Comparison of 3D Algorithms for Non-rigid Motion and Correspondence Estimation*

Pavel Laskov and Chandra Kambhamettu
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716, USA
{laskov,chandra}@cis.udel.edu
http://www.cis.udel.edu/∼vims

Abstract

We address the problem of non-rigid motion and correspondence estimation in 3D images in the absense of prior domain information. A generic framework is utilized in which a solution is approached by hypothesizing correspondence and evaluting the motion models constructed under each hypothesis. We present and evaluate experimentally five algorithms that can be used in this approach. Our experiments were carried out on synthetic and real data with ground truth correspondence information.

1 Introduction

Experimental comparison of algorithms for non-rigid motion and correspondence estimation is highly important. A vast amount of relevant work published in the last decade builds on heterogeneous ideas, yet no single algorithm is known to provide a robust solution under a variety of conditions. In this paper we attempt to cross-investigate a number of algorithms – some well-established ones as well as some promising recent approaches – with the aim of identifying the ideas that may lead to major improvements of current methods of non-rigid motion analysis.

Why is it desirable to further improve current motion analysis techniques? The answer is: some developing applications impose much greater requirements for motion analysis than what current methods are capable of. While specific application-dependent techniques, e.g., left-ventricular surface motion tracking [10] or cerebral cortical surface correspondence estimation [13], have proved to be very successful, in general very little is known about how to robustly recover unrestricted non-rigid motion from observations. The potential benefits of such knowledge are manifold. It may further help the analysis of 3D biomedical images by quantifying what is currently perceived only visually as a motion field between corresponding points. It may facilitate the segmentation of multiple motion by filtering out

*Funding for this work was provided under NSF grants CAREER IRI-9984842 and CISE CDA-9703088.

motions with distinct characteristics. Another interesting potential application is decreasing the bandwidth in transmission of dynamic image sequences: if compact representation of the motion between successive images could be found, only this component would need to be transmitted instead of full images.

The goal of a robust non-rigid motion estimation algorithm can be seen as the following: *in the absence of any prior information other than 3D images before and after motion, recover some meaningful compact representation of the observed motion.* Let us point out the three essential requirements of this scenario:

1. Correspondence between points, or other features, in images is assumed unknown. As part of its job, the algorithm must recover the correspondence, but it is not the only objective of the algorithm.

2. No prior shape information, nor any information about the physical properties, is available.

3. The algorithm must not be limited to specific points in objects with some favorable properties.

The problem of unknown correspondence lies at the heart of non-rigid motion estimation. In some cases, it may be decoupled from the motion estimation, in that the results of an algorithm providing only correspondence can be later used by a motion estimation algorithm that assumes known correspondence. For this reason we also consider the "correspondence only" algorithms in the current investigation. The second requirement prompts us to leave out the physically-based methods as well as the methods utilizing global shape topology. The rationale here is that using a model from an inappropriate domain may lead to erroneous model estimation, which would severely hamper motion analysis. Finally, the requirement for applicability to arbitrary points rules out the techniques that are looking for "feature points", such as points with high curvature, etc.

Comprehensive coverage of non-rigid motion estimation techniques can be found in two literature reviews published in the mid-90's [1, 9]. Experimental cross-evaluation of such techniques, to our knowledge, is the first of a kind. Due to space constraints and heterogeneity of existing algorithms we are only able to cover a small subset thereof. Nonetheless we hope that the findings of this work provide a useful insight in development of more advanced techniques.

2 Generic Framework for Non-rigid Motion and Correspondence Estimation

The abstract setting of the problem of non-rigid motion and correspondence estimation can be seen as follows: *given two sets of n-dimensional points:* $P = \{p_k\}$ *and* $P' = \{p'_k\}$*, for each point of interest* $p_i \in P$ *find the index* j *of the corresponding point* p'_j *and the motion function* $s(p)$ *such that* $p'_j = p_i + s(p_i)$. For 3D images, dimensionality of p is obviously 3. In the classical rigid registration problem correspondence is trivial, i.e., $i = j$, and the motion function $s(p) = Rp + t$ for some rotation matrix R and translation vector t. In non-rigid registration correspondence is continuous, and the motion function is either affine with full 12

degrees of freedom or non-linear, i.g., [11]. Even a non-linear function, however, may not be powerful enough to accurately model the transformation, or it may require prohibitively large time to compute. To cope with this predicament, a *local* motion function s_i may be sought for each point of interest.

It is known from Helmholtz mechanics that infinitesimal transformation can always be represented as a composition of translation, rotation and non-rigid transformation: $s(p) = Kp + t$, where $K = R + D$ or $K = RD$, D being a non-rigid transformation matrix. The recovery of R and D is possible, however, only for a symmetric (in the additive representation), or a positive-definite (in the multiplicative representation) matrix D [4].

In an attempt to overcome the limitations on the kind of deformation, the local topology of an object can be used. In order to do so, we make a simplifying assumption that the objects under investigation are surfaces in 3D. The parametric representation of a surface can be recovered by fitting a second degree polynomial [2, 8], and the differential-geometric information thus obtained can be used to estimate motion and correspondence.

Consider the point p_0 on S, the surface before motion. The problem of finding the counterpart of p_0 naturally precedes modeling the motion between the two. Thus the generic procedure of estimating correspondence and motion can be seen to have the following two steps:

1. Find the point p'_k on S' so as to minimize some correspondence error $E_c(p_0, p'_j)$ among some candidate points p'_j.

2. Recover the motion model between p_0 and p'_k.

One pair of corresponding points may not be enough to uniquely recover the motion. In this case one can consider small neighborhoods around the points of interest and model motion between these neighborhoods. The correspondence error function can also be formulated in terms of the neighborhoods in question.

Within this abstract framework, the differences between individual algorithms lie in the kind of the correspondence error function E_c and in the estimation of the motion model (if any) given the corresponding neighborhoods of interest.

3 Brief Review of Selected Algorithms

We now present a brief review of the algorithms selected in our study. Our goal here is to highlight the features of these algorithms which, in our opinion, have critical impact on their performance in our evaluation. Details can be found in the cited sources.

3.1 Iterative Closest Point Algorithm

The Iterative Closest Point (ICP) algorithm was developed by Besl and McKay for registration of 3D shapes [3]. It has become a popular tool in 3D registration and has received some extensions [5, 6].

To adapt the ICP algorithm to our generic framework we proceed as follows. Let P be the set of points in the neighborhood of the point of interest p_0 on S,

and let X be the set of points around a candidate point p'_j on S'.[1] We run the ICP algorithm on P and X until convergence (usually not more than 5 iterations), and use the resulting registration error as the value of $E_c(p_0, p'_j)$. The process is repeated for all other candidate correspondence points. The point with the smallest registration error is chosen as the correct correspondence, and the rigid transformation associated with this registration is chosen as the motion model.

Despite the use of the rigid model, the main advantage of the ICP algorithm is that instead of assuming a trivial correspondence it proceeds iteratively, by selecting the subset of the closest points, estimating motion, and applying the motion to the initial point set P at each iteration. This idea is similar to Ullman's incremental rigidity scheme [12] known for its ability to approximate small non-rigid deformations.

Since this algorithm has been used in a variety of applications it was interesting to investigate its performance in our framework.

3.2 Homothetic Motion Algorithm

A number of algorithms naturally fitting into our generic framework has been proposed for restricted classes of non-rigid motion[2]. The homothetic motion represents uniform expansion/contraction of the surface, characterized by the stretching coefficient θ. Since our neighborhoods of interest are, in general, small, homothetic motion may be a reasonably close approximation of the undergoing non-rigid motion. The optimal stretching coefficient can be computed from the values of the Gaussian curvature in the neighborhoods η and η' before and after motion [7]:

$$\theta = \sqrt{\frac{\sum_{i \in \eta, \eta'} K_i K'_i}{\sum_{i \in \eta'} (K'_i)^2}}.$$

Once the stretching coefficient is estimated, the correspondence error can be computed as:

$$E_c = \sum_{i \in \eta, \eta'} (K_i - \theta^2 K'_i)^2.$$

The candidate point with the smallest error is selected as the true correspondence, and the respective θ as the motion model.

3.3 Unit Normal Algorithm

Another differential-geometric algorithm is based on the relationship between the unit normals before and after small deformation[3]. Let n and n' denote the unit normals to the surface before and after deformation. Then the following relation is known from classical differential geometry [14]:

$$n' = n - n \times rot\ s, \tag{1}$$

[1] The notation of [3] is directly followed in this exposition. Our implementation uses the original quaternion formulation without acceleration.

[2] A more detailed discussion on this topic can be found in [9].

[3] The original version of this algorithm was proposed in [8]. We present here – without derivation – a modified, computationally more efficient version.

where $rot\ s$ denotes the curl of the motion function s. Assuming the affine motion function

$$s = au + bv + c,$$

one can define the error function as the discrepancy of Eq. (1). For the neighborhood of interest, this error can be shown to be:

$$\epsilon = \sum_{i \in \eta, \eta'} \left(\Delta n_i + \frac{1}{E_i}(n_i \cdot a)(r_u)_i + \frac{1}{G_i}(n_i \cdot a)(r_v)_i \right). \tag{2}$$

Minimizing the squared error $(\epsilon \cdot \epsilon)$ over the neighborhood around the point of interest, one can obtain the system of linear equations in which unknowns are the motion parameters a and b:

$$\begin{bmatrix} \sum_{i \in \eta} \frac{1}{E_i}\ n_i n_i^T & 0 \\ 0 & \sum_{i \in \eta} \frac{1}{G_i}\ n_i n_i^T \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = - \begin{bmatrix} \sum_{i \in \eta, \eta'} \frac{1}{E_i}\ (\Delta n_i \cdot (r_u)_i)\ n_i \\ \sum_{i \in \eta, \eta'} \frac{1}{G_i}\ (\Delta n_i \cdot (r_v)_i)\ n_i \end{bmatrix}. \tag{3}$$

The unit normal algorithm proceeds as follows:

1. Fit second degree polynomials $z(x, y)$ to 3D data tuples (x_i, y_i, z_i) in the neighborhoods of interest before and after motion.
2. By letting $u \equiv x$, $v \equiv y$, construct the parametric representation $r(u, v) \equiv (u, v, z(u, v))$. In this representation compute the partial derivatives $r_u \equiv (1, 0, z_u)$ and $r_v \equiv (0, 1, z_v)$, unit normals $n = r_u \times r_v / |r_u \times r_v|$, and the coefficients of the first fundamental form $E = (r_u \cdot r_u)$, $G = (r_v \cdot r_v)$.
3. Decompose the left-hand side matrix using SVD or LU decomposition[4].
4. For each candidate correspondence point p'_j on the surface after motion, back-substitute the right-hand side vectors obtained from the neighborhood of p'_j (Δn is the only parameter that depends on the surface after motion). Evaluate the hypothesis by plugging the motion model (a_j, b_j) in Eq. (2).
5. Select the motion model and the correspondence with the smallest error.

3.4 Amini's Algorithm

Amini and Duncan addressed the problem of estimation of correspondences in the sequences of 3D images of left ventricle [2]. Their algorithm is inspired by the idealized thin-plate model, that can be adapted to modeling bending deformation, and by the conformal motion model corresponding to non-uniform stretching deformation. The combined bending/stretching energy is defined as

$$\begin{aligned} \epsilon_{ik} = {} & \lambda_{be}\{((\kappa_1)_i - (\kappa'_1)_k)^2 + ((\kappa_2)_i - (\kappa'_2)_k)^2\} \\ & + \lambda_{st} \left\{ \left(\frac{E_i}{E'_k} - \frac{F_i}{F'_k} \right)^2 + \left(\frac{E_i}{E'_k} - \frac{G_i}{G'_k} \right)^2 + \left(\frac{F_i}{F'_k} - \frac{G_i}{G'_k} \right)^2 \right\}, \end{aligned} \tag{4}$$

where κ_1 and κ_2 are the principal curvatures, E, F, G are coefficients of the first fundamental form, λ_{be} and λ_{st} are non-negative constants.

[4]Notice that the expression $n_i n_i^T$ denotes an open product of 3D vectors which is a 3×3 matrix block.

For the fixed correspondence hypothesis j the error is computed by minimizing the energy over the neighborhood of p'_j and summing over the neighborhood of p_0:

$$E_c = \sum_{i \in \eta} \min_{k \in \eta'_j} \epsilon_{ik}. \tag{5}$$

The hypothesis minimizing the error (5) is chosen as correct correspondence.

3.5 Wang's Algorithm

The algorithm of Wang et al. [13] was proposed for 3D surface correspondence, with application to matching of brain images. The algorithm combines Euclidean distance information with the differential-geometric information. The latter, unlike some of the previous methods that require 3D data, is estimated from 2D images.

The error function utilized by Wang's algorithm is the following:

$$E_c = d_{ij} \cdot n_{ij} \cdot f_{ij}. \tag{6}$$

The Euclidean distance measure d_{ij} is defined as

$$d_{ij} = 1 + \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}.$$

The unit normal match measure n_{ij} is defined as

$$n_{ij} = 2 - n_i \cdot n_j.$$

The feature match measure f_{ij} is a heuristic, curvature-based measure applicable only to brain images. To make this algorithm suitable for our generic framework, we have defined a similar measure based on Gaussian curvature:

$$f_{ij} = 1 + \log_{10}\left(\frac{|K_i - K'_j|}{|K_i|} + 1\right).$$

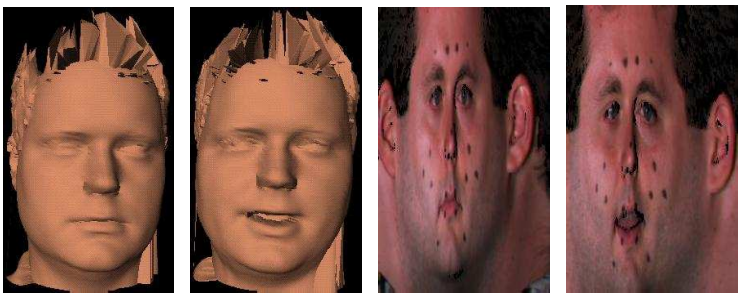
The diacritical feature of Wang's algorithm is the combination of three distinct modalities of correspondence measure in a single multiplicative error function.

4 Experimental Results

Our experimental setup consists of three tiers:

1. Artificial motion on analytical shapes. The shape is a 25×25 quadric $0.1u^2 + 0.1v^2$ with 49 points of interest evenly spaced with step 2 around its apex. Affine motion $au + bv$ with parameters $a = (0.001, 0.002, 0.003)$ and $b = (0.003, 0.001, 0.002)$ is applied after an initial rotation by 5° and translation by 5% of the magnitude of the (x, y, z) vector of the point of interest. The affine component is magnified with the parameter δ in the range of 1 to 1000.

2. Artificial motion on real shapes. The shape is a facial range image obtained from a Cyberware scanner. Motion is applied in the same way as above.



(a) (b) (c) (d)

Figure 1: Examples of 3D shapes with correspondence information: (a), (b) – 3D images before and after deformation; (c), (d) – visual images with correspondence information before and after deformation.

3. Real motion on real shapes. Motion between a pair of facial range images is estimated. The ground truth correspondence is obtained from the visual images of the same shape with color markers (invisible to the range scanner) applied at points of interest. The examples of 3D shapes and their visual images are given in Fig. 1.

In the artificial motion experiments the true correspondence is trivial. Given the computed correspondence (u_c, v_c), and the true correspondence $(\bar{u}, \bar{v})$, the *correspondence error* is defined as $\sqrt{(u_c - \bar{u})^2 - (v_c - \bar{v})^2}$.

One can see from Figures 2 and 3 that the algorithms exhibit different behavior with respect to the motion magnitude. The error of the homothetic and the unit normal algorithms increases with greater motion magnitude, whereas the other three algorithms are much less affected by it. One can also notice similarity of the accuracy of the ICP and Wang's algorithms – it turns out that the latter's error function is dominanted by the distance component. Amini's algorithm had persistent problems in both experiments. This may be due to imbalance of its error function consisting of two distinct terms representing different types of motion. The balance between the terms can vary even across the same data set, and we had to set both weights to 1 to avoid unjustified bias.

The real motion experiments were carried out on 7 data sets based on sequences of 3D images from three subjects. Data sets 2 and 4 represent relatively large motion. In our experiments the 180×524 real images were sub-sampled by 2, and the search window of size 7 was used.

Three numerical criteria were used for evaluation. The correspondence error was defined earlier in this section. The *relative image error* is computed as $\frac{||\bar{p}' - p'_c||}{||\bar{p}' - p_0||}$, where p_0 is the point of interest before motion, $\bar{p}'$ is the true corresponding point, and p'_c is the computed corresponding point. The *image error improvement ratio* is computed as $\frac{||\bar{p}' - p'_{1-1}|| - ||\bar{p}' - p'_c||}{||\bar{p}' - p'_{1-1}||}$, where p'_{1-1} is the point after motion under

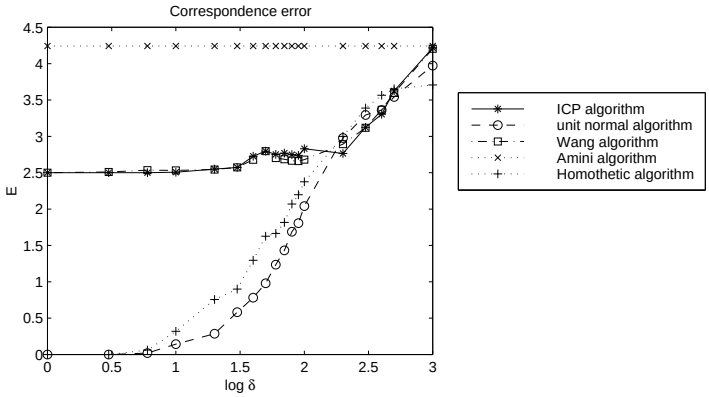


Figure 2: Comparison of the algorithms on the analytical shape and artificial motion.

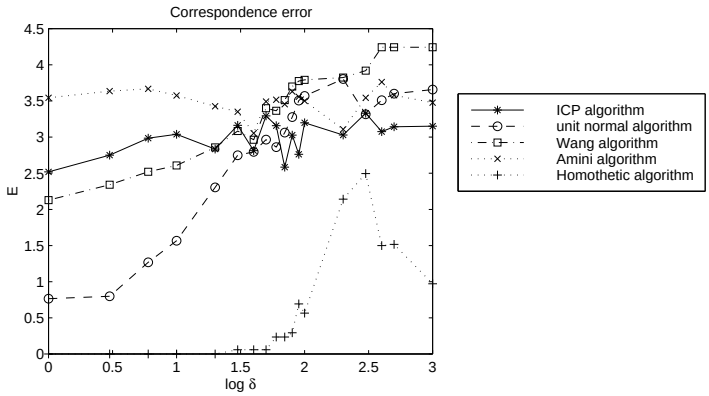


Figure 3: Comparison of the algorithms on the real shape and artificial motion.

trivial correspondence. It shows the relative improvement of error over trivial correspondence. This measure is only applicable when the true correspondence is *not* trivial; however, the latter occurs very rarely.

The results are presented in Figure 4. One can see that the unit normal algorithm is significantly more robust than the other 4 algorithms. On all but one dataset it exhibits the smallest correspondence error, and on all datasets the smallest relative image error and the largest improvement ratio. Also noteworthy is that none of the algorithms is even close to a sub-pixel correspondence error reported on several occasions in biomedical image registration literature. This signifies that the setup of the non-rigid motion/correspondence estimation for arbitrary points of interest is inherently more difficult.

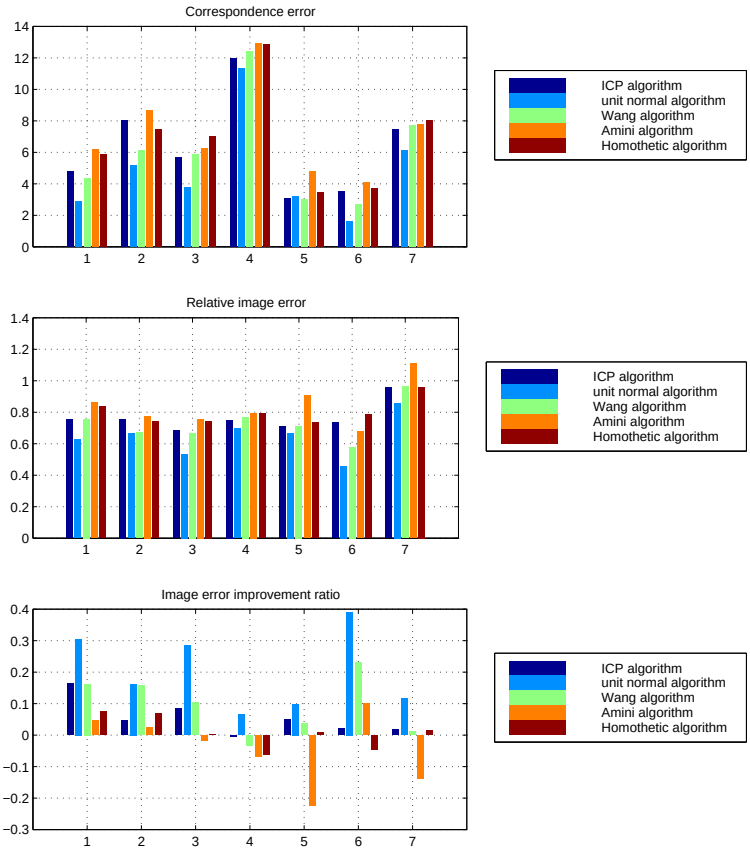


Figure 4: Comparison of the algorithms in the real motion, 7 data sets.

5 Conclusions

Among the five considered algorithms, the unit normal algorithm seems to be the most robust, especially on real data. Its rigorous differential-geometric background is appealing, although it remains to be seen how critical the small motion assumption is on real data. The relative strength of Wang's algorithm suggests that multiplicative error function is a good idea for combining characterization of different factors. The ICP algorithm can still be not the worst choice, although it is the slowest of the five.

Overall, the accuracy of current algorithms has much room for improvement. The observed image error constitutes at least 50% of the motion, and the improvement over the trivial correspondence is it at most 40%. Special effort must be made to perform intelligent hypothesis search: the simple window search used in our experiments has problems with large motion due to quadratic growth of the computational effort.

The future work we envision in this area would address more advanced methods

of motion estimation and building motion models more powerful than the affine. Another potential direction of research is incorporation of smoothness assumption.

References

[1] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: A review. In *Proceedings of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pages 16–22, 1994.

[2] A. A. Amini and J. S. Duncan. Bending and stretching models for LV wall motion analysis from curves and surfaces. *Image and Vision Computing*, 10(6):418–430, 1992.

[3] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb. 1992.

[4] S. Chaudhuri and S. Chatterjee. Motion analysis of a homogeneously deformable object using subset correspondences. *Pattern Recognition*, 24(8):739–745, 1991.

[5] J. Feldmar, G. Malandain, J. Declerck, and N. Ayache. Extension of the ICP algorithm to non rigid intensity-based registration of 3D volumes. *Computer Vision and Image Understanding*, 66(2):193–206, May 1997.

[6] A. E. Johnson and M. Hebert. Surface registration by matching oriented points. In *Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pages 121–128, 1997.

[7] C. Kambhamettu and D. B. Goldgof. Curvature-based approach to point correspondence recovery in conformal non-rigid motion. *Computer Vision, Graphics and Image Processing: Image Understanding*, 60(1):26–43, 1994.

[8] C. Kambhamettu, D. B. Goldgof, and M. He. Determination of motion parameters and estimation of point correspondences in small non-rigid deformations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 943–946, 1994.

[9] C. Kambhamettu, D. B. Goldgof, D. Terzopoulos, and T. S. Huang. *Handbook of Pattern Recognition and Image Processing: Computer Vision*, chapter 11: Nonrigid Motion Analysis, pages 405–430. Academic Press, Inc, 1994.

[10] P. Shi, A. J. Sinusas, R. T. Constable, E. Ritman, and J. S. Duncan. Point-tracked quantitative analysis of left-ventricular surface motion from 3D images sequences. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(1):36–50, Jan. 2000.

[11] R. Szeliski and S. Lavalleé. Matching 3D anatomical surfaces with nonrigid deformations using octree-splines. *International Journal of Computer Vision*, 18(2):171–186, May 1996.

[12] S. Ullman. Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, 13:255–274, 1984.

[13] Y. Wang, B. S. Peterson, and L. H. Staib. Shape-based 3D surface correspoondence using geodesics and local geometry. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 644–651, 2000.

[14] C. E. Weatherburn. *Differential geometry of three dimensions*. Cambridge University Press, 1927.