Bounded approximations of geodesics for triangular manifolds with partially missing data

Wuhrer, Stefanie; Shu, Chang; Bose, P.; Ben Azouz, Zouhour

Report (National Research Council Canada. Radio and Electrical Engineering Division : ERB); no. ERB-1146, 2007-05

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC·CNRC

*Bounded Approximations of Geodesics for Triangular Manifolds with Partially Missing Data**

Wuhrer, S., Shu, C., Bose, P., and Ben Azouz, Z.
May 2007

* 11 pages. May 2007. NRC 49316. ERB-1146.

Canada

Bounded Approximations of Geodesics for Triangular Manifolds with Partially Missing Data

Stefanie Wuhrer[1,2]

Chang Shu[1]

Prosenjit Bose[2]

Zouhour Ben Azouz[1]

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

Abstract

In this paper we present an 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 a maximum relative error bound of the approximation. The error bound is shown to be 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 method to add an object to the embedding computed via least-squares MDS.

1. Introduction

The computation of geodesic distances on a triangular manifold is a common operation in many applications. In computer graphics, geodesic distances were recently applied to produce parameterizations with low distortion, which has direct applications to texture mapping and morphing [ZKK02, ZSGS04, PC05]. Geodesic distances can further be applied to classify isometric shapes [YHHC00, HSKK01, EK03, JZ]. Bronstein et al. [BBK06b, BBK03, BBK05, BBKY06] use geodesic distances for face recognition. The 3D models used in these applications usually come from digitizing real-world objects from a discrete set of measurements using a 3D laser-range scanner or image-based reconstruction. Therefore, the reconstructed surfaces are often incomplete. Despite of efforts on fixing holes in triangular meshes, a general reliable hole-filling algorithm is still unavailable.

In this paper, we explore the problem of computing estimates on geodesic distances with worst-case optimal upper and lower 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 [MMP87, KS98, SSK*05] 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, such as matching isometric objects, 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 the fast marching technique (FMM) introduced by Kimmel and Sethian [KS98]. Note that 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. Furthermore, we 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 [EK03].

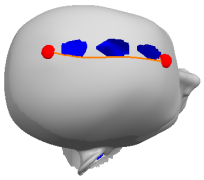


Figure 1: *Paths computed via FMM tracing holes.*

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 [Gow68]. The Euclidean distance between the embedded points approximates the true geodesic distance between the original points.

Section 2 reviews related work on geodesic distance computations on triangular manifolds and on using MDS for geometry processing. Section 3 discusses the theory of canonical forms obtained via MDS and derives a method to add an additional object to the least-squares MDS (LSMDS) embedding. Section 4 presents the algorithm used to obtain approximations of geodesic distances with worst-case tight error bounds and Section 5 shows experimental results. Finally, Section 6 gives concluding remarks.

2. Related Work

Assume we are given a graph G consisting of a set V of n vertices and a set E of m edges. Let each edge be associated with a non-negative weight. The problem of computing the shortest weighted paths between one given vertex v and all the vertices in V, known as the *single source shortest path problem (SSSP)*, is well-studied. When considering a triangular manifold S, the vertex set V corresponds to the vertices $p_i, i = 1,\ldots,n$ of S and the edge set E corresponds to the $O(n)$ edges of S. The resulting shortest paths correspond to geodesic paths on the triangular manifold. Hence, we assume $m = O(n)$ in the following.

A well-known approach to solving the SSSP problem is Dijkstra's graph search algorithm [Dij59]. The algorithm's running time is $O(n \log n)$ for planar graphs. Dijkstra's graph search algorithm gives an approximation, since the shortest paths are always measured along edges of the graph and never cut through faces of the graph.

Kimmel and Sethian [KS98] present an approach called fast marching method on triangular domains (FMM) that solves the SSSP problem by solving the Eikonal equation on a triangular grid based on [Set96]. The algorithm's running time is $O(n \log n)$ and therefore optimal. The algorithm proceeds by iteratively unfolding all of the triangles of the triangular mesh. The shortest paths found using the FMM method only approximate the true geodesic distances on the triangular mesh. 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.

Surazhsky et al. [SSK*05] implemented the algorithm by Mitchell et al. [MMP87] to find the exact geodesic paths on triangular meshes from one source point to all other points of the mesh. Although the worst-case running time of the algorithm is $O(n^2 \log n)$, they found the algorithm's average running time to be much lower. The exact algorithm by Mitchell et al. is then modified to obtain an algorithm that solves the SSSP problem approximately. The approximation ratio is bounded by $(1+\varepsilon)$ for any positive constant ε.

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 illustrated in Figure 1. This results in erroneous geodesic distances. In this paper, we approximate geodesic distances for incomplete surface descriptions by applying MDS to the weighted geodesic distances computed via FMM and by measuring Euclidean distances in the compressed space. Elad and Kimmel [EK03] applied a similar technique to the problem of matching three-dimensional objects. They match the objects based on MDS embeddings called *canonical forms*.

Canonical forms were then used by Bronstein et al. [BBK06b, BBK03, BBK05, BBKY06, BBK06a] for face recognition. Jain and Zhang [JZ] improved the efficiency of surface matching algorithm using canonical forms by computing an approximate canonical form via the Nyström approximation and Jain et al. [JZvK] find one-to-one correspondences of isometric surfaces using canonical forms.

3. Canonical Forms

Elad and Kimmel [EK03] define the *canonical form* of a triangular manifold S as the mapping of S to a low-dimensional Euclidean space, such that the Euclidean distances between the mapped vertices approximate the geodesic distances between the original vertices well. The canonical form is computed via MDS with the geodesic distances between vertices on the triangular manifold as dissimilarities. This has the effect that the canonical form of a non-rigid body is posture invariant. Elad and Kimmel use this invariance of the canonical form to match objects in different postures.

We compute the canonical form of a triangular manifold using a sample set of vertices on the manifold. In the case of incomplete surfaces, geodesic distances are estimated using the property that Euclidean distances between vertices of the canonical form approximate geodesic distances between vertices of the original manifold. To compute the canonical form, MDS is employed with the geodesic distance computed using the FMM algorithm as dissimilarities.

3.1. 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 $X_1,\ldots,X_n$ 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. In the following, two ways to compute MDS are reviewed. Classical MDS, also called Principal Coordinate Analysis (PCO), is a method closely related to Principal Component Analysis that assumes that the dissimilarities are Euclidean distances in a high dimensional space and that 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. LSMDS 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.

3.1.1. Classical MDS

Classical MDS was introduced by Gower [Gow66]. It aims to find points X_i in k dimensions, such that $E_{PCO} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}(\delta_{i,j}^2 - d_{i,j}(X)^2)$ is minimized. Denote the position vector of point X_i by $\vec{x}_i = \begin{bmatrix}x_{i,1} & x_{i,2} & \ldots & x_{i,k}\end{bmatrix}^T$ and denote the point matrix by $X = \begin{bmatrix}\vec{x}_1 & \vec{x}_2 & \ldots & \vec{x}_n\end{bmatrix}^T$. Let $B = XX^T$ be the inner product matrix of X.

The goal is to find X. Note that the quality of the point set X is invariant w.r.t. translations, rotations, and reflections. Therefore, we can choose the centroid of the point set to be the origin of the coordinate system, that is, $\sum_{r=1}^{n} x_{r,i} = 0, i = 1,\ldots,n$. This yields $B = HAH$, where A is a matrix with elements $A_{i,j} = -\frac{1}{2}d_{i,j}(X)^2, 1 \leq i,j \leq n$ and where $H = I - \frac{1}{n}\mathbf{1}$ with I as $n \times n$ identity matrix and $\mathbf{1}$ as $n \times n$ matrix containing 1 at each position [CC01, p.33].

A spectral decomposition is performed on B. Let $B = V\Lambda V^T$, where V is the matrix of eigenvectors of B and Λ is the diagonal matrix of eigenvalues of B, be the spectral decomposition with sorted eigenvalues in decreasing order. Then $X = V\Lambda^{\frac{1}{2}}$ yields the sought points. A practical algorithm is given by Cox and Cox [CC01, p.38-39]. Classical MDS minimizes $E_{PCO} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}(\delta_{i,j}^2 - d_{i,j}(X)^2)$ if $\delta_{i,j}$ are Euclidean distances in a high-dimensional space. This algorithm takes $O(kn^2)$ time and $O(n^2)$ space.

One disadvantage of this method is that the objective function E_{PCO} aims to minimize a Frobenius norm and is not meaningful if the dissimilarities cannot be viewed as Euclidean distances in a high dimensional space. Further, no weights or confidence values can be assigned to the dissimilarities between objects. The main advantage of classical MDS is that the solution cannot get stuck in local minima.

3.1.2. Least-Squares MDS

LSMDS aims to find points X_i in k dimensions, such that $E_{LS} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\omega_{i,j}\left(\delta_{i,j} - d_{i,j}(X)\right)^2$ is minimized, where $\omega_{i,j}$ are non-negative weighting coefficients with $\omega_{i,j} = \omega_{j,i}$. Since the objective function E_{LS} is a complex function, it is easier to iteratively approximate the objective function by a simple function. This approach is pursued in the algorithm *Scaling by Maximizing a Convex Function* (SMACOF) that is explained by Borg and Groenen [BG97, p.146-155] and used by Elad and Kimmel [EK03] to compute canonical forms. SMACOF proceeds by iteratively refining a simple majorization function that bounds the objective function E_{LS} from above.

We can rewrite the objective function [BG97] as

$$E_{LS} = \alpha + \beta - \gamma$$

with $\alpha = \sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{i,j}\delta_{i,j}^2, \beta = \sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{i,j}d_{i,j}(X)^2$ and $\gamma = 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{i,j}d_{i,j}(X)\delta_{i,j}$ and using the Cauchy-Schwartz inequality bound it by

$$E_{LS} \leq \sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{i,j}\delta_{i,j}^2 + tr(X^TVX) - 2tr(X^TB(Z)Z) = \tau,$$

where V is an $n \times n$ matrix with elements

$$V_{i,j} = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\omega_{i,j}(e_i - e_j)(e_i - e_j)^T$$

with e_i as ith column of the $n \times n$ identity matrix, Z is a possible solution for X, and $B(Z)$ is an $n \times n$ matrix with elements

$$B_{i,j} = \begin{cases} -\frac{\omega_{i,j}\delta_{i,j}}{d_{i,j}(Z)} & i \neq j, d_{i,j}(Z) \neq 0 \\ 0 & i \neq j, d_{i,j}(Z) = 0 \\ \sum_{j=1,j\neq i}^{n} B_{i,j} & otherwise \end{cases}.$$

The minimum of τ can be found by setting the gradient of τ to zero. This approach is presented by Borg and Groenen [BG97] and used by Elad and Kimmel [EK03]. Setting the gradient to zero yields an iterative approach to update the current solution of X using the formula $X_{r+1} = V^+B(X_r)X_r$, where X_r is the position of X after r iterations and V^+ is the pseudo-inverse of V. The initial configuration X_0 can be chosen as a random point set. It can be shown that this approach of solving the optimization problem is equivalent to using a gradient descent method to minimize E_{LS} and therefore has only a linear convergence rate. Further, each iteration requires the inversion of a $n \times n$ matrix. The algorithm takes $O(n^2t)$ time, where t is the number of iterations needed until convergence of X.

We minimize τ using a quasi-Newton method instead. The quasi-Newton method used is the *limited-memory Broyden-Fletcher-Goldfarb-Shanno* (LSBFGS) scheme discussed by Bronstein et al. [BBKY06]. This quasi-Newton method offers the advantage of obtaining close to quadratic convergence rates without the need to compute the inverse of the Hessian matrix explicitly in each step. Instead LSBFGS updates an approximation to the inverse Hessian matrix in each iterative step, such that the approximation converges to the true inverse of the Hessian in the limit. Using LSBFGS instead of a simple gradient descent method to minimize τ reduced the number of iterations t required until convergence significantly in our experiments.

The main disadvantage of this method is that the optimization can get stuck in local minima. However, Elad and Kimmel [EK03] found that this method is more accurate in practice than classical MDS.

3.2. Adding an object to the MDS embedding

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. An inefficient approach to this problem is to disregard the previously computed points $X_1,X_2,\ldots,X_n$ and to apply MDS to $n+1$ objects as before. A more efficient approach to this problem for classical MDS was derived by Gower [Gow68]. Assume that we already computed the $n\times k$ matrix X containing the positions of the first n points $X_1,\ldots,X_n$. To find the embedding X_{n+1} of the object O_{n+1} in k-dimensional space, we use the previously computed embedding X and the dissimilarities $\delta_{n+1,1},\ldots,\delta_{n+1,n}$ describing the relationship between O_{n+1} and the remaining objects. Let d_i^2 denote the squared distance between $X_i, i=1,\ldots,n$ and the centroid of X. Since the centroid of X was chosen to be the origin, $d_i^2 = B_{i,i}$, where $B_{i,i}$ is the diagonal element of the previously computed matrix $B = XX^T$. We can compute an $n\times 1$ vector D containing elements $D_i = d_i^2 - d_{n+1,i}^2$ if we use the optimal case $d_{n+1,i}(X) = \delta_{n+1,i}, 1\le i\le n$. The embedding X_{n+1} can then be computed as

$$X_{n+1} = \frac{1}{2}(X^TX)^{-1}X^TD.$$

Note that the matrix X^TX that is inverted has dimension $k\times k$ and that the inverse is therefore fast to compute in typical applications (with $k=2$ or $k=3$). Gower proved that the position of X_{n+1} computed using this method is identical to the position of X_{n+1} when classical MDS is performed for all the $n+1$ objects at once.

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

$$E_{LS}^* = \sum_{i=1}^{n}\omega_{n+1,i}\left(\delta_{n+1,i} - d_{n+1,i}(X)\right)^2,$$

which can be written as

$$E_{LS}^* = \alpha^* + \beta^* - \gamma^*$$

where $\alpha^* = \sum_{i=1}^n \omega_{n+1,i}\delta_{n+1,i}$,
$\beta^* = \sum_{i=1}^n \omega_{n+1,i}(\vec{x}_i - \vec{x}_{n+1})^T(\vec{x}_i - \vec{x}_{n+1})$ and
$\gamma^* = 2\sum_{i=1}^n \omega_{n+1,i}\delta_{n+1,i}\sqrt{(\vec{x}_i - \vec{x}_{n+1})^T(\vec{x}_i - \vec{x}_{n+1})}$.

We can now 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 a LSBFGS quasi-Newton approach. Although the authors are not aware that this addition to the LSMDS embedding was discussed previously, this is not the main contribution of this work.

4. Geodesic Distance Estimation

We use the canonical form to estimate the geodesic distance between any given pair p_i and p_j with $1\le i,j\le n$ of vertices on a triangular manifold S with partially missing data. The main idea of the approach is to compute the canonical form of the manifold based on weighted geodesic distances on S. That is, we use geodesic distances as dissimilarities $\delta_{i,j}\forall i,j\in S$ and we use confidence values $\omega_{i,j} = 1 - \frac{m_{i,j}^h}{m_{i,j}}$, where $m_{i,j}$ is the number of edges on the geodesic path from p_i to p_j on S and where $m_{i,j}^h$ is the number of edges tracing a hole of S on the geodesic path on S from p_i to p_j. Since S is a manifold, we can find the edges of S tracing a hole of S as 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 edges of S have similar lengths. 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 [RBBK06]. Since we wish to extrapolate information using the metric property of the manifold, we weigh distances tracing a hole of S less than accurate distances, but we do not disregard those distances completely.

Let $\hat{S}$ denote the complete surface partially represented by S. Let $\hat{\delta}_{i,j}$ be the geodesic distance between p_i and p_j on $\hat{S}$ and let $\delta_{i,j}$ be the geodesic distance between 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. After computing the canonical form, p_i is associated with a point X_i in embedding space and p_j is associated with a point X_j in embedding space. As before, $d_{i,j}(X)$ denotes the Euclidean distance between X_i and X_j in embedding space.

Note that 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 $\hat{\delta}_{i,j} \forall i,j \in S$ contains therefore 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 with indices in P of S. 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. For real-life data sets, an algorithm using quadratic storage is not feasible. Taking a sample set P of vertices for the embedding has a negative effect on the quality of the results however and the sample set should therefore be large enough to represent the overall shape of S well. We use Voronoi sampling to choose the sample set P. The approach is discussed in Section 4.1.

Note that the canonical form has the property that Euclidean distances in the canonical form approximate geodesic distances on S well according to the optimality measure $\sum_{i\in P}\sum_{j\in P}\left(1-\frac{m^h_{i,j}}{m_{i,j}}\right)\left(\delta_{i,j}-d_{i,j}(X)\right)^2$. 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. In fact, the error made by approximating $\hat{\delta}_{i,j}$ by $d_{i,j}$ can be bounded and the bound is worst-case optimal. The approach used to compute a worst-case optimal error bound for the approximation is discussed in Section 4.2.

A detailed overview of the algorithm suggested to estimate geodesic distances on incomplete triangular manifolds along with an analysis of the algorithm is given in Section 4.3.

4.1. Voronoi sampling

We aim to compute a set P of n_P sample points on S that represent the overall shape of S well. We choose a set of samples with uniformly distributed geodesic distance.

Voronoi sampling, also called Farthest Point Sampling (FPS), allows to obtain uniformly distributed samples in an iterative way. FPS was introduced to image processing by Eldar et al. [ELPZ97]. It starts from a random sample and iteratively computes the next sample as the vertex which is farthest from the samples computed so far. If the metric is chosen to be the Euclidean distance, the new sample can be viewed as the vertex which is the center of the largest circle not containing any sample points. This observation yields an elegant relationship between the sampling approach and Voronoi diagrams that can be used to obtain an efficient sampling algorithm. Eldar et al. show that this sampling approach has favorable properties such as a high data acquisition rate and good anti-aliasing properties.

The FPS approach can be extended to obtaining samples on a 3D surface. In this case, the distance metric to measure uniformity is the distance metric intrinsic to the surface being sampled, that is the geodesic distance on the surface. This yields a relationship between the sampling approach and *generalized Voronoi diagrams*. Generalized Voronoi diagrams are Voronoi diagrams built in an arbitrary metric space. It is well-known that generalized Voronoi diagrams may have disconnected and non-convex cells, which complicates the use of Voronoi diagrams for the sampling approach. Moenning and Dodgson [MD03] combine FMM and FPS to overcome the problem of storing the Voronoi diagram explicitly. The main idea is to run FMM while tracking multiple propagation fronts originating from different sample points. Moenning and Dodgson claim that the resulting algorithm takes $O(n\log n)$ time to compute n_P uniformly distributed sample points on a surface with n vertices, but the algorithm appears to take $O(n_P n\log n)$ time.

Combining FMM and FPS to obtain samples used to perform MDS works well in practice [EK03, JZvK]. Hence, we use FPS to obtain P and compute an MDS embedding.

4.2. Worst-Case Optimal Error Bound of the Approximation

The error made by approximating the geodesic distance $\hat{\delta}_{i,j}$ on the complete surface by the Euclidean distance $d_{i,j}$ between X_i and X_j in embedding space can be bounded by finding an upper bound and a lower bound for $\hat{\delta}_{i,j}$.

A lower bound on $\hat{\delta}_{i,j}$ is given by the Euclidean distance $\hat{d}_{i,j}$ between p_i and p_j, since the shortest path can not be shorter than the straight line segment between the two points.

An upper bound on $\hat{\delta}_{i,j}$ is given by $\delta_{i,j}$. This upper bound 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 however, that the upper bound only exists if there exists a path on S between p_i and p_j. If this is not the case, p_i and p_j are located on two distinctively connected components of S. In this case, the error of the approximation $d_{i,j}$ is not bounded from above. Hence, only an error bound of infinity can be given. Note that this error bound is optimal, because we can always find a surface $\hat{S}$ where the geodesic distance $\hat{\delta}_{i,j}$ between p_i and p_j exceeds any given finite length by building high mountains between the regions corresponding to the distinct connected components containing p_i and p_j on S.

Once a lower bound $\hat{\delta}^{lower}_{i,j}$ and an upper bound $\hat{\delta}^{upper}_{i,j}$ are

known, the *relative error* e of the approximation of $\hat{\delta}_{i,j}$ by $d_{i,j}$ is computed as $e = \frac{\max(|d_{i,j}-\hat{\delta}_{i,j}^{lower}|,|d_{i,j}-\hat{\delta}_{i,j}^{upper}|)}{d_{i,j}}$.

4.3. 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. Since the aim is not to reduce the dimensionality of the data, we choose $k = 3$ and $k = 4$ as dimensions of the embedding space in our experiments. 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 $\delta_{i,j}, i,j \in P$ on S along with confidence values $\omega_{i,j}, i,j \in P$. Note that 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, since the longest edge of S and the widest angle of the triangles in the triangulation of S are small. The minor theoretical flaw of using geodesic distances computed via FMM instead of the exact geodesic distances on S can further be overcome by using any known exact algorithm to compute geodesic distances on S [MMP87, SSK*05]. The main advantage of choosing FMM is that it can be combined with FPS to obtain an elegant and efficient sampling approach [MD03]. Furthermore, FMM is easy to implement and solves the SSSP problem in optimal asymptotic running time on triangular meshes.

The pairwise geodesic distances $\delta_{i,j}, i,j \in P$ on S along with confidence values $\omega_{i,j}, i,j \in P$ are then used to perform LSMDS and to obtain a canonical form in 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 $\max(|\hat{d}_{i,j} - d_{i,j}(X)|, |\delta_{i,j} - d_{i,j}(X)|)$. Recall that this error bound is finite if and only if there exists a path from p_i to p_j on S.

After $O(n_P(n \log n + n_P t))$ preprocessing time, this algorithm reports in $O(n \log n)$ time an approximation of the geodesic distance between p_i and p_j on S along with a worst-case optimal error bound. An overview of the algorithm as chart is given in Figure 2. Note that both the preprocessing time and the approximation time are dominated by the time needed to compute $\delta_{i,j}$ using the FMM algorithm. If we can find a faster way to perform this computation, the algorithm's running time is expected to decrease significantly.

5. 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 3 was modified to simulate the holes present in the models of the Civilian American and European Surface Anthropometry Resource (CAESAR) data base [RDP99] as shown in Figure 3. In the Figure, holes are shown in blue. We found 362 testing samples on the incomplete model. The following distances between each pair of sample points p_i and p_j where the geodesic path between p_i and p_j traces at least 20 boundary edges of S are considered: 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 worst-case optimal relative error bound $e_{i,j}$ computed as proposed in this paper using 4000 samples to compute the canonical form. We used these distances to find the true relative errors of both $\hat{\delta}_{i,j}^{upper}$ and $d_{i,j}$. A histogram of the error bounds $e_{i,j}$ and relative errors of $\hat{\delta}_{i,j}^{upper}$ and $d_{i,j}$ is shown in Figure 4. Figure 4 (a) shows the number of occurrences of relative errors in the bins shown at the x-axis. It is apparent that $e_{i,j}$ overestimates the error of $d_{i,j}$ in most cases. Figure 4 (b) shows the percentage of distances where the relative error of $d_{i,j}$ is smaller than or equal to the relative error of $\hat{\delta}_{i,j}^{upper}$. 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 $\hat{\delta}_{i,j}^{upper}$.

We furthermore evaluated the quality of approximation of the approach using three real-life data sets. The first data set is the model of a bag that was reconstructed from multiple images. The model consists of 7091 vertices, 520 of which were used to compute a canonical form. We used 511 test

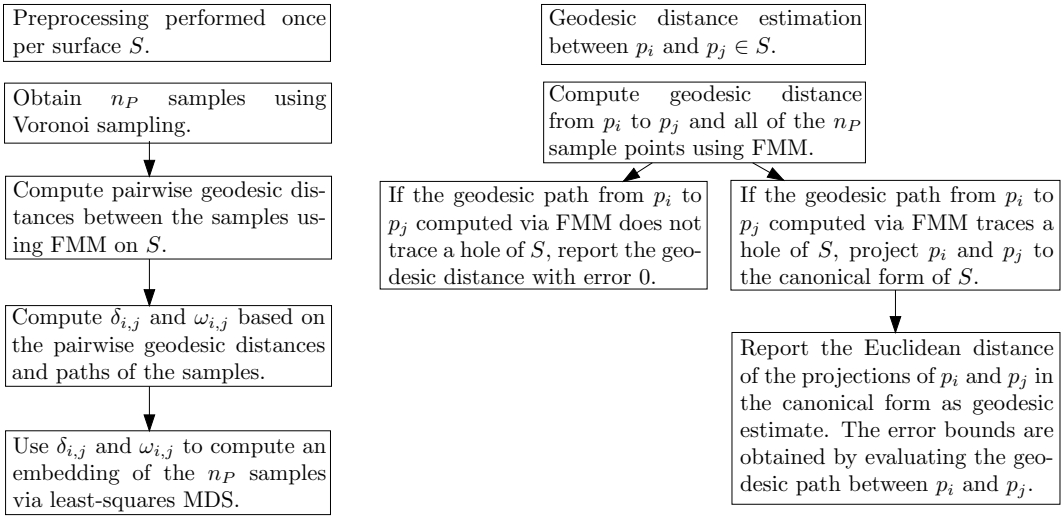
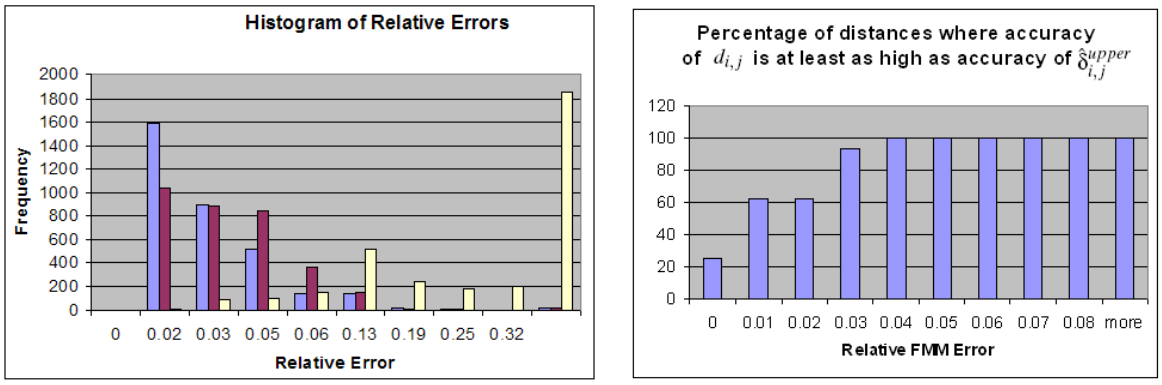


Figure 2: *Chart describing the overview of the proposed algorithm. The left side of the chart describes the preprocessing performed once per surface. The right side of the chart describes the steps taken to compute the geodesic distance between two points on the surface after preprocessing.*



(a) (b)

Figure 4: *(a) Histogram shows the number of times that relative error bound in a certain bin occurs. The occurrence of $\hat{\delta}_{i,j}^{upper}$ is shown in blue, the occurrence of $d_{i,j}$ is shown in purple, and the occurrence of $e_{i,j}$ is shown in yellow. (b) Blue column shows the percentage of distances where the relative error of $d_{i,j}$ is smaller than or equal to the relative error of $\hat{\delta}_{i,j}^{upper}$.*

vertices that were not used to compute the canonical form to evaluate the quality of the geodesic approximations. The model is shown in Figure 5. Vertices used to compute the canonical form are shown in red and vertices used to evaluate the quality of the geodesic approximation are shown in green. Holes in the model are visible in blue. Figure 6 shows the canonical form obtained via LSMDS with $\mathbb{R}^3$ as embedding space obtained for the bag model.

The results of the quality of the approximations of geodesic distances between the green sample points for $\mathbb{R}^3$ and $\mathbb{R}^4$ as embedding space are shown in the first two rows of Table 1. All of the error bounds in Table 1 are given as relative error bounds. The first row discusses the results when LSMDS is performed with $\mathbb{R}^3$ as embedding space. Note that 83.87% of the geodesic distance approximations have a relative error under 0.25. This means that for 83.87% of the geodesic distance approximations, the approximation $d_{i,j}$ is bounded by $\frac{3}{4}\hat{\delta}_{i,j} \leq d_{i,j} \leq \frac{5}{4}\hat{\delta}_{i,j}$. Further, 99.78% of the geodesic distance approximations are bounded by $\frac{1}{2}\hat{\delta}_{i,j} \leq d_{i,j} \leq \frac{3}{2}\hat{\delta}_{i,j}$ and all of the geodesic distance approximations are bounded by $\frac{1}{4}\hat{\delta}_{i,j} \leq d_{i,j} \leq \frac{7}{4}\hat{\delta}_{i,j}$. The second row discusses the results when LSMDS is performed with $\mathbb{R}^4$ as embedding space. Allowing a higher-dimensional embedding space yields higher accuracy until the dimension of the embedding space is sufficient. We can see that the quality of the result only changed insignificantly. Hence, we can conclude that $\mathbb{R}^3$ is a suitable embedding space for the bag.

We further estimated all the distances from a fixed source

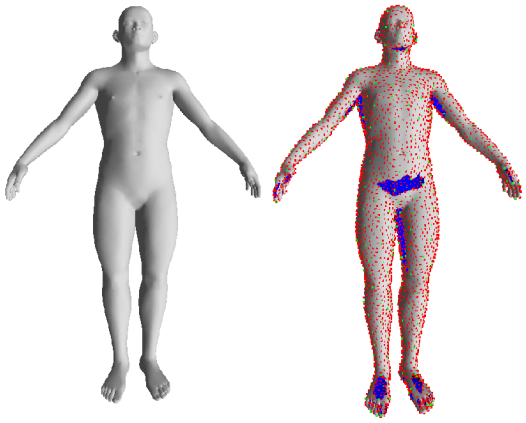


Figure 3: *Model of an artistic human model. Left: Complete model. Right: Modified model. The points used to compute the canonical form are shown in red and the test vertices are shown in green.*

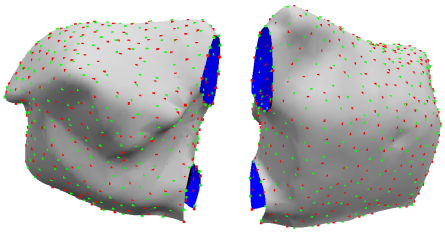


Figure 5: *Model of a bag obtained by image-based reconstruction. The points used to compute the canonical form are shown in red and the test vertices are shown in green.*

point for the bag model. The source point is shown as red dot in Figure 7. Figure 8 (a) shows a color display of the distance estimates from the source point. The color scale shows small distances as red color and far distances as blue color. Figure 8 (b) shows a color display of the corresponding error bounds. Detail views of the hole boundary of the same color-displayed model are shown in Figure 8 (c) and (d), respectively. The color scale shows small error bounds as red color and large error bounds as blue color. We can see that the distance estimate proceeds smoothly around the hole of the bag. However, the error bounds get larger close to the hole of the bag, since some shape information is not available.

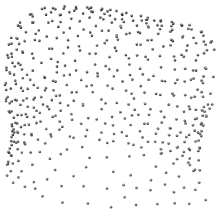


Figure 6: *Canonical form of the bag obtained via LSMDS.*

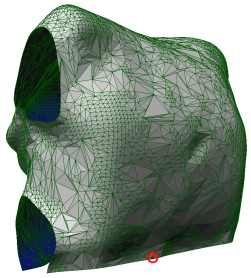


Figure 7: *Bag model with red dot as source point for SSSP computation.*

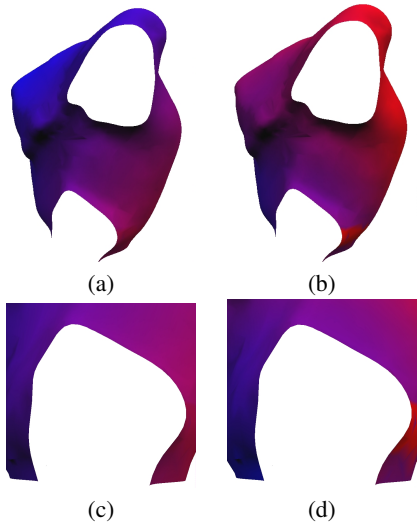


Figure 8: *(a) Color display of the geodesic distance estimates from one source point. Red means small distance and blue means large distance. (b) Color display of the error bounds of the distances shown in (a). Red means small error bound and blue means large error bound. (c) Detail view of (a). (d) Detail view of (c).*

The two other data sets are human models from the CAESAR data base that were acquired using a 3D range scans. The first of the human models consists of 209660 vertices, 3707 of which were used to compute a canonical form. We used 1218 test vertices that were not used to compute the canonical form to evaluate the quality of the geodesic approximations. The model is shown in Figure 9. Vertices used to compute the canonical form are shown in red and vertices used to evaluate the quality of the geodesic approximation are shown in green. Holes in the model are visible in blue. Figure 11 shows the canonical form with $\mathbb{R}^3$ as embedding space obtained for the model. Figure 10 shows a detailed view of holes on the arm and upper body of the model.

The results of the quality of the approximations of geodesic distances between the green sample points for $\mathbb{R}^3$ and $\mathbb{R}^4$ as embedding space are shown in the third and fourth rows of Table 1. The third row discusses the results when

LSMDS is performed with $\mathbb{R}^3$ as embedding space. Note that although less than 2% of the original 209660 vertices of the model were used as samples to compute the canonical form, 99.50% of the bounded geodesic distance approximations have a relative error under 0.75. The fourth row discusses the results when LSMDS is performed with $\mathbb{R}^4$ as embedding space. Again, we can see that $\mathbb{R}^3$ is a suitable embedding space for the model, since the quality of the result only changes insignificantly by allowing an embedding in higher dimensions.

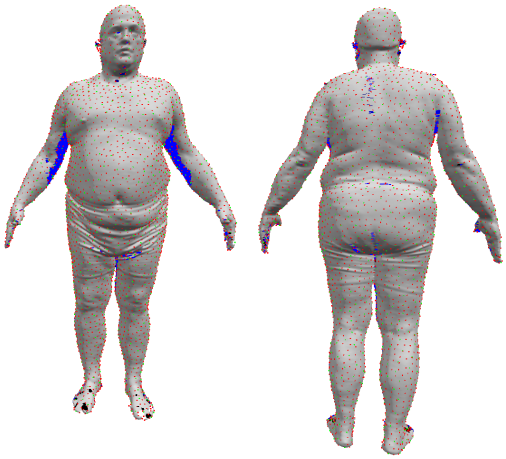


Figure 9: *The points used to compute the canonical form are shown in red and the test vertices are shown in green.*

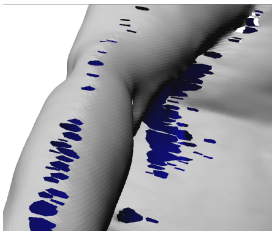


Figure 10: *Detail views of holes on arm and upper body.*

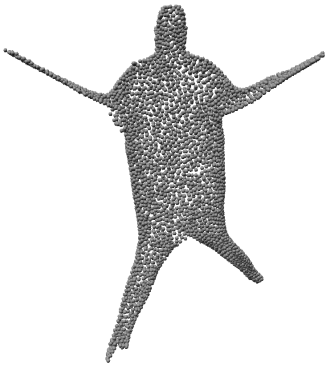


Figure 11: *Canonical form obtained via LSMDS.*

The second of the human models consists of 160190 vertices, 3707 of which were used to compute a canonical form. We used 1218 test vertices that were not used to compute the canonical form to evaluate the quality of the geodesic approximations. The model is shown in Figure 12. Vertices used to compute the canonical form are shown in red and vertices used to evaluate the quality of the geodesic approximation are shown in green. Holes in the model are visible in blue. Figure 13 shows the canonical form with $\mathbb{R}^3$ as embedding space obtained for the model.

The results of the quality of the approximations of geodesic distances between the green sample points for $\mathbb{R}^3$ and $\mathbb{R}^4$ as embedding space are shown in the fifth and sixth rows of Table 1. The fifth row discusses the results when LSMDS is performed with $\mathbb{R}^3$ as embedding space. The sixth row discusses the results when LSMDS is performed with $\mathbb{R}^4$ as embedding space. Again, we can see that a relatively small sample set yields satisfying results with $\mathbb{R}^3$ as embedding space.

Figure 12: *The points used to compute the canonical form are shown in red and the test vertices are shown in green.*

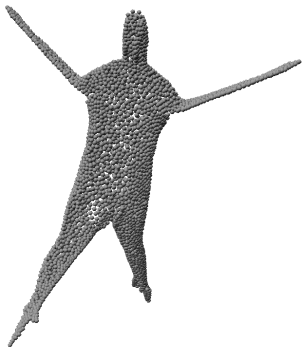


Figure 13: *Canonical form obtained via LSMDS.*

Object Embedding / % of errors under error bound	0.25	0.5	0.75	1
bag				
MDS $\mathbb{R}^3$	83.87	99.78	100.00	100.00
MDS $\mathbb{R}^4$	83.87	99.78	100.00	100.00
csr0001a				
MDS $\mathbb{R}^3$	75.15	93.07	99.50	99.90
MDS $\mathbb{R}^4$	75.14	93.12	99.48	99.95
csr0106a				
MDS $\mathbb{R}^3$	69.64	87.12	98.43	99.75
MDS $\mathbb{R}^4$	70.47	87.67	98.65	99.92

Table 1: *Quality of approximation. The table shows the percentage of estimated errors with finite error bounds within relative error bounds of* $0.25, 0.5, 0.75$*, and* 1*, respectively. A relative error bound of* 0.5 *implies that the true geodesic distance is at least half and at most* 1.5 *times the estimated geodesic distance.*

6. Conclusion

In this paper we presented an algorithm to compute approximate geodesic distances on triangular manifolds with partially missing data. The geodesic distance approximations have worst case optimal error bounds. The geodesic approximation algorithm was shown to be feasible for large data sets with hundreds of thousands of vertices. The experimental results illustrate that the relative error bounds are below 0.25 for most approximations even for surfaces with significant amounts of missing data.

Furthermore, a new method was derived to add an object to the embedding computed via LSMDS. This method is generally applicable to problems involving LSMDS.

References

[BBK03] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Expression-invariant 3d face recognition. In *AVBPA* (2003).

[BBK05] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Three-dimensional face recognition. *IJCV 64*, 1 (2005), 5–30.

[BBK06a] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *PNAS 103*, 5 (2006), 1168–1172.

[BBK06b] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Robust expression-invariant face recognition from partially missing data. In *ECCV* (2006).

[BBKY06] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R., YAVNEH I.: Multigrid multidimensional scaling. *NLAA, Special issue on multigrid methods 13*, 2–3 (2006), 149–171.

[BG97] BORG I., GROENEN P.: *Modern Multidimensional Scaling Theory and Applications*. Springer, 1997.

[CC01] COX T., COX M.: *Multidimensional Scaling, Second Edition*. Chapman & Hall CRC, 2001.

[Dij59] DIJKSTRA E. W.: A note on two problems in connexion with graphs. *Numerische Mathematik 1* (1959), 269–271.

[EK03] ELAD A., KIMMEL R.: On bending invariant signatures for surfaces. *PAMI 25*, 10 (2003), 1285–1295.

[ELPZ97] ELDAR Y., LINDENBAUM M., PORAT M., ZEEVI Y.: The farthest point strategy for progressive image sampling. *TIP 6*, 9 (1997), 1305 – 1315.

[Gow66] GOWER J. C.: Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53* (1966), 325–338.

[Gow68] GOWER J. C.: Adding a point to vector diagrams in multivariate analysis. *Biometrika 55*, 3 (1968), 582–585.

[HSKK01] HILAGA M., SHINAGAWA Y., KOHMURA T., KUNII T. L.: Topology matching for fully automatic similarity estimation of 3d shapes. In *SIGGRAPH* (2001).

[JZ] JAIN V., ZHANG H.: A spectral approach to shape-based retrieval of articulated 3d models. *CAD to appear*.

[JZvK] JAIN V., ZHANG H., VAN KAICK O.: Non-rigid spectral correspondence of triangle meshes. *IJSM to appear*.

[KS98] KIMMEL R., SETHIAN J.: Computing geodesic paths on manifolds. *PNAS 95*, 15 (1998), 8431–8435.

[MD03] MOENNING C., DODGSON N.: *Fast Marching farthest point sampling*. Tech. Rep. 562, University of Cambridge, Computer Laboratory, 2003.

[MMP87] MITCHELL J., MOUNT D., PAPADIMITRIOU C.: The discrete geodesic problem. *SIAM J. Comput. 16* (1987), 647–668.

[PC05] PEYRE G., COHEN L.: Geodesic computations for fast and accurate surface remeshing and parameterization. *Progr. Nonlinear Differential Equations Appl. 63* (2005), 157–171.

[RBBK06] ROSMAN G., BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R.: Topologically constrained isometric embedding. In *MLPR* (2006).

[RDP99] ROBINETTE K., DAANEN H., PAQUET E.: The caesar project: A 3-d surface anthropometry survey. In *3DIM* (1999).

[Set96] SETHIAN J.: A fast marching level set method for monotonically advancing fronts. *PNAS 93*, 4 (1996), 1591–1595.

[SSK*05] 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.

[YHHC00] YAHIA H., HUOT E., HERLIN I., COHEN I.: Geodesic distance evolution of surfaces: A new method for matching surfaces. In *CVPR* (2000).

[ZKK02] ZIGELMAN G., KIMMEL R., KIRYATI N.: Texture mapping using surface flattening via multi-dimensional scaling. *TVCG 8*, 2 (2002), 198–207.

[ZSGS04] ZHOU K., SNYDER J., GUO B., SHUM H.-Y.: Iso-charts: Stretch-driven mesh parameterization using spectral analysis. In *SGP* (2004).