Approximations of Geodesic Distances for Incomplete Triangular Manifolds

Zouhour Ben Azouz[1] Prosenjit Bose[2] Chang Shu[1] Stefanie Wuhrer[1,2]

Abstract

We present a heuristic algorithm to compute approximate geodesic distances on a triangular manifold S containing n vertices with partially missing data. The proposed method computes an approximation of the geodesic distance between two vertices p_i and p_j on S and provides an upper bound of the geodesic distance that is shown to be optimal in the worst case. This yields a relative error bound of the estimate that is worst-case optimal. The algorithm approximates the geodesic distance without trying to reconstruct the missing data by embedding the surface in a low dimensional space via multi-dimensional scaling (MDS). We derive a new heuristic method to add an object to the embedding computed via least-squares MDS.

1 Introduction

The computation of geodesic distances on a given triangular manifold S with n vertices is a well-studied problem in computational geometry and differential geometry. Algorithms computing geodesic distances on polyhedral surfaces can be classified into two approaches. Algorithms following the first approach view S as a graph and algorithms to compute shortest distances on graphs are extended to find geodesic distances on S. Mitchell et al. [9] presented an algorithm that computes the exact geodesic distances from one source point on S to all other points of S in $O(n^2 \log n)$ time. Surazhsky et al. [11] implemented the algorithm and found the algorithm's average running time to be much lower, which makes the algorithm relevant for practical tasks. Algorithms following the second approach view S as a discretized differentiable surface and algorithms from differential geometry are extended to find geodesic paths on S. Kimmel and Sethian [7] presented an approach called Fast Marching Method (FMM) on triangular domains that computes approximations to the geodesic distances from one source point on S to all other points of S by solving the Eikonal equation on a triangular grid. The algorithm's running time is $O(n \log n)$ and therefore optimal. The accuracy of the approach depends on the quality of the underlying triangulation; namely on the longest edge and the widest angle in the triangular mesh.

[1]National Research Council of Canada
[2]Carleton University, Ottawa, Canada

The computation of geodesic distances on a triangular manifold is a common operation in many areas such as computer graphics, computer vision, and pattern recognition [2, 6, 11]. The 3D models used in these application areas usually come from digitizing real-world objects from a discrete set of measurements. For incomplete triangular manifolds, all of the above-mentioned methods compute the geodesic path between two points on opposite sides of a hole by tracing along the boundary of the hole as in Figure 1. This results in erroneous geodesic distances.

We explore the problem of computing estimates of geodesic distances with worst-case optimal upper bounds on a triangular manifold S with partially missing data without attempting to fill the holes of S. To our knowledge, this problem has not been explored so far. The main advantage of this approach compared to previous approaches to compute geodesic distances on triangular manifolds [9, 7, 11] is that the error of the estimate is bounded for incomplete surfaces. The resulting approximated geodesic distances can be used to modify the above-mentioned applications for models with incomplete surface descriptions.

The approximation of the geodesic distance consists of three main steps. First, we compute the geodesic distance $\delta_{i,j}$ between the vertices p_i and p_j for $i, j \in P$, where P is a set of indices of uniformly distributed sample points on S, using FMM. The geodesic path between p_i and p_j computed by FMM may trace a hole of the model and therefore be incorrect, see Figure 1. However, we can compute confidence values $\omega_{i,j} = 1 - \frac{m^h_{i,j}}{m_{i,j}}$, where $m_{i,j}$ is the number of edges on the geodesic path computed by FMM from p_i to p_j and where $m^h_{i,j}$ is the number of edges tracing a hole of S on the geodesic path from p_i to p_j. Second, we use the geodesic distances $\delta_{i,j}$ as dissimilarities and the confidence values $\omega_{i,j}$ as weights to embed the manifold S in a low-dimensional Euclidean space via multi-dimensional scaling (MDS). In this way, we obtain a canonical form of S similar to the one introduced by Elad and Kimmel [2]. Third, we compute an estimate of the true geodesic distance between two arbitrary vertices p_i and p_j on S by projecting p_i and p_j to the canonical form of S using an extension of the technique devised by Gower [5]. The Euclidean distance between the embedded points approximates the true geodesic distance between the original points. A detailed description of the approach is available in the full version of the paper [12].

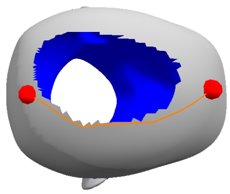

Figure 1: *Paths computed via FMM tracing holes.*

2 Preliminaries

Let S denote an incomplete triangular manifold with n vertices. Let $\hat{S}$ denote the complete surface partially represented by S. Let $\hat{\delta}_{i,j}$ be the geodesic distance between the vertices p_i and p_j on $\hat{S}$ and let $\delta_{i,j}$ be the geodesic distance between the vertices p_i and p_j on S. Note that $\delta_{i,j}$ equals $\hat{\delta}_{i,j}$ if the geodesic path between p_i and p_j on S does not trace a hole of S. Let $\hat{d}_{i,j}$ denote the Euclidean distance between p_i and p_j.

We aim to preprocess S, such that given any pair of query vertices p_i and p_j on S, we can report an estimate of $\hat{\delta}_{i,j}$ along with error bounds of the estimate.

We propose a heuristic solution based on multi-dimensional scaling (MDS). MDS is used to compute a mapping of $p_i, i = 1, \ldots, n$ to a set of points $X = \{X_i, i = 1, \ldots, n\}$ in a low-dimensional Euclidean space, such that the Euclidean distance $d_{i,j}(X)$ between X_i and X_j approximates the geodesic distances between p_i and p_j well for $1 \leq i,j \leq n$. This mapping has previously been called *canonical form* by Elad and Kimmel [2].

2.1 Multi-Dimensional Scaling

MDS is a commonly used technique to reduce the dimensionality of high-dimensional data. Given a set of n objects $O_1, \ldots, O_n$ in d dimensions as well as the pairwise dissimilarities $\delta_{i,j}, 1 \leq i,j \leq n$ with $\delta_{i,j} = \delta_{j,i}$ between objects O_i and O_j, the aim is to find points $X = \{X_1, \ldots, X_n\}$ in k dimensions with $k < d$, such that the Euclidean distance $d_{i,j}(X)$ between X_i and X_j equals $\delta_{i,j}$ for $1 \leq i,j \leq n$. This aim can be shown to be too ambitious, since in general it is not possible to find positions $X_1, \ldots, X_n$ in k dimensions such that $d_{i,j}(X) = \delta_{i,j}$ for all i,j. To find a good approximation, different related optimality measures can be used. Classical MDS [4], also called Principal Coordinate Analysis (PCO), is a method closely related to Principal Component Analysis. It assumes that the dissimilarities are Euclidean distances in a high dimensional space and aims to minimize $E_{PCO} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\delta_{i,j}^2 - d_{i,j}(X)^2)$ by finding a mapping as eigenvectors of a matrix. Least-Squares MDS (LSMDS) [1, p.146-155] aims to minimize $E_{LS} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \omega_{i,j} \left(\delta_{i,j} - d_{i,j}(X)\right)^2$, where $\omega_{i,j}$ is a non-negative weight that can be viewed as a confidence value corresponding to the dissimilarity $\delta_{i,j}$. Therefore, MDS can be viewed as a mapping from arbitrary objects O_i in d dimensions to points X_i in k dimensions with the constraint that an objective function E is minimized.

A question that arises in MDS is how to treat an additional object O_{n+1} in d-dimensional space with corresponding dissimilarities $\delta_{n+1,1}, \ldots, \delta_{n+1,n}$ that becomes available only after the objects $O_1, O_2, \ldots, O_n$ have been mapped to points $X_1, X_2, \ldots, X_n$ in k-dimensional space. Gower [5] proposed an efficient approach to add an object to the PCO embedding. To add an object O_{n+1} to the LSMDS embedding, we are also given the corresponding weights $\omega_{n+1,1}, \ldots, \omega_{n+1,n}$. The technique by Gower does not yield satisfying results, since the objective function minimized for the embedding of the objects $O_1, \ldots, O_n$ is E_{LS}. Instead, we try to minimize the least-squares function $E^*_{LS} = \sum_{i=1}^{n} \omega_{n+1,i} \left(\delta_{n+1,i} - d_{n+1,i}(X)\right)^2$. We can compute the gradient of this objective function w.r.t. the point $\vec{x}_{n+1}$ analytically as $\nabla E^*_{LS} = \sum_{i=1}^{n} 2\omega_{n+1,i}(\vec{x}_{n+1}^T - \vec{x}_i^T)\left(1 - \frac{\delta_{n+1,i}}{d_{n+1,i}}\right)$. This allows us to add the object O_{n+1} to the MDS embedding by minimizing E^*_{LS} using the *limited-memory Broyden-Fletcher-Goldfarb-Shanno* (LSBFGS) scheme, a quasi-Newton approach. Although we are not aware that this addition to the LSMDS embedding was discussed previously, this is not the main contribution of this work.

3 Geodesic Distance Estimation

We use the canonical form to estimate the geodesic distance between a given pair p_i and p_j with $1 \leq i,j \leq n$ of vertices on a triangular manifold S with partially missing data. The main idea is to compute the canonical form of the manifold based on weighted geodesic distances on S. That is, we use geodesic distances as dissimilarities $\delta_{i,j}$ and we use confidence values $\omega_{i,j} = 1 - \frac{m^h_{i,j}}{m_{i,j}}$, where $m_{i,j}$ is the number of edges on the geodesic path from p_i to p_j on S and where $m^h_{i,j}$ is the number of edges on the geodesic path on S from p_i to p_j that pass through triangles of S which share at least one vertex with the boundary of a hole. Since S is a manifold, we can find the vertices of S adjacent to a hole of S as endpoints of edges of degree less than two, since every edge not adjacent to a hole of S has degree two. We chose this measure for $\omega_{i,j}$, since it can be computed more efficiently than the fraction of the length of the path that does not trace the boundary of a hole. When working with data obtained from laser range scanners, $\omega_{i,j}$ is a good approximation of the fraction of the path that does not trace the boundary of a hole, because all of the triangles of S have good aspect ratio. If paths that trace holes of S obtain weight 0, it can be proven that Euclidean distances in embedding space approximate the original geodesics well [10]. Since we wish to extrapolate information using the metric property of the manifold, we give distances tracing a hole of S less weight, but we do not disregard those distances.

The geodesic distances $\hat{\delta}_{i,j}$ form a metric. That is, $\hat{\delta}_{i,j}$ is non-negative, symmetric, and satisfies the triangle inequality. If $\hat{S} = S$, the set of dissimilarities $\hat{\delta}_{i,j}$ contains redundant information. When S is a true subset of $\hat{S}$, we take advantage of this redundancy by weighing well approximated geodesic distances higher than geodesic distances tracing around a hole of S. We use the geodesic distances with confidence values to compute a canonical form of a sample set P of size n_P. This sample set is necessary for objects with hundreds of thousands of vertices, since computing the canonical form is not only computationally expensive but also requires quadratic storage in the number of vertices to embed due to the quadratic number of dissimilarities and weights. Since taking a sample set P of vertices for the embedding has a negative effect on the quality of the results, the sample set should be large enough to represent the overall shape of S well. We use Voronoi sampling [3], also called Farthest Point Sampling (FPS), to choose the sample set P. FPS provides uniformly distributed samples with respect to the geodesic distances in an iterative way. FPS and FMM can elegantly be combined [8] to obtain n_P uniformly distributed sample points, $\delta_{i,j}$, and $\omega_{i,j}$, in $O(n_P n \log n)$ time.

The canonical form has the property that Euclidean distances in the canonical form approximate geodesic distances on S well according to the optimality measure $\sum_{i \in P} \sum_{j \in P} \left(1 - \frac{m^h_{i,j}}{m_{i,j}}\right) (\delta_{i,j} - d_{i,j}(X))^2$. Hence, we expect $d_{i,j}$ to be a good approximation of $\hat{\delta}_{i,j}$ on $\hat{S}$ even if $\delta_{i,j}$ is obtained by a path tracing a hole of S. The error made by approximating $\hat{\delta}_{i,j}$ by $d_{i,j}$ can be bounded as follows. A lower bound $\hat{\delta}^{lower}_{i,j}$ on $\hat{\delta}_{i,j}$ is given by the Euclidean distance $\hat{d}_{i,j}$ between p_i and p_j. Note that $\hat{\delta}^{lower}_{i,j}$ is not necessarily worst-case optimal. An upper bound $\hat{\delta}^{upper}_{i,j}$ on $\hat{\delta}_{i,j}$ is given by $\delta_{i,j}$. The upper bound $\hat{\delta}^{upper}_{i,j}$ is optimal in the worst case, since a path tracing a hole of S can be the shortest path on $\hat{S}$ if $\hat{S}$ has a high mountain where the hole is located on S. Note that $\hat{\delta}^{upper}_{i,j}$ only exists if p_i and p_j are located on a connected component of S. The *relative error* e of the approximation of $\hat{\delta}_{i,j}$ by $d_{i,j}$ is computed as $e = \frac{\max(\left|d_{i,j} - \hat{\delta}^{lower}_{i,j}\right|, \left|d_{i,j} - \hat{\delta}^{upper}_{i,j}\right|)}{d_{i,j}}$.

3.1 Algorithm Overview and Analysis

We now describe the algorithm used to estimate geodesic distances on the incomplete surface S. First, we compute the canonical form of S. A set P of indices of sample points on S is obtained via FPS.

FMM is performed to obtain all of the pairwise geodesic distances $\delta_{i,j}$ on S along with confidence values $\omega_{i,j}$. FMM does not compute exact geodesic distances on S, but approximations. However, the geodesic distances computed via FMM approximate $\delta_{i,j}$ well for surfaces obtained using a laser range scanner in practice. The minor theoretical flaw of using geodesic distances computed via FMM instead of the exact geodesic distances on S can be overcome by either using known exact algorithm to compute geodesic distances on S [9] or by adjusting the error bounds to include the error caused by FMM. The main advantages of FMM are its efficiency and simplicity.

The pairwise geodesic distances $\delta_{i,j}$ on S along with confidence values $\omega_{i,j}$ are then used to perform LSMDS and to obtain a canonical form in the embedding space. Instead of starting with a random point set, we initialize the canonical form to the canonical form computed using classical MDS. This reduces the risk of getting stuck in a local minimum when performing the iterations required for LSMDS, since classical MDS cannot get stuck in local extrema. Computing the $\binom{n_P}{2}$ geodesic paths on the surface S consisting of n vertices via FMM takes $O(n_P n \log n)$ time and computing the canonical form given the weights and dissimilarities takes $O(n_P^2 t)$ time for LSMDS, where t is the number of iterations required for convergence. Hence, this algorithm is computationally expensive. However, computing the canonical form once per surface S can be viewed as a preprocessing step.

Second, we estimate the geodesic distance between any pair p_i and p_j of vertices on S. Note that i and j do not have to be elements of P. To estimate the geodesic distance, we first compute the geodesic distance $\delta_{i,j}$ between p_i and p_j via FMM and analyze the resulting geodesic path. If the path does not trace a hole of S, a valid geodesic path was found. We report the result along with an error bound of zero, since the exact geodesic path was found. Otherwise, the path traces a hole of S. If $i \notin P$ ($j \notin P$ respectively), p_i (p_j respectively) is projected to the canonical form. To project p_i to the canonical form, all of the geodesic distances $\delta_{i,r}, r \in P$ and weights $\omega_{i,r}, r \in P$ are computed via FMM in $O(n \log n)$ time and an optimization problem with k variables is solved using a quasi-Newton method. Once the embedded points X_i and X_j are known, we use the Euclidean distance $d_{i,j}(X)$ in embedding space to approximate the geodesic distance between p_i and p_j on S. The approximation error of $d_{i,j}(X)$ is bounded by $\max(\left|\hat{d}_{i,j} - d_{i,j}(X)\right|, |\delta_{i,j} - d_{i,j}(X)|)$. This error bound is finite iff there exists a path from p_i to p_j on S.

4 Experimental Results

The accuracy of the approximation was evaluated using a synthetic data set. The complete data set of an artist-created human body consisting of 20002 vertices shown in Figure 2 was modified to contain holes as shown in Figure 2. In the Figure, holes are shown in blue, points used to compute the canonical form are shown in red, and test vertices are shown in green. We chose 509 testing samples on the incomplete model. We consider pairs of sample points p_i and p_j where the geodesic path between p_i and p_j crosses at least 20 triangles of S that have a vertex on the boundary of a hole. The following distances are examined: the true geodesic distance $\hat{\delta}_{i,j}$ computed via FMM on the complete surface, the upper bound $\hat{\delta}_{i,j}^{upper}$ computed via FMM on the incomplete surface, and the estimate $d_{i,j}$ along with a relative error bound $e_{i,j}$ computed as proposed in this paper using 4000 samples to compute the canonical form in $\mathbb{R}^3$. We used these distances to find the true relative errors $e(\hat{\delta}_{i,j}^{upper})$ of $\hat{\delta}_{i,j}^{upper}$ and $e(d_{i,j})$ of $d_{i,j}$. Figure 3 shows the percentage of distances where $e(d_{i,j})$ is smaller than or equal to $e(\hat{\delta}_{i,j}^{upper})$. We can see that $\hat{\delta}_{i,j}^{upper}$ is more accurate than $d_{i,j}$ for small relative error bounds. For larger relative error bounds, $d_{i,j}$ is more accurate than $\hat{\delta}_{i,j}^{upper}$.

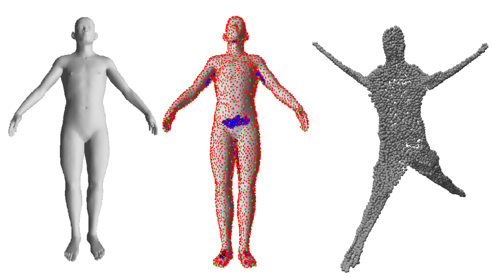

Figure 2: *Left: Complete model. Middle: Modified model. Right: Canonical form.*

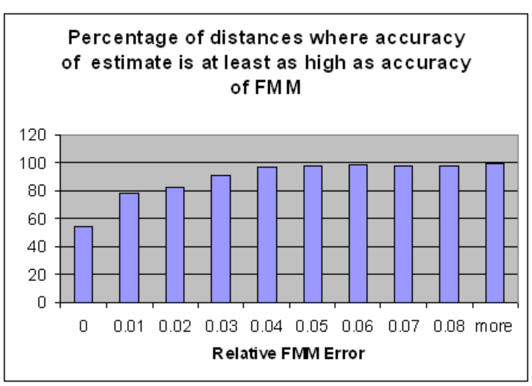

Figure 3: *Blue column shows the percentage of distances where $e(d_{i,j})$ is smaller than or equal to $e(\hat{\delta}_{i,j}^{upper})$.*

5 Conclusion

Taken together, the preceding discussion proves the main theorem. Let n denote the number of vertices of S, let n_P denote the number of sample points computed via FPS, and let t denote the number of iterations performed to compute the canonical form.

Theorem 1 *An incomplete triangular manifold S can be preprocessed in $O(n_P(n \log n + n_P t))$ time, such that given any pair of query vertices p_i and p_j with $1 \leq i, j \leq n$ on S, we can report an estimate of $\hat{\delta}_{i,j}$ along with error bounds $\hat{\delta}_{i,j}^{lower}$ and $\hat{\delta}_{i,j}^{upper}$ in $O(n \log n)$ time. The upper bound $\hat{\delta}_{i,j}^{upper}$ is worst-case optimal.*

An interesting open question is to find an easily computable worst-case optimal lower bound of $\hat{\delta}_{i,j}$. Another open question is how to choose the confidence values $\omega_{i,j}$ to optimize the quality of the estimate.

References

[1] Borg I., Groenen P.: *Modern Multidimensional Scaling Theory and Applications.* Springer, 1997.

[2] Elad A., Kimmel R.: On bending invariant signatures for surfaces. *PAMI 25*, 10 (2003), 1285–1295.

[3] Eldar Y., Lindenbaum M., Porat M., Zeevi Y.: The farthest point strategy for progressive image sampling. *TIP 6*, 9 (1997), 1305 – 1315.

[4] Gower J. C.: Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53* (1966), 325–338.

[5] Gower J. C.: Adding a point to vector diagrams in multivariate analysis. *Biometrika 55*, 3 (1968), 582–585.

[6] Jain V., Zhang H.: A spectral approach to shape-based retrieval of articulated 3d models. *CAD 39*, 5 (2007), 398–407.

[7] Kimmel R., Sethian J.: Computing geodesic paths on manifolds. *PNAS 95*, 15 (1998), 8431–8435.

[8] Moenning C., Dodgson N.: *Fast Marching farthest point sampling.* Tech. Rep. 562, University of Cambridge, Computer Laboratory, 2003.

[9] Mitchell J., Mount D., Papadimitriou C.: The discrete geodesic problem. *SIAM J. Comput. 16* (1987), 647–668.

[10] Rosman G., Bronstein A. M., Bronstein M. M., Kimmel R.: Topologically constrained isometric embedding. In *MLPR* (2006).

[11] Surazhsky V., Surazhsky T., Kirsanov D., Gortler S. J., Hoppe H.: Fast exact and approximate geodesics on meshes. *ACM Trans. Graph. 24*, 3 (2005), 553–560.

[12] Wuhrer S., Shu C., Bose J., Ben Azouz Z.: Bounded Approximations of Geodesics for Triangular Manifolds with Partially Missing Data. Technical Report NRC-49316/ERB-1146, 2007. http://cg.scs.carleton.ca/~swuhrer/geodesics.pdf