Aparajeya, Prashant and Leymarie, Frederic Fol

Point-based Medialness for Animal and Plant Identification


Available at: http://research.gold.ac.uk/11007/

COPYRIGHT
All material supplied via Goldsmiths Library and Goldsmiths Research Online (GRO) is protected by copyright and other intellectual property rights. You may use this copy for personal study or research, or for educational purposes, as defined by UK copyright law. Other specific conditions may apply to individual items.

This copy has been supplied on the understanding that it is copyright material. Duplication or sale of all or part of any of the GRO Data Collections is not permitted, and no quotation or excerpt from the work may be published without the prior written consent of the copyright holder/s.
ABSTRACT

We introduce the idea of using a perception-based medial point description [9] of a natural form (2D static or in movement) as a framework for a part-based shape representation which can then be efficiently used in biological species identification and matching tasks. The first step is one of fuzzy medialness measurements of 2D segmented objects from intensity images which emphasises main shape information characteristics of an object’s parts (e.g. concavities and folds along a contour). We distinguish interior from exterior shape description. Interior medialness is used to characterise deformations from straightness, corners and necks, while exterior medialness identifies the main concavities and inlands which are useful to verify parts extent and reason about articulation and movement. In a second step we identify a set of characteristic features points built from three types. We define (i) an Interior dominant point as a well localised peak value in medialness representation, while (ii) an exterior dominant point is evaluated by identifying a region of concavity subtended by a minimum angular support. Furthermore, (iii) convex points are extracted from the form to further characterise the elongation of parts. Our evaluated feature points, together are sufficiently invariant to shape movement, where the articulation in moving objects are characterised by exterior dominant points. In the third step, a robust shape matching algorithm is designed that finds the most relevant targets from a database of templates by comparing the dominant feature points in a scale, rotation and translation invariant way (inspired by the SIFT method [17]). 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 scales.

Keywords

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

1. INTRODUCTION

In this short communication we introduce our proposed 2D shape representation for biological objects (animals and plants) which is inspired by results and techniques from cognitive psychology, artistic rendering and animation and 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 (here approximate disks of various radii, Fig. 1(a)). Different body movements are characterised by a particular orientation and combination of these primitives. 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 [9].

Inspired with these two approaches to the perception of natural motions, 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. The main
advantage over other classical medial-based representations of 2D shape is one of combined compactness, robustness and capacity of dealing with articulated movements.

Shape representation towards matching has been addressed in many ways by computer scientists in recent years, including by directly characterising and grouping contour points \cite{24, 15}, by contour analysis \cite{2, 10, 5, 21}, using Blum’s medial axis transform (MAT) and its related shock graph \cite{8}, combining contour and skeleton \cite{11}, \cite{1}, or using instead the inner distance \cite{14} (some of the main medial representations are illustrated in Fig. 1 including our proposed method). Most related to our approach are contour enclosure-based symmetries \cite{7} and medial point transform \cite{23}, which compute similar medialness maps, but do not apply these to the retrieval of dominant points and their use in shape matching. Other classical approaches emphasise similarly either boundary information (e.g. Fourier, wavelet and scale-space analyses of closed contours) or interior information (e.g. primitive retro-fitting or approximation) \cite{8}. 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 DB). We propose instead to find an efficient medial 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.

Our current method is to develop a shape matching algorithm which is invariant to translation, scale and rotation and is inspired by the now classic SIFT approach \cite{16, 17}. Our shape representation is derived from the region-based medial point description of shape proposed by Kovács et al. \cite{7} in cognitive science and perception studies. The purpose of evaluating such medialness measurement is to provide a description of the shape which is local, compact, can easily be applied at different spatial scales, and mimicks human sensitivity to contour stimuli. This process maps the whole shape information into a few number of points we call “dominant” and hence makes it compact. Contrarily to classical medial-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). We augment the medial dominant points with main contour points indicating significant convex and concave features, thus bringing together with our notion of medialness the main 2D point-based shape systems proposed over the years in the fields of cognitive psychology and computer vision: the so-called “codons” denoting contour parts \cite{19} and high curvature convexities often used in scale-space analyses \cite{3}.

Mathematically, medialness of a point in the image space is defined as the containment of sets of boundary segments falling into the annulus of thickness $\epsilon$ (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.

Figure 2: From [9, Fig.2] with permission: the $D_\epsilon$ function for a simple shape defined as an accumulation of curve segments falling inside the annulus neighborhood of thickness $\epsilon$ (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.

Mathematically, $D_\epsilon$ is defined as:

$$D_\epsilon(p) = (t_4 - t_3) + (t_2 - t_1)$$

where $t_1 = 0$.

To identify internal dominant points a morphological top-hat transform \cite{13} is applied to isolate peaks in the medialness signal. Peaks are filtered using an empirically derived threshold. The selected peaks are then each characterised by a single representative point. To avoid considering large numbers of nearby isolated peaks which are characteristic of object regions with many small deformations, only peaks at a given minimum distance away from each other are retained. The extraction process of external dominant point is achieved by combining a concavity measure together with length of support on the contour. Again, a spatially localised filtering is applied to isolate representative dominant points. Furthermore, to improve the robustness of our representation, we extracted the set of convex points to capture the blob like structure from the shape. We have observed that the articulation and movement of limbs can be captured via such additional dominant points. Together, the selected dominant points (internal and external) and convex points are then considered as the representative feature points of the shape. Our matching algorithm 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 dominant 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.t the target image. These values are then applied over the query image to find the best possible matching location in the target image.

2. MEDIALNESS MEASURES & FEATURE EXTRACTION

A medial point is defined by computing the $D_\epsilon$ function as a distance metric (to boundary segments). The $D_\epsilon$ value at any point in transformed space represents the degree to which this point is convex and concave associated with a percentage of bounding contour pixels of the ob-
We further impose that no remaining points of locally maximised

els in width) is applied over the top-hat image such that within the

selected medialness points which tend to form clusters. To do so, a

flat circular structuring element of radius

is taken as the smallest distance

between

any pair of

as a grey-level 2D function) with the morphological opening

of this image by a flat structural element (a disk with radius as a pa-

parameter). Opening is a set operator on functions which “removes” small objects from the foreground of an image, placing them in the background (augmenting the local function set values).

This filtering is followed by a thresholding to discard remaining ar-

eas of relatively low medialness significance. Figure 4(b) shows

the result obtained after applying the white top-hat transform on a

dog in a typical standing posture to illustrate how the increasing

value of tolerance (c) operates an averaging effect on medialness,

as previously observed by Kovács et al. [9]; as c increases,

smaller symmetries are discarded in favor of those at a larger scale.

2.1 Dominant Points Extraction

Medialness measurement is currently done separately for internal

and external regions to take advantage of the perceptual figure-

ground dichotomy known to be a powerful cue in humans. This

also enables our method to consider articulated objects as potential

targets in pattern recognition tasks.

2.1.1 Internal Dominant Points

Medialness increases with “whiteness” in our transformed im-

ages (visualisation of medialness). To select points of internal dom-

inance, a “white” top-hat transform is applied, resulting in a series

of bright white areas. The white top-hat transform is defined as the
difference of an input function (here an image of medialness mea-

sures as a grey-level 2D function) with the morphological opening

of this image by a flat structural element (a disk with radius as a pa-

arameter). Opening is a set operator on functions which “removes” small objects from the foreground of an image, placing them in the background (augmenting the local function set values) [23] [9].

This filtering is followed by a thresholding to discard remaining ar-

eas of relatively low medialness significance. Figure 4(b) shows

the result obtained after applying the white top-hat transform on a

medialness image.

We still require to process further the output of the top-hat trans-

form to isolate the most dominant points amongst the remaining

selected medialness points which tend to form clusters. To do so, a

flat circular structuring element of radius \( \epsilon / 2 \) (but of at least 2 pix-

els in width) is applied over the top-hat image such that within the

element it produces only that value which maximises medialness.

We further impose that no remaining points of locally maximised

medialness are too close; this is currently implemented by imposing

a minimum distance of length 2\( \epsilon \) is taken between any pair of

selected points. We have found that in practice this is sufficient to

avoid the clustering of final interior dominant points (Fig 4(c)).

2.1.2 External Dominant 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 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. 5 left). On the other hand, if a point is located very close to the contour, it can easily be due to noise or small perturbations in the boundary. Therefore, to be able to retrieve reliable external dominant points, it is first re-

quired to provide an adapted definition of concavity as a significant shape feature.

Figure 5: Left: External medialness processing on a humanoid. The articulated movement of the left arm changes the location and orientation of the associated external dominant point (at 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. Blue arrows show the local support for concavity while brown ar-

rows indicate the direction of flow of medialness (away from the concavity). Right: Top: Detection of concave regions (on a butterfly object) using angular support. Bottom: Detected Conc-

ave points.

We define a point (contour 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 (\( \epsilon \)). The value

of threshold (\( \theta \)) is tunable but is always less than \( \pi \), which permits

control to the angular limit of the concave region. A point whose

local concavity is larger than \( \theta \) is considered a “flat” point. In our

experiments we tuned the value of \( \theta \) from \( 5\pi / 6 \) to \( 8\pi / 9 \). In as-

association, we define an external circular region (of radius function

of \( \epsilon \)) centered at each locus containing candidate external dominant

points. Each such region may provide only one representative dom-

inant point, where the dominance of a particular point is decided by

the maximum containment of boundary points inside the associated

annulus (of medialness) and corresponds to the maximum length of

support. Finally, we position the representative dominant point to

be near the contour at a fixed distance outside the form (Fig. 5)

right).

2.1.3 Convex Points

Our final shape feature is a set of convex points, where a shape

has sharp local internal bending and gives a signature of a blob-like

Figure 4: Illustration of the three successive steps in isolating internal dominant points: (a) medialness representation of (the interior of) a standing dog figure; (b) corresponding top-hat transform; and (c) internal dominant points illustrated as black dots together with the original object’s boundary.
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 shape [19, 11, 2, 21]. The process of extraction of convex points is similar to the extraction of concave points, the main difference being the value of threshold angle ($\theta$), where $\pi < \theta \leq 2\pi$. In our experiments we have found useful values to be in the range: $5\pi/4$ to $4\pi/3$. Such convex and concave points are complementary to each other 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 [19]. In our case we relate these two sets with the extremities of the traditional medial axis of H. Blum: end points of interior branches correspond to center of positive extrema of curvature and end points 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 [13]. Together, the three sets: concave, convex and interior dominant, form a rich enough point-based description of medialness to allow us to efficiently address applications with articulated movement for real image data.

2.2 Articulation

Anatomically, an animal’s articulated movement is dependent on the point of connection between two bones or elements of a skeleton. Our results show that exterior dominant points (representative of significant concavities) have higher potential to trace such articulations, unless the shape is highly deformed [12]. For usual movements (e.g. walking or jogging), these feature points remain present and identifiable in association to an underlying bone junction and hence can provide a practical signature for it (Fig. 3 bottom).

3. MATCHING ALGORITHM

Our objective is to design a robust matching algorithm that will match dominant points (query to targets) in an efficient way, in both time (or equivalent numerical complexity) and accuracy. First, both the internal and external dominant points are separately extracted from query and target images. Internal dominant points are the keypoints for evaluating scale, rotation and translation of the query image w.r.t. to a target image. After finding the scale, rotation and translation of the query image, the next task is to improve the correctness of the matching algorithm. For this, external dominant points can play a role and improve the final accuracy. The following information is associated with the dominant points:
Stage I finds the scale ($\beta$) and translation of the query image by matching a dominant point. Stage II is a check on the scale $\beta$ 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.

Stage I: Take an element ($q_i$) from set $Q_1$ and match it with each element ($t_j$) of set $T_1$. For each pair of ($q_i,t_j$), the scale ($\beta$) is evaluated as: $\beta = \frac{r_{t_j}}{r_{q_i}}$ (Fig. 5). 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: $q_{i\beta x} = t_j - q_i$ and $q_{i\beta y} = t_j - q_{iy}$, where $q_i = (q_{ix},q_{iy})$ and $t_j = (t_{jx},t_{jy})$.

Stage II: Now take the next element ($q_{i+1}$) from set $Q_1$ and match it again with each element ($t_j$) of set $T_1$. For each pair of ($q_{i+1},t_j$) find the scale $\beta'$. If the ratio of $\beta' > \beta$ under the tolerance level $T_x$, then go to stage III. Otherwise, repeat at another point and check the same tolerance criterion $T_x$ and repeat if necessary until all the elements of $Q_1$ 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. 7).

Stage III: We define the orientation ($\alpha$) 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 $Q_1$ and ($t_j,t_{j+k}$) are matched dominant points in $T_1$, then orientations $\alpha_q$ and $\alpha_t$ are defined as the angle between matching dominant points (Fig. 7). The rotation ($\theta$) of the image $Q$ is thus defined by the difference of orientations, i.e. rotation($\theta$) = $\alpha_t(t_{j+k}) - \alpha_q(q_{i+k})$.

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 ($Q_1, Q_2$ 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 $4 \times 4$ homogeneous matrix $H$ to perform the required linear (rotation and scaling) and affine (translation) transforms for all feature points found in image $Q$.
2. Calculate the modified coordinate positions by matrix multiplication of $H$ with the feature point positions.
3. For each modified $q_i$ in ($q_i \in Q_1$) if there is a $t_j$ ($t_j \in T_1$) within a tolerance radius of $r \times \epsilon$, their $\beta$-value is then compared. If the $\beta$ ratio is within $T_x$, 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 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 convex points (of limbs). We are currently exploring this combined use of external dominant points. Where on its potential advantage.

Figure 7: For both query (test) and target images, the orientation ($\alpha$) 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 ($\theta$) of the query image w.r.to the target is given by the difference in orientations.
verification of invariance under such transformations (Fig. 9), 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 perturbations: small or less perturbed, medium and large or highly perturbed (examples in Fig. 8). 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.

Figure 8: Three levels of (added) perturbation: I - none (originals), II - small, III - medium and IV - large.

To construct different databases we used the standard MPEG-7 [10], ImageCLEF-2013 [4] and Kimia (Brown University’s) [20] datasets. Furthermore, we also initiated our own database where we selected different sequences of animals in motion (from videos). From these datasets, we have collected a total of 1130 samples belonging to Animals other than insects (520 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 (200 samples, including Acer campestre, Acer opalus, Acer platanoides, Acer pseudoplatanus, Acer saccharinum, Anemone hepatica, Ficus carica, Hederahelix, Liquidambar styraciflua, Liriodendron tulipifera, and Populus alba specimens).

We note that we are limiting the sizes of our test databases as we require well segmented binary forms (distinguishing figure from ground) to initiate our medialness transform (e.g. ImageCLEF includes circa 5000 plant images, while we have segmented only 200 of these thus far). 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 decided 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 $1130 \times 4 = 4520$. From this database, we took each sample as a different query, resulting in a total comparison set of $4520 \times 4520 = 20.43 \text{ million}$ forms, 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 9, 10, 11, and 12. In our experiments, the matching algorithm always finds the best fitting shape area for the query in the target image (Fig. 13). For empirical analysis, we performed two individual comparisons: (a) precision obtained on different sets of data types (Fig. 14), and (b) precision obtained at different perturbation levels (Fig. 15). 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.
Figure 10: Top-10 results on some of the samples from the Plant database (ImageCLEF-2013 [4]) with shape perturbations using our $F$-measure as the basis for ranking. NB: The first match is always the desired target.

Figure 11: 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.

Figure 12: 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.

Figure 13: 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 (retrieving the correct pair of individual insects (before juxtaposition) from the DB).
Acknowledgments

This work was partially funded by the European Union (FP7 – ICT, Grant Agreement #258749; CEEDs project). List of external databases used in our research: (i) Brown University’s Shape DB, (ii) MPEG-7 shape DB, (iii) ImageCLEF DB.

6. REFERENCES