Point-based Medialness 2D Shape Description and Identification

Prashant Aparajeya · **Frederic Fol Leymarie**

Abstract We propose a perception-based medial point description of a natural form (2D: static or in articulated movement) as a framework for a shape representation which can then be efficiently used in biological species identification and matching tasks. Medialness is defined by adapting and refining a definition first proposed in the cognitive science literature when studying the visual attention of human subjects presented with articulated biological 2D forms in movement, such as horses, dogs and humans (walking, running). In particular, special loci of high medialness for the interior of a form in movement, referred to as "hot spots", prove most attractive to the human perceptual system. We propose an algorithmic process to identify such hot spots. In this article we distinguish exterior from interior shape representation. We further augment hot spots with extremities of medialness ridges identifying significant concavities (from outside) and convexities (from inside). Our representation is strongly footed in results from cognitive psychology, but also inspired by know-how in art and animation, and the algorithmic part is influenced by techniques from more traditional computer vision. A robust shape matching algorithm is designed that finds the most relevant targets from a database of templates by comparing feature points in a scale, rotation and translation invariant way. The performance of our method has been tested on several databases. The robustness of the algorithm is further tested by perturbing the data-set at different levels.

P. Aparajeya
Deparment of Computing, Goldsmiths, University of London, London SE14 6NW, UK
E-mail: p.aparajeya@gold.ac.uk

F. F. Leymarie
Deparment of Computing, Goldsmiths, University of London, London SE14 6NW, UK
E-mail: ffl@gold.ac.uk

Keywords 2D shape analysis, dominant points, information retrieval, medialness representation, shape compression, planar articulated movement.

1 Introduction

In this communication we describe our proposed 2D shape representation for biological objects (animals and plants) which is informed by results and techniques from artistic rendering and animation, cognitive psychology, as well as computer vision (Fig. 1.(h)). An artist will often draw different poses of an animal in movement by using various combinations of primitive structures of different sizes (*e.g.* approximate disks of various radii, Fig. 1.(a)), a technique referred to as "geometric drawing" [20]. Different body movements are characterised by a particular orientation and combination of these primitives. Each body part can be fleshed out and refined in successive sketches [35, 20]. Such sketches of characters can then be directed using the Line of Action technique from animation: a single curve running through the "middle" of the character, which represents an overall force and direction of movement for the character [10, 17]. From the point of view of psychophysical investigations on the perception of shape movements by humans, Kovács *et al.* have shown that such articulated movements of a biological character can be best captured via a minimal set of dominant features, potentially being represented as isolated points or "hot spots" [26, 25].

Inspired with these two approaches to the perception of natural 2D articulated movements, we have investigated a possible scheme based on the notion of robust medialness presented by Kovács *et al.* that can efficiently capture the important structural *part-based* information commonly used in artistic drawings and animations. While in the papers by Kovács *et al.* a mathematical model of medialness approximating humans perceptual attention is presented, no algorithmic chain is proposed to retrieve feature points. One of our contributions is to precisely propose such an algorithmic chain. We further modify their definition of medialness by introducing orientation to the bounding contour points to reduce halo effects from nearby contour elements constitutive of separate object parts. We also augment (interior) hot spots as proposed by Kovács with significant concave and convex points by locating these at the ends of medialness ridges. This permits us to tie medialness with the "codon" theory of Hoffman, Richards and others, proposed in the 1980's as potent object part separators [44] and often used to characterise shape complexity [21] or in scale-space analyses of contours [37].

Shape representation towards matching has been addressed in many ways by computer scientists in recent years, including by directly characterising and grouping contour points [54, 34], by classical
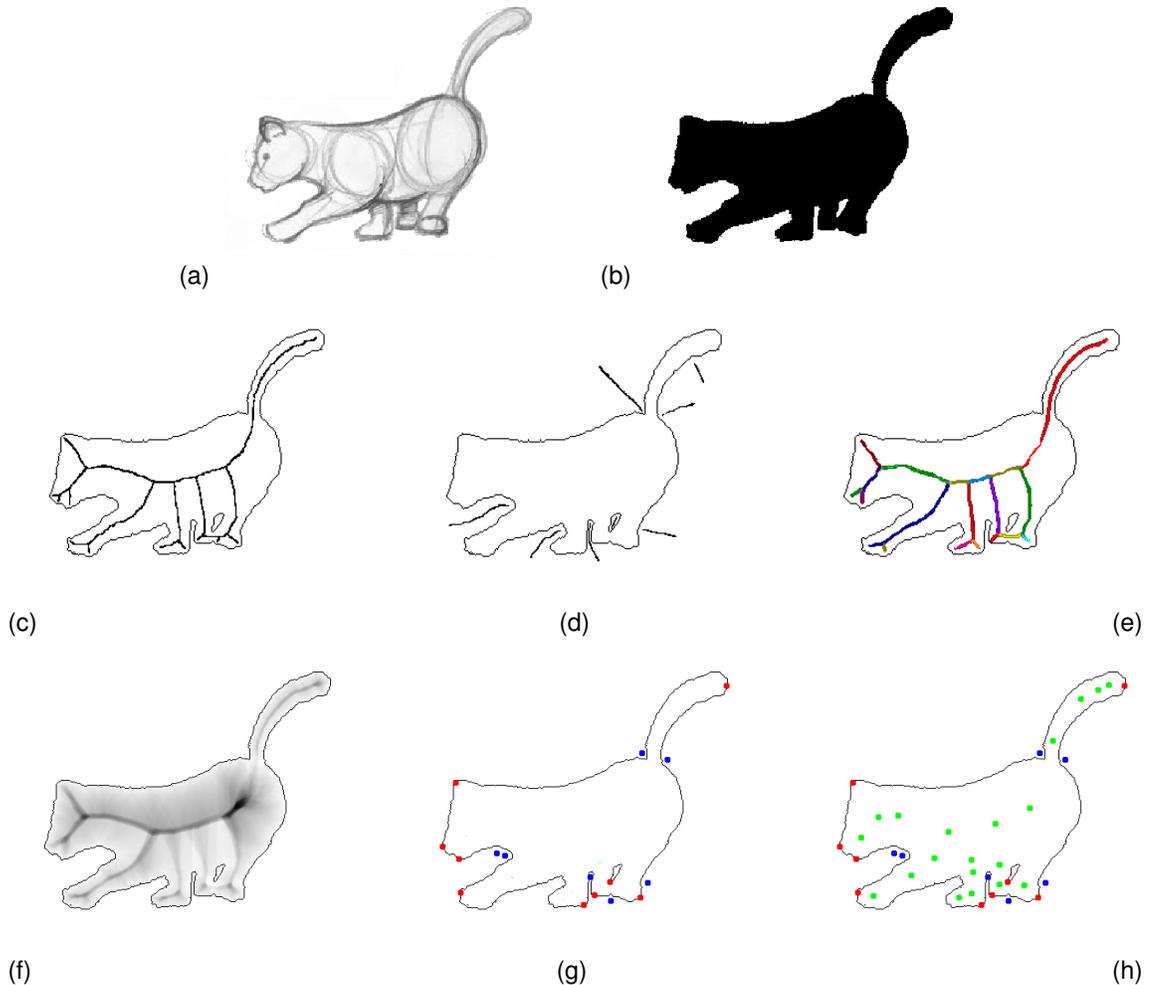
Fig. 1 (a) Sketch of a cat built from a small number of approximate disks (visible sketched lines); (b) corresponding segmented and binarised image; (c) classical internal medial-axis approximation; (d) external medial-axis; (e) 2D shock graph; (f) proposed interior medialness map; (g) recovered concave (blue dots) and convex (red dots) points, and (h) final dominant (medial) point set (in green for internal ones, in blue for external/concave ones, and in red for convex points) obtained via our method.

contour analysis [6,28,12,50], by computing local contour-based "contexts" [5,56,46,43,55],[1] using instead the inner distance to compute contexts invariant to planar articulations and similar deformations [33,19], using Blum's medial axis transform and its related shock graph to combine geometry and topology with contour and interior symmetries [8,24], or combining explicitly contour and medial axis (aka "skeleton") [30,56,3] (some of the main medial representations are illustrated in Fig. 1, including our proposed method). Other classical approaches emphasise either boundary information (*e.g.* Fourier,

---

[1] A "shape context" is defined centered at each feature contour point, by seeking within a circular polar grid proximal contour neighbors and creating a descriptive histogram from orientation-labeled bins [5]. It represents a discrete approximation of relative contour similarity where a region of influence for each considered feature contour point is being evaluated.

wavelet and scale-space analysis of closed contours) or interior information (*e.g.* primitive retro-fitting or approximations including medial-reps and cores) [42,37].[2] In morphometrics applied to biological data, *landmarks — i.e.* an organism-specific characteristic of a morphological sub-part and consisting of a number of feature points, such as the apex, lobes, teeth, veins junctions, insertion point, petiole, when considering a plant leaf — are used to compare various specimens, where the main practical difficulty remains the automatic extraction of such features [13].

Early computer vision methods related to our approach are "contour enclosure-based symmetries" [22] and "medial point transform" [52], which compute similar (robust) medialness maps, but stop short of applying these to the retrieval of dominant points and their use in shape matching. More recent and also related to our approach is the combined contour and medial axis feature points representations of Shen *et al.* which builds a feature vector for later retrieval tasks [49]. Both the contour and the skeletal graph are segmented in fragments and each such fragment is then used to produce a representative "point" for the feature vector.[3] This approach remains classical in spirit as it depends on well defined (thin) contours and skeletal graphs.

The general approach to matching is then to find good ways to put in correspondence the whole shape representation from a query with an equivalent complete shape representation of a target object (*e.g.* extracting a skeleton from a segmented image and defining a process to match it with another skeleton description in the database). We propose instead to define an efficient medialness-based representation which remains discrete (point-based), is (at least approximately) invariant to scaling, rotations and translations, and can be the basis of a feature vector map for efficient query-target matching tasks as produced in the discipline of Information Retrieval. Note that we do not require to have a complete object segmented and thus will also address partial shape matching. Note also that most of the well established shape-based approaches do not consider deformations and articulated movements, while we do [16,19]. Contrarily to classical skeletal-based representations, ours is *not overly sensitive* to small boundary deformations and furthermore gives high response in those regions where the object has high curvature with *large boundary support* and in the vicinity of *joints* (between well-delineated parts, such as the limbs of an animal).

---

[2] Medial-reps or M-reps, previously referred to as cores, are a discrete alternative to Blum's medial axis, developed by Stephen Pizer and his collaborators, where medial atoms (loci) are selected in a sparse sampling of a main or long medial axis and linked to the object boundary via spokes normal to that boundary at their attachment points [42]. Such a sparse connected representation of medialness is well adapted to elongated forms, such as found in medical imaging, when modeling various body tissues [41]. This approach is model based, where a sparse connected skeletal grid is retro-fitted to the outline of an object segmented in an image (in 2D or 3D).

[3] Where the contour fragments are built alike the method of Wang *et al.* [55].
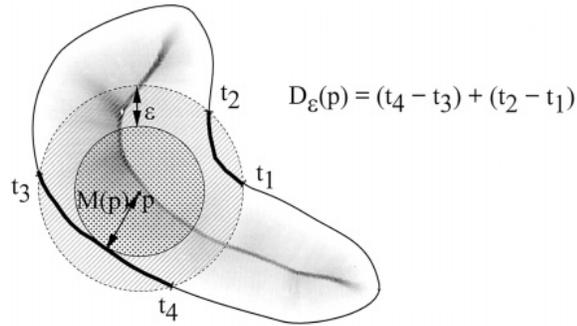
Fig. 2 Adapted from [26, Fig.2] with permission from the lead author: the D_ϵ function for a simple shape defined as an accumulation of curve segments falling inside the annulus neighborhood of thickness ϵ (thick boundary segments within the gray ring) centered around the circle (with center p). $M(p)$ is taken as the minimum radial distance from point p to the nearest contour point.

2 Medialness Measure via Oriented Boundary

According to Kovács *et al.*, the definition of *medialness* of a point in the image space is given by the accumulation of sets of boundary segments falling into an annulus of thickness parameterised by the tolerance value (ϵ) and with interior radius taken as the minimum radial distance of a point from boundary [26] (Fig. 2). Medialness hence defined maps an image (of the interior of an object) to a grey level 2D map where greyness is a direct measure of accumulated medialness measure for each point of the original image under consideration. One can think of such a map as a landscape with peaks and pits, ridges and valleys (*i.e.* a 2D set function). A "dominant point" for this medialness map then intuitively corresponds to a well localised peak.

In order to implement this notion of medialness we first need to characterise the D_ϵ medialness measure. This is a sampling function constrained to annular sectors. Given a minimum radius function $R(p)$ with center locus p which defines the interior shell of the annulus, and $R_{max}(p) = R(p) + \epsilon$ the exterior shell, a given sector S_i is specified by an angular opening θ_i (in radians) with bounding segments defined by the intercepts t_i and t_{i+1}. The area of the sector is then: $A_i = \theta_i(R + 0.5\epsilon) \times \epsilon$. In practice the intercepts bounding a sector $[t_i, t_{i+1}]$ are given by the extremal points of contour or edge segments entering and exiting the annulus, *i.e.* crossing either of its interior or exterior shells (Fig. 2). The medialness measure is then taken as a sampling over the sum of two or more annular sectors containing boundary information (*i.e.* with contour or edge segments); what is actually measured in each sector is application dependent; in our case one can count the amount of boundary information present in each sector, *e.g.* the number of edge pixels. or the contour segment length or, a binary counter for fixed-length elementary sub-sectors (or bins for fixed sub-tended elementary angular steps, $\delta\theta$), or even

as the area of the sector, A_i. We denote the measure taken over an annular sector S_i: $\widehat{S_i}$. Then, we can express the medialness measure first proposed by Kovács *et al.* in a general form as:

$$D_\epsilon = \sum_{k=0}^{k=n} \widehat{S_i} \, , i = 2k + 1 \, , n \geq 1 \, . \tag{1}$$

This formulation implies that at least two separate annular sectors are traversed by boundary information. Notice that when the tolerance value ϵ reduces to zero, D_ϵ reduces to a maximal inscribed disk and leads to the traditional medial axis graph measure.

One noticeable drawback of this definition when seeking to retrieve dominant points is that it does not make a distinction with neighboring boundary segments part of separate object parts and which ought not to be considered in the support annular zones; *e.g.* this may be the case with the fingers of one's hand when these are kept near each other (Fig. 3). That is, by introducing a boundary information catchment area (via the tolerance annular width, ϵ), we loose the implicit property of a maximal disk being inscribed (to one object's interior part). In figure 3 (top) the medialness mapping has been performed using Kovács' original method where the medialness of each finger also is influenced by neighboring fingers and result into extra information being added creating a typical "halo" effect in the vicinity of ridges (shown in dark grey in our results).

This imprecision in D_ϵ can be remedied by introducing a measure of orientation at boundary points when assuming figure-ground segmentation (*i.e.* when knowing the interior versus the exterior of an object); we propose to use such information (*e.g.* when obtained from a classical gradient boundary filter) to modify the medialness function, resulting in D_ϵ^* (equation 3, derived in the following sub-section). This not only reduces the impact of neighboring parts on medialness measurements, but also emphasize the sharpness of medialness at the tips of ridges (Fig. 3), which proves helpful in practice to locate convex and concave features points as will become clearer below.

On completion of medialness measurements, each pixel in the transformed image space holds a local shape information (of accumulated medialness). When assuming figure-ground separation, thickness variations, bulges and necks of an object are captured via *interior* medialness measurement. In the work of Kovács *et al.* it is shown that humans are most sensitive to a small number of localised areas of medialness which coarsely (by visual inspection) correspond to joints for animated bodies [26, 25]. Our equivalent (extended) notion is defined as *dominant points* and can be applied to any objects, animated or not. Dominant points are constrained to be a relatively small number of points of high medialness obtained by filtering out the less informative, redundant and noisy data from the initial medi-
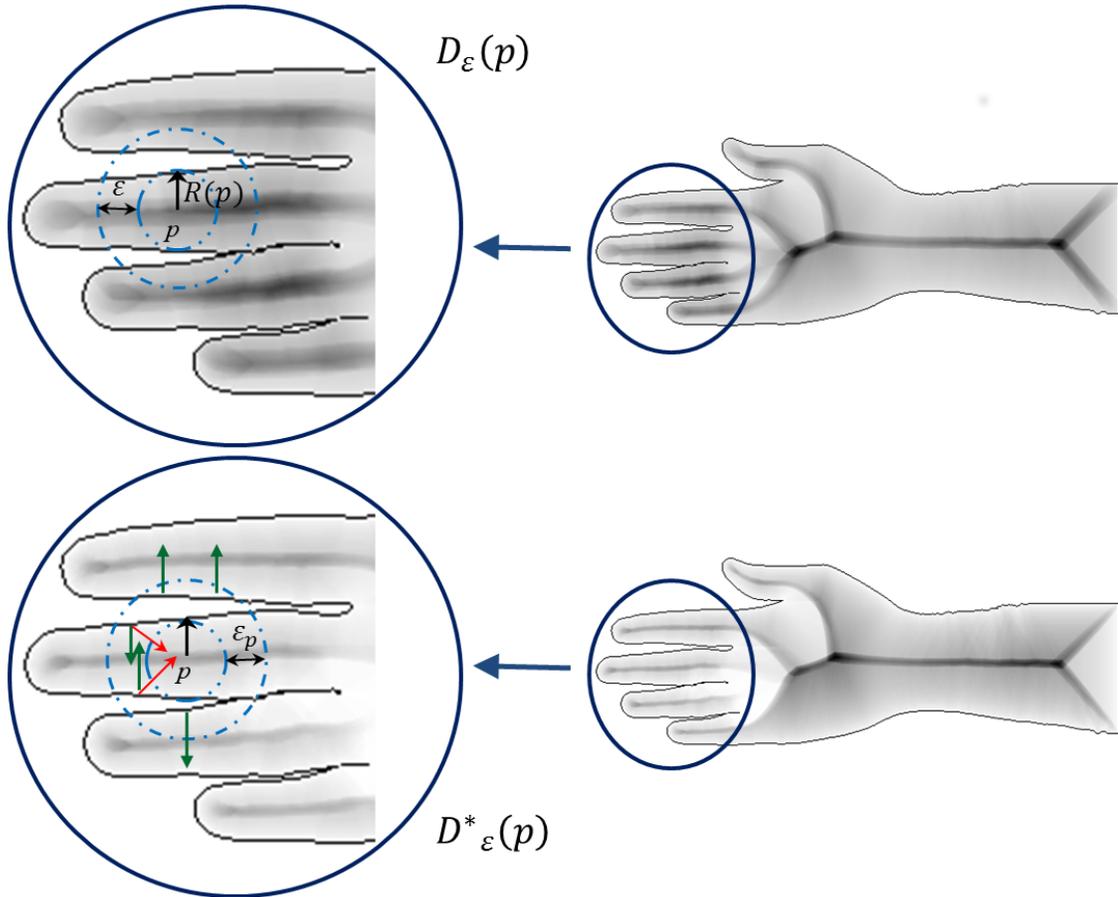
Fig. 3 Comparison of the results of computing medialness functions on the image of a human hand. *Top*: the result of applying the original D_ϵ function as proposed by Kovács *et al. Bottom*: the result of our proposed method which takes into account orientation at the boundary (D_ϵ^*). In computing D_ϵ, the medialness at a point p may take support from neighboring fingers (their boundary information), while when using the proposed D_ϵ^* such extraneous information can be discarded using a simple scalar product of a vector locally pointing towards the interior of the figure (here the fingers) with its associated vector pointing towards the center of the annulus gauge.

alness image space. In distinction to Kovács *et al.* we compute medialness both for the figure (interior) and the ground (exterior) as we seek to also characterise concavities of an object.

To identify dominant points a *morphological top-hat transform* [53] is applied to isolate ridges and their peaks in the medialness 2D function. Ridges are filtered using an empirically derived threshold.[4] The selected ridges are then locally characterised by isolated representative points associated to peaks in the ridge data. To avoid considering large numbers of nearby isolated peaks which are characteristic of elongated object parts or parts with many small deformations, only peaks at a given minimum distance away from each other are retained. The extraction process of *external* dominant points is achieved

---

[4] Currently the threshold value is globally set (the typical value is 10 for the intensity range 0-255) and remains constant for the whole DB.
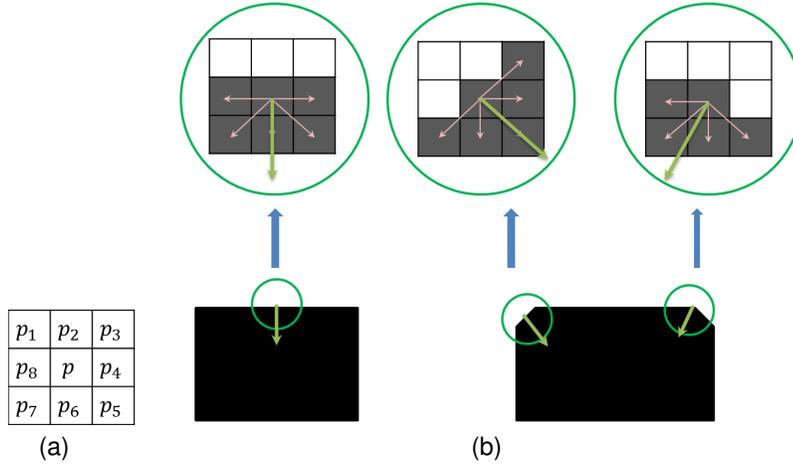
Fig. 4 Examples of the orientation assessment of candidate boundary points. (a) 8-connected neighbor-points of p, and (b) three different examples of boundary points: horizontally straight, right diagonal and at a corner of the object. The orientation of a boundary-point b is illustrated by a green arrow; in practice this can be approximated by the summation from other vectors in its 8-direct neighborhood linking to object pixels (shown as pink arrows), some of which may be null vectors.

by combining a concavity measure at the tip of medialness ridges together with a length of support on the associated contour segments. Again, a spatially localised filtering is applied to isolate representative dominant points. The set of *convex points* is similarly extracted. The details of the recovery of the three types of selected points is given in Section 3.

Together, the three types of selected dominant points are then considered as the *representative feature points* of the shape. Our matching algorithm (detailed in Section 4) is designed in such a way that it first compares internal dominant points of a query object with internal representative dominant points of target shapes in a database. External concave points are then similarly processed and convex points are used in a final refinement step. The matching algorithm first analyses the amount of scale, rotation and translation of the query w.r.to the target image. These values are then applied over the query image to find the best possible matching location in the target image. We next provide the details of the various stages of computation.

2.1 Medialness for Oriented Boundary Segments

We define the computation of 2D medialness (modifying Kovacs *et al.*'s original definition [26]) by introducing orientation to each considered object boundary pixels. The orientation of a boundary pixel is given by a vector $\overrightarrow{v}$ with respect to a positively oriented horizontal X-axis. Given a pixel p, a set of 8-neighbor pixels $\{p_1, p_2, ... , p_8\}$ and a set of vectors $\{\overrightarrow{v_1}, \overrightarrow{v_2}, ... , \overrightarrow{v_8}\}$ joining p and its neighbors $\{\overrightarrow{pp_1}, \overrightarrow{pp_2}, ... , \overrightarrow{pp_8}\}$ (Fig. 4), then $\overrightarrow{v}$ is defined as (the vector sum):
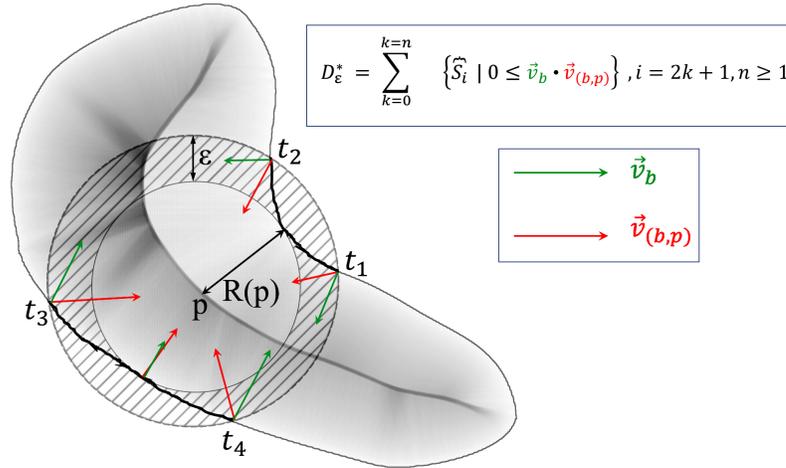
Fig. 5 Illustration of the D_ϵ^* function for a simple shape, defined as an accumulation of boundary segments falling inside an annulus neighborhood of thickness ϵ (shown as darker boundary segments within the annulus' ring) centered around a position p, and such that the associated orientation vector $\vec{v}$ has a dot product with the unit radius vector (taken in the direction of the line from $b(t)$ to p) which is positive. $R(p)$ is taken as the minimum radial distance from p to the nearest contour point.

$$\vec{v} = \sum_{i=1}^{8} \vec{v_i} \tag{2}$$

The medialness gauge at a point p is defined similarly to D_ϵ by adding an orientation constraint such that only those boundary loci which are pointing inward (with respect to the figure) are considered for evaluation (Fig. 5):

$$D_\epsilon^* = \sum_{k=0}^{k=n} \left\{ \widehat{S_i} \mid 0 \leq \overrightarrow{v_b} \cdot \overrightarrow{v_{(b,p)}} \right\} , i = 2k+1 , n \geq 1 . \tag{3}$$

for a point $p = [x_p, y_p]$, vector $b(t) = [(x(t), y(t)]$ describing (pixel) loci along a piece of 2D bounding contour (B) of the object, and such that $\overrightarrow{v_b}$ is the orientation of the boundary point $b(t)$, $\overrightarrow{v_{(b,p)}}$ is the orientation of the line joining $b(t)$ to p.[5] The metric $R(p)$ (minimum radial distance to the interior annular shell) is taken as the smallest available distance between p and a bounding contour element:

$$R(p) = \left\{ \min_t (|p - b(t)|) \mid 0 \leq \overrightarrow{v_b} \cdot \overrightarrow{v_{(b,p)}} \right\} \tag{4}$$

In figure 6 the medialness measurement is performed on a sea-horse profile illustrating the evolution in interior medialness measures when varying the value of tolerance (ϵ), which reflects a smoothing

---

[5] The non-negativeness (of the scalar product $\overrightarrow{v_b} \cdot \overrightarrow{v_{(b,p)}}$) is used to rule out boundary pixels which are oriented away from the given annulus centre. We do not consider the geometry (differential continuity) of a contour other than provided by that gradient orientation. NB: this criterion is efficient if we have reliable figure-ground information. This is a limit of the modified gauge D_ϵ^*; however we can always fall back on the original gauge D_ϵ if object segmentation is not reliable.

$$D^*{}_\varepsilon(p) - Function$$

$\epsilon = 4$ $\epsilon = 8$ $\epsilon = 12$ $\epsilon = 16$
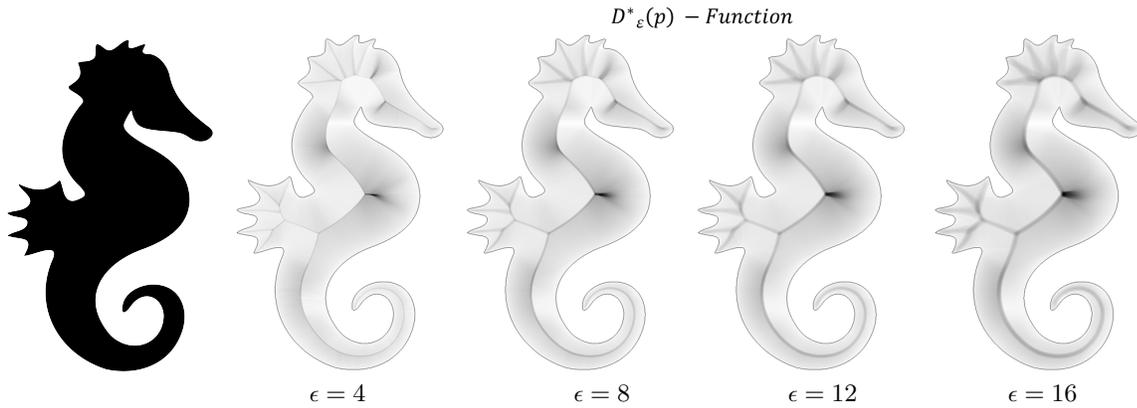
Fig. 6 Medialness and tolerance — *Left*: silhouette of sea-horse (image size: 630×1010); other images to the right show variations in D_ϵ^* for increasing tolerance values (ϵ, in pixel units).

effect: *i.e.* as the tolerance increases, medial symmetries with smaller boundary support are averaged out in favor of larger scale ones — their associated ridges become more dominant.[6]

2.2 Adaptive Tolerance (ϵ) as a Function of the Minimum Radial Distance ($R(p)$)

The medialness measure of a point p varies with the two parameters: $R(p)$ and ϵ, where $R(p)$ is the minimum radial distance between p and the bounding contour, and ϵ is the width of the annulus region (capturing object trace or boundary information). Any boundary point b falling inside this annulus which satisfies the (orientation) condition of equation 3 is added in support for medialness at p. We can think of the width of the annulus as dictated by ϵ as an equivalent to a scaling parameter: the larger the width, the more averaging of nearby contour information is considered. How to set the tolerance ϵ in order to have desirable scaling properties thus needs to be addressed.

First notice that as the minimum radius value $R(p)$ augments, it is useful to augment the corresponding value of ϵ so as to filter away noisy variations in what is an increasing part of a lower curvature (flatter) set of boundary segments. On the other hand, for smaller values of $R(p)$ we should use a sharper, smaller width ϵ in order to more accurately capture the augmenting local boundary curvature variations. Furthermore, when considering the tip of sharp fine elongated limbs, *e.g.* the end of fingers, a decreasing value of ϵ is also beneficial in practice to avoid blurring the corresponding centre of curvature locus. We have therefore studied various ways to set automatically an augmenting value of the tolerance ϵ as the minimum radius value $R(p)$ is increased. The obvious first choice is to use a simple linear relation, such as: $\epsilon_p = (R(p) + 1)/x$, but we have observed in practice that this leads to rapidly

---

[6] The tolerance value (ϵ) is currently set as the elementary pixel size (and so is related to the resolution used).

$\varepsilon = (R(p) + 1)/x$

$\varepsilon = (R(p) + 1)/8$

(a)

$\varepsilon = log_x(R(p) + 1)$

$\varepsilon = log_{1.5}(R(p) + 1)$

(b)
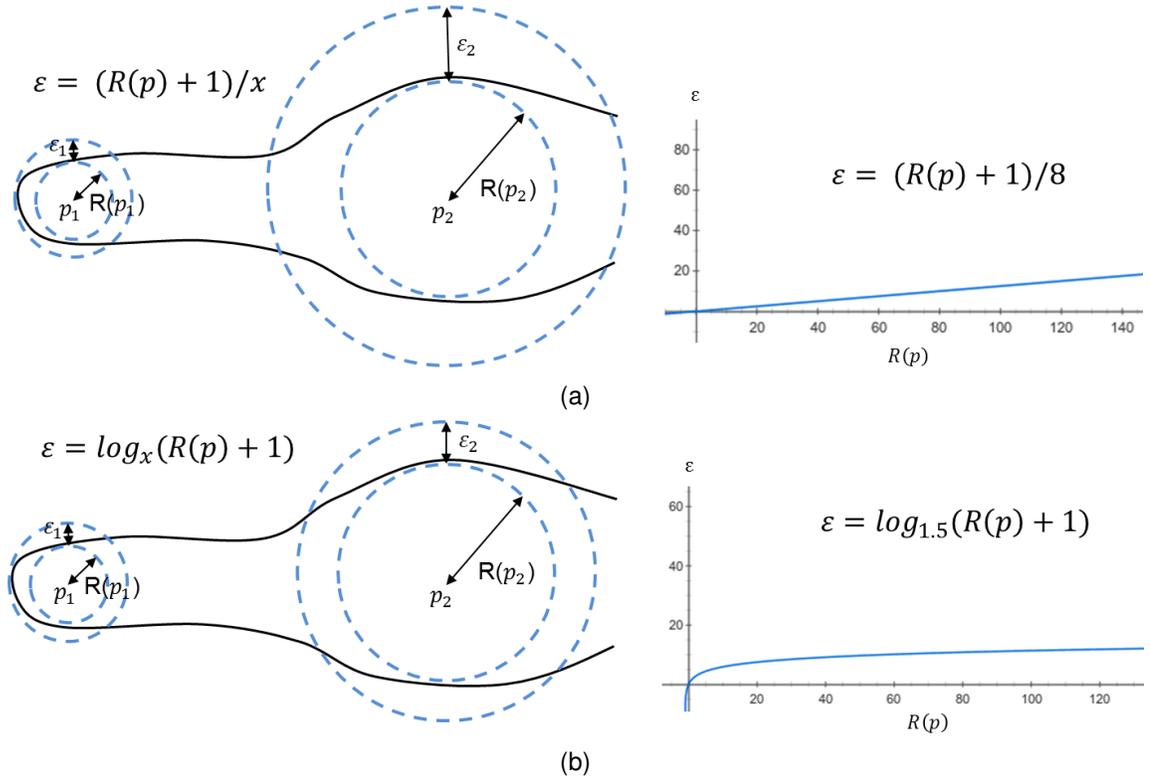
Fig. 7 Tolerance ϵ defined as two different functions of the minimum radial distance $(R(p))$. (a) Linear function: variation of ϵ w.r.to $R(p)$ at a typical value of slope $(1/8)$; and (b) Logarithmic function: variation of ϵ w.r.to $R(p)$ value of base of 1.5.

increasing averaging effects which generate overly thick ridge traces or generate ridge doubling (alike the halo effect when using D_ϵ). To overcome this effect, we select ϵ as a logarithmic function of $R(p)$ (Fig. 8):

$$\epsilon_p = log_x(R(p) + 1). \tag{5}$$

Having selected ϵ to be adaptive as a logarithmic function of $R(p)$, the next step is to define a useful logarithmic *base*. When such a base has a large value it produces very sharp medialness for large radii values response but can emphasise noisy responses. This can be intuitively understood since a relatively large base makes the medialness measure become more and more like a traditional medial axis transform as the relative width of the annulus region becomes smaller and smaller in comparison to $R(p)$ for increasing radii. Setting the logarithmic base to a relatively low value (*i.e.* nearer unity) gives smoother responses, which can augment considerably the halo effect as with a linear function (Fig. 9, first two rows). By empirical investigation we have observed that a good compromise can be obtained for base with values in the range of $[1.4, 2]$. For our reported results we use a logarithmic base of 1.5.

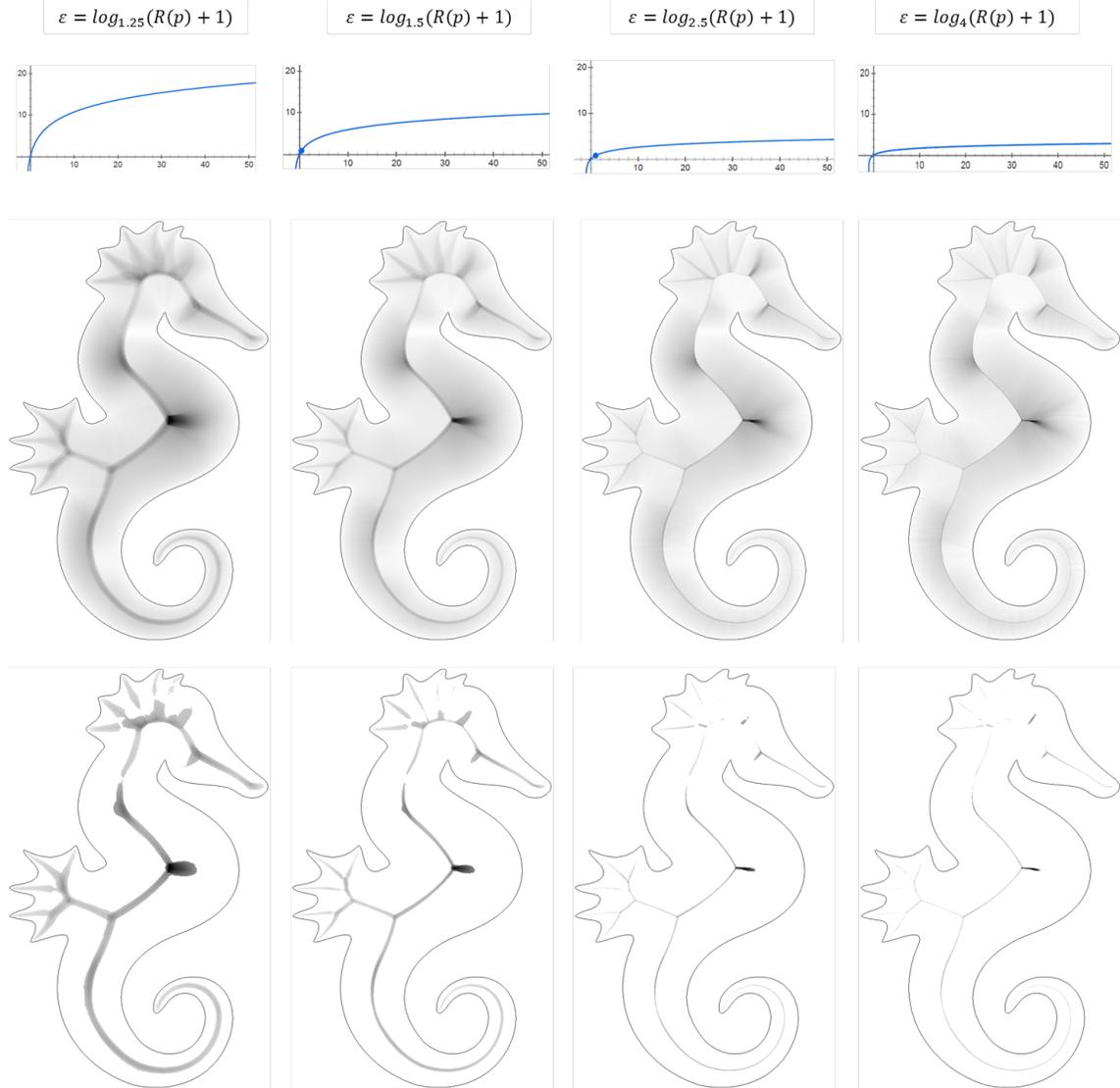$$\varepsilon = log_{1.5}(R(p) + 1) \qquad \varepsilon = (R(p) + 1)/4 \qquad \varepsilon = (R(p) + 1)/8$$
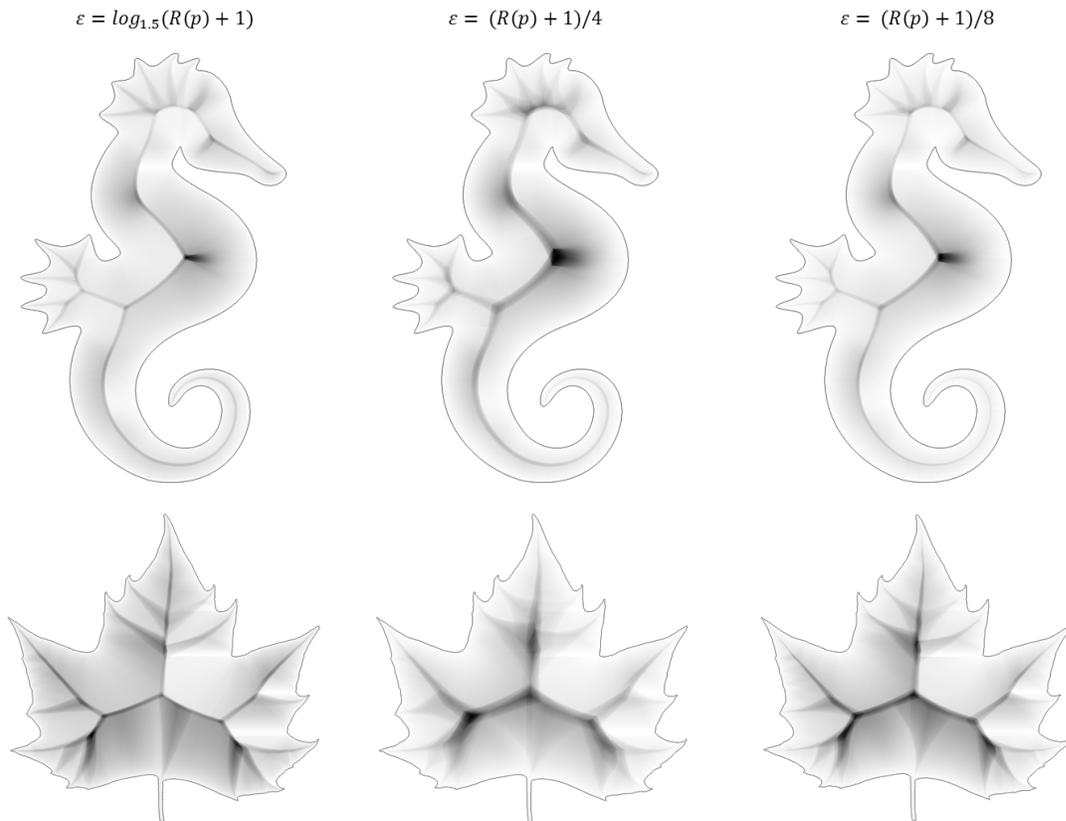
Fig. 8 Comparing (interior) medialness gauge maps for different adapting functions to set the tolerance ϵ. On the left are the results of setting ϵ as a logarithmic function of $R(p)$, while to the right ϵ varies linearly with $R(p)$; notice less sharpness in the medialness ridges and tips near convex corners when using linear functions.

3 Feature Extraction

Medialness measurement is currently done separately for internal and external regions and takes advantage of the perceptual figure-ground dichotomy known to be a powerful perceptual cue in humans [2,29]. This also enables our method to consider more easily articulated objects as potential targets in pattern recognition tasks. On the other hand it requires that some good image segmentation is available prior to initiating our medialness computations.[7] Under these constraints we can now detail how feature points can be retrieved from the medialness map D_ϵ^* with logarithmic base setting of tolerance ϵ.

---

[7] In the recent cognitive science literature, arguments are presented to support the idea the medialness can in fact be the basis of figure-ground segregation [29].

$$\varepsilon = log_{1.25}(R(p) + 1)$$ $$\varepsilon = log_{1.5}(R(p) + 1)$$ $$\varepsilon = log_{2.5}(R(p) + 1)$$ $$\varepsilon = log_4(R(p) + 1)$$

Fig. 9 Comparing internal medialness measure different adaptive tolerance levels ϵ set as logarithmic functions of $R(p)$. We illustrate from left to right how augmenting the logarithmic base impacts the resulting medialness map (2nd row) and associated ridge traces retrieved by applying a morphological hat transform (3rd bottom row). Note that for a base nearer unity, the medialness response suffers more from halo effects making dominant points more difficult to localise, while higher values of the base augments sharpness of the results but also emphasise noisier, less significant (with small support) boundary features.

3.1 Internal Dominant Points

Medialness increases with blackness in our transformed images (used as a means of visualisation of medialness). To select points of internal dominance, a black top-hat transform is applied, resulting in a series of dark black areas which typically correspond to peaks, ridges and passes of the medialness map when considered as a height field. The black top-hat transform is defined as the difference of the
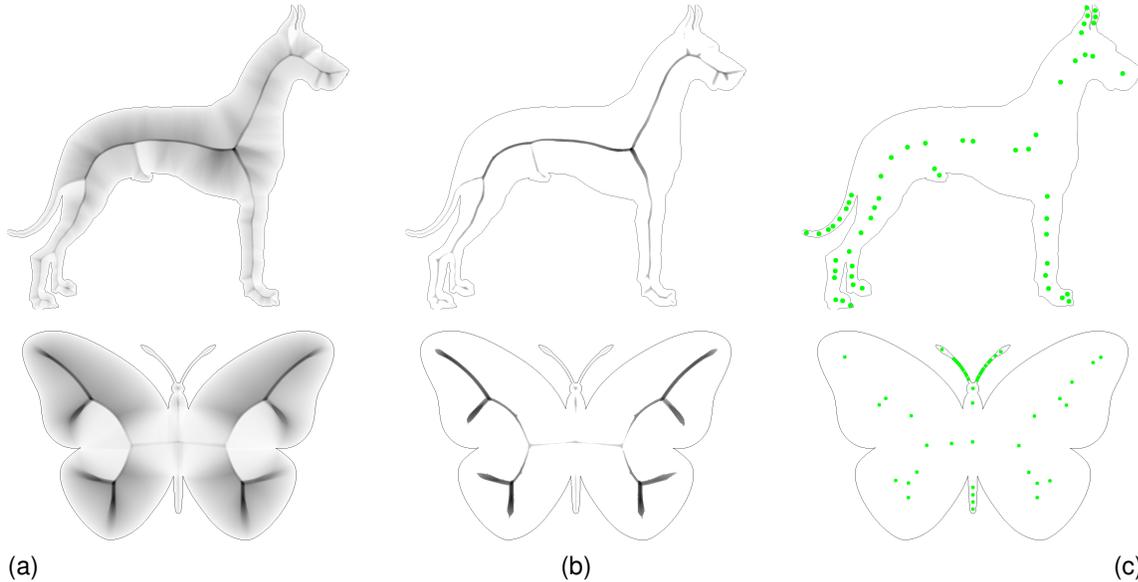
Fig. 10 Illustration of the three successive steps in isolating *internal* dominant points: (a) medialness representation of (the interior of) a standing dog (top) and butterfly (bottom); (b) corresponding top-hat transform; and (c) internal dominant points illustrated as **green** dots together with the original object's pre-segmented boundary.

morphological closing of an input function by a flat structural element (a disk with radius as single parameter). Closing is a set operator on functions which removes small holes from the foreground of an image, placing them in the background (augmenting the local function set values) [48,53,14]. This filtering is followed by a thresholding to discard remaining areas of relatively low medialness significance. Figures 9 (bottom row) and 10 (b) show the result obtained after applying the black top-hat transform on a medialness map.

We still require to process further the output of the top-hat transform to isolate the most dominant points amongst the remaining selected medialness loci. We also consider the cases where the resulting ridges are more like plateaus and thus rather flat at their top. In order to identify isolated representative dominant points for such plateaus we pull-up such flatten regions and map the central locus of a plateau to the highest local peak value (Fig. 11). A simple way to achieve this result is to locally modify the value of medialness at each filtered point of the top-hat image to ensure a non-flat "response-value" ($\psi_\epsilon(p)$):

$$\psi_\epsilon(p) = \left\{ \sum M(q) \mid (|p - q| \leq \epsilon_p) \ \& \ M(q) \leq M(p) \right\} . \tag{6}$$

To provide some control of the possible clustering of dominant points, a flat circular structuring element of radius ϵ_p (but of at least 2 pixels in width) is also applied over the output of a top-hat transform to pick-up maxima. We also impose that no remaining points of locally maximised medialness are too
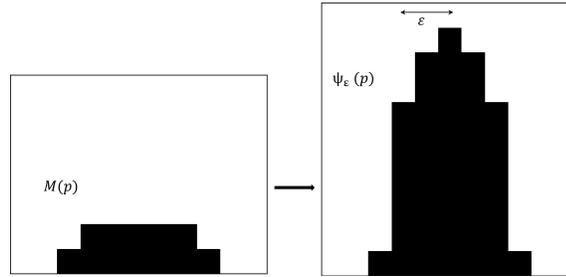
Fig. 11 *Left*: Illustration of a plateau in medialness response following a top-hat transform where no peak locus is isolated. *Right*: Response-value ψ_ϵ after filtering the plateau.

close to each other; this is currently implemented by imposing a minimum distance of length $2\epsilon_p$ taken between any pair of selected points. An example of applying these steps to identify interior dominant points in medialness is given in Figure 10 (c).

3.2 External Concave Points

In practice, if an object can be deformed or is articulated, salient concavities can be identified in association to those deforming or moving areas (such as for joints of a human body). Considering this empirical observation, the location of an external dominant point can be made invariant to this deformation/articulation only up to a certain extent. For example, if the location of an external dominant point in medialness is initially relatively far away from the corresponding contour segment, a slight change in the boundary shape near the movable part (such as an arm movement) can considerably change the position of that associated dominant point (Fig. 12). On the other hand, if a potential concavity point is located very close to the contour, it can easily be due to noise or small perturbations in the boundary reflecting a local maximum of curvature. Therefore, to be able to retrieve reliable external concave points, it is first required to provide an adapted definition of concavity as a significant shape feature.

We define a point of local concavity if it falls under a threshold angular region, under the constraint of length of support which itself depends on the tolerance value (ϵ). The value of the threshold (θ_{ext}) is tunable but is always less than π, which permits to control the angular limit of the concave region. A point whose local concavity is larger than θ_{ext} is considered a flat point. In our experiments we tuned the value of θ_{ext} from $5\pi/6$ to $8\pi/9$. In association, we define an external circular region (of radius function of ϵ) centered at each locus containing candidate external dominant points. Each such region may provide only one representative dominant point, where the dominance of a particular point is decided by the maximum containment of boundary points inside the associated annular gauge (of
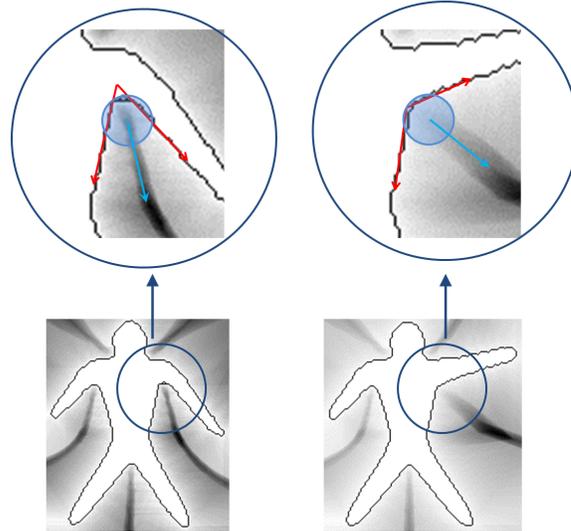
Fig. 12 *Bottom row*: External medialness processing on a humanoid object. The articulated movement of the left arm changes the location and orientation of the associated external dominant point (near the concave curvature peak). If the external dominant point is reasonably far from the contour, then it proves difficult to retrieve a (shape-based) match with the modified form. *Top row*: Red arrows show the local support for concavity while light blue arrows indicate the direction of flow of medialness (away from the concavity).

exterior medialness) and corresponds to (our definition of) the maximum length of support. Finally, we position the representative dominant point to be near the contour at a fixed distance outside the form.[8]

3.3 Interior Convex Points

Our final shape feature is a set of $convex\ points$, where an object's figure has sharp local internal bending and gives a signature of a blob-like part or significant internal curvature structure (*i.e.* a peak in curvature with large boundary support). The goal is to represent an entire protruding sub-structure using one or a few boundary points. Traditionally, such protrusions have a significant contribution in characterising a form's figure [44,30,6,50]. The process of extraction of convex points is similar to the extraction of concave points, the main difference being the value of threshold angle (θ_{int}), where $\pi < \theta \leq 2\pi$. In our experiments we have found useful values to be in the range: $5\pi/4$ to $4\pi/3$ (Fig. 13).

Convex and concave feature points are complementary to each others and have been used in the *codon* theory of shape description: a codon is delimited by a pair of negative curvature extrema denoting concavities and a middle representative positive maximum of curvature denoting a convexity [44]. In our case we relate these two sets with the extremities of the traditional medial axis of H. Blum [8]: end points of interior branches correspond to center of positive extrema of curvature and end points

---

[8] This heuristic, of positioning the representative concavity near the object contour trace is useful both for visualisation and for greater robustness in matching under articulated movements.

of exterior branches are mapped to negative extrema of curvature of the boundary. The repositioning of these extrema near the boundary is alike the end points of the PISA (Process Inferring Symmetry Axis) representation of M. Leyton [32].
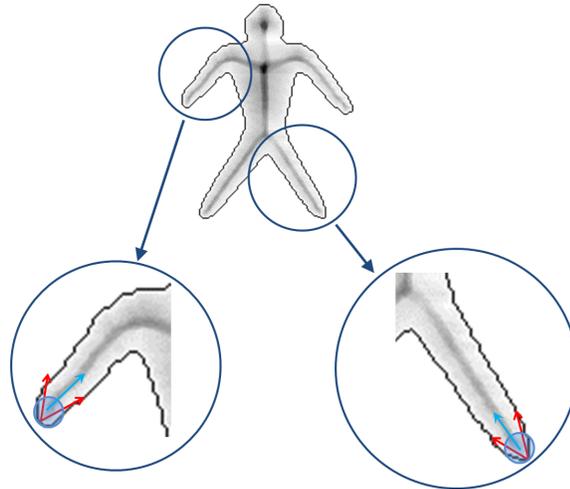

Fig. 13 Illustration of interior medialness processing on a humanoid object to identify significant convexities in the vicinity of the ends of medialness (ridge) trace.
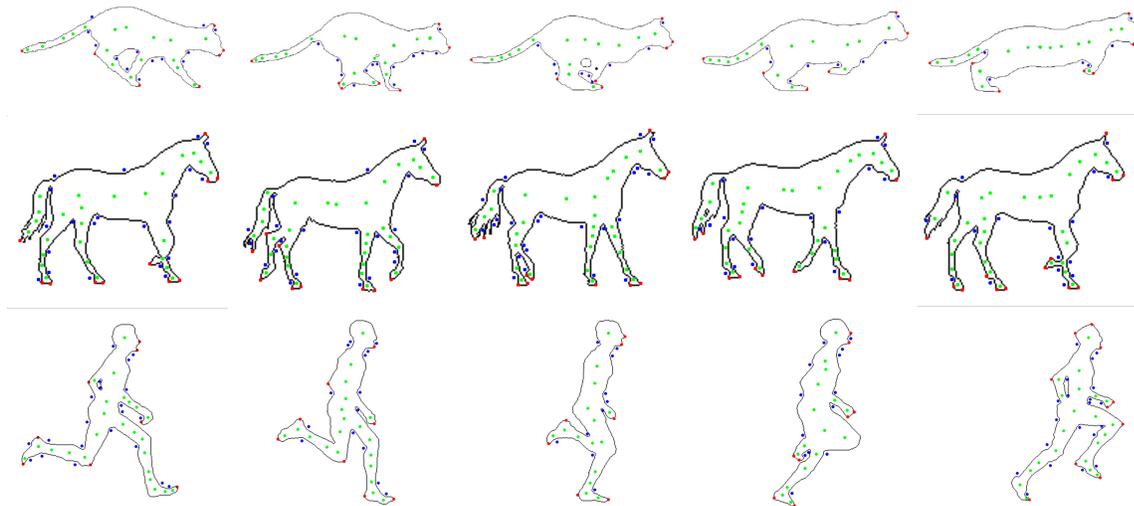
3.4 Feature Point Set and Articulation

Together, the three sets derived from medialness — Interior, Concave and Convex (Fig. 14) — form a rich description of shape which can be used to address the difficult problem of building a shape-based matching system as we will explore in the next section. Such a feature set is also useful to describe anatomic properties of biological forms.

Anatomically, an animal's articulated movement is dependent on joints at the point of connection between bone elements. Our experimental results indicate that a selection of feature points, combining the three sets in describing the articulate part (*e.g.* a limb) should permit to correctly track such movement, as such a feature set remains identifiable in most frames. Upon occlusions (which naturally occur when viewing movement from a typical standing fixed position) while a type of feature may temporarily disappear from a few frames (*e.g.* a concavity or convexity for a given limb), the remaining feature points can be identified and the motion tracking pursued, unless the shape is highly deformed [31]. For usual movements (*e.g.* walking or jogging), these feature points remain present and identifiable in most
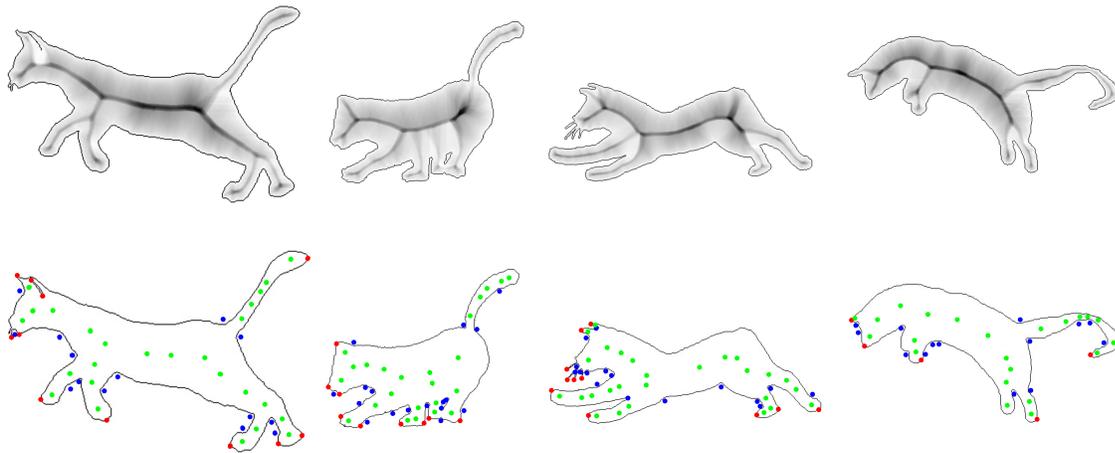
Fig. 14 Top: The D_ϵ^* -function (for the figure/interior only) for a sequential set of frames of the movement of a cat. The maxima (black spots) of the function are good candidates as primitives for biological motion representation. Bottom: Our proposed shape representation in terms of dominant medial (in green) and contour (convex (red) and concave (blue)) points.

Fig. 15 Illustration of how feature points (three types) are preserved through typical articulated movements by: (top) a running cat, (middle) a trotting horse, (bottom) a running man (from a few frames of a running athlete captured by Edweard Muybridge in 1887).

frames (of a typical video sequence) in association to an underlying set of bone junctions and hence can provide a practical signature for these articulated motions (Fig. 15).[9]

---

[9] The precise definition and study of an object part is in itself an important topic which requires a separate presentation, including its precise use in characterising articulated movement. Note also that part perception is studied in psychology where it is shown to be an important cue for human's ability to deal with occlusions and other partial visual information [7].

4 Shape Matching

Our objective is to design a robust algorithm that will match dominant points (query to targets) in an efficient way, in both time (or equivalent numerical complexity) and accuracy. First, feature points are located from query and target images and kept with separate identities as members of the three types: internal, concave and convex. Internal (dominant in medialness) points are used as the keypoints for evaluating scale, rotation and translation of the query image w.r. to a target image. The next step is to improve the correctness of the matching algorithm. For this, concave points are used as additional information; if needed convex points are also considered in a third refinement pass. The following in-formation is associated with the feature points : $< X, Y, R, D_\epsilon >$, where (X,Y) is the 2D location of a feature point p, R is the minimum radial distance of the point from the contour and D_ϵ^* is the medialness measurement. The test case (query Q) is represented as: Q= $\{Q_I, Q_E, Q_c\}$, for internal (Q_I), concave (or external) (Q_E) and convex (Q_C) feature points; while the target image (T) is similarly represented as: $T = \{T_I, T_E, T_C\}$. Each element of Q_I is matched with each element of T_I following four stages.
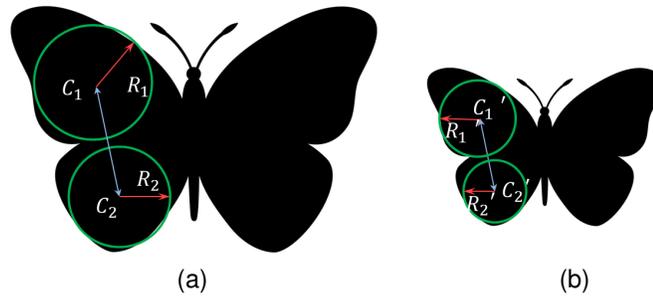
Fig. 16 Illustration of how scaling of an object is evaluated towards matching. Here (a) a butterfly form is used, to be matched with (b) a scaled-down version.

If an object is uniformly scaled by a factor β, we expect the distance between any two similar parts to be scaled by the same factor. In Figure 16, an example of scaling is shown for a butterfly object: consider the top-left (centered at C_1 with radius R_1) and bottom-left (centered at C_2 with radius R_2) parts of a butterfly (a) being scaled-down to corresponding disks: centre $C_1^{'}$ with radius $R_1^{'}$ and centre $C_2^{'}$ with radius $R_2^{'}$ part of the butterfly (b), then:

$$\frac{R_1}{R_1^{'}} = \frac{R_2}{R_2^{'}} \Rightarrow R_2^{'} = R_2 \times \frac{R_1^{'}}{R_1} \tag{7}$$

Here, the scaling factor for butterfly (b) having a disc with radius $R_2^{'}$ w.r.to butterfly (a) having a disc with radius R_2 is $\beta \equiv R_1^{'}/R_2^{'}$. Similarly, their respective distance vectors C_1C_2 and $C_1^{'}C_2^{'}$ get scaled by the same factor, *i.e.*:

$$\frac{R_1}{R_1^{'}} = \frac{C_1C_2}{C_1^{'}C_2^{'}} \Rightarrow C_1^{'}C_2^{'} = C_1C_2 \times \frac{R_1^{'}}{R_1} \tag{8}$$

Stage I is thus used to identify a scale (β) as well as identify the required translation of the query image by matching a dominant point. Stage II is a check on the scale β to make sure at least one more internal dominant point can be used in the matching process (if not, move to a different dominant point not yet considered). In stage III, the rotation of the query image (w.r.to the target) is evaluated. Finally, stage IV modifies the Cartesian positions of each feature point of the query image by applying the evaluated scale, rotation and translation and we proceed to measure a matching performance value.
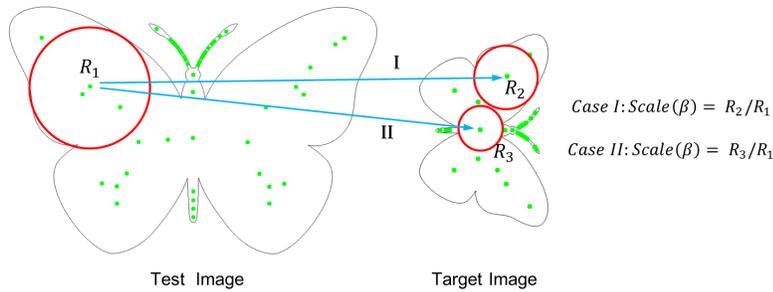
Fig. 17 Illustration of the evaluation of scale (β). In the query image on the left, for a particular dominant point falling inside a red circle, two possible matching locations in the target image, are shown: cases I and II. In each case the circle's radius is dictated by the minimum distance to the contour (from medialness) and the scale (β) is given as the ratio of the minimum radial distances of target vs query.

Stage I: Take an element (q_i) from set Q_I and match it with each element (t_j) of set T_I. For each pair of (q_i,t_j), the scale (β) is evaluated as (Fig. 17):
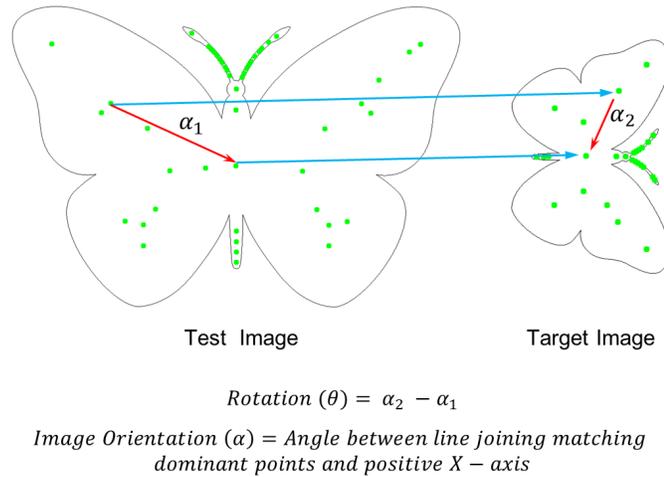
$$scale(\beta) = \frac{R_{t_j}}{R_{q_i}} \tag{9}$$

The scale (of query image w.r. to target) for the matching pair (q_i,t_j) is defined via two translations, one for each axial direction: $\overrightarrow{q_xt_x} = t_x - q_x$ and $\overrightarrow{q_yt_y} = t_y - q_y$, where $q_i = (q_x,q_y)$ and $t_j = (t_x,t_y)$.

Stage II: Now take the next element (q_{i+k}) from set Q_I and match it again with each element (t_j) of set T_I. For each pair of (q_{i+k},t_j) find the scale $\beta^{'}$. If the ratio of $\beta^{'}$ over β is under the tolerance level T_s, then goto stage III. Otherwise, repeat at another point and check the same tolerance criterion T_s and repeat if necessary until all the elements of Q_I are counted. The value $\beta^{'}$ (if it is under the tolerance

level) ensures compatibility under scaling of the query image (with respect to a target) and helps in finding the matching location of the next internal dominant point to consider (red arrows in Fig. 18).

Stage III: We define the orientation (α) of an image by the angle between a line joining two matched dominant points (as obtained from step I and II) and the positive (reference) x-axis. If (q_i, q_{i+k}) are the matching dominant points in Q_I and (t_j, t_{j+l}) are matched dominant points in T_I, then orientations α_q and α_t are defined as the angle between matching dominant points (Fig. 18). The rotation (θ) of the image Q is thus defined by the difference of orientations, *i.e.:*

$$rotation(\theta) = \alpha_{T(t_j, t_{j+l})} - \alpha_{Q(q_i, q_{i+k})} \tag{10}$$

Test Image   Target Image

$$Rotation\ (\theta) =\ \alpha_2\ -\ \alpha_1$$

$$Image\ Orientation\ (\alpha) = Angle\ between\ line\ joining\ matching$$
$$dominant\ points\ and\ positive\ X - axis$$

Fig. 18 For both query (test) and target images, the orientation (α) is the angle between a line joining the two matching internal dominant points (shown with blue arrows) and a positive x-axis. The required rotation (θ) of the query image w.r.to the target is given by the difference in orientations.

Stage IV: Upon obtaining the values of translation, rotation and scale of the image Q (w.r.to T), our next task is to transform the positions of all feature points (Q_I, Q_E and Q_C) of the image Q into the space of image T and finally check for a match. This is done as follows:

1. Construct the 4×4 homogeneous matrix H, of the form $\begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \delta x & \delta y & 0 & 1 \end{bmatrix}$, to perform the required linear (rotation and scaling, represented by the parameters (a, b, c, d)) and affine (translation $(\delta x, \delta y)$) transforms for all feature points found in image Q.[10]

2. Calculate the modified coordinate positions by matrix multiplication of H with the feature point positions.

3. For each modified q_i ($q_i \epsilon Q_I$) if there is a t_j ($t_j \epsilon T_I$) within a tolerance radius of $r \times \epsilon$, their β-value is then compared. If the β ratio is within T_s, then count it as a match.

4. Repeat step 3 for *external* dominant and convex points, *i.e.* each element of Q_E and Q_C with T_E and T_C respectively.

Consider M_I, M_E and M_C as the sets of internally, externally (concave) and convex matching feature points. Intuitively, more shape discrimination is present in internal and external dominant points while convex points add details (end points of protruding parts delimited by external (concave) points). Hence we make use of the following heuristic: our matching metric is biased towards internal and external (concave) dominant points. We express the problem of finding a best matching location of Q in T as the maximization of the F-measure:

$$ F = \frac{2 \times (|M_I| + |M_E|)}{(|Q_I| + |Q_E|) + (|T_I| + |T_E|)} \, , \tag{11} $$

where $|A|$ is the cardinality of set A. To handle the situations where many F-measures have same values, we then compute a percentage of matched convex points, F_C, and maximize this value:

$$ F_C = \frac{2 \times |M_C|}{(|Q_C| + |T_C|)} \tag{12} $$

There is more than one way to combine the information from the three sets of feature points. Currently, we are using first the internal dominant points to reduce the number of targets to further consider for a potential match. Then we add external (concave) dominant points to further reduce the number of candidates, and finally use convex points only if we still have multiple candidates left. Note that we do not use yet in practice additional information implicitly available, such as "codon" structure, *i.e.* how

---

[10] We use the traditional 3D graphics notation when performing affine transformation using 4×4 matrices; as we are only dealing with a 2D problem, one of the spatial dimensions is redundant, but this is not a problem in practice.

pairs of concavities are associated with specific convex points [44]. This structural information would be useful for finer matching under articulated motions (of limbs). We are currently exploring this combined use of concave and convex points and how to relate it to associated internal dominant points (and object parts), and this work will be reported elsewhere.

5 Experimental Results and Discussion

Our current matching algorithm is hierarchical in its use of feature points (internal (medial hot spots) –> external (concave tips) –> internal (convex tips)) and has proven useful even with common body articulated movements, but as pointed out above, this could be refined by using more structural information. At this stage in our research program we have performed an extensive experiment on different heterogeneous databases containing biological forms (large animals, plants and insects) to verify the performance in efficiency and accuracy of our novel proposed method.[11] Four main types of datasets were used for this purpose: (i) animals taking a static posture and in movement, (ii) humans taking a static posture or in action (articulated movement), (iii) insects and (iv) plants (only leaf forms). Also, we performed some random re-scaling, rotating and translating for the verification of invariance under such transformations (Fig. 20), and added a number of occlusions, by performing random cuts, to test the method's response beyond affine transforms. Furthermore, the robustness of the algorithm is also evaluated by applying a "structural noise" — introducing scalable random geometric deformations — which we designed by performing randomised morphological set operations on the segmented (binarised) objects. In our experiments we have defined three levels of such structural perturbations: small or less perturbed, medium and large or highly perturbed (examples in Fig. 19). We note that other methods relying on smooth continuous contours (such as methods based on the use of codons or curvature scale-space, as well as many of the traditional medial-axis methods) will have great difficulty in dealing with such deformations — which are to be expected in noisy image captures and under varying environmental conditions such as due to decay and erosion.[12]

To construct different databases we used the standard MPEG-7 [28], ImageCLEF-2013 [11] and Kimia (Brown University's) [47] datasets. Furthermore, we also initiated our own database where we se-

---

[11] Our first results on such data were presented in a short paper at the 1st International Workshop on "Environmental Multimedia Retrieval" (EMR) held in conjunction with the ACM International Conference on Multimedia Retrieval (ICMR) in Glasgow (UK), April 1, 2014 [1].

[12] We do not claim that this way of perturbing the data is physically accurate in modeling natural decay or erosion. Rather it provides a simple (computational) way to approximate such effects and produce deformations which appear (visually) credible in modeling these.
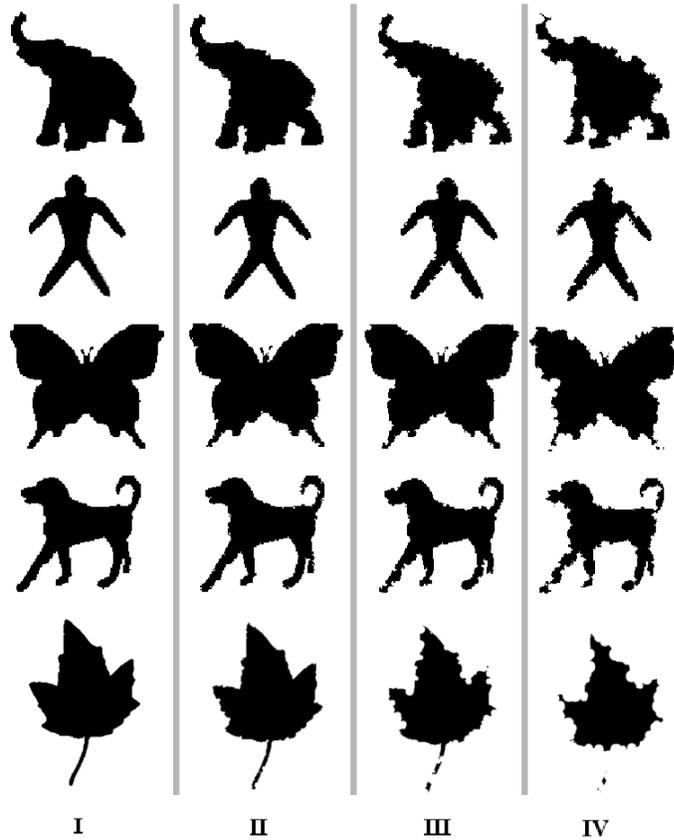
Fig. 19 Three levels of (added) perturbation: I - none (originals), II - small , III - medium and IV - large.

lected different sequences of animals in motion (from videos). From these datasets, we have collected a total of 2215 samples belonging to Animals other than insects (650 samples, including Human, Horse, Rat, Cat, Panther, Turtle, Elephant, Bat, Deer, Dog, and Ray forms), Insects (410 samples, including Butterfly, Bug, Mosquito, Ant, and other miscellaneous insect forms), and Plant leaves (1155 samples, including Acercampestre, Aceropalus, Acerplatanoides, Acerpseudoplatanus, Acersaccharinum, Anemonehepatica, Ficuscarica, Hederahelix, Liquidambarstyraciflua, Liriodendrontulipifer, Populusalba specimens).

We note that we are limiting the sizes of our test databases as we require segmented binary forms (distinguishing figure from ground) to initiate our medialness transform (*e.g.* ImageCLEF includes circa 5000 plant images, among which only 1155 contain clearly distinguishable leaf forms). This is a limitation of our current approach, but we are working on extensions to grey-level and color images (as non-segmented inputs). Also, rather than focus on one type of biological forms, say butterflies, we de-
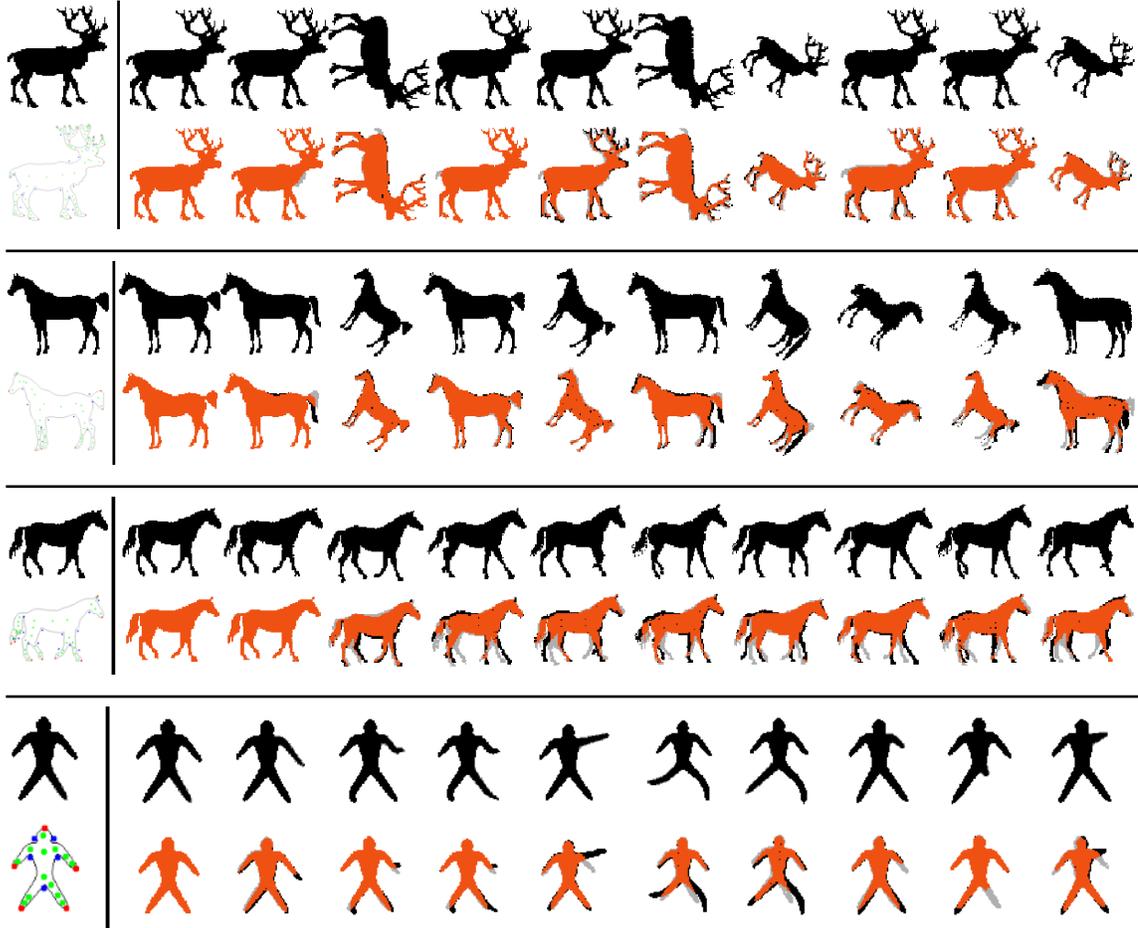
Fig. 20 Top-10 results on some of the collected samples from our extended database (containing 8860 forms): for deer, horse and humanoid forms when using our $F - measure$ as the basis for ranking. The leftmost image for each ranking results is the query, while the images on the top row are the target images matched at successive rank location from best to worst. The bottom row then shows the overlaying (in orange) of the query on the respective target image (after transformation) and their spatial differences (in grey). The top two series show tests for the invariance under scaling, rotation and translation. The third and fourth series show the behavior of matching using the $F - measure$ in the presence of articulated movements.

cided to test and show the potential power of our approach for a number of very different biological forms, from plant leaves, to various species of insects to larger animals (including humans).

Furthermore, to check the robustness of our algorithm, we deformed each such sample at the previously indicated three levels of perturbation, thus bringing our total dataset count to 2215×4 = 8860. From this database, we took each sample as a different query, resulting in a total comparison set of 8860×8860≈78.5 million possible pairings, and exploited our current ranking metric (F alone, or with F_C) to find the returned top-10 matches. Examples of such top-10 results can be seen in Figures 20, 21, 22, and 23.
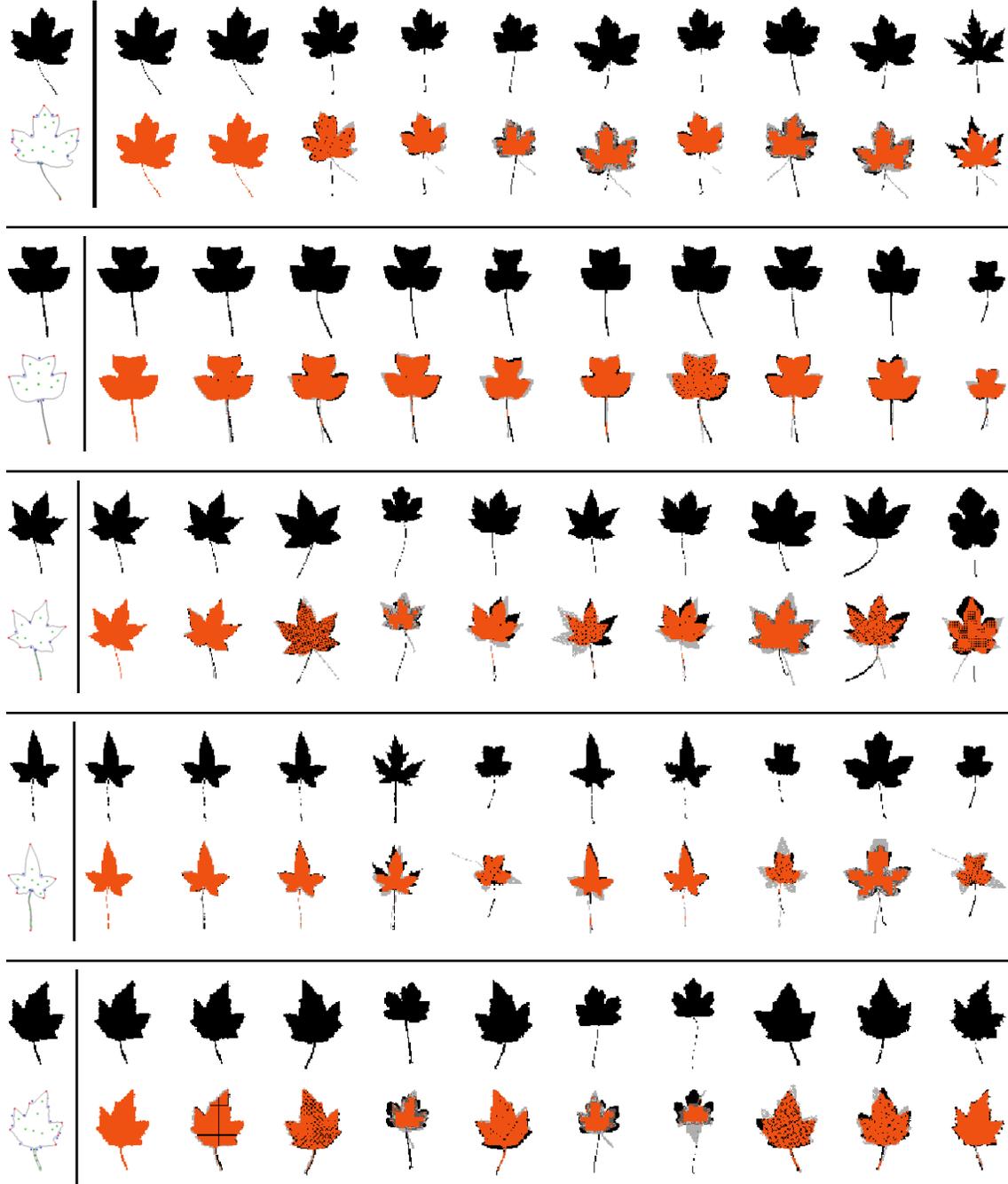
Fig. 21 Top-10 results on some of the samples from the *Plant* database (ImageCLEF-2013 [11]) *with shape perturbations* using our $F - measure$ as the basis for ranking. NB: The first match is always the desired target.
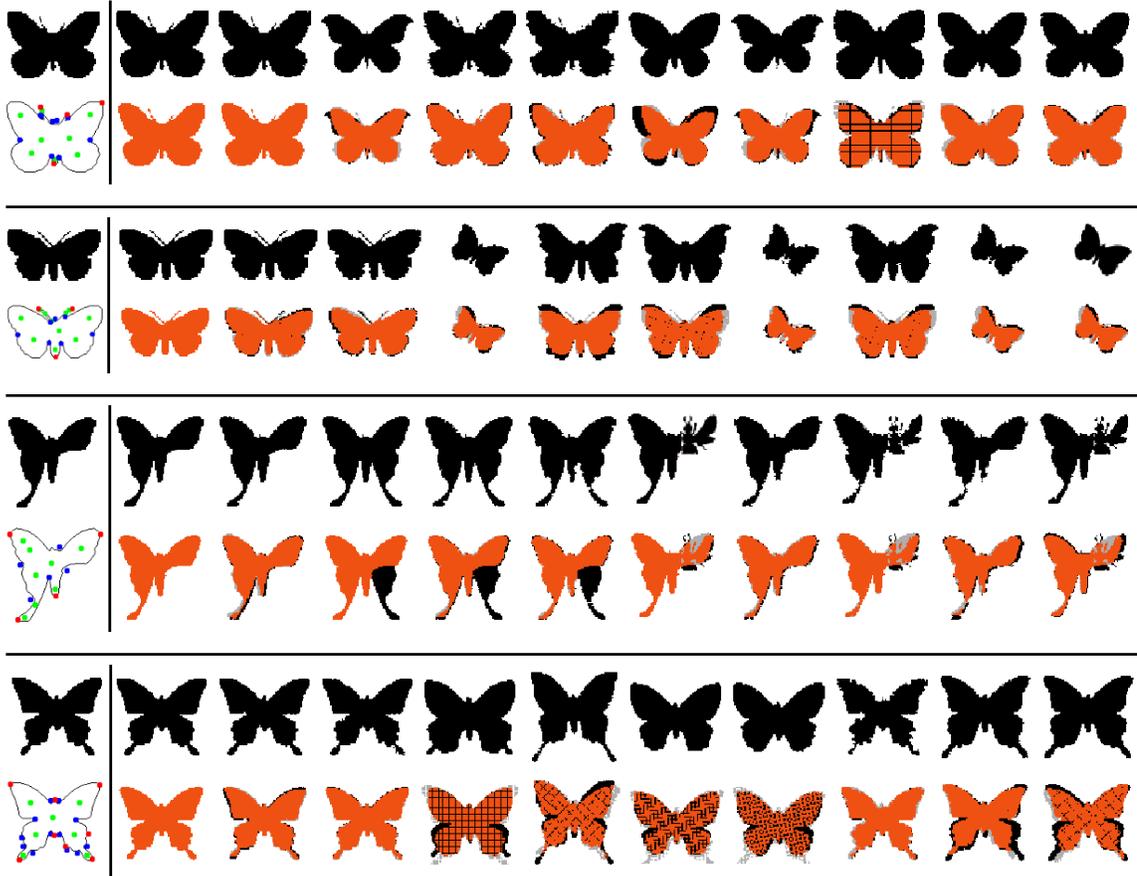
Fig. 22 Top-10 results on some of the samples for *butterfly* forms from the *Insect* database including structural perturbations when using our $F - measure$ as the basis for ranking. NB: A partial shape query (3rd series from the top) returns valid and interesting results.

First, in Figure 20 we notice the ranking follows a gradual degradation in spatial overlays (of query and target), but where the first 10 forms are all of the right type (amongst the current DB of 8860 forms). We also notice a graceful degradation in ranking when dealing with articulated planar movements (of a human and horse figures).

Second, in Figure 21 we study the use of our current matching method on plant leaf queries and also obtain a graceful degradation in rankings with a similar behavior in spatial form overlaps. We note that our feature points can be used to support a higher level semantic description of leafs for better taxonomic characterisation [13], in particular by permitting to identify: the apex, main lobes, and petiole using convex hot spots. The insertion point (where the petiole reaches the leaf) can be characterised by a codon: pair of bounding (significant) concavities together with the intermediate main convexity denoting the petiole). The apex and lobes extent and relative size can be obtained by finding their associated codon information (apex and lobes tips as convexities associated with nearby concavities).
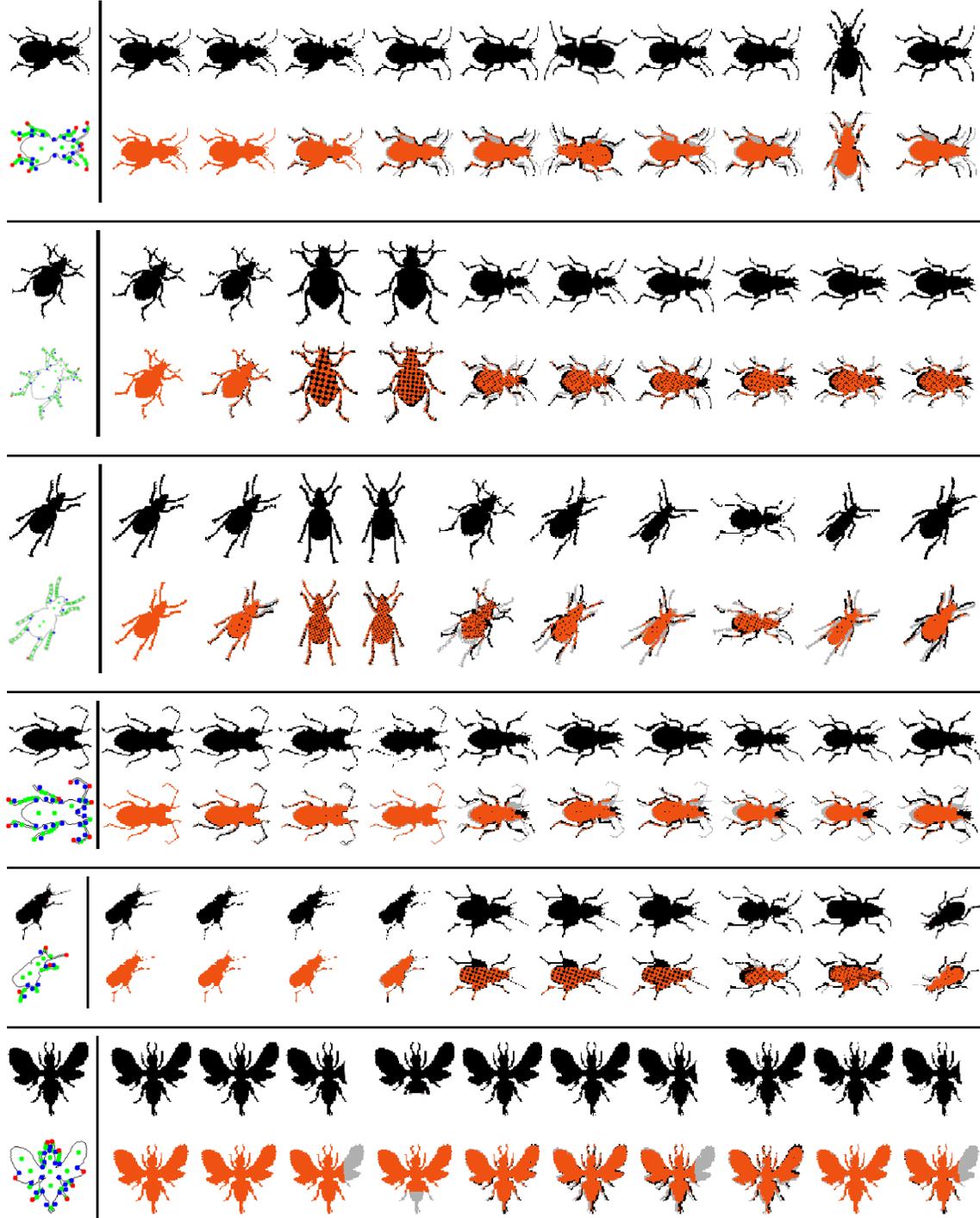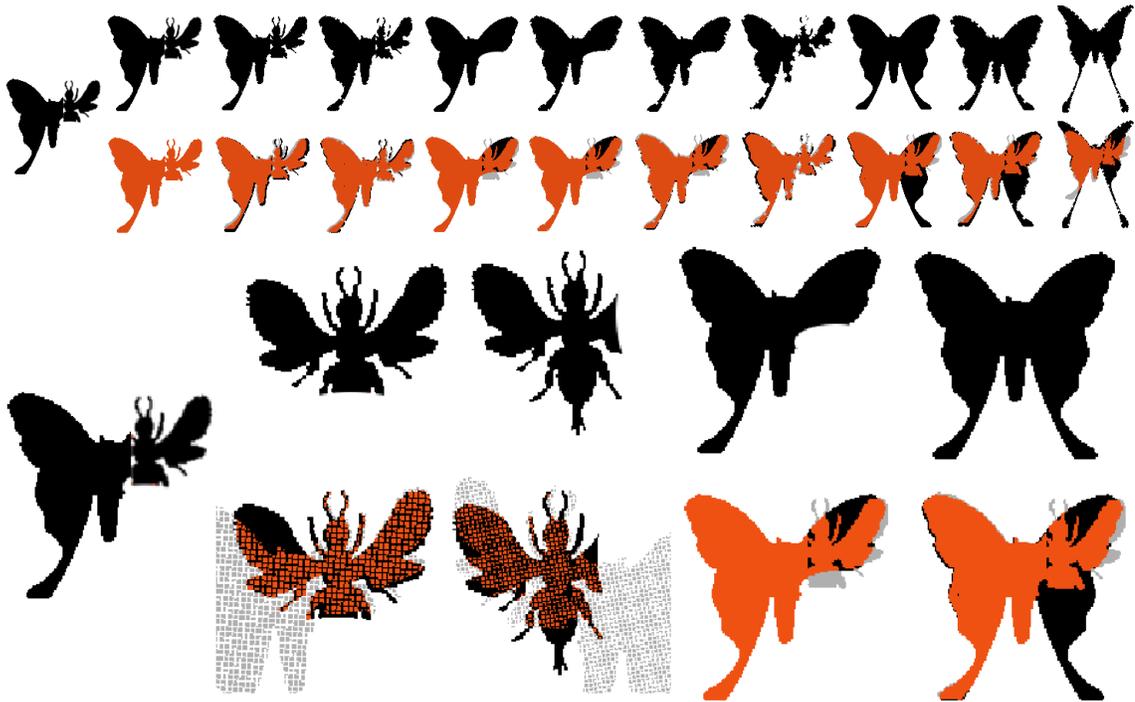
Fig. 23 Top-10 results on some of the samples from the *Insect* database including structural perturbations when using our $F - measure$ as the basis for ranking.

Fig. 24 A special query image obtained as a juxtaposition of two different insect cuts finds a best fitting location in the target image, and results into an interesting series of part-based matches (eventually retrieving the correct pair of individual insects (before juxtaposition) from the DB). The top series shows as usual the top 10 matches. The bottom series shows the interesting matches where the individual original insects are found (ranked 12th and 16th (for the smaller insect) and 4th and 8th for the larger butterfly.

If teeth (at the border of a leaf) are detectable, with sufficient resolution in the input image, then we can select a related scale of analysis, *i.e.* setting the annular gauge with corresponding parameters (R, ϵ), once the other previous main features have been identified (perhaps for a lower resolution input image). It would then be of interest to study the relation between the medial hot spots and main vein layout; the current approach in the literature is to directly process the textured input images to either segment out veins and then characterise their networks [40] or by directly extracting approximate feature points, *e.g.* using the Harris (corner) detector [38]. By simple visual inspection, some leafs with strong (well delimited) lobes and/or teeth (with clear associated convexity tips) may see their dominant veins well approximated by medial hot spots; but this is clearly not the case for many other leaf types with rather smooth outlines [27].

Third, in Figures 22, 23 and 24 we show some results with queries from insects forms. In Figure 22 we study the use of our current matching method on butterfly queries and again obtain a graceful degradation in rankings with a similar behavior in spatial form overlaps as for leaves and mammals. Note that the 3rd series indicated in that figure is for a butterfly with a large part cut away: we either get top

retrievals for butterflies with similar cuts or complete butterflies which represent likely completed forms (including the correct original butterfly, before being cut, ranked no. 3). Good matches are also found for an artificially constructed case of fusing the cut form with another smaller insect (see also Fig. 24). Fig. 23 gives more evidence for the quality of retrieved matches on other insect forms. Fig. 24 illustrates a special case where we fabricated an artificial query by merging together two different insects with large cuts to their original form. The returned top matches are interesting as they include the original two individual insects with varying degrees of cuts.

In our experiments, the matching algorithm always finds the best fitting shape area for the query in the target image (Fig. 24). For empirical analysis, we performed two individual comparisons: (a) precision obtained on different sets of data types (Fig. 25), and (b) precision obtained at different perturbation levels (Fig. 26). When the query image belongs to the original set, the retrieval rate at different ranks is very high, while the performance significantly decreases for high levels of perturbations. Note however that even under a large structural perturbation, a given form which may have lost some of its significant shape features (*e.g.* a limb) can still be matched with perceptually similar targets — as judged by a human observer and validated here as we know the ground truth.

Fig. 27 shows a comparative result analysis by computing precision at different ranking locations with respect to two of the currently most popular image matching methods: SIFT[36] and SURF[4]. The image features used by these methods (such as edge or corner responses) prove not particularly good on their own at capturing the shape structure. As a consequence, their accuracy reduces drastically when there is a slight deformation, occlusion and/or perturbation in the same shape. On our test database, already from the 2nd ranked match, SIFT and SURF both show a rapid reduction in accuracy (to circa 10%) in detecting a relevant target shape. While, on the other hand, our shape descriptor shows a slow fall in overall accuracy — from 85.7% at 2nd rank to 57% at 10th rank (Fig. 27).

For each query image, the system returns a ranked sequence of images. By computing the Average Precision (AP) we can obtain an overall measure to compare results and methods. AP is defined in the Information Retrieval literature as the average of precision measurements evaluated at the position of each of the relevant documents in the ranked sequences [15]:

$$AP = \frac{\sum\limits_{r=i}^{N} P(r)rel(r)}{I_r} , \tag{13}$$

where r is the rank of the current image in the returned image set, N is the number of retrieved images, $rel()$ is a binary function on the relevance of a given rank.[13] $P(r)$ and I_r are respectively the precision and the number of relevant images at the given cut-off rank. Furthermore, to test the performance of our system, we evaluated the Mean Average Precision (MAP) for a given set of queries:

$$MAP = \frac{1}{Q}\sum_{q=i}^{Q} AP(q)\,. \tag{14}$$

Here, Q is the number of queries in the dataset. The obtained comparative MAP graph at different ranking locations on our dataset is shown in Figure 28. MAP values for our method stay well above the 50% mark even for 10-th place in rankings, and decrease smoothly from 100% for 1st place rankings. The SIFT and SURF methods on the other hand show poor MAP results with the values 11% (SIFT) and 13% (SURF) for 10th place in ranking. This is not entirely conclusive as our database remains limited in size; furthermore, in future investigations, especially once implementations are available, we plan to compare our method with other ones which specifically rely on explicit shape measures other that simple edges or corner loci, such as "shape contexts" and its relatives [5,56,39], although most of these techniques typically are not adapted to deformations, articulations, structural perturbations, occlusions and cuts.

6 Conclusion

The strengths of our approach include: (1) emphasizing a representation of 2D shapes based on results from studies on human perception via robust medialness analysis; (2) introducing an algorithmic chain providing an implementation of this shape analysis tested on current reference 2D binary animal and plant databases; (3) mapping of the whole form into a small number of dominant feature points with associated shape significance (in terms of medialness, concavity, convexity); (4) achieving accuracy for top ranked matches (exactness) and with nice (observable) degradation properties for the following highest ranking results; (5) testing and demonstrating robustness under some levels of articulation

---

[13] The binary value of $rel(r)$ is set automatically by parsing a text file which contains the ground truth for each image: *i.e.*, which type it corresponds to, such as a dog, a horse, etc. Thus, the AP measure is only available given ground truth (identification) is provided.
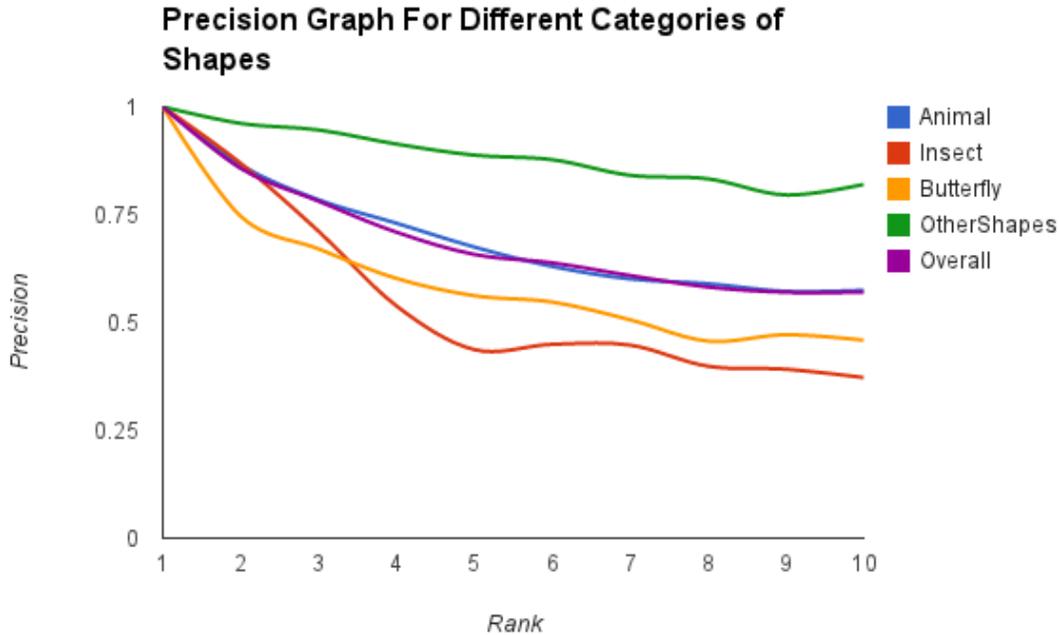
Fig. 25 Result analysis (precision) on the different datasets containing originals of large animal, plant and insect images, using our proposed $F - measure$ as the ranking metric.

and other structural (noise, cuts) perturbations. Our method also shows promising results for part-based matching tasks, in the context of occlusions, cuts and mixed object parts, as well as for shape reconstruction and compression, all important topics which will require further investigations.

The shape representation we have proposed to explore for the purpose of Information Retrieval (on objects present in images) is well rooted in established results from human visual perception (cognitive science [26,25,44,29]) and it also relates to the art of drawing and animation [31]. We are the first to propose and explore a possible algorithmic chain which implements the intuitive notion of "hot spots" first indicated by Kovács *et al.* [26,25]. We have refined and extended this notion in a number of ways: (i) we have introduced orientation to the boundary in the computation of medialness (metric D_ϵ^* (eqn.(3)) which reduces the effect of nearby extraneous contours (halo effect); (ii) we have proposed an adaptive scheme to automatically set the tolerance (or annulus width ϵ) in relation to an increasing disc minimum radius R_p (eqn.(5)); (iii) we have extended the computation of medialness as proposed by Kovács *et al.* to regions exterior to an object, in particular to characterise significant concavities; (iv) we have added to the notion of "hot spots" of Kovács *et al.* significant concave and convex points at the tip ends of ridges of medialness.
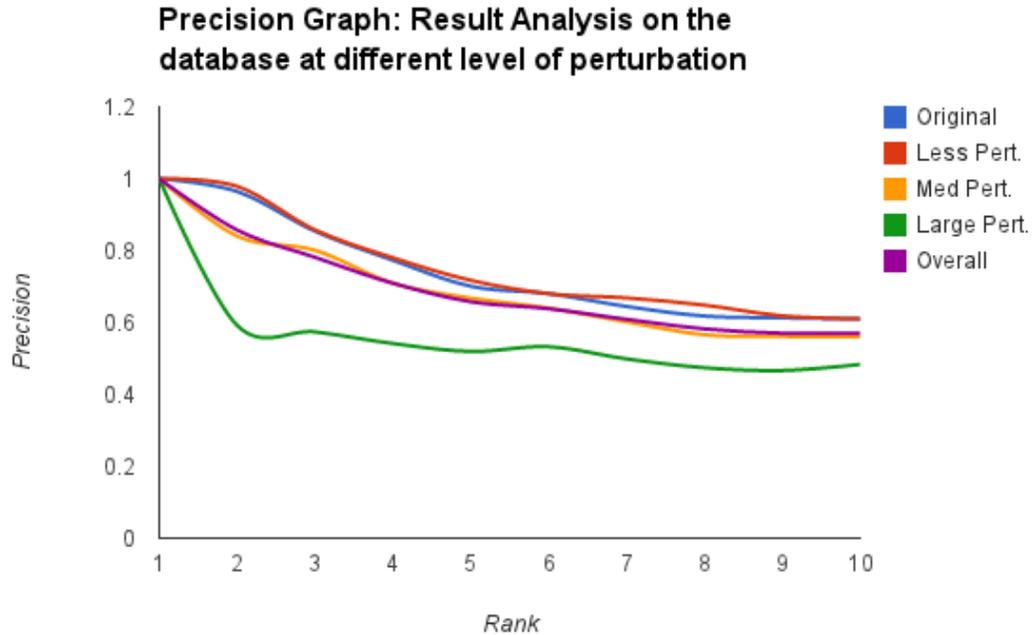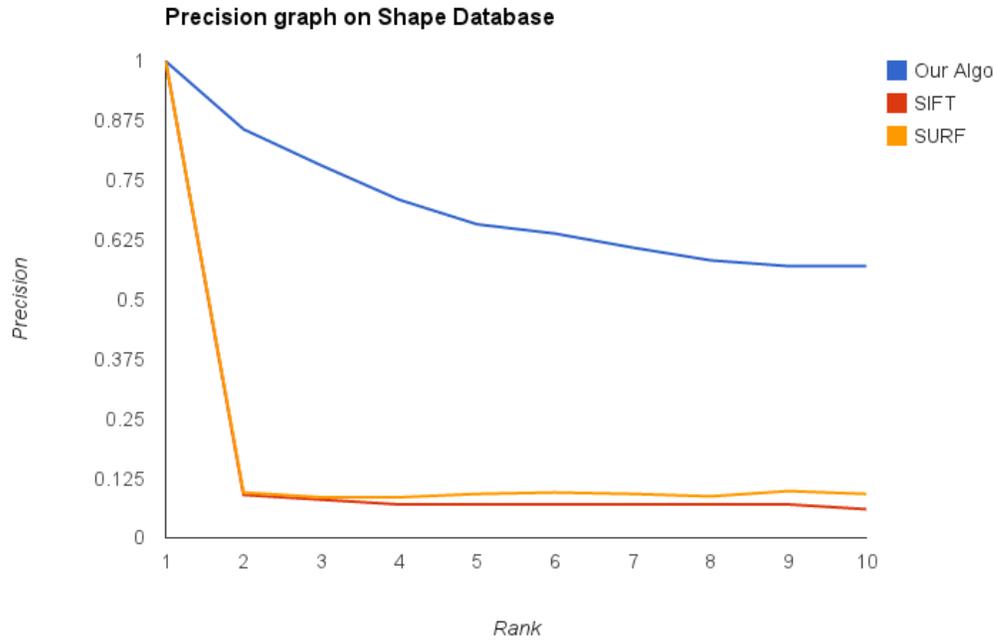
Fig. 26 Result analysis of the entire dataset for different levels of perturbations, using our $F - measure$ as the ranking metric. NB: Many of the images are also randomly rotated, scaled and translated.

The matching algorithm we have proposed represents a first step in exploiting this shape representation for its application in Information Retrieval problems for objects in images. In this paper we describe a simple method based on (a) invariance to affine transformations and (b) applied in a hierarchical manner by focusing first on interior dominant points, followed by concave and convex information. As our method is based on feature points it can naturally be extended to build information vectors in relation to each identified feature point, *e.g.* by considering the associated minimum radius, boundary orientation, curvature support, image intensity or color, local image texture. This sort of extension will be explored as we grow the databases upon which we can apply shape-based representations.

One area of extension to explore is to study how such a point-based representation built from medialness may be combined with other existing more classical Point Pattern Matching (or PPM) techniques [51].[14] This class of methods often rely on large samples of point distributions over grids as in morphometrics [9], and the automatic selection of highly informative feature points as landmarks remains

---

[14] An example of such a combination of methods to exploit their relative strengths has recently been proposed by Nanni *et al.* who use [39]: Shape Contexts, Inner Distance, Height Functions, Shapelets, traditional Curvature, Fast Radial Symmetry Transform, Local Phase Quantization, Histogram of Gradients, Wavelets. Shape descriptors (contour based) are compared using a Weighted Spectral Distance to measure dissimilarity, while image texture descriptors are compared using the Jeffrey divergence operator.

Fig. 27 Comparative result analysis (precision) of the entire dataset at different ranking locations, using our shape descriptor with $F - measure$ as the ranking metric w.r.to SIFT[36] and SURF[4]. NB: Many of the images are also randomly rotated, scaled and translated.

an active area of research [18,23]. Also, rather than comparing our method only with readily available texture-based well-known methods, such as SIFT and SURF which are well suited to deal with large databases of images, we ought to also compare with more closely related methods such as the Inner Distance Shape Context method — one of the few other published method which is capable to explicitly deal with planar articulations [16] — and other recent related methods (*e.g.* [19,55,46]); but notice that it is unclear how these other shape-based techniques scale with a database size (as they are often only tested on relatively small datasets of at most a few hundred targets). This will require accessing good implementations of such methods which remain a limitations when attempting a careful examination of pros and cons of various methods.

We emphasise that our motivation is inter-disciplinary: we wish to be able to address semantic gaps and mimic human performance, and this explains our exploitation of results from cognitive science in order to design a notion of feature points which relate explicitly to established models of the human perception of shape. The original perceptual model of medialness proposed by Kovács *et al.* was studied in
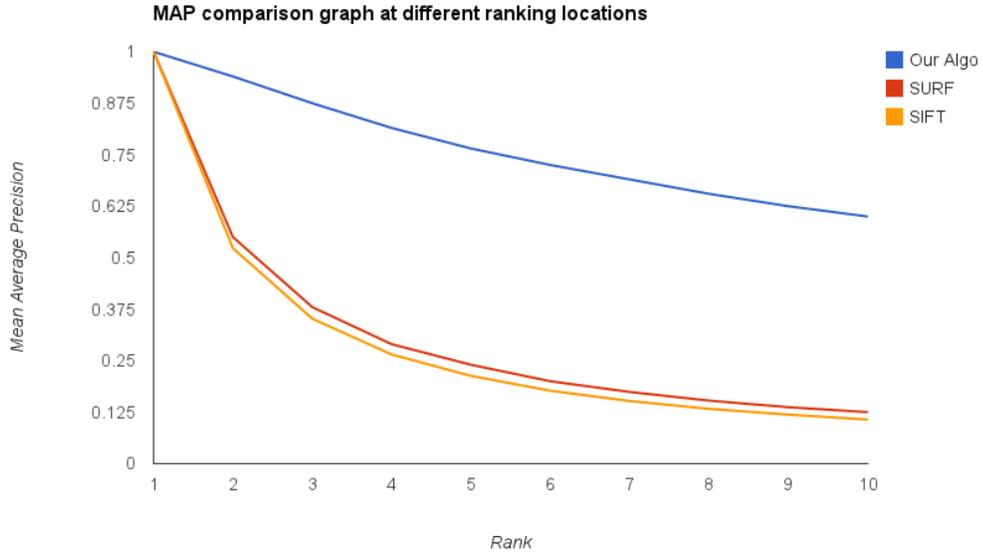
Fig. 28 Result analysis of the entire dataset by evaluating Mean Average Precision (MAP) at different ranking locations, using our $F-measure$ as the ranking metric with SIFT[36] and SURF[4]. NB: Many of the images are also randomly rotated, scaled, translated and perturbed.

the context of the articulated movement of biological forms (such as a running animal) [26]. Our results presented in this communication confirm the potential of this method when applied to the Information Retrieval problem in dealing with image sequences of articulated movement of animals, and is proving also powerful for static scenarios when accessing snapshots of either animals or plants.

A Appendix: F-Measure

In classical Information Retrieval, the balanced F-score is defined as the harmonic mean of $Precision$ and $Recall$ [45]:

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

Our derived formula of $F - measure$ (eqn. 11) is based on this equation:

$$F = \frac{2 \times (|M_I| + |M_E|)}{(|Q_I| + |Q_E|) + (|T_I| + |T_E|)} .$$

The classical definitions of Precision and Recall are given as:

$$Precision = \frac{t_p}{t_p + f_p} \tag{16}$$

$$Recall = \frac{t_p}{t_p + f_n} \qquad (17)$$

where, t_p= true positive, f_p= false positive, and f_n= false negative. Consider the case of internal medial (dominant) and external (concave) points as the input of this evaluation metric, then:

t_p= #(feature points matched correctly) = $|M_I| + |M_E|$,

f_p = #(feature points in the query image that are not matched), and

f_n = #(feature points in the target image that are not matched).

Hence:

$t_p + f_p$ = #(feature points in the query image) = $|Q_I| + |Q_E|$, and

$t_p + f_n$ = #(feature points in the target image) = $|T_I| + |T_E|$.

Putting these values in equations (16) and (17) and further using equation (15) results in our definition of F-measure (equation 11).

References

1. Prashant Aparajeya and Frederic Fol Leymarie. Point-based medialness for animal and plant identification. In Stefanos Vrochidis et al., editors, *Proceedings of the 1st International Workshop on Environnmental Multimedia Retrieval*, volume 1222, pages 14–21, Glasgow, UK, April 2014. CEUR-WS.org.

2. Rudolf Arnheim. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, new version expanded and revised edition of the 1954 original edition, 1974.

3. X. Bai, W. Liu, and Z. Tu. Integrating contour & skeleton for shape classification. In *ICCV Workshops*, pages 360–7, 2009.

4. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

5. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

6. S. Berretti et al. Retrieval by shape similarity. *IEEE Trans. on Multimedia*, 2(4):225–39, 2000.

7. Irving Biederman. Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision*, 13:241–253, 2001.

8. H. Blum. Biological shape and visual science (part I). *Journal of Theoretical Biology*, 38:205–287, 1973.

9. Fred L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

10. C. Bregler et al. Turning to the masters: Motion capturing cartoons. *ACM Trans. on Graphics*, 21(3):399–407, 2002.

11. B. Caputo et al. ImageCLEF 2013: The vision, the data and the open challenges. In *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pages 250–268. Springer, 2013.

12. L. Chen et al. Efficient partial shape matching using Smith-Waterman algorithm. 2008. CVPR Wksh., pp. 1–6.

13. James Cope, David Corney, Jonathan Clark, Paolo Remagnino, and Paul Wilkin. Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*, 39(8):7562–7573, 2012.

14. E.R. Dougherty and R.A. Lotufo. *Hands-on morphological image processing*. Tutorial Texts in Optical Engineering, Vol. TT59. SPIE Publications, 2003.

15. Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88:303–338, 2010.

16. Raghuraman Gopalan, Pavan Turaga, and Rama Chellappa. Articulation-invariant representation of non-planar shapes. In *Computer Vision — ECCV 2010*, volume 6313 of *Lecture Notes in Computer Science (LNCS)*, pages 286–299. Springer.

17. M. Guay et al. The Line of Action: An intuitive interface. *ACM Trans. on Graphics*, 32(6):Article no. 205, 2013.

18. Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, August 2009.

19. Rong-Xiang Hu, Wei Ja, Yang Zhao, and Jie Gui. Perceptually motivated morphological strategies for shape retrieval. *Pattern Recognition*, 45:3222–3230, 2012.

20. Seymour Simmons III and Marc. S. A. Winer. *Drawing: The Creative Process*. Simon and Schuster (Prentice-Hall), 1977.

21. Greet Kayaert, Johan Wagemans, and Rufin Vogels. Encoding of complexity, shape, and curvature by macaque infero-temporal neurons. *Frontiers in Systems Neuroscience*, 5(51), 2011.

22. M. Kelly and M. D. Levine. Annular symmetry operators. In *ICCV*, pages 1016–21, 1995.

23. Johannes Keustermans, Dirk Vandermeulen, Wouter Mollemans, Filip Schutyser, and Paul Suetens. Construction of statistical shape models using a probabilistic point-based shape representation. In *Symposium on Statistical Shape Models and Applications*, number Article 21, Delémont, Switzerland, June 2014. http://shapesymposium.org/.

24. B.B. Kimia. On the role of medial geometry in human vision. *Journal of Physiology – Paris*, 97(2):155–90, 2003.

25. I. Kovács. "Hot spots" and dynamic coordination. In *Dynamic Coord. in the Brain*. MIT Press, 2010. pp. 215–28.

26. I. Kovács, Á. Fehér, and B. Julesz. Medial-point description of shape. *Vision research*, 38(15):2323–33, 1998.

27. Mónica G. Larese et al. Automatic classification of legumes using leaf vein image features. *Pattern Recognition*, 47:158–168, 2014.

28. Longin Jan Latecki and Rolf Lakamper. Shape similarity measure based on correspondence of visual parts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1185–1190, 2000.

29. Oliver W. Layton, Ennio Mingolla, and Arash Yazdanbakhsh. Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Frontiers in Psychology*, 5(Article 972):20 pages, September 2014. Perception Science Series.

30. F. Leymarie and M. D. Levine. Simulating the grassfire transform using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1):56–75, January 1992.

31. F. Fol Leymarie et al. Point-based medialness for movement computing. In ACM Proc. of MOCO, pp.31–6, Paris, France, 2014.

32. Michael Leyton. *Symmetry, Causality, Mind*. MIT Press, 1992.

33. H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE PAMI*, 29(2):286–99, 2007.

34. Z. Liu, J. An, and F. Meng. A robust point matching algorithm for image registration. In *ICMV*, volume SPIE 8350, 2011.

35. Andrew Loomis. *Successful Drawing*. Viking Books, 1951.

36. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

37. Y. Mingqiang et al. A survey of shape feature extract. tech. In *Pattern Rec. Tech. & Applic.* pp. 43–90, InTech, 2008.

38. Sofiene Mouine, Itheri Yahiaoui, and Anne Verroust-Blondet. Advanced shape context for plant species identification using leaf image retrieval. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR)*, number Article 49, page 8 pages, 2012.

39. Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Ensemble of shape descriptors for shape retrieval and classification. *International Journal of Advanced Intelligence Paradigms*, 6(2):136–156, 2014.

40. JinKyu Park et al. Utilizing venation features for efficient leaf image retrieval. *J. of Systems & Software*, 81(1):71–82, 2008.

41. Stephen M. Pizer et al. Deformable M-reps for 3D medical image segmentation. *International Journal of Computer Vision*, 55(2/3):85–106, 2003.

42. Stephen M. Pizer, Kaleem Siddiqi, Gabor Székely, James N. Damon, and Steven W. Zucker. Multiscale medial loci and their properties.

43. Vittal Premachandran and Ramakrishna Kakarala. Perceptually motivated shape context which uses shape interiors. *Pattern Recognition*, 46:2092–2102, 2013.

44. W. Richards and D. D. Hoffman. Codon constraints on closed 2D shapes. *CVGIP*, 31(3):265–81, 1985.

45. C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 2nd edition, 1979.

46. Edgar Roman-Rangel, Carlos Pallan Gayol, and Jean-Marc Odobez. Searching the past: An improved shape descriptor to retrieve Maya hieroglyphs. In *ACM Multimedia*, Scottsdale, Arizona, USA, 2011.

47. T. Sebastian et al. Recognition of shapes by editing their shock graphs. *IEEE PAMI*, 26(5):550–71, 2004.

48. Jean Serra, editor. *Image Analysis and Mathematical Morphology*, volume 2. Academic Press, 1988.

49. Wei Shen, Xinggang Wang, Cong Yao, and Xiang Bai. Shape recognition by combining contour and skeleton into a mid-level representation. In Shutao Li, Chenglin Liu, and Yaonan Wang, editors, *Pattern Recognition*, volume 483 of *Communications in Computer and Information Science*, pages 391–400. Springer, 2014.

50. P. Srestasathiern and A. Yilmaz. Planar shape representation and matching under projective transformation. *CVIU*, 115(11):1525–35, 2011.

51. Jun Tang, Ling Shao, and Simon Jones. Point pattern matching based on line graph spectral context and descriptor embedding. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 17–22, March 2014.

52. G.J. van Tonder and Y. Ejima. Flexible computation of shape symmetries. *IEEE SMC, Part B*, 33(3):535–40, 2003.

53. L. Vincent. Morphological grayscale reconstruction in image analysis. *IEEE Trans. Ima. Process.*, 2(2):176–201, 1993.

54. P. B. Van Wamelen, Z. Li, and S. S. Iyengar. A fast expected time algorithm for the 2-D point pattern matching problem. *Pattern Recognition*, 37(8):1699–711, 2004.

55. Xinggang Wang, Bin Feng, Xiang Bai, Wenyu Liu, and Longin Jan Latecki. Bag of contour fragments for robust shape classification. *Pattern Recognition*, 47(6):2116–2125, 2014.

56. Jun Xie, Pheng-Ann Heng, and Mubarak Shah. Shape matching and modeling using skeletal context. *Pattern Recognition*, 41:1756–1767, 2008.