Shape- and Pose-Invariant Correspondences using Probabilistic Geodesic Surface Embedding

Aggeliki Tsoli and Michael J. Black

Brown University, Providence, RI, USA and
Max Planck Institute for Intelligent Systems, Tübingen, Germany
aggeliki@cs.brown.edu, black@is.mpg.de

Abstract. Correspondence between non-rigid deformable 3D objects provides a foundation for object matching and retrieval, recognition, and 3D alignment. Establishing 3D correspondence is challenging when there are non-rigid deformations or articulations between instances of a class. We present a method for automatically finding such correspondences that deals with significant variations in *pose*, *shape* and *resolution* between pairs of objects. We represent objects as triangular meshes and consider normalized geodesic distances as representing their intrinsic characteristics. Geodesic distances are invariant to pose variations and nearly invariant to shape variations when properly normalized. The proposed method registers two objects by optimizing a joint probabilistic model over a subset of vertex pairs between the objects. The model enforces preservation of geodesic distances between corresponding vertex pairs and inference is performed using loopy belief propagation in a hierarchical scheme. Additionally our method prefers solutions in which local shape information is consistent at matching vertices. We quantitatively evaluate our method and show that is is more accurate than a state of the art method.

1 Introduction

Finding correspondences between non-rigid 3D deformable objects is a critical task for many applications. Examples include object recognition and retrieval, shape deformation and morphing, 3D surface registration, *etc.* By defining correspondences using a structure preservation criterion, we can assess the similarity between two objects based on the amount of structure distortion. For applications involving search for similar 3D object models, it may be critical to have a measure of similarity that is invariant to common variations within a class (e.g. body pose and identity variation). Additionally, mesh alignment, for example of laser scans of human bodies, typically employs surface registration methods like ICP [3], [15] which require an initial set of correspondences. Here we describe a fully automated method for obtaining such correspondences between meshes that vary in shape, pose, and resolution.

Although the problem of establishing correspondences among rigid objects has been addressed in the literature adequately, finding correspondences between

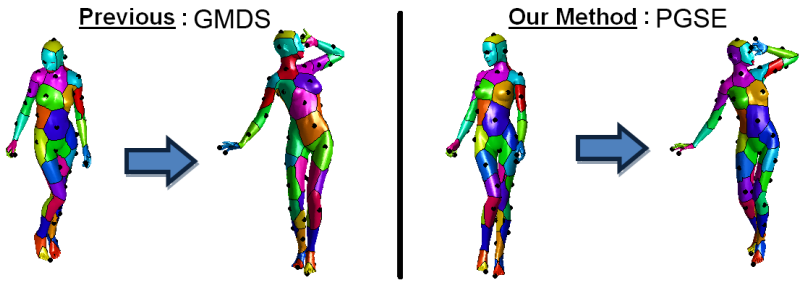


Fig. 1. Local optima in a combinatorial optimization problem for matching objects varying in pose and shape. Previous work, Generalized Multi-Dimensional Scaling (GMDS) [5], relies only on the preservation of geodesic distances and can yield non-meaningful correspondences; e.g. the chest of the body in the left pose is mapped to the back of the body in the right pose (corresponding regions are shown with the same color). Our method, Probabilistic Geodesic Surface Embedding (PGSE), achieves more intuitive results by combining geodesic distances with local surface descriptors in a coarse-to-fine probabilistic optimization framework.

non-rigid deformable objects is still a challenge. Variations in pose and shape change the local geometry of the object's surface increasing the likelihood of a false match. In addition, matching two objects entails solving a combinatorial problem in the exponential space of possible pairwise correspondences. Such an optimization may get stuck in local optima resulting in non-meaningful correspondences. Figure 1 shows an example of non-meaningful correspondences produced by related work, Generalized Multi-Dimensional Scaling (GMDS) [5], where the chest is mapped to the back of the human model and vice versa. This effect is significantly diminished using our method, Probabilistic Geodesic Surface Embedding (PGSE).

Previous methods for matching nonrigid deformable objects with significant variation in pose aim at providing global consistency of correspondences by preserving intrinsic properties of the objects. Usually these methods find deformation-invariant representations of the objects and match the objects in the representation domain. Examples include the use of geodesic distances [5], diffusion distances [6] or representations in the Möbius domain [13].

Although preservation of the intrinsic properties of the objects may be sufficient to assess their similarity, intrinsic-only matching criteria are oblivious to object self-symmetries and may yield non-meaningful correspondences. To overcome this weakness, previous work has explored the use of local surface properties and/or costs of surface deformation. Previous local surfaces properties are either geometric or based on the intrinsic characteristics of the shape or both. For instance, the work in [2] uses oriented histograms describing the distribution of points in local neighborhoods along the object surface (spin images [11]). Dubrovina *et al.* [7] use a local surface descriptor based on the eigenvalues of the Laplace-Beltrami operator which is related to the flow in the mesh representation of the object. Wang *et al.* [16] use descriptors based on curvature

and surface normals targeted towards a specific class of surfaces (brain surfaces). Efforts that also take into account object deformation include [10], [18].

Most previous work considers pose variations of the same object. To the best of our knowledge, only the work in [18] considers variations in shape, but the objects to be matched do not have significant differences in pose. We are concerned with finding correspondences among objects of the same category varying in shape, pose, and resolution. Extending previous approaches for global matching, we rely on preserving normalized geodesic distances to account for the additional variation in shape. We also employ a probabilistic framework for optimization similar to the one in [2]. We enforce stricter geodesic preservation constraints and use alternative local surface descriptors that are invariant to shape, pose, and resolution variations.

Our main contributions can be summarized as follows:

- A method for finding surface point correspondences of a non-rigid object undergoing significant deformation due to *pose* and *shape* variation.
- A method for finding surface point correspondences between objects differing in global/local *resolution* and triangulation, containing up to a small proportion of holes.
- Correspondence search that effectively explores the space of possible correspondences and is more robust to local optima than previous work. It relies on a discriminative probabilistic model that preserves properties related to geodesic distances and uses loopy belief propagation (LBP) for inference.

2 Probabilistic Geodesic Surface Embedding

We consider the problem of finding correspondences between two triangular meshes, a model mesh X and a data mesh Z. The *model mesh* $X = (V^X, E^X)$ is a complete surface consisting of a set of vertices $V^X = (x_1, \ldots, x_{N^X})$ and a set of edges E^X. The *data mesh* $Z = (V^Z, E^Z)$ may contain a modest number of holes (missing data); the vertices and edges are $V^Z = (z_1, \ldots, z_{N^Z})$ and E^Z respectively. Typically the data and model meshes differ in shape, pose, and resolution. Each data mesh vertex $z_k, k = 1, \ldots, N^Z$ is associated with a *correspondence variable* $c_k \in \{1, \ldots, N^X\}$ that specifies the model mesh vertex it corresponds to. The task of finding correspondences is one of estimating the most likely set of all correspondence variables $C = (c_1, \ldots, c_{N^Z})$ given a specific pair of model and data meshes X, Z.

2.1 Probabilistic Model

We cast the problem of finding correspondences as one of finding the most likely embedding of the data mesh Z into the model mesh X encoded as an assignment to all correspondence variables $C = (c_1, \ldots, c_{N^Z})$. More specifically we take a discriminative approach where our goal is to find a configuration of C that maximizes the distribution $p(C|X, Z)$ over all correspondence variables conditioned

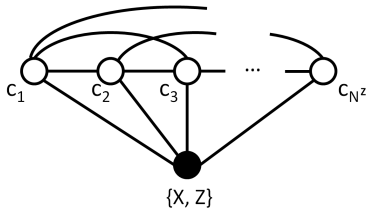


Fig. 2. Conditional Random Field (CRF) model for finding correspondences. The observed variable in the model is a pair of a model mesh $X = (V^X, E^X)$ and a data mesh $Z = (V^Z, E^Z)$. The latent variables are the correspondence variables $C = (c_1, \ldots, c_{N^Z})$ of all data mesh vertices. Edges in the model between latent and observed variables favor correspondences that preserve the intrinsic properties of the data mesh vertices. Geodesic constraints between all possible pairs of correspondence variables are enforced through the edges between latent variables.

on a pair of mesh instances X, Z. Writing this distribution as an undirected graphical model, we get the Conditional Random Field (CRF) model depicted in Figure 2. Each latent variable node in the model denotes the correspondence variable c_k of vertex z_k, $k = 1, \ldots, N^Z$, in the data mesh. The observed variable is a pair of model and data meshes X, Z.

We *approximate* the conditional distribution of the correspondence variables using potential functions, ψ, linking all *pairs* of latent variables and unary potentials, ϕ, linking each latent variable with the data. Formally we approximate the conditional distribution as: $p(C|X, Z) \propto \prod_k \phi(c_k, X, Z) \prod_{k,l} \psi(c_k, c_l, X, Z)$.

The main idea behind our approach is that the geodesic distances between points in the data mesh Z should be the same as the geodesic distances between the corresponding points in the model mesh X. Our method searches for correspondences that satisfy this property. At the same time we want to preserve in the embedding the intrinsic geodesic properties (geodesic signature) of the data mesh vertices. All the abovementioned constraints are enforced using the potentials described below.

Pairwise geodesic potential $\psi(c_k, c_l, X, Z)$: We consider normalized geodesic distances as the invariant used to match meshes that deform non-rigidly due to changes in shape and pose. We calculate exact geodesic distances using the the Fast Marching method described in [12]. For each pair of data mesh vertices z_k, z_l, we define a potential function $\psi(c_k, c_l, X, Z)$ that constrains the pair of correspondences c_k, c_l in the model mesh X to be geodesically consistent with vertices z_k, z_l in the data mesh Z. Let $M = (V, E)$ be a mesh with vertices V and edges E and $h : V \times V \rightarrow \Re$ be a geodesic distance function. Then $h(j, m; M)$ represents the normalized geodesic distance between two vertices j and m in mesh M. The normalization is done by dividing the geodesic distance by the maximum geodesic distance over all pairs of vertices in M. The geodesic potential between a pair of data mesh vertices z_k, z_l is defined as

$$\psi(c_k, c_l, X, Z) = N(h(c_k, c_l; X); h(k, l; Z), \sigma_{kl}^2) \tag{1}$$

where σ_{kl} is a user defined parameter; here $\sigma_{kl} = 0.1 \cdot h(k, l; Z)$.

Geodesic signature potential $\phi(c_k, X, Z)$: We encode a potential that enforces that corresponding vertices c_k in the model mesh have similar intrinsic properties as those in the data mesh z_k. Our goal is to distinguish spatially different areas in the model and data meshes as much as possible. The intrinsic property we use is the mean normalized geodesic distance of vertex z_k over all possible vertices in the data mesh (geodesic signature). The resulting potential can be written as

$$\phi(c_k, X, Z) = N(g(c_k; X); g(k; Z), \sigma_k^2) \tag{2}$$

where $g(j; M = (V, E)) = \frac{1}{|V|} \sum_{m \in V} h(j, m; M)$ is the mean normalized geodesic distance from j to all other vertices m in the mesh M and σ_k is a user defined parameter. The use of geodesic signatures is important because it biases the embedding of the data mesh to the model mesh to match spatially similar areas between the meshes. In practice we observe that this also improves convergence of the optimization procedure described below.

2.2 Inference

Our goal is to find an assignment of the correspondence variables that maximizes the probability $p(C|X, Z)$ as represented by the graphical model. Exact inference is computationally infeasible due to the large number of variables and loops in the graph. Instead we use max-product *loopy belief propagation* (LBP) [17] for approximate inference. Running LBP until convergence yields a set of probabilities over model mesh vertices for each correspondence variable c_k. We compute the optimal correspondence for each data mesh vertex z_k as the model mesh vertex that maximizes the probability distribution of the correspondence variable c_k.

Our inference scheme is performed in two rounds as shown in Figure 3. In the first round, the data mesh is sampled at a coarse level (Figure 3 (a)) using the farthest point sampling method [8]. In a similar way, the model mesh is sampled at a coarse level (Figure 3 (b)) and an initial set of correspondences is obtained using LBP. In the second round, the initial correspondences are refined by restricting the domain for each correspondence variable to be geodesically close to the solution of the first round of inference (Figure 3 (d)). Here we restrict the search to vertices with a geodesic distance up to 1/2 the average geodesic distance between nearby samples in the model mesh. The complexity of each round is $O(K^2L^2)$ where K is the number of samples in the data mesh and L the number of corresponding samples in the model mesh.

3 Results

3.1 Data

We evaluate our algorithm on triangular meshes from the TOSCA nonrigid world database [4] and human bodies generated using the SCAPE model [1]. All the

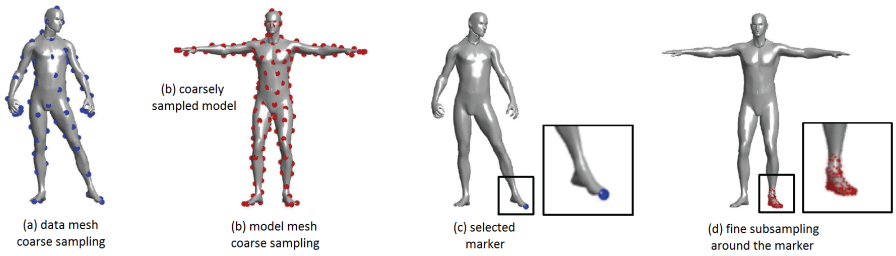


Fig. 3. Illustration of the sampling process during the inference procedure. In the first round, a data mesh and a model mesh are sampled at a coarse level. A coarse sampling of the data mesh to e.g. 75 markers (a) and a coarse sampling of the model mesh to e.g. 150 samples (b) produce an initial set of correspondences. In the second round, for each individual marker in the data mesh (c), the domain of possible correspondences is obtained from finer sampling around the solution found in the first round (d)

objects are represented as triangular meshes and they are simplified to have 2000-4000 vertices to aid comparison with related work. For each pair of meshes we find correspondences of 75-100 surface points. For the following experiments our method requires around 5GB of RAM per pair of meshes. The running time is approximately 1h on a 2.66GHz Intel Xeon processor.

3.2 Evaluation

The meshes we use do not come with any ground truth information about correspondences between their vertices. Typical error metrics in this case measure the degree that geodesic distances are preserved between the data mesh and the model mesh. However, preservation of geodesic distances does not ensure that the correspondences are qualitatively meaningful. The smaller the number of markers used and the larger the number of self-symmetries in the object, the larger the number of possible correspondence configurations with geodesic distances similar to the geodesic distances between data mesh markers. We find that comparing Voronoi regions around the markers and their optimal correspondences provides a more intuitive measure than comparing the degree in which geodesic distances have been preserved. Similar Voronoi regions between the data and model meshes also lead to similar geodesic distances among markers and their optimal correspondences. The opposite is not necessarily true. Comparing Voronoi regions does not only include how well the geodesic distances are preserved, but also how similar the neighborhoods around markers and their optimal correspondences are.

Let $v_s(i)$ be the area of the Voronoi region around marker i and $v_m(c_i^*)$ the area of the Voronoi region around the optimal correspondence c_i^* of marker i in the model. We define the following error metric, T_e, representing the average change in the Voronoi area over all markers and their correspondences.

$$T_e = \frac{1}{|U|} \sum_{i \in U} \left| \frac{v_s(i) - v_m(c_i^*)}{v_s(i)} \right| \tag{3}$$

where U is the set of markers in the data mesh.

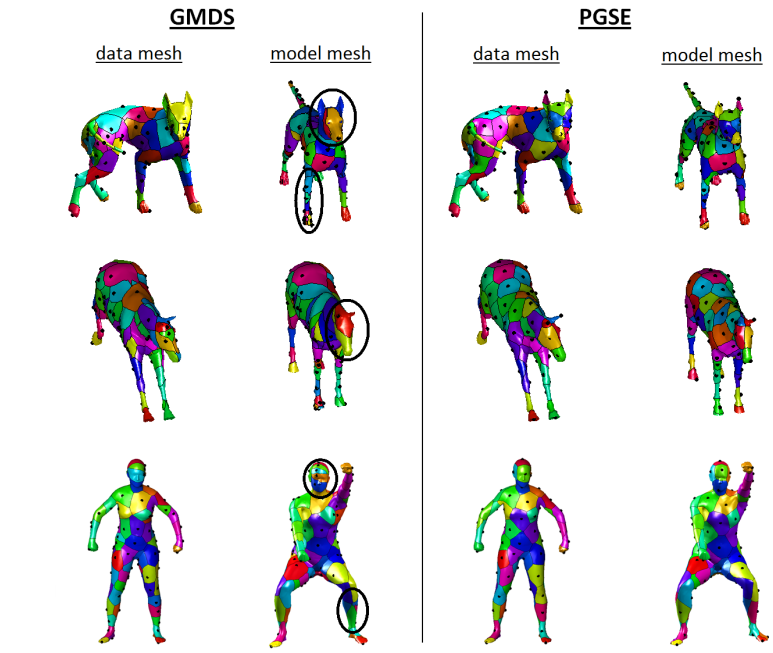
Correspondences in meshes with same topology. We compare our method, PGSE, to the GMDS method presented in [5] using triangular meshes of the same topology. For each object in the TOSCA nonrigid world database, we find correspondences to the canonical object of the category it belongs to. For the SCAPE bodies, we find correspondences between the mean SCAPE body in the canonical pose as defined in the CAESAR dataset [14] and SCAPE bodies varying in pose, shape, and pose and shape together.

Figure 4(a) illustrates the correspondences found with GMDS and PGSE. Evaluating the correspondences using the error metric defined above, we get the error plots shown in Figures 4(b, c, d). For the parameterized bodies generated using the SCAPE model, we sort the results based on pose or shape variation. Pose variation is measured as the average joint angle deviation from the joint angle configuration in the canonical pose. It is weighted by the percentage of mesh vertices each joint controls and it is measured in radians. Shape variation is measured based on the L2-norm of the shape coefficients in the SCAPE model. Given the variety of categories in the TOSCA nonrigid world database, we present only summary statistics of the error over the database. For the case of PGSE, the average T_e error is 0.1410 with standard deviation 0.1059. For the case of GMDS, the average T_e error is 0.2799 with standard deviation 0.1564.

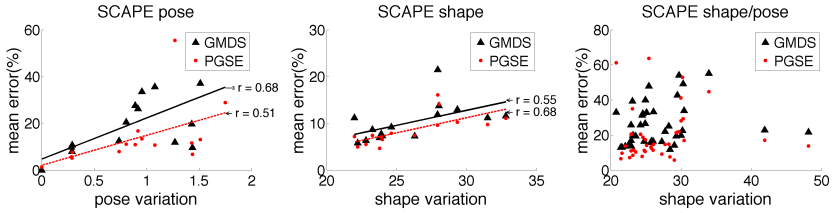
In all cases we see that the error increases as we vary the pose or the shape. Although not reported with error metrics, GMDS performs better on average at preserving geodesic distances; this is not surprising as the GMDS method minimizes exactly this error. In contrast, our method combines the preservation of geodesic distances with local shape matching constraints. Our approach, PGSE, performs better in terms of the maximum discrepancy in geodesic distances between pairs of markers and their correspondences. Evaluating the correspondences using the T_e error (Figure 4), we see that PGSE performs better in all cases. Statistical significance values for the errors per dataset are shown in Table 4(e). Changing the pose yields a bigger increase in the mean error than changing the shape. Changing both shape and pose yields the biggest increase in error as expected.

Correspondences in meshes with different topology. Next we evaluate the effects of changing the global and local resolution of the triangulated meshes used above. We use QSLIM [9] to change the global resolution of the meshes generated based on the SCAPE model and we observe an almost uniform reduction in resolution across the surface of the SCAPE bodies. In this case, we find no significant difference in performance between GMDS and PGSE as a function of mesh resolution.

Often one wants to align an artist-generated template mesh with higher-resolution meshes created by a laser scanner or other structured light system. In this case the meshes have very different topology and resolution. Consequently we find correspondences between the SCAPE bodies varying in shape and pose as above and a custom made template mesh shown as the right mesh in Figure



(a)

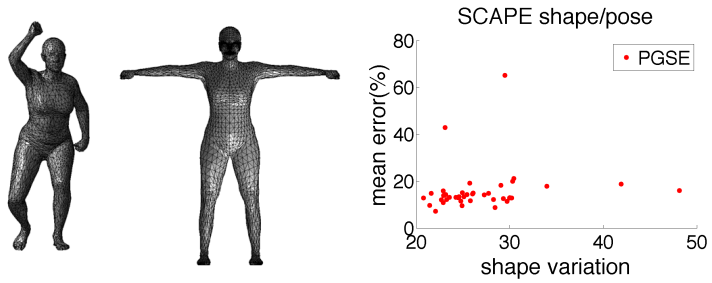


(b) (c) (d)

	TOSCA nonrigid 3D world	SCAPE pose	SCAPE shape	SCAPE shape/pose
T_e	$3 \cdot 10^{-16}$	0.013	$3 \cdot 10^{-4}$	$3.15 \cdot 10^{-4}$

(e)

Fig. 4. (a) Visual correspondences between meshes in the TOSCA nonrigid world database and SCAPE bodies varying in pose and/or shape. Corresponding areas are shown with the same color. Areas where our method, PGSE, performs better than GMDS are circled. Note that correspondences are defined up to intrinsic symmetries in the meshes. (b) Mean Voronoi error plot for the SCAPE bodies varying in pose, (c) shape, and (d) pose and shape. The data points in figures (c,d) are ordered based on shape variation. Table (e) shows the results of the Wilcoxon signed rank test on the errors induced by the GMDS, PGSE correspondences. All the p-values displayed in the table are below the default significance level of 5%.



(a) (b)

Fig. 5. (a) An example pair of meshes with significant differences in local resolution and mesh topology: a SCAPE body and our template mesh. (b) Mean Voronoi-based error for correspondences between the SCAPE bodies varying in shape & pose and the template. To simplify visualization the SCAPE bodies are ordered only based on shape variation. A Voronoi-based error cannot be defined for the case of GMDS due to markers collapsing at the same vertex.

5 (a). This template mesh exhibits significant differences in local resolution and topology compared with the SCAPE bodies. We are unable to quantitatively evaluate GMDS because in most cases the markers collapse to the same vertex on the data mesh surface resulting in Voronoi regions with zero area. In contrast, we observe that even large differences in local resolution between the surface of the data and model meshes does not influence the performance of our algorithm (the error in Figure 5 (b) is similar to the error in Figure 4 (d)) .

4 Conclusions

We present a method that finds correspondences between non-rigid articulated objects varying in pose, shape, and global or local resolution. Our method preserves pairwise normalized geodesic distances between a pair of objects as well as local surface properties also based on geodesic distances. We show improved correspondence over previous work on widely varying mesh models. Additionally using the SCAPE model we are able to separately evaluate accuracy as a function of pose, shape, and resolution variation. We also define a Voronoi-based error measure that better measures correspondences that are intuitively "good." Future work involves making our method robust to noisy surfaces as well as surfaces with missing information. Learning the parameters of our CRF model from training data is another direction for future work.

Acknowledgements. This work was supported by the Office of Naval Research under contract W911QY-10-C-0172.

References

1. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. ACM Transactions on Graphics (TOG) 24(3), 408–416 (2005)

2. Anguelov, D., Srinivasan, P., Pang, H., Koller, D., Thrun, S., Davis, J.: The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In: Advances in neural information processing systems 17: proceedings of the 2004 conference. p. 33. The MIT Press (2005)
3. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Transactions on pattern analysis and machine intelligence pp. 239–256 (1992)
4. Bronstein, A., Bronstein, M., Bronstein, M., Kimmel, R.: Numerical geometry of non-rigid shapes. Springer-Verlag New York Inc (2008)
5. Bronstein, A., Bronstein, M., Kimmel, R.: Generalized Multidimensional Scaling: a framework for isometry-invariant partial surface matching. Proceedings of the National Academy of Sciences of the United States of America 103(5), 1168 (2006)
6. Bronstein, A., Bronstein, M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. International Journal of Computer Vision 89(2), 266–286 (2010)
7. Dubrovina, A., Kimmel, R.: Matching shapes by eigendecomposition of the Laplace-Beltrami operator. In: Proc. 3DPVT. vol. 2 (2010)
8. Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.: The farthest point strategy for progressive image sampling. Image Processing, IEEE Transactions on 6(9), 1305–1315 (2002)
9. Garland, M., Heckbert, P.: Surface simplification using quadric error metrics. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. pp. 209–216. ACM Press/Addison-Wesley Publishing Co. (1997)
10. Huang, Q., Adams, B., Wicke, M., Guibas, L.: Non-rigid registration under isometric deformations. In: Proceedings of the Symposium on Geometry Processing. pp. 1449–1457. Eurographics Association (2008)
11. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(5), 433–449 (2002)
12. Kimmel, R., Sethian, J.: Computing geodesic paths on manifolds. Proceedings of the National Academy of Sciences of the United States of America 95(15), 8431 (1998)
13. Lipman, Y., Funkhouser, T.: Möbius voting for surface correspondence. ACM Transactions on Graphics (TOG) 28(3), 1–12 (2009)
14. Robinette, K., Daanen, H., Paquet, E.: The caesar project: a 3-d surface anthropometry survey. In: 3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on. pp. 380–386. IEEE (1999)
15. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: 3dim. p. 145. Published by the IEEE Computer Society (2001)
16. Wang, Y., Peterson, B., Staib, L.: 3d brain surface matching based on geodesics and local geometry. Computer Vision and Image Understanding 89(2-3), 252–271 (2003)
17. Yedidia, J., Freeman, W., Weiss, Y.: Understanding Belief Propagation and its generalizations. Exploring artificial intelligence in the new millennium 8, 236–239 (2003)
18. Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., Van Kaick, O., Tagliasacchi, A.: Deformation-Driven Shape Correspondence. In: Computer Graphics Forum. vol. 27, pp. 1431–1439. Wiley Online Library (2008)