Hierarchical Three-Dimensional Medical Image Registration

by

Gil J. Ettinger

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1997

Hierarchical Three-Dimensional Medical Image Registration

by

Gil J. Ettinger

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 1997, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

Medical imagery, in particular three-dimensional imagery, provides highly accurate and informative insight into internal anatomical structure and function. While such data is becoming readily available for the clinician, very few tools have been developed to properly exploit the structural and geometric information inherent in the images. A key problem in medical image analysis is *registration*–the process of aligning datasets to the same coordinate frame. Accurate and automated image registration facilitates powerful solutions for such applications as surgical guidance, disease diagnosis, and treatment monitoring. A hierarchical 3D surface-based optimization approach is proposed to address these registration problems. The goals are to register same-subject imagery taken over time in a wide range of situations ranging from rigid to deformable tissues and to evaluate registration accuracy and reliability. Resultant prototype registration systems have been applied to image-guided surgery, change detection, and non-invasive functional brain mapping problems and tested on clinical imagery. In order to characterize registration performance we undertake extensive data and algorithmic studies which validate the accurate and robust performance of the algorithm in a wide range of imaging conditions.

Thesis Supervisor: W. Eric L. Grimson
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

The research described in this thesis was a highly collaborative project among several members of MIT's Artificial Intelligence (AI) Lab and the Surgical Planning Lab (SPL) of the Department of Radiology at Brigham & Women's Hospital. When I started working in this area of medical image analysis over four years ago, this was a new application for the AI Lab and although I was encouraged to pursue my interests, I was initially on my own to identify specific focus areas. I was fortunate, though, to connect with Dr. Ron Kikinis, who runs the Surgical Planning Lab at Brigham & Women's Hospital, and to benefit from the interest in this area by Prof. Eric Grimson and Prof. Tomas Lozano-Perez at the AI Lab soon after I started this project. They and others deserve my appreciation as well as some of the credit for this work.

I first want to thank my advisor, Prof. Eric Grimson, who supported my efforts in many ways. He facilitated my return to MIT after having worked in industry for a number of years. He set up an excellent working environment at the AI lab which not only supported innovative computer vision research, but allowed me the opportunity to pursue new applications such as medical image analysis. And through his interest in this area I greatly benefited from his insight and dedication, not to mention his direct developments on the project. Our frequent discussions at MIT, the hospital, and even in the operating room were vital in developing many of the solutions described here.

Second, I must express my appreciation to Dr. Ron Kikinis whose Surgical Planning Lab brought together a unique combination of computational and medical expertise found in few other places. Dr. Kikinis' work in developing new image-based surgical and diagnostic tools has led to identification of many of the key problems and solution paths in this area. His insight helped define the scope of my work and my time spent at the hospital allowed me to identify and define the medical applications first-hand. The SPL provided a fertile ground for testing our algorithms on a variety of clinical data and for directly interacting with the medical specialists who may benefit from our work. I am grateful for the access to the imagery and medical specialists that the relationship with Dr. Kikinis provided. All the imagery shown in this thesis and used in the experimentation is courtesy of the Department of Radiology and related departments at Brigham & Women's Hospital, except for the mammograms shown in Section 4.3 which are courtesy of Massachusetts General Hospital.

I also want to thank Prof. Tomas Lozano-Perez whose enthusiasm and ideas guided the critical early phases of the research. He not only proposed the framework for our registration solutions, but implemented and tested a number of prototype systems. Especially since this was a completely new application for the AI Lab, the interest, energy, and guidance of both Prof. Grimson and Prof. Lozano-Perez provided continued motivation and helped ensure successful solutions.

Another vital member of our team has been Michael Leventon of the AI Lab. Over the last two years Mr. Leventon revised and greatly extended the user interface and calibration components of our registration and guidance system, resulting in a reliable system that can be used by technicians to support various head registration applications. This user interface was used in generating a number of the images included here. Mr. Leventon also spent countless hours supporting lab and operating room trials.

This work benefited from contributions from many other individuals. Dr. Sandy Wells, both at the SPL and the AI Lab, was a source of creative new ideas for solving both image registration and segmentation problems and continually supported development of our image-guided surgery application. Dr. Steve White, while at TASC, provided expertise and equipment for our laser scanner. Prof. Peter Szolovits supported this work through his Medical Informatics program at MIT. Dr. Langham Gleason, Dr. Tom Moriarty, Dr. Shin Nakajima, and Dr. Hideki Atsumi directly supported the transition of our technology into practical medical applications at Brigham & Women's Hospital. Dr. Peter Black, Dr. Eben Alexander, and Dr. Philip Stieg, of the Department of Neurosurgery at Brigham & Women's Hospital, were eager end-users of our image-guided surgery system who facilitated its use in the operating room. Dr. Laverne Gugino, of the Department of Anesthesiology at Brigham & Women's Hospital, provided the motivation and support for the non-invasive transcranial magnetic stimulation (TMS) system. Other Brigham & Women's researchers who worked on the TMS project included: Wayne Cote, Larry Sprung, Maria Karapelou and Dr. Linda Aglio of the Department of Anesthesiology; Dr. Martha Shenton and Dr. Geoff Potts of the Department of Psychiatry; and Victor Hernandez of MIT. Our multiple sclerosis change detection work was part of a larger effort at Brigham & Women's Hospital spearheaded by Dr. Charles Guttmann. I thank Dr. Bernhard Geiger, formally of INRIA, and Bill Lorensen, of General Electric, for providing surface rendering tools. The extensive data processing efforts at the SPL have been supported by Mark Anderson, Marianna Jakab, and many others. I also thank Tina Kapur, Aparna Lakshmi Ratan, Mike Leventon, Pam Lipson, Liana Lorigo, and Sandy Wells for lending me CPU cycles on their workstations for the many hours of experimentation I performed on this thesis.

Dr. Nicholas Ayache, of Projet Epidaure at INRIA, France, graciously hosted me for a month's visit at his research lab which is doing state-of-the-art medical image analysis work. This visit was highly beneficial, educational, and enjoyable. I appreciate having had the opportunity to partake in this exchange.

MIT, and in particular the AI lab, has been an exciting and productive place for me to pursue my education. I am grateful for the many opportunities to learn and grow at MIT of which I have taken advantage throughout my undergraduate and graduate careers here.

My acknowledgments would not be complete without my expression of gratitude to my family. My lovely wife Linna has been so supportive, understanding, and helpful throughout my time at MIT. Sharing my life with her has not only made this period more enjoyable, but has enriched our many other shared experiences. My daughters, Alyssa and Sharon, have been a constant source of happiness and have made coming home something to look forward to and relieve the day's stress. None of this would have of course been possible without my parents whose constant love, support, and emphasis on education is most appreciated. I love them all very much.

This research was supported in part by DARPA under ONR contract N00014-94-01-0994, in part by the Medical Informatics training grant No. T 15 LM 07092 from the National Library of Medicine, and in part by a Scottish Rite Schizophrenia Grant.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical imaging, and in particular 3D medical imaging, is a quickly growing field which has seen a remarkable proliferation of 3D imaging scanners since the advent of the computed tomography (CT) scanner in 1972. There are now more than 10,000 CT and 3800 magnetic resonance imaging (MRI) scanners in use in North America for detailing internal anatomical structure. Other sensors, such as single photon emission computed tomography (SPECT), are clinically used for functional imaging and new 3D imaging techniques, such as positron emission tomography (PET), 3D optical, 3D ultrasound, and biomagnetism are being developed and deployed. With all this data becoming readily available for the clinician, very few tools have been developed to properly exploit the structural and geometric information inherent in the data. The most common mode of utilizing these 3D datasets is to manually view the separate 2D slices on a lightwall, as is shown in Figure 1.

Tools for better leveraging the structural and geometric information inherent in 3D medical imagery will have a pronounced impact on the delivery of health care. Some of the applications which may benefit from improved use of medical imagery include:

- **Disease diagnosis**. Early detection of disease is often the key to effective treatment. Imagery of internal organs provides evidence of particular pathologies which are manifested through structural or functional changes—for example, tumors, multiple sclerosis lesions, or ligament damage. An example of multiple sclerosis lesion growth detection is shown in Figure 2, where MR images of an individual have been taken at several time points and registered to a uniform coordinate frame. The high resolution and high contrast available with today's scanners support diagnosis of a wide range of problems. Such diagnosis can be improved not only with tools which can highlight such pathologies in single images, but also with tools which allow accurate comparison against baseline images.
- **Treatment monitoring**. By taking multiple images of a subject over time a clinician can effectively track progress or regress of a medical condition and quickly determine impact of treatment regimens. By using initial imagery as a baseline, such subject tracking results in interpretations that are specific to an individual, an important consideration in situations in which “normalcy” covers a wide range of characteristic appearances.
- **Multi-source information integration**. Since different scanners highlight different tissue types and different functions, it is often beneficial to fuse information from different scanners, such as vasculature from angiograms, soft-tissue structures from MRI, or metabolism activity from SPECT. Differing resolutions, aspect ratios, possible distortions as well as differences in coordinate frames pose difficult problems in interpreting combined images. An example of registering a conventional MR image with vasculature from an

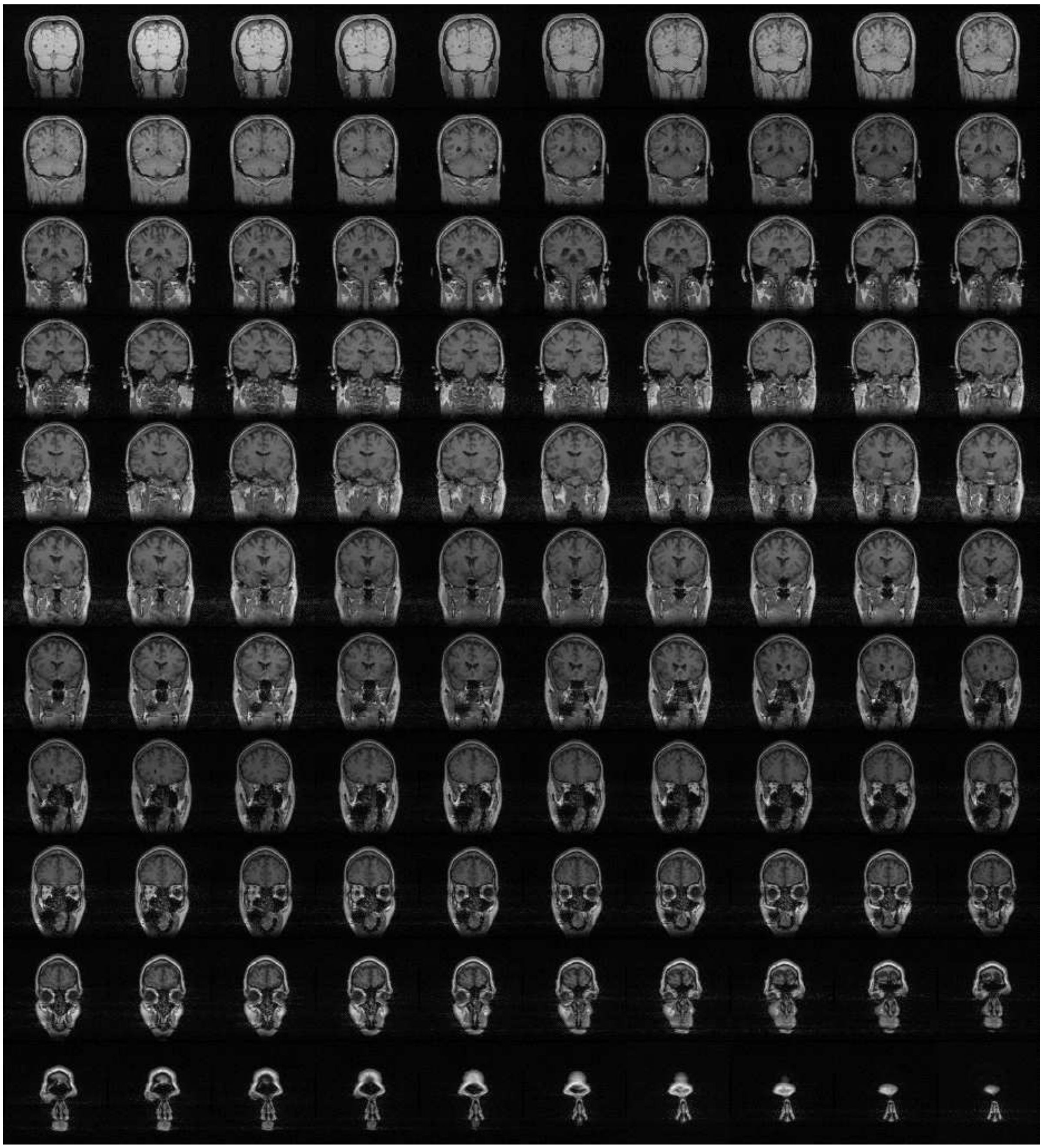


Figure 1 Sample MR image showing the individual slices acquired by the scanner. Such 2D displays are a common mode of examining the inherent 3D structural information in the image.

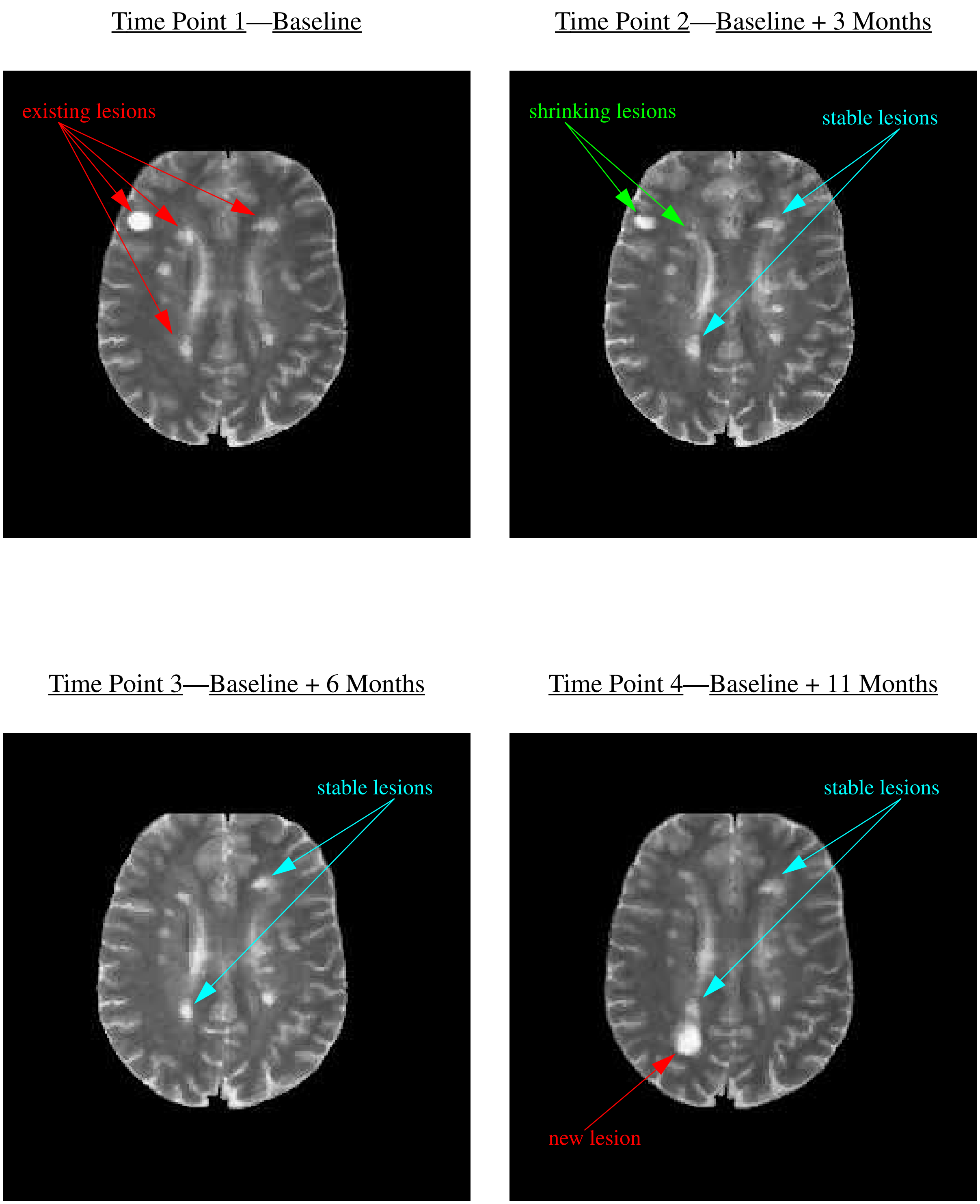


Figure 2 Tracking of multiple sclerosis lesions in MR images of the same subject during the course of a year. Tracking of changes of the same few lesions is shown.

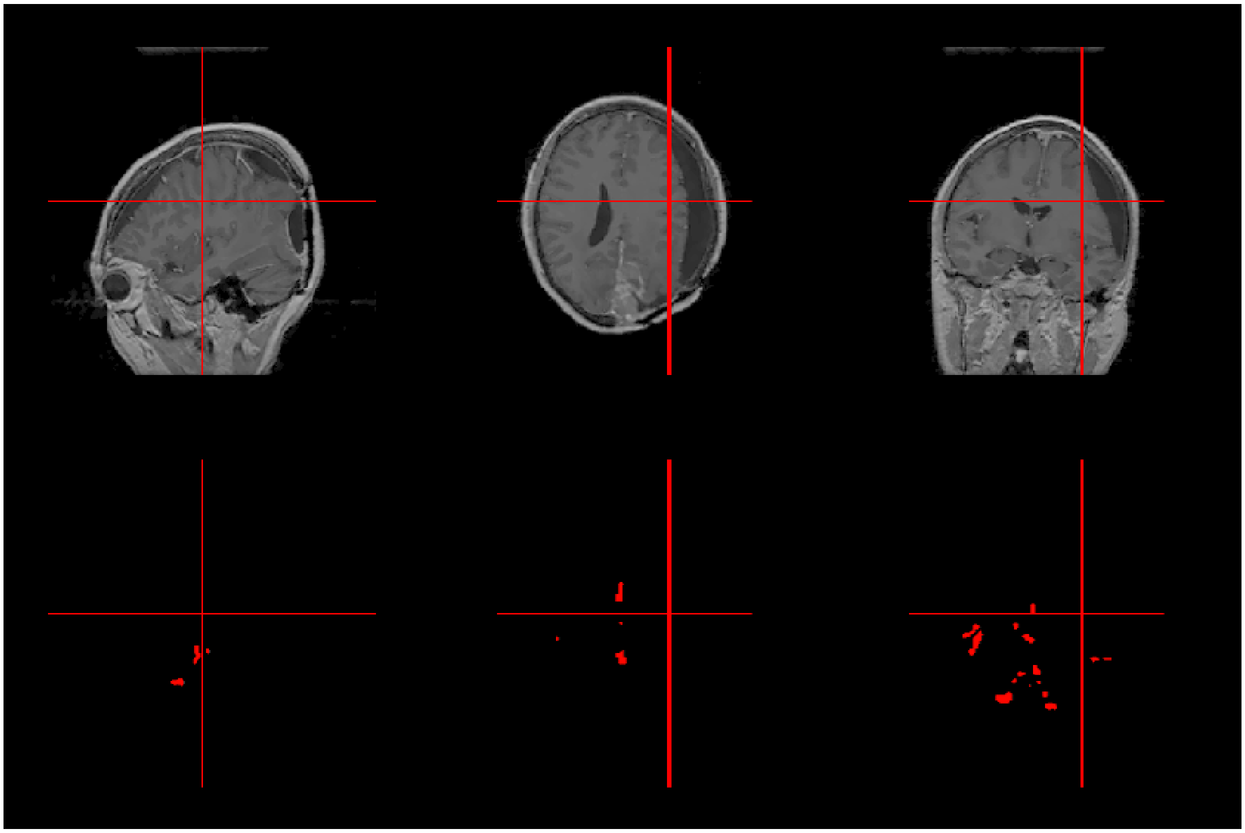


Figure 3 Registration of a conventional MR image (top) with vessels segmented from an MR angiogram (bottom). In both cases three orthogonal slices passing through the interrogated point are shown. The cross-hairs identify corresponding points in both scans. Structures from the different modalities can be overlaid on each other as well, but this display ensures that information from one modality is not obscuring the other modality.

MR angiogram is shown in Figure 3.

- **Surgical field-of-view visualization**. Medical imagery can supplement the surgeon's visual field by providing views of otherwise hidden structures. In many surgeries the surgeon has a limited field-of-view into the operating field. In minimally-invasive surgeries the surgeon is also confronted with difficult hand-eye coordination problems since he is looking at a camera’s view of the surgical field with a totally different reference frame than his own. In other cases, surgeries are complicated by the same visual appearance of different tissues (such as tumor and healthy tissue) although such tissues have high contrast in some medical imagery. Better use of the 3D imagery can improve on all these surgical problems. In particular, *enhanced reality visualization*s, in which the surgeon's field-of-view is augmented with additional structural information, can provide useful guidance in planning and executing the surgery. Such visualizations are generated by overlaying registered rendered anatomical structures onto the surgeon's view of the patient. An example of an enhanced reality visualization is shown in Figure 4.
- **Instrument guidance**. Another application for bringing the imagery into the surgical pro-

Figure 4 An enhanced reality visualization of the internal tumor (green) and ventricles (blue) for a neurosurgery patient. The MR image was registered to the patient's pose on the bed and the internal structures overlaid on a calibrated video view of the patient. In this case the visualization was used for surgical planning—identifying the preferred craniotomy location.

cess is to track surgical instruments relative to prior scans of the patient. For example, the correct placement of biopsy needles and minimally-invasive surgical tools such as laproscopes and remote cutting tools is problematic in limited fields of view. By tracking the 3D pose of these instruments, in conjunction with tracking the 3D pose of the patient, the position of the instrument can be shown on a registered scan of the patient, such as MRI or CT, as shown in Figure 5.

- **Robot-assisted surgery**. In addition to providing guidance to a human surgeon, imagery can be used to guide robotic tool placement. Robotic path planning will generally be performed on a prior MRI or CT scan and will follow that plan after the scan has been registered to the patient.
- **Anatomical modeling**. A novel use of medical imagery is the construction of 3D surface models for surgical planning. Numerical control machines and stereolithography systems can build exact replicas of internal anatomical structures, such as the skull or spine bones, which may then be used in planning corrective surgery.
- **Surgical training systems**. Medical imagery may be used to construct surgical simulators with which students may train for a wide range of conditions. In conjunction with virtual reality technologies such simulators may provide realistic training grounds.

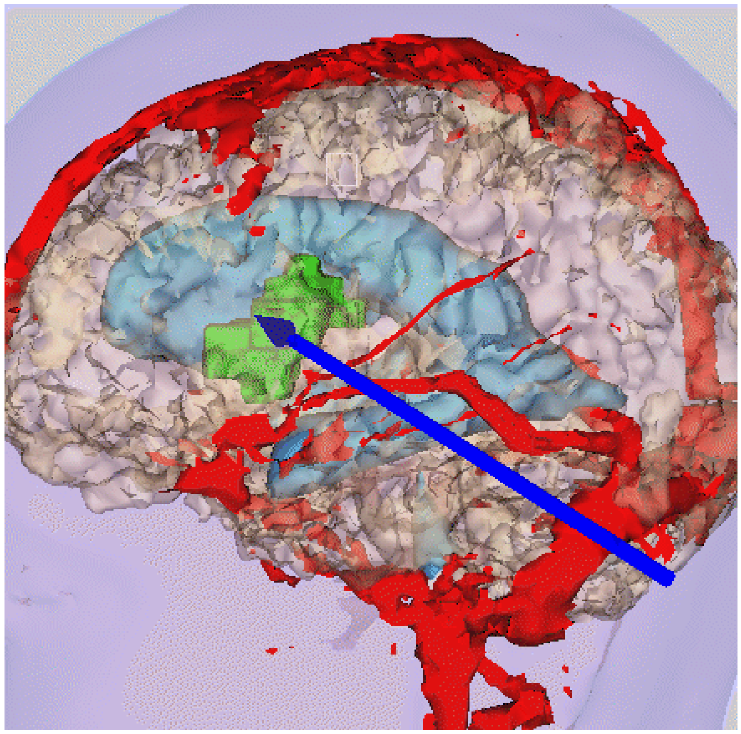


Figure 5 Location and orientation of an instrument from the surgical field (blue) relative to a registered rendering of structures segmented from the patient's MR scan: brain (white), tumor (green), vessels (red), and ventricles (cyan).

Image analysis tools needed to facilitate these application improvements must tackle a number of technical problems:

- **Image segmentation**—classifying the voxels, volumes, or surfaces in the imagery as belonging to particular tissue types, such as bone, fat, vessel, white matter, grey matter, and tumor. Such classification is based on the ability of the sensor to discriminate between tissue types as well as anatomical knowledge of tissue structures and relationships.
- **Image registration**—transforming the raw or segmented imagery to new coordinate frames in which the geometry of the structures is much more usable. Such transformations are required for fusing information from multiple imaging modalities or mapping diagnoses made in 3D images to the physical diseased structures in the patient.
- **Visualization**—rendering and displaying 3D structures such that medical personnel can readily view and interpret the data. Image analysis tools must work synergistically with the clinician, providing sufficient interaction, response, and feedback capabilities.

For most applications, techniques addressing all these problems must be combined to achieve practical solutions. In this thesis, we focus on the development of registration techniques, although segmentation and visualization techniques are integrated as well. Registration often requires high accuracy as well as consideration of the 3D dataset as a whole—i.e., registration must exploit the whole volume as a single entity. As a result, interactive registration in 3D, by manually moving datasets around until they align, is a nearly impossible task. Corresponding landmark points may be identified in the datasets to be registered, but this process relies heavily on the reproducibility of landmark specification. Instead, we base the registration on tissue boundary surfaces, such as scalp or intra-cranial cavity. Three dimensional surfaces offer a reliable and computationally attractive means for registration and are available from most 3D imaging modalities. They allow specification of regions which are expected to remain stable across registered images, support hierarchical processing strategies for increased performance robustness, and support direct measurement of physical distance between datasets. We develop automated surface-based registration solutions which are robust to data variability, such as: imaging distortions, preprocessing perturbations, data outliers, varying resolutions, and initial pose offsets.

1.1 Registration Problem Overview

The registration problem we are considering is depicted graphically in Figure 6. Two 3D surfaces extracted from possibly *different modalities* and at *different times* are extracted from imagery of the *same subject*. We need to compute a 3D transformation, Γ, which transforms points in coordinate system 1 to coordinate system 2. The transformation may be rigid, consisting of six degrees of freedom, or higher order, such as affine, consisting of 12 degrees of freedom. By accurately solving for this transform we may view structures segmented from both images in the same coordinate frame allowing: diagnosis and monitoring of changes over time, detailed proximity analysis of structures visible in different modalities, and guidance of surgical procedures towards specific targets.

In order to solve this problem, our registration goals are:

- **Automated processing—**minimize manual intervention for ease-of-use and high accuracy. Verification tools should provide feedback on the registration accuracy.
- **Pose-space optimization—**model the problem as a minimization of a distance metric between datasets instead of a feature correspondence search. Such an optimization approach avoids the potential problems involved with stable feature extraction and resultant combinatorics of correspondence search. We use surfaces to identify salient and stable structures on which to base the registration solution.
- **Robust performance—**minimize the impact of data deviations. The registration problem is complicated by a number of factors which must be handled by proposed solutions. These factors include:
    - Differing coordinate frames in datasets being registered, resulting in possibly large initial pose offsets.

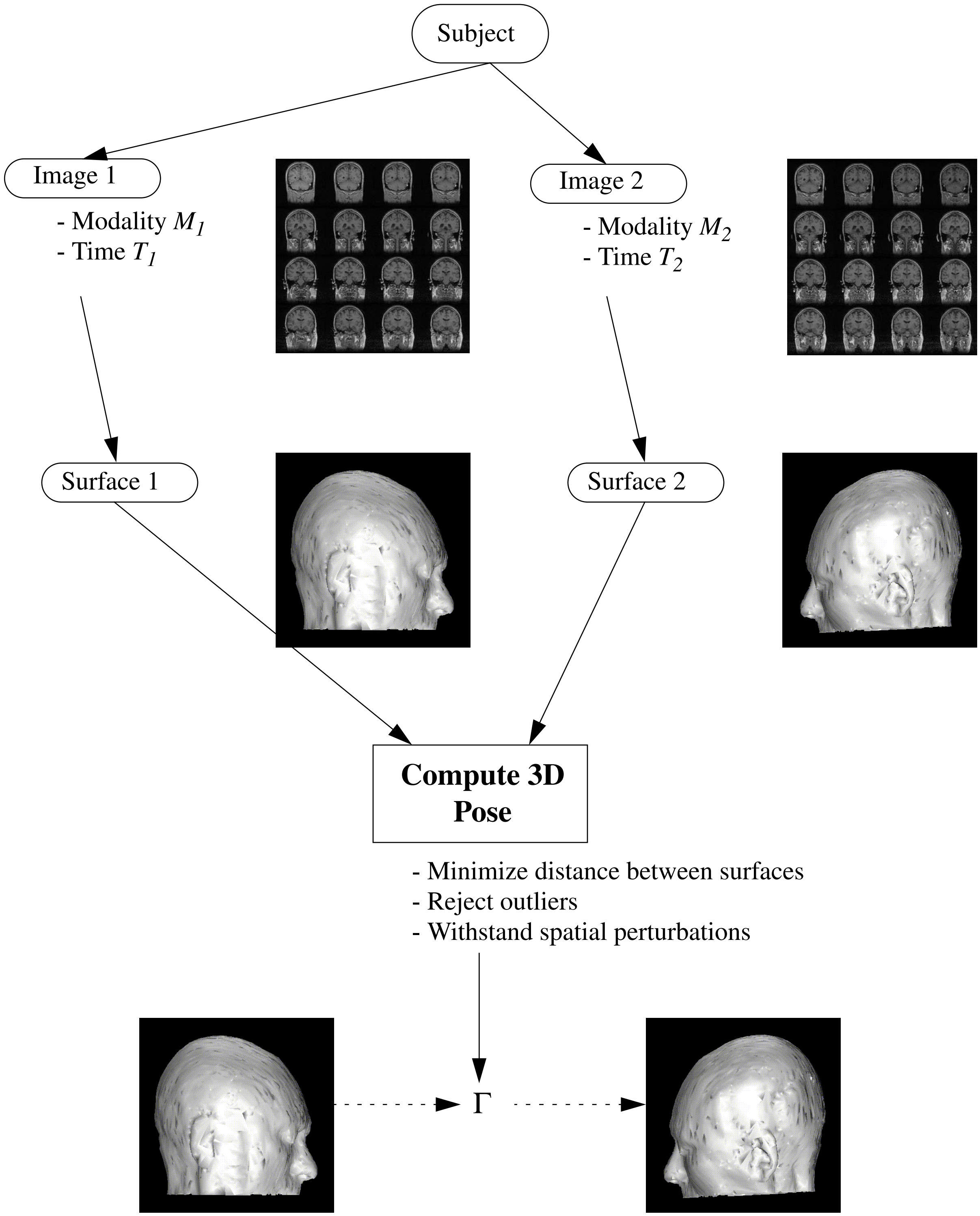


Figure 6 Registration problem overview.

- Sensor distortions which are different across images.
- Differing spatial resolution.
- Different structural coverages—the images may not cover the same physical volumes.
- Tissue may be displaced, articulated, or deformed in one of the images relative to the other.

The proposed solution to these problems consists of a hierarchical approach in which refinements are made along two dimensions:

- **Measurement resolution**—coarse-to-fine representation of surfaces and the metrics measuring distances between them. Coarse metrics are designed to pull in solutions across large pose offsets while not adversely impacted by outliers, missing surface patches, or spatial perturbations. Once close to the final solution, fine metrics are designed to accurately localize the pose which minimizes a more sensitive distance measure taken between the surfaces.
- **Pose complexity**—rigid to deformable transformation of data. While research in other domains is underway in characterizing the nature of tissue motion under a wide range of internal and external forces, models are not yet available to quantify their effect under the many situations encountered in medical applications. As a result, we attempt to minimize the impact of non-rigid distortions, displacements, and deformations by solving for low order alignment transformations which robustly handle such perturbations.

Through these techniques we achieve a smooth objective function which supports effective pose optimization. Local minima are avoided and a large pose region of convergence is achieved, even in the presence of a wide range of data deviations. Through thorough experimentation we have defined our performance limits, thus facilitating practical application to medical problems of interest. As a result we have been able to successfully apply our registration solutions to a number of diagnosis and therapy applications, which have been tested on clinical data.

1.2 Contributions of Research

The 3D medical image registration research described in this thesis develops the following innovations:

1. **Robust 3D surface registration algorithm**. We developed a class of hierarchical 3D surface registration techniques which are well suited to a wide range of medical image analysis applications. The registration algorithms are designed to achieve smooth objective functions which may be effectively optimized over pose solutions. They have been shown to behave robustly in the presence of wide ranges of data variations.
2. **Demonstrated applicability in high-utility medical applications**. We successfully developed a number of medical image analysis applications whose resultant leveraging of 3D imagery may effectively improve health care practice. All applications have been tested on clinical imagery, in some cases directly in the operating room. We are one of the

first groups to have incorporated highly automated and effective image registration into the applications of:

- enhanced reality visualization and instrument tracking for image-guided surgery,
- structural change detection for diagnosis and treatment monitoring, and
- non-invasive functional brain mapping via transcranial magnetic stimulation.

3. **Detailed analysis of registration performance**. We provide a framework for and have conducted extensive experimentation to evaluate the robustness of the registration algorithm relative to:

   - magnitude of pose offset between datasets to be registered,
   - measurement noise,
   - number of outliers,
   - degree of surface coverage,
   - data resolution.

   The results of these experiments have not only shown effective registration performance, but have also given us an understanding of the benefits and limitations of the registration techniques. The resultant algorithm characterization supports effective satisfaction of specific application requirements.

1.3 Roadmap of Thesis

In Chapter 2 we describe the registration problem we tackle, namely intra-subject 3D surface registration. Chapter 3 details the registration algorithm and overviews related registration approaches. In Chapter 4 we describe several medical image registration applications we have developed and demonstrated on clinical data. Chapter 5 provides the results of the detailed sensitivity and accuracy experimentation we have performed on the registration algorithm.

Chapter 2

Registration Problem Description

We focus on the problem of registering multiple 3D datasets of the *same* subject. The 3D datasets may be acquired from multiple modalities and at different times, with physical differences in the imaged structures modeled by either rigid or affine transformations. The imaging modalities, though, may introduce measurement noise, distortions, and incomplete data overlaps which must be handled robustly even if we do not solve for them explicitly. In this chapter we define the key issues in solving the registration problem. We first characterize the medical data with which we work—mainly MRI. Then we describe the registration output in terms of the classes of transformation for which we solve. We motivate the use of surfaces as the basis for registration and conclude with a specification of performance goals.

2.1 Data Characterization

For this thesis we primarily deal with magnetic resonance imagery (MRI). For the applications we develop—neurosurgery, brain mapping, and mammography—the 3D scanner of choice is MRI, due primarily to its high soft tissue contrast. In addition, MRI has minimal known side-effects and provides for a multitude of useful imaging variations. MR images can be tailored to highlight specific tissue types, to use arbitrary slicing planes, and to image blood flow and blood oxygenation. MRI measures the density of atomic nuclei magnetic spins induced by external magnetic fields at specific frequencies. Only nuclei with an odd number of protons or neutrons exhibit such magnetization, but a large number of materials have such properties. For imaging the body, which contains a large amount of water, MRI frequencies are generally tuned to image ^{1}H atoms. The MR image is usually acquired as a set of contiguous slices to yield a volumetric description of magnetization density. Voxel resolution within a slice is usually square, on the order of 1mm, and slice thickness is about 1-3mm. Higher resolution can be achieved, but requires longer acquisition times and a trade-off against signal-to-noise ratio (SNR). A wide range of acquisition parameters yields different acquisition times, ranging from a few minutes to about half an hour. See [15] and [37] for overviews on the physics and mathematics of 3D imaging in general and MRI in particular.

While MRI has many attractive advantages, it suffers from possible geometric distortion which may be caused by several factors:

- **Magnet calibration errors**. The MR scanner consists of a constant-field main magnet and varying-field gradient coils. Scaling errors and geometric distortion occur if the produced magnetic fields vary from the intended field strengths. Techniques for shimming the main magnet and calibrating the complete apparatus have minimized the impact of magnetic field errors. Eddy currents may also be introduced by the gradient coil, but proper shielding minimizes their effect.
- **Chemical shift**. Hydrogen atoms in water and fat respond at slightly different frequencies which results in spatial shifts between fat regions and soft tissue regions. These artifacts, though, can be eliminated by MR pulse sequences which selectively eliminate the fat signal for imaging the water signal only.
- **Magnetic susceptibility variations in the anatomy**. Magnetic susceptibility refers to the induced magnetic field in materials in the presence of an external field. This induced field acts to vary the overall magnetic field at each point in the volume. Since different materials may have different magnetic susceptibilities they may appear offset relative to each other in one direction, the frequency-encoding direction, of the image. This effect is particularly apparent at boundaries of materials with high susceptibility differences. Air/tissue boundaries are a common example of this phenomenon in body imaging, leading to a possible offset of up to 2mm for typical image acquisition parameters. These effects are difficult to correct since they are dependent on the imaged structures, but techniques have been developed to minimize their impact by taking multiple scans [55].

Continued improvements in MR imaging will reduce their geometric distortion, yield higher resolution images, and provide better image contrast. These improvements will simplify registration solutions, particularly approaches based on 3D surfaces such as the ones we develop here. But geometric distortion and resolution differences cannot always be neglected—particularly when registering images from other modalities to MRI. In such cases these artifacts may be handled through explicit deformation or through implicit allowance for measurement noise.

2.2 What is the Result of Registration?

For the registration problem we are considering, the input consists of two 3D surfaces, a reference surface and a test surface that we wish to transform to align with the reference surface. The output is a coordinate transformation matrix or pose, Γ, such that $\Gamma\bar{P}_i^1$ transforms the points, $\bar{P}_i^1$ from the test coordinate frame, surface 1, to the reference coordinate frame, surface 2. The transform, Γ, is at a minimum a rigid 3D transform. Resolution is generally known, so scale is not a required output in many applications. The rigid 3D transform consists of three rotational degrees of freedom, $\{\theta_x, \theta_y, \theta_z\}$, namely rotation about the three orthogonal axes of dataset 1, and three translation degrees of freedom, $\{T_x, T_y, T_z\}$, namely translation along the three orthogonal axes of dataset

1. The rigid transformation matrix, Γ_{rigid}, expressed in homogeneous coordinates is:

$$\begin{bmatrix} \cos\theta_y\cos\theta_z & -\cos\theta_x\sin\theta_z - \sin\theta_x\sin\theta_y\cos\theta_z & \sin\theta_x\sin\theta_z - \cos\theta_x\sin\theta_y\cos\theta_z & T_x \\ \cos\theta_y\sin\theta_z & \cos\theta_x\cos\theta_z - \sin\theta_x\sin\theta_y\sin\theta_z & -\sin\theta_x\cos\theta_z - \cos\theta_x\sin\theta_y\sin\theta_z & T_y \\ \sin\theta_y & \sin\theta_x\cos\theta_y & \cos\theta_x\cos\theta_y & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We allow for deformations by introducing an affine transformation, Γ_{affine}, which replaces the rigid rotation with scaling and shearing parameters to deform a dataset in a linear fashion. $\Gamma_{\text{affine}} = \Psi_{\text{scale}}\Psi_{shear}\Psi_{\text{translate}}$, where Γ_{scale} introduces independent scaling along the coordinates axes:

$$\begin{bmatrix} \alpha_x & 0 & 0 & 0 \\ 0 & \alpha_y & 0 & 0 \\ 0 & 0 & \alpha_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Γ_{shear} introduces shearing:

$$\begin{bmatrix} 1 & \beta_{xy} & \beta_{xz} & 0 \\ \beta_{yx} & 1 & \beta_{yz} & 0 \\ \beta_{zx} & \beta_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\Psi_{\text{translate}}$ is the translation transform:

$$\begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The resultant affine transformation matrix, Γ_{affine}, is thus composed of 12 degrees of freedom as opposed to the rigid transformation with consists of only 6 parameters.

Higher order transformation solutions may be obtained as well, although we question their utility for registration of same-subject imagery and for surface-based registration. In these cases we do not necessarily have a sufficient data sampling within the whole data volume to support higher order solutions and by using surfaces we specifically base the registration on structures that we expect to remain stable over time or across modalities. Instead we suggest the application of the same low-order transformation solutions on local data neighborhoods which may be partitioned according to expected structure articulation or independently moving structures.

Higher order transformations, though, may be necessary for the related problem of alignment of imagery taken from *different* subjects for such purposes as atlas-guided segmentation. In this problem a prototypical scan obtained from a single subject or averaged across several subjects is manually segmented at high fidelity to generate a detailed atlas. This atlas is then registered to new subjects in order to provide structural cues for driving the segmentation of the new subjects. While such a problem may be solved using similar techniques as for the intra-subject registration problem, it is characterized by potentially large differences in physical tissue appearance. Even though anatomical topography is generally consistent across subjects, differences in organ size and geometric detail introduce a high degree of complexity for which it is not clear what type of warping function works best. In particular, it is not clear how pathologies, which are to be diagnosed by such inter-subject registration techniques, affect the warping strategies undertaken by the registration algorithms. This different subject registration problem is not specifically addressed in this thesis.

2.3 Surface-based Registration

Related three-dimensional registration approaches have been based on features (points [58] or curves [31]), surfaces [49][56], and volumes [16]:

- Feature based approaches extract points or curves that are likely to be stable and descriptive and then search for consistent correspondences between the features extracted from both datasets. Sample features are points with maximum Gaussian curvature along isosurfaces of volumetric data.
- Surface registration approaches generally align the two datasets by searching in pose space, namely the components of the transformation, instead of correspondence space, namely the associations of test features to reference features. Surface registration solves for optimal poses which minimize a distance metric between the two surfaces.
- Volume registration approaches also search in pose space, but they minimize a correlation metric computed on the whole volume.

We base the data registration on 3D *surfaces*, rather than features or volumes, for the following reasons:

- Some of the data sources we use produce only surface data and not volumes. For example, the laser scanner, which we use to localize a subject in the surgical planning or operating environment, produces 3D points on the subject's skin surface.
- Surfaces define tissue boundaries along which (1) changes are often expected, or (2) constraints limit expected change. By specifying the particular surfaces to use as a basis for registration, we thus specify the regions we want to lock down. Such registration may then (1) recover particular deformations of interest or (2) lock down regions known to remain rigid in order to look for changes in internal structures where deformations may have happened.

- Surfaces offer a reliable and computationally attractive means for registration. Although surfaces are generally one processing level removed from the raw data—an isosurface routine is required to extract them from volumes—that processing is relatively stable. Furthermore, surfaces may be easily smoothed to provide a hierarchical coarse-to-fine registration mechanism which maintains geometric consistency of the underlying data.

  On the other hand, registration of a complete raw volume is generally performed in intensity space resulting in a large amount of data to match and difficulties in performing hierarchical registration since smoothing and sub-sampling in intensity space may not preserve geometries of imaged structures.

  The third alternative, feature matching, may reduce the complexity of the matching process, but at the cost of feature extraction and correspondence search. Since feature extraction involves several processing steps, such as multiple derivatives for points of high curvature, it is susceptible to noise and may thus require high signal-to-noise ratios as well as high resolution. Depending on the reliability of the feature extraction process, the correspondence search required for performing the registration may also be computationally expensive.

- Surfaces support the measurement of physical distance between two datasets to be matched. Such distances are generally the direct property we wish to minimize. While feature-based approaches use similar metrics, volume matching approaches generally measure correlation (possibly non-linear) between voxel values, a measure of how different the two datasets are, but not necessarily how far apart they are.

We acquire the surface data needed for the registration by two methods:

- Volumetric scanning. The reference dataset is imagery from a medical scanner, such as MRI. The test dataset may be volumetric as well, or generated by surface scanning. The raw data is volumetric, composed of a sequential set of usually contiguous 2D slices. The steps in processing this data are:

  1. Segment the data to classify the voxels into particular tissue types.
  2. Generate an isosurface of the tissue type of interest.
  3. Represent the surface as a set of triangles resulting from the isosurface or a set of points consisting of the triangle vertices.

- Surface scanning. In applications requiring registration of a volumetric scan to a subject's current position we extract skin surface data by using a laser scanner. Other 3D surface scanners, such as probes equipped with optical triangulation systems or magnetic field transmission/reception systems, may also be used, especially for obtaining internal organ surface data in surgery. All these systems directly produce 3D point data of the surface being scanned or probed.

2.4 Automated Robust Registration

One of the key aspects of medical image registration is the need for *automated* registration processing. At issue are not only efficiency concerns, but also accuracy considerations. Manual solution of 3D pose—i.e., manual rotation and translation of a 3D object to align it with another—is a difficult manual manipulation when dealing with six degrees of freedom and becomes almost impossible with higher order transformations. Another common manual solution is the specification of fiducials on which to base the registration. But manual specification of fiducials relies on the accuracy of specifying those points and if a sufficient density of points in the region of interest is not provided such registrations may be difficult to verify for correctness. Efficiency is also an important consideration since surgical applications require effective feedback, and diagnostic applications may require registration of high volumes of data—all further indications of the need for automated processing.

In performing the registration automatically using a surface-based pose optimization approach several issues must be addressed:

- **Robustness in the presence of input data errors.**

  The registration technique must limit the impact of data deviations by identifying key data geometries and detecting possible outliers. Data deviations may result from:

  - surface extraction deviations—e.g., data segmentation errors or surface sampling artifacts;
  - data outliers—e.g., points from spurious structures or surface points present in one dataset but not the other;
  - imaging distortions.

- **Effective performance for varying data resolutions.**

  Since medical scanners as well as surgical tracking tools produce data at a variety of resolutions, registration algorithms must support these variations. Furthermore, since the surface registration approach is structured as a hierarchical process, support for coarse-to-fine resolutions is inherent in the registration method itself.

- **Minimum dependence on initial alignment.**

  Given the pose optimization approach, an initial pose is required to seed the registration. The pose region of convergence of the registration algorithm is defined as the extent to which the initial pose may differ from the optimal pose such that the registration process still converges on that optimal pose. This region of convergence must be sufficiently large to allow a coarse initial alignment. In many medical applications the solution pose is approximately known since subjects are positioned in scanners or in the operating room by medical practitioners. A rough initial alignment may be made using this positioning from which the registration algorithm must converge.

  Additionally, techniques to automatically generate initial alignments may be incorporated using such techniques as interpretation tree search [28] on selected sub-sampled surface points.

- **Avoidance of local minima.**

  Optimization approaches are susceptible to convergence on local minima when optimizing over highly non-convex objective functions, as is generally the case in registration. Techniques such as hierarchical search must pull in solutions from distant points in pose space as well as accurately isolate the global optimum from nearby local minima.

- **Verification of registration results**.

  In performing the registration automatically, the registration process must differentiate acceptable solutions from invalid ones. This process should include qualitative validation tools for visual inspection of the registration results as well as a sufficiently meaningful matching metric to allow the registration algorithm to identify whether the results generated by the algorithm may be acceptable.

In order to evaluate the registration algorithm's ability to address these issues we perform extensive experimentation on MR data which has been perturbed with controlled variations. A selected MR image is taken as a reference dataset. We then register to the reference image a test dataset which is constructed by copying and perturbing the reference image. By using such simulated test images we know both the correct ground truth pose and the magnitude of the applied perturbations. We measure pose error after registration as a function of different perturbations that are applied to the test data prior to registration. The following types of perturbations are applied:

- Outlier addition: Adding random outlier surface points to the test data.
- Surface patch removal: Removing a localized neighborhood of test surface points
- Measurement error: Perturbing the 3D locations of the test surface points from their true locations.
- Data resolution: Decreasing the underlying resolution of the data from which the test surface is extracted.
- Pose offset: Increasing the pose offset between the test and reference datasets.

In addition, we study the relationship of our registration score to pose error in order to support verification metrics.

These simulation experiments are complemented by application experiments in which both the reference and test datasets are representative clinical data. In these cases we do not have exact ground truth pose, but can still measure registration sensitivity to the same type of controlled deviations from an estimate of ground truth. Finally we perform the same type of experiments using different algorithm parameters in order to evaluate the benefits and limitations of particular algorithm configurations.

Chapter 3

Registration Solution Approach

Our registration approach to solving the surface registration problem identified in Chapter 2 is based on hierarchical pose space optimization. Under this approach we continually vary the trade-off between (1) pose localization accuracy and (2) robustness to data perturbations. Robust performance, where outliers and data deviations do not grossly impact results, can be achieved through registration metrics which do not greatly discriminate between similar solutions, thus reducing the impact of possible variations for which we are not directly solving. On the other hand, such metrics generally do not achieve high accuracy since the optimal solution is not highly distinctive. We thus start the registration process with coarse metrics to pull us towards the optimal solution, but continually vary the registration metrics to more discriminating measures that can accurately localize the correct pose even as higher order solution options are introduced.

This hierarchical pose optimization approach is encapsulated in a registration process that consists of the following steps:

1. Surface data extraction—process datasets to generate corresponding surfaces to be registered to each other.
2. Initial alignment—perform a coarse initial match of the surfaces.
3. Hierarchical rigid alignment—solve for rigid alignment of surfaces by sequentially increasing measurement resolution.
4. Flexible registration—solve for affine alignment of surfaces by allowing independent scaling and shearing deformations.

The registration process is depicted graphically in Figure 7. Each of the steps shown is described below.

3.1 Surface Data Extraction

Input data is either volumetric data, originating from a 3D medical scanner, or direct surface data, generated from a 3D localization scanner such as a laser scanner or an optically tracked probe. Direct surface data can be used as is, generally represented as a set of 3D points, possibly with associated normals. Volume data is processed in two steps:

1. Segment the data—label individual voxels in the 3D imagery by tissue type.
2. Extract the bounding surface of the tissue of interest.

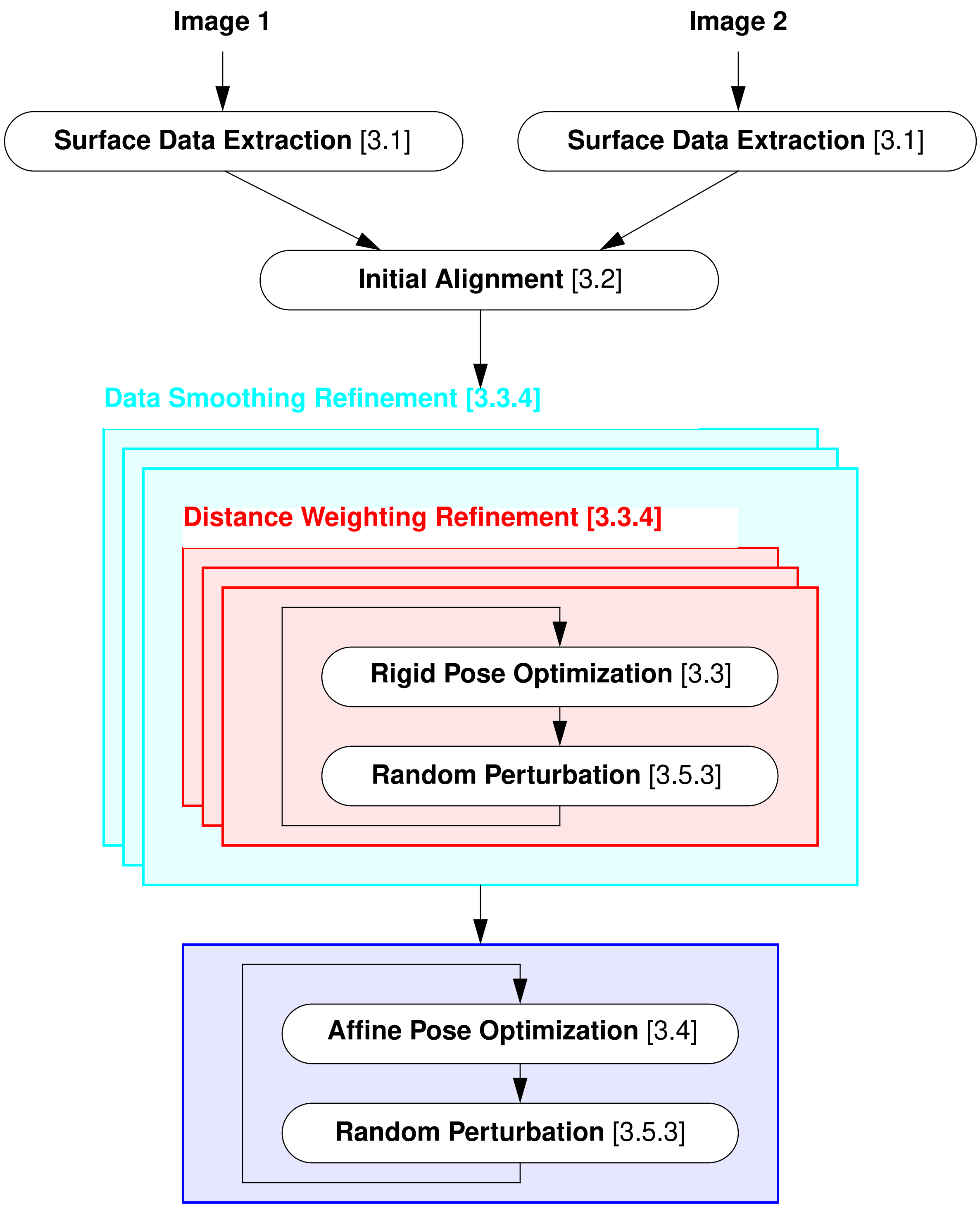


Figure 7 3D registration overview. Number in brackets indicates section in which function is described.

The segmentation process is generally a sensor-specific procedure based on properties of the observed intensities as well as known anatomical information about the imaged organs. While detailed segmentation of internal structures can be a tedious process, for registration we are often interested in bounding surfaces of structures which can be extracted from the imagery in a straight-forward manner. Skin surface, for example, can be extracted using the following automated procedure in relatively noise-free images:

1. Threshold the data to separate background from tissue response.
2. Apply erosion followed by dilation morphological operators [53] to remove background articulation points corresponding to orifices in the skin.
3. Compute connected components of background labels to remove background components internal to the skin surface. Some internal structures, such as sinus cavities, bones, and fluids, may not produce a response in some scanners for selected imaging parameters, thus appearing the same as background. If the registration is aligning just the external surfaces, such internal structures are removed by this step.

In order to extract internal surfaces, such as intra-cranial cavity, either for registration or subsequent visualization, more powerful segmentation techniques are necessary. For such purposes, we use segmentation techniques developed at the Surgical Planning Lab of Brigham and Women's Hospital. These techniques consist of a gain-correcting classifier by Wells [64] and a brain outlining tool by Kapur [39]. The gain-correcting classifier uses statistical knowledge of tissue properties and gain inhomogeneities to dynamically solve for two components at each voxel: tissue labeling and the gain field. The Expectation-Maximization algorithm is used to iterate the solution of the two components. The brain outlining tool applies morphological operators to separate the brain tissue from the rest of the head. The resultant segmentation successfully adapts to intensity variations that are due to spatial inhomogeneities in the sensitivity of the imager, intensity variations associated with different scanners, and those due to the varying appearance of tissues across patients. As a further improvement, methods that correct for spatial distortions due to magnetic susceptibility differences between different tissue types [55] can be incorporated to improve the accuracy of the segmentation. These automated segmentation tools perform well for labeling normal tissue, but interactive tools are generally necessary for segmenting pathological tissue, such as tumors. For these structures, intensity classifiers are interactively constructed along with manual editing tools. Real-time volume rendering tools support effective feedback on the state of the segmentation.

The result of image segmentation is a volume labeled with the tissue classes of interest. In order to extract the bounding surface of the labeled tissue(s) we apply an isosurface routine to construct a faceted approximation to the 3D surface. Two approaches we use are the marching cube algorithm by Lorensen [43], and the Delaunay reconstruction algorithm by Geiger [26], as implemented in his Nuages package. The output facets, generally triangles, can be directly used for registration, or alternatively the surface may be represented by just the facet vertices.

One benefit of the Nuages package is its parameterization of the surface in terms of a maximum surface estimation error. In fitting triangles to the surface it ensures that no point in the output triangular representation is further away from the actual surface than a given threshold. As this error threshold is allowed to grow, the surface is effectively smoothed and its representation will contain fewer, larger triangles. We use this parameter to specify the surface resolution of the data, providing us with one class of coarse to fine descriptions of the surface. Such parameterization is used for both our hierarchical reasoning as well as for resolution sensitivity studies.

3.2 Initial Alignment

The first registration step is to compute a coarse initial alignment to serve as a starting point for the pose optimization process to follow. While initial alignment may not be necessary if the input datasets are sufficiently close in pose, we define initial alignment approaches to use in a fully automated general purpose registration capability. Examples of situations where initial alignment may not be necessary are head registration problems where the subject's head is placed in approximately the same orientation in the MR scanner for each image. In general, our experimentation on head data, described in Chapter 5, has shown that the initial alignment of the datasets to be registered should not be greater than 75° for ideal datasets to avoid converging to the 180° symmetric pose. As coverage overlaps decreases and measurement noise increases, the initial alignment requirements become more stringent. If coverage around most of the head is available, we expect to converge from up to 35° initial offset, but with small coverage areas, such as with laser data from only part of the visible side of the head, our region of convergence reduces to 10°-20°.

Another requirement on the initial alignment process is efficiency. It must run fast relative to the remaining registration steps. At the opposite extreme, if initial alignment is too slow or if it cannot meet our region of convergence constraints, we may resort to sampling pose space, running the registration on each pose sample, and selecting the best solution.

Several initial alignment methods are available to satisfy these relatively loose requirements:

- **Centroid/principal axes alignment**.

  In cases where the datasets are nearly complete overlaps of each other we can align global measures of the surfaces. In particular, we first translate the test dataset so that its centroid aligns with the centroid of the reference dataset. We then rotate the test dataset so that its inertia matrix eigenvectors correspond to the reference eigenvectors. Due to a sign ambiguity and possibly closely valued eigenvalues, up to 24 different initial alignments may have to be considered. These can either be pruned by examining surface distances or can be passed on for refinement by the subsequent registration steps.

- **Alignment method**.

  A more general, but computationally expensive, approach is to align the input surfaces using geometric constraints between a small number of selected surface points. In this method [28], we first select a pair of widely separated surface points from the test dataset and then search over all possible pairs of reference surface points. The search process is accelerated by (1) coarsely sampling the reference dataset, (2) only examining those pairs from the reference dataset whose distance roughly agrees with the distance between the test dataset pair. In addition to distance constraints, normal angle constraints may be used as well. For each valid match of pairs of points, we solve for the transformation that aligns the points and the surface normals, provided such a transformation exists. For those pairings of points with legal transformations, we apply the transformation to all the other points in the dataset, and measure the distance between the transformed and reference surfaces. Sufficiently close initial alignments are passed on for refinement by the subsequent registration steps. If normals are not available we can perform the same constrained search by using triples of points.

- **Manual alignment**.

  Although exact manual registration is often difficult to perform, manual alignment can be used to provide an initial alignment in some medical applications. Such manual alignment may consist of:

  - specification of rough descriptive axis directions, such as coronal, sagittal, and axial slicing directions,
  - specification of correspondences between specific points,
  - manual translation and rotation of one surface relative to the other.

  All these operations may be performed quickly since high accuracy is not required for the initial alignment. More importantly, it is difficult to achieve high accuracy using any of these manual alignment methods even if much time is taken.

3.3 Rigid Alignment

The first registration step, for each initial alignment, is to solve for a rigid transform (three rotation and three translation degrees of freedom) which best aligns the two input surfaces. This problem is cast as an optimization over pose space minimizing an objective function measuring the distance between the two surfaces. The key issues in the specification of the objective function are:

- The metric must be an accurate reflection of the actual displacement between the surfaces. Distance measures used for this purpose include point-to-point and point-to-surface Euclidean distances. Furthermore, the objective function must be robust to outliers, such that outliers do not inordinately affect the registration.
- The objective function must be sufficiently smooth to pull in solutions over a long distance in pose space without getting trapped in local minima. The goal is to achieve a relatively smooth objective function, thus simplifying the optimization algorithm which we apply over that function. A smooth objective function may be achieved by smoothing the surfaces which we are registering or smoothing the objective function itself. A coarse-to-fine approach is generally employed in performing this smoothing in order to first exploit salient properties of the surfaces and then refine the registration using more detailed surface attributes.

We develop two related approaches to rigid alignment: Gaussian-weighted point-to-point optimization and hierarchical point-to-surface optimization. In each case we develop an objective function which estimates surface displacement in a hierarchical and robust manner.

3.3.1 Point-to-Point Optimization

In this approach we represent both surfaces to be registered as sets of 3D points which are randomly (using a uniform distribution) sampled on the surfaces. The objective function sums the distance from each point on surface 1, the test surface we are transforming, to the nearest point or points on surface 2, the reference surface. In order to efficiently retrieve the nearest reference points we construct a spatially indexed table of points. Each table entry represents a small volume of the reference dataset and contains pointers to all reference surface points that are within a given distance of that volume. In order to support non-uniform distribution of points within the reference volume we use an octree, similar to [56], with overlapping cells to hold the reference points. For objective functions which require all nearby reference points, such as the Averaged Nearest Neighborhood metric of Section 3.3.1.3, we may use relatively large minimum size cells. For objective functions which require only the nearest point, as in Sections 3.3.1.1 and 3.3.1.2, we may use smaller cells which will contain fewer points and can therefore be searched linearly for the single nearest point. Alternatively, we have also used hash tables, where we uniformly decompose the reference volume into equally-sized cells to index entries in constant rather than logarithmic time.

3.3.1.1 Saturated RMS Nearest Point Metric

A straight-forward distance measure of two sets of 3D points is the average distance of the closest point from surface 2 to each point in surface 1. The root-mean-square (RMS) of the distances is generally used to represent the average. In order to limit the effect of outliers we saturate the RMS measure at a given distance, $d_{\max}$. Points in surface 1 whose nearest point in surface 2 is further than $d_{\max}$ are considered to be only $d_{\max}$ away from surface 2. The objective function is:

$$E_1(\Gamma) = \sqrt{\frac{1}{N}\sum_i \min\left(\min_j \left|\Gamma\bar{P}_i^1 - \bar{P}_j^2\right|^2, d^2{}_{\max}\right)}$$

where $\bar{P}_i^1$ is a point on surface 1, $\bar{P}_j^2$ is a point on surface 2, Γ is the rigid transformation for which we are solving, and N is the number of points on surface 1. This metric acts like a chamfer matching scheme [38].

3.3.1.2 Gaussian-Weighted Nearest Point Metric

A related optimization metric is the Gaussian-weighted nearest point distance,

$$E_2(\Gamma) = -\frac{1}{N}\sum_i e^{-\frac{\min_j \left|\Gamma\bar{P}_i^1 - \bar{P}_j^2\right|^2}{2\sigma^2}}$$

With this measure, instead of limiting the impact of outliers by saturating the objective function we weight the distance with a Gaussian to gradually limit the impact of outliers. The advantages of this approach over the Saturated RMS measure are:

- It provides the σ parameter for hierarchically solving for pose. Large σ results in a broad objective function that pulls in solutions from far away, but does not necessarily well localize the global optimum. Small σ results in a steeper objective function which better localizes the optimum. But if we start out with the small σ we potentially classify many surface 1 points as outliers and thus converge on the wrong pose, particularly if the starting pose is some distance away from the optimum.
- It results in a potentially larger pose region of convergence since it does not have a hard saturation limit. If the Saturated RMS measure starts from a pose in which many points are saturated, it may be difficult to identify the correct direction in pose space in which to find the global optimum if the points remain saturated under new test poses. The gradual decay of the Gaussian-Weighting measure results in a more robust behavior that can pull in solutions from wide distances in pose space. Another way to look at this benefit is that the derivative of the Saturated RMS Nearest Point metric is discontinuous at the saturation point while the derivative of the Gaussian-Weighted Nearest Point metric remains continuous. As a result, optimization methods that exploit derivatives should have a larger region of convergence with the Gaussian-Weighted Nearest Point metric. A comparison of the two functions showing their continuity relationship is shown in Figure 8.

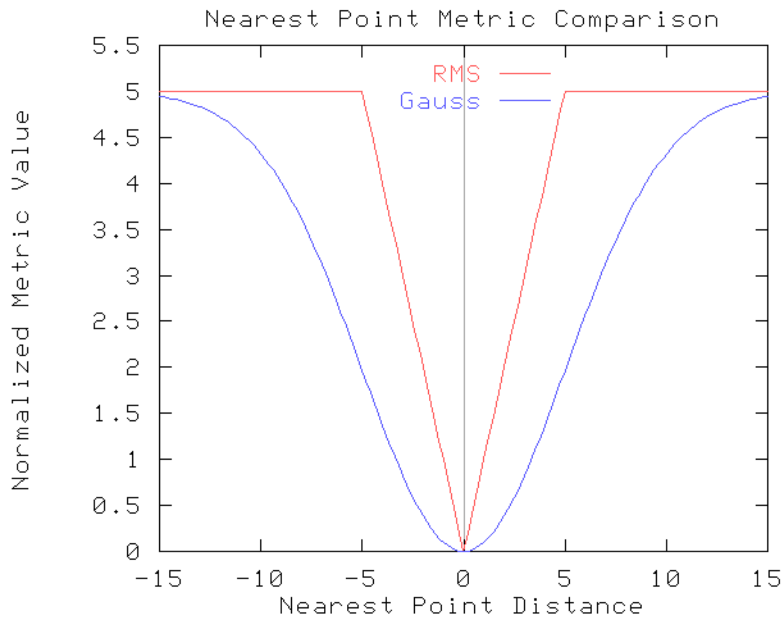


Figure 8 Comparison of Saturated RMS Nearest Point (d_{max} =5) and Gaussian-Weighted Nearest Point (σ =5) metrics, varying the inter-surface distance for one test point. The Gaussian-Weighted Nearest Point function is normalized to the same range as the Saturated RMS Nearest Point metric for the comparison.

3.3.1.3 Averaged Nearest Neighborhood Metric

In order to better approximate the true point to surface distance, we can average the distance from each surface 1 point to all nearby surface 2 points. This averaging is performed using a Gaussian weighting function, where the weighting controls the regions of support over which the distance averaging is computed. More precisely, for each point $\bar{P}_i^1$ on surface 1 and point $\bar{P}_j^2$ on surface 2 we minimize the function,

$$E_3(\Gamma) = -\frac{1}{N}\sum_i \sum_j e^{-\frac{\left|\Gamma \bar{P}_i^1 - \bar{P}_j^2\right|^2}{2\sigma^2}}$$

This function is similar to the posterior marginal pose estimation method used by Wells [63] for model-based recognition applications. By varying σ from large to small values we first achieve a coarse alignment with influence over broad areas and then tighten the fit using smaller but more detailed regions of support.

3.3.2 Point-to-Surface Optimization

If we take the point-to-point distances, $\left|\Gamma \bar{P}_i^1 - \bar{P}_j^2\right|$, in the metrics E_1 through E_3 above as the

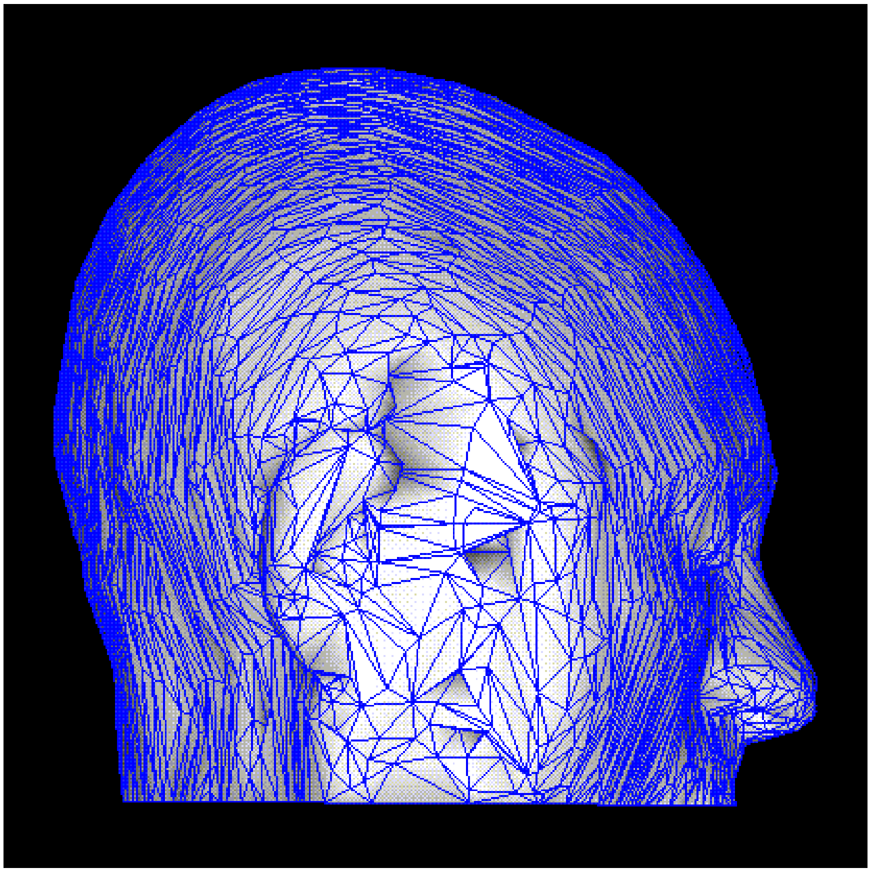


Figure 9 Triangular facets comprising the head surface. The triangle edges are shown in blue—note varying triangle size, depending on underlying surface curvature. A surface rendering of these facets using smooth shading is shown in white.

Euclidean distances between the points we arrive at a distance metric that is inversely proportional to the sampling rate of points on the surface and is thus potentially less reliable for sparser datasets [30]. In order to achieve a more accurate estimate of the distance between the two surfaces being registered we can represent the surface using surface primitives rather than sets of 3D points. Figure 9 shows the triangular surface primitives against which we measure distance with the metrics in this section. The point-to-point metrics described in Section 3.3.1 only use the vertices of the triangles shown in Figure 9. Most isosurface functions available with graphic packages output such triangular facets. The ones we show here were generated with the Nuages package [26].

3.3.2.1 Unidirectional Surface Normal Metric

A simple method of capturing first order surface properties is to introduce the surface normal into the distance measurement. Let surface 1 still be represented as a set of 3D points, $\bar{P}_i^1$, but let surface 2 be represented as a set of 3D points, $\bar{P}_j^{\,2}$, with associated normals, $\hat{N}_j^{\,2}$. We can then measure the distance from point $\bar{P}_i^{\,1}$ to a local surface patch around $\bar{P}_j^{\,2}$ as $\langle\left|\Gamma\bar{P}_i^{\,1}-\bar{P}_j^{\,2}\right|,\hat{N}_j^{\,2}\rangle$, where $\langle\bar{A},\bar{B}\rangle$ denotes dot product of vectors $\bar{A}$ and $\bar{B}$. The metric we want to minimize may then be defined as:

$$E_4(\Gamma) = -\frac{1}{N}\sum_i e^{-\frac{\min_j\langle\left|\Gamma\bar{P}_i^{\,1}-\bar{P}_j^{\,2}\right|,\hat{N}_j^{\,2}\rangle^2}{2\sigma^2}}$$

This measure yields a better distance estimate, but is sensitive to errors in normal computation and the density of the surface points [30]. It is also only applicable if $\Gamma\bar{P}_i^{\,1}-\bar{P}_j^{\,2}$ is relatively small.

3.3.2.2 Bidirectional Surface Normal Metric

One method to reduce the sensitivity of the distance metric to normal measurement errors is to compute a bidirectional distance—such as by taking the average or maximum of the normal distance from surface 1 to surface 2 and the normal distance from surface 2 to surface 1. One such bidirectional metric is thus:

$$E_5(\Gamma) = -\frac{1}{N}\sum_i e^{-\frac{\min_j(\langle\left|\Gamma\bar{P}_i^{\,1}-\bar{P}_j^{\,2}\right|,\hat{N}_j^{\,2}\rangle+\langle\left|\Gamma\bar{P}_i^{\,1}-\bar{P}_j^{\,2}\right|,\hat{N}_i^{\,2}\rangle)^2}{2\sigma^2}}$$

3.3.2.3 Direct Surface Distance Metric

A more direct method for computing point to surface distances is to treat the surface as an explicit set of polygons as is commonly done in computer graphics. Such a representation is usually a set of connected triangles. By representing the reference surface, surface 2, in this manner and the test surface, surface 1, as a set of 3D points, our registration is attempting to minimize the weighted average distance from each of the 3D points and the closest triangle. Thus the metric we minimize is:

$$E_6(\Gamma) = -\frac{1}{N}\sum_i e^{-\frac{\min_j(\mathrm{DIST}(\Gamma\bar{P}_i^{\,1},\Delta_j^{\,2}))^2}{2\sigma^2}}$$

for points, $\bar{P}_i^{\,1}$, in surface 1 and triangles, $\Delta_j^{\,2}$, in surface 2. Surface 1 has a total of N points. The DIST function computes the distance from a 3D point to a 3D triangle. In order to spatially access nearest reference surface triangles to test points we construct a spatially indexed table of triangle vertices and find the nearest reference vertex to a test point. We use an octree for the spatially indexed vertex table in a similar manner to the surface point table used for the point-to-point optimization metrics described in 3.3.1. With each vertex we keep a list of all containing triangles and we measure point-to-triangle distances for each one, returning the minimum. Note that this procedure does not guarantee the smallest point-to-triangle distance since the closest triangle is not necessarily the one with the nearest vertex. Such failures may occur when the surface is highly non-convex and when we have used a high degree of smoothing to generate large triangles. In our applications, though, the surfaces, such as the head or intra-cranial cavity, are relatively convex and our hierarchical approach provides for a high resolution surface representation to be used as the final registration pass, thus minimizing the frequency and impact of such failures.

An available system which generates a triangular surface representation given a segmented data volume is the Nuages system [26]. By using such a direct surface representation we gain the advantages of:

- Reduced sensitivity to subsampling artifacts and normal vector calculation errors.
- Direct support for hierarchical resolution processing by directly smoothing the surface. The Nuages system provides such a smoothing mechanism by applying a maximum surface estimation error parameter in fitting the triangles to the data.

The disadvantage of this approach is the higher computational cost of the point to triangle distance calculations over the point to point distance calculations. But since this metric gives us the most accurate measure of interest we use this measure instead of the surface normal measures in Sections 3.3.2.1 and 3.3.2.2 for all our point-to-surface registration applications.

3.3.3 Non-Spatial Distance Properties

In addition to measuring spatial distance between points we may also include differences between differential geometry properties of the points [58]. Gaussian curvature, for example, is a common property which may be computed to classify points. By attempting to drive points together so that they are not only close to each other, but also have similar properties we may hope to achieve higher accuracy. The drawbacks to incorporating such computations in the registration process, though, are:

- Sufficient data resolution is necessary to compute curvature and related properties.
- Given the generally smooth surfaces of most anatomical surfaces and the noise sensitivity of curvature calculations, the uncertainty associated with the computation of curvature may be as large as the expected variability of curvature across most of the surface.
- Heuristic weights are necessary to combine spatial distances with curvature distances. Bayesian methods may be used to perform such evidential fusion, but any such methods are likely to be highly domain dependent.

As a result of these issues, we propose possible *control* uses of point properties—i.e., use point attributes to prioritize or filter points we are registering. Development of such methods to leverage differential geometry attributes in a reliable manner is not tackled in this thesis, but is a ripe area for future research.

Another measure that may be incorporated into the distance metric is the point location covariance [56]. Generally there is some uncertainty associated with the location of surface points. Contributions to this uncertainty are non-isotropic image resolution, distortions along magnetic susceptibility discontinuities in MR images [55], and errors introduced in the segmentation process. If the location covariance is roughly circular and constant across all points in a given dataset, it will have little impact. On the other hand, for highly non-isotropic voxels, incorporation of such covariances becomes important. Given the high resolution MRI data which we are registering in this thesis, we do not explore the impact of covariances on registration accuracy, but feel this is another important area to follow up in future research.

3.3.4 Hierarchical Refinement

With any of the distance metrics defined in Sections 3.3.1 and 3.3.2 we have several methods available for achieving our smooth objective function goals. The basic idea is to commence the registration with a coarse, but relatively smooth, objective function which has a large region of convergence and few local minima. We then sequentially increase the resolution of the objective function to obtain sharper and more well-defined functions which better localize the optimal solution. The main goal of such processing is to improve accuracy, not efficiency, as is the case with some types of hierarchical processing. In some cases, the processing at the coarser objective function levels may be more efficient since the objective function computation may be simpler. But this is not always the case given that we compute distance transforms or hash the data points at each level. Most of the potential efficiency benefits of hierarchical refinement arise from its processing behavior of gradually focusing on the (hopefully) optimal transformation while avoiding local minima. But since the hierarchical refinement requires additional processing to both generate and process the multiple levels in the hierarchy, we expect efficiency to somewhat decrease.

The multiple methods we combine to achieve a hierarchical refinement strategy are:

1. Distance Measure Weighting Function

   One method of such hierarchical refinement is to sequentially decrease the magnitude of the standard deviation, σ, that is used to weight the distance measures between the surfaces being registered. The impact of reduced values of σ is:

- Small decreases in inter-surface distance are given greater weight.
- Impact of potential outliers is reduced.
- Possibility for local minima is increased, but by sequentially decreasing σ we hope to avoid the local minima.

2. Surface Sampling

   Another method of hierarchical refinement is to sequentially increase the sampling rate of the surfaces. For cases where we represent the surface with a set of 3D points we can start out with a coarse sampling of the data points to get a rough idea of the surface geometry and sequentially add more points to work with a higher resolution representation of the surface. If the sampling is performed randomly, we can effectively reduce the surface resolution, but without performing any smoothing prior to the sampling we may introduce aliasing artifacts of high frequency surface components.

3. Surface Smoothing

   In cases where we represent the surface using patches such as triangles we may directly smooth the surface to get a more accurate scale representation than we might get by sampling. By directly smoothing the surface we can more readily reduce the high frequency components of the surface and retain the low frequency components in which we are interested at the coarser scale. A straight-forward mechanism for achieving the smoother surfaces is to increase the surface estimation error in fitting the surface patches to the underlying data.

The latter two methods we categorize as data smoothing controls, since they act only on the data to generate coarse-to-fine registration measures. The first method acts directly on the distance metric to generate a different set of coarse-to-fine measures.

3.4 Flexible Registration

Once we have obtained an optimal rigid registration, as described in Section 3.3, we may need to further solve for tissue deformations by increasing the complexity of the transformation. Such flexible registration is necessary in cases where:

- High accuracy is needed and sensor distortion is possible, such as distortion in frequency-encoding direction in MR images along magnetic susceptibility discontinuities.
- Tissue may be locally rigid, but is articulated translationally or rotationally due to external forces. The spine or knee are some examples.
- Tissue is actually non-rigid and may deform. Such deformations are expected to be smooth. An example is mammography.

Our approach to such flexible registration is also a hierarchical one in which we gradually incorporate additional parameters into the transformation function which converts from one coordinate system to the other. After the initial rigid registration we apply a global affine transformation to allow shear and scale variations. If higher order transformations are required and sufficient data coverage is available in the volume of interest, the affine transformations may be applied in local volume neighborhoods along with smoothness constraints, similar to [23]. The pose complexity hierarchy therefore consists of two main steps, and an optional third one:

1. Rigid transformation, transforming point $\bar{P}$ to $\bar{P}'$: $\bar{P}' = \Gamma_{\mathrm{rigid}}\bar{P}$. Γ_{rigid} is a transformation consisting of three rotation parameters and three translation parameters. A total of six variables thus describes the transformation. This solution is obtained using a measurement resolution hierarchy as described in Section 3.3.

2. Global affine transformation, transforming point $\bar{P}$ to $\bar{P}'$: $\bar{P}' = \Gamma_{\mathrm{affine}}\bar{P}$. Γ_{affine} is a transformation consisting of three scaling parameters, six shearing parameters, and three translation parameters. A total of twelve variables thus describes the transformation. Bounds on the affine transform may be set by limiting the deformation introduced by the affine transform over the rigid transform to a maximum value for any point. A measurement resolution hierarchy is likely not necessary here since deformations are not expected to be large.

3. Local affine transformations, in which the test volume is decomposed into local neighborhoods, where in each one we solve for optimal affine solutions. We do not apply such local deformations for most of our applications since the surfaces we use are generally not sufficiently dense to support stable alignments over small neighborhoods. A motivating example where such decomposition is feasible and helpful, though, is shown in the mammogram change detection system in Section 4.3 for matching two breasts separately. In this case anatomical knowledge is the key for effective decomposition, allowing us to identify specific well-bounded components that are relatively independent of each other. Other examples include back, or other bone imaging, where individual bones may articulate relative to each other. If the previous registration steps are presumed to be sufficiently close to the correct pose, large-scale anatomical decomposition may be performed to identify such independent local neighborhoods. Anatomical constraints between components may be defined by connectivity or joint degrees of freedom relations, but such constraints are not implemented here.

   Alternatively, iterative volumetric decomposition, such as with octrees, may be performed. In this case we compute the optimal affine transformation in each local neighborhood, or octree cell, and ensure smoothness across neighborhoods by applying each transformation to the center of its neighborhood and interpolating the transformation between cell centers. If the extent of the deformation is limited to relatively small values, simple tri-linear interpolation may be used to smooth the transformation. Otherwise, higher order interpolation functions, such as splines, may be used. We do not implement this alternative here, but it may be well-suited for some applications.

By performing the flexible registration in a hierarchical manner we achieve the following processing behaviors:

- Solve for the most stable components first. The rigid parameters might exhibit small changes as the contributions of the deformations are factored in, but their contributions are not expected to significantly vary at the finer levels.

- Avoid high computational burden and local minima problems of trying to explore the higher dimensional search space of the affine transformation at one time.
- Minimize non-rigid transform contributions. For same-subject registration where physical variations are expected to be small, we do not want registration solutions that introduce complex warping functions that, although align the data, do so in a highly unrealistic manner.

Maintaining transform simplicity is one reason for not including higher order warping, beyond affine, in the registration solution. With higher order transforms it becomes difficult to maintain meaningful displacement functions between the two datasets being registered. If we allow sufficient warping to occur we can transform one surface to perfectly match another, but we will not necessarily know if the resultant displacements are correct. For example, we can unknowingly use higher order warping functions to correct for possible errors that were made in the rigid or affine registration phases, resulting in accurate overlay of the two surfaces being registered but generating a distorted displacement function which is unlikely to be very meaningful. Higher order solutions are also inappropriate for surface registration in which the surfaces do not adequately cover the volume of interest. For example, if we are performing head registration based only on skin surface, it is unlikely that interpolation of high order transformation functions inside the head volume will be accurate without additional registration points inside the volume. We suggest higher order registration solutions for *inter-subject* registration in which multiple surfaces or volumes are the bases for registration.

Our approach, by focusing on surface matching and in particular surfaces that are not expected to exhibit large deformations, is aimed at aligning datasets that have not undergone large deformations. We are primarily looking at identifying the significant components of the coordinate frame transforms between two surfaces. Highly accurate change detection or localization becomes very difficult as the viewed tissue undergoes greater and greater deformations.

3.5 Optimization Methods

In order to apply any of the objective functions described in Sections 3.3 and 3.4 we need to perform a nonlinear optimization over those functions solving for the components of the transformation we are considering. We explore both gradient estimation methods, such as Powell's method, and direct gradient methods, such as quasi-Newton methods.

3.5.1 Powell's Method

Powell's method [51] is basically a conjugate gradient descent method in which the gradient is estimated by performing line minimizations along conjugate directions. Starting at some initial point, this method searches along each conjugate direction for the point which minimizes the objective function along that line. Then from that point it picks a new direction and minimizes along that line until all directions have been scanned. It then minimizes along the direction formed by the initial point and the final point reached, which is the maximum gradient direction estimate. The procedure iterates starting at the new minimum point, but now replacing the conjugate direction which gave rise to the largest objective function decrease with the maximum gradient direction estimate. The iterations continue until decreases in the value of the objective function are below a threshold. The line minimizations performed by Powell use a golden section bracketing approach to narrow in on the desired minimum and a parabolic fitting algorithm to accurately localize the minimum. This technique can avoid relatively small local minima, but is unlikely to behave well with objective functions characterized by high frequency oscillations. We use Powell's method as a baseline optimization technique in comparing the performance of different objective functions since it does not require direct gradient calculations and thus is not biased by gradient estimation errors which may be different for the different objective functions.

3.5.2 Davidon-Fletcher-Powell's Method

Optimization methods that directly use the gradient are generally more efficient than Powell, although they achieve comparable performance [2]. These methods use the gradient to determine new transforms that optimally minimize the objective function. A number of related algorithms have been developed that leverage knowledge of the gradient. We have used the Broyden-Fletcher-Goldfarb-Shanno variant of the Davidon-Fletcher-Powell (DFP) method [51]. This method uses Newton's method to search for a zero of the gradient of the objective function. By using a second order Taylor expansion of the objective function, this process requires an estimate of the Hessian matrix containing the objective function's second order partial derivatives. The method estimates the Hessian matrix through differences in the gradient measures made at sequential updates of the objective function. So although the Hessian matrix is not required from the user, and in fact the exact Hessian is not even desired by the algorithm's authors, the gradient computations are required. We have implemented this optimization technique in our on-line rigid registration applications in case its efficiency benefits were required. Estimates of the gradient may be easily computed for the point-to-point objective functions if we assume that the nearest points remain the same for the computation. We have applied DFP optimization to the Saturated RMS Nearest Point metric and the Averaged Nearest Neighborhood metric. For the Saturated RMS Nearest Point measure we have the following partial derivatives:

$$\frac{\partial E_1}{\partial t_k} = \frac{1}{N}\left(\frac{1}{N}\sum_i \left|\Gamma \bar{P}_i^{\,1} - \bar{P}_j^{\,2}\right|^2\right)^{-\frac{1}{2}} \sum_i \langle \Gamma \bar{P}_i^{\,1} - \bar{P}_j^{\,2}, \frac{\partial}{\partial t_k}\Gamma \bar{P}_i^{\,1}\rangle$$

for each transformation component, t_k, where $\langle \bar{A}, \bar{B}\rangle$ denotes the dot product of A and B. For the translation components, $\frac{\partial}{\partial t_k}\Gamma \bar{P}_i^{\,1}$ is just the unit vector in the translation direction. For the rotation components, $\frac{\partial}{\partial t_k}\Gamma \bar{P}_i^{\,1}$ is estimated as the tangent to the rotation motion, which is the following cross product:

$$\bar{r}_k \times R\bar{P}_i^{\,1}$$

where $\bar{r}_k$ is the axis of rotation for rotation k and R is the rotation matrix component of transformation Γ.

For the Averaged Nearest Neighborhood metric, we have similar partial derivatives:

$$\frac{\partial E_3}{\partial t_k} = \frac{1}{N\sigma^2}\sum_i \sum_j e^{-\frac{\left|\Gamma \bar{P}_i^{\,1} - \bar{P}_j^{\,2}\right|^2}{2\sigma^2}} \langle \Gamma \bar{P}_i^{\,1} - \bar{P}_j^{\,2}, \frac{\partial}{\partial t_k}\Gamma \bar{P}_i^{\,1}\rangle$$

where $\frac{\partial}{\partial t_k}\Gamma \bar{P}_i^{\,1}$ for both the translation and rotation components are the same as above.

3.5.3 Random Pose Perturbations

Independent of the optimization method used and despite attempts to achieve smooth objective functions, we are still subject to being trapped in local minima. As a result, within each hierarchical registration loop, we perform repeated optimizations, starting with the currently best scoring pose, but applying small random perturbations to the best pose and re-optimizing. We perform at least two such iterations and stop if our objective function has dropped below a conservative threshold. Otherwise we continue the random pose perturbation and re-optimization until one of the end conditions is reached:

1. The objective function has dropped below a threshold and at least two iterations have been performed.
2. A given number of iterations have passed since we have seen any substantial (above a threshold) improvement in the objective function.
3. The total number of iterations has eclipsed a maximum number.

All the registration parameters remain unchanged through these iterations, in contrast to the hierarchical registration loops where we do vary some of the registration parameters.

The small random perturbations we apply are chosen from a uniform distribution in the same fixed ranges for all iterations. Since these ranges are usually small, defaults of $\pm 5\text{mm}$ in translation and $\pm 5^{\circ}$ in rotation, we do not vary the shape of the distributions in subsequent iterations. As is shown in Section 5.4.4, a few random iterations do appear to improve the pose region of convergence, but is not clear if more and larger perturbations will continue to expand this capture range. It may be interesting to further explore this issue in future research—specifically analyzing the impact of varying the shape of the distribution, such as Gaussian instead of uniform, as well as the extent of the distribution, gradually decreasing the distribution bounds as in simulated annealing.

3.5.4 A Final Word on Optimization Techniques

Since our focus is on developing registration metrics and objective functions which achieve accurate registration performance in the presence of such real world artifacts as pose offsets, measurement noise and outliers, we have not conducted analysis of the many different optimization techniques that are available. We have found Powell's method and the DFP quasi-Newton method to work well in our applications and defer to extensive work performed on optimization methods to characterize the different approaches [2][51].

3.6 Related Registration Approaches

Image registration, both 2D and 3D, has been an active area of development over the years. Applications of registration include not only medical image analysis but also target recognition, environmental monitoring, and autonomous navigation. As a result, a wide range of registration approaches have been developed. Thorough surveys of these techniques are given in [12] and [59]. We describe registration algorithms which apply to surface-based representations similar to ours.

One of the first algorithms to propose a hierarchical approach to minimizing a distance metric between two datasets to be matched is by Borgefors [10]. In that work a multi-resolution solution is proposed to minimizing an RMS error between edge points in 2D images. By starting with low resolution edge images and proceeding to higher resolution ones, susceptibility to local minima and random noise was reduced. Jiang [38] extended this approach to 3D surface matching and introduced thresholds on the RMS error to minimize the impact of outliers. The outlier threshold was varied to achieve a hierarchical outlier rejection behavior. In both cases, multiple starting points along with some heuristics were used to avoid local minima. Our approach introduces Gaussian weighting to achieve outlier rejection and smooth objective functions, evaluates direct point-to-surface distances, and extends to flexible registration.

Pelizzari *et al* [49] have developed a method that matches retrospective datasets, (MRI, CT, PET), to one another. This work also uses a least squares minimization of distances between datasets, although with a different distance function and with more operator guidance required. This approach requires some operator intervention to set an initial starting position as well as to steer the system toward the correct solution, which our system does not.

Lavallee and Szeliski [56] perform a least-squares minimization of a distance function to match datasets. Here, the distance is weighted by an estimate of the inverse variance of the measurement noise, and the Levenberg-Marquardt method is used to find the minimum. The transform solution iteratively increases in complexity to model deformations. 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 the removal of outliers is sufficient to keep the method from getting trapped into local minima.

Gueziec [31] and Thirion [58] have developed feature correspondence approaches to surface matching. These approaches match ridge lines, which track points of maximum curvature along a surface, and extremal points, which are points of maximum curvature along ridge lines. Ridge lines are matched by indexing into a hash table using five intrinsic parameters of the line. A rigid transform is computed for each possible pair of matched ridge lines and the results are collected in a six-dimensional transform accumulator. Extremal points are matched by alignment techniques in which triplets of points are aligned to obtain initial rigid transforms which are then verified and refined using the remaining points. The complexity of correspondence searching is reduced by applying differential geometry constraints that characterize the points. A potential drawback of these techniques is their requirement for dense data and high signal-to-noise ratios which are needed to reliably extract and characterize the features.

Feldmar [23] has developed a locally affine registration technique which incorporates a hierarchical transform decomposition similar to Lavallee and Szeliski and to our work. The registration is performed by minimizing a least squares distance measure using the iterative closest point algorithm [8][65]. In this algorithm, a set of point-to-point correspondences are first determined by finding the nearest points in the reference surface to the test surface. A transform is then computed which minimizes the distance along the correspondences which are assumed fixed during the transform computation stage. Given the new transform the algorithm iterates, searching for new correspondences followed by new transforms and continues until the distances along the correspondences is below a threshold. Initially a rigid transform is computed, followed by a global affine transform, and finally local affine perturbations. Besides the hierarchically increasing complexity of the transform, no tools are used to reduce occurrences of local minima.

Chapter 4
Medical Application Development

We have applied the developed image registration techniques to a number of applications in surgical planning, surgical guidance, and diagnosis. While registration solutions are not the only keys to solving these problems, they form a main component of the resultant system. Prototype solutions for applications in image-guided neurosurgery, multiple sclerosis change detection, mammogram change detection, and non-invasive functional brain mapping are described below along with demonstrations on MR imagery from actual patients.

4.1 Image-Guided Neurosurgery

Surgeries often involve translating diagnosis of pathologies made on MR or CT imagery into action by treating or removing the structures in question. Although anatomical structures may be clearly visible in the imagery, the surgeon needs to know where those structures are in the patient in front of him. The surgeon needs to know how to approach the structure, where nearby critical structures are located, and how to identify the structures of interest and their extents. At the heart of these issues is translating the coordinate frame of the imagery to the coordinate frame of the patient in the operating room. Difficulties arise when high accuracies are necessary, sensitive tissues dictate complex surgical pathways, or line-of-sight visualization is hindered.

The key issues in achieving an accurate and reliable image-guided surgery system are:

- registration of the MR or CT imagery to the patient,
- maintaining the registration through the surgical procedure,
- tracking medical instruments in the surgical field, and
- visualizing the instruments in the context of the MR/CT imagery.

We have developed a system which addresses these issues in neurosurgical applications. Components of the system have been used in the operating room on over 20 patients, with positive feedback from the surgeons. Validation in the operating room has thus far been qualitative only, but the positive feedback by the surgeons indicates high potential for the system.

4.1.1 Neurosurgery Application Overview

Neurosurgery is an ideal application for image-guided techniques by virtue of the high precision it requires. Reasons for choosing this application include:

- Neurosurgery greatly benefits from knowledge of tissue identity in the vicinity of the surgical field, e.g., the location of hidden blood vessels is often critical to avoiding blood deprivation to the brain.
- Optimal trajectories to deep-seated tumors or lesions may not be the most direct ones, due to the position of vital brain functions. Image guidance can aid in specifying desirable paths.
- Bone removal should be kept to a minimum. Thus careful planning of craniotomy location is desirable.
- Tissue to be resected from the brain is not always visually apparent. Image guidance can aid in specifying tumor or lesion margins by aligning MR imagery in which tissues are distinguished.
- The brain is encased in the skull which by virtue of being rigid requires the simplest form of 3D registration, i.e., a rigid six degree-of-freedom transformation. Although the brain may still move relative to the skull, especially after opening a craniotomy, the skull provides fixed bounds on such motion.

Initial attempts to achieve image-guided neurosurgery consisted of stereotactic frames which directly provided the fiducials for registering the MRI or CT data to the patient [25]. Stereotactic frames, though, are not only uncomfortable for the patient, but are cumbersome for the surgeon. They are limited to guidance along fixed paths and prevent access to some parts of the head. We would like to use a frameless system not only for its simplicity and generality, but also for its potential for use in other parts of the body.

Frameless stereotaxy systems have been pursued by a number of other research groups [1][6][13][19][24][33][45][52][57][62], and usually consist of two components: registration and tracking.

4.1.1.1 Registration for Image-Guided Neurosurgery

The most common form of registration in current use is based on fiducials [1][6][24][33][45][57][62]. These consist of either (1) markers attached to the skin or bone prior to imaging or (2) anatomically extracted salient features on the head. The fiducials are usually manually localized in both the MR or CT imagery and on the patient and the resulting correspondences are used to solve for the registration. Fiducial systems may not be as accurate as frame-based method—Peters [50] reports fiducial accuracy about an order magnitude worse than frame-based methods, but Maciunas [45] reports high accuracy achieved with novel implantable fiducials.

The other approach used by researchers is surface alignment, with a similar problem formulation as developed in this thesis. Ryan [52] generates the patient's scalp surface by probing about

150 points with a trackable medical instrument. Colchester [19] uses an active stereo system to construct the scalp surface. Our approach performs surface alignment, benefiting from its dense data representation, but we use a laser scanner to construct the patient's scalp surface.

4.1.1.2 Tracking for Image-Guided Neurosurgery

Tracking is the process by which changes in object location are dynamically measured. Of particular interest to us is the tracking of medical instruments and the patient's head. The two most common methods of tracking are articulated arms and optical tracking. Articulated arms are attached to the head clamp or operating table and use encoders to accurately compute the angles of its joints and the resulting 3D position of its end point. Such devices, though, may be bulky in the operating room and due to their mechanical nature are not as fault tolerant as other methods. Optical trackers use multiple cameras to triangulate the 3D location of flashing LEDs that may be mounted on any object to be tracked. Such devices are generally perceived as the most accurate, efficient, and reliable localization system [13][17]. Other methods, such as acoustic or magnetic field sensors are being explored as well, but tend to be more sensitive to environmental effects. We use an optical tracking system, the Flashpoint system by IGT Inc., Boulder, CO, USA, due to its accuracy and ease-of-use benefits.

Tracking patient head motion is often necessary since the head is not always clamped to the operating table, the head may move relative to the clamp, the operating table may be moved, or the hardware performing the tracking may be moved to rearrange lights or other equipment in the operating room. Although not all image-guided surgery systems account for patient motion, a number of researchers [1][13][19][45][52][57] solve this problem by attaching trackable markers to the head or clamp. Since a clamp is not always used we choose to attach markers to the head. As we are using an optical system to track instruments, we attach LEDs to the head to track head motion.

4.1.2 The Image-Guided Neurosurgery System

The goal of this application is to provide image-based guidance via two methods:

- Enhanced reality visualization [9]—rendered anatomical structures segmented from MRI or CT are overlaid on the surgeon's view of the patient. Such overlays may be performed on a video monitor whose source camera is looking over the surgeon's shoulder, or ultimately directly in the surgeon's field-of-view using transparent panels, surgical microscopes, or head-mounted displays. These visualizations allow the surgeon to see internal anatomical structures in their correct positions. Visualizing hidden structures in this manner is highly useful in a priori planning of the surgical approach and in guiding the surgery during the procedure.

- Medical instrument tracking—in addition to just overlaying the location of the internal anatomical structures in the surgeon's field-of-view, we can also track the 3D location and orientation of medical instruments that the surgeon is using. Once the imagery is registered to the patient coordinate frame and medical instruments are tracked relative to the patient, we can transform the location of the instruments to the image coordinate frame. With this capability we can overlay the position of the medical instrument tip on the rendered internal structures or view the location of the instrument in orthogonal slices of the image volume. By tracking medical probes we avoid any possible parallax problems with video visualizations or accuracy problems with stereo displays.

The resultant benefits of these guidance methods for the surgeon include:

- accelerated migration to minimally-invasive surgeries via improved hand-eye coordination and better transfer of a priori plans to the patient,
- shorter procedures through increased visualization of the surgical field and more rapid navigation,
- reduced risk of sensitive tissue damage, and
- more accurate and complete tissue resection, ablation, or biopsy.

Our image-guided surgery system is aimed at achieving highly accurate guidance by exploiting as much of the available data as possible. The system, whose architecture is shown in Figure 10, is characterized by the following properties:

- Frameless—no need to attach a frame to the patient.
- Non-fiducial—no need to identify corresponding landmark points in the imagery and on the patient.
- Laser-based—a laser scanner is used to achieve accurate, efficient, and dense depth maps of the patient's scalp.
- Optically-tracked instruments—LEDs are used to dynamically track instruments.
- Optically-tracked head—relative head motion is continually tracked.
- Verifiable—verification tools validate registration and detect possible sensing errors.

The system development effort was a collaborative project between MIT’s Artificial Intelligence Lab and the Departments of Radiology and Neurosurgery at Brigham & Women’s Hospital. Substantial contributions in the development of the system’s user’s interface, which was used in the generation of most of the results shown in this section, were made by Michael Leventon of MIT. The setup of the resultant system is shown in Figure 11. It consists of a portable cart containing a Sun UltraSPARC workstation and the hardware to drive the laser scanner and Flashpoint tracking system. On top of the cart is mounted an articulated extendable arm to which we attach a

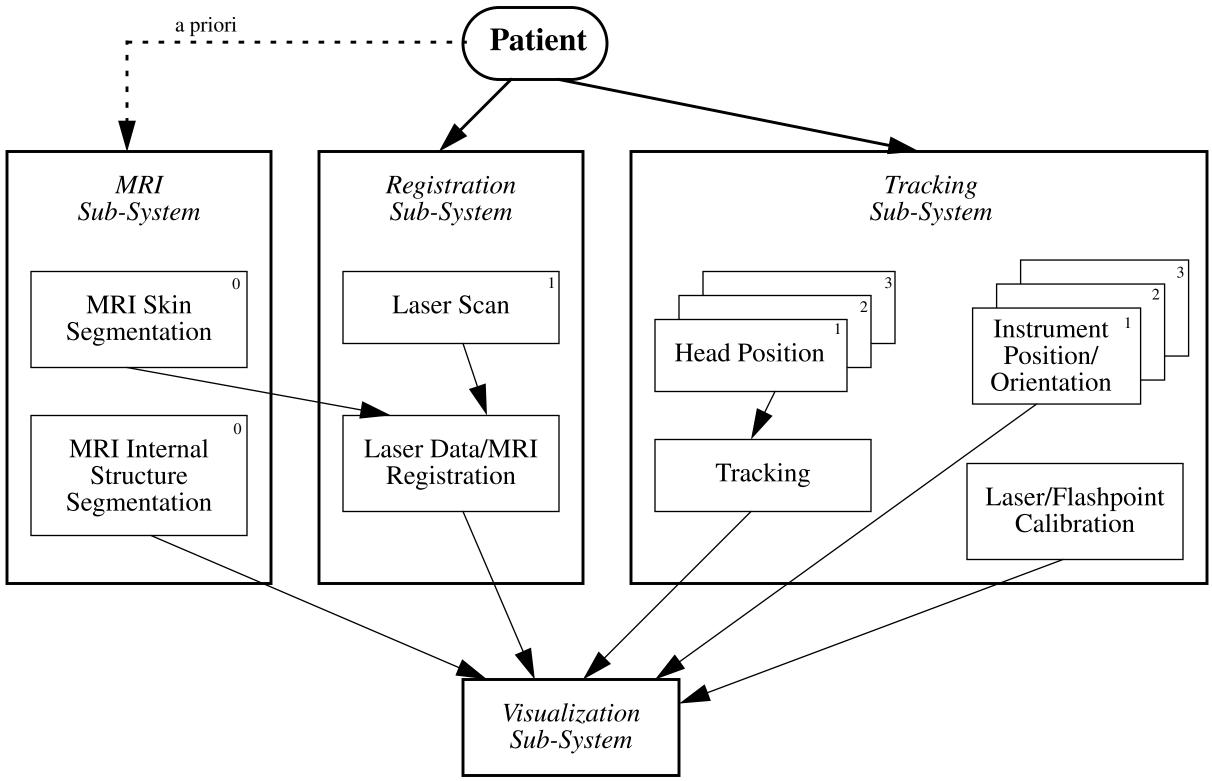


Figure 10 Image-guided surgery system architecture.

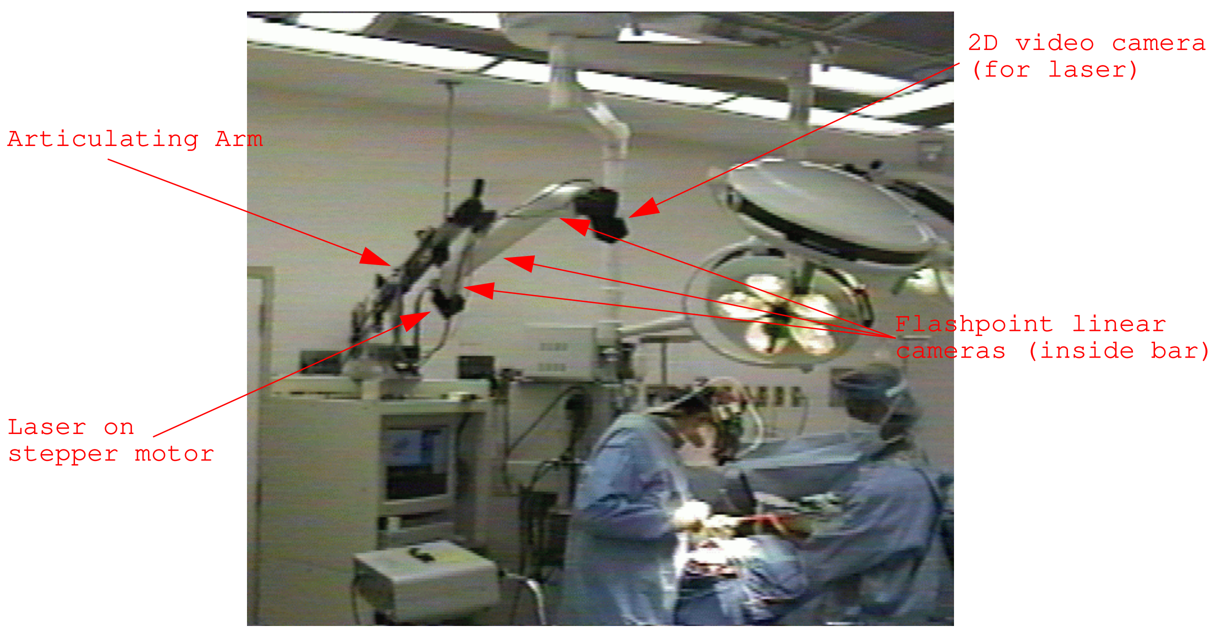


Figure 11 Setup of image-guided surgery system in the operating room.

bar housing the laser scanner and Flashpoint cameras. The joint between the arm and scanning bar has three degrees-of-freedom to allow easy placement of the bar in desired configurations.

The system's principal components are described below.

4.1.2.1 MRI Sub-system

MRI is the prime imaging modality for the neurosurgery cases we support. The images are acquired prior to surgery without a need for special landmarking strategies. The segmentation is performed using automated techniques and semi-automated techniques by specialists at Brigham & Women's Hospital. The segmented tissue types include skin, used for registration, and internal structures such as brain, tumor, vessels, and ventricles. The segmented structures are processed by the Marching Cube algorithm [43] to construct isosurfaces consisting of connected triangular patches. These isosurfaces are used for visualization via surface rendering graphics routines, and in the case of skin, for registration as well.

4.1.2.2 Registration Sub-system

We use a laser scanner to collect 3D data of the patient's scalp surface as positioned on the operating table. The scanner is a laser striping triangulation system consisting of a laser unit (low power laser source and cylindrical lens mounted on a stepper motor) and a video camera. The laser projects a plane of light at a fixed set of angles. The light reflected from underlying surfaces is viewed by a video camera which, via triangulation, localizes the 3D position of the underlying surface points to high accuracy. The laser source is at one end of the sensor bar and the video camera is on the other. The laser is calibrated a priori by using a calibration gauge of known dimensions to calculate the camera parameters and the sweeping angles of the laser. Camera calibration is performed a priori by optimizing for the six location and orientation camera parameters along with focal length using correspondences between 2D image coordinates and 3D world coordinates of laser points on the calibration gauge that are easily automatically extracted in the image. Powell’s method [51] is used for optimization. The laser sweeping planes are found by measuring three 3D point coordinates on the gauge at elevation and width extremes which lie on each individual laser plane. Detail on the laser calibration process is in [41]. Laser scanning time is about 5 seconds and accuracy is about 1mm at the 1m standoff from the scanner which we normally use.

In the operating room the laser scanner is placed to maximize coverage of the salient bony features of the head, such as the nose and eye orbits. To ensure accurate registration we can supple-

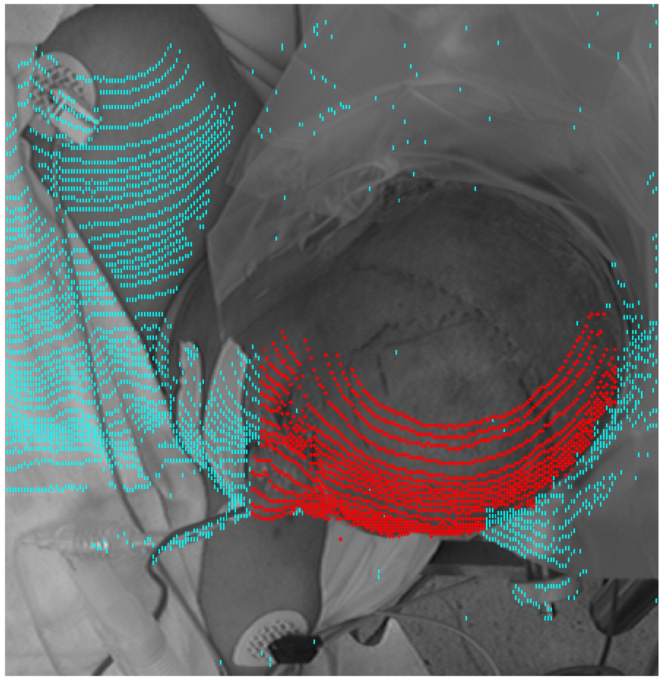


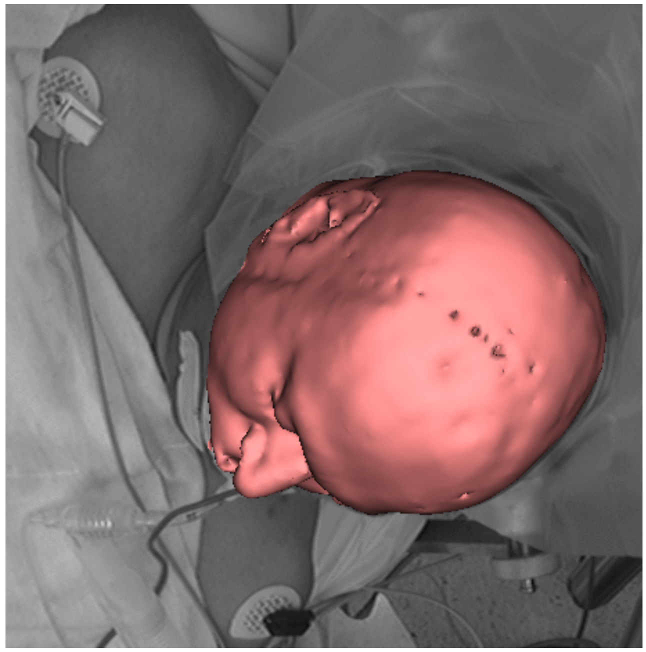
Figure 12 Laser scan of a patient in the operating room. Selected region of interest is shown in red, indicating laser points to be registered to patient's MR image.

ment the laser data with points probed with a Flashpoint pointer, similar to [52], to include skin points that are not visible to the laser in the registration. The acquired laser data is overlaid on the laser scanner's video image of the patient for specification of the region of interest. This process uses a simple mouse interface to outline the region of the head on which we want to base the registration. This process need not be perfect since the registration is designed to deal robustly with outliers. Figure 12 shows the laser data thus obtained from a patient in the operating room. The laser scan and specification of region of interest takes about 30 seconds once the sensor bar is appropriately placed above the patient.

The registration process, as applied to this problem, uses a manual initial alignment step followed by an iterative rigid alignment. Accurate manual alignment can be very difficult, but we aim only to be within 10°-20° of the correct transformation. This process, described in [41], is performed using two displays (see Figure 13) and takes about 60 seconds:

1. Video/MRI overlay—The rendered MRI skin is overlaid on the laser scanner's video view of the patient. The MRI data is rotated and translated in 3D to achieve a qualitatively close alignment.
2. Laser/MRI overlay—The laser data is projected onto three orthogonal projections of the MRI data. The projected MRI data is colored such that intensity is inversely proportional to distance from the viewer. In each overlay view, the laser data may be rotated and trans-

A)



B)

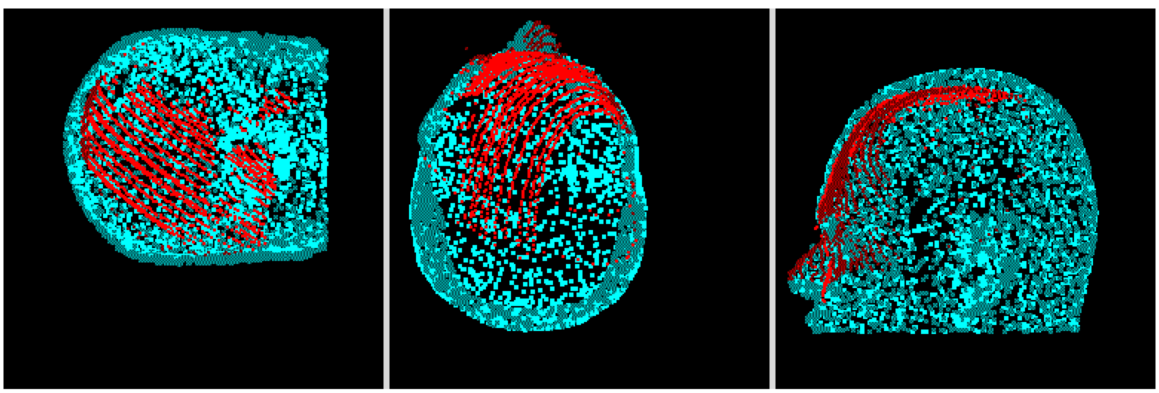


Figure 13 Initial alignment of laser and MRI data.
(A) Overlay of the MRI skin data on laser scanner's video image.
(B) Overlay of the laser data projected on each of three orthogonal projections of the MRI skin data.

lated in 2D to align the projections.

The registration algorithm currently uses the point-to-point optimization techniques described in Section 3.3.1 since they are the most efficient. Both the Averaged Nearest Neighborhood (Section 3.3.1.3) and Saturated RMS Nearest Point (Section 3.3.1.1) metrics have been used in the image-guided surgery system, either sequentially or separately. In all cases we solve for a rigid transform. The registration process runs in about 60 seconds on a Sun UltraSPARC workstation.

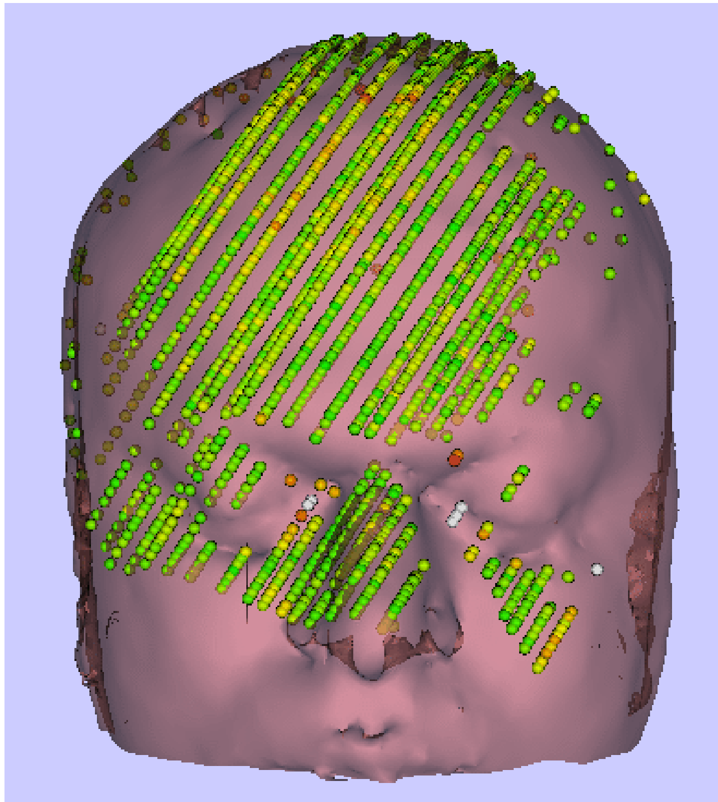


Figure 14 Verification of laser-to-MRI registration by inspection of residual error. The laser data is shown as spheres that are color-coded by their distance to the underlying MRI skin on a continuous scale: green=0mm to yellow=2.5mm to red=5mm. White shows error > 5mm.

Two verification tools are used to inspect the registration results as the values of the objective functions themselves may not be sufficient to guarantee the correct solution. One verification tool overlays the MRI skin on the video image of the patient (Figure 13A), except that we animate the visualization by varying the blending of the MRI skin and video image. A second verification tool overlays the laser data on the MRI skin by color-coding the laser data according to residual distance between the laser points and the nearest MRI skin points, as in Figure 14. Such a residual error display identifies possible biases remaining in the registration solution. Analysis of laser registration performance is given in Section 5.5.1.

By basing the registration process on the scalp of the patient we are assuming that brain motion is minimal relative to the skull. If the patient is approximately in the same pose as he was in the MR scanner, this assumption is generally valid. The more serious problem is brain motion after the craniotomy is performed, as discussed in Section 4.1.2.3.

4.1.2.3 Tracking Sub-system

The Flashpoint system we use for tracking objects in the surgical field is an off-the-shelf product consisting of three high resolution linear cameras which are synchronized with blinking IR LEDs. A large number of LEDs can be supported by sequentially blinking each one of them and viewing each LED through the cameras. The linear cameras are arranged orthogonal to each other such that each one solves for one of the 3D positional coordinates of the blinking LED. The linear cameras are calibrated at the time of installation. The orientation of a rigid object can be tracked by attaching multiple LEDs to it. Two LEDs allow solution of five degrees of freedom, up to twist about the axis connecting the LEDs. Three LEDs allow solution of the six degrees of freedom in a 3D rigid transform. This active tracking system is highly reliable, with an accuracy of about 1mm at a 1m standoff from the cameras.

To deal with possible patient movement, we either attach LEDs to the head or to the clamp which rigidly holds the head during the surgery. We would ideally like to track the actual structures of interest—the brain in this case. But attaching stable markers to the brain is not feasible and usually the craniotomy is too small for effective localization of the brain surface directly. Tracking the skull is the next best option, but that assumes that brain motion is minimal. Given the nature of brain tissue though, it can swell and shrink during the operation in addition to the motion introduced by the surgeon's instruments. The motion can easily be on the order of a few millimeters [14] [35]. We are currently not correcting for such brain motion.

For tracking the skull we attach five LEDs to the head. In our trials the surgeon sutured LEDs to the skin and attempted to minimize skin motion relative to the skull. Alternatively, LED sockets that can be screwed into the skull can be used [45] [52]. Ideally we would like to commence tracking once the registration is performed, but we need to address sterility issues. The registration is performed when the whole head is still visible and the surgical field is not sterilized. At this point we mark and acquire the location of the five LEDs on the patient's scalp around the planned location of the craniotomy, but do not actually attach the LEDs. The marks are scratched in the skin to ensure they are not washed off. Once the patient is sterilized and draped, but before the skin flap is opened, we attach sterile LEDs to the marked locations, by suturing them onto the skin. The LED sockets are notched flat disks designed to easily and securely attach to the skin.

Since in most neurosurgeries the head of the patient is rigidly clamped to the operating table we can avoid skin motion problems and simplify sterility issues by attaching LEDs to the head clamp. We attach at least three LEDs to a rigid bar that connects to the head clamp such that the bar does not interfere with the surgeon’s work, but is generally visible to the Flashpoint cameras. The positions of these LEDs are acquired during registration and are tracked during the procedure. The LEDs are draped as part of the sterilization procedure. The only drawback to this head tracking solution is that we are now one more, possibly unknown, transformation removed from the region of interest—although the head clamp is generally securely attached to the head, it can sometimes unknowingly slip.

We require direct line-of-sight from the Flashpoint cameras to at least three head-tracking LEDs at times when the surgeon requires image-guidance. In order to maintain such line-of-sight we can re-locate the scanning bar such that it is out of the way of the surgeon but maintains visibility of the LEDs. Such dynamic reconfiguration of the scanning bar is one of the benefits of the head tracking process.

We evaluate confidence of the head tracking by computing residual errors between initial and current LED positions. The head tracking process computes a rigid transform from current head position to initial head position at the time of laser-to-MRI registration using Horn’s closed-form least-squares formulation [36]. Movement of LEDs relative to each other appears as large residual errors between their transformed and initial positions. If redundant LEDs are visible, the redundancy will highlight the offending LED which may be de-activated.

Instrument tracking is performed by attaching two LEDs to a sterile pointer. The two LEDs allow us to track the 3D position of the tip of the pointer as well as its orientation, up to the twist angle which is not needed for this application. Figure 15 shows the surgeon using the trackable pointer in the opened craniotomy. The head is being tracked by the attached LEDs.

4.1.2.4 Visualization Sub-system

Two types of visualizations are provided to the surgeon on the workstation monitor. One is an enhanced reality visualization in which internal structures are overlaid on the video image of the patient. The video image, obtained from the laser scanner's camera, is set up to duplicate the surgeon's view of the patient. Any segmented internal MR structures may be displayed at varying colors and opacities. Figure 16 shows a sample enhanced reality visualization from a surgery.

A second visualization, Figure 17, shows the location of the pointer tip in a 3D rendering of selected MRI structures and in three orthogonal MRI slices. These visualizations are updated about once per second as the pointer is moved.

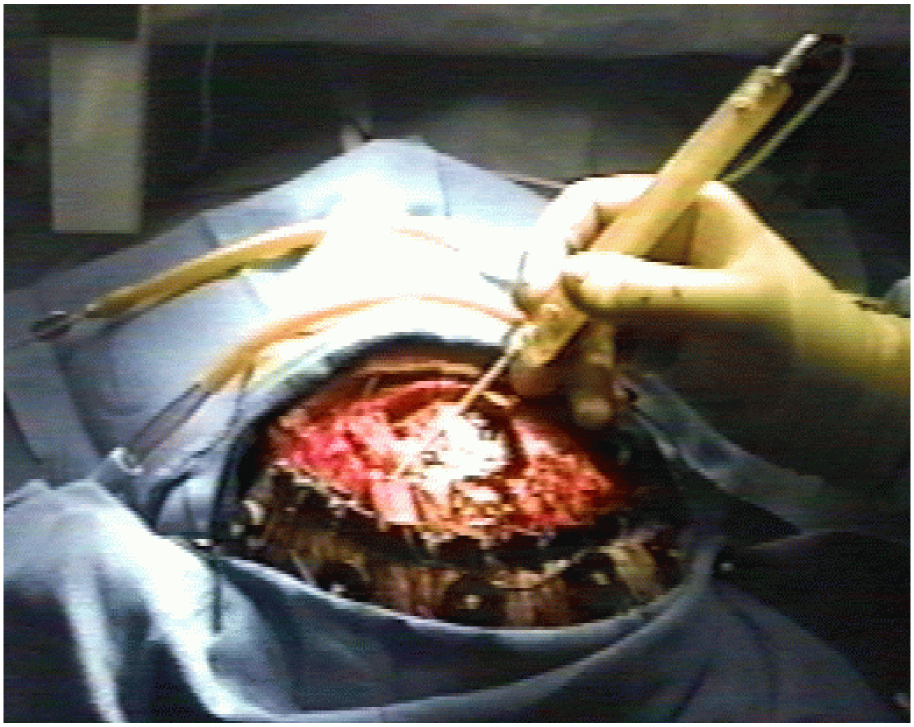


Figure 15 Surgeon using our trackable pointer to identify structures inside the craniotomy. The head-tracking LEDs are in the center of the round disks around the bottom of the skin flap.

4.1.3 Operating Room Procedure

From the surgeon's perspective, using our system involves:

1. Prepare patient for surgery as per usual procedure up to head sterilization process. Head is still visible.
2. Register MRI to patient by placing our scanner bar over patient's head such that patient's face is clearly visible in laser's video image. The bar is generally about 1.5 meters away from the head. Scan patient's head with laser and possibly acquire additional 3D skin points with non-sterile Flashpoint pointer. Added overhead of this step is about 3 minutes.
3. Simultaneously acquire position of head clamp or marked skin points with Flashpoint system to use for head tracking. For direct head tracking, mark location on the scalp of five LEDs to be visible around skin flap after opening. An LED, whose outline is scratched into the skin, is used to acquire 3D position of these points. Added overhead of this step is about 2 minutes.
4. The Flashpoint/laser bar may be re-positioned at any point to avoid interference with lighting or other equipment and to maintain visibility of LEDs.
5. Sterilize and drape patient.
6. For direct head tracking suture five sterile LEDs onto the marked locations. These have notches to duplicate the marked positions. Added overhead of this step is about 5 minutes.

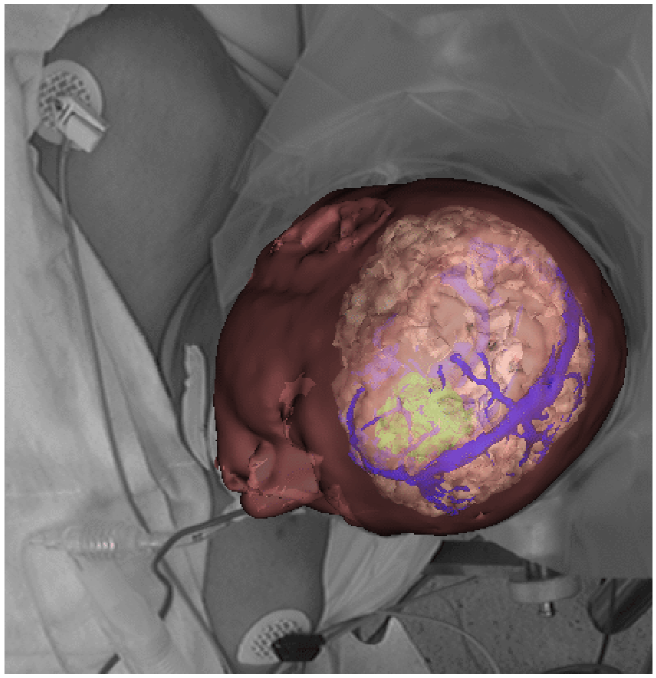


Figure 16 Enhanced reality visualization of the patient showing hidden internal structures overlaid on the surgeon's view of the patient. Tumor is shown in green, blood vessels in blue and brain in white.

7. Proceed with craniotomy and surgical procedure.
8. At any point, use sterile Flashpoint pointer to explore structures in the MR imagery.

Total overhead of performing the registration is about 10 minutes. The registration itself can take place in parallel with the surgeon's work. It generally runs in about 3 minutes, including verification time.

4.1.4 Operating Room Results

We have used the described image-guided neurosurgery system on over 20 tumor resection and vascular anomaly patients at Brigham & Women's Hospital. These patients were mostly tumor resection cases, consisting of metastatic tumors (cancer), both benign and malignant gliomas, and meningiomas. Typical tumor size was about 3cm x 3cm x 3cm. Some patients exhibited vascular anomalies, such as aneurysms or arterio-venous malformations. Some surgical targets were on the surface of the brain, but some were deep targets requiring unusual approaches, such as through the mouth. Most tumors were in the frontal, temporal, or parietal lobes.

The system effectively supported the surgery by providing guidance in the following tasks:

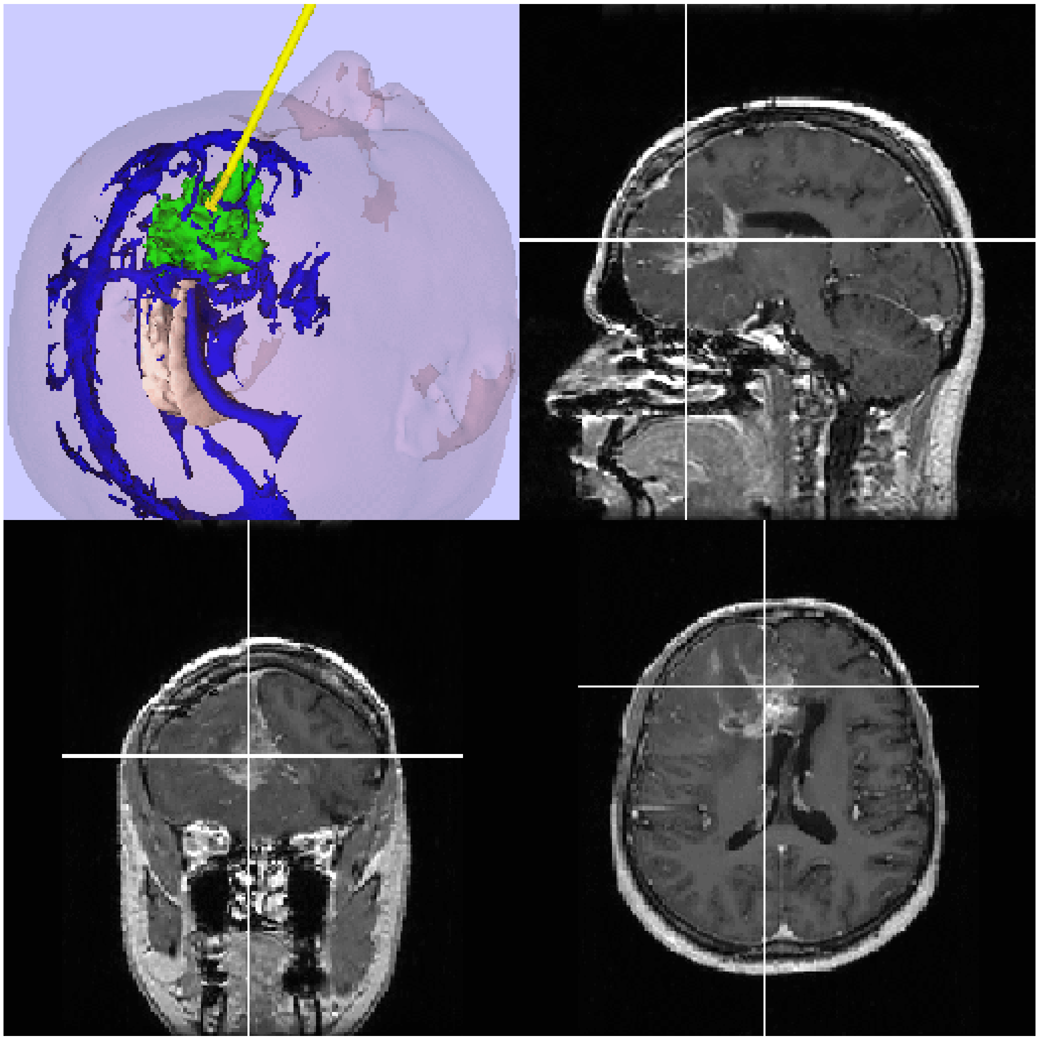


Figure 17 Pointer tracking in 3D MRI rendering and three orthogonal MRI slices.

- Planning bone cap removal—the craniotomy location should minimize bone cap removal while providing access to the tumor or anomaly in a trajectory which avoids sensitive brain structures.
- Identifying margins of tumor—in some cases the tumor's appearance is substantially different than brain tissue, but in others appearance is similar and guidance is helpful in ensuring complete resection.
- Localizing key blood vessels—the proximity of possible hidden blood vessels dictates resection strategy.
- Orienting the surgeons frame of reference—the pose of the head is not always aligned with the axes of the operating table; after draping the head the surgeon may benefit from guidance on how the head is oriented in the surgical field.

To qualitatively validate the system's performance the surgeon placed the pointer on several known landmarks: skull marks from previous surgeries, ventricle tip, inner skull bones such as eye orbits, sagittal sinus, and small cysts or necrotic tissues. He then estimated their position in the MRI scan, and we compared the distance between the expected position and the system's tracked position. In most cases, this error was within two voxels (MRI resolution was 0.9375mm x 0.9375mm x 1.5mm). When tracking the head via the skin-sutured LEDs, skin motion did appear to cause larger errors, up to 1 cm in a few cases, but such problems will be mitigated by using skull screws to mount head-tracking LEDs or by tracking the head clamp.

Feedback from the surgeons who have used the system has been highly favorable, with the instrument tracking visualization providing more useful information than the enhanced reality visualization. The instrument tracking capability provides more quantitative information than the enhanced reality visualization and can be tailored to answer a variety of questions, such as *how close is structure X to the pointer?* rather than just *where is the pointer?*.The resulting much improved insight into the surgical field increased the surgeon's confidence and resulted in shorter surgeries. In one case, one of the surgeons estimated that an otherwise eight hour procedure only took five hours.

4.2 Multiple Sclerosis Change Detection

A growing use of clinical imagery is the identification of medically-significant tissue changes over time. In studying conventional single images, medical professionals can locate possibly anomalous structures based on their knowledge of anatomy and pathology. But by *comparing current images against baseline or previous images*, clinicians may also be able to estimate the rate of change in the progression of a disease or as a reaction to a treatment. These improvements should facilitate earlier and more accurate diagnosis. The availability of relatively high resolution 3D raster datasets from diagnostic scanners and the growing emphasis on prevention and early detection of disease combine to underscore the need for accurate change detection technology.

4.2.1 Multiple Sclerosis Application Overview

One such change detection application is in tracking changes in multiple sclerosis (MS) brain lesions in MR images. These lesions, possibly numbering in the hundreds, are small patches of diseased tissue in the brain's white matter. The size and structure of these lesions change over time and indicate the state and severity of the disease. Tracking changes in the lesions can thus provide critical information in understanding the progression of the disease and monitoring treatment. By automating the change detection process we achieve the following benefits:

- Minimizing the cost of performing change detection procedures.
- Ensuring analysis of the complete dataset.
- Minimizing error introduced by manual intervention.
- Increasing the efficiency of change detection evaluations.

This change detection problem was a central issue in an NIH study conducted at the Department of Radiology at Brigham & Women's Hospital. The goal of the study was to achieve reliable high resolution change detection to enable the measurement of biologically significant changes in MS lesion burden. The study covered 50 patients exhibiting different stages of the disease, including chronic progressive, exacerbating-remitting, relapsing-remitting progressing, and inactive. The approximate MR imaging schedule for each patient was: weekly for the first eight weeks, bi-weekly for the next 16 weeks, and monthly for the next 26 weeks. Initial analysis of this data, without temporal registration, included collection of only global statistics over time, such as the changes in total MS lesion volume. But the accurate registration of our system facilitated tracking of the individual lesions over time and allowed study of local changes. By analyzing local changes

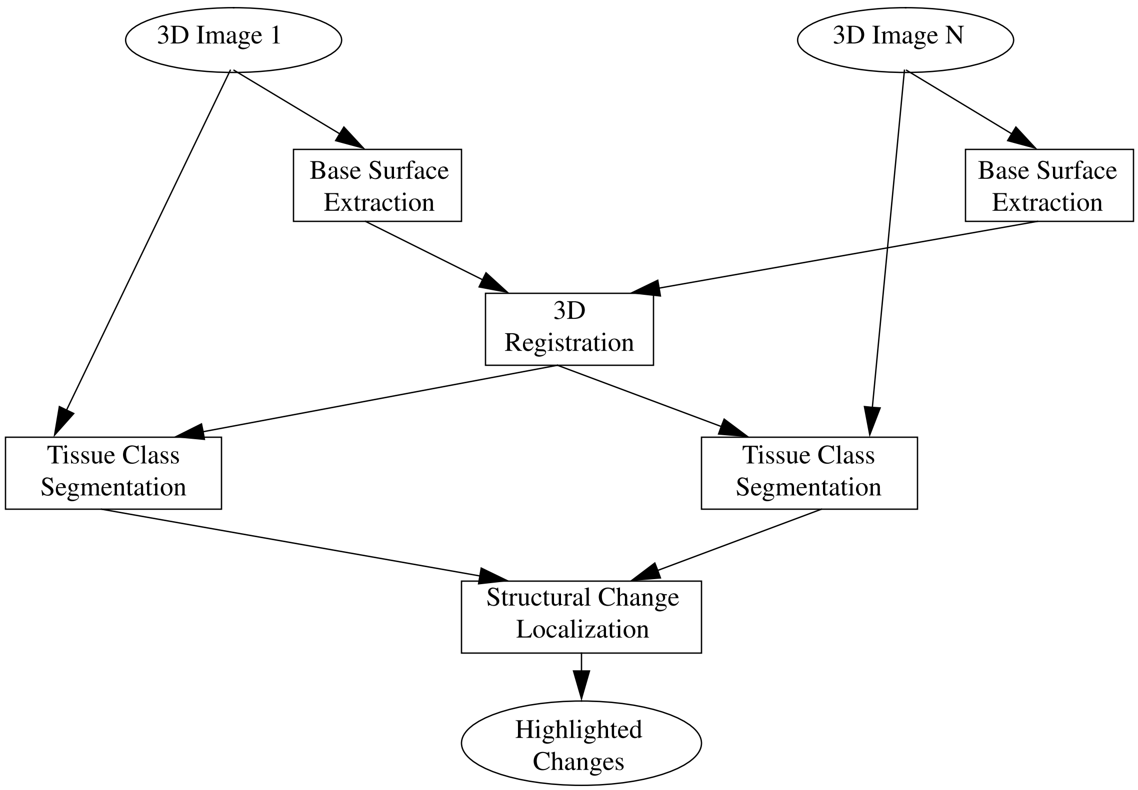


Figure 18 Change detection system architecture.

we can distinguish the independent contributions to total lesion volume of lesions evolving asynchronously—one lesion may be growing while at the same time another may be shrinking.

4.2.2 The Change Detection System

Our change detection system consists of four basic steps, as depicted in Figure 18:

- **Base Surface Extraction**—localization of surfaces of the same structure in both input images.
- **3D Registration**—compute a transformation that maps points from one dataset into the coordinate frame of the other dataset. The transformation is applied to all images such that they all have the same coordinate frame—i.e., a given 3D coordinate corresponds to the same physical location in all the scans.
- **Tissue Class Segmentation**—classify voxels in the 3D imagery according to their tissue type, such as grey matter, white matter, cerebrospinal fluid, and lesions.
- **Structural Change Localization**—identify changes in the tissue structure via 4D (space and time) connected component analysis, image differencing, local statistics, or direct visualization.

Our primary focus is on automating the registration process since that step generally depends most on global calculation consistency (i.e., small errors are often magnified) and can be difficult to perform manually with high accuracy. Semi-automated techniques can often be used for the other steps with less efficient but often adequate results. Errors in surface extraction can often be removed as outliers or canceled out in the registration process. Segmentation, while a difficult problem in general, can be less critical in change detection problems since imaging parameters can be tuned to highlight specific tissues of interest and changes can also be located with the raw imagery. The final step, change localization, is usually a straightforward visualization or analysis operation which is not difficult to automate.

4.2.2.1 Base Surface Extraction

We base the registration on the surface of the positionally stable intra-cranial cavity (ICC). This is the surface of the brain including the surrounding cerebrospinal fluid. Since the patients are positioned in approximately the same pose in the MR scanner for each data collection, brain motion relative to the ICC is minimal. As part of the NIH study all the patient scans were segmented using the automated segmentation tools in conjunction with semi-automated verification tools that are used at Brigham & Women's Hospital. The skin is not as useful a base surface since it is not completely imaged in all the scans and it requires pruning to remove soft tissue regions.

The ICC surfaces were segmented by constructing a non-parametric intensity classifier for tissues that are internal and external to the ICC and then applying morphological filtering techniques to refine the ICC mask [18] [27] [40]. Double channel MR imagery was used for the MS study at a resolution of 0.9375mm x 0.9375mm x 3.0mm, sliced axially. The two channels were proton density and T2 weighting.

4.2.2.2 3D Registration

The registration metric used in this study was the Saturated RMS Nearest Point measure (Section 3.3.1.1), solving for a rigid pose. For initial alignment we had the option of performing a principal axis alignment (Section 3.2), but since the patients were placed in approximately the same pose in the MR scanner each time, such alignment was not necessary. In fact, initial head position error was analyzed by applying the inverse of the registration transform for each scan, which we verified visually, to a set of points. If the registration transform for scan i, was $\boldsymbol{T_i}$, we applied $\boldsymbol{T}_i^{-1}$ to a set points, $\bar{p}_j$, on the scalp of the patient in the reference frame, and measured $\left|\bar{p}_j - \boldsymbol{T}_i^{-1}\bar{p}_j\right|$. The resultant distances determine how far corresponding points were in different images without registration. These distances averaged 9mm (maximum was 18.24mm), indicating that initial position was sufficiently close to seed the registration optimization process, but certainly not sufficiently close to forego the registration process.

The reference scan for each patient was chosen to be roughly in the middle of the collection period. All scans are registered to the reference scan and transformed to the reference coordinate frame. The transformed scan is interpolated to have the same coordinate axes as the reference. We use trilinear interpolation on the grayscale intensity to achieve the desired interpolation. Refinements on the interpolation may be necessary if the trilinear function proves inadequate.

4.2.2.3 Tissue Class Segmentation

For change detection applications we are usually interested in localizing and quantifying changes in particular tissues. Attempting to identify changes directly from the raw (MR) imagery is often problematic due to global and local sensor gain artifacts and magnetic susceptibility variations. By performing tissue segmentation we can systematically treat these problems, focus on the tissues of interest, and also provide improved context for change detection by introducing spatial knowledge of related anatomical structures.

A four class tissue segmenter [64] is applied to the MR imagery after applying the registration transform to the original image. The four tissue classes are: gray matter, white matter, cerebrospinal fluid, and MS lesions. Segmentation is applied to the imagery *after* registration since the registration transformation requires interpolating the transformed imagery to the same coordinate frame as the base image. The interpolation is more accurately applied to the grayscale imagery, where we apply tri-linear interpolation, than to the segmented imagery, where we would generally need to use nearest voxel labeling. The segmenter is the same one used for the image-guided surgery application, Section 4.1.2.1, but in this case we apply it to a two channel problem, proton density and T2 weighting.

4.2.2.4 Structural Change Localization

To identify changes over time we can look at simple image differencing to identify negative and positive changes in lesion volume. But since the number of lesions may be large and in this case we had up to 24 time points, a lesion tracking mechanism proved more useful as a tool to collect statistics on individual lesions. We apply 4D connected component analysis [46] to detect lesions with volume above a given threshold and track them over time. We then collect statistics on rate of change of lesions' volume and shape, where the statistics can be categorized by lesion location, initial lesion size, and disease state.

4.2.3 Change Detection Results

The MS change detection system demonstrated almost fully automated MS lesion monitoring that was successfully applied to a large trial dataset. Fifty patients were analyzed, each one being imaged 20-24 times during the course of a year. Results of the study are presently being analyzed by medical experts in order to determine relationship of lesion volume changes to other clinical indicators and to determine preferred change quantitation measures.

We analyzed the registration results for one patient for which 21 MRI datasets were available. For the 20 registration runs we used 7060 points sampled on the reference ICC surface and an average of 1415 points on the transformed ICC. The RMS measure was saturated at 5mm. The average final RMS error was 1.92mm. For these runs if we only ran the principal axes alignment the average RMS error would have been 2.38mm, indicating probable benefit to the surface registration approach. The RMS error is of course not a measure of deviation from ground truth, which is not available for these tests. Registration accuracy and region of convergence experiment results are given in Section 5.5.2.

The registration results for a sample slice position for a single dataset pair are shown in Figure 19. Changes are shown for the intensity images to avoid incorporating possible segmentation errors when evaluating registration accuracy. The main source of change inside the brain is due to MS lesion growth. Some change is apparent along the ICC surface, reflecting some possible ICC extraction errors. But such errors do not appear to generate registration biases that would manifest in more consistent positive or negative changes along the ICC contour.

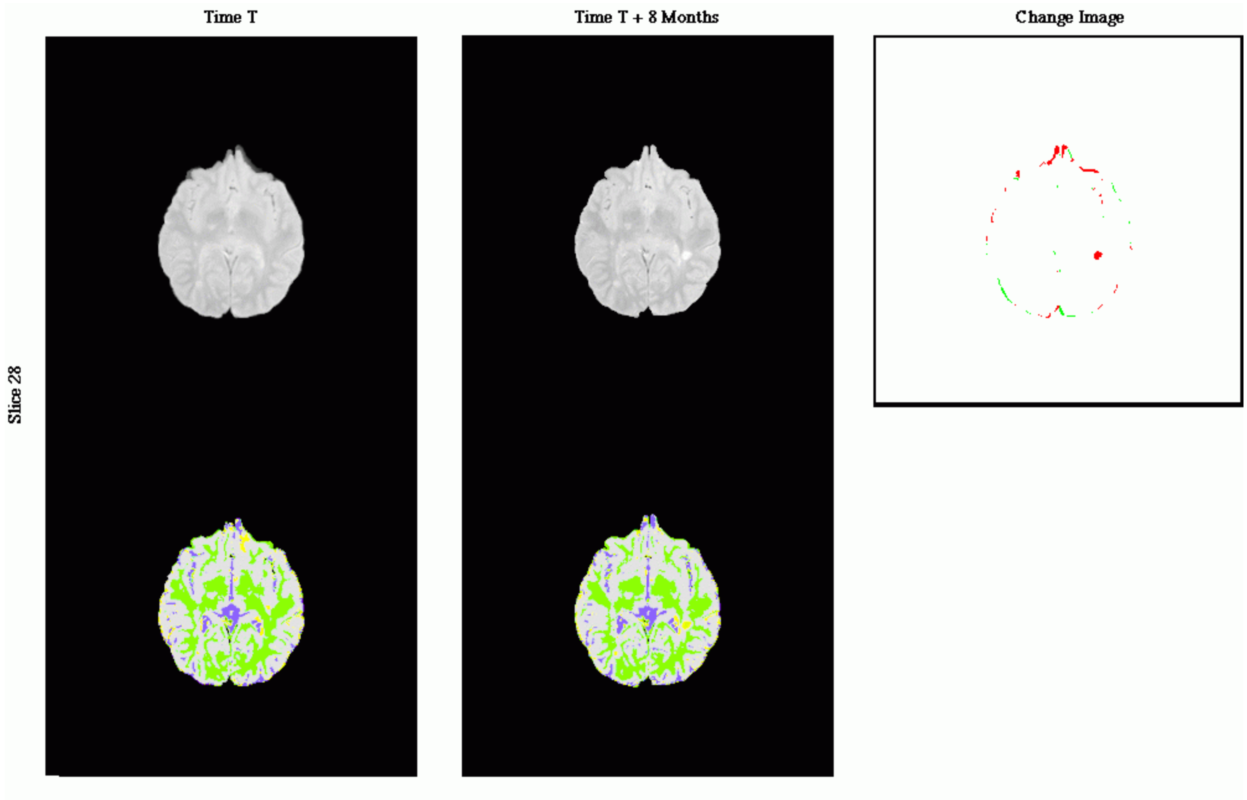


Figure 19 Changes in the same slice position. First column is transformed data which was registered to second column. Image in second column was taken eight months later. Top row is normalized intensity imagery from one of the two MR data channels, along with the absolute difference between the two: green is positive intensity change and red is negative intensity change. Bottom row is the segmented imagery: gray=gray matter, green=white matter, blue=cerebrospinal fluid, yellow=lesions.

4.3 MR Mammogram Change Detection

4.3.1 MR Mammography Overview

The mammogram change detection application focuses on detecting and diagnosing breast cancer. Conventional mammography uses 2D x-rays to image the breast. Although such images have high resolution, their 2D nature limits their ability to detect calcifications and lesions in dense tissue, particularly in younger women. There is clearly a need to reduce false negatives in order to allow early detection and treatment, but a reduction in false positives will also reduce the many unnecessary biopsies that are performed each year. As a result, interest in 3D mammography, such as with MRI, is growing [3][32][34]. While MRI is more expensive than x-ray mammography, potential cost savings lie in improved decisions in surgical management. Accurate definition of tumor size, number, and margins, as is feasible with high resolution 3D imaging, is critical for evaluating treatment options and for staging tumors for preferable breast conservation treatment. Furthermore, MR imaging is the method of choice for analyzing silicone implants. So although MR mammography is not yet routinely used, medical researchers are exploring its introduction into clinical practice.

Conventional MR does not exhibit high contrast between benign and malignant breast tissue, but the introduction of contrast agents, such as gadolinium-DTPA (Gd-DTPA), has greatly improved the discriminability of MR images [3][32][34]. These contrast agents accumulate in regions of increased vascularity, such as in the periphery of malignant tumors. As a result, malignant tumors appear enhanced in MR images, appearing brighter in T1-weighted pulse sequences. By comparing images prior to injection of the contrast agent and after injection, regions of enhancement and possible malignancy can be identified. The timing of the images is important since the most critical factor is the rate of uptake of the contrast agent. Cancers enhance quickly after the injection, whereas most benign disorders exhibit delayed enhancement.

As a result, there is a need to detect enhancements via change detection from the pre-injection image and the post-injection image(s). Although the breast is a highly flexible structure, substantial motion is minimized by avoiding patient motion [34]. The patient lies prone, with the chest wall supported by the table. The contrast agent is administered through an existing tube to avoid disturbing the patient. Image acquisition is fast—images of up to 1mm resolution in all three directions, which appears to be required for effective detection, can be acquired in about 5 minutes, although signal-to-noise may be sacrificed. The overall examination time is generally less than 30 minutes. During this time, we will have some motion, though, which must be corrected through image registration to maximize accuracy. It is expected that such motion will not be extensive and can be modeled via relatively simple transformations, such as affine. These image registration capabilities can also be applied to images of the same subject taken at different examinations, such as at annual check-ups, but a greater degree of motion is likely in these cases, thus requiring better modeling of the nature of the deformation.

4.3.2 The Mammogram Change Detection System

We consider the case of registering two MR images of the breast, one taken prior to injection, and another taken after injection. The system we use for this procedure is similar to the MS change detection application in Section 4.2, but here we solve for non-rigid transformations instead of the rigid poses which were sufficient for the head. The same steps shown in Figure 18 are followed here. The base surface we use for registration is the skin surface of the breast. It is extracted via thresholding, removal of small connected components, and directional tracking of neighboring skin points using the general orientation of the subject to initialize expected skin point orientation. The registration metric used in this study was the Direct Surface Distance measure, solving first for a rigid transform and then for an affine transform.

Once we register the two images, we transform one of the images to the coordinate frame of the other. We use trilinear interpolation to resample the transformed image to the coordinates of the reference. Differences may then be generated by subtracting grayscale values after normalization.

4.3.3 Mammogram Registration Results

We show results for pre- and post-injection MR images of a cancer patient. Each image consists of 60 slices of 2.5mm thickness and 1.25mm in-slice resolution. The two images were taken about 1 hour apart. Figure 20 shows representative slices of these images. In Figure 21 we render the skin surface extracted from one of the images. The segmentation was performed automatically—it includes some measurement noise, particularly on the non-breast surfaces. No internal structure segmentation was performed in this case—all differentiation was performed on the grayscale images.

It is not clear how much motion and deformation has occurred between the two acquisitions. We thus overlay the two images and evaluate pre-registration differences by subtracting the gray-

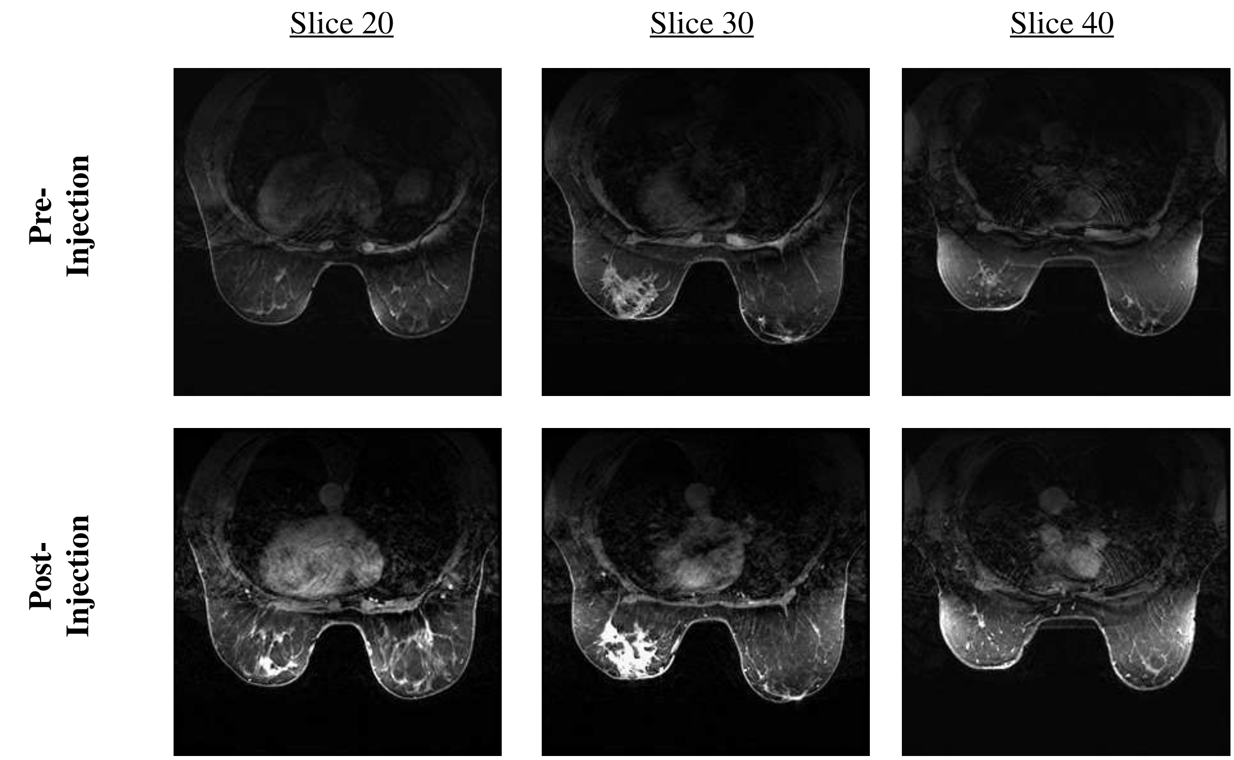


Figure 20 Selected slices from MR mammograms before and after contrast agent injection.

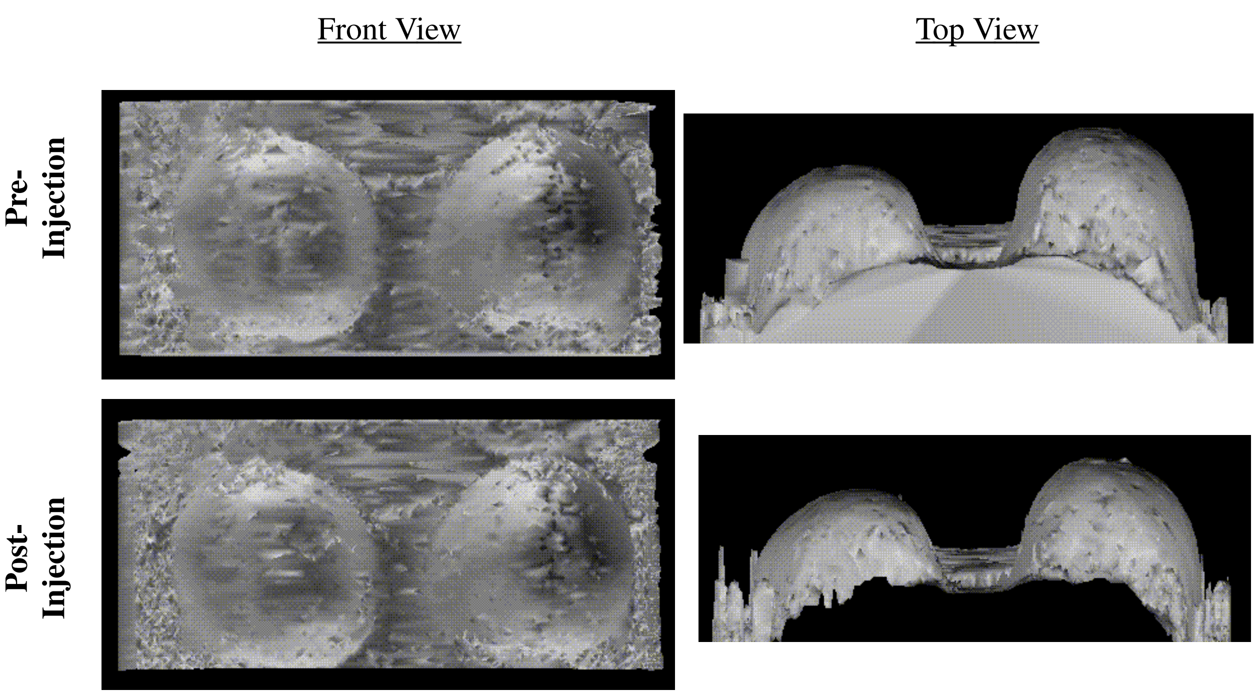


Figure 21 Rendering of skin surface segmented from the pre- and post-injection images. Front and top views are shown.

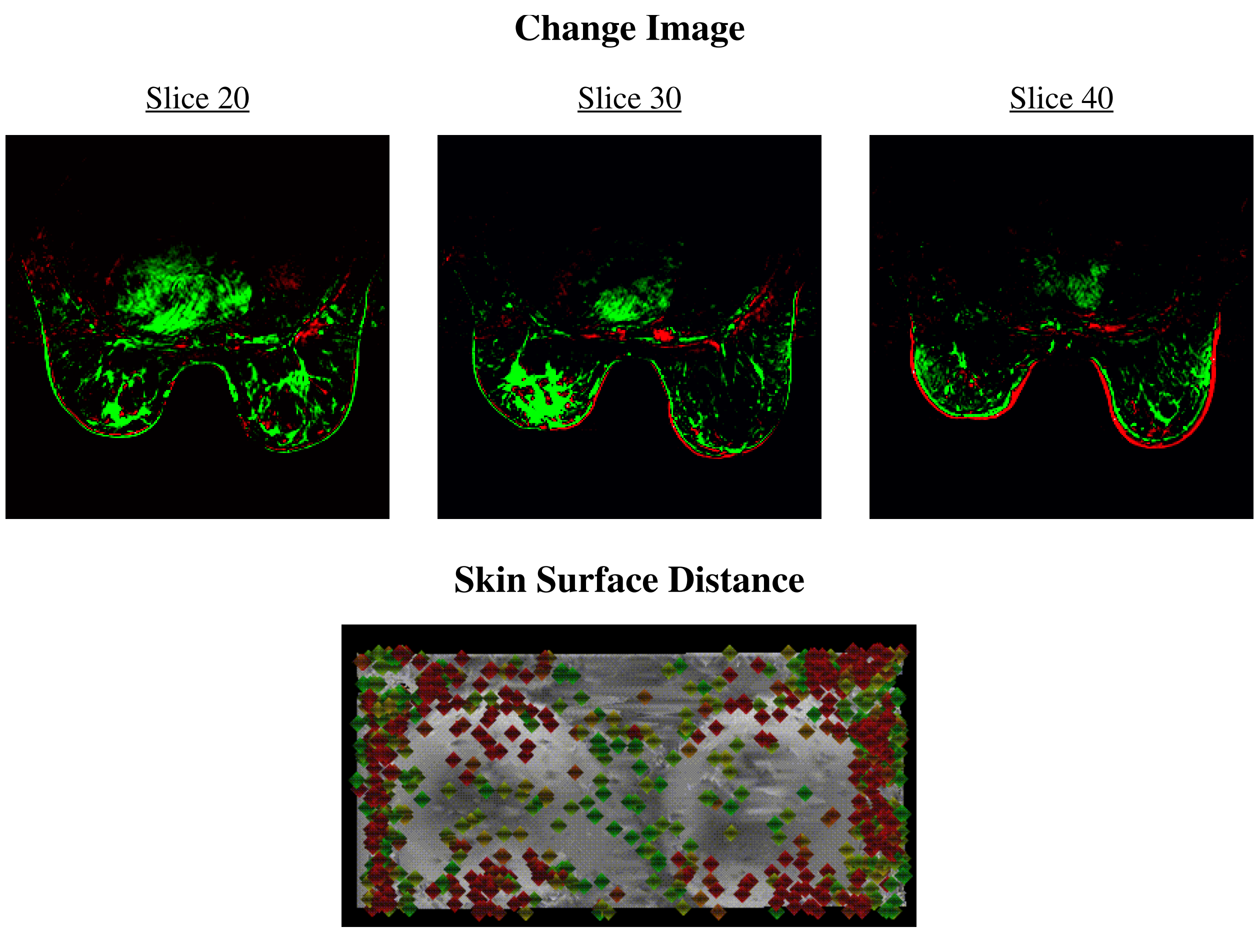


Figure 22 Image differences *prior to registration*. On the top we show representative slices of the grayscale difference image where green indicates positive intensity change and red indicates negative change. The brightness of the color is proportional to the magnitude of the change. On the bottom we show the color-coded distance from sampled points on the pre-injection skin surface to the post-injection skin surface which is rendered underneath. Distance ranges from green=0mm to yellow=1.5mm to red=3+ mm.

---

scale images and computing the distance from points sampled on one skin surface to the other skin surface. Both displays are shown in Figure 22. In the change image, we focus on the breast skin surface for measuring displacement. Since the skin is characterized by a relatively bright contour, mismatches will appear as bright green or red contours, but not necessarily both. If the pre-injection skin surface lies inside the post-injection surface we will see a bright green contour, while in the opposite case we will see a bright red contour. In the latter case we will also tend to see a concentric green contour since the post-injection image is generally brighter. In short, in the change image it is the *brightness* of the skin contour that indicates mismatch—perfectly matched skin surfaces will appear black.

We see that even without registration in some slices little change is evident, but in slice 40 the skin surfaces are clearly offset. In the distance display we clearly see significant skin displacement on the side surfaces of the breast. But it is also clear that we do not have large-scale deformations in this case and thus should be able to correct the small-scale deformations we do have via low order transformations.

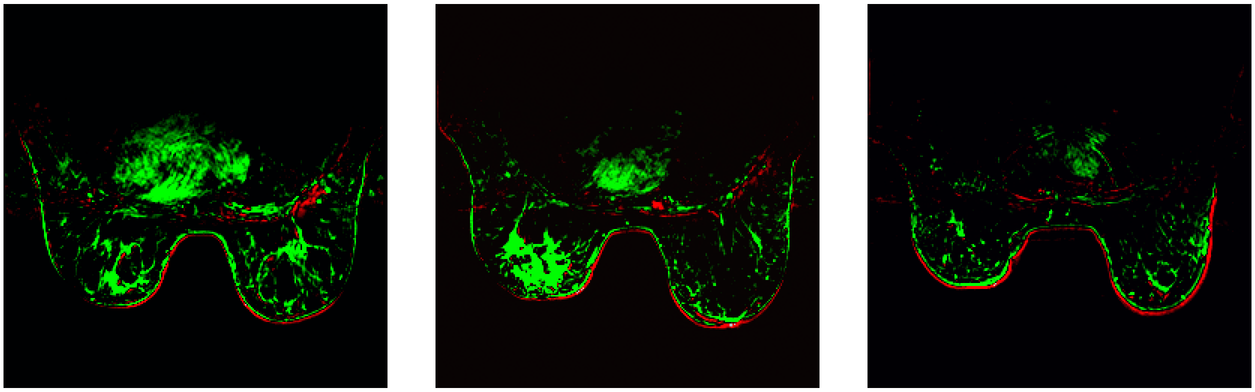
We perform the registration by transforming the pre-injection dataset to align with the post-injection image. The pre-injection skin surface is therefore represented by a set of points randomly sampled on the surface (1000 points in this case), and the post-injection image is represented by a set of triangular surface primitives (with maximum 0.5mm estimation error). We walk through the intermediate registration steps in order to gauge the benefits of our hierarchical registration approach. We first solve for the optimal rigid registration of the two surfaces. The results are shown in Figure 23. The rigid solution was a rotation of 0.6° and a translation of 6.3mm. We see that the displacements present in the initial pose have been reduced with the new rigid pose. Slice 40, which exhibited the greatest initial displacement clearly is better matched now. We also observe that the side surfaces of the breast now have smaller residual errors. The base of the breasts and the body skin surfaces have larger displacements due primarily to skin segmentation errors.

The rigid pose solution is then used by the system to seed the affine registration step. Here we solve for the optimal global affine transformation that minimizes the same point-to-surface distance metric. The affine solution is shown in Figure 24. It is clear from the skin surface distance display that the affine registration reduced the skin displacement present in the rigid solution. The change image, though, shows an interesting result of the linear scaling performed by the affine solution. We reduced the displacement on the inside walls of both breasts by increasing the horizontal scale of the pre-injection image, but this resulted in larger displacements on the outside walls of the breasts. The resultant affine transform appears to be an overall improvement over the rigid solution, but it cannot completely compensate for all the deformation in the image. This is a good example of where knowledge of the imaged structures can be used to decompose the problem into meaningful *local* affine (or rigid) registration problems. We can clearly obtain better results if we register each breast separately since their small-scale motions for which we are solving are independent of each other.

We therefore perform one additional step in which we create two local point neighborhoods in the pre-injection image, consisting of the points on each breast, and perform an affine registration

Change Image

Slice 20 Slice 30 Slice 40



Skin Surface Distance

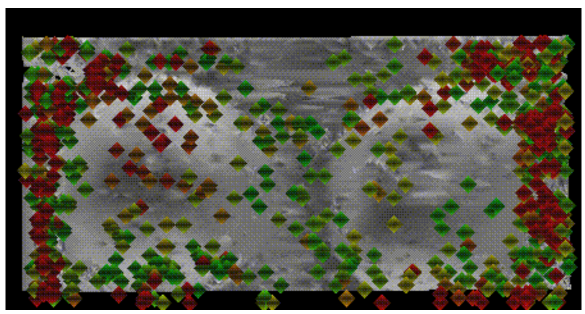


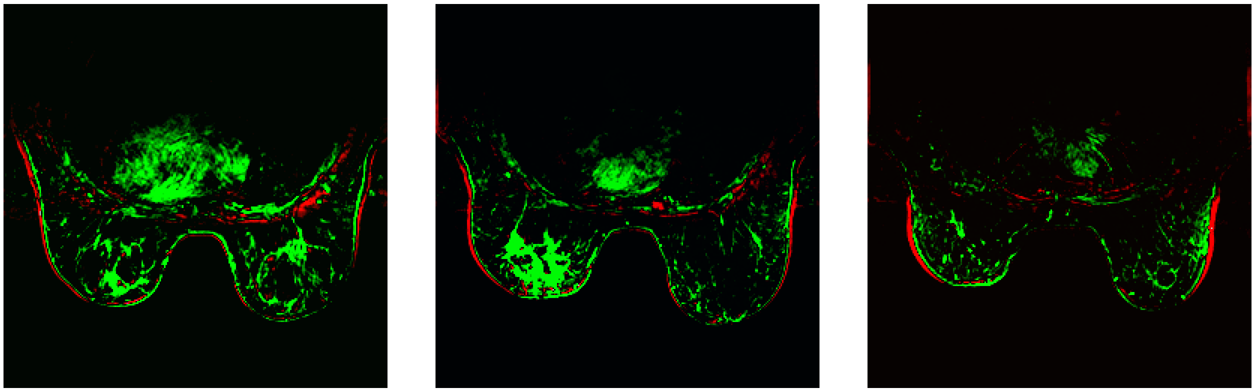
Figure 23 Image differences after *rigid* registration. On the top we show representative slices of the grayscale difference image where green indicates positive intensity change and red indicates negative change. The brightness of the color is proportional to the magnitude of the change. On the bottom we show the color-coded distance from sampled points on the pre-injection skin surface to the post-injection skin surface which is rendered underneath. Distance ranges from green=0mm to yellow=1.5mm to red=3+ mm.

for each point set. The decomposition was performed by manually specifying bounding boxes around each breast. The registration results are shown in Figure 25. We see that most displacements have been dramatically reduced, although not completely eliminated. We note that in this case, not only did the local nature of the registration improve the results, but the removal of the non-breast skin outliers along the body surface allowed us to focus on just the region of interest.

In conjunction with the improvement in registration accuracy we observe similar improvements in the value of the Direct Surface Distance metric that we optimize as we traverse the pose complexity hierarchy. The initial score prior to registration was -0.73 . The score improved to -0.78

Change Image

Slice 20 | Slice 30 | Slice 40



Skin Surface Distance

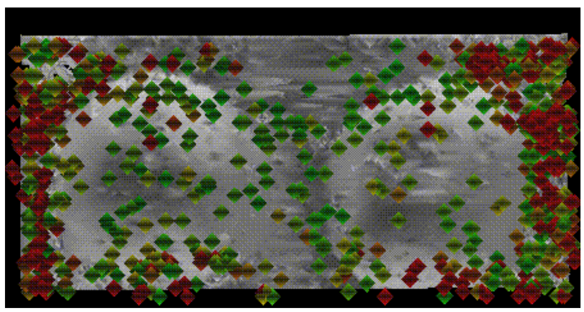


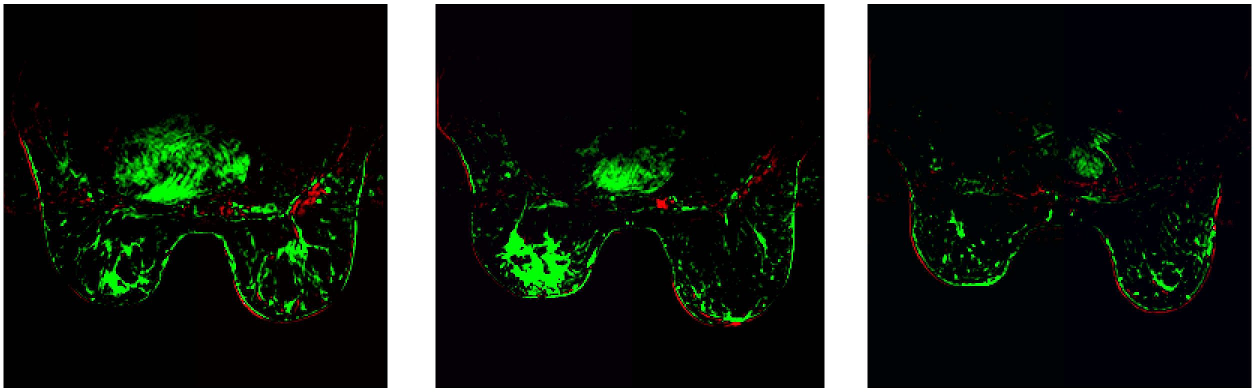
Figure 24 Image differences after *global affine* registration. On the top we show representative slices of the grayscale difference image where green indicates positive intensity change and red indicates negative change. The brightness of the color is proportional to the magnitude of the change. On the bottom we show the color-coded distance from sampled points on the pre-injection skin surface to the post-injection skin surface which is rendered underneath. Distance ranges from green=0mm to yellow=1.5mm to red=3+ mm.

after rigid registration and to -0.82 after global affine registration. The scores for the local affine registrations were -0.96 and -0.97. A perfect match would have a score of -1.0.

After completing the registration we perform the change detection, which in this case we have been continually computing through the change images. In these displays the internal enhancement regions are clearly evident in green. Of course, accurate registration was not necessary to detect such large regions. But in practice, higher resolution, shorter latency after contrast agent injection, and smaller pathologies will require the high registration accuracy we have demonstrated here.

Change Image

Slice 20 | Slice 30 | Slice 40



Skin Surface Distance

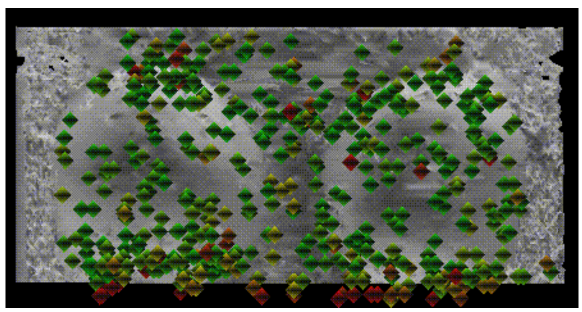


Figure 25 Image differences after *local affine* registration. On the top we show representative slices of the grayscale difference image where green indicates positive intensity change and red indicates negative change. The brightness of the color is proportional to the magnitude of the change. On the bottom we show the color-coded distance from sampled points on the pre-injection skin surface to the post-injection skin surface which is rendered underneath. Distance ranges from green=0mm to yellow=1.5mm to red=3+ mm.

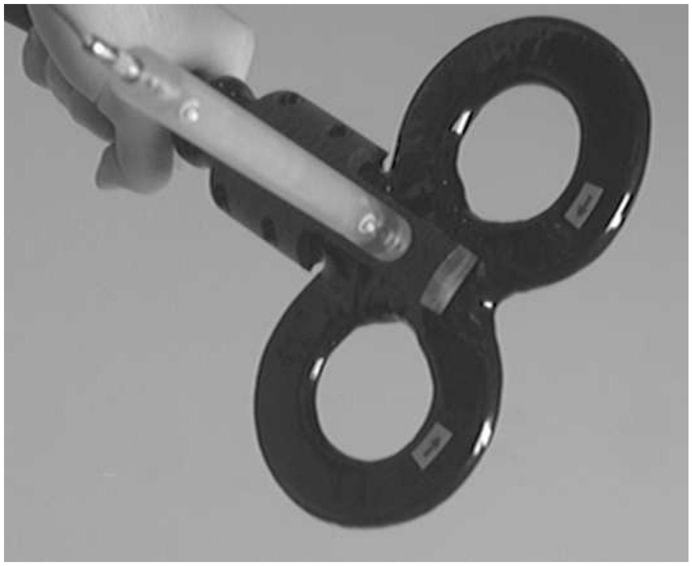


Figure 26 Stimulation coil used for TMS mapping.

4.4 Transcranial Magnetic Stimulation Brain Mapping

Functional brain mapping, consisting of the association of motor, sensory, and perception functions with different regions of the brain, is currently an active research area with a wide range of potential applications. Current non-invasive techniques for functional brain mapping utilize 3D medical scanners to image the brain while the subject undergoes an activity aimed at activating the functional area of interest. Scanners currently used for this purpose are single photon emission computed tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI). The ability of these scanners to capture brain activity results from their sensitivity to such factors as metabolism rate and blood oxygenation. The benefit of such scanners is their ability to quickly capture 3D snapshots of the complete brain activity. Some of their limitations, however, are high cost, use of radioactive agents, and passive control of functional activation.

4.4.1 TMS Application Overview

A promising approach to avoiding the limitations of classical functional brain mapping techniques is the use of a transcranial magnetic stimulation device (Cadwell Laboratories Inc., Washington and Magstim Company Ltd, England) for actively stimulating different parts of the brain. Such devices, first introduced for stimulating the human motor cortex by Barker [5], consist of a circular or figure-8 shaped coil, termed the TMS probe, which can deliver single magnetic field pulse stimuli or pulse trains. An example coil is shown in Figure 26.There is no direct electrical contact with the subject—the device works by inducing small electrical currents (estimated at < 0.25 A) in tissue using brief time-varying magnetic pulses that are focused in front of the coil. The peak magnetic fields are similar to those used with MRI scanners. Resulting energy dissipation in tissue is minimal (< 2 J/sec) [5]. The magnetic field generated by the coil passes through the scalp and skull with little attenuation and causes excitation of cortical neurons. This is one of the main advantages of magnetic stimulation over direct electrical stimulation—electrical impulses are highly attenuated by the skull and therefore can only be used when the brain is directly exposed. Magnetic stimulation is also able to penetrate to considerable depths without causing large electrical fields at the surface. As a result, no pain is associated with the stimulations.

Applications for TMS brain mapping include:

- Surgical planning—identifying the proximity of tumors to vital functional brain regions.
- Surgical guidance—tracking surgical activity relative to vital functional brain regions.
- Neuroanatomy research—building functional anatomical atlases and correlating functional maps with disease processes.
- Diagnosis—evaluating damage to functional activity of the brain.
- Therapy—using TMS as a therapeutic tool in treating such conditions as depression and akinesa [7].

We have explored stimulation of both the motor cortex and visual cortex. Excitations in the motor cortex result in peripheral responses of the affected muscles. Excitations of the visual cortex briefly inhibit interpretation of the affected visual field. The advantages of such a device are:

- Low cost and ease of use—the device is highly portable with few constraints on applicability.
- Active functional activation—rather than trying to spot brain activity when the subject performs different actions, the TMS attempts to directly stimulate certain brain regions and monitor resulting impact. By using latency measures to track response times, most voluntary responses are eliminated. In principle, this leads to a functional mapping that is highly localized both spatially within the brain and temporally for ease of acquisition.

While research is on-going on the biological implications of such a device, the physics of the generated magnetic field, and the development of psychophysical experiments which gauge brain function, we are exploring the technical problems of converting the locations of TMS probe stimulations and associated muscular/sensory responses to a 3D functional brain map. The key steps of our problem are:

1. Register the subject's MRI scan, the subject's position during stimulation, and the TMS probe positions/orientations to the same coordinate frame. For maximum accuracy we would like to avoid a fiducial based system.

2. Track the subject's head motion in order to maintain the registration. Head clamps are to be avoided both for the subject's comfort and to allow free access to the whole head region.
3. Combine the TMS responses, TMS-to-MRI transform, and head motion to generate a functional brain mapping on any 3D surface rendered from the MRI scan. Real-time visualization of the current probe pose in the MRI coordinate frame along with an encoding of previously probed locations is to be used to guide the acquisition of subsequent stimulation points. Such guidance avoids redundant probings, reduces acquisition time, and supports accurate delineation of “hot spot” boundaries.

A few research groups have addressed these problems in different ways. Wassermann *et al* [61] used a least squares surface alignment algorithm similar to ours for generating functional brain maps from TMS with a reported accuracy of 6mm. They used a Polhemus magnetic digitizer to collect surface points along the scalp for registration to the MRI skin surface. Epstein *et al* [20], Levy *et al* [42], and Meyer *et al* [47] have also shown overlaid TMS functional maps on MRI brain surfaces, but the registration was performed using fiducials or stereotactic frames. In other related work, medical researchers (Barker *et al* [5], Amassian *et al* [4], Maccabee *et al.* [44], Wassermann *et al* [60], and Pascual-Leone *et al* [48]) have generated functional maps by essentially drawing grids on the scalp, but did not appear to transfer the maps directly to the cortical surface.

4.4.2 The TMS Brain Mapping System

We developed a system to generate the TMS brain maps in a collaborative effort between MIT’s Artificial Intelligence Lab and the Departments of Radiology, Anesthesiology, Psychiatry, and Neurosurgery at Brigham & Women’s Hospital. The architecture of this system is shown in Figure 27. The components of this system are similar to our image-guided neurosurgery system presented in Section 4.1.2, where they are presented in greater detail. We work with three different coordinate systems:

1. **MRI**: The MRI data is captured in the scanner coordinate frame. The data itself is segmented into skin surface for registration and internal structures for brain mapping visualization.
2. **Laser**: The laser scanner provides 3D data of the subject's scalp surface as positioned for transcranial magnetic stimulation. Here the coordinate frame of the acquired points is centered at a fixed point within the working volume of the laser system.
3. **Flashpoint**: This LED tracking system is used to track subject motion and localize the TMS probe. We mount two LEDs on the TMS coil (specifying its 3D position and orientation, with twist the only degree of freedom not measured) and tape five LEDs on the sub-

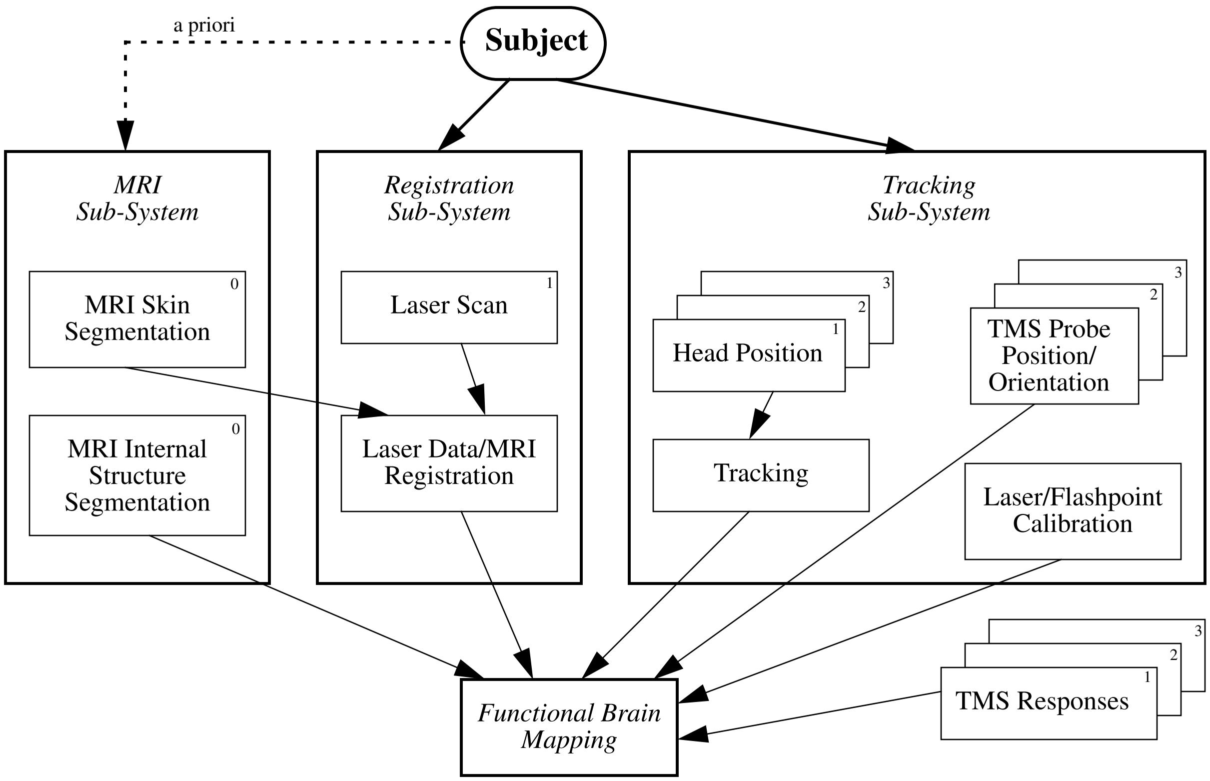


Figure 27 TMS functional brain mapping system architecture.

ject's scalp for tracking head motion. Redundant LEDs are used for tracking head position in case motion is great enough to block up to two of the LEDs. The position and orientation information returned by the system are represented in a coordinate frame centered at a fixed point within the working volume of the Flashpoint cameras.

The goal of the system is to integrate all of these coordinate systems into a single reference frame. That is, we need to relate sampled TMS probe points to the corresponding points in the MRI scan, which we do by using the laser coordinate system as an intermediary. The laser scanner (laser and its associated camera) and Flashpoint system (three linear cameras) are mounted on the same bar which is attached to a movable arm for ease of placement, as shown in Figure 11. Since we fixate the laser and Flashpoint systems relative to each other, we perform an off-line calibration to obtain the Flashpoint-to-laser transform. This transform is then constant for all subsequent TMS data collections. The transform from laser data to MRI coordinates, though, must be computed for each TMS data collection since it depends on the subject's pose during the mapping session.

4.4.2.1 Mapping Procedure

A sample data collection procedure, from the perspective of the subject, is:

1. Acquire an MR image of the subject prior to the TMS session. Segment the scan, using automated techniques [39] [64] and semi-automated techniques used at Brigham & Women's Hospital, into desired anatomical structures, generally skin, gray matter, white matter, and, if necessary, tumor.
2. Set up TMS data collection:
   - For motor cortex mapping, place muscle activity sensors on muscles of interest. Muscle activations with latencies of about 15 to 30 msec indicate successful stimulation.
   - For visual cortex mapping, place subject in front of a computer screen which will flash sequences of letters in the right and left half of the visual field that the subject will attempt to read. Magnetic stimulation is synchronized to be about 0.1 seconds after flashing the letters. False or no readings indicate successful visual activity suppression.
3. Place Flashpoint LEDs on rigid points of the subject's scalp. Currently these are loose LEDs taped in widely separated locations on the skin such that they will not interfere with the TMS probing. For visual cortex mapping we place the LEDs on a tight-fitting swim cap so that they are visible during stimulation on the back of the head.
4. Laser scan the subject—the laser plane is swept across the subject's head collecting 3D positional data of visible skin surfaces. At the same time the positions of the five LEDs taped to the subject's head are acquired by the Flashpoint system.
5. Collect TMS data—the TMS probe is placed at various points on the subject's scalp, as in Figure 28. At each point, the TMS generates a brief magnetic pulse and the responses from the muscle sensors or subject's view of the computer screen are recorded. The position and orientation of the TMS probe are recorded by the Flashpoint system at the same time. 3D renderings of the subject's MRI skin superimposed with TMS points are generated during the data collection to chart progress and guide continued stimulations.

4.4.2.2 Laser Data / MRI Registration

The key to achieving high accuracy in this application is the registration of the MRI data with the actual subject. We accomplish this task by aligning skin surface from the MRI data to skin surface from the subject acquired with a laser scanner. Accuracy requirements are relatively high as many of the active brain centers being studied are on the order of a few mm^3 in volume. Thus the overall accuracy, including any tracking errors, should be within about 1-2mm, which is generally not much larger than the voxel resolution of the MRI scan. We use the same registration techniques developed for the image-guided neurosurgery system as described in Section 4.1.2.2. These algorithms solve for a rigid transform using our Averaged Nearest Neighborhood and Saturated RMS Nearest Point optimization metrics applied sequentially or separately.

Figure 28 Subject set up for TMS stimulation. LEDs mounted on the rod, orthogonal to the coil, track the position and orientation of the coil. LEDs taped to the face are used for tracking head motion.

4.4.2.3 Probe and Head Tracking

The 3D position of the TMS probe is directly obtained by tracking the Flashpoint LEDs that are rigidly attached to it. In order to track the head motion we record the reference position of the LEDs taped to the patient's head at the time we perform the laser data/MRI registration. When the TMS probe is stimulated we record the new position of the head-mounted LEDs and compute the transform necessary to return the head to its reference position. This transform is applied to the position/orientation of the TMS probe at the time of the corresponding stimulation in order to apply the laser data/MRI transformation. Since we may have up to five LEDs to track the head we use Horn's closed form least-squares solution [36] for the tracking transform.

4.4.2.4 Functional Map Generation

We combine the registration and tracking data to obtain the functional brain mapping using these transforms:

- F_L—transformation from Flashpoint coordinates to laser coordinates; calibrated a priori.
- L_M—transformation from laser coordinates (head at time 1) to MRI coordinates; computed from the dynamic registration procedure.
- $H^t{}_r$—transformation of head from time t to reference position at time 1.

We have also collected the following TMS data:

- $C^t{}_p$, $C^t{}_o$—position and orientation of TMS coil at time t, $t \in [1, T]$, in Flashpoint coordinates
- For motor cortex mapping we receive $r^t{}_j$, the measured first response of muscle j to stimulation t, and $w^t{}_j$, the latency from stimulation to the response.
- For visual cortex suppression we record the letters the subject saw flashing on the computer screen. The set of letters are separated into left and right fields of view for separating the processing of the two sides by the brain.

In order to compute the brain mapping we need to map the TMS responses to the brain surface using the measured coil positions/orientations and associated transformations. To perform this mapping, for each stimulation t, we process those MRI surface points, $S[i]$, that are sufficiently close to $\boldsymbol{L_M F_L H^t{}_r} C^t{}_p$ to have been possibly stimulated by the pulse. For each $S[i]$ we compute the distance, $d^t[t]$, to the line defined by the point $\boldsymbol{L_M F_L H^t{}_r} C^t{}_p$ and the orientation $\boldsymbol{L_M F_L H^t{}_r} C^t{}_o$. For motor cortex mapping we are currently using a Gaussian weighting function proportional to that distance to "spread" the response $r^t{}_j$ to the points $S[i]$. The purpose of this simple weighting function is to interpolate across the stimulations to obtain a smooth and visible map. If we let $\mathrm{map}^t{}_j[i]$ represent the mapping of response j to stimulation t on the selected surface, then $\mathrm{map}^t{}_j[i] = G(d^t[i], \sigma) r^t{}_j$, with G being the Gaussian weighting function. We then let $\mathrm{map}_j[i]$, the composite mapping from all stimulations, be the maximum $\mathrm{map}^t{}_j[i]$ over all t, which are then normalized over i.

For visual cortex mapping we treat the responses as binary and thus directly map the points that caused positive suppression in the left and/or right fields of view to the MRI surface points, $S[i]$.

4.4.3 Brain Mapping Results

We have performed the TMS brain mapping on ten motor mapping subjects and five visual mapping subjects. Three of the subjects were neurosurgery patients. We use two methods for visualizing the motor cortex mapping. One method, shown in Figure 29, displays the minimum latency response (above threshold) for all the muscles evaluated. The other method, shown in Figure 30, displays the response strength for individual muscles. The responses are localized in their anatomically expected locations. In Figure 30 we see all the stimulations that were collected during the session. Since the stimulation points are visualized in real time, such a display provides feedback for guiding subsequent probes, avoiding redundant probes in no response areas and accurately

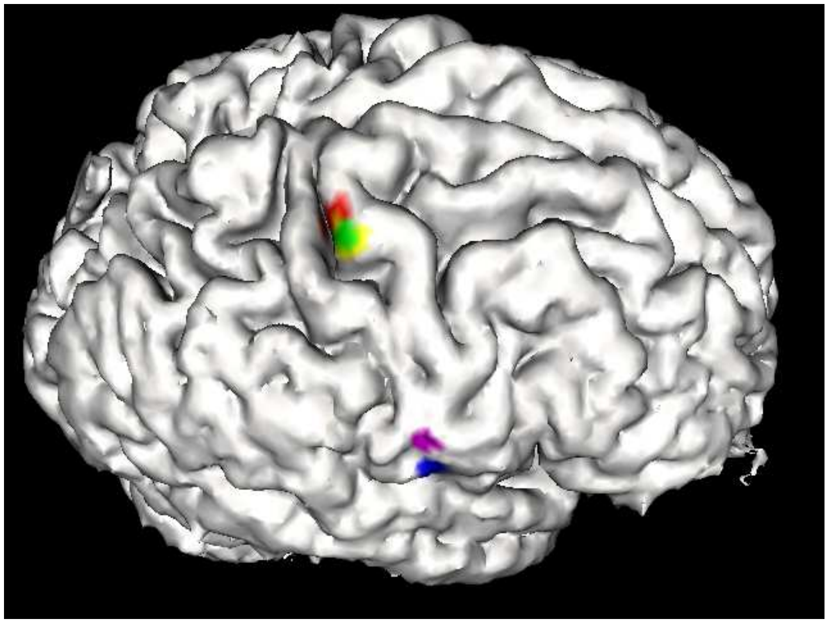


Figure 29 Motor cortex mapping showing minimum latency response for: index finger (red), fore-arm (yellow), biceps (green), jaw (blue/purple).

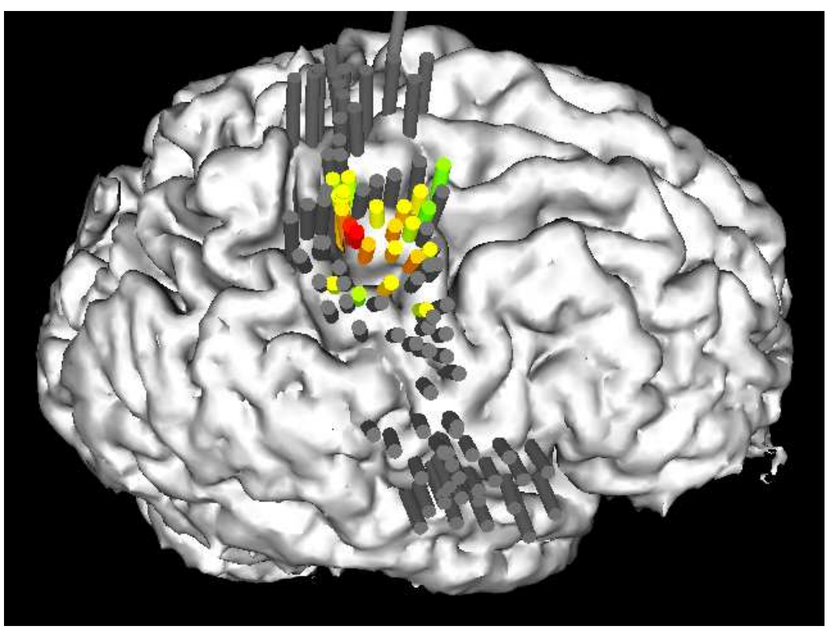


Figure 30 Motor cortex mapping showing response strength for index finger muscle. Oriented stimulation probes are color-coded by response strength, varying from none to strongest in the order: black, green, yellow, red.

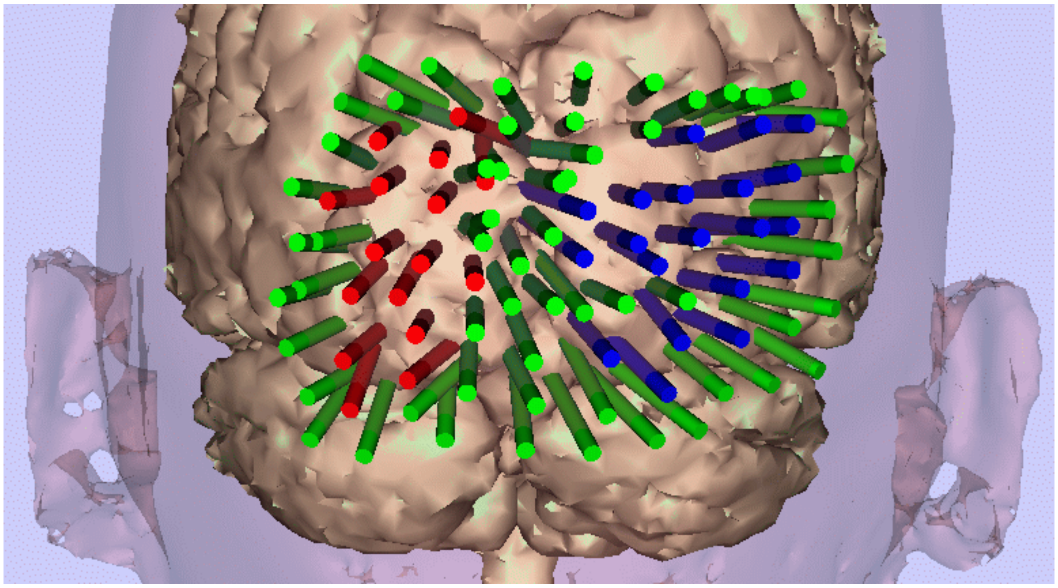


Figure 31 Visual cortex mapping showing stimulation points color-coded by visual suppression response: red = right visual field; blue = left visual field; green = neither.

delineating response region boundaries.

Figure 31 shows results of visual cortex mapping from one of our subjects. We see that the visual suppression effects are well localized and separated across the right and left fields of view.

4.4.4 TMS Validation Testing

Although our functional mappings were positively reviewed by radiology and neurology specialists, it is difficult to validate the results as we have no functional map ground truth for the tested subjects. We are thus pursuing two avenues for validating our results:

- **Map repeatability**: At a minimum we would like to duplicate functional mapping results on the same subject at different time points. Such results are not definitive indicators of mapping accuracy, but do gauge the reliability of our registration and tracking techniques in the context of functional mapping. We have run such repeatability trials on two subjects in which we mapped the same side of the motor cortex at two different time points. Figure 32 shows the biceps muscle maps generated for one of the subjects at two different times. Initial measurements on distances between minimum latency stimulations of the same muscle resulted in repeatability errors of about 1 cm or less. Future analysis will consist of more accurate methods of evaluating repeatability such as weighting multiple small latency stimulations by their response amplitudes and using the same stimulation grid pattern which is saved for each subject with his MRI scan.
- **Surgical validation**: An exact validation can be obtained in the operating room in the case of craniotomy surgeries. Surgeons currently use electrical stimulators to directly stimulate

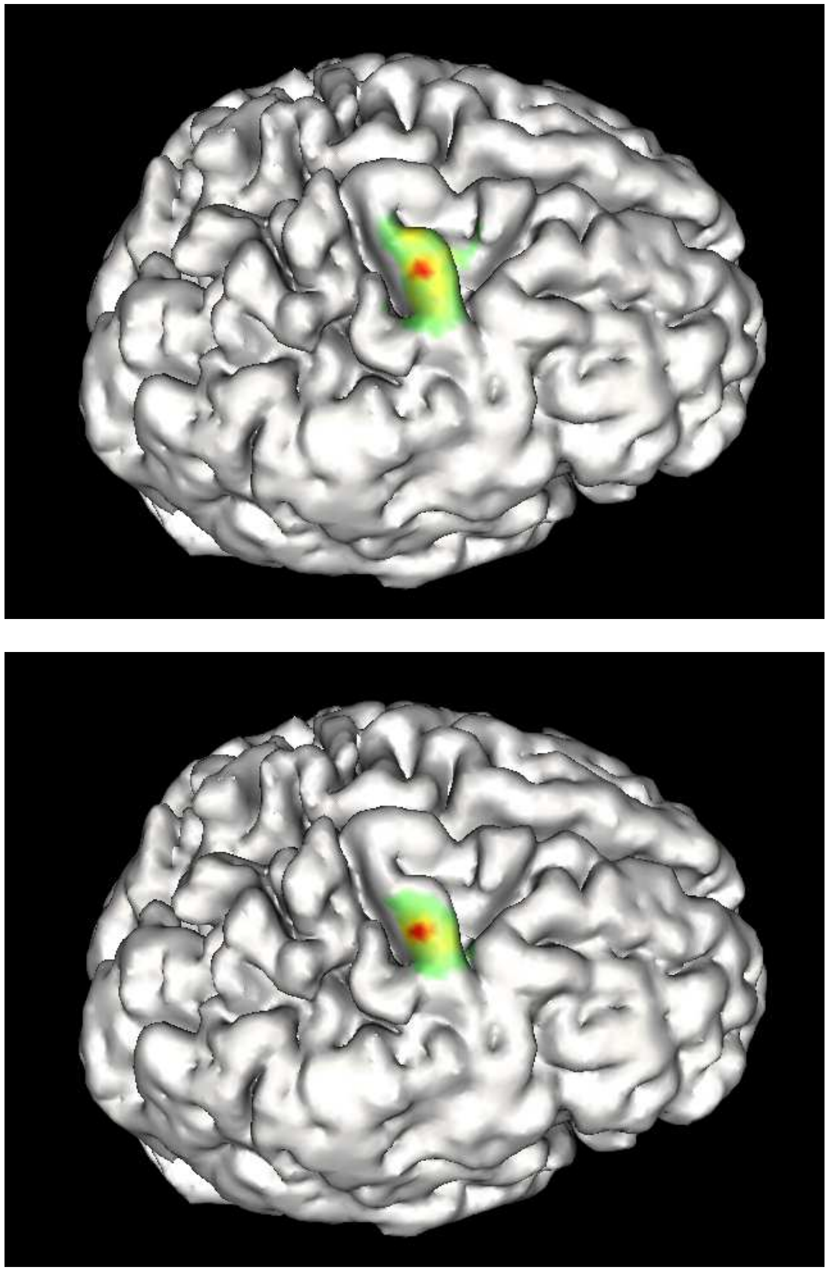


Figure 32 Comparison of the biceps amplitude response mapping of the same subject at two time points.

the brain surface when they are operating near the motor or sensory cortex. By tracking the position of such stimulators relative to an MR scan on which we have overlaid the functional mapping we can verify the maps. Such tracking in the operating room can be performed using our image-guided neurosurgery system. Future trials will explore such quantitative validation, but in a neurosurgery case which we had mapped, the surgeon qualitatively validated our results as accurate. The localization of the motor cortex relative to the tumor of this patient, which was used for surgical planning is shown in Figure 33.

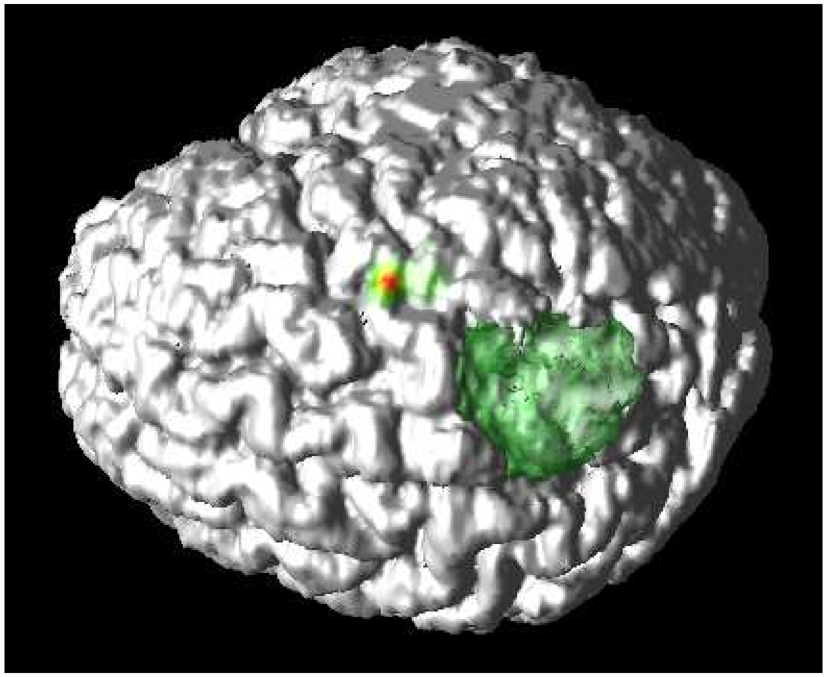


Figure 33 Finger muscle amplitude response mapping (strongest response in red) of a neurosurgery patient. Tumor is large green structure.

4.5 Application Development Summary

We have shown the application of automated surface registration techniques to a number of high-utility medical applications: neurosurgical planning and guidance, diagnostic brain and breast change detection, and non-invasive functional brain mapping. These applications require high registration accuracy and benefit from stable and robust registration performance. The resultant benefits of better leveraging 3D medical imagery include:

- Improved surgical planning through non-invasive brain mapping and integration of multiple imaging modalities.
- Improved visualization into the surgical field which facilitates better migration of surgical plans to the patient, reduces probability of errors, and results in overall shorter procedures.
- Possibility of more minimally-invasive surgeries due to the reduction of hand-eye coordination problems and improved navigation guidance.
- Earlier disease diagnosis and treatment effectiveness indications through change detection analysis.
- Quantitative and localized analysis of structural and functional brain properties for neuroanatomy research.

We have shown promising results in achieving such benefits by testing our implemented prototypes on clinical data and in clinical settings. But more quantitative evaluation is required for more routine use of these prototype systems.

Chapter 5

Registration Algorithm Evaluation

In order to achieve accurate and reliable registration performance in the applications described in Chapter 4 as well as related applications we need to understand the benefits, limitations, and sensitivities of the registration algorithms. Our stated registration goals, given in Chapter 2, were to achieve:

1. **Robustness in the presence of input data errors**, such as surface extraction deviations, data outliers, and imaging distortions.
2. **Effective performance for varying data resolutions**.
3. **Minimum dependence on initial alignment**.
4. **Avoidance of local minima**.
5. **Verification of registration results**.

Our approach to achieving these goals is to characterize empirically the registration algorithms in terms of how well they satisfy the goals and to identify which components of the algorithm prove to have the most impact on solving these problems. Once we have a good characterization of registration performance we can evaluate whether that performance will satisfy the requirements of a particular application.

In characterizing registration *performance*, a range of measures may be used. We employ the following terminology to describe some of the different aspects of performance that may be analyzed:

- **Accuracy**: The degree to which physically corresponding points on the reference and test surface have been brought into alignment. Or similarly, the degree to which the unknown pose has been recovered by the registration pose. We measure pose recovery accuracy as described in Section 5.2.2.
- **Stability**: The degree to which the same results are obtained in a given class of registration problems. We measure stability as the standard deviation of the accuracy metric.
- **Robustness**: The ability to withstand a range of data variations such as initial pose offset, measurement error, or incomplete surface coverages. We measure robustness as a region of convergence across a controlled data variation parameter. We measure the region of convergence as the set of data variation parameter values in which accurate and stable results are obtained, generally about 1-2mm with standard deviations of less than 2mm, although such measures are in general application dependent. For example, pose region of convergence defines the extent to which pose offsets between test and reference datasets can be recovered. But we can also define regions of convergence for maximum acceptable measurement error or minimum required surface coverage. In determining the region of

convergence we are interested in not only the size of the region, but also in the manner by which accuracy degrades as we approach and cross the convergence region boundary. In some cases we may experience a sharp drop off in performance at a certain point, while in others, performance degrades more gradually. That is, the region convergence may sometimes have a well-defined boundary, while in other cases it may be much fuzzier. Gradual degradation may be preferable, although sharp degradations may sometimes be easier to automatically detect—i.e., the system may know it has generated an incorrect result.

- **Success rate**: The percentage of trials in which an accurate result is achieved. We do not explicitly calculate success rate since it requires an accuracy threshold which may be highly application dependent and drastically impact perceived performance results. More importantly, the performance continuum from success to failure may involve a smooth evenly populated histogram of results, making it difficult to determine a reasonable separation of success from failure. Since we do not use a success rate measure, our accuracy results are a compilation over all trials in each run and we use the stability and robustness measures as indicators of success rate.

Ideally we would like to perform our experimentation on real subject data, but this approach is problematic due to the lack of detailed ground truth. Phantom studies can provide us with the ground truth we desire, but sufficiently realistic phantoms are difficult to construct. Instead we perform our experiments on real data, but with controlled data variations. The two types of such experiments are:

- **Simulated test data**.

  The reference dataset is an actual MR image. The test dataset is constructed by copying the reference dataset and applying various perturbations to it. In this manner we know the exact correct registration and we can directly control the perturbations against which we want to measure our sensitivity.

- **Application test data**.

  The reference dataset is an actual MR image. The test dataset is also actual data of the same subject, but from a different sensor or from a different time point. In order to estimate ground truth we run the registration algorithm under the most ideal matching conditions we can achieve and manually verify the registration. We then apply controlled perturbation from this registered condition to measure the registration sensitivity to those perturbations. These tests are designed to more closely resemble operating conditions in the actual applications. Fiducial markers can potentially be attached to the subjects to support this type of experimentation, but were not available for our test imagery. Additional MR or CT scanning was beyond the scope of this research. Fiducial markers are also impractical for images that are taken over long durations for which reproducible fiducial placement may be difficult and accurate localization is required.

The experiments are structured to answer the following data-dependent questions:

1. What is the sensitivity of the registration to incomplete overlaps of the surfaces being registered? That is, how does performance degrade as the overlap between the two images decreases or as outlier points are introduced. Outliers may be introduced by the sensor or

segmentation algorithm. This question addresses part of goal 1, robustness to data deviations.

2. What is the sensitivity of the registration to measurement errors of the basis surfaces? Such errors may be introduced by the sensor or segmentation algorithm and may be a function of the image acquisition parameters. This question addresses part of goal 1, stability in the presence of input data errors.
3. What is the sensitivity of the registration to data resolution? In particular, how does registration performance degrade as the resolution of the reference surface decreases. This question addresses goal 2, effective performance for varying data resolutions.
4. What is the registration algorithm pose region of convergence? That is, how does registration performance degrade as initial pose offset between the two datasets grows. This question addresses goals 3 and 4, minimum dependence on initial alignment and avoidance of local minima.
5. How does the objective function relate to registration accuracy? None of the objective functions are true measures of registration accuracy—they are just measures of alignment. Accurate registrations have good alignments, but good alignments may occur at multiple poses such that the incorrect alignments have nearly as good an objective function measure as the correct alignment. While it may be difficult to determine objective function thresholds in order to make registration success decisions, we want to evaluate the degree to which the objective function qualitatively reflects registration accuracy. That is, as the registration accuracy improves, the objective function should improve as well. This question, in conjunction with the qualitative visualization tools shown in Chapter 4, addresses goal 5, verification of registration results.

In addition we are interested in answering the following algorithm-dependent questions:

6. How do the point-to-point and point-to-surface objective functions compare in satisfying questions 1-4?
7. How do the different hierarchical refinement strategies compare in satisfying questions 1-4? These strategies consist of: distance weighting, surface sampling, and surface smoothing.

5.1 Experimentation Roadmap

After a description of the experimentation framework and procedure in Section 5.2 we describe the results of the experiments. We first describe the data sensitivity results which measure relative performance against separately-controlled data variations. We then present comparisons of the different objective functions and hierarchical refinement techniques. Finally, we describe results of experiments from several actual applications. All tests are for head registration applications, with different modalities and structure surfaces used for registration. A summary of the experiments is given in Table 1.

Table 1 Experiment roadmap.

Experiment Category	Experiment Description	Document Section	Key Conclusions
Data Sensitivity Simulation	Outlier addition	5.3.1.1	Only slight degradation is observed with increased outliers.
	Surface patch removal	5.3.1.2	At least 40% surface coverage is necessary for stable and accurate convergence from 15° rotation offset.
	Measurement error	5.3.2	Only graceful degradation is observed with increased measurement noise.
	Data resolution	5.3.3	Registration remains stable under reduced resolution, but accuracy bias is observed.
	Pose region of convergence	5.3.4	Accurate and stable registration results are obtained up to a 75° initial rotation offset with no other data perturbations.
	Objective function score comparison with pose error	5.3.5	Objective function score may predict pose error under fixed, known data conditions.
Algorithm Analysis	Objective function comparison	5.4.1	Gaussian-Weighted Nearest Point and Direct Surface Distance metric have the best performance.
	Distance weighting hierarchical refinement	5.4.2.1	Improved region of convergence and accuracy results are achieved, as compared with non-hierarchical refinement.
	Data smoothing hierarchical refinement	5.4.2.2	Improved region of convergence results are achieved, as compared with non-hierarchical registration. Overall, not as beneficial as distance weighting.
	Pose complexity refinement	5.4.3	Affine deformations are recoverable with a rigid/affine registration hierarchy.
	Iterative random perturbation	5.4.4	A small number of pose perturbation and re-optimization iterations improves accuracy for relatively large pose offsets.
Application Experiments	Laser/MRI registration	5.5.1	Sufficient laser data coverage (> 25%) is necessary for stable registration.
	Intra-cranial cavity registration	5.5.2	Accurate and stable registration is achievable for up to 45° rotation offset.
	MRI/angiogram registration	5.5.3	Accurate and stable registration is achievable for up to 50° rotation offset.

5.2 Experimentation Procedure

In order to perform the experiments presented in this chapter we set up an automated test environment for controlling the perturbations we wish to make. Each experiment is defined by a specification file which identifies the data to register, the algorithmic parameter settings, and the range of the perturbations to be performed. When executing the experiment, multiple random perturbations are made to the data across the given ranges with each data instantiation registered and saved to a results file. The starting condition of each experiment is 0 pose error between the reference and test datasets. For the simulation experiments we know the reference and test datasets are correctly aligned with each other since the test data starts out as a duplicate of the reference data. For the application experiments we manually align the reference and test data to start the test runs. The possible types of data perturbation are described in Section 5.2.1 and the results we collect are described in Section 5.2.2.

5.2.1 Experiment Specification Definition

For each experiment we set up a given number of test runs. Each run consists of a given number of simulations in which we apply a set of random perturbations. The perturbations apply to one or more data parameters, but always at the same magnitude within each run. For example, for testing pose region of convergence we want to vary the rotation offset applied to the test data from $0°$ to $116°$ at increments of $4°$, resulting in a total of 30 runs. Each run may then consist of 30 simulations. In each simulation we apply a rotation offset of the magnitude prescribed for that run, but around a randomly chosen 3D axis. All the possible perturbations are defined below in the data selections listing.

The experiment specification consists of key algorithm parameter selections and data perturbation parameters. The algorithm selections are:

- **Objective function variant**.
- Sequence of hierarchical refinements to be made during the registration:
  - **Distance weighting**—Range (start, decrement, stop) of Gaussian σ used to weight the point-to-point or point-to-surface distances.
  - **Surface sampling**—Range (start, increment, stop) of size of reference surface point set used with point-to-point registration metrics only.
  - **Surface smoothing**—Range of reference surface smoothing values used with point-to-surface registration metrics only. The parameter we use is the maximum allowable vertex-to-nearest-surface-point error distance for the vertices of the triangular facets

generated by the Nuages surface generation package. Larger allowable surface estimation errors result in coarser surfaces with fewer facets. The range specification for this parameter is thus initial acceptable error, decrement, and final acceptable error.

For point-to-point registration metrics, the finest surface smoothing value is the one used for extracting the surface, after which surface sampling is applied.

The data selections are:

- **Number of test runs**.
- **Number of simulations per run**.
- **Reference surface data.**
- **Test surface data** (possibly the same as the reference).
- **Number of test surface points**.
- **Smoothing of test surface**—Initial value and increment for each successive run, specifying the maximum surface estimation error for the test surface. For each test run we generate a test surface using the given smoothing value—larger surface estimation errors yield smoother surfaces. The test surface is composed of triangular facets whose vertices we use as the set of test points to register to the reference surface. The reference surface may also be smoothed as part of the hierarchical process so that any point during the registration the test data may be at a coarser of finer description level than the reference data.
- **Rotation magnitude offset of test surface**—Initial value and increment for each successive run, specifying a 3D rotation by which to rotate the test data from its initial registered state. For each simulation at a given rotation magnitude we randomly pick a 3D axis around which to apply the rotation. Since the rotation magnitudes are specified to be non-negative we generate 3D rotation axes around the whole sphere. A candidate axis vector is generated by choosing three coordinates in the range [-1,1] using a uniform distribution. The coordinates are used to define our rotation axis vector if the magnitude of the resultant vector is less than or equal to 1.0. Otherwise we pick another set of random coordinates until we find a suitable one.
- **Translation magnitude offset of test surface**—Initial value and increment for each successive run, specifying a 3D translation by which to translate the test data from its initial registered state. Translation is applied after rotation. For each simulation at a given translation magnitude we randomly pick a 3D direction along which to apply the translation. Since the translation magnitudes are specified to be non-negative we generate 3D translation directions around the whole sphere. The direction vector is generated in the same manner as the rotation axis vector for the rotation offsets above.
- **Outliers added to test surface**—Initial value and increment for each successive run, specifying the fraction of outlier points added to the test dataset as a function of the number of original points in the test dataset. Outliers are added after rotation and translation offsets are applied. For each simulation at a given outlier rate we add outliers at random 3D locations in the bounding box of the original test dataset. A uniform distribution inside the bounding box volume is used.
- **Removal of test surface patch**—Initial value and increment for each successive run,

specifying the fraction, R in the range [0,1], of the total surface to remove from the test dataset. The patch removal simulates incomplete coverage of the reference surface by the test surface. Surface removal is applied after outlier addition. To easily approximate such surface removal we model the test surface as a sphere. For each simulation at a given removal rate we randomly pick a 3D direction in the sphere, $\overline{U}_r$, and remove all test points, $\overline{P}_i^1$, which satisfy $\langle P_i^1, \overline{U}_r \rangle \geq \cos(\pi R)\left|P_i^1\right|$. $\overline{U}_r$ is generated in the same manner as the rotation axis vector for the rotation offsets above.

- **Location perturbation of test surface points**—Initial value and increment for each successive run, specifying the standard deviation, σ, of a 0-mean Gaussian distribution used to select a spatial offset by which to perturb the test surface points. Location perturbation is applied after surface patch removal. For each simulation at a given location perturbation σ and for each point on the test surface we generate a perturbation magnitude from the Gaussian distribution. We also pick a 3D direction for each point along which to apply the perturbation. The 3D direction vector is generated in the same manner as the rotation axis vector for the rotation offsets above.

5.2.2 Results Definition

The key registration result in which we are interested is the remaining pose error after registration relative to the initial 0 pose offset prior to our perturbation. We want to measure the degree to which the registration algorithm overcomes the data perturbations which we apply. Pose error is not measured in pose space due to the difficulty of quantifying displacements in pose space in terms of physical displacement errors. We therefore measure pose error as the distance between a point location in the 0 pose state and after application of the registration transform, doing so for selected points in the volume of interest. As a worse case, we take as our test points the corners of the bounding box of the pre-perturbed test surface, as shown in Figure 34. In addition, by measuring pose error at points off the surface being registered we better evaluate recovery of non-rigid transforms. Given (1) the eight corner points, $\overline{C}_i^1$, of the test surface bounding box prior to registration, (2) the perturbation matrix, Ψ, comprising the rotation and translation offsets applied during the test run simulation, and (3) Γ, the resultant registration transform, we then compute four values:

- **Initial average bounding box corner error**, prior to registration:

$$\frac{\sum_i^8 \left|\overline{C}_i^1 - \Psi\overline{C}_i^1\right|}{8}$$

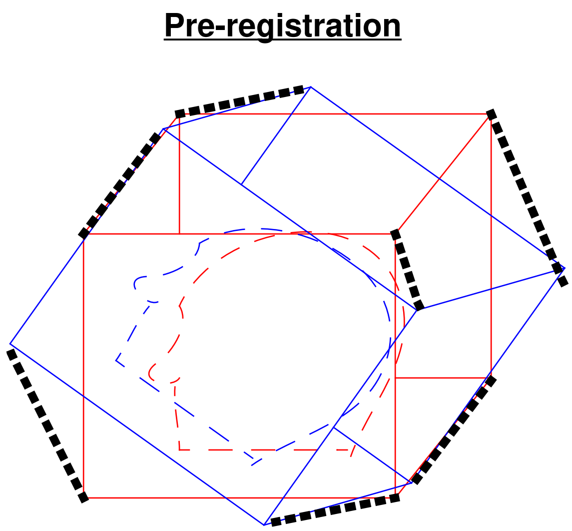


Figure 34 Pose accuracy measurements in terms of distances between corresponding corners of the bounding box of the test data. On left is offset data prior to registration and on right is the measurement of residual error after registration.

- **Initial maximum bounding box corner error**, prior to registration:

$$\max_i \left|\overline{C}_i^1 - \Psi\overline{C}_i^1\right|$$

- **Final average bounding box corner error**:

$$\frac{\sum_i^8 \left|\overline{C}_i^1 - \Gamma\Psi\overline{C}_i^1\right|}{8}$$

- **Final maximum bounding box corner error**:

$$\max_i \left|\overline{C}_i^1 - \Gamma\Psi\overline{C}_i^1\right|$$

These measures are computed for each test run simulation. In addition we record the final objective function value obtained for each simulation. Means and standard deviations are computed for each measure across all the simulations in each test run. We also tabulate the 95th percentile of the final maximum bounding box corner error to characterize the nature of region of convergence degradation. We use the final average bounding box corner error as our primary measure of pose error. Its mean provides a measure of accuracy and its standard deviation provides an indication of stability. For most applications we require registration accuracy on the order of 1-2mm, which is approximately the resolution of most of our test imagery.

5.2.3 Test Environment & Execution Times

All the experiments described in this chapter were run on Pentium Pro 200Mhz PC's running the Linux operating system. Although some of the PC's had multiple CPU's each experiment ran on a single processor at a time. The registration software was written in C without any special effort to optimize for efficiency. In particular we used Powell's optimization method since it readily applied to all objective function tests, but is likely not the most efficient optimization method.

The execution times were a function of many variables, including number of points in the test surface, location of test surface points relative to reference surface, and degree of pose offset. The overall characterization of the two surfaces relative to each other affected the convergence behavior of the optimization process which was the over-riding factor in determining execution time. As the surfaces become less similar, not just in pose but in general surface shape, time to convergence increases. There were also other factors which increased the difficulty of measuring execution time, such as load of the computer.

Experiments with highly overlapping reference and test datasets derived from the same head skin surface with 500 test surface points, less than $15°$ pose offset, spatial perturbation with a Gaussian distribution of $\sigma < 3$, and less than 10% test surface patch removal resulted in the following average execution times for non-hierarchical registration:

- Saturated RMS Nearest Point objective function: 26.3 seconds
- Gaussian-Weighted Nearest Point objective function: 31.3 seconds
- Averaged Nearest Neighborhood objective function: 69.7 seconds
- Direct Surface Distance objective function: 74.1 seconds.

When running the hierarchical registration variants, registration times increased, although not necessarily linearly with the number of hierarchical match loops since convergence within each hierarchical loop is generally faster once we have already found a coarse solution. Hierarchical registration execution times with a three-level distance weighting refinement ($\sigma = \{8, 5, 2\}$) under the same data conditions were:

- Saturated RMS Nearest Point objective function: 35.4 seconds
- Gaussian-Weighted Nearest Point objective function: 64.0 seconds
- Averaged Nearest Neighborhood objective function: 181.6 seconds
- Direct Surface Distance objective function: 202.9 seconds.

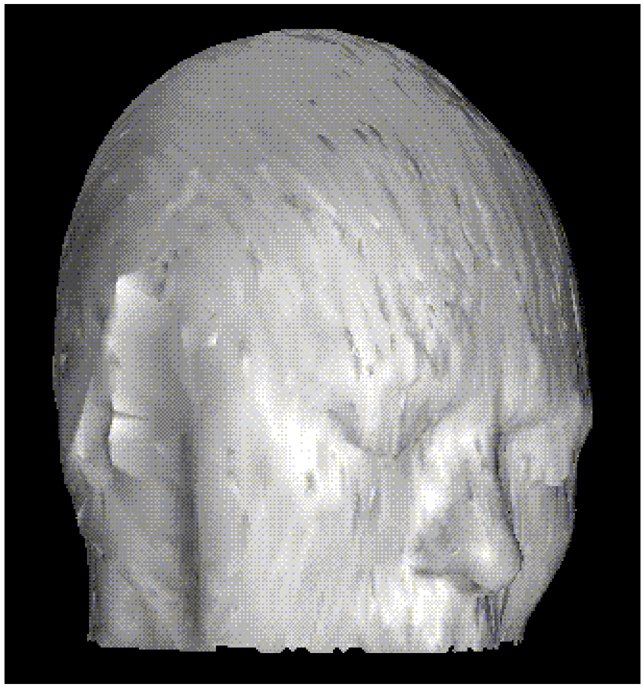


Figure 35 Skin surface used for data sensitivity simulation experiments. Whole head was used as the reference surface and a perturbed instance of it used as the test surface.

5.3 Data Sensitivity Simulation Experiments

These experiments analyze the behavior of the registration techniques under controlled data variations. We analyze the behavior of four objective functions in the presence of incomplete surface overlaps, measurement errors, differing data resolutions, and initial pose offsets, but with only one type of perturbation applied at a time. The test data is constructed by taking the head skin surface segmented from an MRI scan as the reference surface and a perturbed instance of the same surface as the test surface. The reference surface is shown in Figure 35. It was extracted from a sagittal image with voxel resolution of 0.9375mm x 0.9375mm x 1.5mm using the Nuages surface extraction system. The volume had 124 slices, each one with 256 x 256 pixels. The bounding box of the head skin surface was 189.4mm x 191.3mm x 154.4mm. By constructing the test data from the reference data we know the correct registration and we can directly control the perturbations against which we want to measure our sensitivity. The specific perturbations are described with each experiment. All the tests in this section involve no hierarchical registration refinement, but do consist of at least two final pose perturbations and re-optimizations. For instances where convergence is not reliably reached, up to 30 such iterations may be performed as long as improvement in the objective functions is seen. In Section 5.4 we analyze registration performance in the presence of multiple perturbation types applied simultaneously as well as the benefits of hierarchical reasoning. All the simulation experiments in this section apply rigid registration. Comparisons of rigid and affine registration are given in Section 5.4.3.

The benefit of these simulation experiments is in setting upper bounds on performance. In practice, when many factors are at play, we expect to see performance degrade from what we measure here with only single factors varied. But by setting upper bounds we at least have some measures to dictate if we may, but not necessarily will, satisfy particular application requirements.

5.3.1 Incomplete Surface Overlap Sensitivity

In these experiments we evaluate the ability of the registration algorithm to withstand either outliers added to the test surface or surface patches removed from the test surface. Outliers are added as a function of the number of surface points in the test set. Surface patch removal is performed by removing all points in a randomly selected neighborhood. In both cases, a random 15° rotation is applied to offset the test dataset from the reference dataset. The procedure for applying these perturbations is described in Section 5.2.1. For each experiment we examine the initial average bounding box corner error, the final average bounding box corner error, and the 95th percentile of the final maximum bounding box corner error, as described in Section 5.2.2.

5.3.1.1 Outlier Addition Sensitivity

The parameters for the outlier addition experiment are given in Table 2. The results are shown in Figure 36 and Figure 37. These results show that the Saturated RMS Nearest Point and Gaussian-Weighted Nearest Point metrics have little difficulty in converging to the correct pose from the 15° rotation offset even with an additional 380% randomly distributed outliers. Performance degradation is very slow for these metrics. The Averaged Nearest Neighborhood metric also exhibits very slow degradation, but it is not as accurate as the other point-to-point metrics. The Saturated RMS Nearest Point and Gaussian-Weighted Nearest Point metrics had average bounding box corner errors of about 0.2mm while the Averaged Nearest Neighborhood metric had corresponding errors of about 2.0mm. The Direct Surface Distance metric exhibited a somewhat faster performance degradation, although still relatively slow. It exhibited accuracy of less than 1mm for outliers rate of up to 80%, but the pose error gradually increased with additional outliers.

The excellent performance of the Saturated RMS Nearest Point and Gaussian-Weighted Nearest Point metrics is explained in part by the nature of this experiment in which the test points are

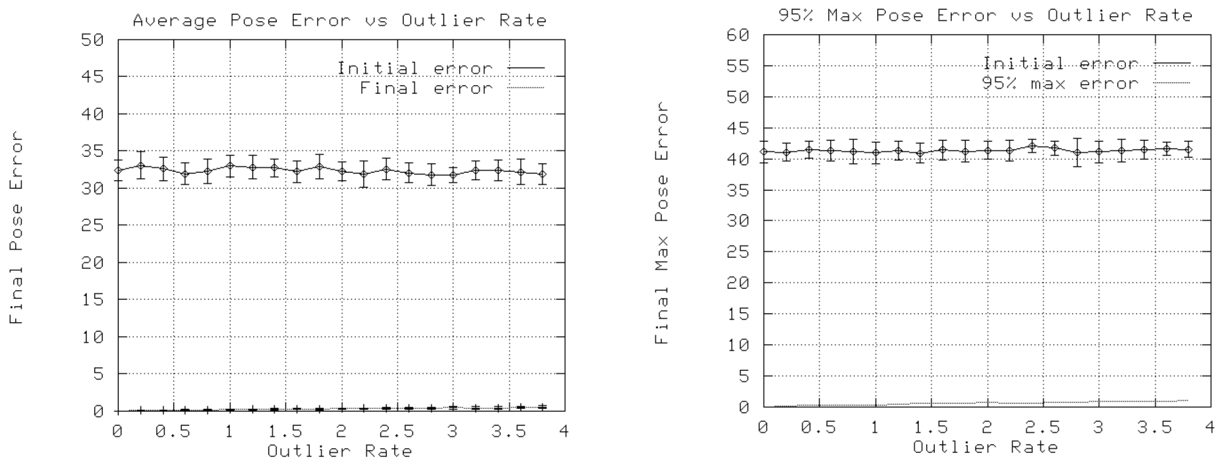
Table 2 Experiment specification for outlier addition experiment.

Experiment Specification Parameter	Value
Objective function variants	Saturated RMS Nearest Point, Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 6949 (fixed to use all points) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **15**° Increment per run: 0°
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: **20%**
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

exact replicas of the reference points. Since these metrics search for the single nearest reference point for each test point, they are able to lock on to the correct set of point correspondences. In general, such exact correspondences do not exist. We see in the experiments of Sections 5.4 and 5.5 that these metrics still perform well when the test and reference datasets are not exact replicas of each other. The slightly reduced accuracy of the Averaged Nearest Neighborhood metric is not unexpected since its smoothing behavior prevents it from localizing the minimum exactly. The Direct Surface Distance metric, which also does not leverage possible point correspondences, performs well in the presence of outliers.

Given that the results of the Saturated RMS Nearest Point and Gaussian-Weighted Nearest Point metrics are very similar, we only show the results of the Gaussian-Weighted Nearest Point metric for the remainder of the simulation experiments in this section. A close-up comparison of the two metrics for the outlier test is shown in Figure 38 where we see that the Gaussian-Weighted Nearest Point metric achieves slightly better results.

Saturated RMS Nearest Point Metric



Gaussian-Weighted Nearest Point Metric

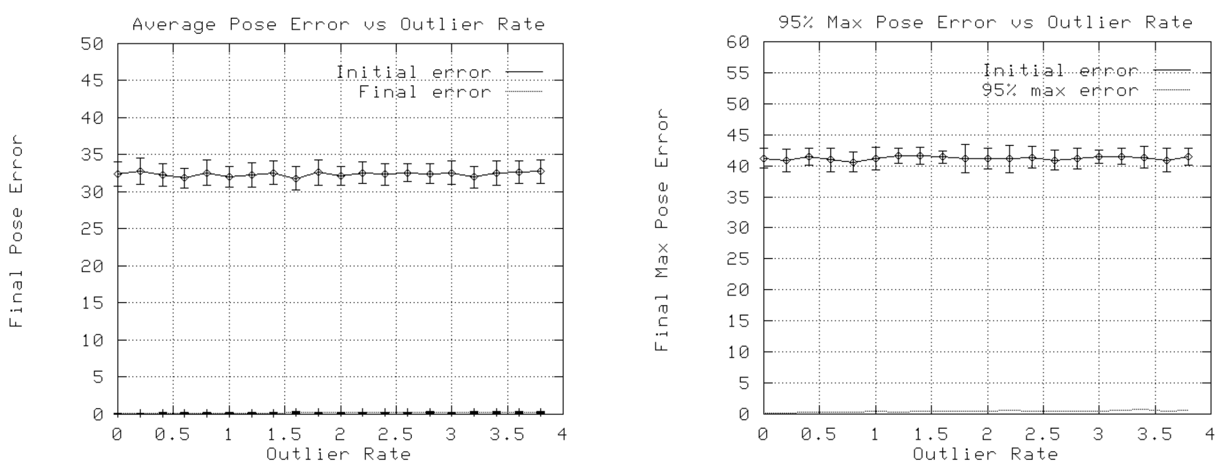
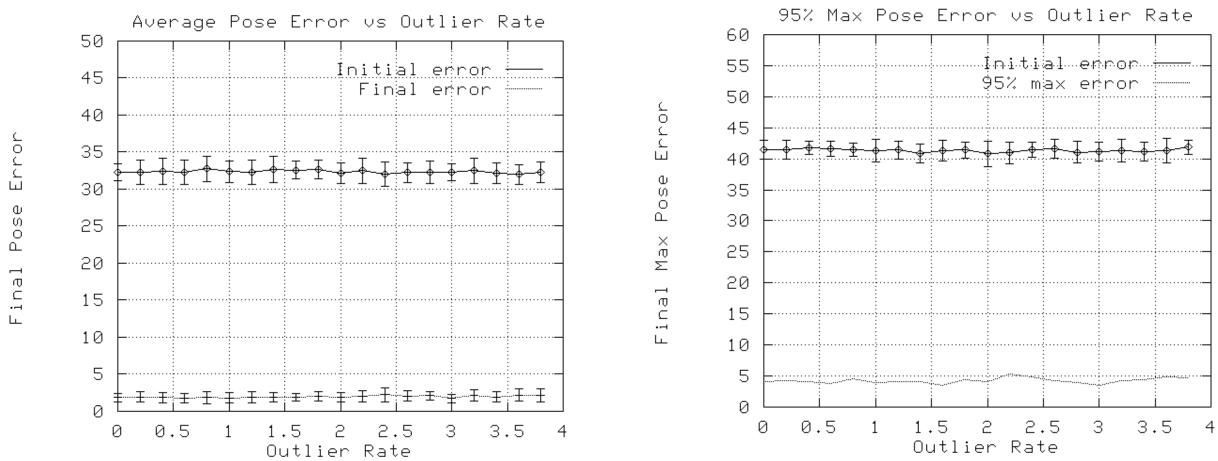


Figure 36 Results (1 of 2) of outlier addition experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

5.3.1.2 Surface Patch Removal Sensitivity

The parameters for the surface patch removal experiment are given in Table 3. In this case we run with more test points to maintain sufficient number of points as surface patches are removed. We also subsample the reference surface points (for the point-to-point metrics) to 72% of the total number of points to explore our sensitivity to reduced reference surface resolution. The results are shown in Figure 39. We first observe that the accuracy of the Gaussian-Weighted Nearest Point metric decreases with the reference surface sampling. With no outliers nor surface patch removal, the average bounding box corner error was 0.07mm (Figure 36 top-left at 0 outlier rate) when no

Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

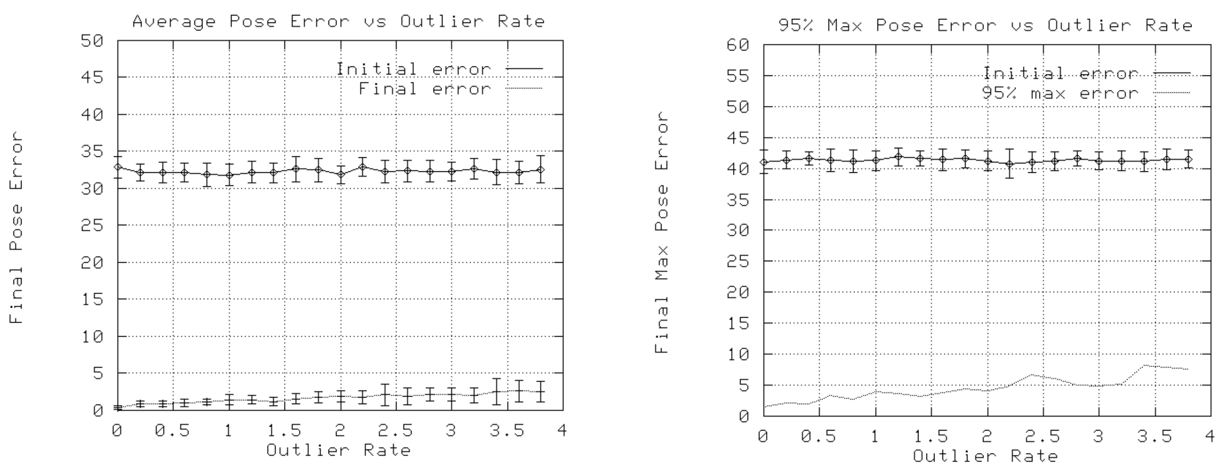


Figure 37 Results (2 of 2) of outlier addition experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

teed to be found in the reference points. But the same error measure increases to 0.8mm (Figure 39 top-left at 0 occlusion rate) when we sample only 72% of the reference points and increase the number of test points from 500 to 2000, introducing a non-zero probability that we do not have exact replicas for a test point. In our simulations, the test point sampling and reference point sampling are performed on the original surface rather than sampling the test set from the reference set. As a comparison the Direct Surface Distance metric has a corresponding error of 0.3mm for the same test surface sampling that resulted in the Gaussian-Weighted Nearest Point metric yielding a 0.8mm error.

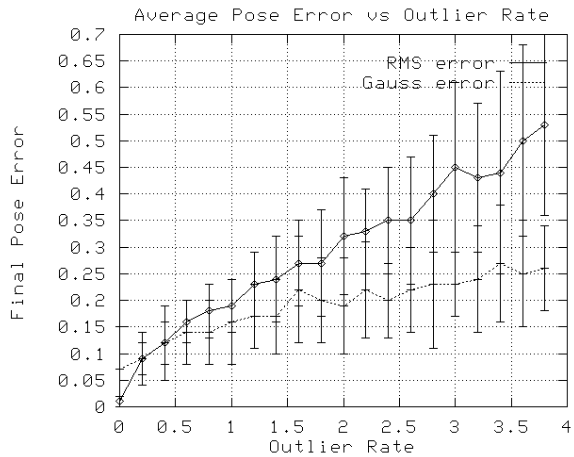
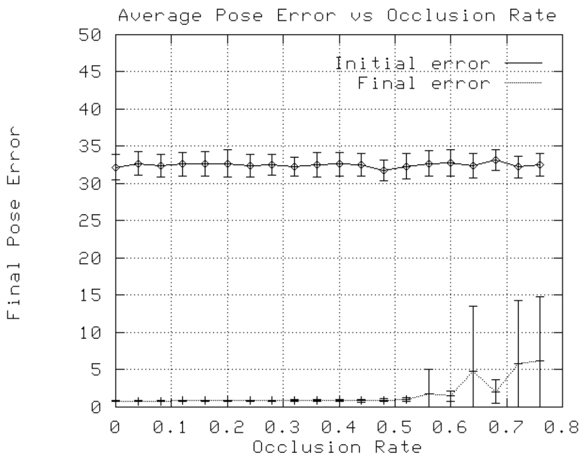


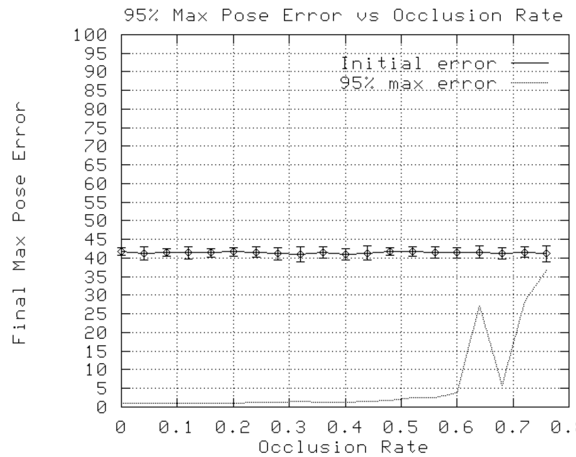
Figure 38 Comparison of average bounding box corner error for Saturated RMS Nearest Point metric (top curve) and Gaussian-Weighted Nearest Point metric (bottom curve).

Table 3 Experiment specification for surface patch removal experiment.

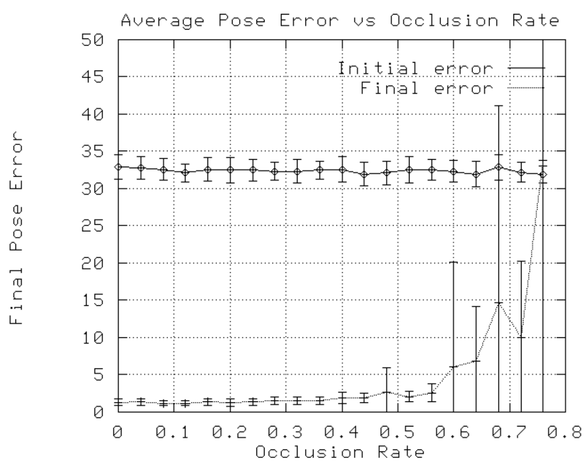
Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 5000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	2000
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **15°** Increment per run: 0°
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: **4%**
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

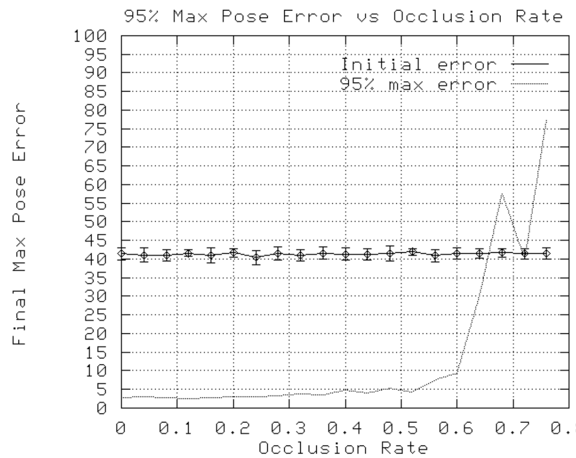
Gaussian-Weighted Nearest Point Metric



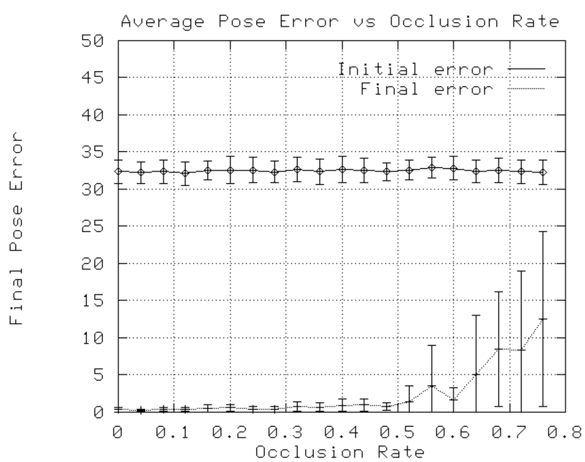


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



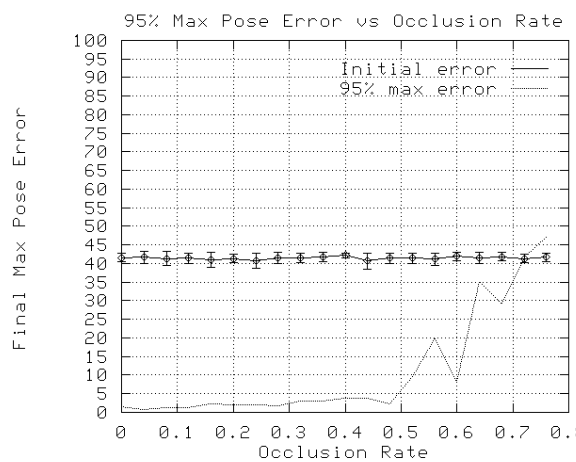


Figure 39 Results of surface patch removal experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

More importantly, we see highly stable performance for surface patch removal (occlusion) rates of up to 50%. Both the Gaussian-Weighted Nearest Point and Direct Surface Distance metrics have average bounding box corner errors of less than 1mm for occlusion rates of less than 50%. The Averaged Nearest Neighborhood metric has errors of 1-2mm in this occlusion range. Around 60% occlusion we notice a marked drop in the reliability of the registration results as the 95th percentile maximum error sharply increases. This result indicates that, at least for head surface data starting with a 15° rotation offset, convergence on the correct pose may be difficult to achieve if less than 40% of the head surface is available in the test dataset. This result is not difficult to imagine—since the crown of the head is somewhat spherical, registering small patches of that surface will not be reliable.

5.3.2 Measurement Error Sensitivity

In this experiment we evaluate the ability of the registration algorithm to withstand measurement noise in the form of surface point locations that are spatially perturbed from their physically correct location. Spatial perturbation is performed by applying a spatial offset to the test points in a random direction whose magnitude is given by a 0-mean Gaussian distribution. There is thus no bias in the test data. The procedure for applying these perturbations is described in Section 5.2.1. We apply a random 15° rotation to offset the test dataset from the reference dataset. We examine the initial average bounding box corner error, the final average bounding box corner error, and the 95th percentile of the final maximum bounding box corner error, as described in Section 5.2.2.

The parameters for the measurement error experiment are given in Table 4. The results are shown in Figure 40, in which we see a slow graceful degradation of accuracy through spatial perturbations of a Gaussian standard deviation of 19mm. The slowest performance degradation is with the Gaussian-Weighted Nearest Point metric. At a perturbation standard deviation of 3mm its registration error is only 1.0mm and it increases to an error of 1.7mm at a perturbation standard deviation of 19mm. The corresponding registration errors for the Direct Surface Distance is 1.2mm and 3.2mm respectively.

5.3.3 Data Resolution Sensitivity

In this experiment we evaluate the effect of varying data resolution on the registration algorithm. We vary the resolution of the test surface by varying the maximum surface estimation error in the Nuages surface generation system. Larger surface estimation errors lead to larger and fewer sur-

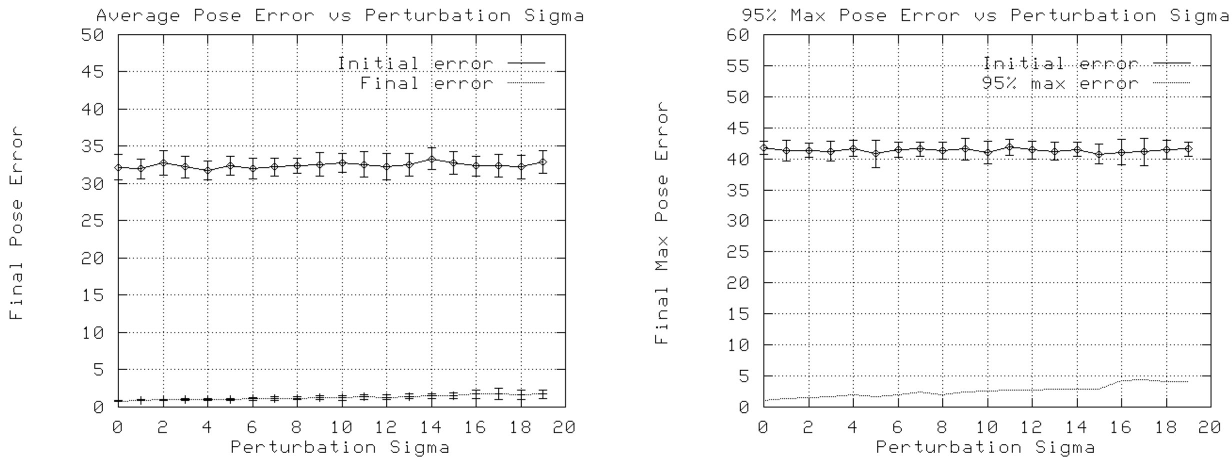
Table 4 Experiment specification for measurement noise experiment.

Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 5000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	2000
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **15°** Increment per run: 0°
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: **1mm**

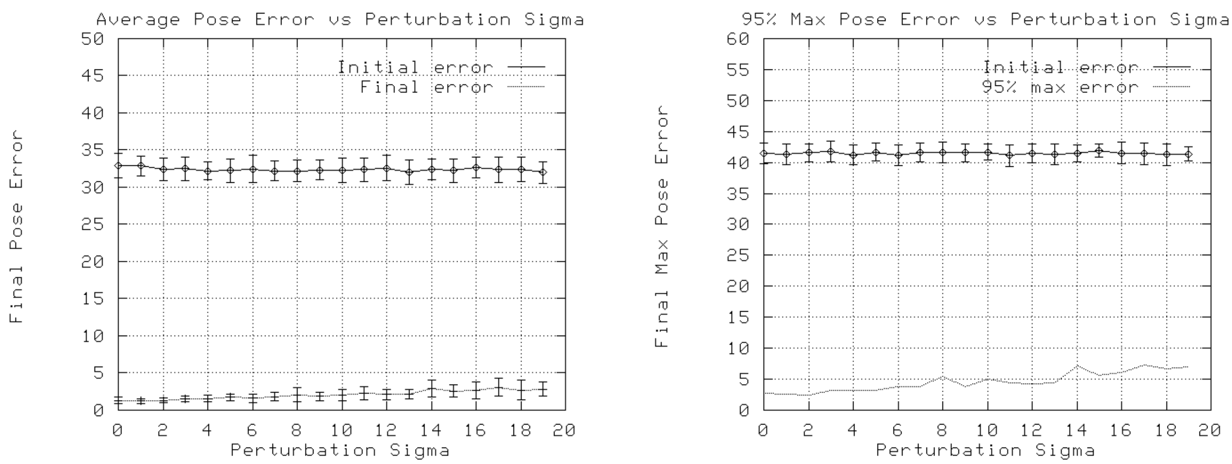
face triangles which have removed some of the higher frequency detail of the surface. The test surface points are taken as the vertices of the resultant surface triangles. The procedure for applying these perturbations is described in Section 5.2.1. We apply a random 15° rotation to offset the test dataset from the reference dataset. We examine the initial average bounding box corner error, the final average bounding box corner error, and the 95th percentile of the final maximum bounding box corner error, as described in Section 5.2.2.

The parameters for the measurement error experiment are given in Table 5. Although we requested 2000 test surface points, as we smooth the surface we generate fewer and fewer test points. At the surface approximation accuracy of 1.7 we have only 1871 points and the number continues to decline to 759 points at the coarsest surface approximation accuracy value of 8.1. The results are shown in Figure 41. We observe a significant increase in registration error from the finest resolution surface tested (0.5) to the next finest value (0.9). We suspect that we lose significant surface detail at that point since it corresponds to the resolution of the original MR image. For further coarser descriptions of the surface we observe only a gradual increase in registration error. Given the high stability of the results, it is apparent that the reduced resolution leads to a

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

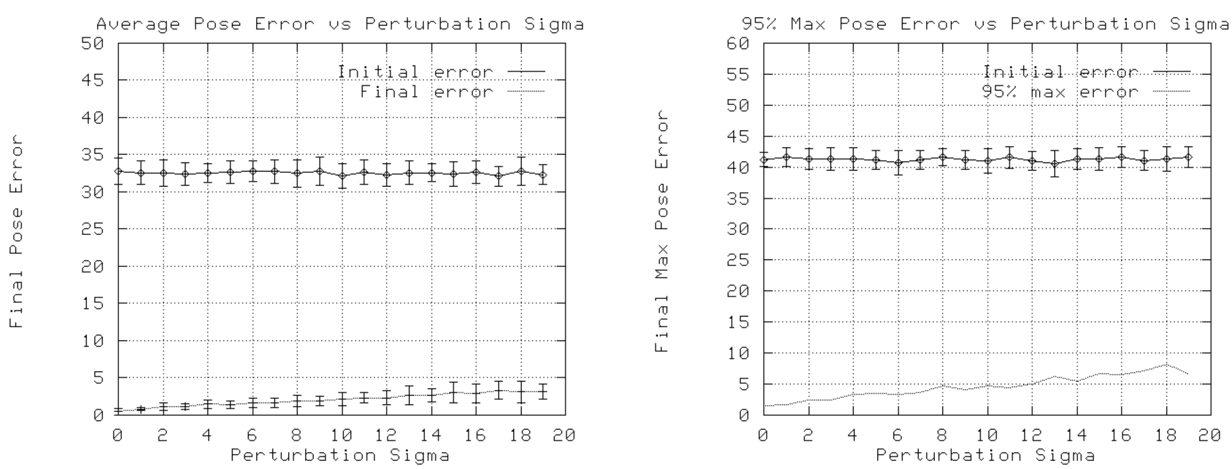


Figure 40 Results of measurement noise experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

Table 5 Experiment specification for data resolution experiment.

Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 5000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	2000
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: **0.4mm**
Rotation magnitude offset of test surface	Initial: **15°** Increment per run: 0°
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

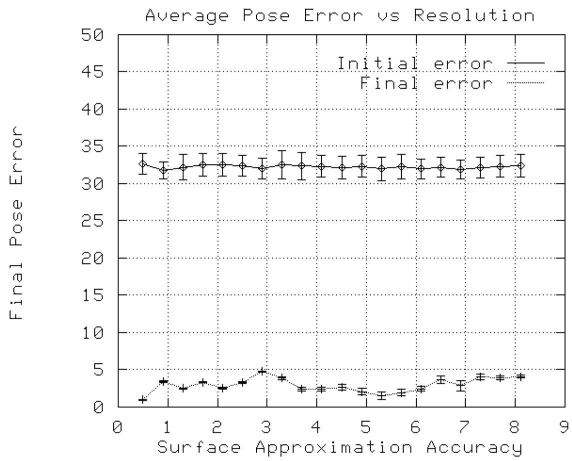
pose bias. The bias likely results from the surface facetization which increases misalignments proportionally to the estimation error at a given point. As is the case with the other experiments, the Averaged Nearest Neighborhood metric exhibits slightly reduced accuracy than the other two metrics. The actual surface representations used to generate some of the test datasets are shown in Figure 42. Note the reduced level of detail in the coarser surface descriptions obtained with the larger surface estimation errors.

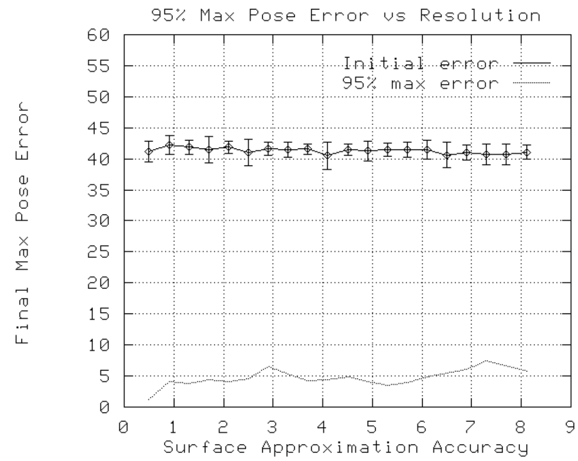
5.3.4 Pose Region of Convergence

In this experiment we evaluate the effect of initial pose offset on the registration algorithm. We continually apply larger 3D rotation offsets to the test data and measure how well the registration converges. The procedure for applying rotation offsets is described in Section 5.2.1. We examine the initial average bounding box corner error, the final average bounding box corner error, and the 95th percentile of the final maximum bounding box corner error, as described in Section 5.2.2.

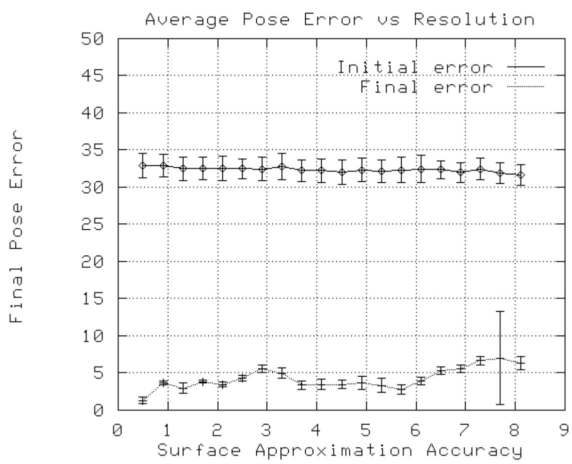
The parameters for the pose region of convergence experiment are given in Table 6. The results are shown in Figure 43. We observe accurate pose convergence for rotation offsets as high

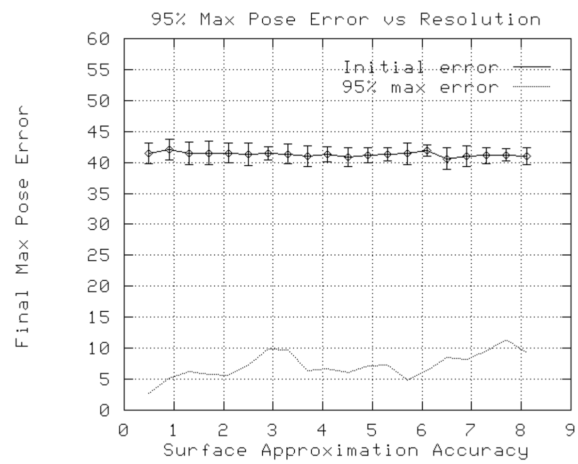
Gaussian-Weighted Nearest Point Metric



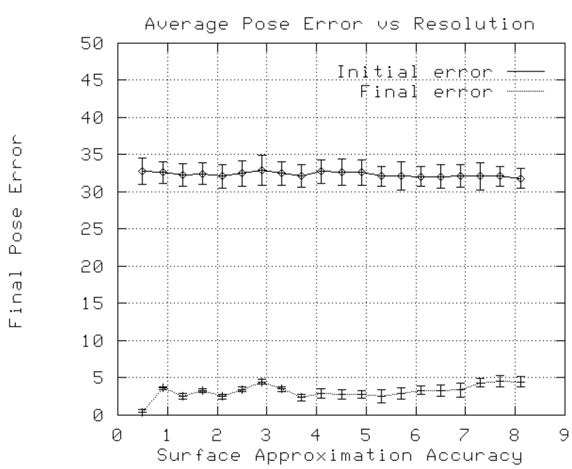


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



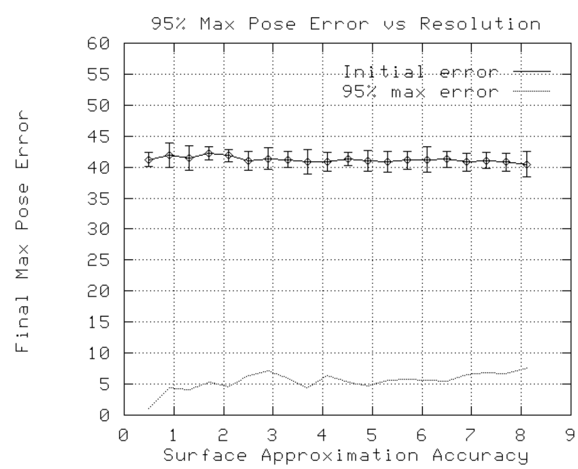


Figure 41 Results of data resolution experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

Smooth Shading | **Flat Shading**

Surface Estimation Accuracy: 0.5mm

Surface Estimation Accuracy: 0.9mm

Surface Estimation Accuracy: 2.5mm

Surface Estimation Accuracy: 4.5mm

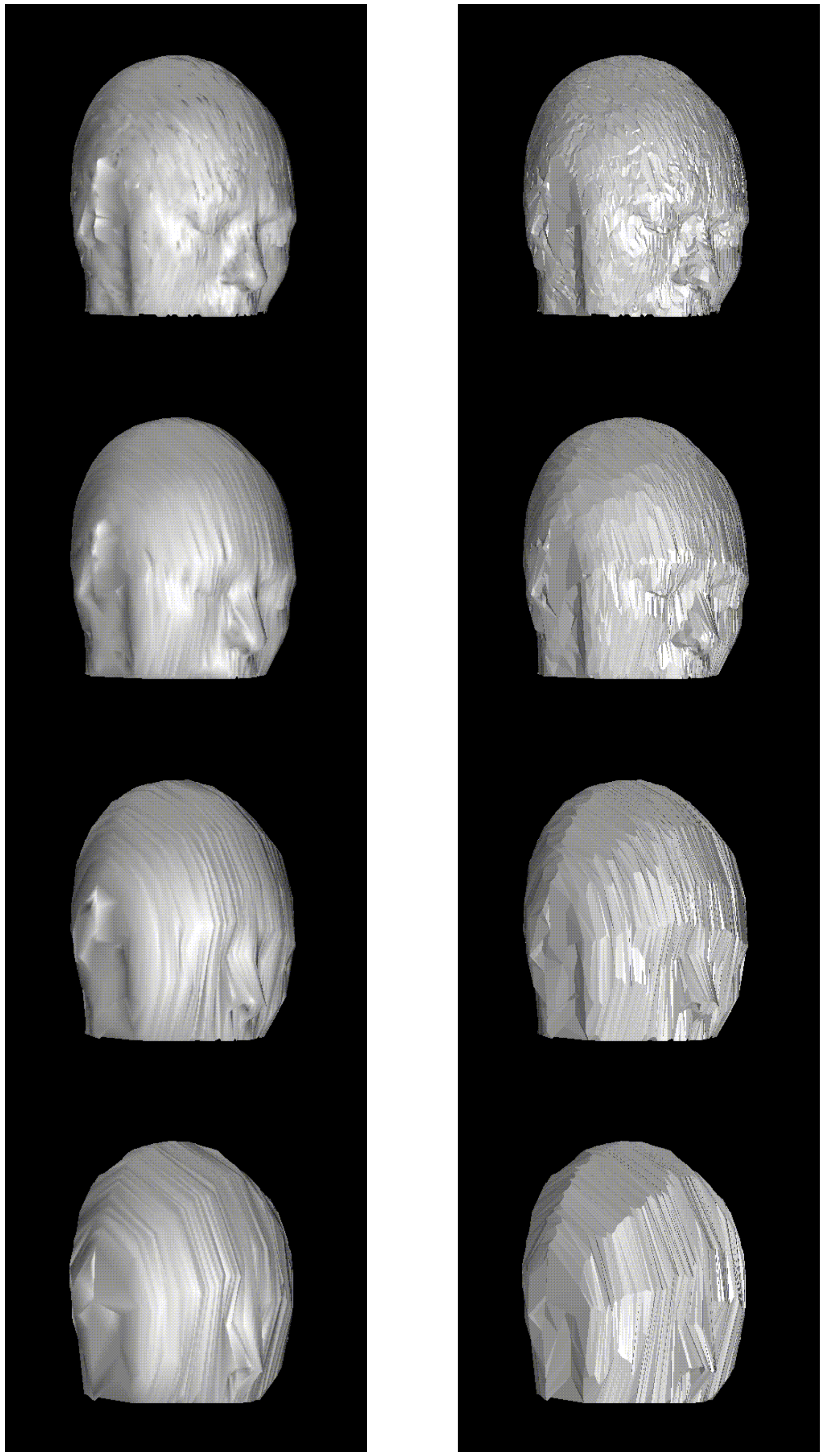


Figure 42 Surface representations used to extract varying resolution test datasets. The flat shading shows the actual surface triangles whose vertices formed the test point set.

Table 6 Experiment specification for pose region of convergence experiment.

Experiment Specification Parameter	**Value**
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 6949 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	2000
Number of test runs	30
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **0**° Increment per run: **4**°
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

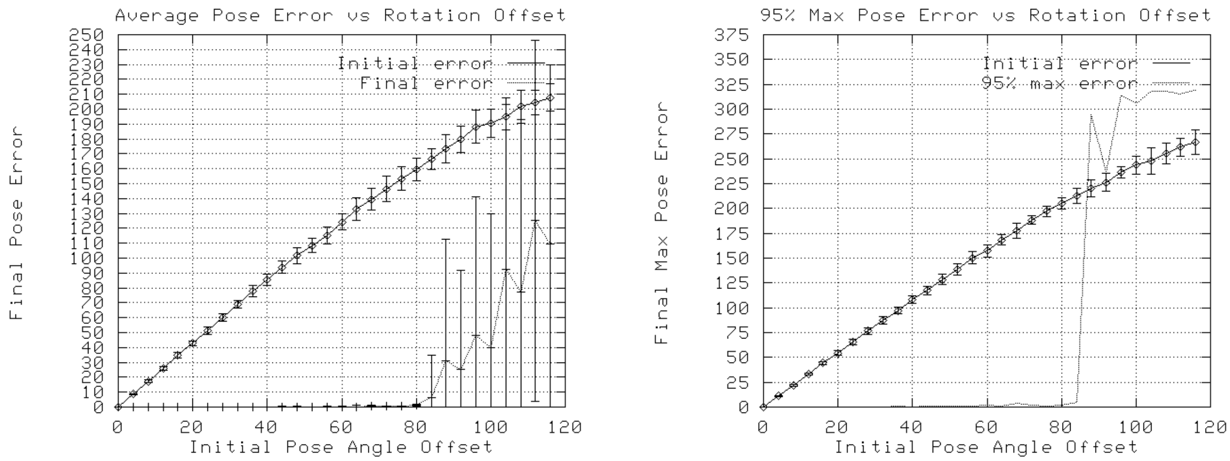
as 75° for the Gaussian-Weighted Nearest Point and Direct Surface Distance metrics and about 55° for the Averaged Nearest Neighborhood metric. While inside the region of convergence we see stable accurate behavior, we also see catastrophic failures outside the region of convergence, as shown in the 95th percentile errors of about 310mm. That value is the diagonal of the bounding box of the head surface. These failures are convergences to the 180° mirror image of the head surface, across the sagittal plane of symmetry. Such failures highlight the need for identification of the region of convergence so we may know to not cross it.

Although in this experiment, the test data is an unperturbed replica of the reference data, we still achieve large regions of convergence in tests in which the test data is perturbed, Section 5.4.1.1, or is altogether derived from a different image of the same subject, Section 5.5.3.

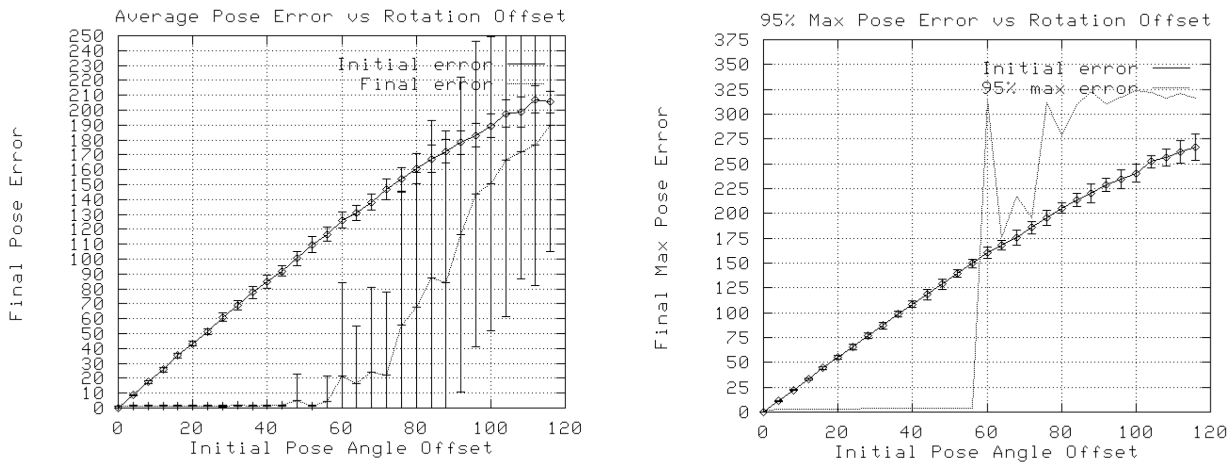
5.3.5 Score Vs. Accuracy

We earlier pointed out that none of the objective functions are true measures of registration accuracy—they are just measures of alignment. While we rely on the fact that accurate registrations have good surface alignments, there are many possible other surface alignments that may be

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

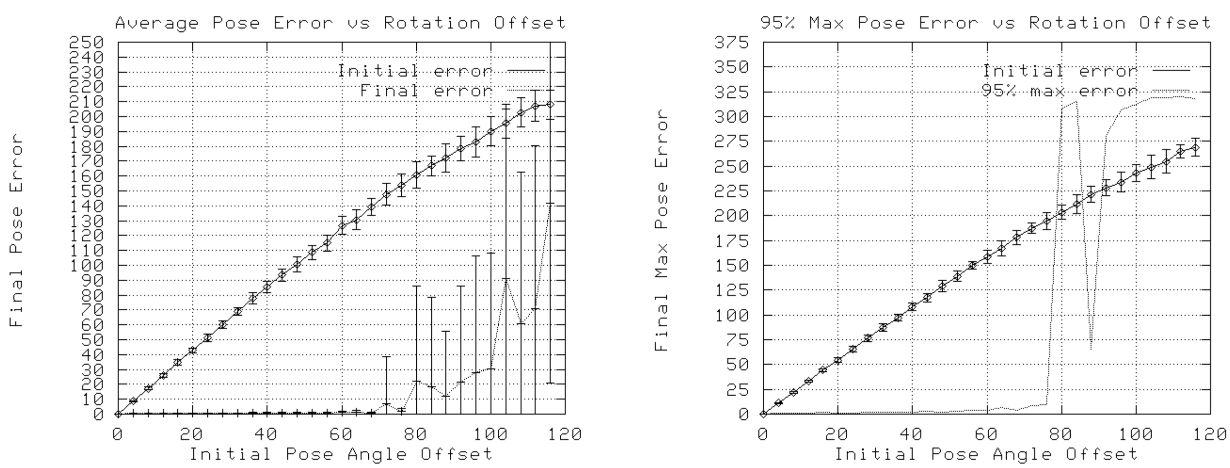


Figure 43 Results of pose region of convergence experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

almost as good. In Section 5.4.1.3 we take a closer look at the actual shape of the objective function surface over pose, but in this section we evaluate how well the reported objective function value reflects the actual registration accuracy. That is, when we do not find the correct pose can we know that by the reported value of the objective function. While it may be difficult to determine objective function thresholds in order to make such registration success decisions, we want to evaluate the degree to which the objective function qualitatively reflects registration accuracy. We examine the reported objective function values for the outlier addition, measurement noise, and pose region of convergence experiments.

The scores for the outlier addition experiment are shown in Figure 44. Here we see that although we consistently converged on the correct pose, the reported objective function score continually increased. This result is not unexpected since the added outliers increase the distance measure even with if their impact is saturated. The score curves appear logarithmic since the marginal impact of outliers decreases as their number increases.

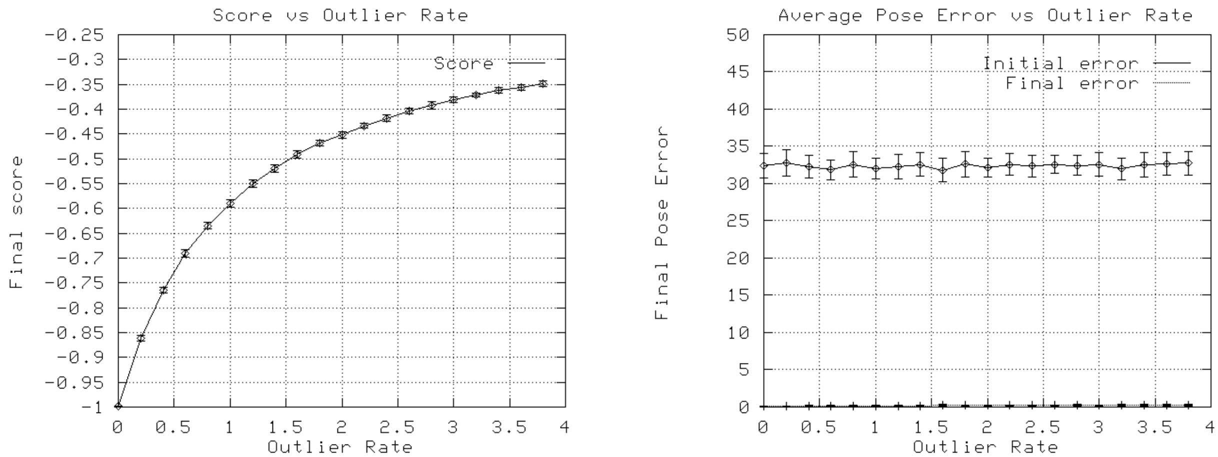
The scores for the measurement noise experiment are shown in Figure 45. In this case the reported objective function values also appear to grow disproportionately with the small increases of registration error. Here the scores grow approximately linearly since the average spatial perturbation value is a linear function of the Gaussian standard deviation used in defining the perturbation distribution. In the outlier addition and measurement noise experiments we see the difficulty of relating score to pose error.

Figure 46 shows the scores for the pose region of convergence experiment. In this experiment the reported objective function curve faithfully follows the registration error curve indicating that if we somehow can estimate such factors as number of outliers and measurement noise we may be able to infer accuracy from the objective function score. Although Bayesian techniques may help address this problem, in practice it is difficult to set simple objective function thresholds. Simon [54], though, has developed some techniques for accuracy estimation based on analysis of the solution stability.

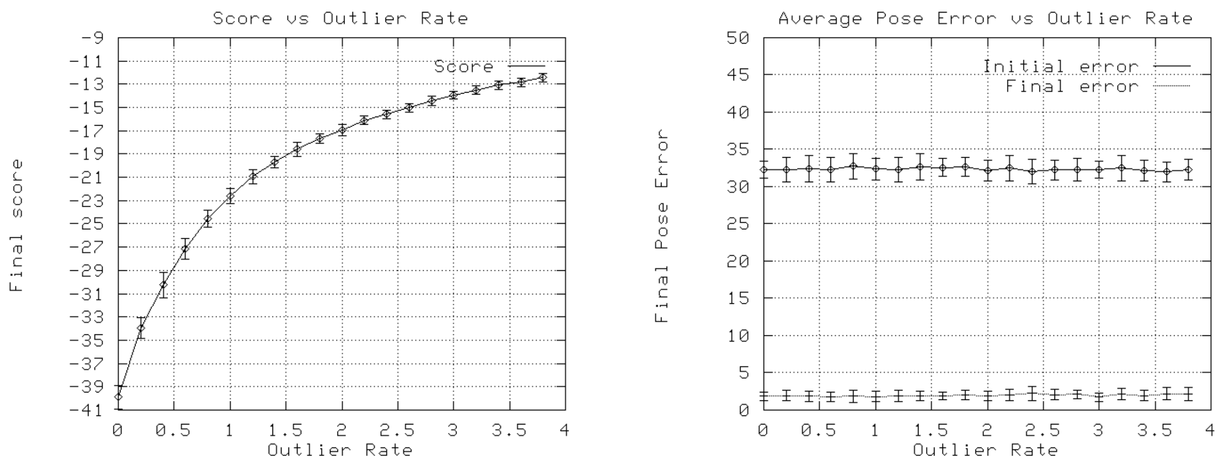
5.3.6 Simulation Experimentation Conclusions

The experiments in this section have set expected upper bounds on registration performance. Although these performance conditions are just preliminary, and testing on more data and conditions is necessary for better characterization, the results provide us with working estimates on the types of problems we may be able to handle. The performance bounds we identified are:

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

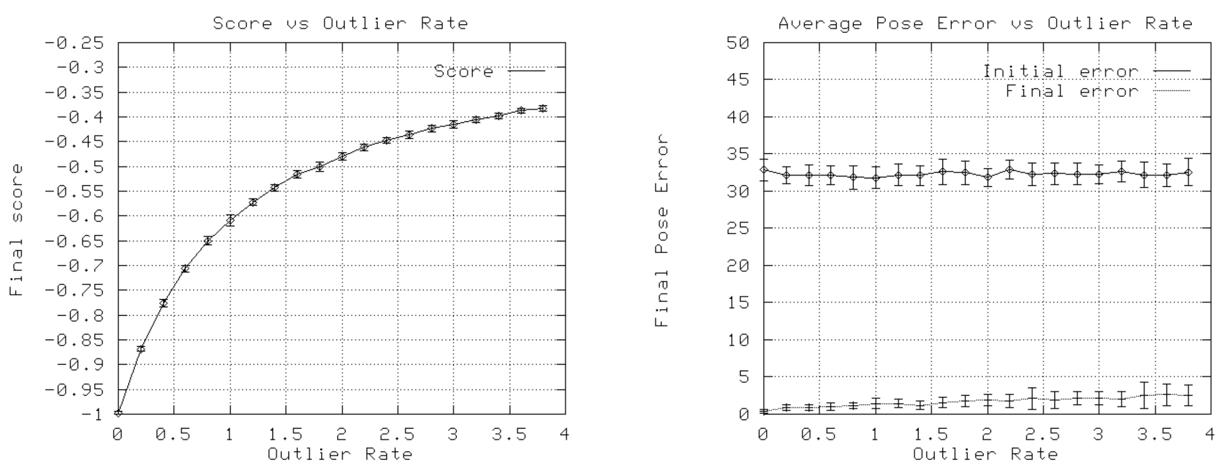
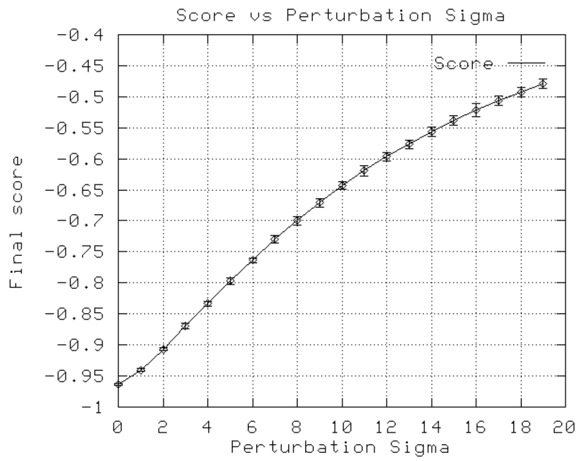
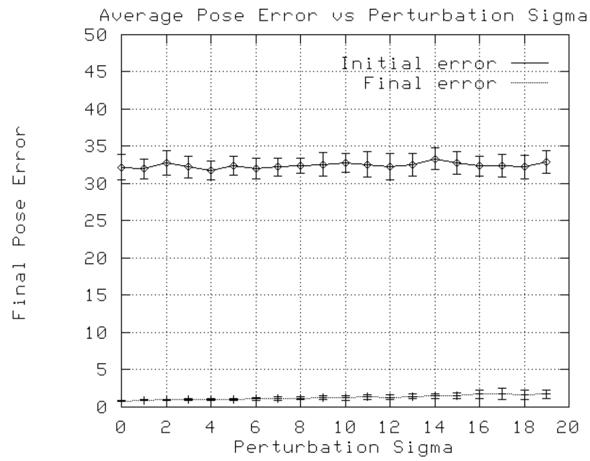


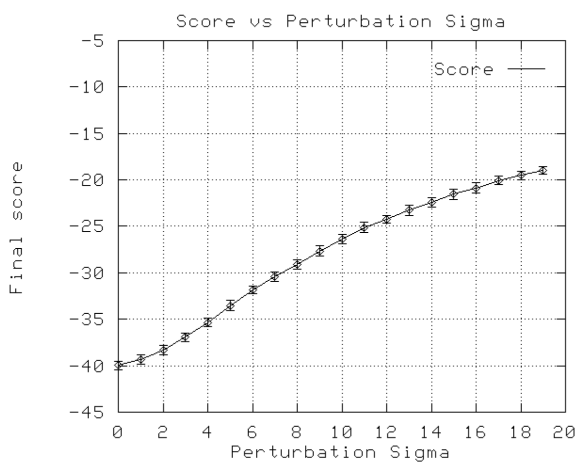
Figure 44 Comparison of reported objective function values (left) to accuracy results (right) for outlier addition experiment.

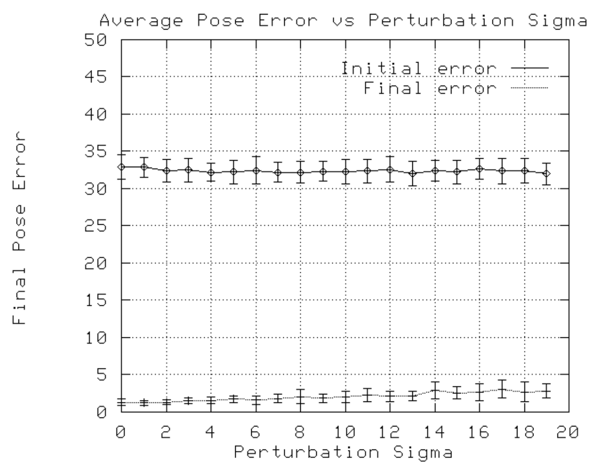
Gaussian-Weighted Nearest Point Metric



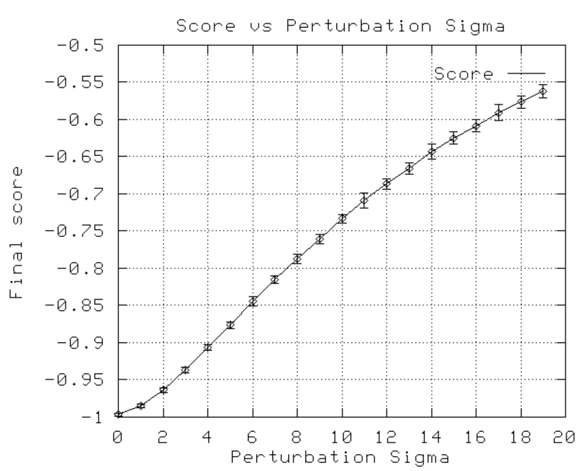


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



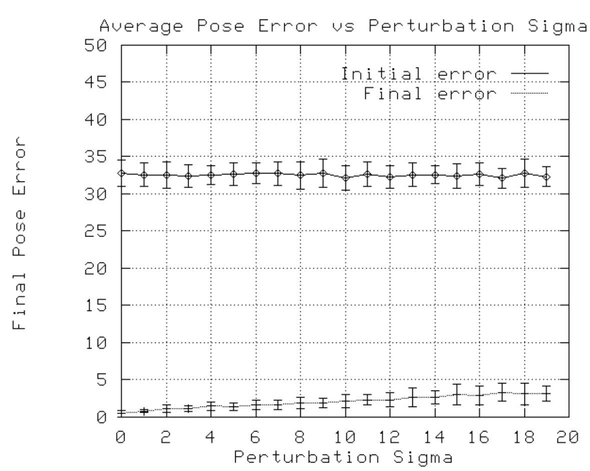
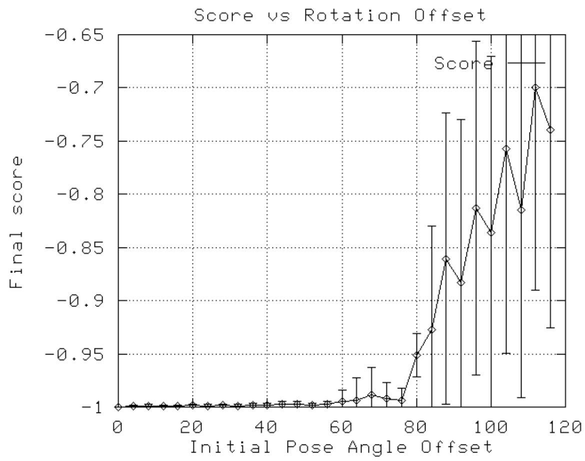
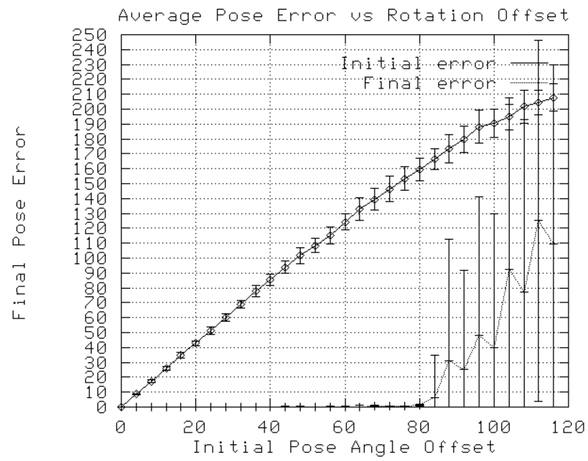


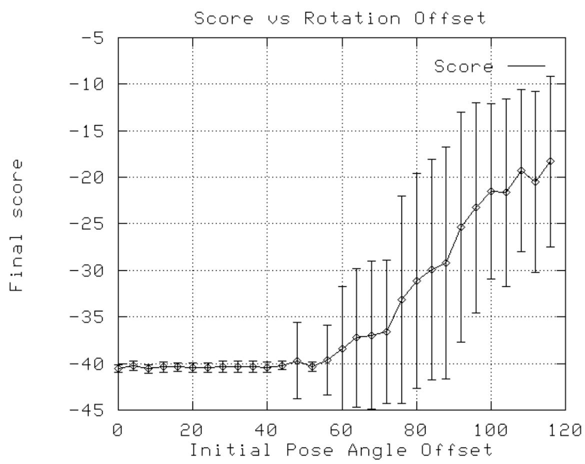
Figure 45 Comparison of reported objective function values (left) to accuracy results (right) for measurement noise experiment.

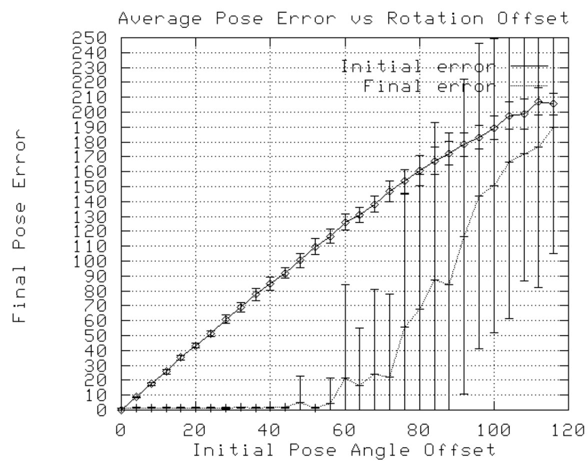
Gaussian-Weighted Nearest Point Metric



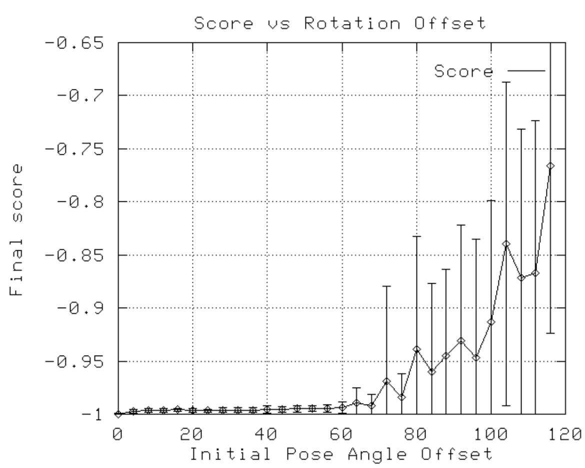


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



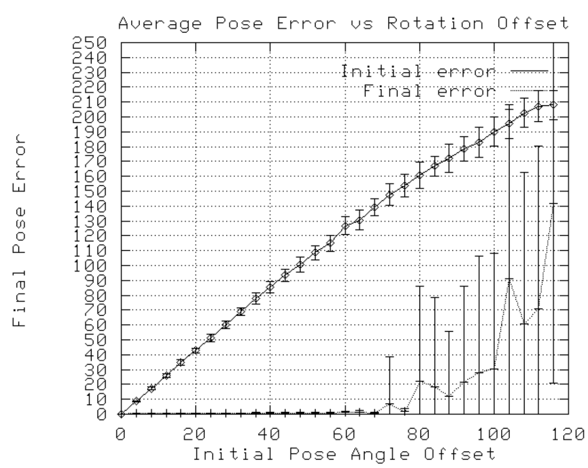


Figure 46 Comparison of reported objective function values (left) to accuracy results (right) for pose region of convergence experiment.

- Randomly placed outliers have only little impact on registration accuracy.
- At least 40% reference surface coverage is necessary for reliable performance from 15˚ rotation offset.
- Increased measurement noise yields gradually larger registration errors. With a 0-mean Gaussian distribution of measurement error, we achieve a registration accuracy within the whole registered volume of less than 1mm with Gaussian standard deviations of up to 3mm and 15˚ rotation offsets. But registration accuracy grows slowly with increased measurement noise. At a Gaussian standard deviation of 10mm, registration error is only about 2mm.
- A surface estimation accuracy of 0.5mm (for the 0.9375mm x 0.9375mm x 1.5mm imagery) is necessary to achieve sub-millimeter registration accuracy. Reduced surface data resolution leads to registration errors of near 5mm, but the errors remain at that value for even very coarse test surfaces.
- Rotation offsets of more than 75˚ are difficult to recover and may lead to 180˚ symmetrically wrong poses.

5.4 Algorithm Analysis

In this section we use the same data and experimentation procedure as in Section 5.3, but focus on registration performance as a function of algorithm parameters instead of data parameters. We compare the four objective function metrics we implemented, evaluate the benefits of performing the registration in a hierarchical manner, and analyze the need for the iterative random perturbations and re-optimizations.

5.4.1 Objective Function Comparison

In comparing the objective functions we evaluate the region of convergence, registration accuracy, and the shape of the objective function surface. We do so with somewhat more realistic test data perturbations than in Section 5.3 by simultaneously applying a rotation offset, translation offset, spatial perturbation, and surface patch removal. The procedure for applying these perturbations is described in Section 5.2.1. We examine the initial average bounding box corner error, the final average bounding box corner error, and the 95th percentile of the final maximum bounding box corner error, as described in Section 5.2.2. The parameters for these experiments are given in Table 7. Note that these experiments are non-hierarchical and apply a rigid transform only.

5.4.1.1 Region of Convergence

We first compare the objective functions in terms of their region of convergence. In this case we are not just measuring pose region of convergence since we are simultaneously applying other

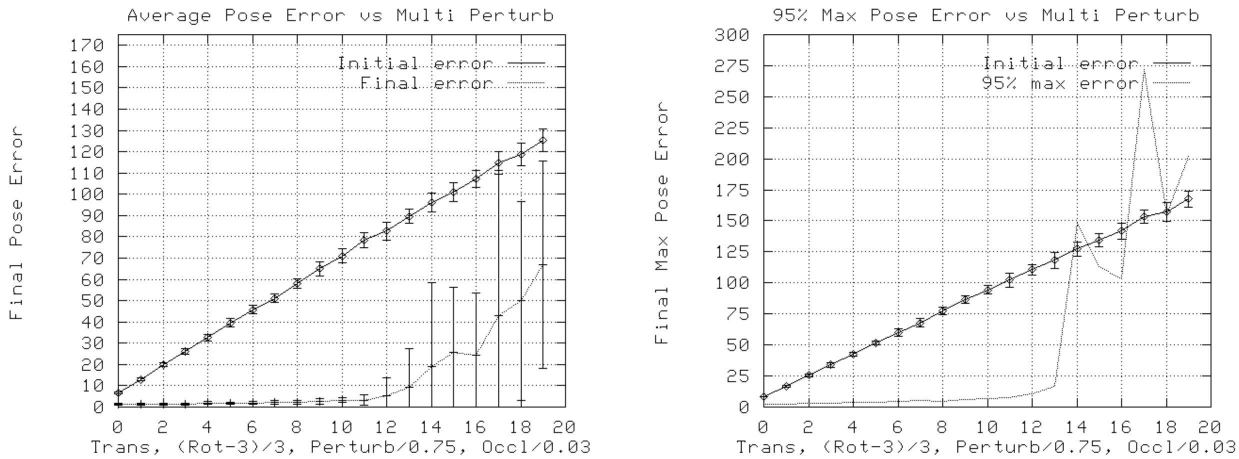
Table 7 Experiment specification for objective function comparison experiments.

Experiment Specification Parameter	Value
Objective function variants	Saturated RMS Nearest Point, Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 2000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**

perturbations, but we will express the region of convergence primarily in terms of rotation. The results of these comparisons are shown in Figure 47 and Figure 48. One unit along the X-axis corresponds to a rotation offset increment of 3°, a translation offset increment of 1mm, a surface patch removal of 3%, and a spatial perturbation increment of 0.75mm. With the initial values as given in Table 7, the first point only has a 3° rotation offset and no other perturbations. The final point has a rotation offset of 60°, a translation offset of 19mm, removal of 57% of the surface area, and spatial perturbation of 14.25mm.

The region of convergence is about 36° for the Saturated RMS Nearest Point, Gaussian-Weighted Nearest Point, and Direct Surface Distance metric. At this point we also have a translation offset of 11mm, occlusion of 33% and spatial perturbations of 8.25mm. This result compares to the 75° rotation region of convergence we had shown in Figure 43 where we only had a rotation offset without any of the other perturbations. The 36° still provides us with sufficient flexibility in generating initial poses for most medical applications. The Averaged Nearest Neighborhood metric has a somewhat smaller region of convergence of about 21°.

Saturated RMS Nearest Point Metric



Gaussian-Weighted Nearest Point Metric

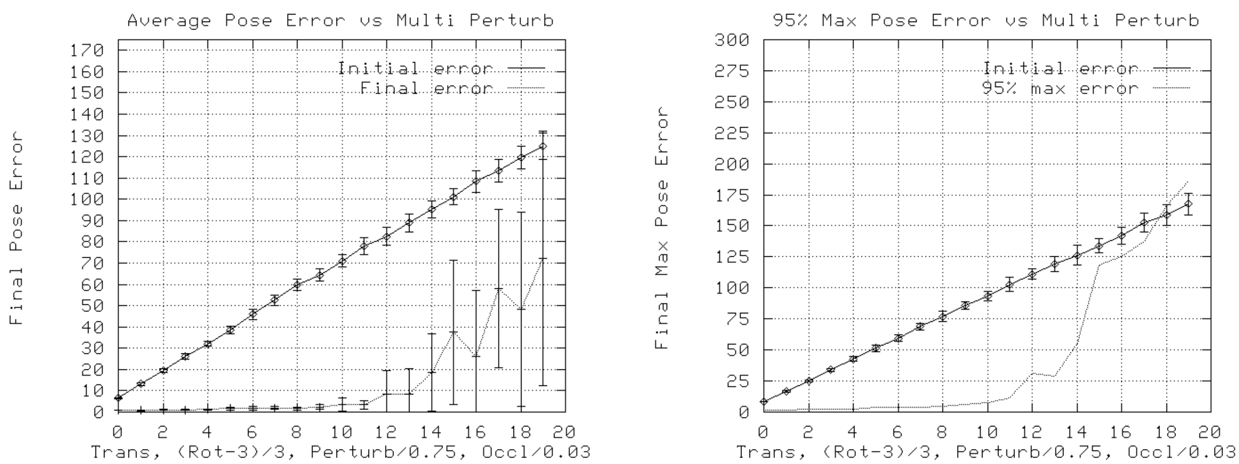


Figure 47 Results (1 of 2) of objective function comparison experiments. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

In analyzing the region of convergence we also evaluate the impact of the point sampling. In the results shown in Figure 47 and Figure 48 we sampled the test surface with 500 points and the reference surface with 2000 points. We compare those results to the case in which we sample 2000 test surface points and 5000 reference surface points as shown in Figure 49 for three objective functions. We increase the sampling rate for the reference surface to increase the probability that sampled test points have close correspondences. These results show that the Direct Surface Distance results are very similar for the two test surface sampling rates, indicating little degradation as we reduce the number of test surface points while maintaining the same surface coverage. Note that the reference surface sampling does not affect the Direct Surface Distance since it uses

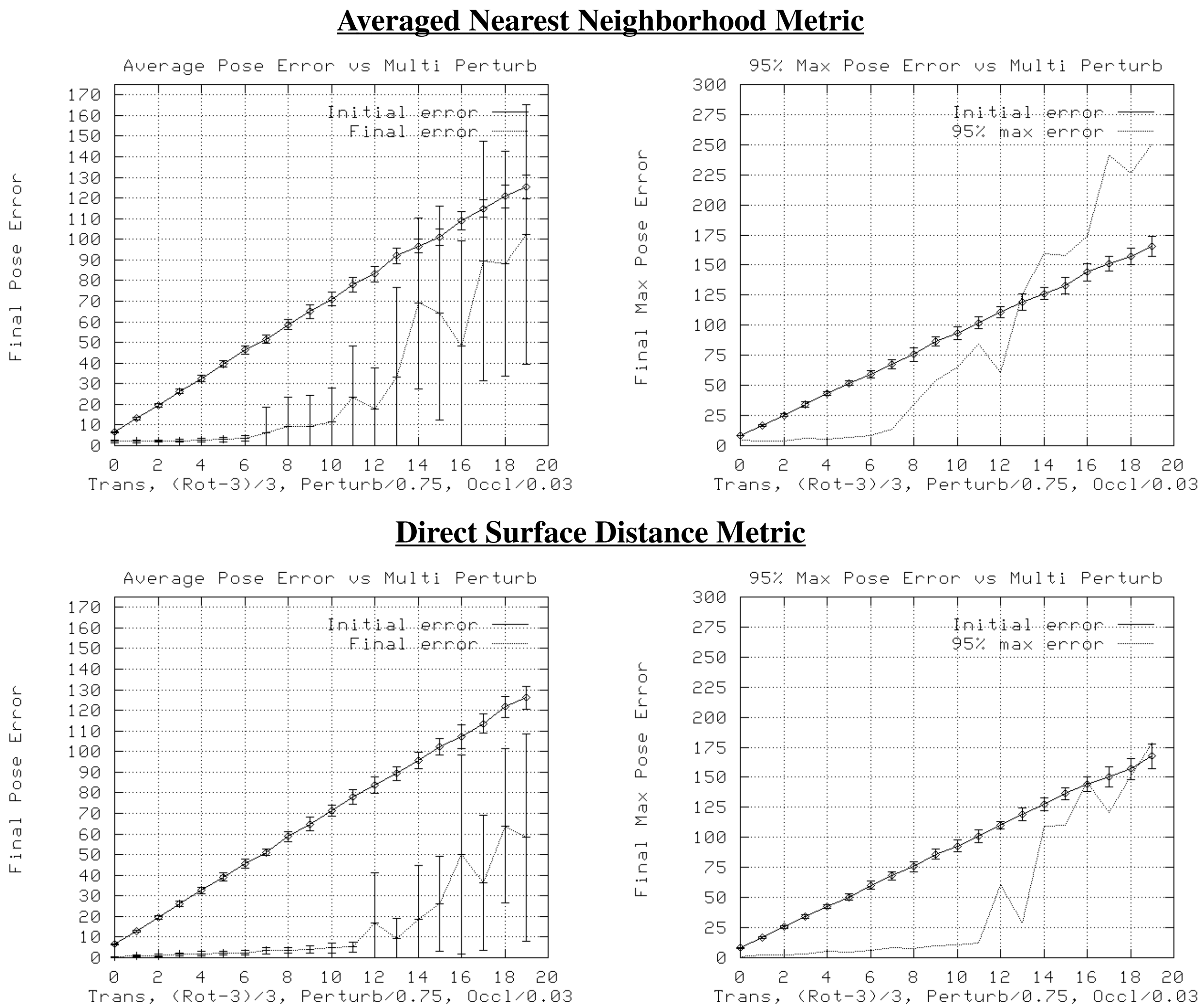


Figure 48 Results (2 of 2) of objective function comparison experiments. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

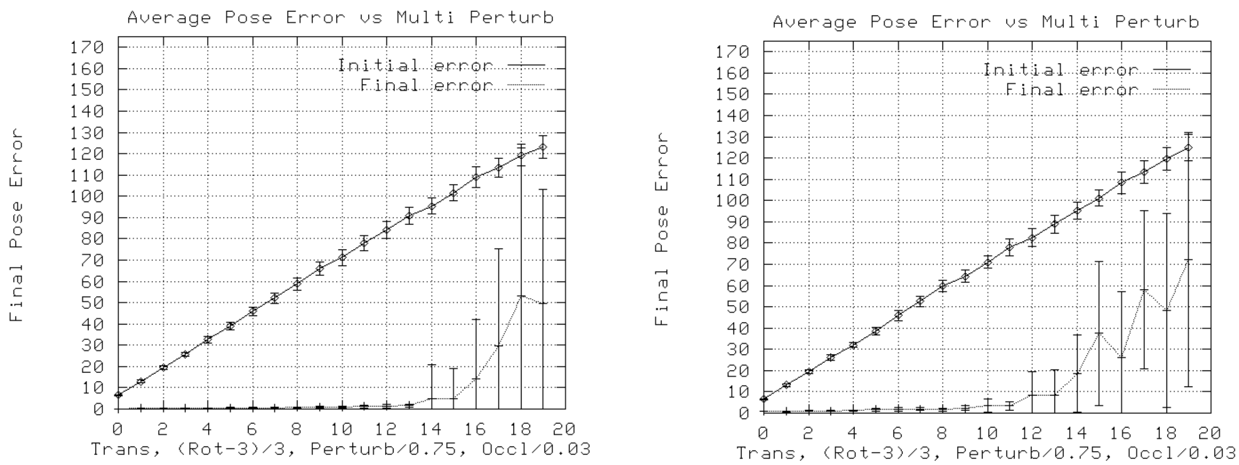
the surface triangles—only the test point sampling affects this metric. The Gaussian-Weighted Nearest Point and Averaged Nearest Neighborhood metric demonstrate measurable improvements in both pose region of convergence and accuracy with increased surface sampling, thus showing greater performance degradation with reduced sampling rates. This result is consistent with the similar observations made in Section 5.3.1.2.

5.4.1.2 Registration Accuracy

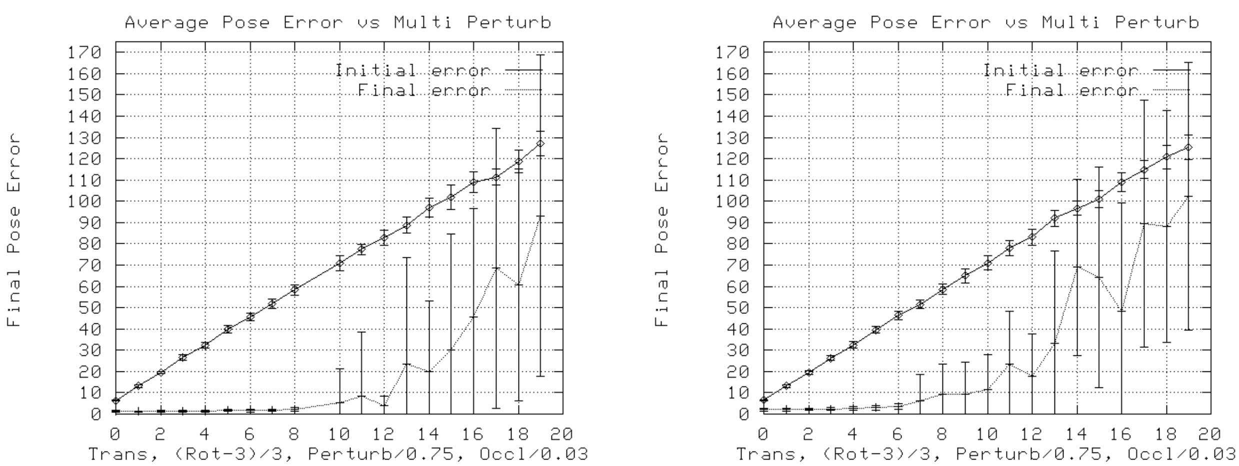
In order to analyze registration accuracy we focus on the pose error within the region of convergence. Figure 50 shows the average bounding box corner error for the four objective functions analyzed in Section 5.4.1.1. The curves are just close-ups of the graphs shown in Figure 47 and

2000 Test Surface Points *500 Test Surface Points*

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

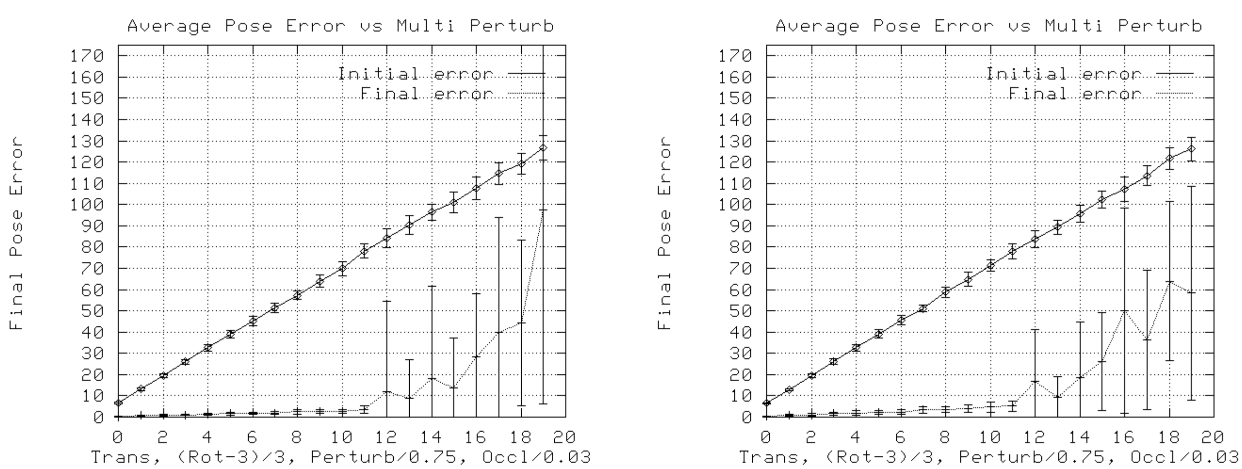


Figure 49 Results of test surface sampling comparison. Test surface sampled with 2000 points on left and 500 points on right. Means and std dev of average pose error measure are shown.

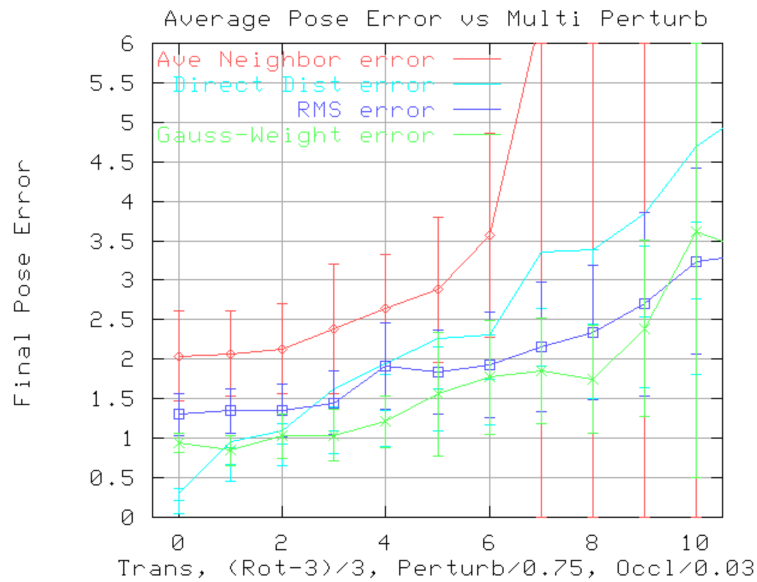


Figure 50 Accuracy comparison of four objective functions.

Figure 48 within the region of convergence. From these results we conclude:

- The Gaussian-Weighted Nearest Point metric has better accuracy than the Saturated RMS Nearest Point metric. The improvement is not statistically significant at all points, but is sufficient to prefer the Gaussian-Weighted Nearest Point. In Figure 47 we see a slightly better region of convergence for the Saturated RMS Nearest Point metric, but that result is not enough to offset the improved accuracy over a wide data perturbation range of the Gaussian-Weighted Nearest Point metric.
- The Averaged Nearest Neighborhood metric has consistently worse accuracy than the other three objective functions.
- The Direct Surface Distance metric has the best accuracy for small perturbations, but lags a little behind the Gaussian-Weighted Nearest Point for large perturbations.

5.4.1.3 Objective Function Surface

In order to analyze performance differences across the objective functions we would like to visualize the objective function landscape to determine the distinctiveness of the correct pose and the possibilities that we may get trapped in local minima. Employing the head surface data that was used for the experimentation in this section we vary the rotation of the test data by $\pm 20°$ around the correct pose and map out each objective function. In order to determine the best directions along to which to visualize the objective function values we use the following procedure:

- For each of the three rotation parameters:
    - Apply a rotation offset in the range [-20˚, +20˚]
        - Optimize along the three translation dimensions only.

Using this procedure we obtain objective function surfaces that have factored out translation effects. At each point in the objective function surface we therefore measure the best surface alignment at the given rotation.

The objective function surfaces for the four functions tested are shown in Figure 51 and Figure 52. For the Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, and Direct Surface Distance we show the objective function surface at $\sigma = 2$ and $\sigma = 8$ to show the impact of the Gaussian weighting. The Saturated RMS Nearest Point metric has only one applicable plot. Note that the Saturated RMS metric does not reach 0 due to the surface sampling present in the data. We draw the following conclusions form the objective function surface plots:

- The sharpest valley is exhibited by the Gaussian-Weighted Nearest Point metric at $\sigma = 2$, which is to be expected for cases where the two point sets arise from the same data source. As we move away from the solution, this metric increases faster than the others. This sharp valley supports accurate localization of the correct pose and thus helps explain the slightly higher accuracy that was achieved in the outlier addition experiment, Section 5.3.1.1 and the pose region of convergence experiment, Section 5.3.4, where all the reference data points were used.
- The smaller Gaussian sigmas result in narrower valleys making it easier to accurately localize pose solutions. But away from the correct solution, the larger sigma yields a smoother surface with a more consistent slope that we would expect to be less susceptible to local minima. Note the small perturbations which start appearing on the $\sigma = 2$ plots as we move away from the correct pose.

Additional objective function surfaces, with laser data as the test sets, are shown in Section 5.5.1.

5.4.2 Hierarchical Refinement of Rigid Solutions

In Section 3.3.4 we described two classes of techniques to hierarchically refine the rigid registration solution: distance weighting and data smoothing. In distance weighting we sequentially reduce the value of the Gaussian standard deviation used to weight the contribution of each test point to the overall score. In data smoothing we sequentially increase the resolution of the reference surface by either surface sampling (for point-to-point measures) or direct surface smoothing (for point-to-surface measure). Here we evaluate the benefits of these approaches, focusing on accuracy and region of convergence. We test the three objectives functions: Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, and Direct Surface Distance.

5.4.2.1 Distance Weighting

We first analyze the need for hierarchical distance weighting refinement by examining the regis-

Saturated RMS Nearest Point | **Gaussian-Weighted Nearest Point**

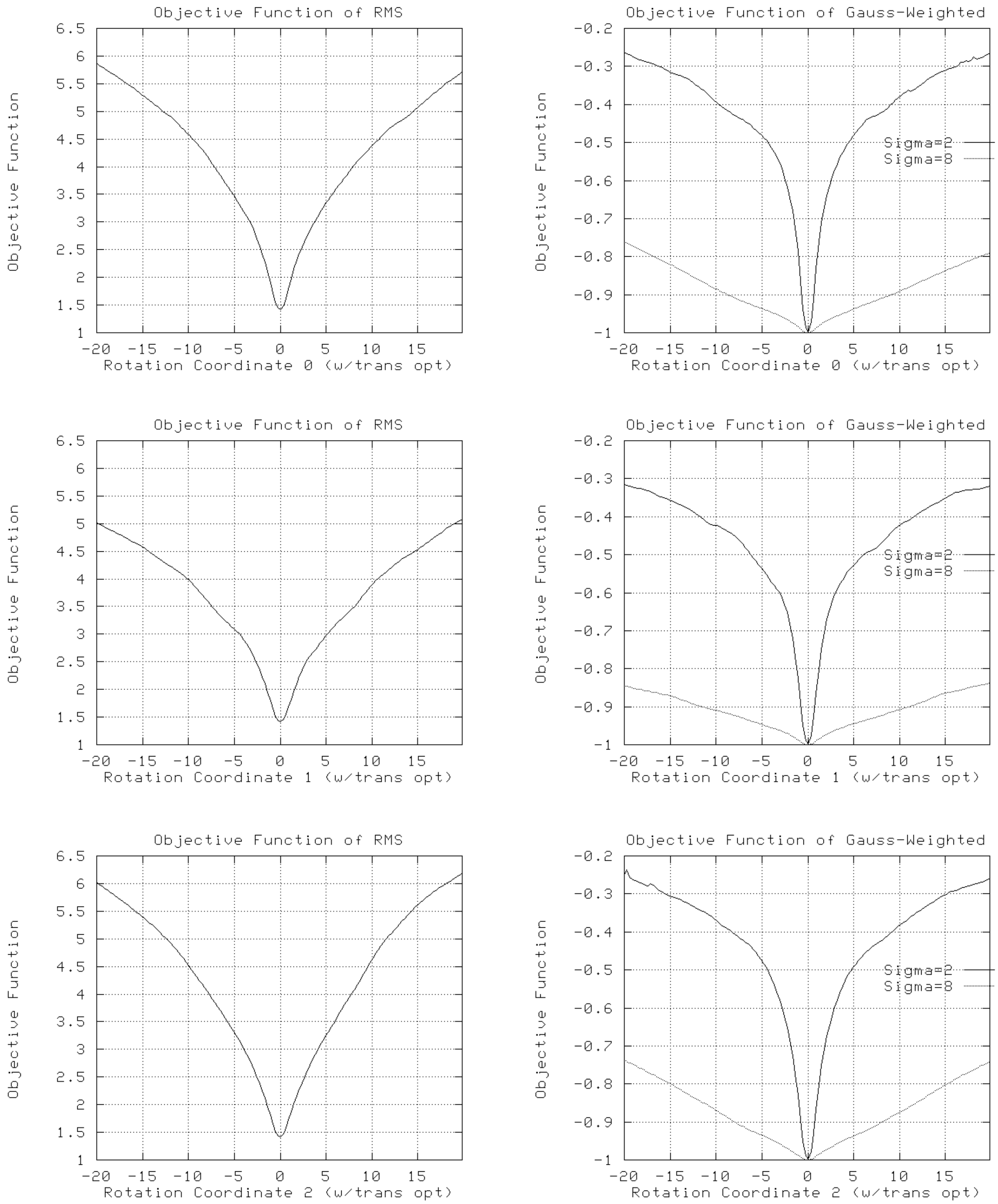


Figure 51 Objective function surface for the Saturated RMS Nearest Point and the Gaussian-Weighted Nearest Point metrics. Surface cutouts along the three rotation angles are shown. X-axis is deviation from the correct pose.

Averaged Nearest Neighborhood **Direct Surface Distance**

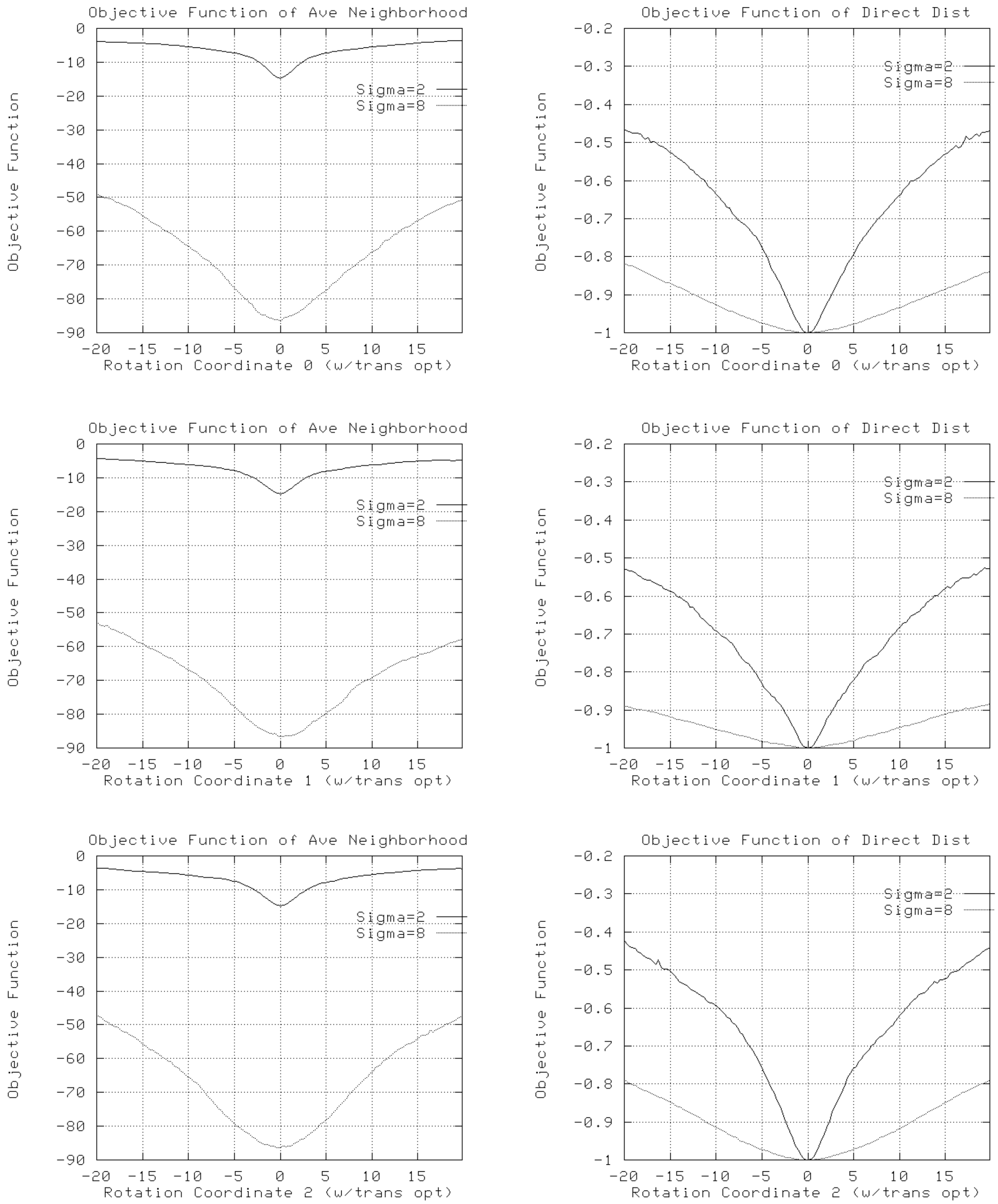


Figure 52 Objective function surface for the Averaged Nearest Neighborhood and the Direct Surface Distance metrics. Surface cutouts along the three rotation angles are shown. X-axis is deviation from the correct pose.

Table 8 Experiment specification for distance weighting comparison experiment.

Experiment Specification Parameter	Value
Objective function variants	Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 **5mm (first trial); 2mm (second trial)** 2000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**

Distance Weighting = 2 **Distance Weighting = 5**

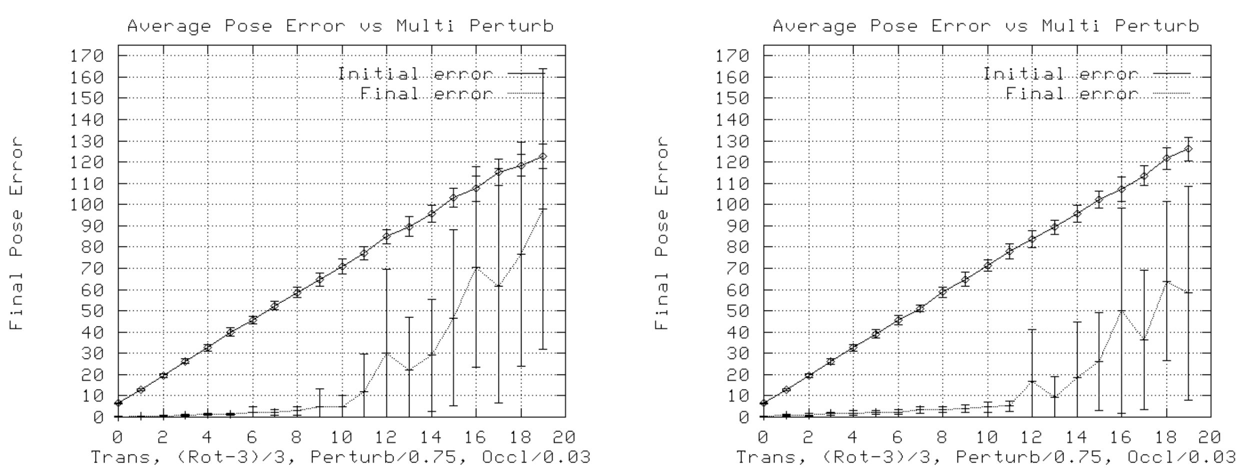


Figure 53 Comparison of distance weighting standard deviations on registration performance.

tration results obtained at different distance weighting values when run separately. We consider the same experiment as was run in Section 5.4.1, but with two values for the distance weighting standard deviation. The parameters for these two trials are given in Table 8. The Direct Surface Distance metric is used. Results are shown in Figure 53. The region of convergence is better with the larger σ, but we achieve better accuracy with the smaller σ. The accuracy benefits are better

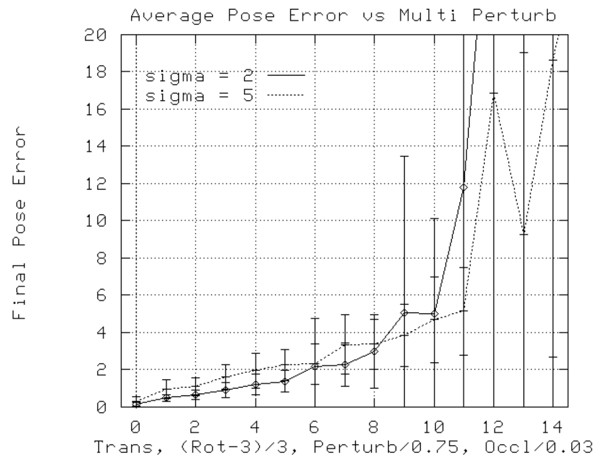


Figure 54 Accuracy comparison of distance weighting standard deviations within the region of convergence.

shown in the close-up of the region of convergence in Figure 54, in particular, for the first seven data points. We now explore the potential of achieving the combined benefits in a single hierarchical framework.

To test the hierarchical distance weighting we run the experiment whose parameters are given in Table 9. Here we apply the same test data perturbations as in Section 5.4.1, but now sequentially vary the distance weighting value from 8 to 5 to 2. The results are shown in Figure 55. Comparing these results to the performance at the single distance weighting value of 5 (the middle value in the hierarchical experiment) as previously shown in Figure 47 and Figure 48 we do see a marked improvement in performance. These improvements are tabulated in Table 10. The region of convergence is now demarcated by a more gradual degradation. To highlight this point we specify the boundary of the region of convergence in the table as a range signifying the point immediately before where performance starts to gradually drop and the point immediately before where performance sharply falls off. The latter point is particularly apparent in the 95th percentile pose errors. The pose errors are specified as the average boundary box corner errors. The hierarchical registration results demonstrate an improvement in both region of convergence and accuracy. The accuracy improvements in the hierarchical version over the single registration loop with $\sigma = 2$ (the finest hierarchical level) and the single registration loop with $\sigma = 5$ are shown in Figure 56 for the Direct Surface Distance metric. Up to 18° offset the pose error of the hierarchical version is identical to the single loop $\sigma = 2$ version. With larger offsets we see improvement with the hierarchical version, which although not outside the one standard deviation bounds indicates a beneficial trend.

Table 9 Experiment specification for hierarchical distance weighting experiment.

Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 **{8, 5, 2}** 2000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**

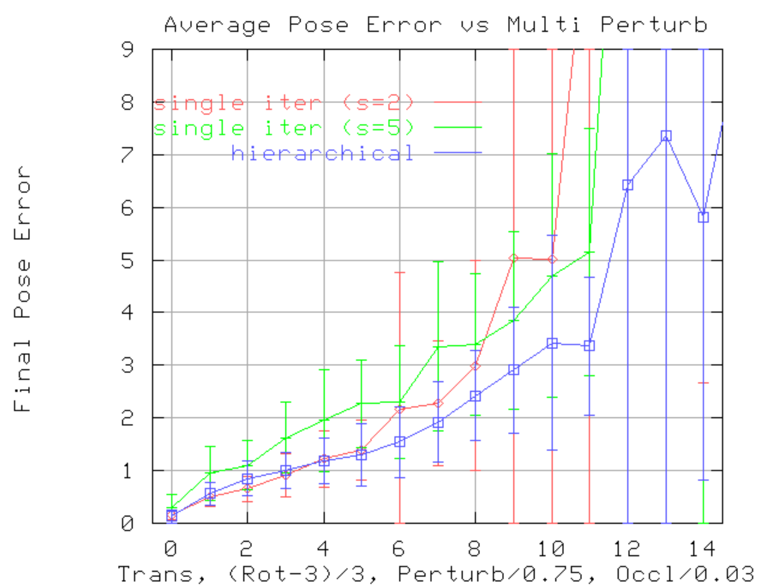
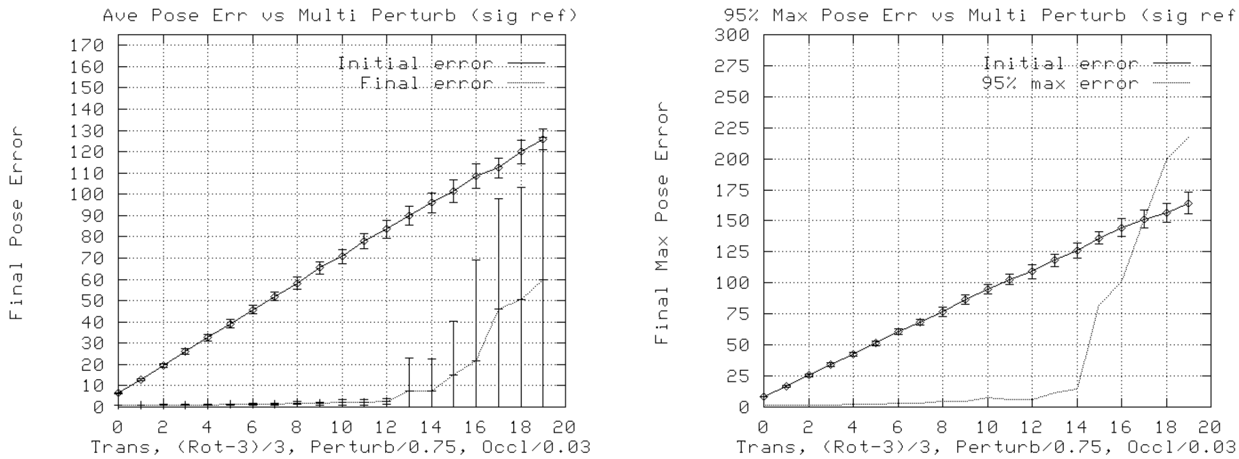
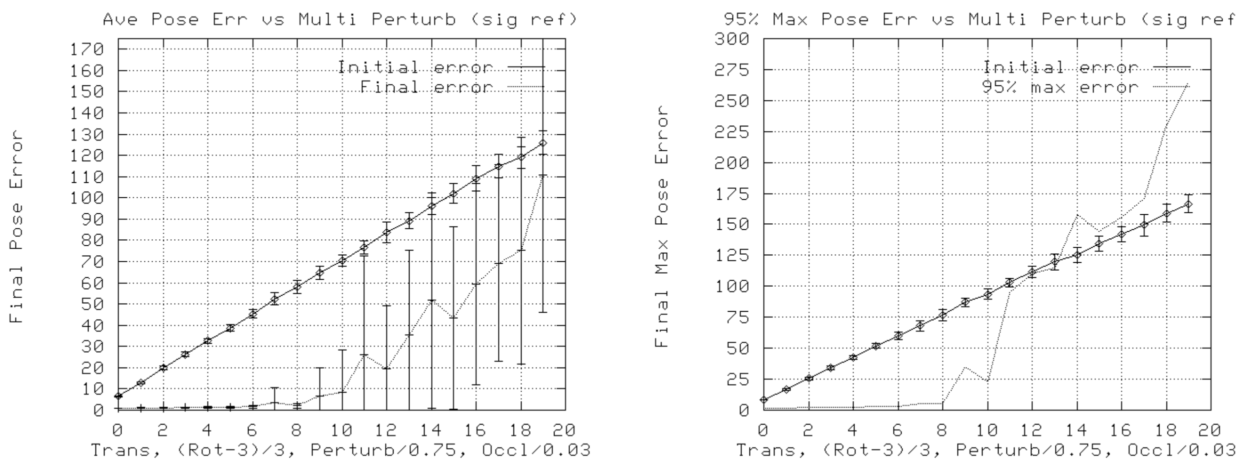


Figure 56 Accuracy comparison of the hierarchical distance weighting registration algorithm and single loop registration at weighting standard deviations of 2mm and 5mm.

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

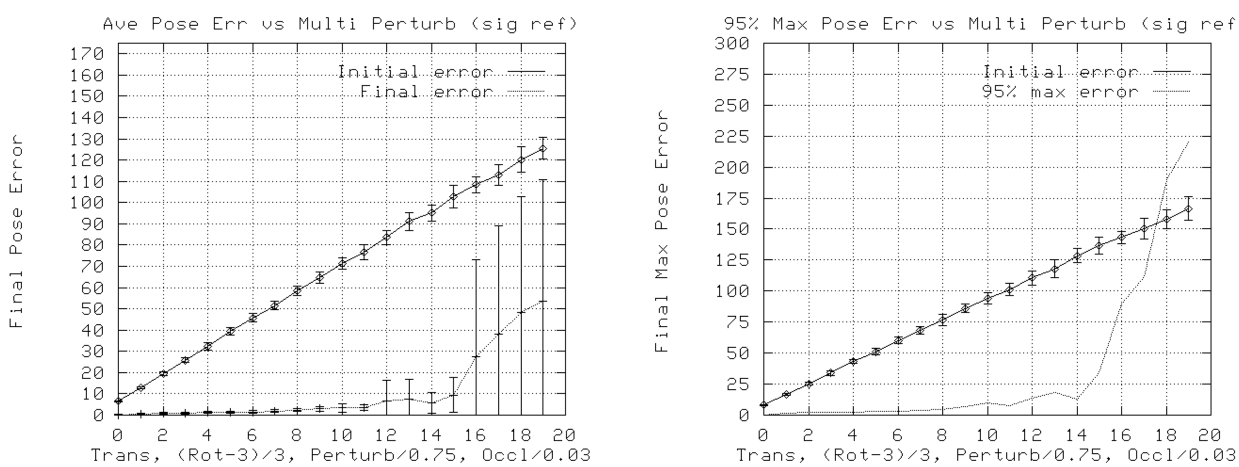


Figure 55 Results of hierarchical distance weighting experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

Table 10 Hierarchical distance weighting vs. single weighting value at mid-point of hierarchy.

		Hierarchical	**Non-hierarchical**
Gaussian-Weighted Nearest Point	Approximate region of convergence	39°-45°	36°-42°
	Pose error measure at rotation offset 3°	0.9mm ±0.0	0.9mm ±0.1
	Pose error measure at rotation offset of 12°	1.0mm ±0.2	1.0mm ±0.3
	Pose error measure at rotation offset of 21°	1.3mm ±0.4	1.8mm ±0.7
	Pose error measure at rotation offset of 30°	1.7mm ±0.6	2.4mm ±1.1
Averaged Nearest Neighborhood	Approximate region of convergence	27°-33°	21°-24°
	Pose error measure at rotation offset 3°	1.0mm ±0.1	2.0mm ±0.6
	Pose error measure at rotation offset of 12°	1.2mm ±0.2	2.4mm ±0.8
	Pose error measure at rotation offset of 21°	1.6mm ±0.5	3.6mm ±1.3
	Pose error measure at rotation offset of 30°	6.6mm ±13.4	9.3mm ±15.1
Direct Surface Distance	Approximate region of convergence	39°-48°	36°
	Pose error measure at rotation offset 3°	0.1mm ±0.1	0.3mm ±0.3
	Pose error measure at rotation offset of 12°	1.0mm ±0.4	1.6mm ±0.7
	Pose error measure at rotation offset of 21°	1.5mm ±0.7	2.3mm ±1.1
	Pose error measure at rotation offset of 30°	2.9mm ±1.2	3.9mm ±1.7

5.4.2.2 Data Smoothing

To test the hierarchical data smoothing we also ran the same experiment as in Section 5.4.1, but with multiple loops varying the surface sampling or smoothing parameters. The experiment parameters are given in Table 11. The surface sampling hierarchy is used for the Gaussian-Weighted Nearest Point and Averaged Nearest Neighborhood metric and the surface smoothing

Table 11 Experiment specification for hierarchical data smoothing experiment.

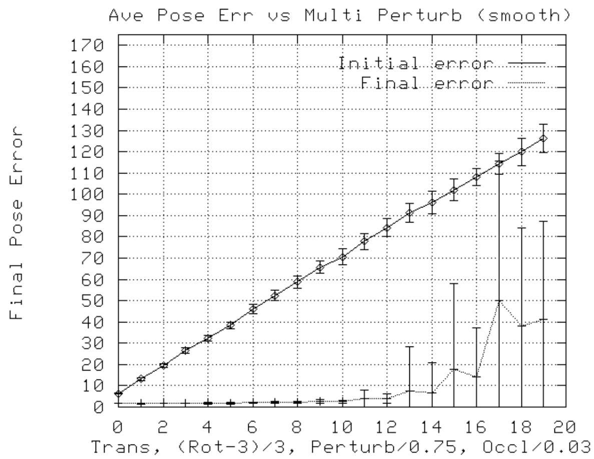
Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) **{500, 1500, 2500}** **{4.5mm, 2.5mm, 0.5mm}**
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**

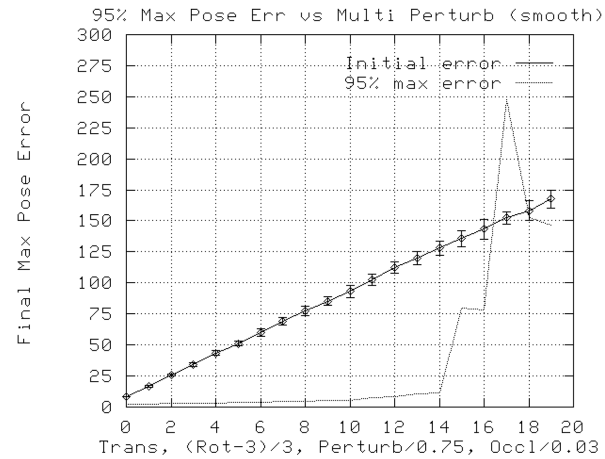
hierarchy is used for the Direct Surface Distance metric. The results of this experiment are shown in Figure 57. We compare these results against the non-hierarchical registration experiment run with the same perturbation parameters shown in Figure 47 and Figure 48. The single loop experiments used a surface sampling of 2000 points and a surface smoothing of 0.5mm, similar to the end points of the hierarchical experiment. In comparing the hierarchical and non-hierarchical cases we see only a small improvement in region of convergence, with the accuracy measures inside the region of convergence remaining the same. The improvement in the region of convergence is about one test data point in the tests in each of the objective functions—corresponding to about 3°.

5.4.2.3 Multi-Dimensional Hierarchical Reasoning

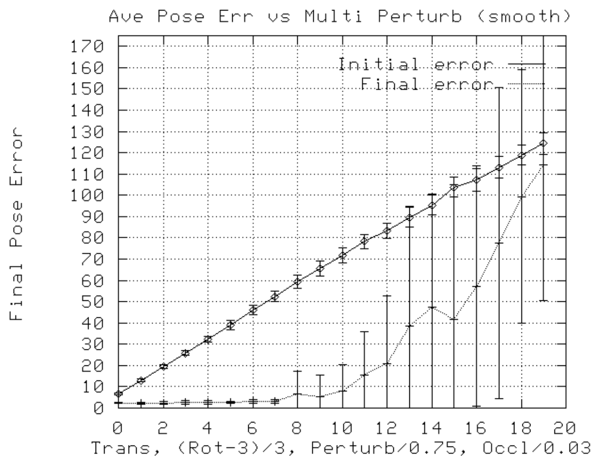
As a final hierarchical reasoning experiment we combine the distance weighting and data smoothing hierarchies into a single registration run. Data smoothing is run on the outer loop and distance weighting on the inner loop. We therefore start with the coarsest reference surface description and

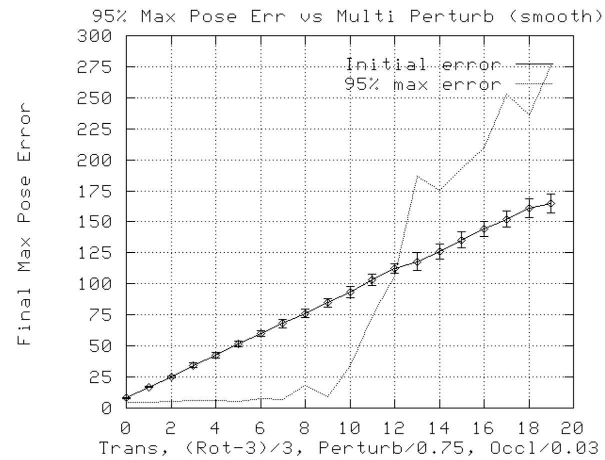
Gaussian-Weighted Nearest Point Metric



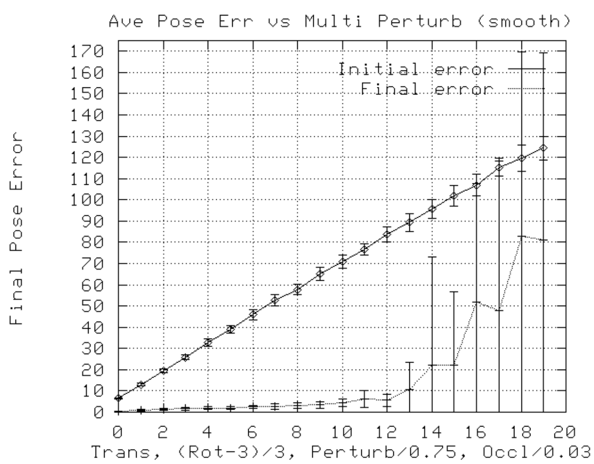


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



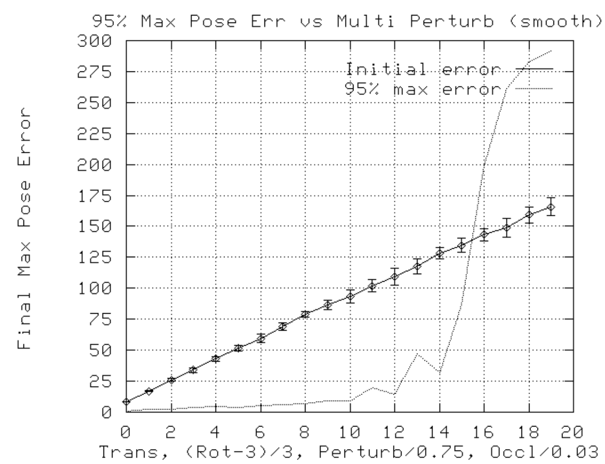


Figure 57 Results of hierarchical data smoothing experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

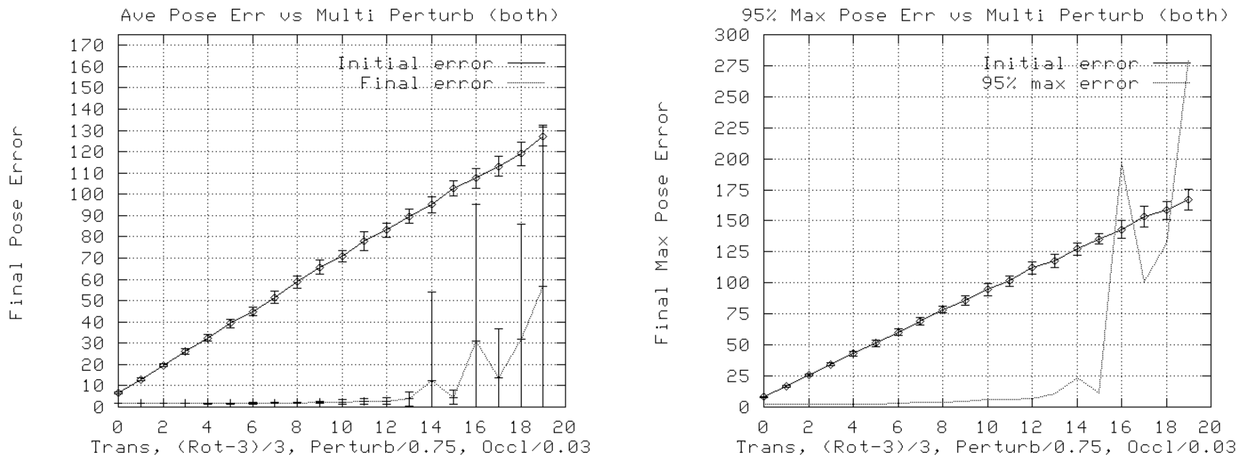
Table 12 Experiment specification for hierarchical data smoothing and distance weighting experiment.

Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 **{8, 5, 2}** **{500, 1500, 2500}** **{4.5mm, 2.5mm, 0.5mm}**
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3**° Increment per run: **3**°
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**

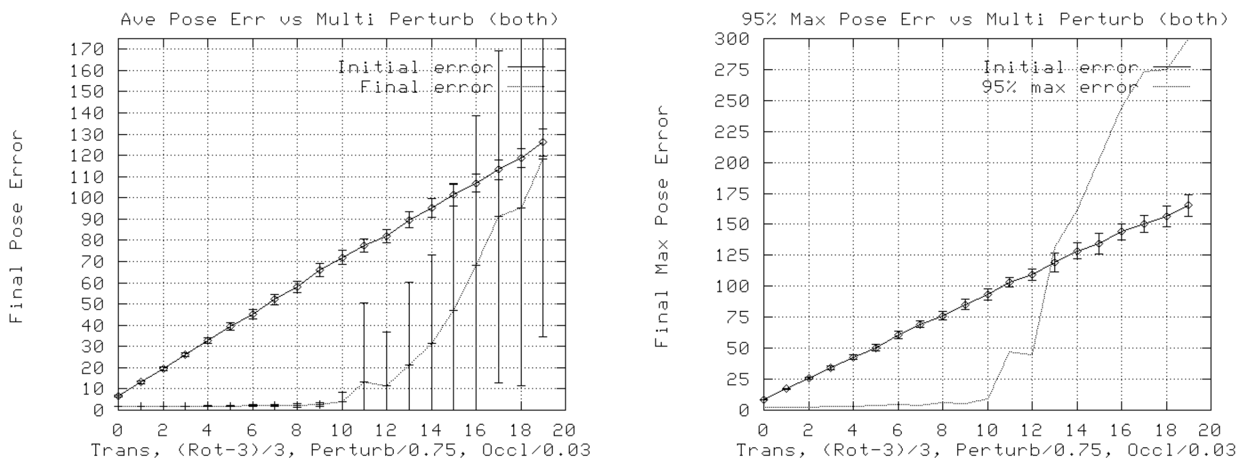
the largest distance weighting standard deviation. The result of that registration pass is then the input for the next registration loop which uses the next smallest distance weighting standard deviation. In this manner we run through all the distance weighting values and the result of the last one is fed into a registration loop with the next finest reference surface description and the largest distance weighting parameter again. We continue until we reach the finest reference surface description and smallest distance weighting standard deviation.

We again ran the same experiment as in Section 5.4.1, but with multiple outer loops varying the surface sampling or smoothing parameters and multiple inner loops varying the distance weighting parameter. The experiment parameters are given in Table 12. The results of this experiment are shown in Figure 58. In comparing these results to the results obtained with the distance weighting only hierarchical registration in Figure 55 we see little change. That is, adding a data smoothing hierarchy to the distance weighting hierarchy provides little improvement. Given the results of the data smoothing only hierarchy experiment, in Section 5.4.2.2, and the combined hierarchical reasoning we do not foresee the need for the data smoothing hierarchy.

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

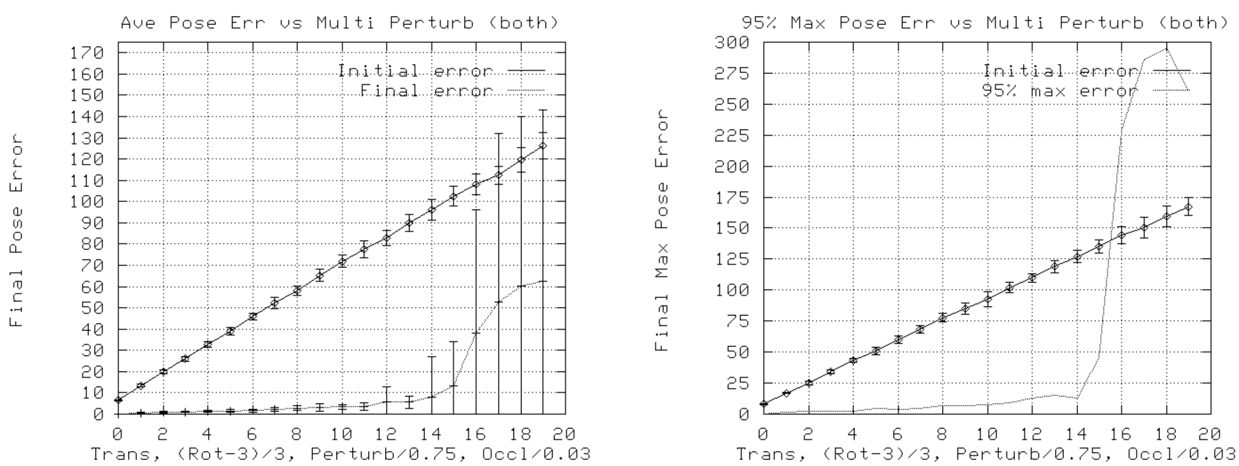


Figure 58 Results of combined hierarchical data smoothing and distance weighting experiment. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

5.4.3 Incorporation of Affine Deformations

We now incorporate an affine transformation into the region of convergence test. Specifically, we build upon the multi-perturbation experiment described in Section 5.4.1.1 in which we simultaneously varied the rigid pose offset, surface patch removal rate, and spatial perturbation. We now follow all those perturbations with random scaling and shearing perturbations of the test points. This additional affine offset is applied by pre-multiplying each test point with the homogeneous transformation matrix:

$$\begin{bmatrix} 1+\gamma_x & \beta_{xy} & \beta_{xz} & T_x \\ \beta_{yx} & 1+\gamma_y & \beta_{yz} & T_y \\ \beta_{zx} & \beta_{zy} & 1+\gamma_z & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where γ_i is a random scaling value in the range [-0.1,0.1] for each i, β_j is a random shearing value in the range [-0.1,0.1] for each j, and T_k is a random translation value in the range [-0.1,0.1] for each k. The same affine matrix is applied to all test points for each simulation. A sample affine distortion for parameters in this range is shown in Figure 59. In this figure the three scale parameters, γ_i, are [0.02, -0.02, -0.03], the six shearing parameters, β_j, are [-0.10, -0.01, 0.00, -0.10, -0.07, 0.04], and the translation parameters are all 0.

We are thus interested in not only how well we recover the rigid components of the transformation, but also the scaling and shearing components. Although we expect the affine solution to warp the surface points to align reasonably well, we are more interested in how well we recover the ground truth transformation. By evaluating the pose error at the bounding box corners of the test points we obtain a more reasonable measure of pose error than by looking at the just the surface points themselves. If through our registration process we generated the wrong rigid transform, which was at least partly compensated for by the affine transform, it is unlikely that the bounding box corner points will align well even if many of the test surface points are close to the reference surface.

The parameters for the affine pose region of convergence experiment are given in Table 13. Here we only experiment with the Direct Surface Distance metric. The results are shown in Figure 60. In the pose error graph we plot both the final pose error, after affine registration, as well as the intermediate solution, after we complete the rigid registration pass. We make the following observations:

- We are able to recover the affine offset reasonably well for small test offsets—the average

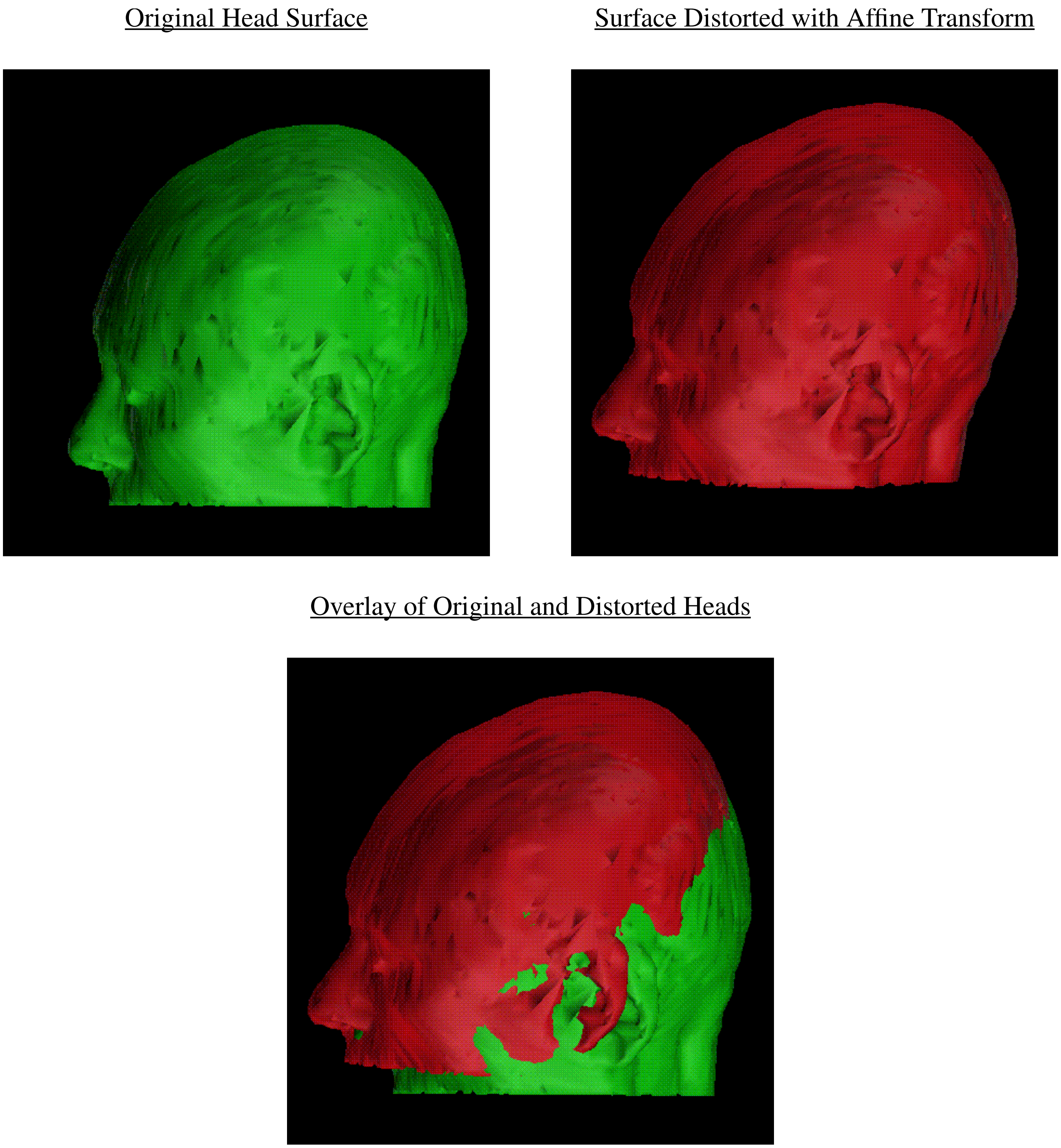


Figure 59 Effect of affine offset we are attempting to recover in this experiment. Original head is shown in green and distorted head is shown in red.

pose error measure is less than 10mm up to 12° offset (data point 3).

- The affine pose error remains approximately 10mm through 27° (data point 8). The intermediate rigid error is relatively stable in that same range.
- The rigid error is around 20mm through 27°, whereas in the same experiment in Section 5.4.1.1 where we did not apply any affine offsets, our error was 3mm or less (see Figure

Table 13 Experiment specification for affine pose region of convergence experiment.

Experiment Specification Parameter	Value
Objective function variants	Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 5mm (fixed) 6949 (unused) 0.5mm (fixed)
Transform complexity	**Rigid, followed by affine**
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **3°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: **0%** Increment per run: **3%**
Location perturbations of test surface points	Initial: **0mm** Increment per run: **0.75mm**
Magnitude of affine scaling offsets	Initial: **0.1** Increment per run: **0**
Magnitude of affine shearing offsets	Initial: **0.1** Increment per run: **0**

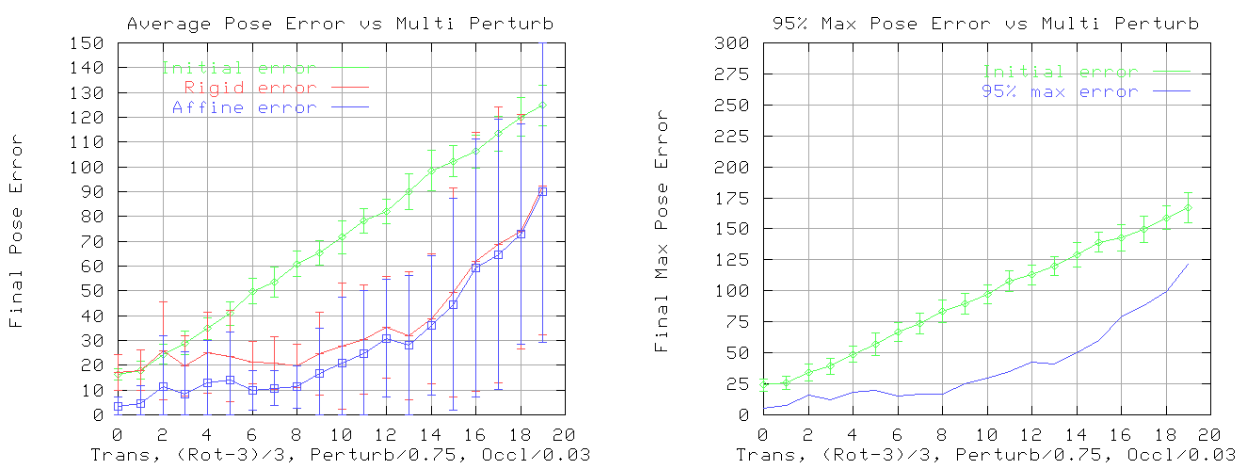


Figure 60 Results of affine pose region of convergence experiment using the Direct Surface Distance metric. Means and std dev of average pose error measure are shown on left for both final affine transform and intermediate rigid transform. 95th percentiles of maximum pose error measure are shown on right for final affine result.

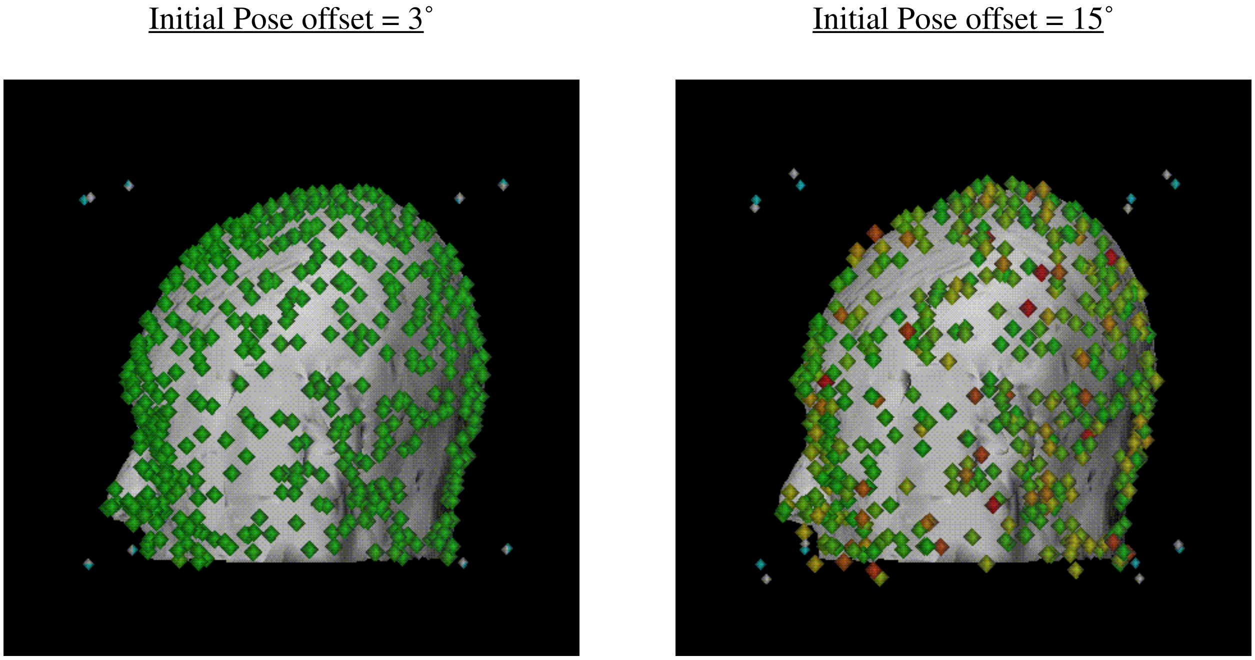


Figure 61 Residual errors for affine registration results. Distance from points on transformed test surface to reference surface is shown color-coded from green=0mm, to yellow=2.5mm, to red=5+ mm. Also shown are the bounding box corners: small cyan dots are box corners of reference surface and small white dots are box corners of transformed test surface.

48). The shift in the initial error (prior to any registration) between the experiment in which we applied no affine offset and here is in the 1-9mm range. That is, the rigid solution cannot totally compensate for the affine offsets.

- The region of convergence in this case appears to be about 27˚, whereas in the comparable experiment with no affine offset, the region of convergence was at least 36˚.
- The difference between the rigid and affine solutions decreases as the magnitude of the error increases. That is, outside our region of convergence, the rigid solution does not pull us close enough to a point where the affine registration can converge to the correct answer. The affine registration makes little difference in this case. As the rigid error and affine error curves converge we are in fact just as likely to have a larger error with the affine solution as with we are with the rigid solution.

We examine two cases of the affine registration tests more closely in order to study the relationship of the bounding box corner error measure and the residual error of the registered skin points. The two cases are taken from run 0, in which we have a 3˚ rotation offset with no other perturbations, and from run 4, in which we have a 15˚ rotation offset along with a 4mm translation offset, spatial perturbation with a Gaussian standard deviation of 3mm, and surface patch removal of 12%. We show the residual errors and the locations of the bounding box corners for these cases in Figure 61. In the first case, with initial rotation offset of 3˚, the average residual distance after registration along the skin surface is 0.08mm while the average bounding box corner error is 0.45mm. In the second case, with initial rotation offset of 15˚, the average residual distance after registration along the skin surface is 1.27mm while the average bounding box corner error is 6.39mm. In this latter case, the test surface points were also spatially perturbed, resulting in the broad distribution of residual errors. We thus observe that under affine registration the bounding box corner error is generally larger than the residual error along the surface. That is, with the affine transformation, any registration errors are magnified at the bounding box corners, truly highlighting the worst-case nature of this pose error measure. With rigid registration, the bounding box corner error is generally the same or somewhat smaller than the surface residual error.

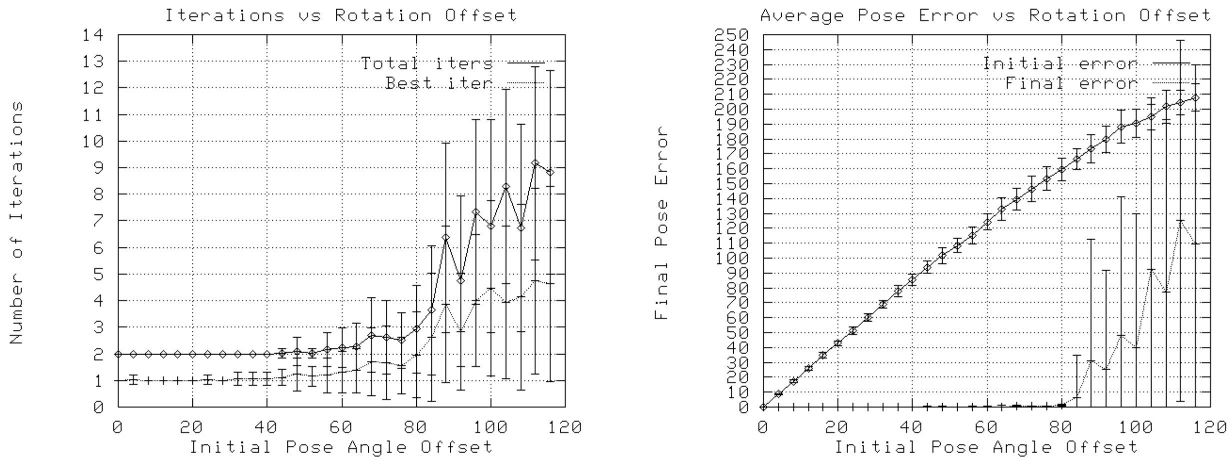
5.4.4 Iterative Reasoning

Within every registration loop in the hierarchical reasoning or within a non-hierarchical registration run we verify the result we generated by applying a random perturbation to it and re-optimizing. In such a way we check whether we converge to the same pose or if possibly our earlier pose solution was a local minimum. At issue is how many such iterations are generally necessary—i.e., how often do we converge on the ultimately reported answer on the first iteration. We analyze such convergence behavior for the pose region of convergence experiment described in Section 5.3.4. In this experiment we always perform at least two iterations, but allow up to 30 if we are able to continually reduce the objective function through the additional iterations. The exact stopping criteria are described in Section 3.5.3. In Figure 62 we show the iteration results in terms of the total number of iterations that were performed for each registration as we increased the rotation offset along with the iteration number which first achieved a score sufficiently close[1] to the score of the reported pose solution. The pose error measures are shown alongside to compare the number of iterations to resultant accuracy. For the Gaussian-Weighted Nearest Neighbor and Direct Surface Distance metrics, which achieved large pose regions of convergence, we observe the following behaviors:

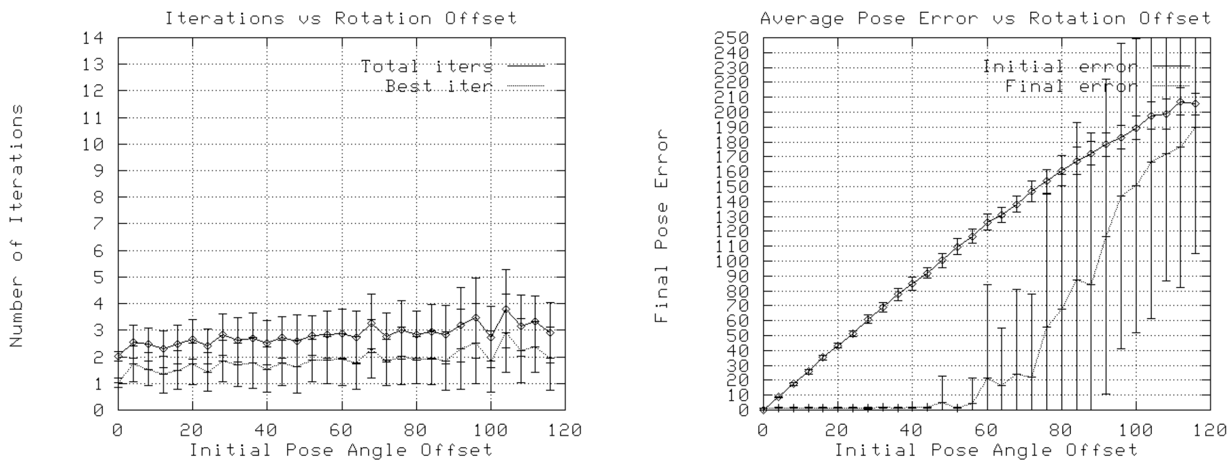
- Within the first part of the region of convergence, [0˚, 30˚], we generally converge on the correct pose on the first iteration and additional iterations are probably not required.
- For the remaining interval of the pose region of convergence, [30˚, 75˚], we are able to achieve highly accurate results through an additional one or two iterations. That is, for the more difficult problems which are probably characterized by variations other than just pose offset, the multiple random pose perturbations and re-optimization do improve the

1. Scores within 0.005 were deemed identical in this experiment.

Gaussian-Weighted Nearest Point Metric



Averaged Nearest Neighborhood Metric



Direct Surface Distance Metric

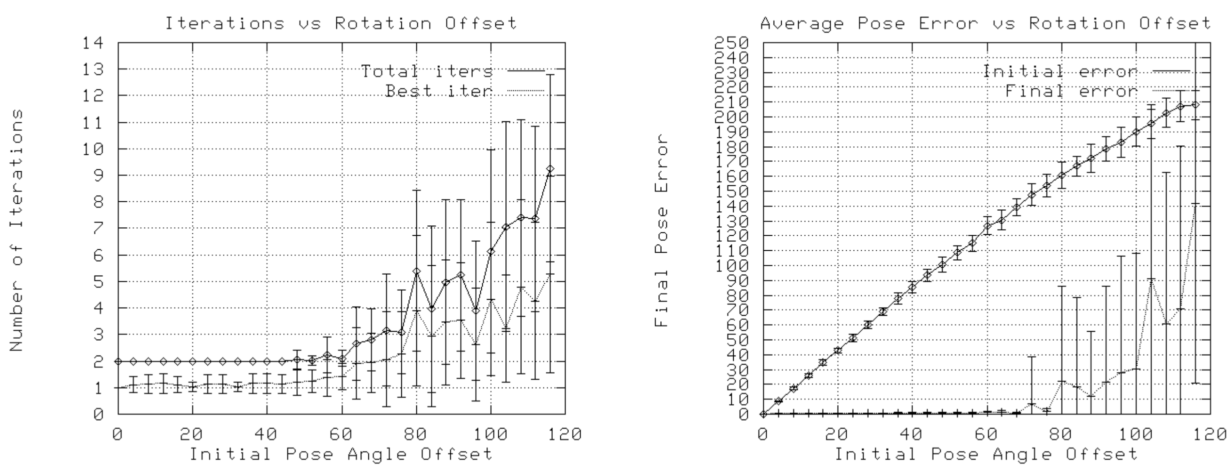


Figure 62 Number of optimization iterations performed, left side, in order to reach average pose error measures on right. Results are for pose region of convergence experiment.

results. Since we used a relatively small score tolerance, 0.005, to determine whether iterations resulted in the same result, the improvements we achieved through the multiple iterations were not necessarily always highly significant, but some payoff is apparent.

- As we move outside the pose region of convergence, we see that many more iterations were performed, although with little benefit as we were still not able to converge on the correct pose. In these cases, the global minimum is so far away in pose space, that the small random perturbations apparently just result in other local minima.

5.4.5 Algorithm Analysis Conclusions

In this section we evaluated registration performance as a function of algorithm parameters. The key issues analyzed were: performance comparison of four objective function metrics, evaluation of hierarchical registration, and analysis of the need for iterative random perturbations. We made the following conclusions:

- The Gaussian-Weighted Nearest Point and Saturated RMS Nearest Point metrics achieve similar performance, but the Gaussian-Weighted Nearest Point metric achieves slightly better accuracy.
- The Gaussian-Weighted Nearest Point and Direct Surface Distance metrics achieve good accuracy, 1-2mm, over a similarly wide range of test data perturbations. Overall, the Gaussian-Weighted Nearest Point metric exhibited slightly better accuracy, but also greater sensitivity to sampling artifacts.
- The Averaged Nearest Neighborhood metrics achieves reduced accuracy and a smaller region of convergence than both the Gaussian-Weighted Nearest Point and Direct Surface Distance metric.
- Hierarchical distance weighting provides measurable improvement in both accuracy and region of convergence. These conclusions are supported by the shapes of the objective functions surface as we vary the distance weighting factor.
- Hierarchical data smoothing does not provide measurable performance improvements.
- Performing a small number of random local pose perturbations and re-optimizations provides improved performance for complex registration problems, but only up to a certain point beyond which any number of perturbations probably will not help.

5.5 Data Sensitivity Application Experiments

In this section we analyze registration performance, using experiments similar to Section 5.3, but using actual case studies from the applications described in Chapter 4. In these experiments the reference dataset is an actual MR image. The test dataset is also actual data of the same subject, but from a different sensor or at a different time point. In order to estimate ground truth we manually initialize the registration at a point close to the expected solution, run the registration algorithm using hierarchical distance weighting and the Direct Surface Distance metric, and manually verify the registration using our residual error display. We then apply controlled perturbations from this registered condition to measure the registration sensitivity to those perturbations.

5.5.1 Laser/MRI Registration

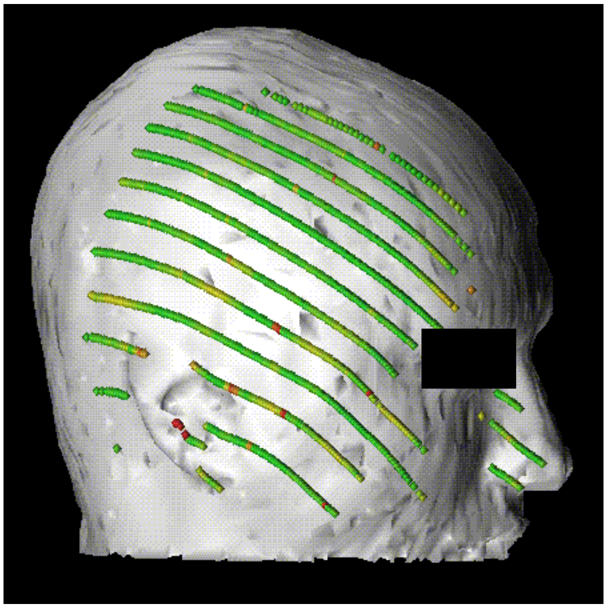
We first evaluate the performance of registering laser data of a subject's scalp to skin surface segmented from an MR image. This type of registration is used in the image-guided surgery application, Section 4.1, and transcranial stimulation application, Section 4.4. The laser data was collected as part of some of the image-guided surgery trials. The laser data for each subject included in this experiment at the ground truthed pose is shown in Figure 63. We see that the ground truth registration is reasonably close, although a few outliers remain due to possible registration error, segmentation error, relative skin motion, or MR sensor distortion.

5.5.1.1 Pose Region of Convergence

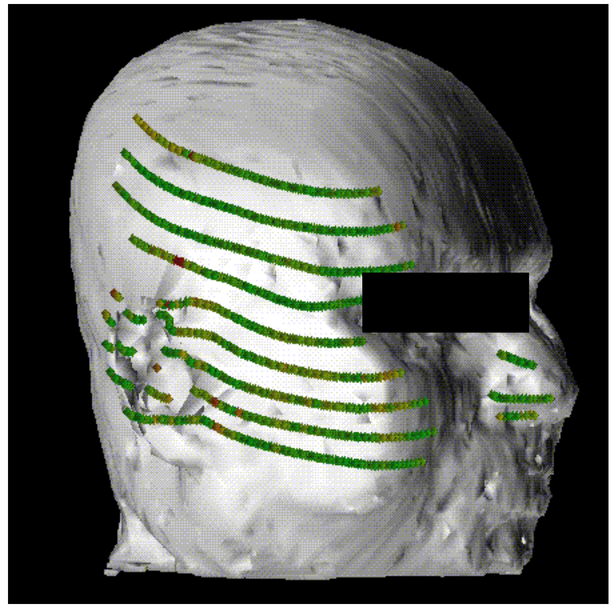
Starting at the ground truth pose for subject 1 we measure our pose region of convergence by offsetting that pose by increasing amounts and re-registering. We then evaluate our registration accuracy by measuring the distance between the corners of the laser data bounding box at the ground truth pose and the newly registered pose. This experiment is similar to the one performed in Sections 5.3.4, but here we perform the registration using hierarchical distance weighting. We also apply a small translation offset in addition to the rotation offset. The experiment parameters are given in Table 14. A total of 1200 laser points were available for this subject and all were used.

The registration results are shown in Figure 64. The Averaged Nearest Neighborhood metric performs poorly. It appears to converge on a different solution than the other two metrics, but the high standard deviation on its pose error indicates that the solution is not very stable. The other two metrics, on the other hand, achieve stable solutions for at least 10° of pose offset, as demonstrated by their relative small pose error standard deviation. The accuracy of the solutions in this range consists of up to 8mm of error, but that error is at the bounding box of the laser data and includes possible ground truth specification errors. The actual error in the region of interest, the central region of the laser scan, is likely to be smaller. We achieve similar results with the Saturated RMS Nearest Point metric, as shown in [30], where it is also shown that errors in the region of interest are generally an order of magnitude less than the resultant RMS error. We also note that one reason for the smaller region of convergence than what we achieved in Section 5.4.1.1 is the much smaller surface overlap between the test and reference data. Here we approximately cover

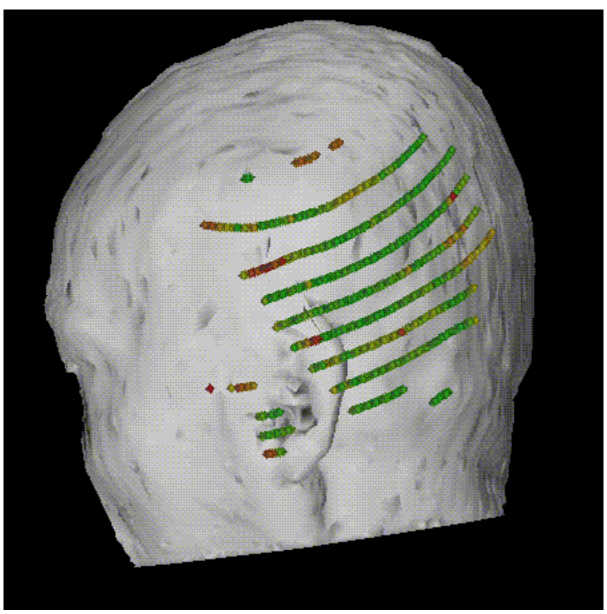
Laser Subject 1



Laser Subject 2



Laser Subject 3



Laser Subject 4

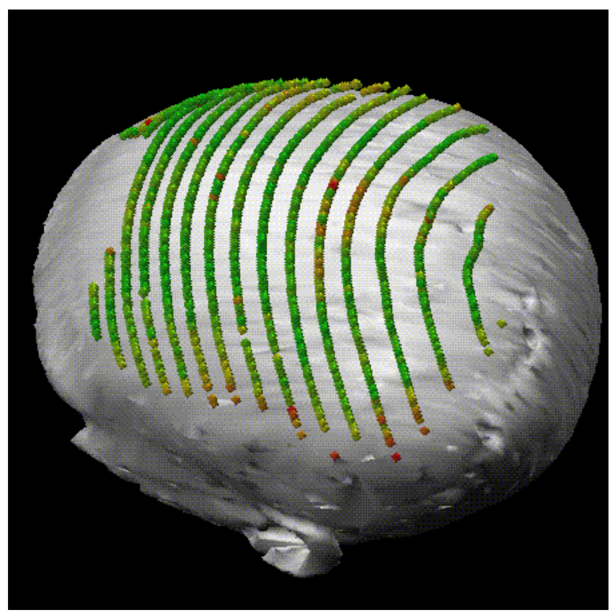


Figure 63 Four subjects used in laser registration experiments. White surface is skin surface segmented from MRI. Colored dots are laser data which has been registered to the MRI skin. Color denotes residual error between laser dots and MRI skin: green=0mm, yellow=2.5mm, and red=5+ mm.

Table 14 Experiment specification for laser data pose region of convergence experiment.

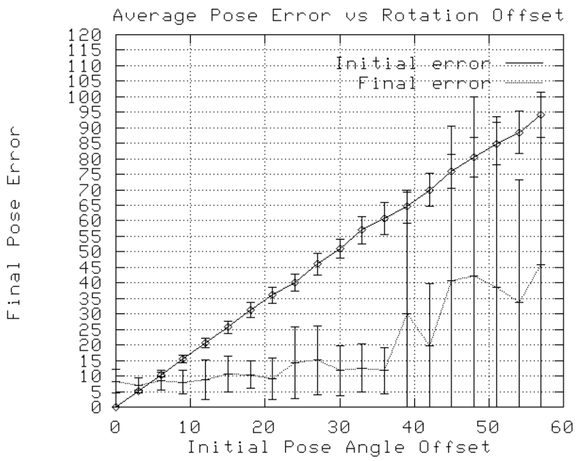
Experiment Specification Parameter	Value
Objective function variants	Gaussian-Weighted Nearest Point, Averaged Nearest Neighborhood, Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 **{8, 5, 2}** 5000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	500
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **0°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

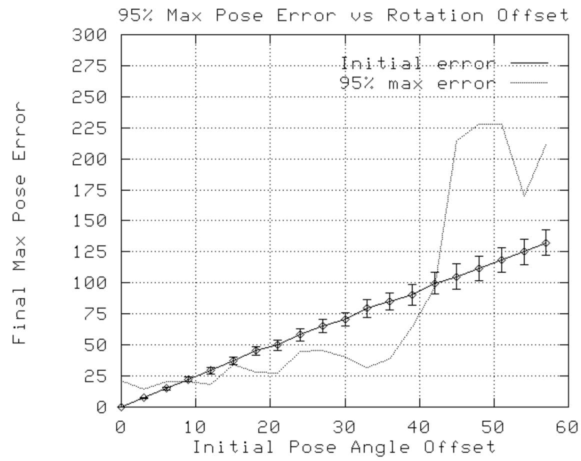
25% of the head surface with the laser and we earlier saw that with less than 40% surface overlap our registration performance with 15° offset starts to degrade.

For comparison we run the same experiment on the other three subjects shown in Figure 63 using the Direct Surface Distance metric. The results are shown in Figure 65. We see similar results with subject 2 as we saw with subject 1, in part since they have similar skin surface coverage. Subject 3 shows reduced accuracy and a minimal pose region of convergence, but that subject had only about 10% head coverage with the laser. Such small skin area coverage is clearly insufficient for registration. Subject 4 shows comparable results to subjects 1 and 2, with possibly an even larger region of convergence. This subject has laser coverage on most of the top of the head which although highly symmetric is also sufficiently large.

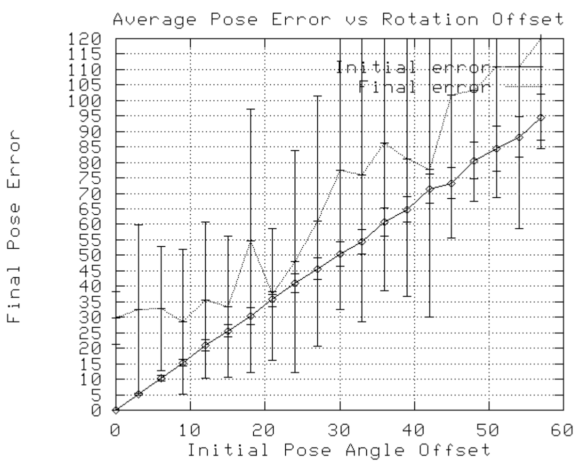
In order to better understand the laser registration performance on these subjects we generated the objective functions surfaces for two of the subjects in the same manner as in Section 5.4.1.3. These objective function surfaces are shown in Figure 66 for the Gaussian-Weighted Nearest Point metric, in Figure 67 for the Averaged Nearest Neighborhood metric, and in Figure 68 for the Direct Surface Distance metric. We see that the objective function surface for subject 1 is much

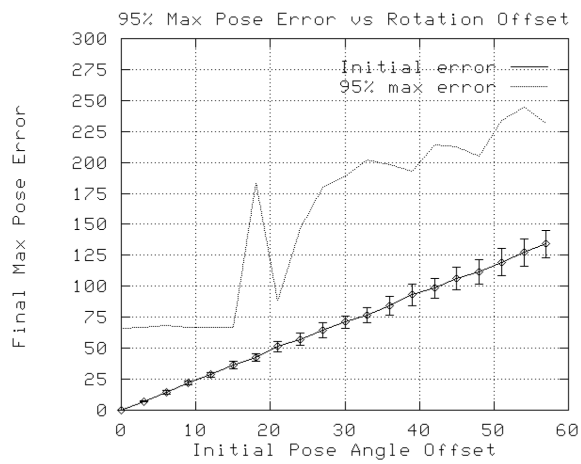
Gaussian-Weighted Nearest Point Metric



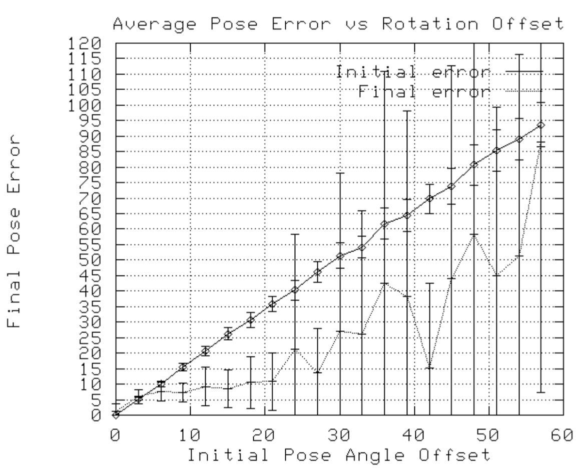


Averaged Nearest Neighborhood Metric





Direct Surface Distance Metric



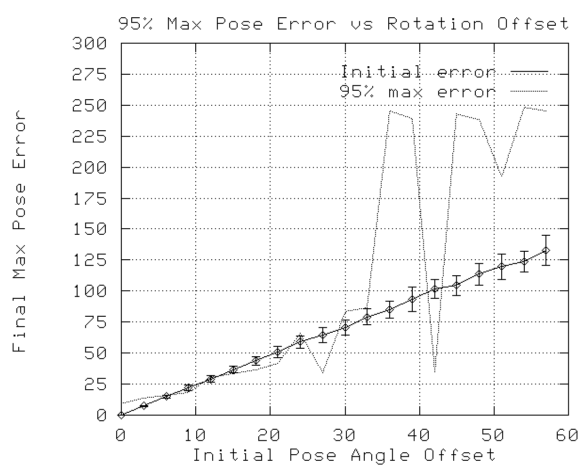
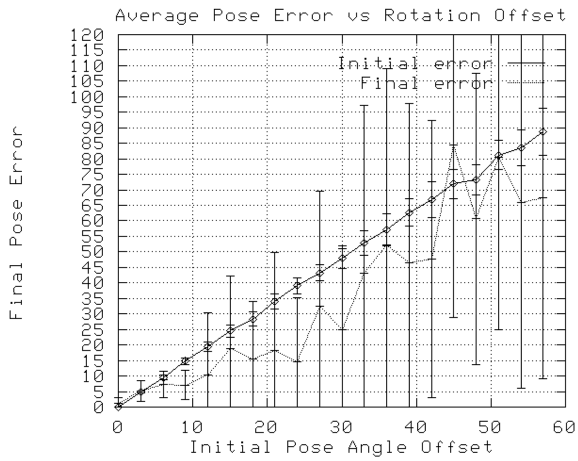
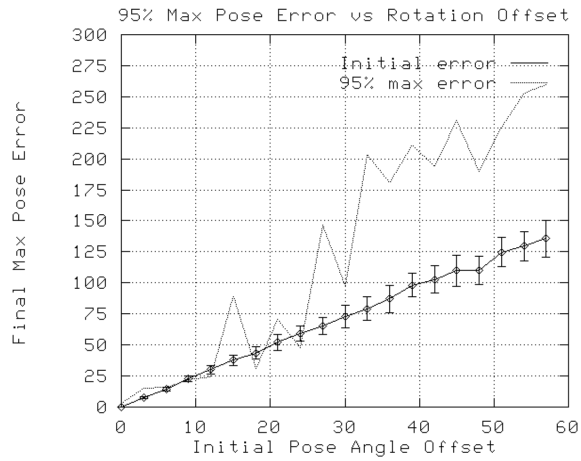


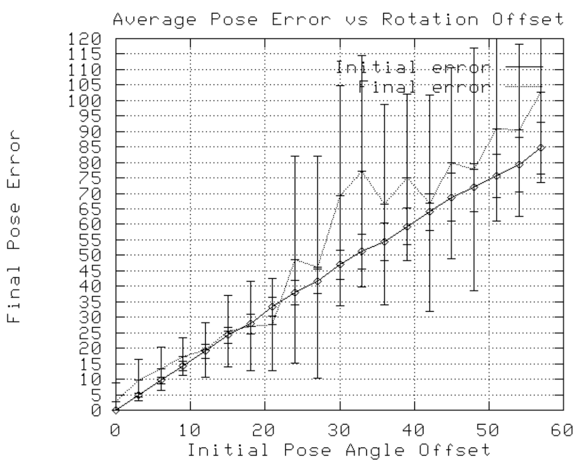
Figure 64 Results of laser pose region of convergence experiment on laser subject 1. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

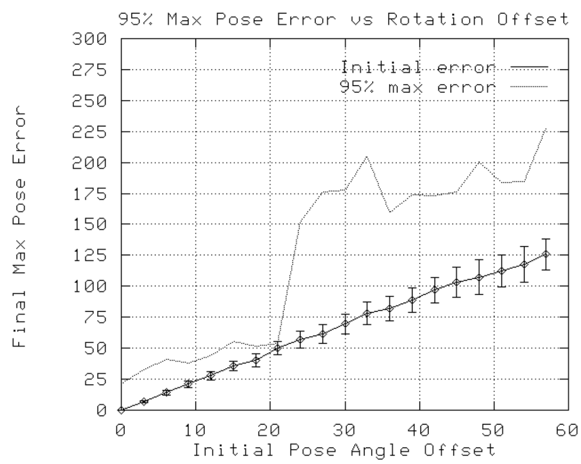
Laser Subject 2 Direct Surface Distance Metric



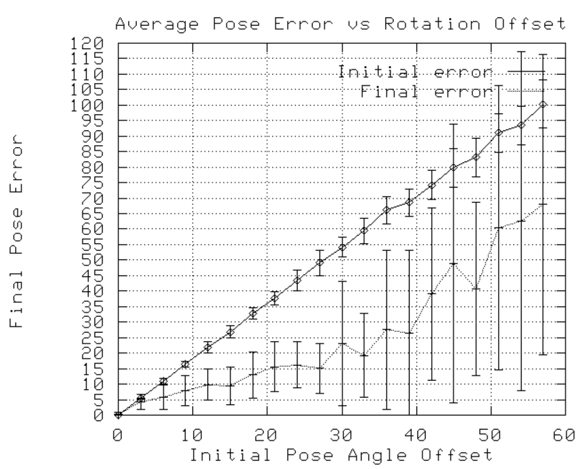


Laser Subject 3 Direct Surface Distance Metric





Laser Subject 4 Direct Surface Distance Metric



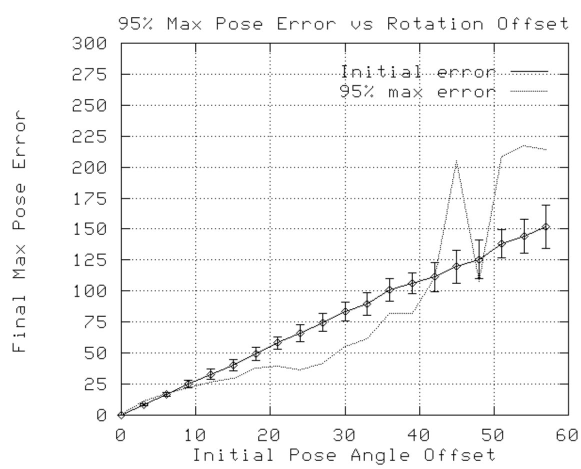


Figure 65 Results of laser pose region of convergence experiment for laser subjects 2, 3, and 4 using the Direct Surface Distance metric. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

Laser Subject 1 **Laser Subject 3**

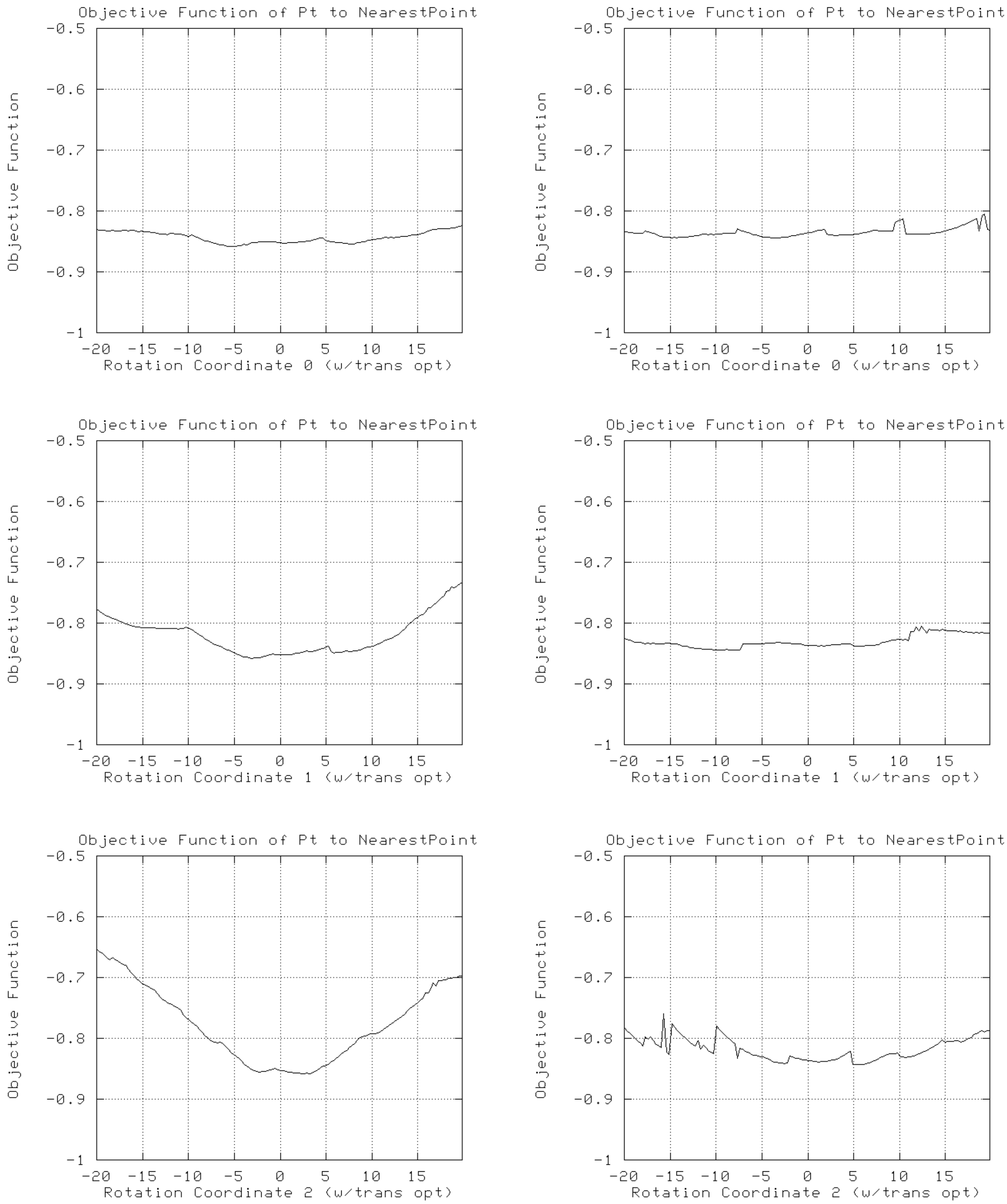


Figure 66 Objective function surface for laser subjects 1 and 3 using the Gaussian-Weighted Nearest Point metric. Surface cutouts along the three rotation angles are shown. X-axis is deviation from the correct pose.

Laser Subject 1 **Laser Subject 3**

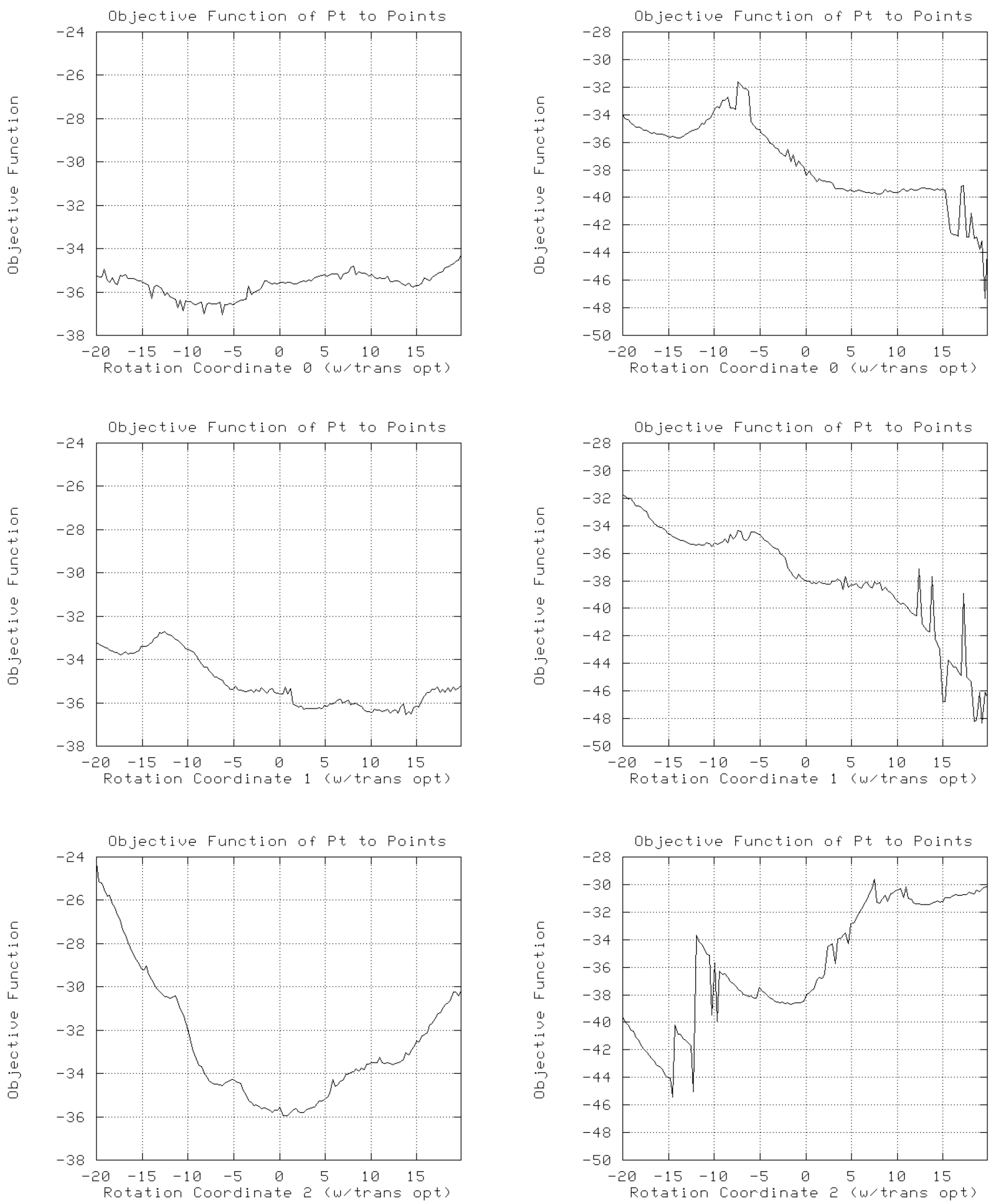


Figure 67 Objective function surface for laser subjects 1 and 3 using the Averaged Nearest Neighborhood metric. Surface cutouts along the three rotation angles are shown. X-axis is deviation from the correct pose.

Laser Subject 1 **Laser Subject 3**

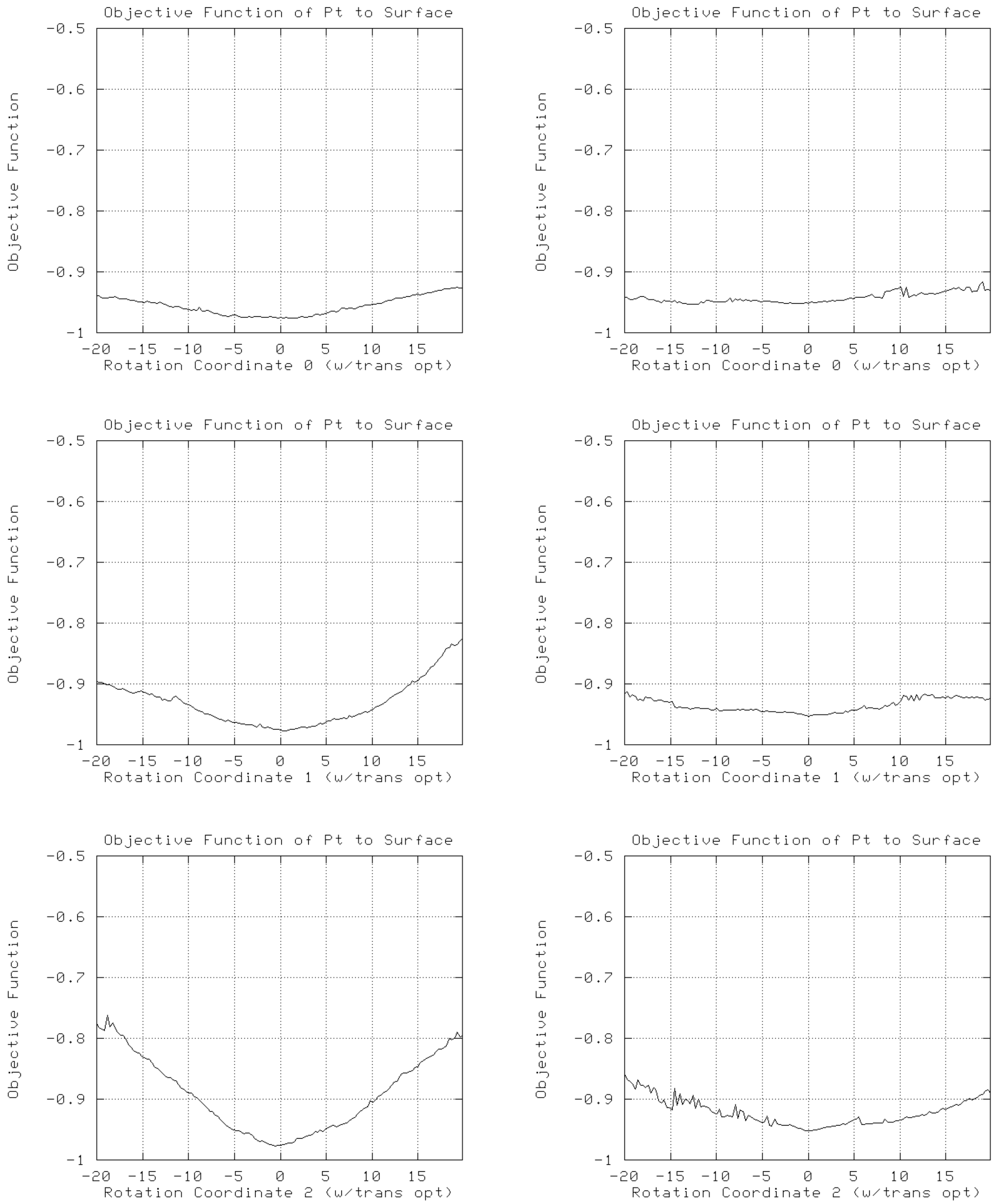


Figure 68 Objective function surface for laser subjects 1 and 3 using the Direct Surface Distance metric. Surface cutouts along the three rotation angles are shown. X-axis is deviation from the correct pose.

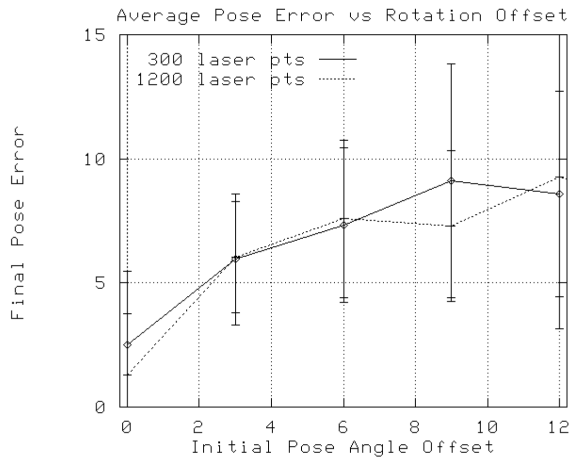


Figure 69 Average bounding box corner error for laser registration of subject 1 with 300 sampled laser points and 1200 laser points.

more concave and smooth than for subject 3. This is particularly true for the Direct Surface Distance metric. The Averaged Nearest Neighborhood metric appears to have a distinct global minimum, but it is at a different pose than for the other two metrics. It is now quite evident that the laser data from subject 3 does not provide enough discrimination across pose.

5.5.1.2 Laser Data Resolution

In Section 5.5.1.1 we saw the importance of the area of coverage of the laser scanner. We are also interested in knowing whether the resolution of the laser data, in terms of laser points per unit surface area, is critical. In this experiment we repeat the pose region of convergence experiment on laser subject 1, but with fewer laser points. Earlier we used all 1200 laser points for registration. In this experiment we randomly sample 300 points and repeat the same experiment as defined in Table 14. We overlay the results from the two registration runs within the region of convergence in order to compare their accuracy, as shown in Figure 69. We see that the results are almost identical for both cases, indicating that the laser data coverage is a much more critical factor than laser data resolution.

5.5.1.3 Augmenting the Laser Registration

In performing laser registration we are often limited in the skin surface we can scan by hair, surgical head clamps, or other surgical equipment. We therefore consider supplementing the laser data with a small number of skin points that we can probe with the Flashpoint 3D tracking system. In general we do not want to solely rely on such skin sampled points due to possible errors intro-

Table 15 Experiment specification for laser data supplementation experiment.

Experiment Specification Parameter	Value
Objective function variants	Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 {8, 5, 2} 2000 (fixed) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	300
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **20**$^{\circ}$ Increment per run: **0**$^{\circ}$
Translation magnitude offset of test surface	Initial: 0mm Increment per run: 0mm
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

duced in the manual skin probing and the potential long time it would take to carefully probe a large set of points around the head. But by combining the laser data with manually probed points we potentially benefit from the high accuracy of the laser data and the more complete surface coverage of the probed points. We explored this possibility by supplementing the subject 1 laser point set with a sequentially larger number of points randomly sampled from the MRI skin surface, the same one used as the registration reference surface. No measurement noise is introduced on these points. In practice, since we would expect higher spatial uncertainty on the manually probed points we could weigh the contributions of each registered test point by the inverse of its variance as is done in [56].

We ran this experiment with the parameters in Table 15. For each test run we add three randomly selected new skin points to the 300 sampled laser points that are used. The results of this experiment are shown in Figure 70. Since we start with a 20° pose offset our pose errors are relatively high without any additional sampled MRI skin points. But as we add skin points, our accuracy and its stability both increase. By about 40 additional MRI skin points we achieve accurate and stable results.

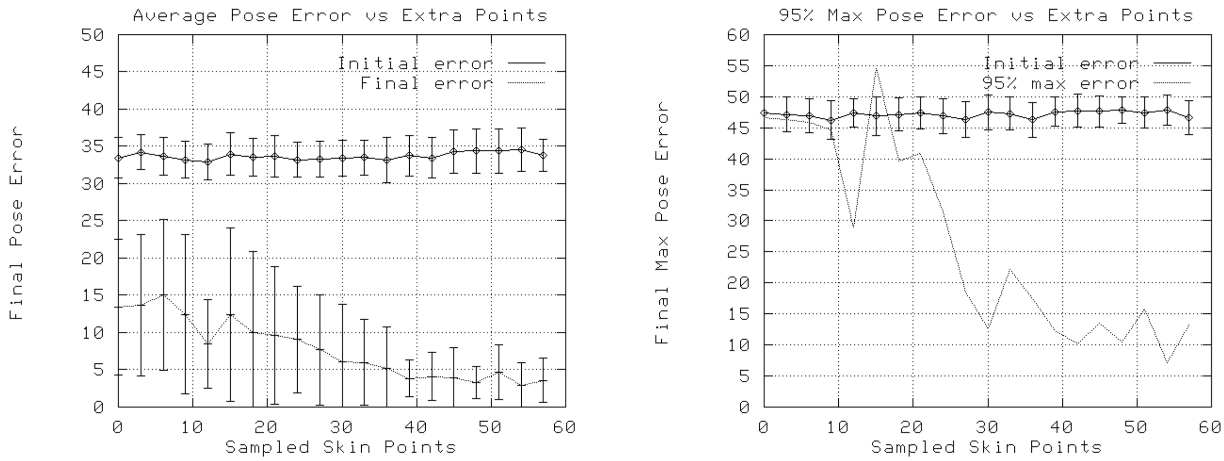


Figure 70 Results of laser data supplementation experiment on laser subject 1 using the Direct Surface Distance metric. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

5.5.2 MRI/MRI Registration

In Sections 5.3 and 5.4 we extensively tested the registration system on MRI head surface data in which the test dataset was a perturbed replica of the reference dataset. In this section we replicate the pose region of convergence experiment, from Section 5.3.4, but with the test data generated from another MR image taken of the same subject as the reference image. We also use a different surface as the basis for registration—the intra-cranial cavity (ICC) instead of the skin surface. The ICC is the outer surface of the brain and its surrounding membranes. The images we register are from the Multiple Sclerosis change detection application described in Section 4.2. The MR images consist of 54 axial slices with resolution of 0.9375mm x 0.9375 x 3.0mm.

We estimate the ground truth pose by running the registration algorithm using hierarchical distance weighting and the Direct Surface Distance metric and manually verifying the registration using our residual error display. This verification is shown in Figure 71. We see that the residual error is small for most points, except where the surface is tangential to the slice plane. Since these images had relatively thick slices, 3mm, we were more susceptible to segmentation errors at those points due to partial voluming and quantization artifacts. For comparison we show the initial configuration of the two ICC surfaces prior to any registration in Figure 72. We see that although the subject was placed in approximately the same pose in the MR scanner each time, there is significant pose error. This amount of initial pose error, though, is easily handled by our registration algorithm.

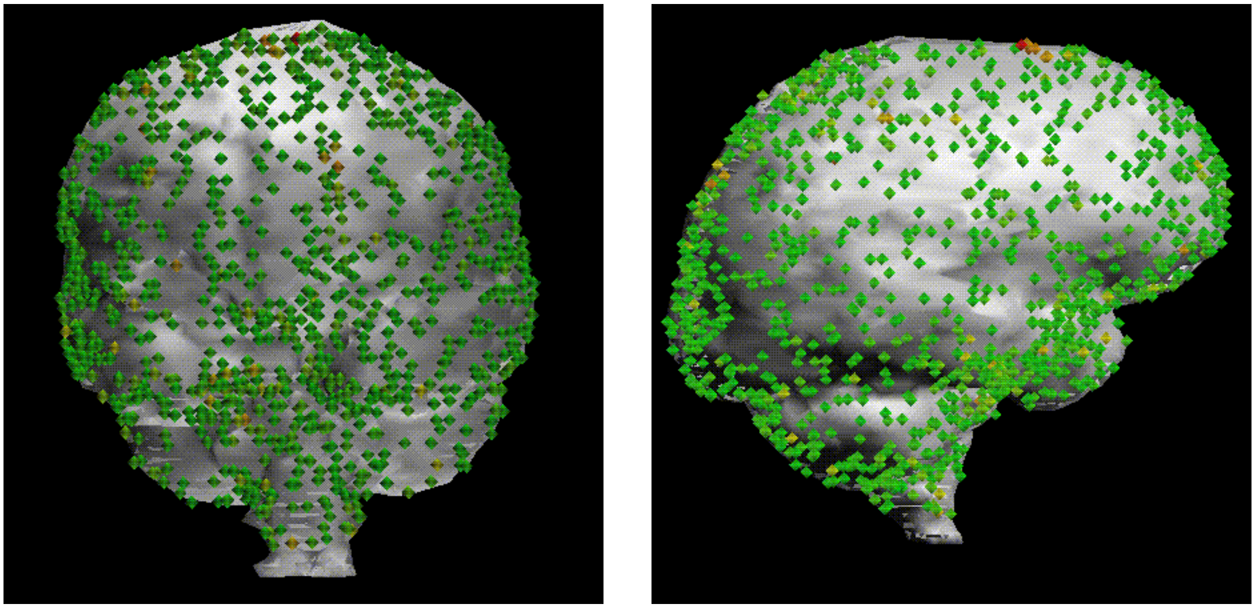


Figure 71 Best pose identified by the registration system for the ICC surfaces of the same subject from two MR images. Front and side views are shown. The reference ICC surface from one scan is shown as the white surface. The test ICC surface is shown as the set of colored points, where the color denotes residual error from the test point to the underlying reference ICC surface: green=0mm, yellow=2.5mm, red=5+ mm.

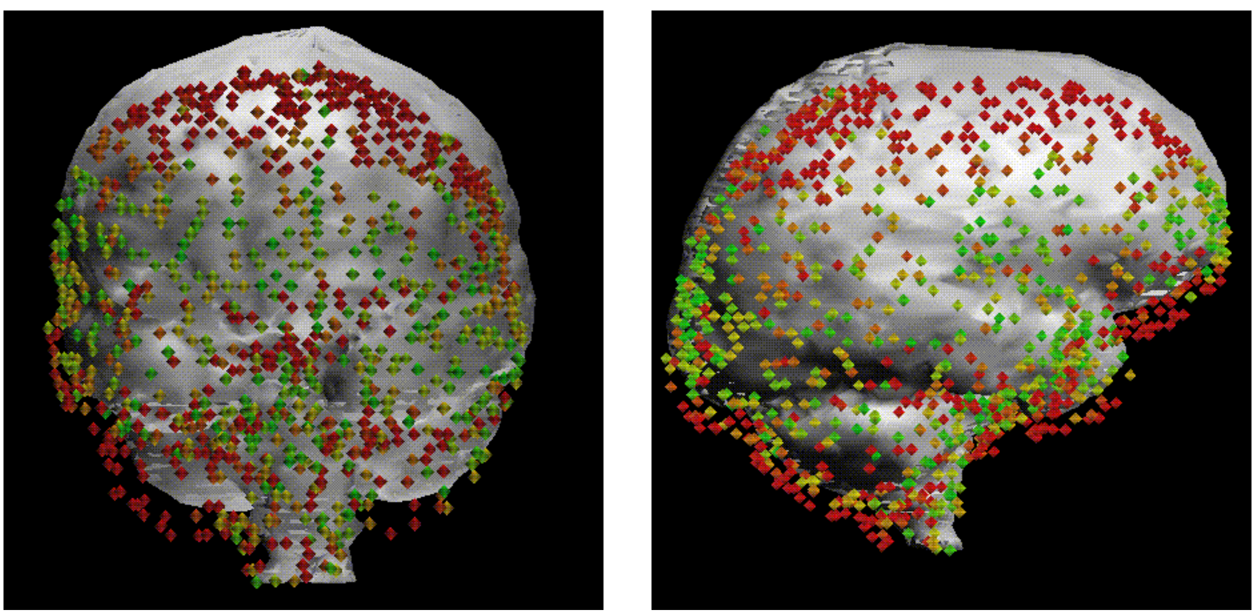


Figure 72 Initial pose of the two ICC surfaces to be registered. Front and side views are shown. The reference ICC surface from one scan is shown as the white surface. The test ICC surface is shown as the set of colored points, where the color denotes distance from the test point to the underlying reference ICC surface: green=0mm, yellow=2.5mm, red=5+ mm.

Table 16 Experiment specification for ICC pose region of convergence experiment.

Experiment Specification Parameter	Value
Objective function variants	Direct Surface Distance
Hierarchical refinement Distance weighting—standard deviation Surface sampling—total reference points Surface smoothing—max reference error	 **{8, 5, 2}** 5000 (unused) 0.5mm (fixed)
Transform complexity	Rigid
Number of test surface points	1000
Number of test runs	20
Number of simulations per run	30
Smoothing of test surface	Initial: 0.5mm Increment per run: 0mm
Rotation magnitude offset of test surface	Initial: **0°** Increment per run: **3°**
Translation magnitude offset of test surface	Initial: **0mm** Increment per run: **1mm**
Outliers added to test surface	Initial: 0% Increment per run: 0%
Removal of test surface patch	Initial: 0% Increment per run: 0%
Location perturbations of test surface points	Initial: 0mm Increment per run: 0mm

From the identified ground truth pose we then apply controlled pose offsets to measure the pose region of convergence. The experiment parameters are given in Table 16. The results of this experiment are shown in Figure 73. We observe that high accuracy is achieved, 1mm or less average bounding box corner error, through 45° with only a small number, outside the 95th percentile, of inaccurate solutions beyond that range.

5.5.3 MRI/MRA Registration

Another registration example is information fusion from images which each highlight different tissue classes. In this case we register a conventional MR image of the whole head with a high resolution MR angiogram which details the vasculature. The MRI contains 124 sagittal slices whose resolution is 0.9375mm x 0.9375mm x 1.5mm. It provides good soft tissue contrast and identifies pathologies such as tumors. The MR angiogram contains 124 axial slices whose resolution is 0.6645mm x 0.6645mm x 1.0mm and provides coverage of only part of the head, focusing on the region around the tumor.

We conduct an experiment in which we extract the skin from both the MRI and MR angiogram to use as the basis for registration. We estimate the ground truth pose as in the laser registra-

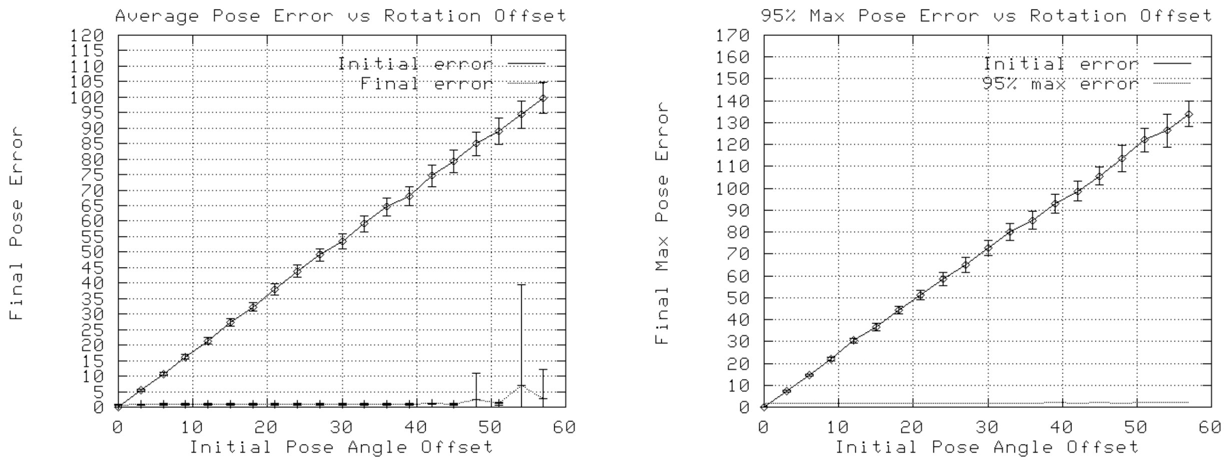


Figure 73 Results of ICC registration experiment using the Direct Surface Distance metric. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

tion experiment, Section 5.5.1, by initializing the registration at a point close to the expected solution, running the registration algorithm using hierarchical distance weighting and the Direct Surface Distance metric, and manually verifying the registration using our residual error display. This verification is shown in Figure 74. The partial coverage of the MR angiogram data over the MRI is clearly evident. For the ground truth registration we observe only small residual errors around most of the head, with the only region of poorer registration around the left ear. Due to the flexibility of the ear and the MR scanner receiver coil often impinging on the ear, those points are not reliably aligned.

We use the same experiment procedure as for the ICC registration using the parameters in Table 16. From the identified ground truth pose we apply controlled pose offsets to measure the pose region of convergence. The results of this experiment are shown in Figure 75. We observe that stable registration is achieved to at least 50° pose offset. The registration accuracy is not perfect, due in part to possible errors in specification of the ground truth. At the first test data point (3° rotation offset) our bounding box corner error measure is only 1mm. Through 51° offset, the error measure is around 3mm, but consistently with a small standard deviation of about 1mm. Beyond that a small error measure increase is observed, to 4mm, along with a 4mm standard deviation.

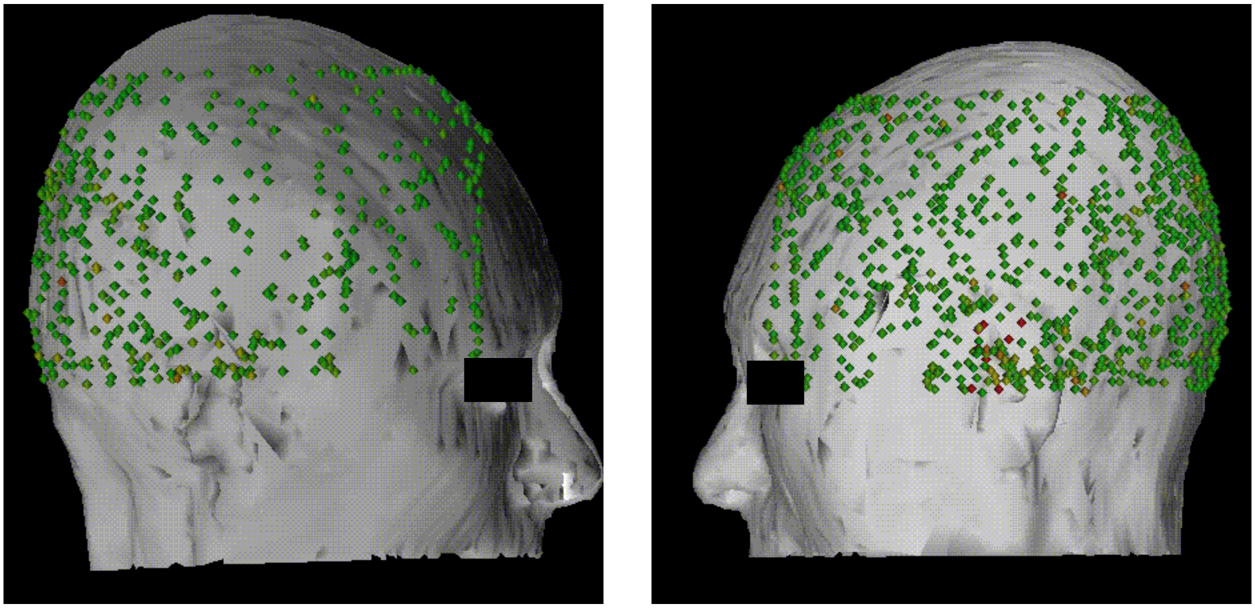


Figure 74 Best pose identified by the registration system for the MRI and MR angiogram of the same subject. Right and left views are shown. The MRI skin surface is shown as the white surface. The MR angiogram skin surface is shown as the set of colored points, where the color denotes residual error from the test point to the underlying reference ICC surface: green=0mm, yellow=2.5mm, red=5+ mm.

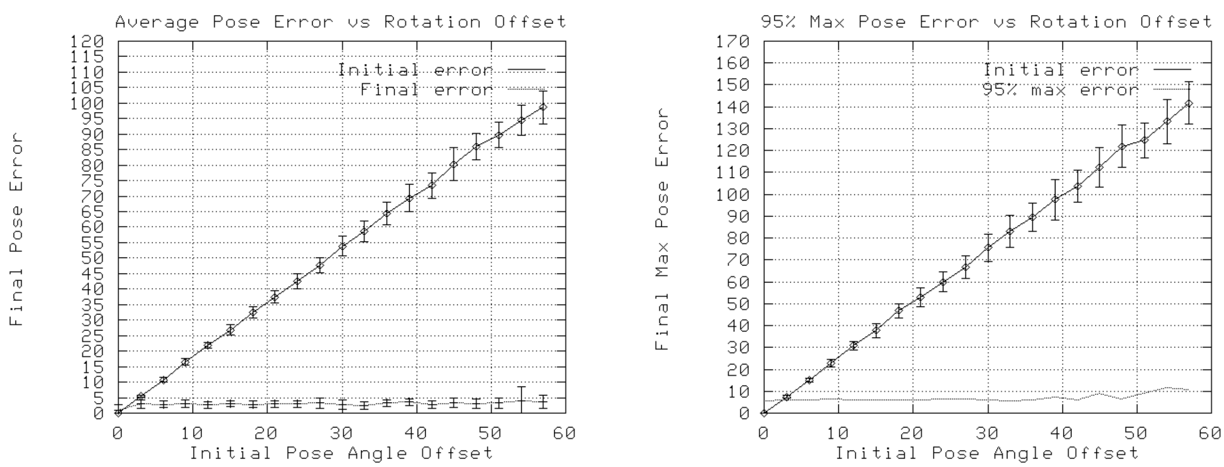


Figure 75 Results of MRI/MRA registration experiment using the Direct Surface Distance metric. Means and std dev of average pose error measure are shown on left and 95th percentiles of maximum pose error measure are shown on right.

5.5.4 Application Experimentation Conclusions

In this section we analyzed the performance of real application data. Given the nature of the data we had less control with these experiments, particularly in defining ground truth pose, but such testing is important to validate our algorithms and show their applicability to real-world problems. Our observations from these experiments are:

- Sufficient surface overlap is necessary to achieve accurate registration with large region of convergence. With about 25% surface coverage we achieved a 10° pose region of convergence. The distinctiveness of the covered surface is of course important as well.
- The Direct Surface Distance metric has the most well-behaved objective function surface.
- Using real application data instead of simulated test data appears to still uphold the accuracy and stability shown with the simulation experiments in Sections 5.3 and 5.4.
- Effective laser registration of small surface patches is likely to require surface data supplementation by such means as a 3D localization probe.
- Applications such as MS change detection or multi-modality fusion are effectively handled by surface registration techniques developed here.

5.6 Registration Evaluation Summary

In this chapter we described the extensive data and algorithm sensitivity experiments we performed on both simulated and actual imaging conditions. We return to the questions posed at the beginning of the chapter, providing some answers based on the experimentation results. In general, we deem accurate registration performance for average bounding box corner errors of less than 2mm and stable performance if the standard deviation of this measure is less than 2mm.

1. What is the sensitivity to incomplete surface overlap?

   For head registration in which our datasets are offset by about 15°, at least 40% surface coverage is necessary for accurate convergence. For smaller pose offsets we can achieve reliable performance with smaller surface coverages, as demonstrated with the laser data, but with less than about 25% head coverage we do not achieve stable convergence in all cases.

   The introduction of spurious data points has minimal impact on registration performance if they are randomly placed in the volume of interest using a uniform distribution. A 300% outlier rate, with no other artifacts, reduces accuracy by less than 2mm when starting from a 15° rotation offset. More biased outliers may of course have a larger impact if, for example, a large set of test data outliers align with the reference surface at the wrong pose.

2. What is the sensitivity to measurement errors?

   Increased measurement error yields gradually larger registration errors. With a 0-mean Gaussian distribution of measurement error, we achieve a head registration error of less than 1mm with measurement error standard deviation of up to 3mm. This condition is achieved with a 15° rotation offset when the test data is a duplicate of the reference data. Larger errors are expected for less ideal data conditions or for more biased measurement error, but overall performance is expected to degrade slowly with measurement error.

3. What is the sensitivity to data resolution?

   We measure data resolution as the maximum distance a point on the representative surface is estimated to be from the actual surface. For the 0.9375mm x 0.9375mm x 1.5mm head data we are using, a surface estimation accuracy of 0.5mm is necessary to achieve sub-millimeter registration accuracy. Such results are obtained with a 15° rotation offset when the test data is a duplicate of the reference data. Reduced surface data resolution leads to registration errors of up to 5mm for the same test conditions, but remain stable at that value. That is, this method of data resolution reduction leads to a pose bias.

4. What is the pose region of convergence?

   For head data in which the test data is a duplicate of the reference data we achieve accurate and stable results for up to a 75° rotation offset. Beyond that range we risk converging on the 180° wrong pose. For more realistic simulations in which we have some measurement noise and missing surface patches, the pose region of convergence is near 36°. For real data tests of comparable head registration our pose region of convergence is actually better, at about 45°.

5. What is the relationship of the objective function to registration accuracy?

   The behavior of the output objective function value is a good relative indicator of accuracy, but its absolute value depends on a number of application-specific factors, such as fraction of outliers and measurement noise. Visualization tools, such as residual error displays, are useful in complementing the objective function for making decisions on the adequacy of a registration result.

6. Which objective function performs best?

   The Direct Surface Distance metric yields the best overall results in terms of accuracy, stability, region of convergence, and its support for hierarchical refinement. The Gaussian-Weighted Nearest Point metric, though, achieves nearly similar results and is generally more efficient.

7. What is the benefit of hierarchical refinement?

   We achieve both improved accuracy and region of convergence by using distance weighting refinement—gradually decreasing the effective acceptable neighborhood size in which we may find matching reference surface patches for test points. When comparing such hierarchical registration against a single pass registration using the average hierarchical distance weighting, registration errors are reduced by up to 1mm and the pose region of convergence is increased by about 7° when using the Direct Surface Distance metric.

   Hierarchical smoothing of the surface data provides region of convergence improvements over non-hierarchical registration, but the performance benefits are smaller than with the distance weighting method.

   In addition, hierarchical refinement of pose complexity allows us to recover non-rigid deformations in the presence of translation, rotation, scaling, and shearing.

> Finally, complementing the registration process with a small number of random local pose perturbations and re-optimizations at each hierarchical processing loop improves avoidance of local minima if large steps in pose space have been taken.

Most of the operating points given in questions 1-4 above are for ideal situations in which only the given parameter is controlled with the others remaining at the best case condition. When multiple parameters are varied, as they are in general, our performance degrades, but the degradation is graceful. When we simultaneously introduced an increase in pose offset, a reduction in surface coverage, and an increase in measurement noise we achieved accurate and stable performance for up to a 36° rotation offset, surface patch removal of up to 33%, and spatial measurement noise of Gaussian $\sigma = 8.25$mm. Only selected combinations of rotation offsets, surface patch removal rates, and measurement noise were tested so this is only a small sample in the overall performance space of permutations of different test conditions. But this operating point provides us with reasonable guidelines for developing feasible registration applications.

Chapter 6

Conclusion

6.1 Summary

In this thesis we described the development of a robust hierarchical 3D surface registration algorithm and demonstrated its applicability to several high-utility medical applications. The main aspects of the registration algorithm are:

- **Same subject data.** The datasets being registered are of the same subject, but may be collected by different sensors and/or at different times.
- **Surface-based registration.** Surfaces, rather than volumes or features, are used as the basis for registration. Surfaces provide stable structures which are computationally attractive to register. They also support measurement of physical distance between two datasets which is the specific property we wish to minimize during registration. As a result, distance-based verification tools have been developed to view registration solutions.
- **Automated processing.** The complete registration process is accomplished automatically.
- **Hierarchical pose refinement.** We iteratively refine our registration in order to minimize dependence on initial alignments, avoid local minima, and achieve accurate registration in the presence of data variabilities. Two basic hierarchies are used:
    - **Measurement resolution**—coarse-to-fine representation of surfaces and the metrics measuring distances between them. Coarse metrics are designed to pull in solutions across large pose offsets while not adversely impacted by outliers, missing surface patches, or spatial perturbations. Once close to the final solution, fine metrics are designed to accurately localize the pose which minimizes a more sensitive distance measure taken between the surfaces.
    - **Pose complexity**—rigid to deformable transformation of data. We perform global and possibly local affine registration after obtaining a rigid alignment to account for tissue deformation and displacement.

A focus of the research was applying the registration algorithm to medical applications which may benefit from improved utilization of 3D imagery. We developed prototype applications which were tested on clinical imagery in the following domains:

- **Image-guided neurosurgery.** The system generates enhanced reality visualizations for showing the surgeon internal and hidden structures and tracks medical instruments to identify structures at the instruments' tips. Such a system is useful for both a priori surgical planning and dynamic guidance in the operating room. It has been successfully used in the operating room on over 20 patients.
- **Multiple sclerosis change detection.** The system registers to the same coordinate system MR images of the same subject taken at different times for identification of structural MS lesion changes. By automating the process, a large data throughput is achieved—over

1000 images were processed. The results allow localized analysis of changes to the possibly hundreds of lesions.

- **MR mammogram change detection.** The system solves for the optimal non-rigid mammogram registration in order to detect enhancing tissue after the injection of a contrast agent. Only limited testing has been performed on this application thus far.
- **Non-invasive brain mapping.** The system uses a transcranial magnetic stimulation device to actively map functional brain areas. The spatial localization, low cost, and ease of use of such a system are important advantages which lead to applications in surgical planning, surgical guidance, neuroanatomy research, diagnosis and therapy. We have mapped the motor cortex of ten subjects, including two neurosurgery patients, and the visual cortex of five subjects.

An important aspect of the research was the characterization of registration performance in terms of both data and algorithm parameters. Extensive experimentation evaluated registration accuracy, stability, and robustness across many types of inputs. We demonstrated effective performance for the following data variabilities:

- Large 3D initial pose offsets.
- Data outliers, introduced by the sensor or surface extraction algorithm.
- Differing data coverages, introduced by differing fields of view of the sensors.
- Measurement noise, introduced by sensor distortions or segmentation errors.
- Differing data resolutions.
- Non-rigid tissue deformation, introduced by physical structure motion between data collections.

Through our experiments we set performance limits across these data variabilities and characterized registration performance for meeting application requirements. We established the benefits of hierarchical registration refinement and determined that of the four objective functions implemented and tested, the Direct Surface Distance metric achieved the best overall performance.

6.2 Future Directions

We highlight several potentially useful direction in which to extend the work of this thesis. These directions are divided into algorithm development and application development areas.

- **Algorithm development:**
  - Integration of surface and volumetric registration.
    Surfaces provide a distance measure on the structures we wish to align, while volumes provide an appearance similarity measure which we wish to maximize. Combining the two measures may lead to a more powerful registration algorithm. A wide range of options exist for such combination: from keeping surface and volume registration sep-

arate with feedback between the two to integrating the two measures into a single evidence source. A Bayesian scheme may be used to combine surface and volume evidence since likelihood derived from each may be considered independent to a first approximation.

- Integration of segmentation and registration techniques.

  Surface registration relies on structure segmentation for extraction of the appropriate surface. Such segmentation, though, may be include some uncertainty which is a function of image resolution, image contrast, and partial voluming artifacts. Such uncertainty may be represented statistically, resulting in a Bayesian distance measure to capture spatial uncertainty, but we may also provide a feedback mechanism between the registration and segmentation process to update the segmentation and account for label uncertainty.

- Bayesian modeling of surface properties.

  We currently base the registration on the distance between two surfaces. We have commented (in Section 3.3.3) on the potential uncertainty of incorporating addition properties, such as curvature, into the objective function, primarily due to the noise sensitivity and high SNR requirements of these measures. If these uncertainties can be characterized, though, certain applications may benefit from their incorporation. A Bayesian modeling approach is a natural choice for capturing the uncertainty of these measures and for appropriately weighting their contribution. Building the Bayesian models, though, requires a thorough understanding of the impact of tissue structure, imaging parameters, and pre-processing operations on the surface characterization. The availability of application-specific image databases may spur such development.

- Result reliability estimates.

  We observed in Section 5.4.1.3 that accurate and stable registration is obtained for well-behaved objective functions which are characterized by concave surfaces. Conversely, non-concave objective functions lead to un-reliable results as shown in Section 5.5.1.1. By characterizing the shape of the objective function near the generated solution we may evaluate the reliability of the result. Simon [54] has developed some solutions to this particular problem.

- **Application development:**

  - Integration of physical models of tissue changes.

    We earlier indicated (in Section 3.4) that higher order transforms, beyond affine, may result in undesirable “over-fitting” of the data. But if constraints are available to bound the possible classes of deformations, such higher order transformation may lead to effective results. For example, Bro-Nielsen [11] has developed models for bone growth to generate predictive capabilities for a child’s jaw bone.

  - Integration of coarse anatomical knowledge for structural decomposition.

    In Section 4.3 we observed the benefits of decomposing a registration problem into independent components. In applications were such separate motion or articulation is possible, it may be beneficial to perform such decomposition. The constraints between the components, such as a rotation joint, must be modeled to achieve consistent results. The form of these constraints and their application must be specified. One possible solution for components which must remain continuous is to smooth the pose solution across the articulation point.

- Characterizing imagery and application to assess result adequacy from objective function value.

  In Section 5.3.5 we observed that if we know the outlier rate and measurement noise, the reported objective function score may be a good indicator of pose accuracy. By appropriately characterizing the application as well as the residual distance statistics at the optimal solution we may be able to derive such pose accuracy estimates.

- Quantitative application-level accuracy tests.

  While we applied our registration algorithm to several clinical applications, as described in Chapter 4, testing in the clinical setting has been mostly qualitative. Quantitative testing has focused on the registration function, not the complete end-to-end system. Quantitative testing through the use of ground truthed anatomical markers or realistic phantoms is necessary prior to fully operational use of these systems.

- Improved region of interest tracking.

  In the image-guided surgery application (Section 4.1), we developed mechanisms to track the patient in order to discount any relative motion between the patient and sensor. With these methods we attach tracking LEDs to the head clamp, the patient's skin, or possibly the patient's skull. But we are really interested in tracking the brain, which may move relative to all these reference frames. The craniotomy is generally too small to perform reliable surface registration, but such registration may at least solve for perpendicular motion to account for local brain swelling or shrinking. Alternatively, we may attempt to align the sulci/gyri pattern or blood vessel pattern of the brain through video views of the craniotomy. The sulci/gyri pattern, though, may not be sufficiently unique, and the blood vessels may be too small to reliably extract from MR angiograms. As a result, the use of internal imaging sensors, such as ultrasound, may be the most feasible solution.

References

[1] L. Adams, A. Knepper, D. Meyer-Ebrecht, R. Ruger, W. van der Brug. “An Optical Navigator for Brain Surgery”. *IEEE Computer*, **29**(1):48-54, Jan. 1996.

[2] P.R. Adby, M.A.H. Dempster. *Introduction to Optimization Methods*. Chapman and Hall, London, 1974.

[3] D.D. Adler, R.L. Wahl. “New Methods for Imaging the Breast: Techniques, Findings, and Potential”. *AJR*, **164**:19-30, January 1995.

[4] V.E. Amassian, R.Q. Cracco, P.J. Maccabee, J.B. Cracco, A. Rudell, L. Eberle. “Suppression of Visual Perception by Magnetic Coil Stimulation of Human Occipital Cortex”. *Electroencephalography and Clinical Neurophysiology* ,**74**: 458-462, 1989.

[5] A.T. Barker, L.L. Freeston, R. Jalinous, J.A. Jarratt. “Magnetic Stimulation of the Human Brain and Peripheral Nervous System: An Introduction and the Results of an Initial Clinical Evaluation”. *Neurosurgery* ,**20**(1): 100-109, 1987.

[6] G. Barnett, D. Kormos, C. Steiner. “Frameless Stereotaxy Using a Sonic Digitizing Wand: Development and Adaptation to the Picker ViStar Medical Imaging System”. R.J. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[7] R.H. Belmaker, A. Fleischmann. “Transcranial Magnetic Stimulation: A potential new frontier in Psychiatry”. *Biol. Psychiatry,* **38**:419-412, 1995.

[8] P. Besl, N. McKay. “A Method for Registration of 3D Shapes”. *IEEE PAMI*, **14**(2), February 1992.

[9] P. McL. Black, R. Kikinis, W. Wells, D. Altobelli, W. Lorensen, H. Cline and F. Jolesz. “A New Virtual Reality Technique for Tumor Localization”. *Program Book of the Annual Meeting of the Congress of Neurological Surgeons*, 1993.

[10] G. Borgefors. “Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm”. *IEEE PAMI*, **10**(6): 849-865, November 1988.

[11] M. Bro-Nielsen, C. Gramkow, S. Kreiborg. “Non-Rigid Image Registration Using Bone Growth Model”. *Proceedings First Joint CVRMED/MRCAS*, Grenoble France, March 1997.

[12] L.G. Brown. “A Survey of Image Registration Techniques”. *ACM Computing Surveys*, **24**(4):325-376, December 1992.

[13] R.D. Bucholtz, K.R. Smith. “A Comparison of Sonic Digitizers Versus Light Emitting Diode-Based Localization”. R.J. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[14] R.D. Bucholz, D.D. Yeh, J. Trobaugh, L.L. McDurmont, C.D. Sturm, C. Baumann, J.M. Henderson, A. Levy, P. Kessman. “The Correction of Stereotactic Inaccuracy Caused by

Brain Shift Using an Intraoperative Ultrasound Device". *Proceedings First Joint CVRMED/ MRCAS*, Grenoble France, March 1997.

[15] Z.H. Cho, J.P. Jones, M. Singh. *Foundations of Medical Imaging*.John Wiley & Sons, New York, 1993.

[16] G.E. Christensen, M.I. Miller, M.W. Vannier, U. Grenander. "Individualizing Neuroanatomical Atlases Using a Massively Parallel Computer". *IEEE Computer*, **29**(1): 32-38, January 1996.

[17] P. Cinquin et al. "Computer Assisted Medical Interventions". *IEEE EMB*, 254-263, May/ June 1995.

[18] H.E. Cline, W.E. Lorensen, R. Kikinis, F. Jolesz. "Three-Dimensional Segmentation of MR Images of the Head Using Probability and Connectivity". *Journal Computer Assisted Tomography*, **14**(6):1037-1045, 1990.

[19] A.C.F. Colchester et al. "Development and Preliminary Evaluation of VISLAN, a Surgical Planning and Guidance System Using Intra-Operative Video Imaging". *Medical Image Analysis*, **1**(1):73-90, March 1996.

[20] C.M. Epstein, D.G. Schwartzberg, K.R. Davey, D.B. Sudderth. "Localizing the Site of Magnetic Brain Stimulation in Humans". *Neurology* ,**40**: 666-670, 1990.

[21] G.J. Ettinger, W.E.L. Grimson, T. Lozano-Perez, W.M. Wells III, S.J. White, R. Kikinis. "Automatic Registration for Multiple Sclerosis Change Detection". *Proceedings IEEE Workshop on Biomedical Image Analysis*, Seattle WA, pp. 297-306, June 1994.

[22] G.J. Ettinger, M.E. Leventon, W.E.L. Grimson, R. Kikinis, V. Gugino, W. Cote, L. Sprung, L. Aglio, M. Shenton, G. Potts, E. Alexander. "Experimentation with a Transcranial Magnetic Stimulation System for Functional Brain Mapping". *Proceedings First Joint CVRMED/MRCAS*, Grenoble France, March 1997.

[23] J. Feldmar, N. Ayache. "Locally Affine Registration of Free-Form Surfaces". *Proceedings IEEE Computer Vision and Pattern Recognition*, Seattle WA, pp. 496-501, June 1994.

[24] R.L. Galloway. "Interactive Image Guided Neurosurgery". *IEEE Trans. Biomed. Eng.*, **39**(12):1226-1231, Dec. 1992.

[25] R. Galloway. "Stereotactic Frame Systems and Intraoperative Localization Devices". R. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[26] B. Geiger. *Three-dimensional Modeling of Human Organs and its Application to Diagnosis and Surgical Planning*. Ph.D. Thesis, Programme 4: Robotique, Image et Vision, INRIA, France, 1993.

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

[28] W.E.L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press, Cambridge, 1990.

[29] W.E.L. Grimson, G.J. Ettinger, S.J. White, P.L. Gleason, T. Lozano-Perez, W.M. Wells III, R. Kikinis. "Evaluating and Validating an Automated Registration System for Enhanced Reality Visualization in Surgery". *Proceedings First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, Nice France, pp. 3-12, April 1995.

[30] W.E.L. Grimson, G.J. Ettinger, S.J. White, T. Lozano-Perez, W.M. Wells III, and R. Kikinis. "An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization". *IEEE TMI*, **15**:(2), April 1996.

[31] A. Gueziec, N. Ayache. "New Developments on Geometric Hashing for Curve Matching". *Proceedings IEEE Computer Vision and Pattern Recognition*, 1993.

[32] S.E. Harms, D.P. Flamig, W.P. Evans, S.A. Harries, S. Bown. "MR Imaging of the Breast: Current Status and Future Potential". *AJR*, **163**:1039-1047, November 1994.

[33] M.P. Heilbrun, S. Koehler, P. McDonald, W. Peters, V. Sieminov, C. Wiker. "Implementation of a Machine Vision Method for Stereotactic Localization and Guidance". R.J. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[34] S.H. Heywang-Kobrunner. "Contrast-Enhanced Magnetic Resonance Imaging of the Breast". *Investigative Radiology*, **29**(1):94-104, January 1994.

[35] D.L.G. Hill, C.R. Maurer Jr, M.Y. Wang, R.J. Maciunas, J.A. Barwise, J.M. Fitzpatrick. "Estimation of Intra-Operative Brain Surface Movement". *Proceedings First Joint CVRMED/MRCAS*, Grenoble France, March 1997.

[36] B. Horn. "Closed-form Solution of Absolute Orientation Using Unit Quaternions". *JOSA* A, **4**:629-642, April 1987.

[37] H.G. Jacobson. "Fundamentals of Magnetic Resonance Imaging". *JAMA*, **258**(23):3417-3423, December 18, 1987.

[38] H. Jiang, R. Robb, K. Holton. "A New Approach to 3D Registration of Multimodality Medical Images by Surface Matching". *Visualization in Biomedical Computing*, 1992.

[39] T. Kapur, W.E.L. Grimson, R. Kikinis. "Segmentation of Brain Tissue from MR Images". *Proceedings First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, Nice France, pp. 429-433, April 1995.

[40] R.M. Kikinis, Shenton, F.A. Jolesz, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. Guttmann, R.W. McCarley, W. Lorensen, H. Cline. "Routine Quantitative Analysis of Brain and Cerebrospinal Fluid Spaces with MR Imaging".*Journal MRI*, **2**:619-629, 1992.

[41] M.E. Leventon. *A Registration, Tracking, and Visualization System for Image-Guided Surgery*. Master's Thesis, MIT Department of Electrical Engineering and Computer Science, 1997.

[42] W.J. Levy, V.E. Amassian, U.D. Schmid, C. Jungreis. “Mapping of Motor Cortex Gyral Sites Non-invasively by Transcranial Magnetic Stimulation in Normal Subjects and Patients”. *Magnetic Motor Stimulation: Basic Principles and Clinical Experience (EEG Suppl. 43)*, Editors: W.J. Levy, R.Q. Cracco, A.T. Barker, J. Rothwell. Elsevier Science Publishers, B.V., pp. 51-75, 1991.

[43] W.E. Lorensen, H.E. Cline. “Marching Cube: A High Resolution 3-D Surface Construction Algorithm”. *Computer Graphics*, **21**(3): 163-169, 1987.

[44] P.J. Maccabee, V.E. Amassian, R.Q. Cracco, J.B. Cracco, A.P. Rudell, L.P. Eberle, V. Zemon. “Magnetic Coil Stimulation of Human Visual Cortex: Studies of Perception”. *Magnetic Motor Stimulation: Basic Principles and Clinical Experience (EEG Suppl. 43)*, Editors: W.J. Levy, R.Q. Cracco, A.T. Barker, J. Rothwell. Elsevier Science Publishers, B.V., pp. 111-120, 1991.

[45] R. Maciunas, J. Fitzpatrick, R. Galloway, G. Allen. “Beyond Stereotaxy: Extreme Levels of Application Accuracy Are Provided by Implantable Fiducial Markers for Interactive Image-Guided Neurosurgery”. R.J. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[46] D. Metcalf, R. Kikinis, C. Guttmann, L. Vaina, F. Jolesz. “4D Connected Component Labelling Applied to Quantitative Analysis of MS Lesion Temporal Development”. *Proceedings IEEE EMBS*, October 1992.

[47] B.U. Meyer, R. Diehl, H. Steinmetz, T.C. Britton, R. Benecke. “Magnetic Stimuli Applied Over Motor and Visual Cortex: Influence of Coil Position and Field Polarity and Motor Responses, Phosphenes, and Eye Movements”. *Magnetic Motor Stimulation: Basic Principles and Clinical Experience (EEG Suppl. 43)*, Editors: W.J. Levy, R.Q. Cracco, A.T. Barker, J. Rothwell. Elsevier Science Publishers, B.V., pp. 121-134, 1991.

[48] A. Pascual-Leone, E. Gomez-Tortosa, J. Grafman, D. Alway, P. Nichelli, M. Hallett. “Induction of Visual Extinction by Rapid-Rate Transcranial Magnetic Stimulation of Parietal Lobe”. *Neurology*, **44**: 494-498, 1994.

[49] C. Pelizzari, G. Chen, D. Spelbring, R. Weichselbaum, C. Chen. “Accurate three-dimensional registration of CT, PET, and/or MR images of the brain”. Journal of Computer Assisted Tomography, **13**(1), 1989.

[50] T. Peters, B. Davey, P. Munger, R. Comeau, A. Evans, A. Olivier. “Three-Dimensional Multimodal Image-Guidance for Neurosurgery”. *IEEE TMI*, **15**(2):121-128, April 1996.

[51] W.H. Press, S.A. Teukolsky, S.T. Vetterling, B.P. Flannery. *Numerical Recipes in C, The Art of Scientific Computing, Second Edition*, Cambridge University Press, 1992.

[52] M.J. Ryan, R.K. Erickson, D.N. Levin, C.A. Pelizzari, R.L. Macdonald, G.J. Dohrmann, “Frameless Stereotaxy with Real-Time Tracking of Patient Head Movement and Retrospective Patient-Image Registration”. *J. Neurosurgery*, **85**:287-292, August 1996.

[53] J. Serra. *Image Analysis and Mathematical Morphology.* London, Academic Press, 1982.

[54] D.A. Simon. *Fast and Accurate Shape-Based Registration*, Ph.D. Thesis Carnegie Mellon University, Robotics, 1996.

[55] T.S. Sumanaweera, G.H. Glover, T.O. Binford, J.R. Adler. “MR Susceptibility Misregistration Correction”. *IEEE TMI*, **12**(2): 251-259, June 1993.

[56] R. Szeliski, S. Lavallee. “Matching 3D Anatomical Surfaces with Non-Rigid Deformations using Octree-Splines”. *Proceedings IEEE Workshop on Biomedical Image Analysis*, Seattle WA, pp. 144-153, June 1994.

[57] S. Tebo, D. Leopold, D. Long, S. Zinreich, D. Kennedy. “An Optical 3D Digitizer for Frameless Stereotactic Surgery”. *IEEE Comp. Graph. Appl.*, **16**(1):55-64, Jan. 1996.

[58] J.P. Thirion. “Extremal Points: Definition and Application to 3D Image Registration”. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Seattle WA, pp. 587-592, June 1994.

[59] P.A. van den Elsen, E.D. Pol, M.A. Viergever. “Medical Image Matching--A Review with Classification”. *IEEE Engineering in Medicine and Biology*, **12**(1):26-39, March 1993.

[60] E.M. Wassermann, L.M. McShane, M. Hallett, L.G. Cohen. “Noninvasive Mapping of Muscle Representations in Human Motor Cortex”. *Electroencephalography and Clinical Neurophysiology*, **85**: 1-8, 1992.

[61] E.M. Wassermann, B. Wang, T.A. Zeffiro, N. Sadato, A. Pascual-Leone, C. Toro, M. Hallett. “Locating the Motor Cortex on the MRI with Transcranial Magnetic Stimulation and PET”. *Neuroimage,* **3**: 1-9, 1996.

[62] E. Watanabe, “The Neuronavigator: A Potentiometer-Based Localizing Arm System”. R. Maciunas, ed., *Interactive Image-Guided Neurosurgery*, Amer. Assoc. Neur. Surg., 1993.

[63] W.M. Wells. *Statistical Object Recognition*, MIT AI Technical Report 1398, January 1993.

[64] W.M. Wells. “Adaptive Segmentation of MRI Data”. *IEEE TMI*, **15**(4):429-442, August 1996.

[65] Z. Zhang. “Iterative Point Matching for Registration of Free-form Curves and Surfaces”. *International Journal of Computer Vision*, **13**(2), 1994.