An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization

W.E.L. Grimson[1] T. Lozano-Pérez[1] W.M. Wells III[1,2] G.J. Ettinger[1,3]
S.J. White[3] R. Kikinis[2]

Abstract

There is a need for frameless guidance systems to help neurosurgeons to plan the exact location of a craniotomy, to define the margins of tumors and to precisely identify locations of neighboring critical structures. We have developed an automatic technique for registering clinical data, such as segmented MRI or CT reconstructions, with the patient's head on the operating table. A second method calibrates the position of a video camera relative to the patient. The combination allows a visual mix of live video of the patient with the segmented 3D MRI or CT model, enabling enhanced reality techniques for planning and guiding neurosurgical procedures, and to interactively view extracranial or intracranial structures non-intrusively. Extensions of the method include image guided biopsies, focused therapeutic procedures and clinical studies involving change detection over time sequences of images.

1 Motivating Problem

Many surgical procedures require highly precise localization on the part of the surgeon, in order to extract targeted tissue while minimizing collateral damage to adjacent structures. The problem is exacerbated by the fact that this 3D localization often requires isolating a structure deeply buried within the body. While methods exist (e.g. MRI, CT) for imaging and displaying the 3D structure of the body, the surgeon must still relate what she sees on the 3D display with the actual anatomy of the patient.

Current solutions in neurosurgery typically involve presurgically attaching a stereotactic frame to the patient's skull, then imaging the skull and frame as a unit. This allows the surgeon to determine, from the imagery, the location of the tumor or other target relative to a coordinate system attached to the stereotactic frame, and thus to the patient's head. As well, the frame typically allows the positioning of a probe at any orientation relative to the patient, letting the surgeon mark a planned angle of entry that localizes the expected extraction of material. Unfortunately, stereotactic frames are cumbersome to the surgeon, and involve considerable discomfort to the patient. As well, such frames can have limited flexibility, especially should surgical plans change in the middle of the procedure, e.g. if the line of attack is found to pass through critical regions, such as the motor strip.

1.1 An Ideal Solution

Ideally, one would prefer a system that automatically registers 3D data sets, and tracks changes in a data set's position over time, without requiring the attachment of any devices to the patient. An ideal system should support: real-time, adaptive, enhanced reality patient visualizations in the operating room; dynamic image-guided surgical planning; image guided surgical procedures, such as biopsies or minimally invasive therapeutic procedures; and registered transfer of *a priori* surgical plans to the patient in the OR.

While our group is actively developing all aspects of such a system, this paper focuses on one key component, the registration of different data sources to determine relevant coordinate frame transformations.

1.2 Contributions to the Ideal Solution

We have created a system that registers clinical image data with the position of the patient's head on the operating table at the time of surgery, using methods from visual object recognition. The method does not require a previously attached stereotactic frame. The method has been combined with an enhanced reality technique [7, 2, 19], in which we display a composite image of the 3D anatomical structures with a view

[1] AI Lab, MIT, Cambridge MA
[2] Dept. of Radiology, Brigham and Womens Hospital, Harvard Medical School, Boston MA
[3] The Analytic Sciences Corporation, Reading MA

[0] This report describes research supported in part by DARPA under Army contract number DACA76-85-C-0010 and under ONR contracts N00014-85-K-0124 and N00014-91-J-4038 and in part by the Medical Informatics training grant No. T 15 LM 07092 from the National Library of Medicine.

of the patient's head. This registration enables the transfer to the operating room of preoperative surgical plans, obtained through analysis of the segmented 3D preoperative data [4], where they can be graphically overlaid onto video images of the patient. Such transfer allows the surgeon to apply carefully considered surgical plans to the current situation, and to mark internal landmarks used to guide the progression of the surgery. Extensions of our method include adaptively re-registering the video image of the patient to the 3D anatomical data, as the patient moves, or as the video source moves, as well as other surgical applications such as image guided biopsy, or focused therapeutic procedures such as laser disc fusion or tumor ablation. We have also recently demonstrated the use of our system in clinical settings, by registering data sets acquired over extended time periods, thereby enabling the detection of changes in anatomy over time.

2 An Example Scenario

The following scenario demonstrates our approach:

(1) A patient requiring surgery is scanned by a 3D, high resolution, internal anatomy scanner, such as Magnetic Resonance Imaging (MRI) or Computed Tomography (CT). The scan is segmented into tissue types.

(2) The patient is placed in the operating room. Prior to draping, the patient is scanned by a laser range scanner. The 3D locations of any table landmarks are also calculated to identify their location relative to the patient. The current MRI or CT scan is automatically registered to the patient skin surface depth data obtained by the laser range scanner. This provides a transformation from MRI/CT to patient.

The position and orientation of a video camera relative to the patient is determined, by matching video images of the laser points on an object to the actual 3D laser data. This provides a transformation from patient to video camera. The registered internal anatomy is displayed in enhanced reality visualization [7, 2, 19] to "see" inside the patient. In particular, the two previously computed transformations can be used to transform the 3D model into the same view as the video image of the patient, so that video mixing allows the surgeon to see both images simultaneously.

The patient is draped and surgery is performed. The enhanced reality visualization does not require the surgeon to do anything different from normal, but rather provides her with additional visualization information to greatly expand her limited field of view.

(3) The location of table landmarks can be continually tracked to identify changes in the position of the patient's attitude, relative to the visualization camera. Visualization updates are performed by updating the MRI/CT to patient transformation. Viewer location can be continually tracked to identify any changes in the position of the viewer. For a stationary video camera, this is straightforward, though for head-mounted displays such tracking is both more relevant and more challenging. Visualization updates are performed by updating the patient to viewer transformation.

The surgical procedure is executed with an accurately registered visualization of the anatomy of the patient, thus reducing side effects.

3 Details of Our Approach

Methods currently exist for Part 1 [4, 6]. In this paper we focus on part 2, where the key step is the registration of data obtained from the patient in the operating room with previously obtained data and surgical plans. Part 3 is part of our planned future work. The basic steps of our method are outlined below.

3.1 Model input

We obtain a segmented 3D reconstruction of the patient's anatomy, (e.g. CT or MRI). Current segmentation techniques are generally semi-automatic, typically by training an intensity classifier on a user selected set of tissue samples, where the operator uses knowledge of anatomy to identify the tissue type. Once initial training is completed, the rest of the scans can be automatically classified on the basis of intensities in the scanned images, and thus segmented into tissue types [4, 6]. Removing gain artifacts from the sensor data [20], and correcting for distortions due to magnetic susceptibility differences between different materials [17] can both improve the segmentation.

This 3D anatomical reconstruction is referred to as the model, and is represented relative to a model coordinate frame, whose origin is the centroid of the points.

3.2 Data input

We obtain a set of 3D data points from the patient's skin surface using a Technical Arts laser range scanner. It operates by scanning a laser beam using an oscillating mirror through an optical mechanism that results in a controlled plane of light. A video camera is placed at an angle to this plane such that a portion of the plane is in the camera field of view. When an object is placed in this visible region such that it intersects the laser plane, points in the camera image illuminated by the laser unambiguously correspond to fixed 3D scene points. In general, the correspondences

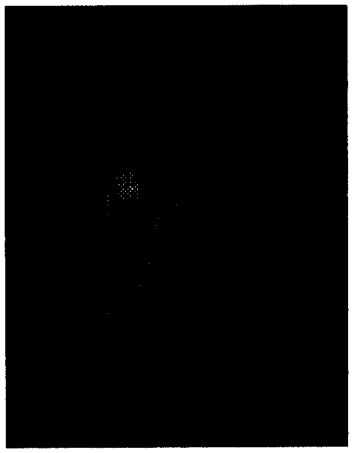

Figure 1: **Example of laser data (shown as large dots) overlaid on CT model of a plastic skull, after an initial alignment of the two point sets. Note the transparent laser points which are actually lying inside the skull.**

between the scene points and image points are calculable by using a nonlinear projective transform, which can be determined by scanning an object of known form. Since the deflection of the beam in the image is in a known direction, one can process many such points in a single scan (or position of the laser light plane). In this case, the device actually produces 240 3D measurements for any single scan. The measurements are accurate to within 0.003".

The plane of the laser can be arbitrarily controlled, so that data points from a 3D volume are obtained. The laser can either be moved by small increments to obtain a dense sampling of data, or the laser plane can be moved by larger increments, to obtain a small number (5-10) of planar slices of data from the scene.

This information is referred to as the data, and is represented in a coordinate frame attached to the laser, which reflects the position of the patient in a coordinate frame that exists in the operating room. Our problem is to determine a transformation that will map the model into the data in a consistent manner.

3.3 Matching data sets

We match the two data sets as follows:

(1) To initiate the matching, we have several options. We can use a simple graphical interface to roughly align the laser data with the 3D model, providing an estimate of the view direction of the model. In this case, we extract a sampled set of visible points of the model, using a z-buffer. Given a pixel size for the z-buffer and given an estimate for the view direction, we project all of the model points into a plane orthogonal to the view direction, where the plane is tesselated in pixels of the given side. Within each pixel, we keep only the point closest to the viewer. This gives us a temporary model, which we can use for matching.

Alternatively, we sample a set of evenly spaced directions on the view sphere. For each view, we use the z-buffer method described above to extract a sampled set of visible points of the model. For each such model, we execute the matching process described below.

(2) Next, we separate laser data of the patient's head from background data. Currently we do this with a simple user interface, in which three orthogonal views of the laser data are presented, and the user places a bounding box around the data that actually comes from the patient. Note that this process need not be perfect, we simply want to remove gross outliers from the data. Given this segmented laser data, we find three widely separated laser points.

(3) We use constrained search [9, 8] to examine all ways of matching the three laser points to three points selected from the sampled MRI model. For each association, the method tests whether the pairwise distances between model points and laser points are roughly the same. If all such tests are valid, the match is kept, and we compute the coordinate frame transformation that maps the three laser points into their corresponding model points. These transformations form a set of hypotheses. Note that due to the sampling of the model data, the actual object points corresponding to the selected laser points may not exist, so these hypothesized transformations are at best approximations to the actual transformation.

In the example of Figure 4, there are 481 laser sample points, and the skull model has 35,265 sample points. Given an estimated view, and a coarsely sampled z-buffer, there are 409 model points in the sampled view. In principle, there are $\approx 2.02e14$ possible hypotheses, but using simple distance constraints, only 16,945 possible hypotheses remain for further testing.

(4) We use the Alignment Method [12] to filter these hypotheses. For each hypothesis, we verify that the fraction of the laser points, transformed by the hypothesized transformation, without a corresponding model point within some predefined distance is less than some predefined bound. We discard those hypotheses that fail this verification. For efficiency, we use two levels of sampling of the laser points, first verifying that a coarsely sampled set of laser points are in agreement, then further verifying, for those that pass this test, that all the laser points are in agreement.

Figure 1 shows an example of the model and laser data after a verified alignment. Note that some of the

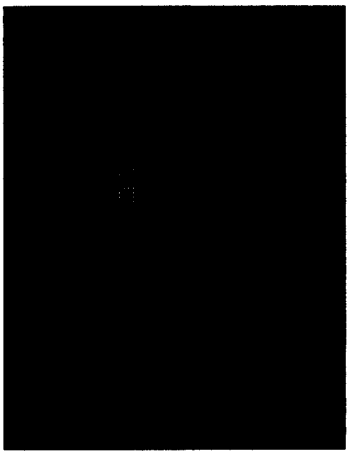

Figure 2: Final alignment of data and model.

laser points are partially buried in the CT model (displayed as partially transparent), indicating that the initial alignment is not sufficiently accurate.

(5) Evaluate each verified hypothesis as follows:

(5.1) Sum, for all transformed laser points, a term that is a sum of the distances from the transformed point to all nearby model points, where the distance is weighted by a Gaussian distribution [18]. This Gaussian weighting is a method for roughly interpolating between the sampled model points to estimate the nearest point on the underlying surface to the transformed laser point. More precisely, if vector ℓ_i is a laser point, vector m_j is a model point, and T is a coordinate frame transformation, then the evaluation function for a particular pose (or transformation) is

$$E_1(T) = \sum_i \sum_j -e^{-\frac{|T\ell_i - m_j|^2}{2\sigma^2}}. \qquad (1)$$

This function is similar to the posterior marginal pose estimation (PMPE) method of [18]. Because of its formulation, the objective function is quite smooth, and thus facilitates "pulling in" solutions from moderately removed locations in parameter space.

(5.2) Iteratively maximize this evaluation function using Powell's method. This yields an estimate for the pose of the laser points in model coordinates.

(5.3) Execute stages 5.1 and 5.2 with a multiresolution set of Gaussians. A broad Gaussian is used to allow influence over large areas, resulting in a coarse initial alignment, which can be reached from a wide range of starting positions. Then, narrower Gaussian distributions are used to focus on only nearby model points to derive the pose.

(5.4) Starting from the resulting pose, repeat the evaluation process, using a rectified least squares distance measure. In particular, perform a second sampling of the model from the current viewpoint, using a much more finely sampled z-buffer. Relative to this finer model, use Powell's method to minimize the evaluation function:

$$E_2(T) = \sum_i \min\left\{d_{\max}^2, \min_j |T\ell_i - m_j|^2\right\} \qquad (2)$$

where $d_{\max}$ is some preset maximum distance. This objective function is essentially the maximum a posteriori model matching scheme of [18]. It acts much like a robust chamfer matching scheme (e.g. [13]). The expectation is that this second objective function is more accurate locally, since it is composed of saturated quadratic forms, but it is also prone to getting stuck in local minima. Hence we add one more stage.

(5.5) We observe that while the above method always gets very close to the best solution, it can get trapped into local minima in the minimization of E_2. To improve upon this, we take the pose returned by the above step, and perturb it randomly, then repeat the minimization. We continue to do this, keeping the new pose if its associated RMS error is better than our current best. We terminate this process when the number of such trials that have passed since the RMS value was last improved becomes larger than some threshold.

(5.6) The final result is a pose, and a measure of the residual deviation of the fit to the model surface. An example is shown in Figure 2.

We collect such solutions for each verified hypothesis, over all legal view samples, and rank order them by smallest RMS measure. The result is a highly accurate transformation of the MRI data into the coordinate frame of the laser sensor.

3.4 Camera Calibration

Once we have such a registration, it can be used for surgical planning. A video camera can be positioned in roughly the viewpoint of the surgeon, i.e. looking over her shoulder. By calibrating the position and orientation of this camera relative to the laser coordinate system, we can render the aligned MRI or CT data relative to the view of the camera. This rendering can be mixed with the live video signal, giving the surgeon an enhanced reality view of the patient's anatomy [7, 2, 19]. This can be used to plan a craniotomy or a biopsy, or to define the margins of an exposed tumor for minimal excision. Figure 3 shows an alignment of a CT model and an actual image of a skull in a calibrated video camera.

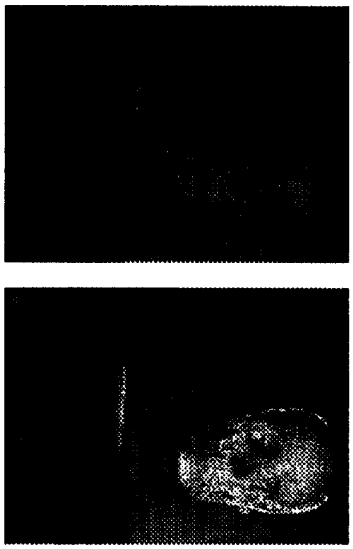

Figure 3: **Example of video image, and overlay of a registered 3D (CT) object model with the real object in that image.**

We have investigated two methods for calibrating the camera position and orientation. In the first case, a calibration object of known size and shape is placed in the common field of view of the laser ranging system and the video camera. Landmark points on the object are identified, and measured in camera coordinates by extracting the landmark points in the video image, and in laser coordinates by fitting a model of the calibration object to the laser data. The camera parameters are calculated by using Powell's method to minimize the distance between transformed laser points and matched image points.

A second method does not rely on a known calibration object. Instead images of the laser slices are taken with the video camera. Straight line segments are located in the video images and matched to corresponding straight line segments in the laser data. If three such matching segments are found, they can be used to solve for an approximation to the perspective projection transformation, and thus for the pose of the camera. Using this as a starting point, Powell's method can again be used to optimize the pose estimate to best bring all of the laser data into projective alignment with the corresponding video data. Thus, one can use the patient directly to calibrate the camera, and thus this process can be repeated throughout the surgical procedure as needed (e.g. if the position of the camera relative to the operating table is perturbed).

Figure 4: **Example of registered laser data (shown as large dots) overlaid on CT model.**

4 Testing and Applications

As a first controlled experiment, we have registered a CT reconstruction of a plastic skull with laser data extracted for a variety of viewpoints. We have run the system both with an initial pose estimate, and by sampling a range of views on the viewing sphere. In all cases, the system finds a correct registration (Figure 4), with typical residual RMS errors of 1.6 mms.

For a typical set of views of the model, the number of points in a coarse model was $\approx$ 500, and led to initial hypotheses ranging from 16,945 to 114,062. Only one hypothesis survived Alignment verification, in which case a finer model of 9428 points was refined relative to the laser data. This led to a single correct solution with an RMS residual of 1.5mm. It is worth commenting on the RMS error, since interactively viewing the overlaid results on a 3D display suggests that the registration is much more accurate than an RMS of 1.5 mm would suggest. First, the method we use to extract a surface model from the MRI or CT data is overly simple. For each individual slice of the data, we extract exterior boundary points. When the underlying surface is oriented nearly tangential to the image slice, this method will undersample the skin surface. Thus, between two adjacent slices, there may be considerable gaps between extracted skin points. A better solution would be to interpolate a dense skin surface, and ensure that model skin points are uniformly sampled. The effect of not doing this is that occasionally we can have transformed laser points that lie quite close to the model surface, but for which the nearest sampled point is some distance away along the surface. Thus the tail of the histogram can pull the overall RMS value to higher value than is correct (Figure 5).

Note that the resolution of the CT scan is $1 \times 1 \times 2mms$. Thus, the model points lie at the nodes of a

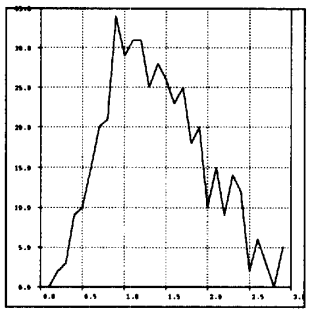

Figure 5: Histogram of residual errors for the final pose of Figure 2.

discrete lattice, and since the laser points are not constrained to lie on the same lattice, this descretization will also contribute to the reported RMS errors.

We have also successfully run trials matching laser data against an MRI scan of one of the authors, an example of which is shown in Figures 6 and 7. The resolution of this MRI scan is $0.9375 \times 0.9375 \times 1.5mms$.

Recently we have run a series of trials with actual neurosurgery patients. An example registration of the laser data against an MRI model of the patient is shown in Figure 8. Note that while most of the scalp had been shaved for surgery, a patch of hair was left hanging down over the patient's temple. As a result, laser data coming from the hair cannot be matched against the segmented skin surface in the MRI model, and this shows up as a set of points slightly elevated above the patient's skin surface in the final registration. We can automatically remove these points, and reregister the remaining data. As well, the tumor and the ventricles of the patient are also highlighted. The RMS error in this case was 1.9mm. Finally, given the

Figure 7: Visualisation image of brain merged with video view.

registration between the patient and the model (by matching the laser data in this manner) we can transform the model into the coordinate system of a second video camera, and overlay this model on top of the camera's video view. This is shown in Figure 9.

Besides applications for surgical planning and guidance, including tumor excision and biopsy, the method has other applications, including the registration of multiple clinical data sets such as MRI versus CT. As a demonstration of this, we have registered equences of MRI scans of the same patient, taken over a period of several months, and used differences in the registered scans to visualize and measure changes in anatomy [5]. These scans are part of an ongoing NIH study of multiple sclerosis (MS) at Brigham and Womens Hospital aimed at determining the optimal frequency for performing MR imaging of MS patients.

5 Related Work

Several other groups have reported methods similar to ours. Of particular interest are three such approaches. Pelizzari et al. [16, 15] have developed a method that matches retrospective data sets, (MRI, CT, PET), to one another. This work also uses a least squares minimization of distances between data sets,

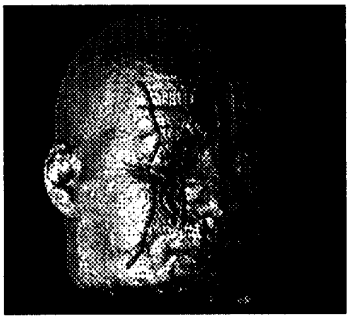

Figure 6: Example of registered laser data (shown as large dots) overlaid on MRI model.

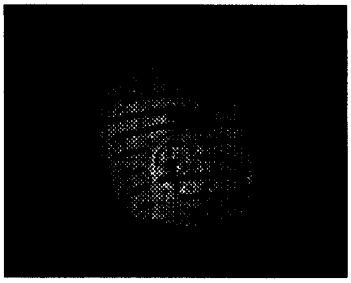

Figure 8: Example of registered laser data (shown as large dots) overlaid on an MRI model. This is a case of registration of an actual neurosurgical case, with the patient fully prepped for surgery before the laser data is acquired.

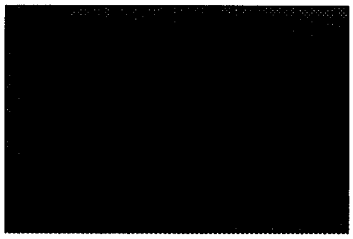

Figure 9: Using the results of Figure 8, and given a calibration of a video camera relative to the laser, we can overlay parts of the MRI model on top of a video view of the patient, providing an enhanced reality visualisation of the tumor. In this figure, the tumor is shown in green, and the ventricles are displayed as a landmark in blue.

although with a different distance function. Typical reported RMS errors are 3-5mms. This approach does require some operator intervention to set a decent initial starting position, which our system does not. It also apparently requires some operator intervention to steer the system towards the correct solution, suggesting that local minima are a potential problem. Our system avoids this difficulty by randomly perturbing near final solutions to find better nearby minima.

A second related approach [3, 14] also does a least-squares minimization of a distance function to match data sets. Here, the distance is weighted by an estimate of the inverse variance of the measurement noise, and a Levenberg-Marquardt method is used to find the minimum. The method presently requires a reasonable initial starting position, though the authors observe that sampling over the view sphere could remove this restriction. Once an initial solution is found, points with large errors are removed and the minimization is repeated to refine the pose. It is unclear whether removing outliers is sufficient to keep the method from getting trapped into local minima.

A third approach [1, 10, 11] performs automatic rigid registration of 3D surfaces by matching ridge lines which track points of maximum curvature along the surface. This method is not directly suitable for dealing with sparse data, such as the laser input.

References

[1] Ayache, N., J.D. Boissonnat, L. Cohen, B. Geiger, J. Levy-Vehel, O. Monga, P. Sander, "Steps Toward the Automatic Interpretation of 3-D Images", In *3D Imaging in Medicine*, Fuchs, Hohne, & Piser (eds) NATO ASI Series, Springer-Verlag, 1990, pp 107-120.

[2] Black, P., R. Kikinis, W. Wells, D. Altobelli, W. Lorensen, H. Cline F. Jolesz, "A New Virtual Reality Technique for Tumor Localisation" *Cong. Neurological Surgeons*, 1993.

[3] Champleboux, G., S. Lavallee, R. Szeliski, L. Brunie, "From accurate range imaging sensor calibration to accurate model-based 3D object localization", *CVPR*:83–89, 1992

[4] Cline, H., W. Lorensen, R. Kikinis, F. Jolesz, 1990, "3D Segmentation of MR Images of the Head Using Probability and Connectivity." *JCAT* 14(6):1037-1045.

[5] Ettinger, G., E. Grimson, T. Lozano-Perez, "Automatic Registration for Multiple Sclerosis Change Detection", *CVPR Workshop Biomed. Image Anal.*, Seattle, 1994.

[6] Gerig, G., W. Kuoni, R. Kikinis, O. Kübler, 1989, "Medical Imaging and Computer Vision: an Integrated Approach for Diagnosis and Planning," *Proc. 11'th DAGM Symposium*, Hamburg FRG, Springer, pp. 425–443.

[7] Gleason, L., R. Kikinis, D. Altobelli, W. Wells, E. Alexander, P. Black, F. Jolesz, "A New Virtual Reality Technique for Non-Linkage Stereotactic Surgery" *Society for Stereotactic & Functional Neurosurgery, Ixtapa, Mexico*, 1993.

[8] Grimson, W.E.L., 1990, *Object Recognition by Computer: The role of geometric constraints*, MIT Press, Cambridge.

[9] Grimson, W.E.L. and T. Lozano-Pérez, "Localising Overlapping Parts by Searching the Interpretation Tree", *IEEE PAMI*, 9, No. 4, 469 – 482, 1987.

[10] Gueziec, A., N. Ayache, "Smoothing and Matching of 3-D Space Curves", *Second ECCV*, May 1992, pp 620-629.

[11] Gueziec, A., N. Ayache, "New Developments on Geometric Hashing for Curve Matching", *CVPR*, 1993, pp 703-704.

[12] Huttenlocher, D. and S. Ullman, 1990, "Recognising Solid Objects by Alignment with an Image," *Int. J. Comp. Vis.* 5(2):195–212.

[13] Jiang, H., R. Robb, K. Holton, 1992, "A New Approach to 3D Registration of Multimodality Medical Images by Surface Matching", *Visualiz. in Biomed. Comp.*: 196–213.

[14] Lavallee, S., R. Szeliski, L. Brunie, "Matching 3d smooth surfaces with their 2d Projections using 3d distance maps", *SPIE – Geom. Methods in Comp. Vis.*, 1991, pp. 322–336.

[15] Levin, D., X. Hu, K. Tan, S. Galhotra, C. Pelizzari, G. Chen, R. Beck, C. Chen, M. Cooper, J. Mullan, J. Hekmatpanah, J. Spier, "The Brain: integrated three-dimensional display of MR and PET images", *Radiology* 172(3):783–789, 1989.

[16] Pelizzari, C., G. Chen, D. Spelbring, R. Weichselbaum, C. Chen, "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain", *J. Computer Assisted Tomography* 13(1):20–26, 1989.

[17] Sumanaweera; Glover; Binford; Adler, "MR Susceptibility Misregistration Correctoin", *IEEE TMI* 12, 1993.

[18] Wells, W. M., 1993, *Statistical Object Recognition*, Ph.D. Thesis, MIT. (MIT AI Lab TR 1398)

[19] Wells, W., R. Kikinis, D. Altobelli, W. Lorensen, G. Ettinger, H. Cline, L. Gleason, F. Jolesz, 1993, "Video Registration using Fiducials for Surgical Enhanced Reality" *Proc. 15th Conf. IEEE Engin. in Med. Biol. Soc.*, IEEE.

[20] Wells, W., R. Kikinis, F. Jolesz, E. Grimson, 1994, "Statistical Gain Correction and Segmentation of Magnetic Resonance Imaging Data", in preparation.