Geometric Constraint Analysis and Synthesis: Methods for Improving Shape-Based Registration Accuracy*

David A. Simon^{1,2} and Takeo Kanade¹

¹Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 ²Center for Orthopaedic Research, Shadyside Hospital Pittsburgh, PA 15232

Abstract. Shape-based registration is a process for estimating the transformation between two shape representations of an object. It is used in many image-guided surgical systems to establish a transformation between pre- and intra-operative coordinate systems. This paper describes several tools which are useful for improving the accuracy resulting from shape-based registration: constraint analysis, constraint synthesis, and online accuracy estimation. Constraint analysis provides a scalar measure of sensitivity which is well correlated with registration accuracy. This measure can be used as a criterion function by constraint synthesis, an optimization process which generates configurations of registration data which maximize expected accuracy. Online accuracy estimation uses a conventional root-mean-squared error measure coupled with constraint analysis to estimate an upper bound on true registration error. This paper demonstrates that registration accuracy can be significantly improved via application of these methods.

Keywords: geometric constraint analysis, geometric constraint synthesis, online accuracy estimation, shaped-based registration, intra-operative registration.

1 Introduction

The registration process is a fundamental component of most image-guided surgical systems. Registration estimates a spatial transformation between two coordinate systems: a pre-operative system used to construct plans or simulations based upon medical data (e.g., CT, MRI, or X-ray images), and an intra-operative system in which the surgical procedure is performed (e.g., relative to a robot, navigational guidance system, etc.) Any image-guided surgical procedure which spatially relates pre-operative data to intra-operative execution requires solution of the registration problem.

There are many approaches to registration for image-guided surgery and an excellent review can be found in [5]. A class of registration methods referred to as *shape-based* methods uses representations of object shape to estimate the required transformation. Representations are constructed using data collected in the two coordinate systems (i.e., pre- and intra-operative). Registration estimates a transformation which aligns one shape representation with the other in a manner which minimizes a measure of the distance between them.

Several factors affect shape-based registration accuracy, including: errors in the shape representations due to sensor noise or shape reconstruction errors [11]; the quantity of registration data; and the locations on the registration object from which the data are collected [10]. This paper addresses the problem of improving shape-based registration

^{*} This work was supported in part by a National Challenge grant from the NSF (IRI-9422734). Please address all correspondence to the first author at das@ri.cmu.edu.

accuracy via intelligent selection of registration data and online estimation of accuracy. Intelligent data selection (IDS) is comprised of *geometric constraint analysis* which provides a sensitivity measure shown to be well correlated with registration accuracy; and *geometric constraint synthesis*, an optimization process which generates data configurations which maximize the sensitivity measure for a fixed quantity of data. IDS uses the pre-operative shape representation to generate a *data collection plan* (DCP) which can be used during surgery to guide the acquisition of registration data. *Online accuracy estimation* provides an upper bound on true registration accuracy based upon a conventional root-mean-squared error.

The proposed methods have been investigated in-vitro on cadaveric specimens and via simulation studies and are currently being incorporated into a clinical image-guided orthopaedic surgical application [9]. The current paper describes the methods, reports encouraging results, and suggests approaches for incorporating the methods into clinically viable registration systems.

2 Methods

This paper focuses on a special case of shape-based registration: surface-based registration with discrete point data. One shape representation (the "Model") is a triangle mesh surface model of the registration object constructed from CT images. The other representation (the "Data") is a set of discrete point data collected from the registration object during surgery using a digitizing probe.

2.1 Constraint Analysis

Most approaches to shape-based registration attempt to minimize an error measure such as the following least-squared measure:

$$\min_{\boldsymbol{T}} \sum_{i} \|\boldsymbol{M}_{i} - \boldsymbol{T}(\boldsymbol{D}_{i})\|^{2} \tag{1}$$

where each D_i represents a point in the Data, each M_i represent a point in the Model, and T is a 3-D transformation which minimizes the expression. Details of shape-based registration methods can be found in [2][3][5], and descriptions of the methods used in this work appear in [8]. Due to fundamental similarities among shape-based registration solution methods, the techniques proposed in this paper are independent of the particular registration solution method used.

Solving the registration problem results in an estimate, T_{est} , of the true (and usually unknown) registration transformation, T_{true} . The error resulting from a single registration trial can be expressed as:

$$\boldsymbol{T}_{err} = \boldsymbol{T}_{true} \cdot \boldsymbol{T}_{est}^{-1} \tag{2}$$

where T_{err} is a transformation which represents the difference between estimated and true transformations. The goal of constraint analysis is to provide a scalar measure of sensitivity which is a good predictor of T_{err} for a given Model and Data without performing registration, and without the need to know T_{true} .

Derivation of the Method

The point-to-surface distance in (1) is defined as the length of the shortest line joining a point and a surface. In general, there is no closed form analytical expression for this

distance given an arbitrary surface; however, the following local approximation has been proposed [12]:

$$D(x) = \frac{F(x)}{\|\nabla F(x)\|} \tag{3}$$

where x is a point which may or may not lie on the surface, F(x) = 0 is the implicit equation of the surface, $\|\nabla F(x)\|$ is the magnitude of the gradient of F at x, and D(x) is the approximate distance. It can be shown that D(x) is a first order approximation of the true point-to-surface distance, and is exact when the surface is a plane.

Given a point x_s which lies on the surface, a small transformation, T_s , will perturb this point from its resting position. T_s can be represented by a homogeneous transformation which is a function of the 6 parameter vector,

$$t = \begin{bmatrix} t_x & t_y & t_z & \omega_x & \omega_y & \omega_z \end{bmatrix}^{\mathbf{T}}$$
 (4)

in which $(\omega_x, \omega_y, \omega_z)$ are rotations about the X, Y, and Z axes respectively, and (t_x, t_y, t_z) are translations along the newly rotated X, Y, and Z axes. The gradient of D with respect to t is a 6-vector,

$$V(x_s) = \frac{\partial}{\partial t} D(T_s(x_s)) = \begin{bmatrix} n \\ x_s \times n \end{bmatrix}$$
 (5)

where n is the unit normal to the surface at the point x_s [8]. This result can be extended to consider the effect of perturbing a *collection* of points with respect to the surface:

$$E_{P}(T_{s}(x_{s})) = dt^{T} \left[\sum_{x_{s} \in P} V(x_{s}) V^{T}(x_{s}) \right] dt$$
$$= dt^{T} \Psi_{P} dt$$
(6)

The scalar quantity $E_P(T_s(x_s))$ is a first order approximation of the least-squared error of (1). It measures the error which would result by perturbing a set of discrete points, P, initially assumed to be on the surface, by the small transformation T_s . The matrix Ψ_P is a symmetric, positive semi-definite 6x6 scatter matrix which contains information about the distribution of the original $V(x_s)$ vectors over the points in the set P. Performing principal component analysis [4], Ψ_P is transformed into an expression which is more easily interpreted:

$$E_{P}(\boldsymbol{T}_{s}(\boldsymbol{x}_{s})) = d\boldsymbol{t}^{\mathrm{T}} Q \Lambda Q^{\mathrm{T}} d\boldsymbol{t}$$

$$= \sum_{i=1}^{6} \lambda_{i} (d\boldsymbol{t}^{\mathrm{T}} \boldsymbol{q}_{i})^{2}$$
(7)

where $\Lambda = diag[\lambda_1 \dots \lambda_6]$ is a diagonal 6x6 matrix of the eigenvalues of Ψ_P in which $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \lambda_5 \geq \lambda_6$; Q is a 6x6 matrix whose columns are the eigenvectors of Ψ_P ; and each q_i is an eigenvector corresponding to the eigenvalue λ_i which represents a differential transformation 6-vector. This result is similar to one presented in the context of industrial inspection [6].

From (7) it can be seen that the eigenvector q_1 corresponding to the largest eigenvalue, represents the *direction of maximum constraint*. Perturbing the points in the set P in the

direction of q_1 will result in the largest possible change in E_P from among all possible directions of perturbation. Similarly, the differential transformation represented by the eigenvector q_6 corresponds to the *direction of maximum freedom*. Perturbing the points in this direction will result in the smallest possible change in E_P from among all possible directions of perturbation. In general, an eigenvalue λ_i is proportional to the rate of change of E_P induced by a differential transformation in the direction q_i .

A special situation occurs when some of the λ_i are close to or equal to zero. For each such eigenvalue, a singularity exists such that perturbing the points in the direction of the corresponding eigenvector will result in no change in E_P . Such singularities are undesirable in registration since it is impossible to localize the object in the direction(s) of the singularity(s). As demonstrated below, sets of discrete points, P, which have well-conditioned scatter matrices, Ψ_P , are preferable to sets which have ill-conditioned scatter matrices for achieving accurate registration. In this work, the noise amplification index (NAI) [7] is used as a measure of matrix conditioning and is defined as

$$\frac{\lambda_6}{\sqrt{\lambda_1}}$$
 (8)

This quantity is the product of the inverse condition number and the square root of the minimum eigenvalue, and provides an upper bound on the amplification of residual errors (e.g., discrete point Data measurement noise, and errors in the Model) to the estimated parameters (e.g., registration transformation parameters) [7].

Scale and Coordinate System Dependences

In constraint analysis, there is an implicit weighting factor related to object size which determines the relative importance of rotational versus translational errors. Due to the x_s term on the right hand side of (5), if constraint analysis is applied to two objects which differ only in size, the resulting NAI values will differ. The larger object will weight rotational components more heavily since the corresponding x_s terms will be larger. A solution to this problem is to pre-normalize the Model so that the average radius as measured about the origin is unity. This has the effect of weighting rotational and translational components equally, on average. A complete discussion of the scale dependence problem can be found in [8].

Constraint analysis has a dependence upon the location of the origin of the Model coordinate system arising from the x_s term on the right hand side of (5). For a given Model, it can be shown that maximal sensitivity of constraint analysis is achieved when the constraint analysis coordinate system origin is located at the centroid of the Model [8].

2.2 Constraint Synthesis

The goal of constraint synthesis is to automatically generate Data sets which maximize the NAI for a given Model and a fixed number of points. The resulting *data collection plan* (DCP) can then be used to guide the acquisition of Data during the intra-operative Data collection process. More formally, the constraint synthesis problem is to:

Select M discrete points from a set, V, and place them into the set, P, of (6) such that the NAI is maximized.

In this paper, the set *V* is composed of all vertices of a given triangle mesh Model. In general, any sufficiently dense sampling of a surface can be used for *V*. If there are regions of the Model in which Data cannot be collected (e.g., because of limited access

during surgery), points in these regions can be excluded from *V*. The number of points, M, in *P* is fixed for a given trial of constraint synthesis. Finding Data configurations of *minimum size* which satisfy registration accuracy requirements is discussed below.

Constraint synthesis is a combinatorial search problem, and for all but artificially small problems the solution space is too large to search exhaustively. A search algorithm for solving constraint synthesis which combines hillclimbing and a non-deterministic optimization method is described below. A complete description of constraint synthesis solution methods can be found in [8].

Hybrid PBIL / Hillclimbing Search Algorithm

In *next ascent hillclimbing* (NAH), M vertices are chosen from the set of possible vertices, V, and placed into the "selected" set P. Let NAI(P) represent the value of the NAI computed from (5) - (8) using the points in P. Randomly select a vertex, \mathbf{v}_p , from P, and a vertex, \mathbf{v}_v , from V. Substitute \mathbf{v}_p with \mathbf{v}_v in P, and compute the new value of NAI(P). If this substitution results in an improvement in the NAI, then implement the substitution and iterate the process. If the substitution does not improve the NAI, then recompute NAI(P) with new randomly selected vertices \mathbf{v}_p and \mathbf{v}_v . Continue iterating until there are no additional substitutions which improve the NAI. The maximum number of NAI evaluations per iteration is N(M-1) where N and M are the number of vertices in the sets V and P respectively, although such a large number of evaluations is rarely reached in practice. For an average size problem (e.g., $N \approx 5000$, $M \approx 50$), NAH usually converges within 1000 iterations. The number of NAI evaluations is typically small during initial iterations, and increases during the later iterations when there are fewer possible substitutions which increase NAI(P).

In high-dimensionality optimization problems, hillclimbing methods such as NAH are susceptible to local minima in the search space. Genetic algorithms (GAs) are biologically motivated adaptive systems based upon principles of natural selection and genetic recombination which attempt to avoid local minima. A simplified model of the GA called Population-Based Incremental Learning (PBIL) was recently introduced [1]. For the purposes of this paper, PBIL can be thought of as a black-box with the following inputs: the set of allowable vertices, V; the number of points in the configuration set, P; a function which computes NAI(P) based upon the surface Model; and a stopping criterion. The output of PBIL is the particular configuration which maximizes the NAI among all configurations evaluated by PBIL within a given trial.

While PBIL is good at avoiding local minima in the constraint synthesis search space, the resulting solutions may not be locally optimal. Likewise, hillclimbing methods are good at ensuring local optimality, but usually don't converge to globally optimal configurations. By combining these two approaches, it is possible to take advantage of the strengths of each. In the hybrid search algorithm, PBIL is run, followed by a run of NAH initialized at the configuration found by PBIL.

Data Configuration Stability

Data collection plans (DCPs) generated by constraint synthesis can be used to guide acquisition of registration data. Since the precise object location is unknown before registration, it is impossible to acquire the exact points specified by constraint synthesis. Due to this uncertainty and to noise in the sensing process, the *effective* NAI value (i.e., computed from the collected Data) may be smaller than the *ideal* NAI (i.e., computed from the DCP). Certain Data configurations are more stable than others (i.e., there is

less NAI variation as the points in *P* are perturbed about the DCP positions). Attempts to incorporate stability criteria into the constraint synthesis process have resulted in exponential complexity [8]. Nevertheless, improved stability can be achieved via two methods: navigational guidance during Data collection and high curvature filtering.

During the data acquisition process, it is possible to use the current registration transformation estimate to provide navigational guidance to the human Data collector. Guidance is provided by displaying a 3-D graphical rendering of the registration object and overlaying icons representing the locations of the desired point and the sensor tip. The sensor location icon is dynamically updated in real-time based upon measurements, and is derived from the registration transformation estimate. The goal of the Data collector is to align the two icons. Each time additional Data is collected, uncertainty in the collection process is reduced by refining the registration transformation estimate.

The primary cause of Data configuration instability is disparity between the surface normals of desired and collected Data points. Constraint synthesis may select a given Data point because its surface normal strongly contributes to constraint in a given direction (see (5)). However, if the Data point actually collected has a significantly different surface normal, the resulting NAI value may be reduced. This effect can be reduced by initially focussing data collection in regions of low curvature so that surface normals of the collected points are more likely to be similar to those of the desired points. After low curvature points are collected and collection uncertainty is reduced, points in higher curvature regions can be collected. To implement this, several DCPs are synthesized, some with points in regions of low curvature and others in regions of higher curvature [8]. The resulting DCPs can then be used to guide the collection process.

3 Results

This section demonstrates significant improvement in registration accuracy due to the proposed methods. Three Models are used in the reported experiments: a cube with edge length of 100 mm, a human cadaveric femur, and a human cadaveric pelvis. Models of the femur and pelvis with superimposed Data collection plans are shown in Fig. 1.

The registration error measure used to report results in this section is the maximum correspondence error (MCE) [8][11]. The MCE is computed by transforming all vertices in a Model by T_{err} of (2), computing distances between each transformed vertex and its un-transformed correspondence, and selecting the largest distance. The MCE specifies the largest single point displacement within a registration object resulting from T_{err} .

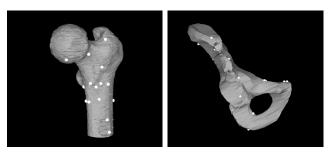


Fig. 1. Surface Models of the femur and pelvis with overlaid DCPs.

3.1 Constraint Analysis Experiments

Registration trials were conducted using simulated Data to demonstrate the relation between registration error and NAI. Data points were generated by applying known random transformations to nominal Data configurations and adding zero-mean Gaussian noise. Since the true transformations, T_{true} , are known, the error transformations, T_{err} , can be computed. Fig. 2 shows a plot of MCE vs. NAI for the three nominal cube configurations shown on the right of the figure. Configuration C1 contains 24 points per face, while C2 and C3 contain 4 points per face each. For each configuration, the mean, standard deviation, minimum and maximum MCE over 500 registration trials are plotted. The parameters for generating noise and random transformations are shown in the plot. The trend from the plot is clear: configurations with larger values of NAI result in smaller registration error. In particular, note that configuration C2 has smaller values of MCE (and a larger NAI) than C3, despite having the same number of points.

Fig. 3 demonstrates differences in noise sensitivity as a function of NAI for the cube configurations. The graphs show how MCE varies as a function of expected noise magnitude. For each datum, 500 registration trials were performed and the mean values for these trials are plotted. In the absence of noise, all three configurations perform equally well. As noise increases, configurations with smaller values of NAI are clearly more sensitive. This illustrates that the utility of intelligent data selection is dependent upon the magnitude of sensor noise (among other factors).

3.2 Constraint Synthesis Experiments

Table 1 demonstrates the efficacy of the constraint synthesis search algorithms for the pelvis. Data configurations were synthesized using 4 configuration sizes and 4 methods of generation. Five configurations were generated for each size-method combination,

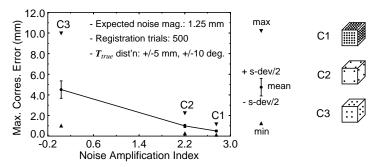


Fig. 2. MCE vs. NAI for 3 configurations on a cube.

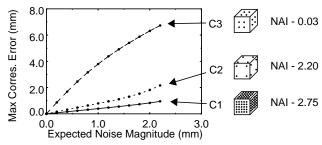


Fig. 3. MCE vs. expected noise magnitude for 3 cube configurations.

except for the random method for which 1000 configurations were generated. The maximum and minimum NAI values over the generated configurations are shown. For each configuration size, the hybrid PBIL/NAH method produced the best results.

|--|

Method	Configuration Size - M			
	10	25	50	75
Random	0.41 / 0.00	1.18 / 0.08	1.52 / 0.33	1.68 / 0.53
NAH	1.42 / 1.28	2.62 / 2.43	3.97 / 3.76	4.90 / 4.79
PBIL	1.41 / 1.35	2.70 / 2.58	3.88 / 3.81	4.84 / 4.76
PBIL + NAH	1.52 / 1.36	2.75 / 2.65	4.02 / 3.92	4.94 / 4.90

Fig. 4 compares 5 random and 5 synthesized configurations of size 25 for the pelvis in a plot of MCE versus NAI. For each configuration, a set of registration trials was performed using the indicated parameters. In this graph, the 5th and 95th percentiles of MCE are plotted instead of the minimum and maximum values. When generating the simulated registration Data, a second noise component was added which models the uncertainty associated with Data collection. This noise perturbs a point from its nominal location by a uniform random distance *along the surface*. For this experiment, the radius of uncertainty was 5.0 mm. From the graph, it is clear that the synthesized configurations are superior to the randomly generated ones in terms of both NAI and MCE.

Fig. 5 shows similar results for the femur Model using 5 random and 5 synthesized configurations of size 10. The figure demonstrates the effect of high curvature filtering; no filtering results in unstable Data configurations and larger errors.

3.3 In-vitro Cadaver Experiments

We performed registration trials using Data collected from a cadaveric femur. For these experiments, estimation of T_{true} is a challenging engineering problem which our group has solved using a highly accurate fiducial-based registration method [11]. Using a filtered version of the femur Model, DCPs of 6 and 50 points were synthesized, each a total of 5 times. The corresponding Data points were collected on the actual femur using a digitizing probe. Each synthesized configuration was independently collected 5 times. In addition, 50 manually-selected Data sets were collected for each configuration size. To guide the collection process, the navigational guidance mechanism described above

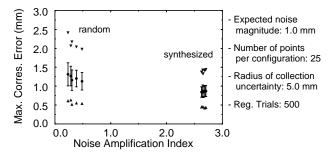


Fig. 4. MCE vs. NAI for 5 random and 5 synthesized configurations on pelvis Model.

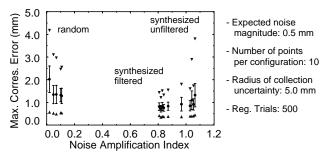


Fig. 5. Femur: MCE vs. NAI - random and synthesized points.

was used. Initial values of T_{est} were computed using manually selected anatomical landmarks and point-to-point registration [2].

Experimental results are shown in Fig. 6. Each graph plots the MCE value resulting from registration versus the *effective* NAI value computed after registration using the closest Model points (M_i of (1)) to solve for n and x_s of (5). From the graphs it is clear that the synthesized point configurations are superior to the manually selected ones for both configuration sizes. Six points is the theoretical minimum number required to solve the shape-based registration problem without correspondence. As seen, selecting 6 well-conditioned Data points is a difficult task for humans. Note that some synthesized configurations for the 6-point results have small NAI values and large MCE values due to data collection uncertainty. However, using the online accuracy evaluation method described below, such configurations can easily be identified and additional Data can be collected to improve the result.

To be useful, an online accuracy estimate must relate a quantity which can be measured during the registration process, to a second quantity which has physical meaning to the task for which registration is being performed. Fig. 7 shows a plot of MCE versus RMS error (definition in the figure). It is shown in [8] that the slope of the line which relates worst case MCE to RMS error is independent of sensor noise, the number of Data points, and Data collection uncertainty, assuming that the effective NAI value is slightly greater than zero. Furthermore, it is shown that the slope of this line can be determined from simulated registration experiments such as those of Section 3.2. Therefore, during the registration process, online measurement of RMS error can be used to estimate an upper bound on MCE. This estimate can then be used to determine if accuracy requirements are satisfied, and additional Data collection can be requested if not.

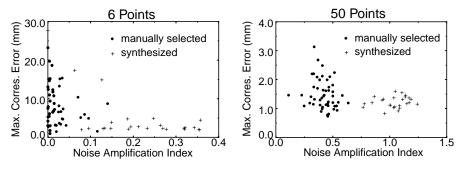


Fig. 6. MCE vs. NAI - physically collected Data on femur. Note scale differences.

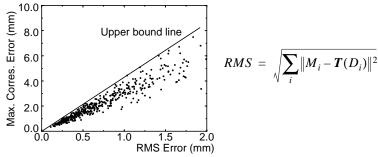


Fig. 7. Basis of online accuracy estimation - plot of MCE vs. RMS error.

By coupling online accuracy estimation with intelligent data selection, it is possible to collect *minimally-sized* Data sets which satisfy accuracy requirements. This is done by pre-synthesizing multiple NAI-optimal configurations of increasing size, each of which is a superset of the previous. During the collection process, a Data set is collected and registered, and accuracy is estimated. This process is continued until accuracy requirements are satisfied, or until all of the synthesized sets have been collected.

4 Conclusions

The methods described in this paper show promise as tools for analyzing and maximizing accuracy in shape-based registration. Intelligent data selection is likely to be most useful when data collection is expensive and sensor noise is high. Online accuracy estimation is likely to be useful with and without intelligent data selection. Work is currently in progress to evaluate the practicality of these methods in clinical situations.

References

- [1] S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithm. In A. Prieditis, ed, *Proc. Int'l Conf. Mach. Learning*, pp. 38–46, San Mateo, CA, 1995. Morgan Kaufmann Publishers.
- [2] P. Besl and N. McKay. A method for registration of 3-d shapes. IEEE. PAMI, 14(2):239–256, Feb 1992.
- [3] E. Cuchet et al., Registration in neurosurgery and neuroradiotherapy applications. In Proc 2nd Int'l Symp. MRCAS, pp. 31–38, Baltimore, Nov. 1995.
- [4] M. G. Kendall and A. Stuart. Canonical Variables, chap 43, pp. 320-369. Griffin, London, 4th ed., 1977.
- [5] S. Lavallee. Registration for computer-integrated surgery: Methodology, state of the art. In R. H. Taylor, et al., eds, Computer-Integrated Surgery, chap 5, pp 77–97. MIT Press, Cambridge, Massachusetts, 1995.
- [6] C.H. Menq, H.T. Yau, and G.Y. Lai. Automated precision measurement of surface profile in CADdirected inspection. *IEEE Trans. Robotics and Automation*, 8(2):268–278, April 1992.
- [7] A. Nahvi and J.M. Hollerbach. The noise amplification index for optimal pose selection in robot calibration. In *Proc IEEE Int'l Conf. Robotics and Automation*, Minneapolis, April 1996.
- [8] D. A Simon. Fast and Accurate Shape-Based Registration. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, December 1996.
- [9] D. A. Simon, et al., Development and validation of a navigational guidance system for acetabular implant placement. In *Proc. 1st Joint CVRMed / MRCAS Conference*, Grenoble, March 1997.
- [10] D. A. Simon, et al., Accuracy validation in image-guided orthopaedic surgery. In Proc. 2nd Int'l Symp. MRCAS, Baltimore, Nov. 1995.
- [11] D. A. Simon, M. Hebert, and T. Kanade. Techniques for fast and accurate intra-surgical registration. *Journal of Image Guided Surgery*, 1(1):17–29, April 1995.
- [12] G. Taubin. Est'n of planar curves, surfaces, and nonplanar space curves defined by implicit eqn's with applications to edge and range image segmentation. *IEEE Trans PAMI*, 13(11):1115–1138, Nov 1991.