Skeleton-based recognition of shapes in images via longest path matching

Gulce Bal[⋆], Julia Diebold[⋆⋆], Erin Wolf Chambers[⋆⋆⋆], Ellen Gasparovic[†], Ruizhen Hu[‡], Kathryn Leonard[§], Matineh Shaker[¶], and Carola Wenk[‖]

Abstract We present a novel image recognition method based on the Blum medial axis that identifies shape information present in unsegmented input images. Inspired by prior work matching from a library using only the longest path in the medial axis [3], we extract medial axes from shapes with clean contours and seek to recognize these shapes within "noisy" images. Recognition consists of matching longest paths from the segmented images into complicated geometric graphs, which are computed via edge detection on the (unsegmented) input images to obtain Voronoi diagrams associated to the edges. We present two approaches: one based on map-matching techniques using the weak Fréchet distance, and one based on a multiscale curve metric after reducing the Voronoi graphs to their minimum spanning trees. This paper serves as a proof of concept for this approach, using images from three shape databases with known segmentability (whale flukes, strawberries, and dancers). Our preliminary results on these images show promise, with both approaches correctly identifying two out of three shapes.

[⋆] Dept. of Computer Engineering, Middle East Technical University, gulcebal@gmail.com.

[⋆⋆] Dept. of Computer Science, Technical University of Munich, julia.diebold@in.tum.de.

[⋆⋆⋆] Dept. of Mathematics and Computer Science, Saint Louis University, echambe5@slu.edu. Research supported in part by NSF grants CCF-1054779 and IIS-1319573.

[†] Dept. of Mathematics, Duke University, ellen@math.duke.edu.

[‡] Dept. of Mathematics, Zhejiang University, ruizhen.hu@gmail.com.

[§] Dept. of Mathematics, California State University Channel Islands, kleonard.ci@gmail.com. Research supported in part by NSF grant IIS-0954256.

[¶] Dept. of Electrical Engineering, Northeastern University, shaker@ece.neu.edu.

[‖] Dept. of Computer Science, Tulane University, cwenk@tulane.edu. Research supported in part by NSF grant CCF-0643597.

1 Introduction

We present a method and proof-of-concept for image recognition based on information extracted from the Blum medial axis. Shape recognition and matching based solely on contour points have been shown to perform weakly in the presence of occlusion, partial data, and noise [22, 13, 4]. Unorganized point sets [5] representing boundaries of shapes are often matched using assignment algorithms for graph matching [10]. Another class of methods which use Hausdorff distance to match the edge maps [13] has the advantage of not requiring correspondences of edge features, but they do not necessarily preserve the integrity of shape parts. Global shape representations which are translation, rotation, or scale invariant such as coefficients of Fourier descriptors [18] may result in incorrect matchings due to noise or occlusion. Historically, approaches based on the medial axis have suffered from its instability and complexity in the presence of noise and pixelation. Our approach is designed to bypass those problems while preserving the strengths of the medial axis as a shape descriptor, including meaningful decomposition into parts and stability despite occlusion. Furthermore, our matching techniques are designed to be near-invariant to Euclidean motions (translation, rotation, and scaling).

While shape recognition based on the medial axis has been well-studied for presegmented shapes [26], this project is among the first to perform recognition using the medial axis on an unsegmented unknown image. The basic concept builds on previous work which recognizes objects by matching longest paths in the medial axis, but only in the limited setting where the input is a "nice" shape taken from a particular hand-drawn catalog [3]. Here, we apply a similar philosophy to match shapes in the much more challenging domain where the input is an arbitrary image. As a result, we must apply edge detection and other techniques in order to identify significant shape information present in the image. Additionally, whereas [3] uses both the medial skeleton and the radius function, our current results use only the skeleton because extracting reliable radius information from arbitrary edges in an image presents additional challenges.

Since there is no common frame of reference between shapes from our canonical library of possibilities and our input image, we must match an arbitrary path (the longest path from the canonical image) into a messy geometric graph (the Voronoi diagram of the edges detected from our image). We use two different approaches in this work, one based on map-matching using the weak Fréchet distance and the other based on a multiscale curve matching into the minimum spanning tree of the graph computed from the input image edges.

Our initial results indicate that both matching methods perform reasonably well, clearly matching two of our three initial tests to the correct image. The algorithms are reasonably efficient, although the map-matching approach is more computationally intensive due to the exhaustive set of rotations and transformations that must be tested. Testing on a larger database than our three-object set is required to determine the full power of these methods.

2 Background

2.1 The medial axis

The *medial axis* of an object is the set of points which have more than one closest point on the object's boundary. It was first introduced by Blum as a tool for recognizing shapes in biological images [6]. It is known that the medial axis has the same homotopy type as the original shape [17], and therefore it gives a topologically accurate but simpler representation of the shape of an object. In addition, the geometry of the boundary curve is encoded in the geometry of the medial skeleton and its radius function. The *medial axis transform* is the set of points in the medial axis annotated with the radius of the largest inscribed ball centered at each point. This structure can be used to recover the entirety of the original shape. Applications and algorithms using this structure are numerous; see for example the survey by Leymarie and Kimia and the many other references in [16].

2.2 Shape recognition using the medial axis

One of the main motivations for this work is the fact that medial-axis based structures such as the *shock graph* have had notable success with the problem of image recognition among a large database [26, 23, 24]. Each of these algorithms catalogs a set of canonical shape categories by computing the shock graph (an annotated version of the medial axis) for each of the shape instances. The next step is to read input images and attempt to match the shock graphs of the input images against the library of known shapes. These algorithms are based on dynamic programming, and work efficiently since the shock graph is a tree whenever the input shape is simply connected.

Another line of research motivating our work does not use the entire structure of the medial axis, but instead does the matching strictly based on the longest path in the medial axis and its associated radius function. Bai et. al [3] implemented and tested on a library of shapes containing 56 images total, with 4 objects per shape class [2]. Their approach of removing a shape from the library and testing to get the correct classification resulted in a success rate of 98.2%. In addition, they implemented and tested their method on a larger dataset [23] with 94.4% accuracy.

Although this matching is naturally less successful for images with high radial symmetry, they nonetheless successfully match input shapes to the correct class for the vast majority of tested images. This is perhaps surprising, given how much rich information about the medial axis is lost when only considering the single longest path. However, the work has so far been applied only to catalogs of images with hand-drawn, clean contours. In this paper, we apply a related method to recognize a shape contained in an arbitrary (noisy) input image.

2.3 Map-matching

Given a graph G embedded in Euclidean space $\mathbb{R}^d$ (most often $\mathbb{R}^2$) and a polygonal curve γ also embedded in $\mathbb{R}^d$, the map-matching problem asks for the path in G which is closest to γ, generally under some distance measure such as the Fréchet distance or weak Fréchet distance. Recently, this problem has been considered in both theoretical and applied settings due to its utility in GIS applications [1, 7, 8]. In this setting, one often has a trajectory (such as is given by a GPS unit placed in a vehicle) which needs to be matched to the closest path on a known road network, modeled as the graph G.

Our setting is slightly different: although the graphs we work with are extracted from images and thus have embeddings in $\mathbb{R}^2$, our input paths are not embedded in the same frame of reference since the scales and orientations of the arbitrary input images can be quite different from the reference images from the library. This variation is somewhat similar to the notion of a graph isomorphism, but here, our input graphs are geometric graphs rather than arbitrary ones. While fast algorithms for Fréchet distance to a geometric graph have been looked at in some limited settings, such as for trees [12], no one previously has considered the problem where the input path is not given as an embedding into the same frame of reference as the graph G, which adds considerably to the difficulty of the problem.

We perform map-matching via the weak Fréchet distance. Let $\gamma_1, \gamma_2 : [0,1] \to \mathbb{R}^2$ be two curves in the plane. The weak Fréchet distance δ_{wF} between them is defined as:

$$\delta_{wF}(\gamma_1, \gamma_2) = \inf_{\alpha_1, \alpha_2 : [0,1] \to [0,1]} \max_{t \in [0,1]} \|\gamma_1(\alpha_1(t)) - \gamma_2(\alpha_2(t))\|,$$

where α_1 and α_2 range over all continuous reparametrizations with $\alpha_1(0) = \alpha_2(0) = 0$ and $\alpha_1(1) = \alpha_2(1) = 1$, and $||.||$ denotes the Euclidean norm. The weak Fréchet distance is a well-suited distance measure for comparing curves as it takes into account the continuity of the curves. In our setting, we consider the set $\mathscr{T}$ of translations, rotations, and scalings. And the related map-matching problem that we address is to find for a geometric graph G, a curve γ, and any admissible transformation $T \in \mathscr{T}$, the path in G that minimizes the weak Frechet distance to any $T(\gamma)$.

2.4 $H^{1/2}$-type multiscale curve metric

Our other method of matching an input path into a geometric graph is via the $H^{1/2}$ multiscale curve metric, first introduced in [20], evaluated on a curve extracted from the graph and the known longest medial path. The last decade has produced a substantial body of work on finding shape metrics that respect the underlying geometry of shape space, where a shape is modeled as a curve in $\mathbb{R}^2$ possibly modulo a group of transformations [25, 19, 14]. Unfortunately, these metrics are computationally expensive and can be unwieldy to implement in any realistic setting. The $H^{1/2}$-type metric is a middle-ground: a weakened linearization of a Riemannian metric that

is computationally fast. In other words, it computes distances based on geometric quantities whereas the Fréchet distance does not.

For ease of exposition, results here are given for plane curves as objects in $\mathbb{C}$ instead of $\mathbb{R}^2$. We trust the reader can move naturally between these two representations. Given a smooth arclength-parameterized open plane curve $\gamma(s)$, define an $H^{1/2}$ "norm"[9] as:

$$\|\gamma\|_{\frac{1}{2}}^2 = \int_0^L \int_0^{\min(s,L-s)} \beta(s,t)^2 \, dt \, ds,$$

where L is the length of the curve, and the angle $\beta(s,t)$ between the rays joining $\gamma(s)$ to $\gamma(s+t)$ and $\gamma(s-t)$ is given by:

$$\beta(s,t) \equiv \arg \frac{\gamma(s+t)-\gamma(s)}{\gamma(s)-\gamma(s-t)}.$$

Moreover, β gives rise to a metric on curves. Let Σ be the set of homeomorphisms $\sigma : [0,1] \to [0,1]$ and γ_1, γ_2 be Lipschitz curves. Then:

$$L(\gamma_1,\gamma_2) = \inf_{\sigma\in\Sigma} \int\int (\beta_1(s,t) - \beta_2(\sigma(s),t))^2 \, ds \, dt$$

gives the metric:

$$d^2(\gamma_1,\gamma_2) = L(\gamma_1,\gamma_2) + L(\gamma_2,\gamma_1).$$

For a discretized curve sampled over dyadic intervals, we have:

$$\|\gamma\|_{\frac{1}{2}}^2 = \sum_{n=1+2^{k-1}}^{N-2^{k-1}} \sum_{k=1}^{K} \beta(n,k)^2 2^{-k},$$

where N is the number of sampled points, K determines the maximum number of dyadic intervals, and the angle β is:

$$\beta(n,k) = \arg \frac{\gamma(n+2^{k-1})-\gamma(n)}{\gamma(n)-\gamma(n-2^{k-1})}.$$

If γ is an arclength parameterization of a Lipschitz graph, then the angles $\beta(n,k)$ are, in a distributional sense, the same as the wavelet coefficients of γ over the same dyadic interval system. In this way, the collection of angles $\{\beta(n,k)\}$ provides a multiscale analysis of the curve γ and, in turn, the Haar coefficients of γ' provide a fast computation for $\{\beta(n,k)\}$ based on scaled second differences:

$$\sum_{n=1+2^{k-1}}^{N-2^{k-1}} \sum_{k=1}^{K} \beta(n,k)^2 2^{-k} = \sum_{n=1+2^{k-1}}^{N-2^{k-1}} \sum_{k=1}^{K} \left(\gamma(n+2^k) - 2\gamma(n) + \gamma(n-2^k)\right) 2^{-k}. \quad (1)$$

[9] We are not viewing the space of plane curves as linear, but the integral defined is analogous to Sobolev norms on function spaces and the integrand is analogous to a wavelet decomposition of γ. Additionally, the "norm" gives rise to a metric on curves in the standard way.

If γ_1 and γ_2 are sampled by $M \leq N$ points, respectively, then $\sigma : \{1,\ldots,M\} \to \{1,\ldots,N\}$ and scales are limited by $K \leq \log_2 M$ and we obtain the discrete approximation to the continuous metric:

$$L(\gamma_1,\gamma_2) \approx \min_{\sigma\in\Sigma_{M,N}} \sum_{m=1+2^{k-1}}^{M-2^{k-1}} \sum_{k=1}^{K} \frac{1}{k^2} |\beta_1(m,k) - \beta_2(\sigma(m),k)|^2$$

which in turn can be computed using second differences as above.

The metric as defined is naturally translation invariant. In the discrete case, rotation invariance is introduced by rotating the line joining $\gamma(n+2^k)$ and $\gamma(n-2^k)$ to be horizontal (a coarse approximation to the tangent line at $\gamma(n)$) and scale invariance is introduced by normalizing the average inter-point distances to be one. See [20] for details and full generality of results.

3 Method

3.1 Extracting medial axes from "known" images

In general, the medial axis of an object in a natural image is difficult to extract automatically, as it requires segmenting the image, extracting the points on the boundary of the object of interest, then computing the medial axis. We select three image databases with known segmentability: whale flukes, strawberries, and dancers. We use k-means clustering to extract an initial binary representation of the object of interest, then apply morphological techniques to obtain a clean boundary. We extract the centers and radii of the circumcircles of the Delaunay triangulation of the boundary points and retain only those centers and radii corresponding to the interior of the object, thereby obtaining the interior medial axis. See Figure 1 for an illustration of this process. For more details on this process, see [15].

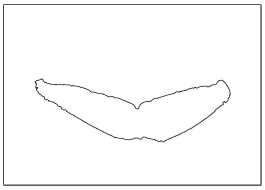
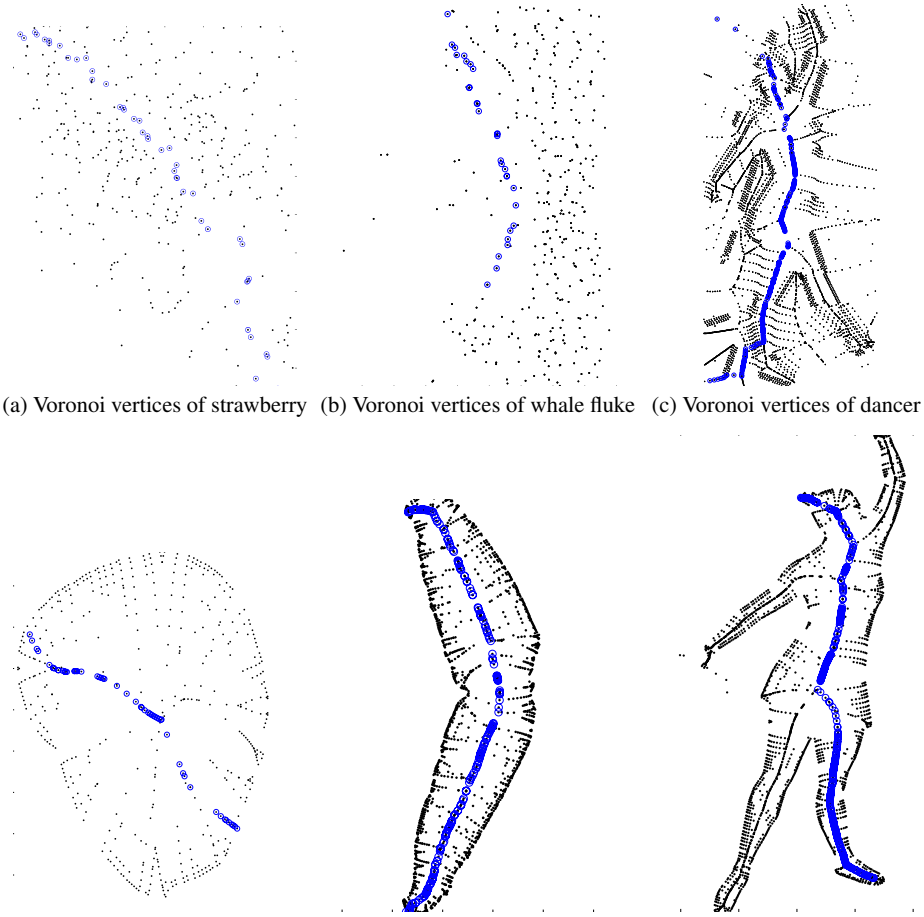
(a)  (b)  (c) 

Fig. 1: Intermediate steps for extracting the medial axis from the whale image. The original image can be seen in Figure 6. Images above are (a) the initial cluster containing the whale fluke resulting from k-means clustering, (b) the segmented whale fluke after morphological processing, and (c) the resulting boundary points. The medial axis with longest path resulting from the boundary displayed here can be seen in Figure 2(e).

To extract the longest path within the axis, we apply Dijkstra's algorithm to find the point P on the axis that is farthest from a randomly selected medial point, then repeat Dijkstra's algorithm to find the medial point Q farthest from P. Retracing steps from Q to P generates the sequence of medial points along the longest path in the medial axis. See Figure 2 for an illustration of this process on our 3 test images.



(a) Voronoi vertices of strawberry (b) Voronoi vertices of whale fluke (c) Voronoi vertices of dancer

(d) Medial skeletons of strawberry (e) Medial skeletons of whale fluke (f) Medial skeletons of dancer

Fig. 2: Top row: Voronoi vertices with longest path highlighted. Bottom row: Medial skeletons with longest path highlighted. These longest paths are shown matched to one another in Figures 12 - 14.

3.2 Extracting Voronoi edges from "unknown" input images

Giraffe Smoothed Giraffe Torbreck Smoothed Torbreck

Fig. 3: Examples of smoothed images.

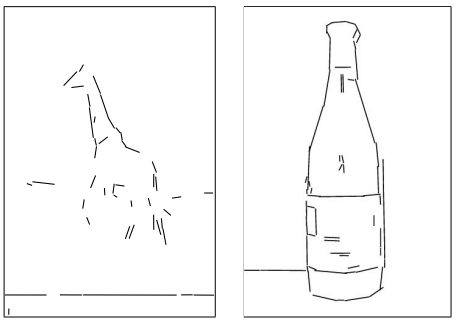


Fig. 4: Results of edge detection on smoothed images of a giraffe and a bottle.

Given an input image, we smooth it as in [21]. See Figure 3. Let f denote the noisy input image and u the denoised (smooth) version. We obtain u by minimizing the energy:

$$E(u) = \int_{\Omega} (f-u)\, dx + \lambda \int_{\Omega} |\nabla u|\, dx, \tag{2}$$

where Ω denotes the image domain and $\lambda \in \mathbb{R}_{>0}$ a weighting factor. The first term ensures that u is similar to f and the second term forces u to be smooth everywhere except at strong edges.

Next, we run a line segment detector (LSD) algorithm [11] on the smoothed version in order to extract prominent edges and thus a likely boundary of a shape. LSD locally detects straight contours on the image, giving subpixel results while controlling the number of false detections per pixel. Contours are naturally defined by the image gradient and level lines of the image which divide the transition region

from dark to light or the opposite. The algorithm works by finding the unit vectors tangent to the level lines, thus computing the level line angle at each pixel. The resulting vector field is then segmented into connected regions that share the same level line angle up to a threshold. Each connected region is represented by a geometrical object such as a rectangle. The principal axis of this object defines the main direction which is chosen as the line segment.

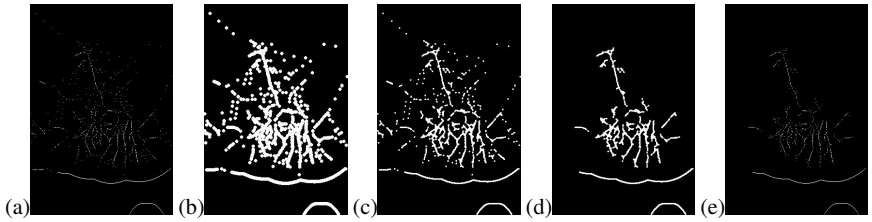


Fig. 5: Removing "outlier" medial points for giraffe image. (a) Original image with all medial points, followed by the resulting images after (b) dilation, (c) erosion, and (d) point deletion. (e) Final image with remaining medial points.

The output is a set of edges with noise, as in Figure 4, which we process into a Voronoi diagram to extract potential medial points. In doing so, we remove "outlier" medial points (including points in the region external to the shape) by a dilation and erosion process, as depicted in Figure 5. That is, we first thicken the medial points to form many connected point clusters and subsequently erode them (while still maintaining connected structures). We then identify and delete all point clusters in the processed image of a sufficiently small area, and/or those points that are greater than a certain small distance away from the largest connected structures in the image. We then compare the resulting image with the original input image and delete all medial points in the input image corresponding to deleted points in the processed image, yielding the desired image without outliers.

Next, our objective is to match the single longest path from each of our initial image instances into the graph which, in each case, approximates the medial axis of the shape that is present in each input image. We pursue this problem in two different ways as outlined in the following two subsections.

3.3 Matching via weak Fréchet distance

Our first method of matching is based on map-matching via the weak Fréchet distance. The related map-matching problem is to find for a geometric graph G and a curve γ a path in the graph that minimizes the weak Fréchet distance to γ. For a polygonal curve γ with n vertices and a graph G with a total number of m edges and vertices, the map-matching problem can be solved in $O(mn\log(mn))$ time [27]. This algorithm constructs a "free space graph" which is essentially a combinatorial rep-

resentation of the product space of (parameterizations of) the curve and the graph. Each vertex-edge pair is assigned a weight that equals their Euclidean distance, and then a shortest path algorithm in this "free space graph" (where the length of the path is computed as the maximum of the weights) computes a path with minimum weak Fréchet distance. Please see [27] for more details.

In our setting, we consider the set $\mathscr{T}$ of translations, rotations, and scalings. And the related map-matching problem that we address is to find for a geometric graph G, a curve γ, and any admissible transformation $T \in \mathscr{T}$, the path in G that minimizes the weak Frechet distance to any $T(\gamma)$. We sample $\mathscr{T}$ by applying a fairly exhaustive set of scalings, translations, and rotations to the curve, and for each such transformation we run the map-matching algorithm of Wenk et al. [27]. In particular, we sample the transformation space as follows: We consider rotations by 0, 90, 180, and 270 degrees. We hold the aspect ratio constant and apply a single scaling factor; the maximum scaling factor is determined such that the width of the (possibly rotated) path equals the width of the graph, and the minimum scaling factor is chosen to be half the maximum factor; this range is sampled in steps of 0.2. The two-dimensional translation space is determined to consist of all translations such that the bounding box of the (possibly scaled and rotated) path fits entirely inside the bounding box of the graph; the translation space is sampled in steps of 10 pixels. As described in Section 4, the dimensions of each bounding box are several hundred pixels by several hundred pixels. The resulting range of scales where between 1 and 2.4 for the strawberry, between 0.38 and 0.78 for the whale fluke, and between 0.6 and 1.2 for the dancer. We note that this method is computationally intensive for each example, involving multiple tests for different possible orientations and sizes of the path.

3.4 Matching via an $H^{1/2}$-type metric

Our second method of matching addresses the fact that the Fréchet based algorithm described in Section 3.3 is especially difficult because the input medial axis graph can be quite noisy and messy depending on how well our edge detection and smoothing algorithms are able to isolate prominent shapes. Additionally, the second method applies a metric that is invariant under Euclidean motion.

We first simplify the Voronoi graph to a tree to avoid cycles when computing the longest path in the graph. We choose the minimum spanning tree because it appears to capture the prominent shape features quite well, though other ways of simplifying the input graph may be worth investigating. Note that in converting the graph to a tree we may lose segments on the longest path. Suppose γ_1 is a discrete representation of the longest path in the medial axis of a known object, and we are given the Voronoi edges from an unknown image. Our method is as follows, with curve matching running in $O(MN\log M)$ time:

1. Compute the minimal spanning tree for the Voronoi edges.
2. Extract γ_2, the longest path in the Voronoi tree.

3. Resample γ_1 and γ_2 to have $N = M = 128$ equally spaced points.
4. Normalize scale so that each curve has an average inter-point distance of one.
5. Compute second differences as described in Equation 1.
6. Extract second differences corresponding to every fourth point on γ_1 to allow for flexible point matching (otherwise the points are matched one-to-one in order), following the procedure outlined in Section 2.4.
7. Apply dynamic programming to find the matching of points of γ_1 to γ_2 that minimizes the distance $d^2(\gamma_1, \gamma_2)$ between the curves.
8. Sum scaled second differences corresponding to the optimal matching to obtain approximation to $d^2(\gamma_1, \gamma_2)$.

4 Results

The three images we use are of a strawberry, a whale fluke, and a dancer. Here we match the medial axis extracted as described in Section 3.1 to the Voronoi diagram of the same image extracted as described in Section 3.2. The dimensions of the bounding boxes of the Voronoi diagrams are 479×367 for the strawberry, 618×418 for the whale fluke, and 540×239 for the dancer. Results from the two matching methods are comparable, and both seem promising.

Fig. 6: Input images: a strawberry, a whale fluke, and a dancer.

4.1 Weak Fréchet map-matching distance results

For the dancer and the whale fluke, the transformation that minimized the weak Fréchet distance over all sampled transformations was found correctly, see Figures 8 and 9. The point matching computed by the weak Fréchet distance also appears to be of good quality. The distance for the minimum transformation computed for the dancer is so small (2.5 pixels), that the transformed dancer path and the resulting

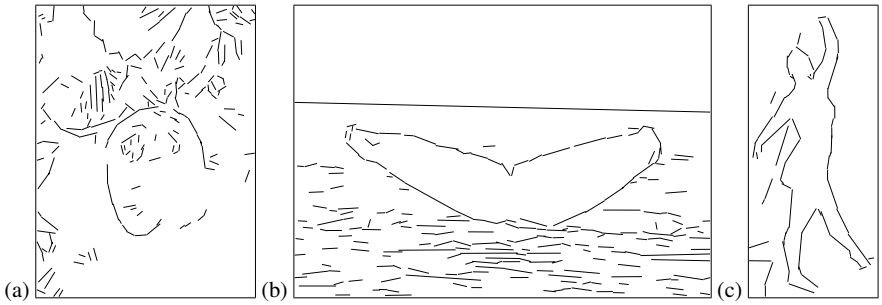


Fig. 7: Results of edge detection: (a) strawberry, (b) whale fluke, and (c) dancer.

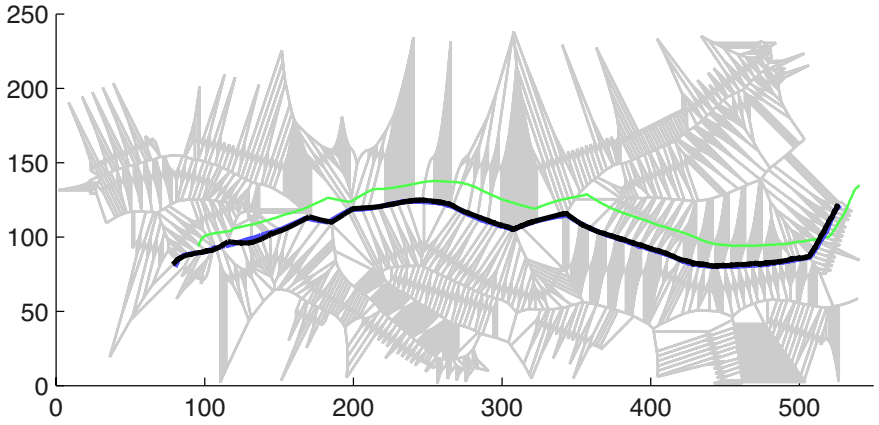


Fig. 8: Matching the dancer path into the dancer graph. The graph edges are shown in light gray, and the path is shown in green. The algorithm finds the correct transformation with minimum Fréchet distance 2.5 pixels (at scale 1.0 with no rotation). The transformed path is shown in black, and the corresponding path in the graph in blue.

matched path in the graph almost coincide. For the whale fluke, the minimum computed transformation (11.2 pixels) is very close to the transformation with the third smallest distance of 11.6 pixels which applies an additional 180 degrees transformation to the whale path. For the strawberry, the path is found for multiple small scales at multiple positions at small distances (ranging between 8.1 to about 13) in the graph, see Figure 8. Out of the $4,368$ sampled transformations per sampled scale, 9.2% of the transformations at scale 1.0 have a distance less than 15. At scale 1.2, this reduces to 5.5%, and at scale 2.0 this reduces to only 0.07%. We believe that this is an artifact caused by the almost grid-like dense edge pattern in the strawberry graph in combination with the very straight shape of the strawberry path.

We also compared the dancer path to the strawberry graph, the whale fluke graph, and the dancer graph. We computed the minimum weak Fréchet distance over all sampled transformations. The computed minimum distances were 8.9 pixels for the

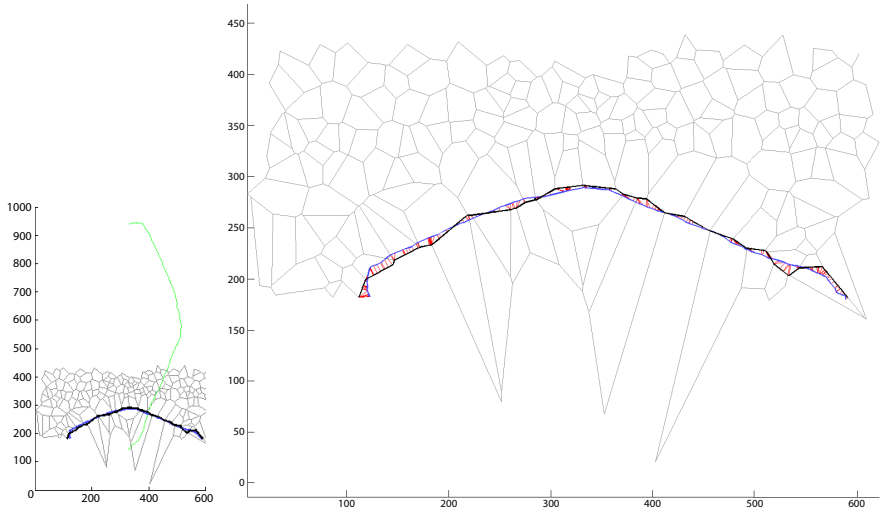


Fig. 9: Matching the whale path into the whale graph. The graph edges are shown in light gray, and the path is shown in green. The algorithm finds the correct transformation with minimum Fréchet distance 11.2 pixels (at scale 0.6 with 90 degrees rotation). The transformed path is shown in black, and the corresponding path in the graph is blue. The red lines show the optimal point matching.

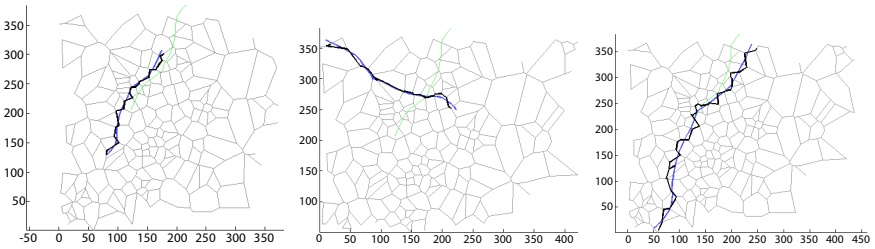


Fig. 10: Matching the strawberry path into the strawberry graph. The algorithm finds too many occurrences of the path at a small scale. The graph edges are shown in light gray, and the path is shown in green. Results are shown for the minimum distance (8.1) at scale 1.0 (with a rotation of 180 degrees), the minimum distance (8.9) at scale 1.2 (with a rotation of 270 degrees), and the minimum distance (13.4) at scale 2.0 (with a rotation of 180 degrees). The transformed path is shown in black, and the corresponding path in the graph in blue.

strawberry graph, 9.7 pixels for the whale fluke graph, and 2.5 for the dancer graph. The dancer path therefore correctly determined the dancer graph as the graph it matches best with, see Figures 8 and 11.

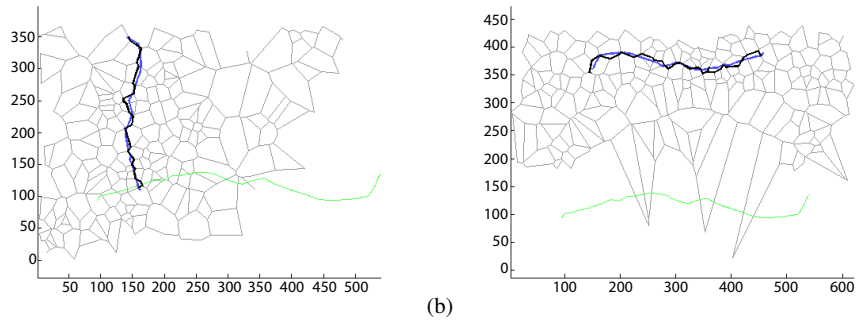


Fig. 11: Matching the dancer path into the strawberry graph (distance 8.9) and into the whale fluke graph (distance 8.7). Both distances are larger than the distance into the dancer graph (2.5), see Figure 8.

4.2 $H^{1/2}$ metric results

Initial results for matching the medial longest path to the Voronoi tree longest path are correct for the two instances where the longest Voronoi path contains the desired medial points. Apart from the strawberry image, where the Voronoi tree longest path fails to contain edges belonging to the medial axis of the strawberry, the closest match corresponds to the correct classification. In addition, the optimal matching between points performs reasonably well. See Figures 12 - 14. Note that the scale of the curves has changed. This is because of the scale invariance we introduced by normalizing inter-point distances to be one.

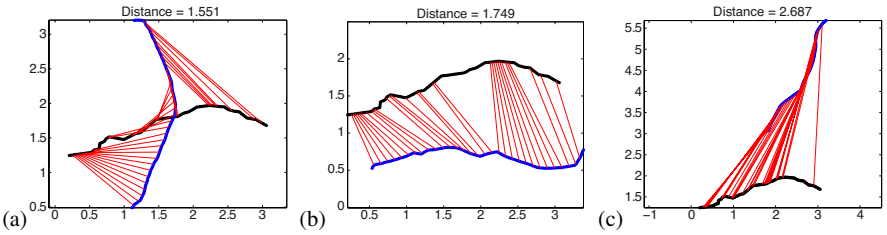


Fig. 12: Matching medial longest path from (a) whale (distance = 1.551), (b) dancer (distance = 1.749), (c) berry (distance = 2.687) into the Voronoi tree longest path of the whale fluke. Lines show the optimal point matching. The minimum distance into the messy graph correctly classifies the unknown image as a whale.

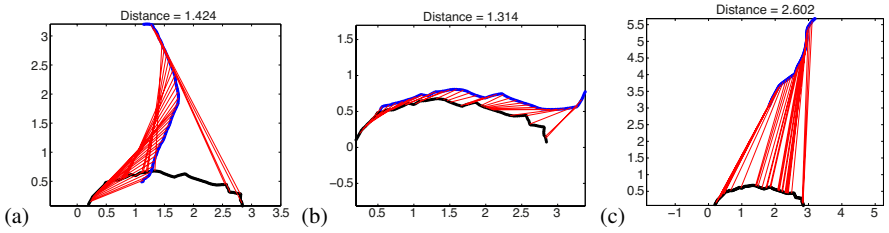


Fig. 13: Matching medial longest path from (a) whale (distance = 1.424), (b) dancer (distance = 1.314), (c) berry (distance = 2.602) into the Voronoi tree longest path of the dancer. Lines show the optimal point matching. The minimum distance into the messy graph correctly classifies the unknown image as a dancer.

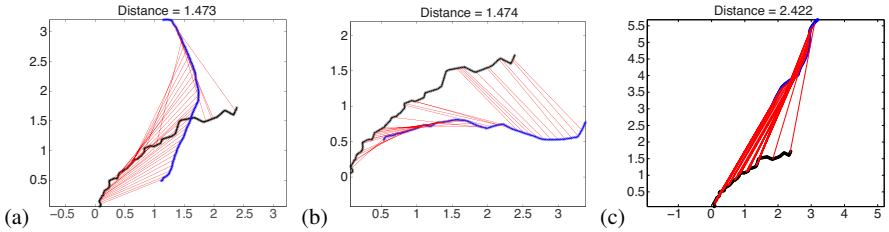


Fig. 14: Matching medial longest path from (a) whale (distance = 1.473), (b) dancer (distance = 1.474), (c) berry (distance = 2.422) into the Voronoi tree longest path of the strawberry. Lines show the optimal point matching. The minimum distance into the messy graph incorrectly classifies the unknown image as a whale. Note that the longest path in the Voronoi tree for the strawberry image does not contain any edges from the medial axis of the strawberry itself.

5 Discussion and Future Directions

Our matching techniques show enough promise to merit additional investigation. We are curious about the success of the algorithms when the Voronoi diagram is dense or grid-like, or where the longest path in the medial axis does not trace a prominent shape feature (such as when the input image is nearly round with radial symmetry).

5.1 Analysis of the weak Fréchet map-matching distance

Sampling the transformation space to minimize the weak Fréchet map-matching works well for the dancer and the whale fluke. The strawberry graph exhibits a grid-like dense edge pattern which causes the strawberry path to be found in many locations in the graph, in particular for small scales. While this behavior is extreme in the strawberry, it is also present in the whale fluke data, where the path with the second smallest distance is located at a different location with an additional 180

degrees rotation. We believe that the "small scale" problem could be overcome by analyzing the distribution of distances for fixed scale and varying translations and rotations, in order to identify transformations with significant distances. We will investigate this direction in future research.

For the dancer path, the distance into the dancer graph was much smaller than into the strawberry graph and the whale fluke graph. The minimum weak Fréchet distance into the messy graph therefore correctly classifies the unknown image as a dancer.

5.2 Analysis of the $H^{1/2}$-type metric

Matching longest paths using the $H^{1/2}$-type metric performs well for the two cases, whale and dancer, where the longest path in the Voronoi tree contains edges corresponding to the medial axis of the object of interest. Not surprisingly, it fails for the third image, the strawberry, where no medial edges appear in the Voronoi tree longest path. The strawberry image is particularly challenging, as the berry itself contributes very few edges to the very complicated edge map seen in Figure 7(a) and contains several spurious edges in its interior. This illustrates the need for an additional evaluation of relative importance of Voronoi vertices, perhaps through classification of vertices as belonging to the foreground or background or noise.

In addition, the optimal matching of points between the two longest paths currently seems to favor matches that map the medial path into the length of the Voronoi path. For example, in Figure 12 the medial axis for the whale in the Voronoi path starts at about the halfway point whereas the optimal matching begins at the left of the path. Because the optimal matching can skip enough points to avoid highly mismatching segments, it seems likely that a longer match will often be lower cost for medial curve matching. At the same time, two curves that are identical up to a point can correspond in a match that is too short. Figure 15 illustrates this issue. For the larger scales, the differences in the β angles (and their second difference approximations) will grow as the points on the circles approach the points on the line. Hence the lowest cost match avoids points toward the end of the circle attached to the line. Penalizing skips that are longer than an average skip, or adding the difference in the radius function values to the cost of matching two points may improve the medial point correspondence.

5.3 Future work

Our initial proof of concept for this approach is promising. Based on our results and prior work in this area [3], we speculate that this approach will also work well to capture the same shape in a different pose (such as a dancer in different positions). Future work will consider a larger library of shapes as well as input images in differ-

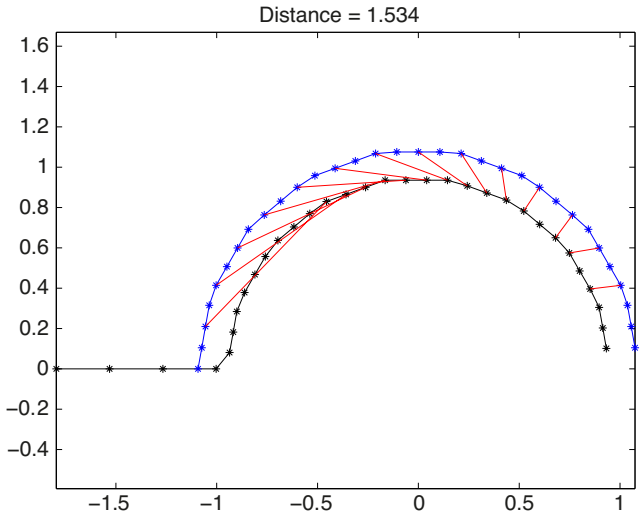


Fig. 15: Matching a semicircle into a curve composed of the union of a semicircle and a line. Instead of matching semicircle to semicircle, the semicircle is matched to points away from the line.

ent poses. There is also potential to include the radius function as well as the longest path for improved recognition results, but it is not clear if the Voronoi graphs from input images will prove too noisy to reliably calculate this information.

Both approaches perform better with a simpler Voronoi graph. We are currently exploring methods for evaluating the saliency of either a particular Voronoi vertex or (equivalently) an edge pair associated to a Voronoi vertex. In addition, both approaches would benefit from using the information in the radius function on the medial and Voronoi points that gives the distance to the corresponding edge points. We anticipate substantial improvement from the combination of these modifications.

We also hope to reduce the cost of learning additional shape classes once a sufficient number of classes have been learned. Learning the visual models for classification of test objects requires a significant number of training samples. In the method of one-shot learning [9], the information from previously learned categories is used for training new categories, using a Bayesian prior and maximum a posteriori (MAP) estimation. This model could be used to optimize and extend learning for the current methods.

Acknowledgments

The authors would like to thank the Institute for Pure and Applied Mathematics, the Association for Women in Mathematics, Microsoft Research, the National Science Foundation, and the National Geospatial Agency for support, financial and other-

wise, of this collaboration. Kathryn Leonard thanks Matt Feiszli for providing the initial Matlab code for the $H^{1/2}$ metric for closed curves which was modified for this project.

References

1. Helmut Alt, Alon Efrat, Günter Rote, and Carola Wenk. Matching planar maps. *J. Algorithms*, 49(2):262–283, November 2003.
2. Cagri Aslan and Sibel Tari. An axis-based representation for recognition. In *ICCV*, pages 1339–1346. IEEE Computer Society, 2005.
3. X. Bai, X. Yang, D. Yu, and L. J. Latecki. Skeleton-based shape classification using path similarity. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 22(4):733–746, 2008.
4. Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.
5. Serge Belongie, Greg Mori, and Jitendra Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006.
6. H. Blum. A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form*, pages 362–80, 1967.
7. Sotiris Brakatsoulas, Dieter Pfoser, Randall Salas, and Carola Wenk. On map-matching vehicle tracking data. In *Proceedings of the 31st International Conference on Very Large Data Bases*, VLDB '05, pages 853–864. VLDB Endowment, 2005.
8. Daniel Chen, Anne Driemel, Leonidas J. Guibas, Andy Nguyen, and Carola Wenk. Approximate map matching with respect to the Fréchet distance. In Matthias Müller-Hannemann and Renato Fonseca F. Werneck, editors, *ALENEX*, pages 75–83. SIAM, 2011.
9. Li Fei-Fei, Robert Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.
10. Steven Gold and Anand Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):377–388, 1996.
11. Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: a Line Segment Detector. *Image Processing On Line*, 2012, 2012.
12. Joachim Gudmundsson and Michiel Smid. Fréchet queries in geometric trees. In Hans L. Bodlaender and Giuseppe F. Italiano, editors, *Algorithms - ESA 2013*, volume 8125 of *Lecture Notes in Computer Science*, pages 565–576. Springer Berlin Heidelberg, 2013.
13. Daniel P. Huttenlocher, Gregory A. Klanderman, and William J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.
14. S. Kushnarev. Teichons: Solitonlike geodesics on universal Teichmüller space. *Experimental Mathematics*, (18):325–336, 2009.
15. K. Leonard, R. Strawbridge, D. Lindsay, R. Barata, M. Dawson, and L. Averion. Minimal geometric representation and strawberry stem detection. In *Computational Science and Its Applications (ICCSA), 2013 13th International Conference on*, pages 144–149, June 2013.
16. Frederic F. Leymarie and Benjamin B. Kimia. From the infinitely large to the infinitely small: Applications of medial symmetry representations of shape. In Kaleem Siddiqi and Stephen Pizer, editors, *Medial Representations: Mathematics, Algorithms and Applications*, pages 327–351. Kluwer Academic Publishers, 2006.
17. André Lieutier. Any open bounded subset of $\mathbb{R}^n$ has the same homotopy type as its medial axis. *Comput.-Aided Des.*, 36(11):1029–1046, September 2004.
18. C.C. Lin and Rama Chellappa. Classification of partial 2-D shapes using Fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):686–690, 1987.

19. A. Trouve M. I. Miller and L. Younes. On metrics and Euler-Lagrange equations of computational anatomy. *Ann. Rev. Biomed. Engng*, (4):375–405, 2002.
20. Kathryn Leonard Matt Feiszli, Sergey Kushnarev. Metric spaces of shapes and applications: Compression, curve matching and low-dimensional representation. *Geometry, Imaging, and Computation*, to appear.
21. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
22. Thomas B. Sebastian and Benjamin B. Kimia. Curves vs. skeletons in object recognition. *Signal Processing*, 85(2):247–263, 2005.
23. Thomas B. Sebastian, Philip N. Klein, and Benjamin B. Kimia. Shock-based indexing into large shape databases. In *Proceedings of the 7th European Conference on Computer Vision-Part III*, ECCV '02, pages 731–746, London, UK, 2002. Springer-Verlag.
24. Thomas B. Sebastian, Philip N. Klein, and Benjamin B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):550–571, May 2004.
25. E. Sharon and D. Mumford. 2d-shape analysis using conformal mapping. *International Journal of Computer Vision*, (70):55 75, 2006.
26. Nhon H. Trinh and Benjamin B. Kimia. Skeleton search: Category-specific object recognition and segmentation using a skeletal shape model. *International Journal of Computer Vision*, 94(2):215–240, September 2011.
27. Carola Wenk, Randall Salas, and Dieter Pfoser. Addressing the need for map-matching speed: Localizing global curve-matching algorithms. In *Proceedings of the 18th International Conference on Scientific and Statistical Database Management*, SSDBM '06, pages 379–388, Washington, DC, USA, 2006. IEEE Computer Society.