves the estimation of the similarity between the two corresponding multivariate point sets. The problem is solved by resorting to multivariate statistical approaches appropriately used to provide similarity indices. Several statistical indices have been proposed in the past, as a means to measure similarity between vector distributions. These distances (for instance, the Kolmogorov–Smirnov distance (KS) the Chi-square test ( $\chi^2$ -statistic), etc. [28]), measure how unlikely it is that one distribution is drawn from the population represented by the other.

The Wald–Wolfowitz (WW) multivariate statistical test originally proposed to judge whether two sets of multivariate observations come from the same distribution, is adopted here to compare the two vector sets. The WW-test has also been directly incorporated for retrieval processes [32] in large image libraries [29], with the great advantage

of being, by design, suitable for dealing with multivariate distributions. The suggested test assumes the two sets of multidimensional points as distributions, and consequently it is not influenced by the order of the points. Therefore, rotated versions of points lead to the same similarity measurement.

The above test, dealing with the nonparametric “Multivariate Two-Sample Problem” [22], is the multivariate extension of the classical statistical one-dimensional test of Wald and Wolfowitz. A brief description of its operation is as follows: Given two sets of vectors  $\{\mathbf{x}_i\}$  and  $\{\mathbf{y}_i\}$  the overall minimal spanning tree (MST) is constructed and the  $W$  statistic quantifies if the two sets are evenly placed in the branches of MST. Applied to leaf recognition using the MESS method, the more positive the value of  $W$  the more similar the two sequences and the corresponding shapes are. The  $W$ -index is computed upon the  $R$  value which counts the number of subsequences on MST graph with the same identity. A more detailed description of the WW-test is given in Appendix A. A detailed demonstration of the WW-test is given in Fig. 7.

At this point two significant observations should be made: (1) the two sets to be compared can differ in size (as the only restriction is the same dimensionality), (2) the  $W$ -index carries an absolute meaning and therefore, can be utilized in comparisons under different embedding conditions.

## 4 Evaluation of the MESS method

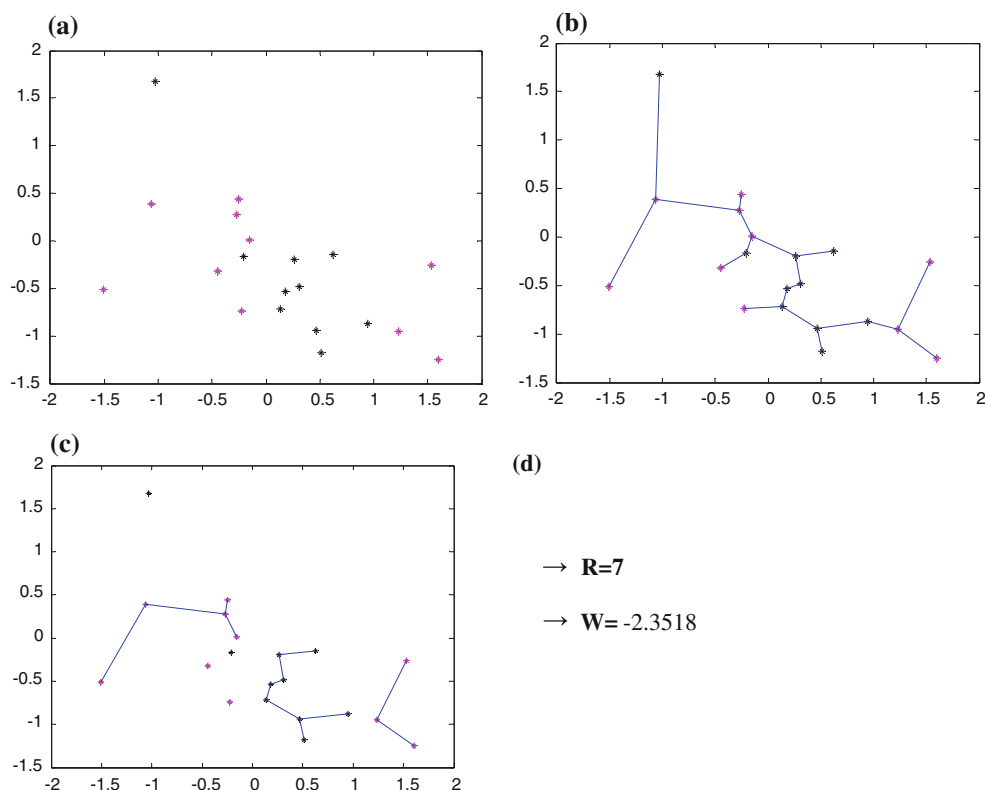
### 4.1 The leaves image database

The employed experimental set of leaves was extracted from the site: <http://herbarium.cs.columbia.edu/data.php>. It includes 40 species (found in the Greek area) with 25 leaves from each species. Therefore, the whole database consists of 1,000 leaf images, a typical number used in retrieval studies [6, 8, 10]. Although this number is not very large it is considered sufficient to demonstrate the improvement on classification that is achieved over existing methods. The database consists of some leaves with occlusions or distortions. Although quantification of such measurements is not possible, selection of these leaves is essential to obtain more realistic results. All leaf species used, are named in Table 2.

#### 4.1.1 Image preprocessing

Image preprocessing is an essential step to image analysis. Initially, a thresholding procedure is applied to each image. Then, all images are turned into binary. For denoising

**Fig. 7** Visual configuration of the algorithmic procedure of the multivariate WW-test for two equally sampled distributions: **a** position the samples in the corresponding feature space, **b** construct the overall MST, **c** remove the edges ending at different sample and **d** finally define the test statistic  $R$ , and the value  $W$



purposes, a median filtering is applied to each image, to eliminate the existence of veins. Afterward, an edge detection algorithm is performed, and finally the contour trace algorithm computes the leaf boundary coordinates in an anti-clockwise manner.

#### 4.2 Evaluation criteria: description

Three frequently used performance measures in shape retrieval applications, that is recognition rate, correct result rate and precision and recall diagram were employed in the evaluation procedure. In all evaluation measures, every single leaf of the database is used as query image. To simulate a retrieval problem, the query image has been excluded from the database at each trial, so all methods described below belong to the leave-one-out experimentation processes.

- recognition rate
- The recognition rate is defined as the ratio of the correct hits to the highest possible number of correct hits among the retrieved images. Every single leaf of the database is assumed as query and a number of the most similar images are retrieved. The final recognition rate is the average of all queries.
- correct result rate
- In the correct result rate a hit is assumed if one correct image is found in the retrieved set. Actually this

**Table 2** Database species

Acer_saccharinum	Ligustrum_vulgare
Acer_negundo	Lamium_purpureum
Aster_cordifolius	Lonicera_japonica
Asplenium_platyneuron	Melilotus_alba
Aster_divaricatus	Ostrya_virginiana
Acer_saccharum	Platanus_occidentalis
Ailanthus_altissima	Passiflora_lutea
Carpinus_caroliniana	Portulaca_oleracea
Crataegus_crus-galli	Parthenocissus_quinquefolia
Catalpa_bignonioides	Quercus_montana
Cercis_canadensis	Quercus_alba
Cynanchum_laeve	Quercus_rubra
Datura_stramonium	Saxifraga_virginiensis
Eupatorium_serotinum	Silene_stellata
Galium_aparine	Smilax_glauca
Hibiscus_laevis	Tilia_americana
Humulus_japonicus	Ulmus_rubra
Ipomoea_lacunosa	Urtica_dioica
Liriodendron_tulipifera	Viburnum_prunifolium
Lysimachia_nummularia	Xanthium_strumarium

experiment evaluates the time needed (expressed in retrieved images) for correct response.

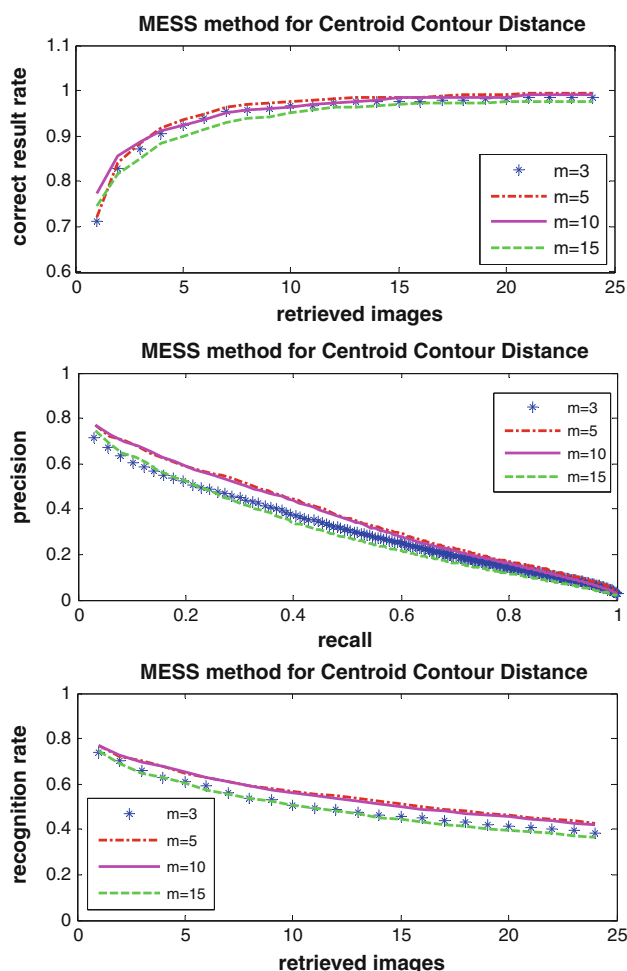
- precision and recall diagrams

- To evaluate a retrieval process, a commonly used performance measurement is the precision and recall pair. Precision is defined as the ratio of the correct hits to the total number of retrieved images. The final precision of retrieval is the average precision of all queries. Precision measures the accuracy of the retrieval.

Recall is defined as the ratio of the correct hits to the correct database's hits. Recall measures the robustness of the retrieval performance.

#### 4.3 Setting the optimum embedding dimension

As it was mentioned in Sect. 3.3, an indication of the optimum embedding dimension is provided by experimentation in retrieval performance for various values of  $m$ . For that purpose all three evaluation criteria mentioned above were performed for both distances (CCD, AC). The derived results are given in Figs. 8 and 9, for CCD and AC distances, respectively.



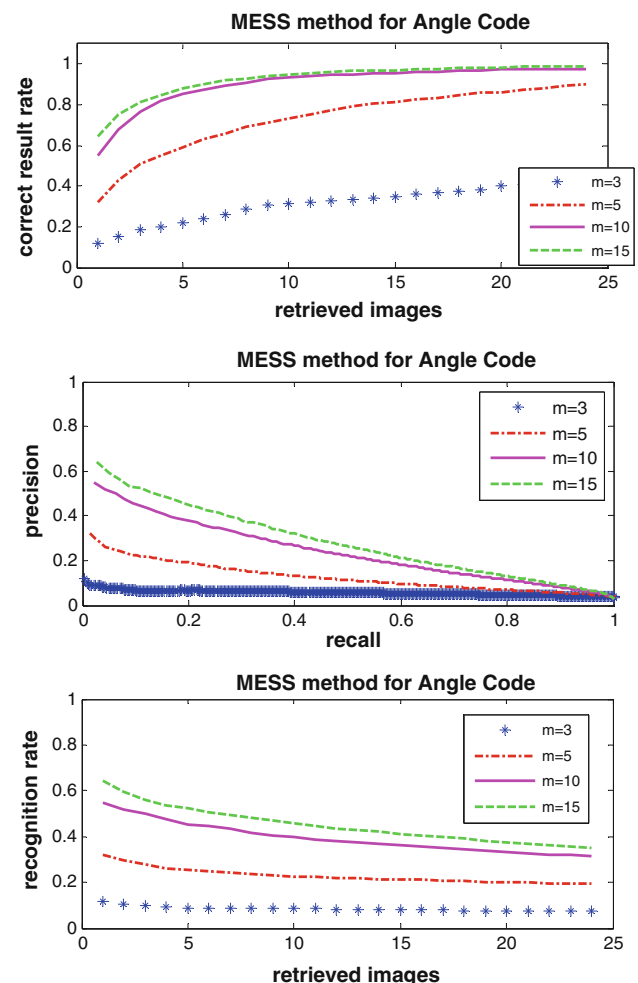
**Fig. 8** Performance evaluation for various values of embedding dimension  $m$ , using MESS method for the CCD sequence

As we can observe, the performance of the MESS method for CCD distance is better for low embedding dimensions. On the other hand, in the case of the MESS method using the AC distance, higher dimensions lead to better results. Therefore, indicative optimum values which are adopted for the embedding dimension in our experimentation are  $m = 5$  and  $m = 15$  for CCD and AC sequences, respectively.

These values of  $m$  are in line with our discussion in Sect. 3.3 where smooth trajectories are represented with fewer parameters in the phase space [8]. The AC distance is actually a measure of the contour tangent, i.e., the local derivative which has much more variability than the CCD.

#### 4.4 Retrieval performance: experimentation

In this paragraph experimental results are shown indicating that under all measures discussed in Sect. 4.2, the MESS method improves shape matching and the overall retrieval



**Fig. 9** Performance evaluation for various values of  $m$ , using MESS method for AC sequence



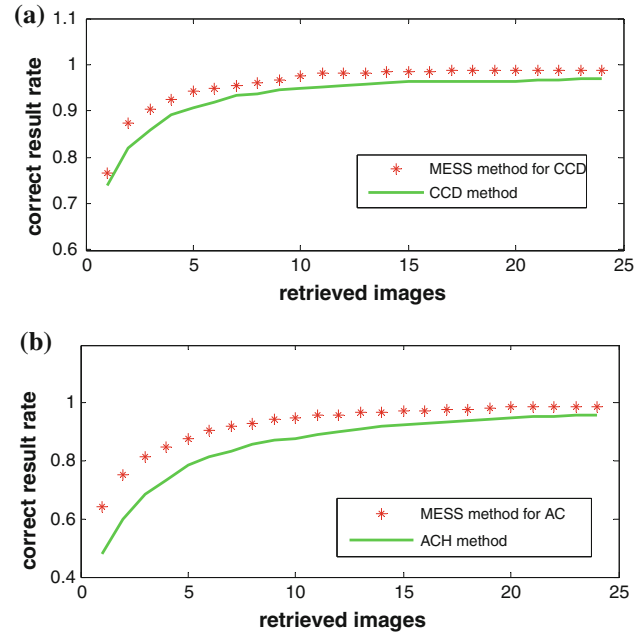
performance. In addition, a comparison between CCD and AC sequences is shown, demonstrating that CCD sequences are more informative than the AC sequences. Finally, the proposed method is compared with two common techniques, Dynamic Time Warping (DTW) [16, 17] and Inner Distance Shape Context (IDSC) [19].

Results are given in Figs. 10, 11, 12 for all the evaluation criteria discussed in Sect. 4. These results strongly indicate that in both cases (i.e., CCD and AC) the MESS method improves the retrieval performance.

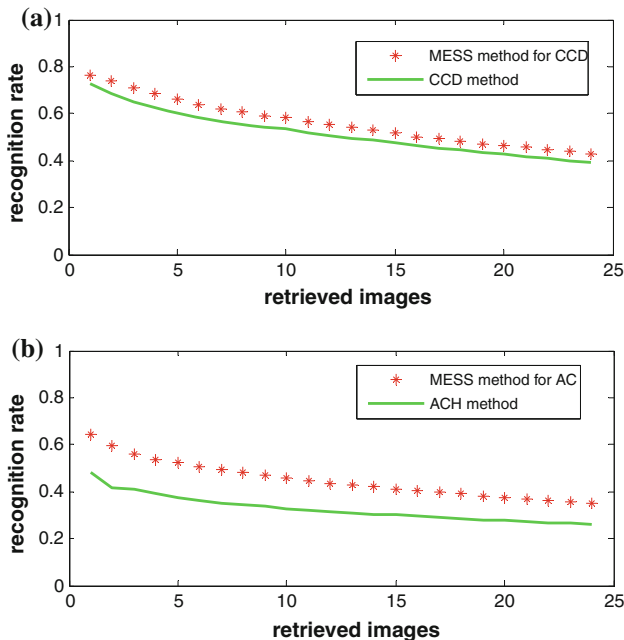
Direct comparison of the CCD and AC sequences is given in Fig. 13. All three evaluation criteria were used and the results prove that the CCD sequences perform better indicating that the shape contour is better described by CCD than the AC sequence. This is expected since the CCD distance includes both global information, related to the leaf area and shape as well as local, related to contour details. Therefore, in their combined use, which is expected to further improve the retrieval performance, the CCD should be emphasized by setting to it a higher weight [6].

A subset of the database <http://herbarium.cs.columbia.edu/data.php> is used to compare the MESS method with the state-of-art methods such as DTW and IDSC. In Table 3 comparative results of the above methods are shown. The outperformance of MESS over DTW respecting recognition rate and mean time processing is obvious. Concerning IDSC + DP (Dynamic Programming) it is shown that this method leads to very high recognition rate. However, acquisition of these retrieval results is followed by the

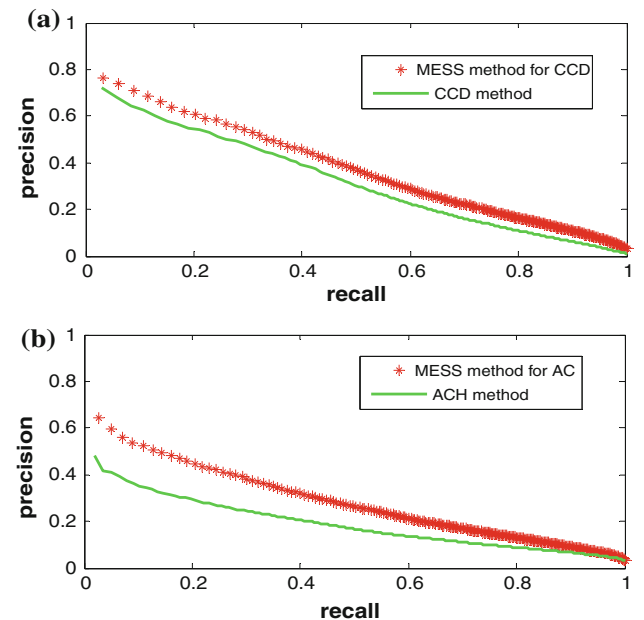
disadvantage of a high computational cost, as it is 3.5 times slower than the MESS method. This is due to the several stages that consist of the complete method.



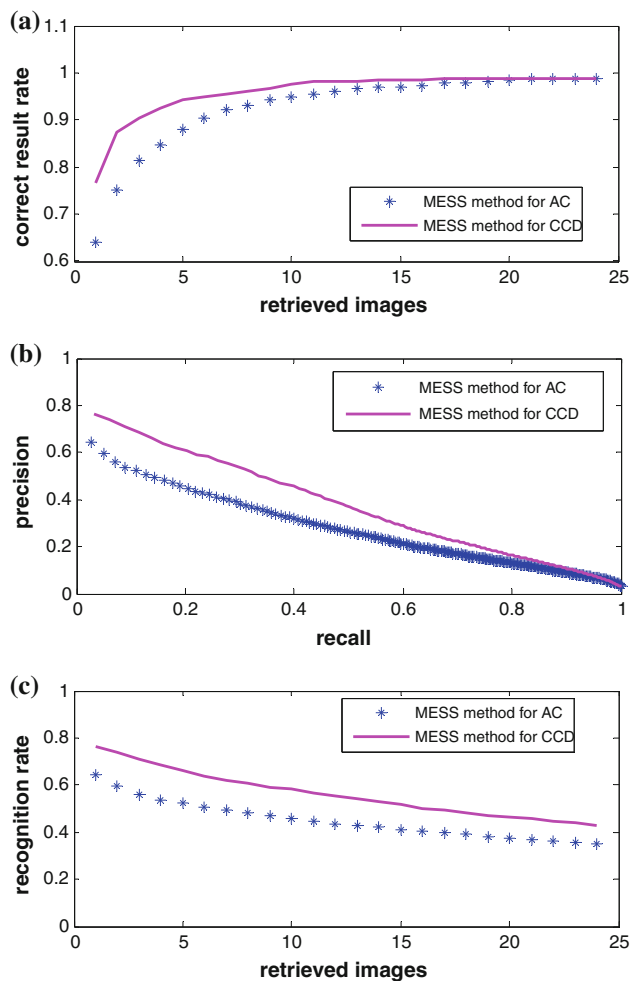
**Fig. 11** Correct result rate for the Centroid Contour Distance (a) and Angle Code (b) sequences. The MESS method outperforms the classical methods. The embedding dimension in MESS method is  $m = 5$  and  $m = 15$ , respectively



**Fig. 10** Recognition rate for the CCD (a) and AC (b) sequences. The improvement with MESS method is noticeable. The embedding dimension in MESS method is  $m = 5$  for CCD and  $m = 15$  for AC



**Fig. 12** Precision versus recall for the Centroid Contour Distance (a) and Angle Code (b) sequences. The MESS method performance is better than the other two methods. The embedding dimension in MESS method is  $m = 5$  and  $m = 15$ , respectively



**Fig. 13** Performance evaluation for the two sequences using our MESS method. Correct result rate (a), precision versus recall (b), and recognition rate (c), were used for evaluation. In all three cases the CCD sequences performs better which proves that it is more informative than AC sequence. The embedding dimension is  $m = 5$  for CCD and  $m = 15$  for AC

**Table 3** Recognition rates among common techniques and mean time processing comparison, over top 24 retrieved leaves

Method	Recognition rate	Normalized mean time processing
DTW using CCD	0.5898	2
MESS method using CCD	<b>0.6222</b>	1
DTW using AC	0.5213	2
MESS method using AC	<b>0.5855</b>	1
IDSC + DP	0.7770	3.5

#### 4.5 Computational complexity

The complexity of the MESS algorithm for feature extraction is relatively low, as it only depends on the

number of  $N$  points of the original sequence. As far as  $x(n)$  time series is obtained, transformation of the sequence into a set of  $N$   $m$ -dimensional vectors in  $R^m$ , needs only an index defining the values used to form vectors in the reconstructed phase space. Consequently, its complexity is  $O(N)$ . In our experiments, the value  $N$  is in the order of 200.

In the matching process, the WW-test is nonparametric and for  $N$  points its complexity is  $O(N^2)$  or  $O(N \log N)$  depending on the algorithm used to implement the MST, resulting in low complexity of the overall system. The tradeoff between down sampling and discriminating accuracy is also an option which could reduce the computational cost a great deal.

In our experiments, using Matlab 2009b, on an Intel core i7 950 8G PC, for two given sequences with  $N = 200$ , extraction of vectors in phase space followed by WW distance computation between them takes about 7 and 40 ms, respectively. Comparative results for the overall performance are given in Table 3.

## 5 Discussion

The MESS method is a two-stage procedure which finds similarity between 1D shape representation sequences. The first stage is the sequence embedding in a higher phase space and the second one is the efficient measure of similarity through a multivariate statistical test. The MESS method demonstrates that any sequence representing a shape is better portrayed if reconstructed in an appropriate phase space.

The incorporation of MESS technique in leaf image retrieval proved very promising. The presented method improves a great deal the usage of existing features like the CCD and AC sequences, in content-based shape-image retrieval applications. Further advantages of the MESS method are its tolerance to several distortions of the shape, as well as its ability to retain invariance with respect to rotation. However, when severe occlusions take place then the computed centroid is shifted away from the actual centroid of the shape and the performance might be degraded.

Finally, the suggested method can be regarded as a stand-alone method, but it can be easily incorporated to already existing methods to enhance the overall performance. Phase space representation is suitable for the fusion of various descriptors originating from differently parameterized time series or descriptors that encode diverse information, such as AC, CCD.

The application to different shape classification problems like character-recognition and hand writing style identification is almost straightforward.

## 6 Originality and contribution

Our paper addresses the problem of shape similarity by introducing a two-stage approach : a novel feature extraction step followed by a suitable and efficient matching procedure. The Multidimensional Embedding Sequence Similarity (MESS) method is evaluated over the challenging problem of leaf-shape recognition and retrieval.

The feature extraction step aims at the shape representation of leafs through 1D sequences. There exist many methods that employ (or result to) a time series as a shape descriptor. In our work the descriptor is built, from a scalar sequence derived from the leaf outline, by forming a set of ‘time-lagged’, properly ‘embedded’ vectors. This set of multidimensional points in phase space retains most of the properties of the original sequence. By exploiting the descriptive power of this encoding scheme, that translates different shapes into distinct scatter plots in a common phase space, we can achieve enhanced retrieval performance.

In the matching process, the Wald–Wolfowitz statistical test is adopted. The test quantifies the similarity between leaves by comparing sets of multidimensional points in phase space. With the least effort, shape similarity becomes invariant to translation, scale and rotation transformations and robust to occlusions.

The originality of the suggested paper stems from the coupling of feature extraction stage with the appropriate matching process. To the best of our knowledge, shape representation in phase space using ‘time-delay embedding’ method in conjunction with WW-test for similarity measurement between leaf shapes is considered for the first time in leaf image retrieval.

## Appendix A. The multivariate Wald–Wolfowitz test

The Wald–Wolfowitz multivariate statistical test [22] assesses the commonality between two different sets of multivariate observations.

The output of the test can be expressed as the probability that two point-samples are coming from the same distribution. Its great advantage is that it is model-free and this stems from the graph–theoretic origin of the test, which is actually based on the concept of MST graph [29, 30]. The MST is a spanning tree containing exactly  $(N - 1)$  edges, for which the sum of edge weights is minimum. In WW-test, the graph is built over points in Rd: a single node corresponds to every given point, the weight associated with every possible edge is the corresponding interpoint Euclidean distance. WW-test can be used to test the hypothesis  $H_0$ , whether any two given multidimensional point samples  $\{X_i\}_{i=1:m}$  and  $\{Y_i\}_{i=1:n}$  are coming

from the same multivariate distribution. A great advantage is that no a priori assumption about the distribution of points in the two samples is a prerequisite [31].

In the first step, the sample identity of each point is not taken into account and the MST of the overall sample is constructed.

Then, based on the sample identities of the points, a test statistic  $R$  is computed.  $R$  is the total number of runs, while a run is defined as a consecutive sequence of identical sample identities. Rejection of  $H_0$  is for small values of  $R$ . The null distribution of this statistic has been derived, based on combinatorial analysis. It has been shown that the quantity

$$W = \frac{R - E[R]}{\sqrt{\text{Var}[R]}}$$

approaches (asymptotically) the standard normal distribution, while the mean  $E[R]$  and variance  $\text{Var}[R/C]$  of  $R$  depend on the sizes  $m$  and  $n$  of the two point-samples and can be computed using the following analytical expressions:

$$E[R] = \frac{2mn}{N} + 1$$

$$\text{Var}[R/C] = \frac{2mn}{N(N-1)} \times \left\{ \frac{2mn - N}{N} + \frac{C - N + 2}{(N-2)(N-3)} [N(N-1) - 4mn + 2] \right\}$$

where  $N = m + n$ ,  $C$  is the number of edge pairs sharing a common node defined as  $C = \frac{1}{2} \sum_{i=1}^N d_i(d_i - 1)$  and  $d_i$  is the degree of the  $i$ th node.

The above analysis enables the computation of the significance level (and  $p$  value) for the acceptance of the hypothesis  $H_0$ .

## References

1. Nam Y, Hwang E, Kim D (2005) CLOVER: a mobile content-based leaf image retrieval system. In: Digital libraries: implementing strategies and sharing experiences. LNCS, vol 3815, pp 139–148. doi:[10.1007/11599517\\_16](https://doi.org/10.1007/11599517_16)
2. Belhumeur PN, Chen D, Feiner S, Jacobs DW, Kress WJ, Ling H, Lopez I, Ramamoorthi R, Sheorey S, White S, Zhang L (2008) Searching the world’s herbaria: a system for visual identification of plant species. In: ECCV, part IV. LNCS, vol 5305, pp 116–129. doi:[10.1007/978-3-540-88693-8\\_9](https://doi.org/10.1007/978-3-540-88693-8_9)
3. Kebapci H, Yanicoglou B, Unal G (2010) Plant image retrieval using color, shape and texture features. Comput J. doi:[10.1093/comjnl/bxq037](https://doi.org/10.1093/comjnl/bxq037)
4. Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recogn 37(1):1–19
5. Mehtre BVM, Kankanhalli MS, Lee WF (1997) Shape measures for content based image retrieval: a comparison. Inf Process Manag 33(3). doi:[10.1016/S0306-4573\(96\)00069-6](https://doi.org/10.1016/S0306-4573(96)00069-6)

6. Wang Z, Chi W, Feng D (2003) Shape based leaf image retrieval. *IEE Proc Vision Image Signal Process* 150(1):34–43. doi:[10.1049/ip-vis:20030160](https://doi.org/10.1049/ip-vis:20030160)
7. Mokhtarian F, Abbasi S (2004) Matching shapes with self-intersections: application to leaf classification. *Proc IEEE Trans Image* 13(5):653–661. doi:[10.1109/TIP.2004.826126](https://doi.org/10.1109/TIP.2004.826126)
8. Nam Y, Hwang E, Kim D (2008) A similarity-based leaf image retrieval scheme: joining shape and venation features. *Comput Vis Image Underst* 110(2):245–259. doi:[10.1016/j.cviu.2007.08.002](https://doi.org/10.1016/j.cviu.2007.08.002)
9. Lee C-L, Chen S-Y (2006) Classification of leaf images. *Int J Imaging Syst Technol* 16(1):15–23
10. Cabalero C, Aranda M (2010) Plant species identification using leaf image retrieval. In: *Proceedings of the ACM international conference on image and video retrieval*, Xi'an, China, pp 327–334. doi:[10.1145/1816041.1816089](https://doi.org/10.1145/1816041.1816089)
11. Casanova D, Junior J Sa, Bruno Od (2009) Plant leaf identification using Gabor wavelets. *Int J Imaging Syst Technol* 19(3):236–243. doi:[10.1002/ima.v19.3](https://doi.org/10.1002/ima.v19.3)
12. Park J, Jun E, Nam Y (2008) Utilizing venation features for efficient leaf image retrieval. *J Syst Softw* 81(1):71–82. doi:[10.1016/j.jss.2007.05.001](https://doi.org/10.1016/j.jss.2007.05.001)
13. Shen Y, Zhou C, Lin K (2005) Leaf image retrieval using a shape based method. In: *IFIP international federation for information processing 187/2005*, pp 711–719. doi:[10.1007/0-387-29295-0\\_77](https://doi.org/10.1007/0-387-29295-0_77)
14. Peng HL, Chen SY (1997) Trademark shape recognition using closed contours. *Pattern Recogn Lett* 18(8):791–803. doi:[10.1016/S0167-8655\(97\)00050-0](https://doi.org/10.1016/S0167-8655(97)00050-0)
15. Gdalyahu Y, Weinshall D (1999) Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. In: *IEEE transactions on PAMI*, vol 21(12), pp 1312–1328. doi:[10.1109/34.817410](https://doi.org/10.1109/34.817410)
16. Santosh KC (2010) Use of dynamic time warping for object shape classification through signature. *Kathmandu Univ J Sci Eng Technol* 6:33–49. doi:[10.3126/kuset.v6i1.3308](https://doi.org/10.3126/kuset.v6i1.3308)
17. Tak YS (2007) A leaf image retrieval scheme based on partial dynamic time warping and two level-filtering. In Wei D, Miyazaki T, Paik I (eds) *Proceedings of the 7th IEEE international conference on computer and information technology—CIT 2007*. IEEE Computer Society, Los Alamitos, pp 633–638. doi:[10.1109/CIT.2007.158](https://doi.org/10.1109/CIT.2007.158)
18. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape context. *IEEE Trans Pattern Anal Mach Intell* 24(4):509–522. doi:[10.1109/34.993558](https://doi.org/10.1109/34.993558)
19. Liang H, Jacobs D (2007) Shape classification using the inner-distance. *IEEE Trans Pattern Anal Mach Intell* 29(2):286–299. doi:[10.1109/TPAMI.2007.41](https://doi.org/10.1109/TPAMI.2007.41)
20. Wu SG, Bao FS, Xu EY, Wang Y-X, Chang Y-F, Xiang Q-L (2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In: *IEEE 7th international symposium on signal processing and information technology* <http://flavia.sourceforge.net/>, pp 11–16. doi:[10.1109/ISSPIT.2007.4458016](https://doi.org/10.1109/ISSPIT.2007.4458016)
21. Chang C, Hwang S, Buehrer D (1991) A shape recognition scheme based on relative distances of feature points from the centroid. *Pattern Recogn* 24(11):1053–1063. doi:[10.1016/0031-3203\(91\)90121-K](https://doi.org/10.1016/0031-3203(91)90121-K)
22. Friedman JH, Rafsky LC (1979) Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann Stat* 7(4):697–717. doi:[10.1214/aos/1176344722](https://doi.org/10.1214/aos/1176344722)
23. Abarbanel HDI (1996) *Analysis of observed chaotic data*. Springer Verlag, New York
24. Chan H-L, Fang S-C, Chao P-K, Wang C-L, Wei J-D (2009) Phase-space reconstruction of electrocardiogram for heartbeat classification. In: *WC IFMBE proceedings*, vol 25(4), pp 1234–1237. doi:[10.1007/978-3-642-03882-2\\_327](https://doi.org/10.1007/978-3-642-03882-2_327)
25. Laskaris N, Zafeiriou S, Garefa L (2009) Use of random time-intervals (RTIs) generation for biometric verification. *Pattern Recogn* 42(11):2787–2796. doi:[10.1016/j.patcog.2008.12.028](https://doi.org/10.1016/j.patcog.2008.12.028)
26. Peyre G (2009) Manifold models for signals and images. *Comput Vis Image Underst* 113(2):249–260. doi:[10.1016/j.cviu.2008.09.003](https://doi.org/10.1016/j.cviu.2008.09.003)
27. Cao L (1997) Practical method for determining the minimum embedding dimension of a scalar time series. *Phys D* 110(1–2):43–50
28. Rubner Y, Puzicha J, Tomasi C, Buhmann JM (2001) Empirical evaluation of dissimilarity measures for color and texture. *Comput Vis Image Underst* 84:25–43
29. Theoharatos Ch, Laskaris N, Economou G, Fotopoulos S (2005) A generic scheme for color image retrieval based on the multivariate Wald–Wolfowitz test. *IEEE Trans Knowl Data Eng* 17(6):808–819. doi:[10.1109/TKDE.2005.85](https://doi.org/10.1109/TKDE.2005.85)
30. Zahn CT (1971) Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans Comput C* 20(1). doi:[10.1109/T-C.1971.223083](https://doi.org/10.1109/T-C.1971.223083)
31. Dijkstra EW (1959) A note on two problems in connection with graph. *Numer Math* 1(1):269–271. doi:[10.1007/BF01386390](https://doi.org/10.1007/BF01386390)
32. Theoharatos Ch, Laskaris N, Economou G, Fotopoulos S (2004) A similarity measure for color image retrieval and indexing based on the multivariate two sample problem. In: *Proceedings of EUSIPCO*, Vienna, Austria.