

Can Walking and Measuring along Chord Bunches Better Describe Leaf Shapes?

Bin Wang¹, Yongsheng Gao¹, Changming Sun², Michael Blumenstein³, John La Salle⁴

¹School of Engineering, Griffith University, Australia

²CSIRO Data61, PO Box 76, Epping, NSW 1710, Australia

³FEIT, University of Technology Sydney, Australia

⁴Atlas of Living Australia, CSIRO National Research Collections Australia

bin.wang@griffith.edu.au, yongsheng.gao@griffith.edu.au, changming.sun@csiro.au,
michael.blumenstein@uts.edu.au, john.lasalle@csiro.au

Abstract

Effectively describing and recognizing leaf shapes under arbitrary deformations, particularly from a large database, remains an unsolved problem. In this research, we attempted a new strategy of describing shape by walking along a bunch of chords that pass through the shape to measure the regions trespassed. A novel chord bunch walks (CBW) descriptor is developed through the chord walking that effectively integrates the shape image function over the walked chord to reflect the contour features and the inner properties of the shape. For each contour point, the chord bunch groups multiple pairs of chord walks to build a hierarchical framework for a coarse-to-fine description. The proposed CBW descriptor is invariant to rotation, scaling, translation, and mirror transforms. Instead of using the expensive optimal correspondence based matching, an improved Hausdorff distance encoded correspondence information is proposed for efficient yet effective shape matching. In experimental studies, the proposed method obtained substantially higher accuracies with low computational cost over the benchmarks, which indicates the research potential along this direction.

1. Introduction

The human vision system is efficient and has the strength of recognizing objects only by their shapes. For machine vision, shape is originally represented as a binary mask or the contour is further extracted from the binary pattern after the object is segmented from the image. The task of shape analysis is to capture the geometrical information that is independent of the transformational (scaling, rotation, mirror, articulation, etc.) effects for the ultimate goal of object recognition. However, from the available binary mask or contour representation of the shape to the successful recognition of the object, there exist many challenges such as rigid and non-rigid deformation of the object, noises, partial occlusion, intra-class variations, and

efficient demands of various real time applications.

Developing a method to address all the above issues may be unrealistic. For example, part-based algorithm is much effective for handling articulation deformation [31]. It is however highly sensitive to segmentation errors especially for noisy data. Those methods that can deal with partial occlusion may not be suitable for real time application due to their expensive computing cost. Therefore, developing a method which can balance the demands of many aspects instead of satisfying all the demands is achievable.

Automatic plant leaf identification is a significant applications of computer vision and has received considerable studies [1][3][5][6][9][10][37][38] in recent years. However due to the small interclass difference (see Fig. 5 and Fig. 6), the large intra-class variations (see Fig. 7) and particularly the natural self-overlap (see the middle leaf image shown in the first row of Fig. 1) of the leaf shapes, the issue of effectively and efficiently recognizing the leaf shapes is still not well-addressed. For example, we try the state-of-the-art approaches, shape contexts [15], inner distance shape contexts [9], and height function [36] respectively to distinguish the leaf shapes shown in the first row of Fig. 1 and they all consider that the middle one is more similar to the right one than the left one.

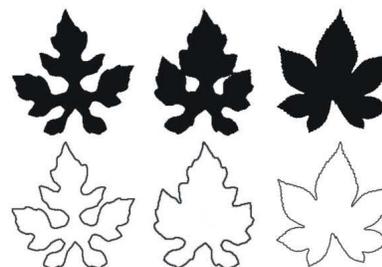


Fig. 1. Three leaf images (the left two leaves belong to the same species) and their corresponding outer contours. If only from their outer contours, the middle one which is self-overlapped may be identified as the species of the right one.

This work aims to develop a shape description and matching method for effectively and efficiently recognizing leaf shape. A hierarchical framework, using chord bunch

walks (CBW) descriptors, is proposed. In this framework, the chord integrals over the shape region, termed the chord walks, are used to capture the inner structural information of the shape, and the angles formed by the chord pair walk are included for reflecting the boundary properties. The multiple chord pair walks emanated from a common contour point are grouped to provide a coarse-to-fine description for the shape. The proposed CBW descriptors are invariant to rotation, scaling, translation, and mirror transforms and possess the potential to handle the self-overlapped leaf shapes. The experimental study demonstrates its better performance over the state-of-the-art approaches.

2. Related Work

Many shape descriptors have been proposed and can be categorized into global approaches and local approaches. The form one extracts the contour features, usually represented as a 1-D function, or the shape region features, usually represented as a 2-D function and summary them to yield a feature vector, or a sequence of features as a whole shape description. A shape distance can then be measured in a norm metric, e.g. L_2 distance. The classical descriptor for this group is the Zernike moments (ZM) [2] generated by projecting the shape image function on a set of orthogonal basis functions defined in the interior of a unit circle. Yap et al. [12] proposed a novel set of 2D transforms, termed polar harmonic transforms (PHTs), for rotation invariant features extraction. Like the Zernike moments, PHTs are based on a set of orthogonal projection bases. However its computing complexity is significantly lower than that of the Zernike moments. Recently, the well-known Radon transform and its generalization, Trace transform, have been extensively used for shape description and matching [7][11][18][29][30][33]. These methods finely capture the inner structural information via line integrals over the shape region. More recently, Hong and Soatto [34] develop an invariant and robust shape descriptor using a series of isotropic kernels.

The above mentioned methods derive the shape descriptors from the whole shape region. There are also another group of global methods extract the features only from the contour. The popular ones are various spectral methods, like Fourier descriptor [8][21] and wavelet descriptor [22]. Recently, Hu et al. [3] propose a novel descriptor, termed, multiscale distance matrix (MDM) [3], which uses the distances between the contour points at multiple scales to build a matrix for reflecting the spatial relationship of the contour points, where the distance between the contour points can be Euclidean or other metric such as inner distance [9]. To provide a more effective while efficient method for shape retrieval, the more recent work, hierarchical string cuts [10], is proposed to extract a group of geometrical features, termed string cuts, for characterizing the spatial configuration of the curve

segments relative to their strings.

For local approaches, a descriptor is extracted for each visual primitive such as point, contour fragments and polygon's edge of the shape. The local descriptors associated with the primitives of the compared shapes are used for building correspondence between them and the resulted matching cost is taken for the shape distance measure. Shape contexts [15] builds a histogram for each contour point to encode the information of the relative spatial distribution (distance and orientation) of all the contour points to it. The inner-distance shape contexts [9] replaces the Euclidean distance with the inner-distance, defined as the length of the shortest path between landmark points within the shape silhouette, for effectively capturing the partial structure of the shape. Distance sets [16] proposes a rich local descriptor for an image point using the spatial configuration of feature points surrounding that point, where the spatial arrangement is characterized by the set of the distance between the that point with the feature points of its neighbourhood. Height function [36] uses the height values (perspective distance) from all the contour sample points to the tangent line across a given contour point to yield local descriptor for capturing the geometric relationships of the contour points to the given point.

The above methods focus on extracting the spatial arrangement information of the contour points of the shape. There are also many methods that pay attentions on characterizing the curvature property or bend potential at the shape contours. Alajlan et al. [17] propose to use the area of the triangles formed by the contour points to measure the convexity/concavity of each point at different scales. Integral invariants [13] introduces a class of functionals of integral operators conducted on the shape for making the resulted descriptors robust to high-frequency noise and small deformations. It has been effectively applied to leaf identification [19]. Contour flexibility [23] makes effort to represent the deformable potential at each contour point and argues that both local and global features can be finely extracted by this descriptor. Many other similar methods can be found in [24][25][27][28].

3. The Proposed Method

We first define chord walk pair for encoding the local discriminative information of the shape and then introduce the chord bunch walks to present a coarse to fine description for the shape. The invariances derived from the chord bunch walks are presented and discussed. They are finally used for shape distance measure.

3.1. Chord Pair Walk

Given a binary shape image $f(x, y)$, the shape regions (denoted as D) are formed by a subset of pixels in the image plane \mathbb{R}^2 . Let Ω be the outer contour of the shape enclosing all pixels in D , which can be represented as an arc-length

parameterization form [23]: $z(t) = (\bar{x}(t), \bar{y}(t)), t \in [0,1)$, where $(\bar{x}, \bar{y}) \in \Omega$. Since Ω is a closed contour, we have $z(t+1) = z(t)$ and $z(t-1) = z(t)$.

For a contour point $p = z(t)$, we walk along a pair of chords of $\overline{pp'}$ and $\overline{pp''}$ whose end points are $p' = z(t+s)$ and $p'' = z(t-s)$ respectively, where $s \in (0, \frac{1}{2}]$. The paths of walks sometimes fall inside the shape region D , and sometimes fall outside the shape region D . We record the lengths of walking inside D (\hat{l}_t^s and \hat{l}_t^{-s}) and the lengths of walking outside D (\check{l}_t^s and \check{l}_t^{-s}), which can be mathematically expressed as

$$\begin{cases} \hat{l}_t^s = \int_0^{l^s} f(\bar{x}(t) + \tau \cos \theta_t^s, \bar{y}(t) + \tau \sin \theta_t^s) d\tau \\ \hat{l}_t^{-s} = \int_0^{l^{-s}} f(\bar{x}(t) + \tau \cos \theta_t^{-s}, \bar{y}(t) + \tau \sin \theta_t^{-s}) d\tau \end{cases} \quad (1)$$

and

$$\begin{cases} \check{l}_t^s = \int_0^{l^s} (1 - f(\bar{x}(t) + \tau \cos \theta_t^s, \bar{y}(t) + \tau \sin \theta_t^s)) d\tau \\ \check{l}_t^{-s} = \int_0^{l^{-s}} (1 - f(\bar{x}(t) + \tau \cos \theta_t^{-s}, \bar{y}(t) + \tau \sin \theta_t^{-s})) d\tau \end{cases}, \quad (2)$$

where l^s and l^{-s} , θ_t^s and θ_t^{-s} are the lengths and orientations of the chord pair $\overline{pp'}$ and $\overline{pp''}$, respectively.

The above defined \hat{l}_t^s and \hat{l}_t^{-s} are the integrals of the shape image function over the chord pairs $\overline{pp'}$ and $\overline{pp''}$ while \check{l}_t^s and \check{l}_t^{-s} are the integrals of the complement function of the shape image over the chord $\overline{pp'}$ and $\overline{pp''}$. Then we use them to generate a quintuplet as

$$W_t^s = (\check{l}_t^{-s}, \check{l}_t^{-s}, \sin(\theta_t^{-s} - \theta_t^s), \hat{l}_t^s, \hat{l}_t^s). \quad (3)$$

We call it chord pair walk. A graphical illustration of the chord pair walk is given in Fig. 2.

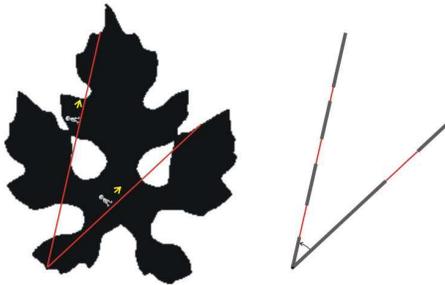


Fig. 2. An example illustration of the concept of chord walk pair. Left: a pair of chords emanated from a contour point are put on the shape image plane for walking through. Right: the obtained chord walk pair (the portions in bold represent the paths of the chord walking inside the shape region while the remains are the paths of chord walking outside the shape region).

The chord pair walk W_t^s is a local descriptor to capture the geometrical features of the neighborhood of the contour point $p = z(t)$. Its first two elements and last two elements reflect the property of the left neighborhood and the right neighborhood of the contour point p , respectively, and its middle element $\sin(\theta_t^{-s} - \theta_t^s)$ measures the curvature property at the contour point in which the value of $\sin(\theta_t^{-s} - \theta_t^s)$ being > 0 , < 0 and $= 0$ indicates the convex, concave and flat properties respectively. The chord pair walk W_t^s captures the rich geometrical information around the contour point p including not only the contour features, but also the structure information of the inner region. This desirable characteristic make the proposed chord pair walk descriptor obviously outperform those approaches such as integral invariants [13], triangle features [6][7], arch height [25], and contour flexibility [23] which only capture the curvature feature or the bendable property of the contour.

3.2. Chord Bunch Walks

In the previous section, for a given contour point $p = z(t)$, we start from it to walk towards its right neighbor contour point $p' = z(t+s)$ and its left neighbor contour point $p'' = z(t-s)$, respectively and obtain a chord pair walk descriptor W_t^s for capturing the geometrical features of its neighborhood. A single chord pair walk may be not discriminative enough, while a group of chord pair walks can largely improve the discriminative ability of the descriptor. Varying the parameter s and let it take the values $2^{-1}, \dots, 2^{-K}$, we can obtain K chord pair walks emanating from the common point p and group them to form a chord bunch walks (CBW) defined as

$$\tilde{W}_t^{(K)} = \bigcup_{k=1}^K \{W_t^s, s = 2^{-k}\}. \quad (4)$$

An example illustration of the concept of CBW is shown in Fig. (3).

Now, for each contour point $p = z(t)$, we have built a CBW descriptor to encode the multiscale shape geometrical information associated with it. In fact, the CBW descriptor is obtained by grouping the chord pair walks of different size s . Further, observing the Fig. 3, we can find that the whole geometrical structure of the chord bunch varies with its located contour point which indicates that the spatial relationships between the chord walk pairs in the chord bunch are also useful for shape discrimination. To include this valuable information to the CBW descriptor, we further extract the spatial relationship features from the chord bunch as follows.

For two adjacent chord pair walks, $\{W_t^s, s = 2^{-k}\}$ and $\{W_t^s, s = 2^{-(k+1)}\}$, in the chord bunch, a pair of angles $\vartheta_{t,k}^{(R)}$ and $\vartheta_{t,k}^{(L)}$ between them can be obtained by

$$\vartheta_{t,k}^{(R)} = \theta_t^{2^{-k}} - \theta_t^{2^{-(k+1)}}, \quad \vartheta_{t,k}^{(L)} = \theta_t^{-2^{-(k+1)}} - \theta_t^{-2^{-k}}. \quad (5)$$

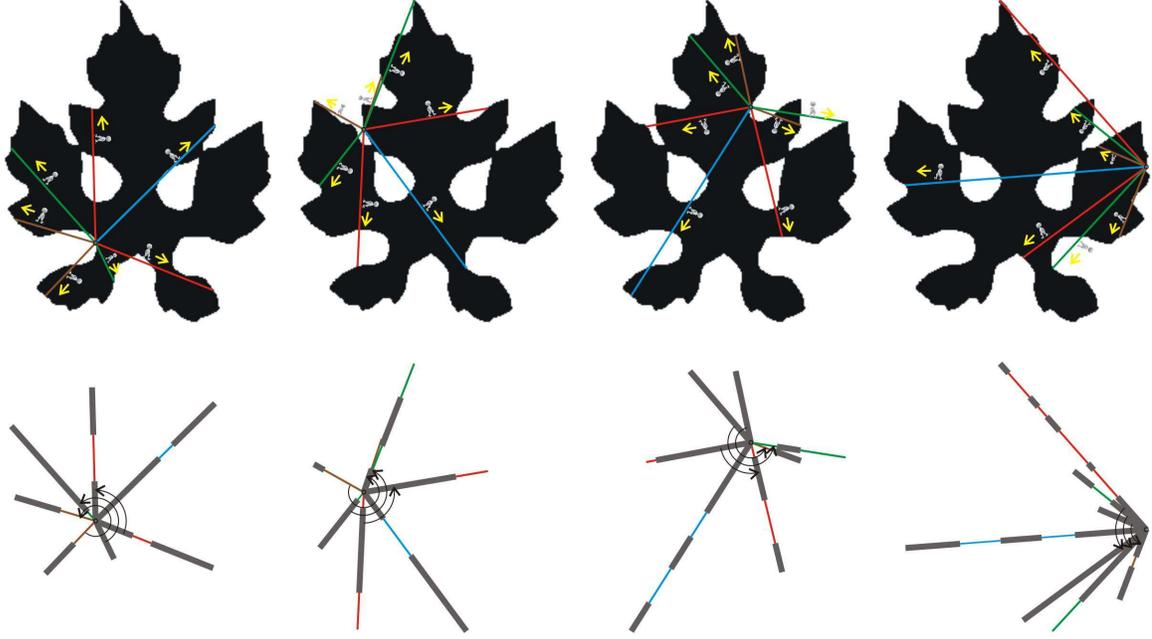


Fig. 3. An example illustration of the concept of chord bunch walks. The upper figure: the chord bunches of size $K=4$ emanated from different contour points are put on the image plane for walking through (the chord pair walks are marked by the same color). The figure below: the obtained chord bunch walk descriptors.

Since there are total K chord pair walks in CBW, $K-1$ angle pairs are available. In Fig. 4, we give an example to illustrate these angle pairs and use them to generate a vector as

$$\delta_t^K = \bigcup_{k=1}^{K-1} \{\sin(\vartheta_{t,k}^{(R)} - \vartheta_{t,k}^{(L)})\}, \quad (6)$$

which is used to reflect the whole spatial arrangement of the chord bunch. Then the CBW descriptor defined in Eq. (4) can be further extended by

$$\tilde{W}_t^{(K)} = \left(\bigcup_{k=1}^K \{W_t^s, s = 2^{-k}\} \right) \cup \left(\bigcup_{k=1}^{K-1} \{\sin(\vartheta_{t,k}^{(R)} - \vartheta_{t,k}^{(L)})\} \right). \quad (7)$$

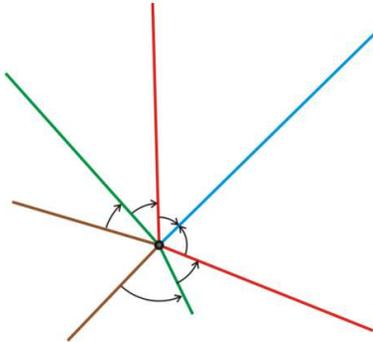


Fig. 4. An example illustration of the angle pairs $\vartheta_{t,k}^{(R)}$ and $\vartheta_{t,k}^{(L)}$ derived from the chord bunch shown on the top left of Fig. 3.

At this time, we have built a descriptor $\tilde{W}_t^{(K)}$ of $5K-3+K-1=6K-4$ dimensions for each contour point $p = z(t)$.

3.3. Invariance and Normalization

A good descriptor is expected to be translation, scale, rotation, and mirror invariant. In the following, we discuss the invariant properties of the proposed CBW descriptor. Since a translation of the shape function $f(x, y)$ results in the same translation for all the points of the shape region, including the contour points. Thus the proposed CBW descriptor has the intrinsic invariance to translation.

When the shape $f(x, y)$ is rotated by angle θ , the length features $\hat{l}_t^{-s}, \tilde{l}_t^{-s}, \hat{l}_t^s$ and \tilde{l}_t^s of the chord pair walk W_t^s are invariant and only its direction angles θ_t^s and θ_t^{-s} are changed by $\theta_t^s + \theta$ and $\theta_t^{-s} - \theta$ respectively. So, the angle $\theta_t^{-s} - \theta_t^s$ is invariant and W_t^s is accordingly invariant. The angle pair $\vartheta_{t,k}^{(R)}$ and $\vartheta_{t,k}^{(L)}$ defined in Eq. (5) are likewise invariant to the rotation. So, the CBW descriptor is inherently invariant to rotation. For the scaling, all the angle related features in Eq. (3) and Eq. (5) are invariant, and only the length features $\hat{l}_t^s, \tilde{l}_t^s, \hat{l}_t^{-s}, \tilde{l}_t^{-s}$ are changed. We derive the normalized version $\hat{\hat{l}}_t^s$ of \hat{l}_t^s to scaling as

$$\hat{\hat{l}}_t^s = \frac{\hat{l}_t^s - \min_{0 \leq t < 1} \{\hat{l}_t^s\}}{\max_{0 \leq t < 1} \{\hat{l}_t^s\} - \min_{0 \leq t < 1} \{\hat{l}_t^s\}}. \quad (8)$$

The other three features $\tilde{l}_t^s, \hat{l}_t^{-s}, \tilde{l}_t^{-s}$ can be normalized to be scale invariant in the same way.

For the mirror transform, without losing generality, we assume that the shape is reflected about the y -axis, i.e., its image function and contour parameter equation are changed into $f(x, y) = f(-x, y)$ and $z(t) = (-\bar{x}(1-t), \bar{y}(1-t))$, $t \in [0, 1]$, respectively. The chord pair walk $W_t^s = (\hat{l}_t^{-s}, \check{l}_t^{-s}, \sin(\theta_t^{-s} - \theta_t^s), \hat{l}_t^s, \check{l}_t^s)$ will be changed by

$$\begin{aligned} \hat{W}_t^s &= (\hat{l}_{1-t}^s, \check{l}_{1-t}^s, \sin(\pi - \theta_{1-t}^s - \pi + \theta_{1-t}^{-s}), \hat{l}_{1-t}^{-s}, \check{l}_{1-t}^{-s}) \\ &= (\hat{l}_{1-t}^s, \check{l}_{1-t}^s, \sin(\theta_{1-t}^{-s} - \theta_{1-t}^s), \hat{l}_{1-t}^{-s}, \check{l}_{1-t}^{-s}). \end{aligned} \quad (9)$$

Namely, the parametrized contour is flipped and in each chord pair walk, the first two elements are swapped with the last two elements, and the middle element is invariant.

Next, we check what happens to the angle pair $\vartheta_{t,k}^{(R)}$ and $\vartheta_{t,k}^{(L)}$ defined in Eq. (5). From their definitions, we can conclude that their mirror transformed versions are

$$\begin{aligned} \hat{\vartheta}_{t,k}^{(R)} &= \pi - \theta_{1-t}^{-2-k} - \pi + \theta_{1-t}^{-2-(k+1)} = \theta_{1-t}^{-2-(k+1)} - \theta_{1-t}^{-2-k} \\ &= \vartheta_{1-t,k}^{(L)} \end{aligned}$$

and

$$\hat{\vartheta}_{t,k}^{(L)} = \pi - \theta_{1-t}^{-2-(k+1)} - \pi + \theta_{1-t}^{-2-k} = \theta_{1-t}^{-2-k} - \theta_{1-t}^{-2-(k+1)} = \vartheta_{1-t,k}^{(R)}$$

So, from Eq. (6), we have

$$\begin{aligned} \delta_t^K &= \prod_{k=1}^{K-1} \{\sin(\hat{\vartheta}_{t,k}^{(R)} - \vartheta_{t,k}^{(L)})\} = \prod_{k=1}^{K-1} \{\sin(\vartheta_{1-t,k}^{(L)} - \vartheta_{1-t,k}^{(R)})\} \\ &= -\delta_{1-t}^K. \end{aligned} \quad (10)$$

Namely, the mirror transform only changes the sign of δ_t^K .

Then we can conduct following normalizations:

$$\begin{aligned} W_t^s &= (\max(\hat{l}_t^{-s}, \check{l}_t^s), \max(\check{l}_t^{-s}, \hat{l}_t^s), \sin(\theta_t^{-s} \\ &\quad - \theta_t^s), \min(\hat{l}_t^{-s}, \check{l}_t^s), \min(\check{l}_t^{-s}, \hat{l}_t^s)) \end{aligned} \quad (11)$$

and

$$\delta_t^K = \prod_{k=1}^{K-1} \{\sin(\vartheta_{t,k}^{(R)} - \vartheta_{t,k}^{(L)})\}. \quad (12)$$

For removing the effect of flipping the contour, we treat the contour as a set of points and ignore the order relationship of the contour points. Through the above processes, the CBW has become completely invariant to scaling, rotation, translation and mirror transforms.

3.4. Implementation and Time Complexity

Given a binary digital image $I_{N \times N}$ of an object, like those approaches in [3][9][13][17][20][36], we adopt the sample scheme of uniform spacing and sampling its outer contour into a sequence of T points: $p_t(x_t, y_t)$, $t = 1, \dots, T$, where $T = 2^{K+1}$ and K is a given integer, t is the index of the sample point according to the order along the contour in counter-clockwise direction, (x_t, y_t) are the coordinates of the sample point p_t and we have $p_{t+T} = p_t$.

For each contour point p_t , we group K chord pairs

$\{\overline{p_t p_{t+s}}, \overline{p_t p_{t-s}}\}_{s=2^1, \dots, 2^K}$ and for each chord $\overline{p_t p_{t+s}}$, we uniformly sample it into $V-1$ points $\{p'_i(x'_i, y'_i), i = 1, \dots, V-1\}$, where $V = \max(|x_t - x_{t+s}|, |y_t - y_{t+s}|)$, namely, the chord $\overline{p_t p_{t+s}}$ is segmented into V parts of equal length. Let $p_t(x_t, y_t)$ be $p'_0(x'_0, y'_0)$ and $p_{t+s}(x_{t+s}, y_{t+s})$ be $p'_V(x'_V, y'_V)$. Since a line segment $p'_i p'_{i-1}$ is fallen in the shape region, if and only if $f(x'_i, y'_i) f(x'_{i-1}, y'_{i-1}) = 1$, the variable \hat{l}_t^s defined in Eq. (1) can be calculated by

$$\hat{l}_t^s = \left(\frac{\sum_{i=1}^V f(x'_i, y'_i) f(x'_{i-1}, y'_{i-1})}{V} \right) l, \quad (13)$$

where l is the length of the chord $\overline{p_t p_{t+s}}$ and can be calculated by the Euclidean distance between the point p_t and the point p_{t+s} . Another variable \check{l}_t^s defined in Eq. (2) can be simply calculated by $l - \hat{l}_t^s$. Since $1 \leq V \leq N$, calculating Eq. (13) requires time $O(N)$. While calculating other variable \check{l}_t^s and θ_t^s only requires time $O(1)$. So, calculating a chord pair walk defined in Eq. (3) requires time $O(N)$. The time complexity of calculating the chord bunch walks defined in Eq. (4) is $O(KN)$. Since calculating Eq. (5) only requires time $O(1)$ and time $O(K)$ is enough for calculating Eq. (6), calculating the extended version of CBW descriptor defined in Eq. (7) requires time $O(KN + K) = O(KN)$. So, the time cost of calculating the CBW for all the T contour points is $O(TKN)$. Additional cost for normalizing the CBW of each contour point using Eqs. (8), (11) and (12) is $O(TK)$. So, the cost of calculating the final version of CBW for all the contour points is still $O(TKN)$. Since $K = \log_2 T - 1$, it can be rewritten as $O(TN \log_2 T)$.

3.5. Shape Dissimilarity Measure

We have built a CBW descriptor $\tilde{W}_t^{(K)}$ for each contour point p_t , $t = 1, 2, \dots, T = 2^{K+1}$. It is natural to conduct point matching, namely matching the contour point of one shape to the one of another shape by comparing their corresponding CBW descriptors for shape dissimilarity measure. The cost of comparing the CBW descriptors of two points is measured using the L_1 distance between them. Many methods try to find an optimal correspondence between the contour points of the compared shapes for the shape distance measure. A limitation of this scheme is its expensive computational cost (usually more than $O(T^3)$, where T is the number of the contour points). In additional, optimal correspondence based shape dissimilarity measure does not always work well for those shapes (e.g. compound leaf shapes) whose local details are hard to be matched in pairs. Here, we propose an improved Hausdorff distance for an economical shape dissimilarity measure.

Given two point sets $P = \{p_1, p_2, \dots, p_M\}$ and $Q = \{q_1, q_2, \dots, q_N\}$, where M and N are their sizes respectively. the Hausdorff distance between them is defined as

$$H(P, Q) = \max(h(P, Q), h(Q, P)), \quad (14)$$

where

$$h(P, Q) = \max_{p_i \in P} \left(\min_{q_j \in Q} (d(p_i, q_j)) \right) \quad (15)$$

and $d(p_i, q_j)$ is a norm metric, for example L_1 distance. This is the definition of the original version of Hausdorff distance. Dubuisson and Jain [14] proposed a modified version by changing the definition of $h(P, Q)$ as

$$h(P, Q) = \frac{1}{M} \sum_{p_i \in P} \min_{q_j \in Q} d(p_i, q_j). \quad (16)$$

Unlike those one-to-one correspondence based distance measures, Hausdorff distance is a local correspondence matching scheme, so there exists the case of multiple points in one set to be matched to the same point in another set. To enhance the performance the Hausdorff distance, we proposed an improved version in which the global correspondence are somewhat considered.

For each point $q_k \in Q$, those points $p_i \in P, i = 1, \dots, M$ who are mapped to it are collected to form a set

$$\sigma(q_k) = \{p_i: \pi(p_i) = q_k, i = 1, \dots, M\},$$

where $\pi(p_i)$ is defined as

$$\pi(p_i) = \arg \min_{q_j \in Q} (d(p_i, q_j)).$$

Let $\bar{Q} = \{q_j: |\sigma(q_j)| > 1, j = 1, \dots, N\}$ be the set of the points from the set Q that are matched by multiple points from set P and $\omega(q_j) = \frac{1}{|\sigma(q_j)|} \sum_{p_i \in \sigma(q_j)} d(p_i, q_j)$ be the average cost over all the mappings to the point q_j from the set P , where $|\sigma(q_j)|$ is the cardinality of the set $\sigma(q_k)$. Then an extra cost of considering the case of multiple mappings to a same point is defined as

$$\epsilon(P, Q) = \sum_{q_j \in \bar{Q}} \left((|\sigma(q_j)| - 1) \cdot \omega(q_j) \right). \quad (17)$$

The above defined extra cost $\epsilon(P, Q)$ encodes some global correspondence information and is used to modify the Eq.

(16) by

$$h(P, Q) = \frac{1}{M} \left(\epsilon(P, Q) + \sum_{p_i \in P} \min_{q_j \in Q} d(p_i, q_j) \right). \quad (18)$$

Then we use the Eq. (14) and Eq. (18) to measure the dissimilarity between two shapes.

4. Experimental Results

To evaluate the effectiveness of the proposed CBW approach, an experimental study has been conducted on two public available leaf datasets, MEW2012 and ICL leaf datasets and their performances (both accuracy and computation speed) are compared against six state-of-the-art methods. Among them, the Shape Context [15] and Inner Distance Shape Context [9] are classical shape analysis approaches, which are widely used benchmarks for performance comparison. The Height Function [36], MDM [3] and HSC [10] are recently proposed contour based shape analysis methods with state-of-the-art performances on the MPEG-7 CE-1 database and leaf databases.

4.1. MEW2012 Leaf Dataset

Middle European Woody plants (MEW2012) [37] is a public available leaf dataset which contains native or frequently cultivated trees and shrubs of the central Europe Region. There are a total of 9745 leaf images which belong to 153 species with at least 50 samples in each one. One typical sample for each category is shown in Fig. 5. It can be seen that so many species are in the dataset and very minor differences between them make the retrieval task very challenging.

Mean average precision (MAP)¹[4] is a standard measure for evaluating the performance of information retrieval systems, which is widely used in evaluating systems for image retrieval [5][6][32][35], speech index [40], and video retrieval [26][39]. Each leaf image in the dataset is taken as a query to retrieve the similar ones from all the samples in

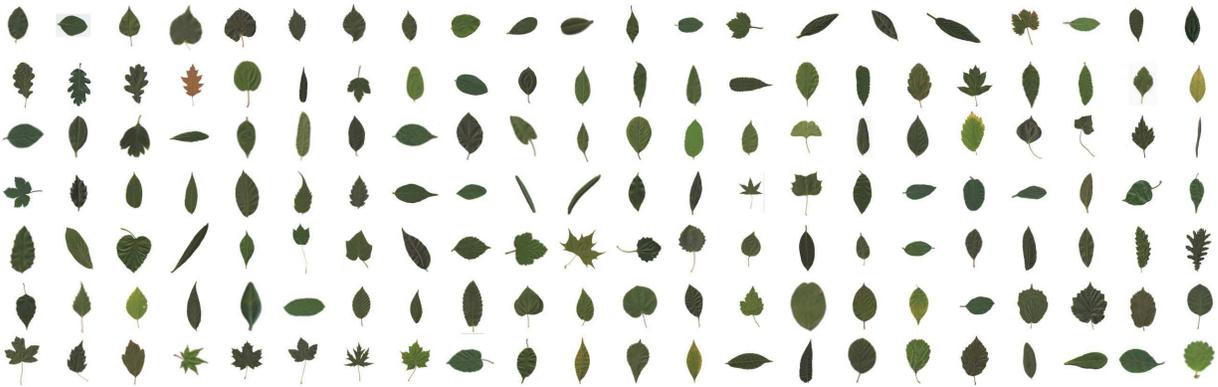


Fig. 5. 153 typical samples from the MEW2012 leaf dataset [37], one sample is shown for each species.

¹ https://en.wikipedia.org/wiki/Information_retrieval.



Fig. 6. 220 typical samples from the ICL leaf dataset [3], one sample is shown for each species.

the dataset. The MAP scores obtained by the proposed method and six state-of-the-art approaches are summarized in Table 1. We also record the computation time of matching one query with all the 9745 samples including the feature extraction time for the query shape, and the average retrieval time for all the comparative methods are reported in this table.

Table 1. MAP score on the MEW2012 leaf dataset.

Algorithm	MAP Score (%)	Average Retrieval Time (s)
IDSC [9]	45.36	85.23
SC+DP [15]	47.94	91.47
MDM-CD-RM [3]	39.14	1.56
MDM-ID-RA [3]	33.71	1.89
Height Function [36]	49.76	97.81
HSC [10]	54.98	0.17
Proposed CBW	62.05	6.59

It can be seen that in this challenging dataset, the proposed method achieves the best MAP score of 62.05% which is 7.07% higher than the second best method HSC and is more than 16% over the other compared methods. These exciting results demonstrate that the proposed CBW has the powerful ability to distinguish plant species by leaf shape features. Comparing the retrieval speed of the proposed method with the other methods, we can find that the proposed method is more than 12 times faster than the methods IDSC [9], SC+DP [15], and Height Function [36]. Although the proposed method is slower than the three other methods, it better balances the effectiveness and efficiency for the retrieval task.

4.2. ICL Leaf Dataset

To further examine the potential application ability of the proposed CBW method to plant leaf identification. Another public leaf dataset called ICL dataset [3] which is built by Intelligent Computing Laboratory (ICL) at the Institute of Intelligent Machines, Chinese Academy of Sciences, is used. This dataset is very large. Currently, it contains 16851 samples from 220 species with each individual species having from 26 to 1078 samples. To the best of our knowledge, it may be the largest dataset currently available for research. One typical samples for each species are shown in Fig. 6. Example leaf shapes are also given in Fig. 7 for showing the rich intra-class varies. All the leaves of the ICL dataset are grown in China which is different from MEW2012 dataset whose leaves are from middle Europe. So, there are many different species between the MEW2012 dataset and the ICL dataset. In particular, there exists 11 species of compound leaves which contain multiple leaflets in the ICL leaf dataset (see the first row of Fig. 8, and the numbers of their samples are 63, 97, 49, 69, 90, 41, 48, 54, 26, 62, and 55 respectively).

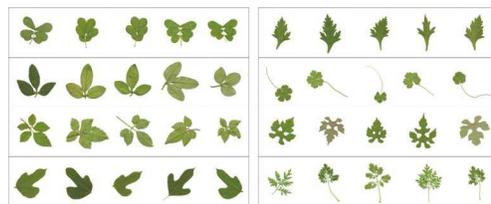


Fig. 7. Example leaves from the ICL leaf dataset for showing intra-class varies including natural deformation and self-intersections.

METHOD												AVG.
MDM-CD-RM	48.6	46.0	88.8	30.3	29.9	98.5	33.2	65.2	40.8	84.2	39.4	55.0
MDM-ID-RA	46.4	37.5	66.2	23.2	29.9	72.9	26.7	52.9	43.3	49.1	28.1	43.3
HSC	55.2	75.4	100.0	53.4	34.6	96.3	62.7	77.4	69.3	73.9	66.6	69.5
Proposed CBW	67.8	78.9	100.0	68.0	37.3	97.6	70.4	81.2	82.9	85.1	70.6	76.3

Fig. 8. The average MAP scores (%) of the proposed CBW method and the other three compared methods on retrieving compound leaves in the ICL leaf dataset.

Many studies [3][25][38] have been done on the ICL dataset. However, they only chose part of leaf samples (no more than 6000) for testing and some of them include pretreatment by cutting off the footstalks from all the leaf samples (called clean leaf samples) in the dataset. To maintain the challenges of the original ICL dataset, all the samples in the ICL dataset are taken for testing and no pretreatment is done in our experiment. Since the computational costs of the point-based matching methods, IDSC+DP [9], SC+DP [25], and Height Function [36], are too high. The fastest method of them also requires more than 24 days to finish all the retrieval tasks (16851 times of retrieval). So we have to omit them in this experiment. The MAP scores of the remaining methods are summarized in Table 2.

Table 2. MAP score on the ICL leaf dataset.

Algorithm	MAP Score (%)	Average Retrieval Time (s)
IDSC+DP [9]	~	~
SC+DP [15]	~	~
MDM-CD-RM [3]	37.45	3.10
MDM-ID-RA [3]	34.18	3.92
Height Function [36]	~	~
HSC [10]	51.65	0.26
Proposed CBW	55.02	9.63

As can be observed, the proposed CBW method consistently keeps the leading position among all the competing methods. In this challenge dataset, the proposed CBW method achieves 55.02% of MAP score which is 3.37%, 17.57%, and 20.84% higher than HSC [10], MDM-CD-RM [3], and MDM-CD-RA [3] methods. We are also concerned with the retrieval results of the compound leaves. So, we particularly show the average MAP scores of all the compound leaf species in Fig. 8, where the reported MAP score for each species is the average of the MAP scores of all the samples included. As can be observed that almost for all the compound leaf species the proposed CBW method achieves the best MAP scores among all the competing methods and only for the sixth species (from left to right in Fig. 8), the proposed method is no more than 1.0% lower than the MDM-CD-RM [3] method. While comparing the average MAP scores over all the species of all the competing methods, the proposed method achieves 76.3% of MAP score which is 6.8%, 21.3%, and 33.0% higher than

HSC [10], MDM-CD-RM [3], and MDM-ID-RA [3], respectively. These comparative results clearly indicate the outstanding performance of the proposed method.

It is worth mentioning that the proposed CBW method report very low MAP scores (no more than 40%) on the fifth compound leaf species relative to the other compound species. We check the ICL dataset and find that among all the 90 samples of this species, the footstalks of the 46 samples are kept, while the footstalks of the other 44 samples are removed. This situation is really a great challenge because the serious occlusion occurred. Although only 37.3% MAP score is achieved for the proposed method, it is still more than 2.7% over the other methods.

We also report the retrieval speed of the proposed method and the other three compared methods in Table 2. It can be seen that the method HSC [10] works much faster than the proposed method and the two versions of MDM method [3], and the proposed method has the same order of retrieval speed as the two versions of MDM method [3]. Considering the balance between the effectiveness and efficiency, the proposed method is however desirable for leaf image retrieval.

5. Conclusion

We have presented a novel approach, termed chord bunch walks, for shape description and matching. Each chord pair walk integrates the shape image function over the walked chord to capture the contour features and reflect the inner properties of the shape. The chord bunch walks are used to set up a hierarchical framework for providing a coarse-to-fine shape description. It has several desirable properties: (1) Invariance with respect to the group of transformation including translation, rotation, scaling, and reflection; (2) Effectively recognizing the self-overlapped leaf shapes; (3) Finely characterizing the local shape geometric properties at multiple scales which is useful to identify the leaf shapes under large intra-class variants and small interclass difference. The substantially higher accuracies with low computational cost over the state-of-the-art methods on two challenging leaf shape image datasets indicate the research potential along this direction.

Acknowledgements

This work was supported in part by the Australian Research Council under Grant DP140101075 and the National Natural Science Foundation of China under Grant 61372158.

References

- [1] S.G. Wu, F.S. Bao, E.Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. In Proc. IEEE International Symposium on Signal Processing and Information Technology, pp. 11-16, 2007.
- [2] A. Khotanzad and Y.H. Hong. Invariant image recognition in Zernike moments. IEEE Trans. Pattern Anal. Machine Intell., 12(5): 250-261, 1990.
- [3] R. Hu, W. Jia, H. Ling, and D. Huang. Multiscale distance matrix for fast plant leaf recognition. IEEE Trans. Image Processing, 21(11): 4667-4672, 2012.
- [4] C.D. Manning, P. Raghavan, and H. Schütze. Introduction to information retrieval. Cambridge University Press 2008.
- [5] H. Laga, S. Kurtak, A. Srivastava, M. Golzarian, and S. Miklavcic. A Riemannian elastic metric for shape-based plant leaf classification. In Proc. Int. Conf. Digital Image Computing: Techniques and Applications (DICTA), pp. 1-7, 2012.
- [6] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet. A shape-based approach for leaf classification using multiscale triangular representation. In Proc. the 3rd ACM Int. Conf. Multimedia Retrieval, pp. 127-134, 2013.
- [7] T.V. Hoang and S. Tabbone. Invariant pattern recognition using the RFM descriptor. Pattern Recognition, 45(1): 271-284, 2012.
- [8] D. Zhang and G. Lu. Study and evaluation of different Fourier methods for image retrieval. Image Vis. Comput., 23(1): 33-49, 2005.
- [9] H. Ling and D.W. Jacobs. Shape classification using the inner-distance. IEEE Trans. Pattern Anal. Machine Intell., 29(2): 286-299, 2007.
- [10] B. Wang and Y. Gao. Hierarchical string cuts: a translation, rotation, scale and mirror invariant descriptor for fast shape retrieval. IEEE Trans. Image Processing, 23(9): 4101-4111, 2014.
- [11] Y.W. Chen and Y.Q. Chen. Invariant description and retrieval of planar shapes using radon composite features. IEEE Trans. Signal Processing, 56(10): 4762-4771, 2008.
- [12] P.-T. Yap, X. Jiang, and A.C. Kot. Two-Dimensional polar harmonic transforms for invariant image representation. IEEE Trans. Pattern Analysis and Machine Intell., 32(7): 1259-1270, 2010.
- [13] S. Manay, D. Cremers, B.-W. Hong, A.J. Yezzi and S. Soatto. Integral invariants for shape matching. IEEE Trans. Pattern Anal. Mach. Intell., 28(10): 1602-1617, 2006.
- [14] M.-P. Dubuisson and A. Jain. A modified Hausdorff distance for object matching. In Proc. The Twelfth Int. Conf. Pattern Recognition, pp. 566-568, 1994
- [15] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Machine Intell., 24(4): 509-522, 2002.
- [16] C. Grigorescu and N. Petkov. Distance sets for shape filters and shape recognition. IEEE Trans. Image Processing, 12(10): 1274-1286, 2003.
- [17] N. Alajlan, I.E. Rube, M.S. Kamel and G. Freeman. Shape retrieval using triangle-area representation and dynamic space warping. Pattern Recognition, 40(7): 1911-1920, 2007.
- [18] B. Wang and Y. Gao. Structure integral transform versus Radon transform: A 2D mathematical tool for invariant shape recognition. IEEE Trans. Image Processing, 25(12): 5635-5648, 2016.
- [19] N. Kumar, P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, and J.V.B. Soares. Leafsnap: A computer vision system for automatic plant species identification. In Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 502-516, 2012.
- [20] M. Liu, B.C. Vemuri, S.-I. Amari, and F. Nielsen. Shape retrieval using hierarchical total Bregman soft clustering. IEEE Trans. Pattern Analysis and Machine Intell., 34(12): 2407-2419, 2012.
- [21] C.T. Zahn and R.Z. Roskies. Fourier descriptors for plane closed curves. IEEE Trans. Comput. 21(3): 269-281, 1972.
- [22] Q.M. Tieng and W.W. Boles. Recognition of 2-D object contours using the wavelet transform zero-crossing representation. IEEE Trans. Pattern Anal. Machine Intell., 19(8): 910-916, 1997.
- [23] C. Xu, J. Liu, and X. Tang. 2D shape matching by contour flexibility. IEEE Trans. Pattern Anal. Machine Intell., 31(1): 180-186, 2009.
- [24] T. Adamek and N.E. O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. IEEE Trans. Circuits and Systems for Video Technology, 14(5): 742-753, 2004.
- [25] B. Wang, D. Brown, Y. Gao, and J. La Salle. MARCH: Multiscale-arch-height description for mobile retrieval of leaf images. Information Sciences, 302: 132-148, 2015.
- [26] C.-L. Chou, H.-T. Chen, and S.-Y. Lee. Pattern-based near-duplicate video retrieval and localization on web-scale videos. IEEE Trans. Multimedia, 17(3): 382-395, 2015.
- [27] N. Arica and F. Vural. BAS: a perceptual shape descriptor based on the beam angle statistics. Pattern Recognition letters, 24(9-10): 1627-1639, 2003.
- [28] F. Janan and M. Brady. Shape description and matching using integral invariants on eccentricity transformed images. In J. Comput. Vis., 113(2): 92-112, 2015.
- [29] S. Tabbone, L. Wendling and J.-P. Salmon. A New Shape Descriptor Defined on the Radon Transform. Computer Vision and Image Understanding, 102(1): 42-51, 2006.
- [30] X. Wang, B. Xiao, J.-F. Ma, and X.-L. Bi. Scaling and rotation invariant analysis approach to object recognition based on Radon and Fourier-Mellin transforms. Pattern Recognition, 40(12): 3503-3508, 2007.
- [31] R. Gopalan, P. Turaga, and R. Chellappa. Articulation-invariant representation of non-planar shapes. In Proc. Euro. Conf. Computer Vision, pp. 286-299, 2010.
- [32] W. Dong, Z. Wang, M. Charikar and K. Li. Efficiently matching sets of features with random histograms. In Proc. ACM International Conf. Multimedia, pp. 179-188, 2008.
- [33] A. Kadyrov and M. Petrou. The trace transform and its applications. IEEE Trans. Pattern Anal. Machine Intell., 23(8): 811-828, 2001.
- [34] B.-W. Hong and S. Soatto. Shape matching using multiscale integral invariants. IEEE Trans. Pattern Anal. Machine Intell., 37(1): 151-160, 2015.
- [35] J. Panda, M.S. Brown, and C.V. Jawahar. Offline mobile instance retrieval with a small memory footprint. In Proc. IEEE Int. Conf. Computer Vision, pp. 1257-1264, 2013.
- [36] J. Wang, X. Bai, X. You, W. Liu, and L.J. Latecki. Shape matching and classification using height functions. Pattern Recognition Letters, 33(2): 134-143, 2012.

- [37] P. Novotny and T. Suk. Leaf recognition of woody species in Central Europe. *Biosystems Engineering*, 115(1): 444-452, 2013.
- [38] C. Zhao, S. S.F. Chan, W. Cham, and L.M. Chu. Plant identification using leaf shapes—A pattern counting approach. *Pattern Recognition*, 48(1): 3203-3215, 2015.
- [39] Y.-H. Yang, W.H. Hsu, and H.H. Chen. Online reranking via ordinal informative concepts for context fusion in concept detection and video search. *IEEE Trans. Circuits and Systems for Video Technology*, 19(12): 1880-1890, 2009.
- [40] Y.-C. Pan and L.-S Lee. Performance analysis for lattice-based speech indexing approaches using words and subword units. *IEEE Trans. Audio, Speech, and Language Processing*, 18(6): 1562-1574, 2010.