Automated Registration for Enhanced Reality Visualization in Surgery

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

Abstract

There is a need for frameless guidance systems to aid neurosurgeons in planning the exact location of a craniotomy, to define the margins of tumors and to precisely locate neighboring critical structures. We have developed an automatic technique for registering clinical data, such as segmented MRI or CT reconstructions, with the actual head of the patient 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. This registration enables us to employ enhanced reality techniques for planning and guiding neurosurgical procedures by merging live video images with the 3D reconstructions, and to interactively view extracranial or intracranial structures in a non-intrusive manner.

1 Motivating Problem

Neurosurgical procedures, such as biopsy or tumor ablation, require highly precise localization on the part of the surgeon, in order to attain the desired extraction of diseased tissue while minimizing damage to adjacent structures. The problem is exacerbated by the fact that the localization is three dimensional in nature, and often requires isolating a structure deeply buried within the cranium. While methods exist (e.g. MRI, CT) for imaging and displaying the 3D structure of the head, this still leaves the surgeon with the problem of relating what she sees on the 3D display with the actual anatomy of the patient.

Current solutions 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 locate the tumor or other target relative to a coordinate system attached to the stereotactic frame, and thus to the patient's head. Unfortunately, stereotactic frames are cumbersome, involve considerable discomfort to the patient, and have limited flexibility, especially should surgical plans have to change in the middle of the procedure.

[1]Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA

[2]Department of Radiology, Brigham and Womens Hospital, Harvard Medical School, Boston MA

[3]The Analytic Sciences Corporation, Reading MA

1.1 An Ideal Solution

An ideal system would automatically register 3D data sets, and track changes in the position of a data set over time, without requiring the attachment of any devices to the patient. Such an ideal system should support: real-time, adaptive, enhanced reality patient visualizations in the operating room; dynamic image-guided surgical planning and 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 of such a system, 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 performs the registration of 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 has been combined with an enhanced reality technique [11] in which we display a composite image of the 3D anatomical structures with a view 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 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 describes the use of our methods.

(1) A patient requiring surgical therapy is scanned by a 3D, high resolution scanner, such as MRI or CT.

(2) The scan is segmented into tissue type in a semi-automatic manner, typically by training an intensity classifier on a user-selected set of tissue samples.

(3) Prior to draping, the patient is scanned by a laser range scanner. The 3D locations of any table landmarks are also calculated, relative to the patient.

l) The MRI or CT scan is automatically registered to patient skin surface depth data from the laser ranger. ı provides a transformation from MRI/CT to patient.

ı) The position and orientation of a video camera rela- to the patient is determined, by matching video images :e laser points on an object to the actual 3D laser data. . provides a transformation from patient to camera.

) The registered internal anatomy is displayed in en- :ed reality visualization [11] to "see" inside the patient, the two computed transformations are used to trans- ı the 3D model into the same view as the video image ıe patient, so that video mixing allows the surgeon to ›oth images simultaneously.

) The patient is draped and surgery is performed. The .nced reality visualization does not interfere with the :on, but provides her with additional visualization in- ation to greatly expand her limited field of view.

) The location of table landmarks can be continually :ed to identify changes in the position of the patient's ude, relative to the visualization camera. Viewer ion can also be continually tracked to identify any ges. Visualization updates are performed by updat- he patient to viewer transformation.

) In general, the surgery is executed with an accurately tered enhanced visualization of the entire relevant pa- anatomy, and thus with reduced side effects.

Details of Our Approach

1 of this scenario is standard practice. Methods exist 'art 2 [4]. Parts 8–9 are part of our planned future . Here, we focus on parts 3–7, where the key step is egistration of data obtained from the patient in the ıting room with previously obtained data.

: use a multi-stage matching and verification of a 3D set acquired at the time of the surgery with 3D clini- ıta sets acquired previously. The central ideas are to laser striping device to obtain 3D data from the pa- s skin, and to use a sequence of recognition techniques ıtch this data to segmented skin data from the MRI Γ reconstruction. These techniques allow us to accu- r register the clinical data with the current position of atient, so that we can display a superimposed image : 3D structures overlaid on a view of the patient.

e basic steps of our method are outlined below.

Model input

btain a segmented 3D reconstruction of the patient's ›my, for example using CT or MRI. The segmentation ›ically done by training an intensity classifier on a :elected set of tissue samples, where the operator uses ledge of anatomy to identify the tissue type. Once l training is completed, the rest of the scans can be ıatically classified on the basis of intensities in the ed images, and thus segmented into tissue types [4]. natically removing gain artifacts from the sensor data e used to improve the segmentation [12].

is 3D anatomical reconstruction is referred to as the l, and is represented relative to a model coordinate . For simplicity, the origin of the coordinate system e taken as the centroid of the points.

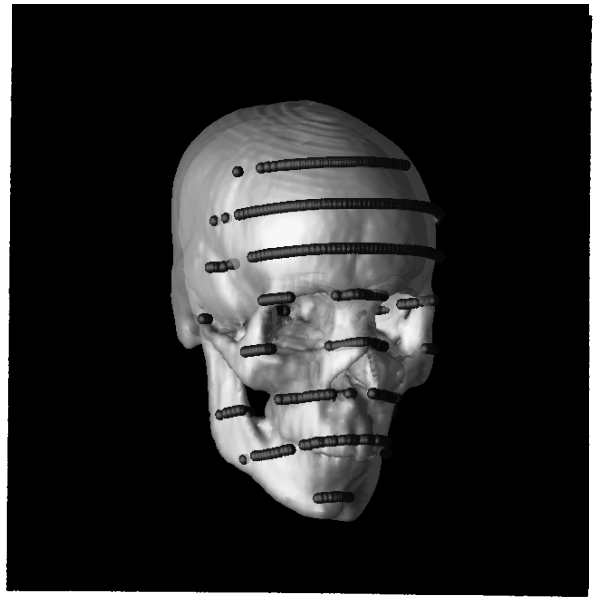


Figure 1: Example of registered laser data (shown as large dots) overlaid on CT model.

3.2 Data input

We obtain a set of 3D data points from the surface of the patient's skin by using a laser striping device. For our purposes, the laser simply provides a set of accurate 3d point measurements, obtained along a small (5-10) set of planar slices of the object (roughly 240 points per slice). 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 using the following steps.

(1) First, we sample a set of views of the model. For each view, we use a z-buffer method 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. Note that this process need not be perfect, we simply want to remove gross outliers from the data. From this data, we find three widely separated points that come from the head.

(3) We use constrained search [6] to match triples of visible sampled model points to the three selected laser points. The method finds all such matches consistent with the relative geometry of each triple, and for each 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 sampling, these hypothesized transformations are at best approximations to the actual transformation.

In examples such as Figure 1, there are typically ≈ 1000 laser sample points, and the model has typically 40,000

sample points. Given a view, and a coarsely sampled z-buffer, there are typically 1000 model points in the sampled view. In principle, there are $\approx 10^{15}$ possible hypotheses, but using simple distance constraints, there are usually $\approx 100,000$ possible hypotheses that remain.

(4) We use the Alignment Method [7] to filter out those hypotheses, by transforming all the laser points by the hypothesized transformation, and verifying that the fraction of the transformed laser points without a corresponding model point within some predefined distance is less than some predefined bound. We discard those hypotheses that do not satisfy this verification.

(5) For each verified hypothesis, we refine as follows:

(5.1) Evaluate the current pose. Thus, if ℓ_i is a vector representing a laser point, m_j is a vector representing a model point, and $\mathcal{T}$ is a coordinate frame transformation, then the evaluation function for a particular pose is

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

This objective function is similar to the posterior marginal pose estimation (PMPE) method used in [10]. This Gaussian weighted distribution 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. Because of its formulation, the objective function is generally quite smooth, and thus facilitates "pulling in" solutions from moderately removed locations in parameter space. As well, it bears some similarity to the radial basis approximation schemes used for learning and recognition in other parts of computer vision (e.g. [2]).

(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 this refinement and evaluation process using a multiresolution set of Gaussians.

(5.4) Using the resulting pose of this refinement, repeat the pose evaluation process, now using a rectified least squares distance measure. In particular, perform a second sampling of the model from the current viewpoint, using a finer sampled z-buffer. Relative to this finer model, evaluate each pose by measuring the distance from each transformed laser point to the nearest model point, (with a cutoff at some predefined maximum distance). Evaluate the pose by summing the squared distances of each point. Minimize using Powell's method to find the least-squares pose solution. Here the evaluation function is

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

where $d_{\max}$ is some preset maximum distance. This objective function is essentially the same as the MAP matching scheme of [10], and acts much like a robust chamfer matching scheme (e.g. [8]). This second objective function is more accurate locally, since it is composed of saturated quadratic forms, but it is also prone to sticking in local minima. Hence we add one more stage.

(5.5) To avoid local minima traps, randomly perturb the solution and repeat the least squares refinement. We continue, 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.

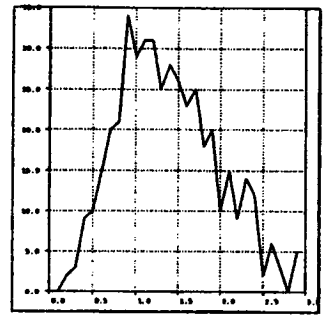


Figure 2: Histogram of residual errors for pose of Figure 1.

(5.6) The final result (Figure 1) is a pose, and a measure of the residual deviation of the fit to the model surface.

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 laser coordinate frame.

3.4 Camera Calibration

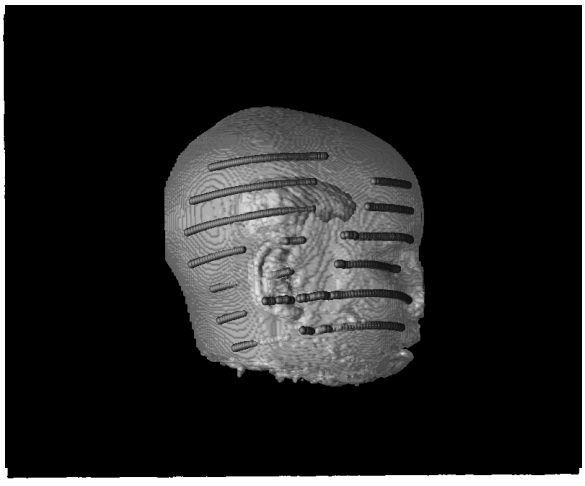
Once we have such a registration, it can be used for surgical planning. A video camera can be positioned in roughly the viewpoint of the surgeon, i.e. looking over her shoulder. If one can calibrate the position and orientation of this camera relative to the laser coordinate system, one can then render 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 [11]. This can be used for tasks such as planning a craniotomy or a biopsy, or defining the margins of an exposed tumor for minimal excision.

We have investigated two methods for calibrating the camera position and orientation, one using a calibration object of known size and shape, and one using an arbitrary object (such as the patient's head). In each case, matching 3D laser features against video images of those features allows us to solve for the position of the camera.

4 Testing the Method

We have run a series of controlled experiments, in which we have registered a CT reconstruction of a plastic skull with laser data extracted for a variety of viewpoints. In all cases, the system finds a correct registration, with typical residual RMS errors of 1.6 millimeters.

We have also 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 3. Note that in this case, 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



re 3: Example of registered laser data (shown as large overlaid on an MRI model. This is a case of registration actual neurosurgical case, with the patient fully prepped ırgery before the laser data is acquired.

Figure 4: Using the results of Figure 3, 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.

ts, and reregister the remaining data. Also displayed he internal positions of the tumor and the ventricles. RMS error in this case was 1.9mm. Finally, given the tration between the patient and the model (by matchhe laser data in this manner) we can transform the el into the coordinate system of a second video camand overlay this model on top of the camera's video . This is shown in Figure 4.

. mentioned earlier, the method has applications for cal planning and guidance, including tumor excision biopsy. The method has broader application, however, ding the registration of multiple clinical data sets such RI versus CT. A companion paper [5] discusses the cation of our method to change detection studies for ing lesion growth in patients with multiple sclerosis.

Related Work

al other groups have reported methods similar to Of particular interest are three such approaches. other groups use alternative least squares minimizanethods [9; 3] with some operator input to initialise o guide the search. A third group [1] perform regisn by matching ridge lines on surfaces.

ferences

Ayache, N., J.D. Boissonnat, L. Cohen, B. Geiger, J. Levy-Vehel, O. Monga, P. Sander, "Steps Toward the Automatic nterpretation of 3-D Images", In *3D Imaging in Medicine*, edited by H. Fuchs, K. Hohne, S. Piser NATO ASI Series, Springer-Verlag, 1990, pp 107-120.

Brunelli,R and T. Poggio, 1991, "HyberBF Networks for Real Object Recognition", *IJCAI*, Sydney, Australia.

Champleboux, G.; Lavallee, S.; Sseliski, R.; Brunie, L., "From accurate range imaging sensor calibration to accurate model-based 3D object localisation", *IEEE Conf. Comp. Vis. Patt. Recog.*:83–89, 1992

Cline, H.E. and W.E. Lorensen and R. Kikinis and F. Joless, 1990, "Three-Dimensional Segmentation of MR Images of the Head Using Probability and Connectivity." *JCAT* 14(6):1037-1045.

[5] Ettinger, G.J. , W.E.L. Grimson, and T. Lozano-Péres, 1994 "Automatic 3D Image Registration for Medical Change Detection Applications", AAAI 1994 Spring Symposium Series, *Applications of Computer Vision in Medical Image Processing.*

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

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

[8] Jiang, H., R.A. Robb and K.S. Holton, 1992, "A New Approach to 3-D Registration of Multimodality Medical Images by Surface Matching", in *Visualization in Biomedical Computing – SPIE*: 196–213.

[9] Pelizzari, C.A.; Chen, G.T.Y.; Spelbring, D.R.; Weichselbaum, R.R.; Chin-Tu Chen, "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain", *JCAT* 13(1):20–26, 1989.

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

[11] W. Wells, R. Kikinis, D. Altobelli, W. Lorensen, G. Ettinger, H. Cline, P.L. Gleason and F. Joless, 1993, "Video Registration using Fiducials for Surgical Enhanced Reality" *15th Conf. IEEE Eng. Med. Biol. Soc.*.

[12] Wells, W.M., R. Kikinis, F.A. Joless, and W.E.L. Grimson, 1993, "Statistical Gain Correction and Segmentation of Magnetic Resonance Imaging Data", in preparation.