massachusetts institute of technology — artificial intelligence laboratory

2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images

Lilla Zollei

AI Technical Report 2002-001 August 2001

© 2002 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images

by

Lilla Zöllei

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

Masters in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2001

© Massachusetts Institute of Technology 2001. All rights reserved.

Certified by: W. Eric L. Grimson
Bernard Gordon Chair of Medical Engineering and
Professor of Computer Science and Engineering, MIT AI Lab
Thesis Supervisor

Certified by: William M. Wells III.
Research Scientist, MIT AI Lab
Assistant Professor of Radiology,
Harvard Medical School and Brigham and Womens' Hospital
Thesis Supervisor

Accepted by: Arthur C. Smith
Chairman, Departmental Committee on Graduate Students

2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images

by
Lilla Zöllei

Submitted to the Department of Electrical Engineering and Computer Science on August 10, 2001, in partial fulfillment of the requirements for the degree of Masters in Electrical Engineering and Computer Science

Abstract

The registration of pre-operative volumetric datasets to intra-operative two-dimensional images provides an improved way of verifying patient position and medical instrument location. In applications from orthopedics to neurosurgery, it has a great value in maintaining up-to-date information about changes due to intervention. We propose a mutual information-based registration algorithm to establish the proper alignment. For optimization purposes, we compare the performance of the non-gradient Powell method and two slightly different versions of a stochastic gradient ascent strategy: one using a sparsely sampled histogramming approach and the other Parzen windowing to carry out probability density approximation.

Our main contribution lies in adopting the stochastic approximation scheme successfully applied in 3D-3D registration problems to the 2D-3D scenario, which obviates the need for the generation of full DRRs at each iteration of pose optimization. This facilitates a considerable savings in computation expense. We also introduce a new probability density estimator for image intensities via sparse histogramming, derive gradient estimates for the density measures required by the maximization procedure and introduce the framework for a multiresolution strategy to the problem. Registration results are presented on fluoroscopy and CT datasets of a plastic pelvis and a real skull, and on a high-resolution CT-derived simulated dataset of a real skull, a plastic skull, a plastic pelvis and a plastic lumbar spine segment.

Thesis Supervisor: W. Eric L. Grimson
Title: Bernard Gordon Chair of Medical Engineering and Professor of Computer Science and Engineering, MIT AI Lab

Thesis Supervisor: William M. Wells III.
Title: Research Scientist, MIT AI Lab

Assistant Professor of Radiology,
Harvard Medical School and Brigham and Womens' Hospital

2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images

by
Lilla Zöllei

Submitted to the Department of Electrical Engineering and Computer Science on August 10, 2001, in partial fulfillment of the requirements for the degree of Masters in Electrical Engineering and Computer Science

Abstract

The registration of pre-operative volumetric datasets to intra-operative two-dimensional images provides an improved way of verifying patient position and medical instrument location. In applications from orthopedics to neurosurgery, it has a great value in maintaining up-to-date information about changes due to intervention. We propose a mutual information-based registration algorithm to establish the proper alignment. For optimization purposes, we compare the performance of the non-gradient Powell method and two slightly different versions of a stochastic gradient ascent strategy: one using a sparsely sampled histogramming approach and the other Parzen windowing to carry out probability density approximation.

Our main contribution lies in adopting the stochastic approximation scheme successfully applied in 3D-3D registration problems to the 2D-3D scenario, which obviates the need for the generation of full DRRs at each iteration of pose optimization. This facilitates a considerable savings in computation expense. We also introduce a new probability density estimator for image intensities via sparse histogramming, derive gradient estimates for the density measures required by the maximization procedure and introduce the framework for a multiresolution strategy to the problem. Registration results are presented on fluoroscopy and CT datasets of a plastic pelvis and a real skull, and on a high-resolution CT-derived simulated dataset of a real skull, a plastic skull, a plastic pelvis and a plastic lumbar spine segment.

Acknowledgments

First and foremost, I would like to say thank you to my thesis supervisors Prof. Eric Grimson and Prof. Sandy Wells. Both of them greatly supported me in achieving my goals throughout these two years and were there to talk to me whenever I had questions or doubts. Prof. Grimson, thank you for your knowledgeable advice regarding research issues, class work and summer employment. Sandy, thank you for being so patient with me and being open for a discussion almost any time. I learned a lot while working with you!

My special thanks go to my third (and unofficial) thesis supervisor, Eric Cosman, the author of my precious *Thesis Prep Talk*. I really appreciated all of our valuable conversations throughout the past year and thanks for keeping me inspired even through a nice and sunny summer. Notice, I managed not to forget how much *I prefer neon to sunlight!*

I sincerely appreciate all the help that I got from our collaborators at the SPL, the Brigham and from the ERC group. In specific, I would like to mention the people who helped me obtaining the majority of my 2D and 3D acquisitions: Ron Kikinis, Dr Alexander Norbash, Peter Ratiu, Russ Taylor, Tina Kapur and Branislav Jaramaz.

Thank you to all the people in the AI Lab for all your valuable suggestions and advice and special THANKS to those who took time to read through my paper and/or thesis drafts: Lauren, Raquel, Kinh and Dave. Lily, thanks for the Canny edge-detection code!

Tina, I would also like to express how greatly I appreciate your never-ending enthusiasm for research and the trust that you invested in me since the first day I got to MIT. I have truly enjoyed collaborating with you!

And at last but definitely not at least I would like to express my appreciation for the constant encouragement that came from my parents and my bother even if the former have been thousands of miles away... *Anyu, Apu és Pisti! Végtelenül köszönöm az önzetlen bizalmat és fáradhatatlan biztatást amit tőletek kaptam nap mint nap!*

This work was supported by the Whiteman Fellowship and the NSF ERC grant (JHU Agreement #8810-274).

Contents

1 Introduction **12**
- 1.1 2D-3D Registration . . . 12
- 1.2 Medical Applications . . . 14
  - 1.2.1 *3D Roadmapping* . . . 14
  - 1.2.2 Orthopedics . . . 15
- 1.3 Problem Statement . . . 16
- 1.4 Thesis Outline . . . 17

2 Background and Technical Issues **18**
- 2.1 2D-3D Rigid-Body Registration . . . 18
  - 2.1.1 Medical Image Modalities . . . 19
  - 2.1.2 Digitally Reconstructed Radiographs . . . 21
  - 2.1.3 Similarity Measures . . . 24
  - 2.1.4 Optimization . . . 26
  - 2.1.5 Number of Views . . . 27
  - 2.1.6 Transformation Representation . . . 27
  - 2.1.7 Transformation Parameterization . . . 29
  - 2.1.8 Other Notations . . . 31
- 2.2 Outline of Our Registration Approach . . . 31
- 2.3 Summary . . . 33

3 The Registration Algorithm **34**
- 3.1 The Transformation Parameter . . . 34
- 3.2 The Objective Function . . . 36
  - 3.2.1 Definition of MI . . . 36
  - 3.2.2 MI in the Registration Problem . . . 37
- 3.3 Probability Density Estimation . . . 38
  - 3.3.1 Parzen Windowing . . . 39
  - 3.3.2 Histogramming . . . 39
- 3.4 The Optimization Procedures . . . 40

3.4.1 Powell's Method 40
3.4.2 Gradient Ascent Strategy 41
3.4.3 Defining the Update Terms 41
3.5 Gradient-based Update Calculations 42
3.5.1 Partial Derivatives of Density Estimators 43
3.5.2 Partial Derivatives of Volume Intensities *wrt* T . 46
3.6 Summary 49

4 Experimental Results **50**
4.1 Probing Experiments 50
4.2 Summary of the Registration Algorithm 53
4.2.1 Step 1: Preprocessing 53
4.2.2 Step 2: Initialization 54
4.2.3 Step 3: Optimization Loop 54
4.3 Registration Results 56
4.3.1 Registration Error Evaluation 56
4.3.2 Objective Function Evaluation 58
4.4 CT-DRR Experiments 58
4.4.1 CT-DRR Registration 58
4.4.2 Multiresolution Approach 61
4.4.3 Robustness, Size of Attraction Basin 64
4.4.4 Accuracy Testing 65
4.4.5 Convergence Pattern 69
4.4.6 Registration Parameter Settings 71
4.5 CT-Fluoroscopy Experiments 72
4.5.1 Experiments with X-Ray Images of Gage's Skull 74
4.5.2 Experiments with Fluoroscopy of the Phantom Pelvis 82
4.6 Summary 82

5 Concluding Remarks **86**
5.1 Summary 86
5.1.1 Controlled Experiments 87
5.1.2 CT - X-ray Registration 87
5.2 Future Research Questions and Ideas 88
5.2.1 Coupling Segmentation and Registration 88
5.2.2 View and Number of Fluoroscopic Acquisitions . 89
5.2.3 Defining Automatic Stopping Criterion for Gradient Optimization Protocols 90
5.2.4 Truncation/Limited Field of View 90
5.2.5 Distortion Effects & Dewarping 90
5.2.6 Histogram Characteristics 91

5.2.7 Code Optimization 91
5.2.8 Improving MI 91

6 APPENDIX **93**
6.1 Small Angle Rotation 93
6.2 The Story of Phineas Gage 94

List of Figures

2.1 Lateral and AP acquisitions of X-ray fluoroscopic images of the pelvis phantom. 20
2.2 Orthogonal slices of a head CT acquisition: axial, sagittal and coronal views 21
2.3 CT-derived DRR images produced by the ray-casting algorithm 22
2.4 CT-derived DRR images produced by the voxel-projection algorithm 24
2.5 The transformation parameter T which relates the coordinate frames of the imaging environment and the data volume; $T = D_c \circ R \circ D_d$. 28

4.1 Results of two probing experiments evaluating (a) mutual information and (b) pattern intensity on the skull dataset. Displacement range of $+/-20$ (mm) and rotational range of $\approx +/-45$ (deg) were specified. 52
4.2 Results of two probing experiments evaluating a cost function on (a) the original and (b) the downsampled and smoothed version of the same phantom pelvis dataset. Displacement range of $+/-20$ (mm) and rotational range of $\approx +/-45$ (deg) were specified. 53
4.3 Single-view simulated fluoroscopic images from the controlled experiments. 60
4.4 Registration results of a phantom pelvis controlled experiment with the *Reg-Pow* method: contours of registration results are overlaid on the observed DRR images 70
4.5 Sample output from a controlled set of *Reg-Hi* experiments. Dataset: plastic pelvis. Initial offsets: (a) $y = 20$ (mm) and (b) $\beta = 15$ (deg). Plots display the magnitude of displacement error, rotation angle and the MI estimate at each iteration. 71

4.6 Real X-ray fluoroscopy of the phantom pelvis and real X-ray images of Phineas Gage's skull 73
4.7 Error distribution based upon the results of 30 experiments with random initial offset on a given interval. Row 1 displays plots with respect to error terms d_e and r_e while row 2 demonstrates errors in D_d and R 76
4.8 Error distribution based upon the results of 30 experiments with random initial offset on a given interval. Row 1 displays plots with respect to error terms d_e and r_e while row 2 demonstrates errors in D_d and R 77
4.9 Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pz* method. 79
4.10 Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pz* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images 80
4.11 Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pow* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images 81
4.12 Registration results of an experiment on real X-ray fluoroscopy and CT of the phantom pelvis dataset using the *Reg-Pow* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images. 83
4.13 Registration results of an experiment on real X-ray fluoroscopy and CT of the phantom pelvis dataset using the *Reg-Hi* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images. 84

List of Tables

4.1 CT dataset specifications (for Plastic Pelvis, Real Skull, Gage's Skull, Plastic Skull, Real Head and Plastic Lumbar Spine); sm1: smoothed volume on hierarchy level 2; sm2: smoothed volume on hierarchy level 3; sm3: smoothed volume on hierarchy level 4 59
4.2 Computing resources – machine specifications. 61
4.3 Timing measurements to contrast registration running time on different hierarchical levels. 63
4.4 Controlled, registration accuracy tests using the *Reg-Hi* method; No hierarchy; 65
4.5 Registration results of methods *Reg-Pz*, *Reg-Hi* and *Reg-Pow* on controlled experiments of a phantom pelvis and a real skull 67
4.6 Timing results of registration methods *Reg-Pz*, *Reg-Hi* and *Reg-Pow* on controlled experiments of a phantom pelvis and a real skull 68
4.7 Experiment description for the X-ray fluoroscopy and CT registration experiments on the Gage skull dataset . 78
4.8 Error measurements for the X-ray fluoroscopy and CT registration experiments on the Gage skull dataset . . . 78

Chapter 1

Introduction

1.1 2D-3D Registration

Recently, there has been a growing number of medical experts who advocate a minimally invasive approach to surgery. Their aim is to reduce the physical stress applied to the human body due to medical treatment/procedures and also to reduce treatment costs, for example, by minimizing the size and number of incisions. Unfortunately, in comparison to open procedures, these approaches restrict the surgeon's view of the anatomy. This leads to an increasing need for advanced imaging techniques that would help them not only with diagnosis, but also with planning and guiding interventions.

Pre-operative images provide an excellent source of detail about the anatomy in question. The widely used three-dimensional image modalities such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) contain high resolution information about the imaged body part. Other imaging techniques such as Positron Emission Tomography (PET) and Functional MRI (fMRI) complement that knowledge with metabolic and functional information. All these datasets can greatly assist in establishing diagnosis and planning procedures preoperatively or evaluating an intervention post-operatively. The same set of images can be conveniently utilized in surgery as well. However, they have the drawback that they may not completely reflect the surgical situation, since they are static.

In some applications it is important to use intra-operative images to follow the changes caused by the procedure or to visualize the location of a tool. In the operating room or interventional suite, it is mostly 2D

images that are available to record details about the current anatomical state. X-ray, X-ray fluoroscopy and portal images are all good examples of image modalities used for this purpose. Two-dimensional acquisitions are often taken instead of volumetric datasets because of timing, radiation-related and technological arguments. First, acquiring several 3D volumetric images during a procedure takes too long to make it practical compared to 2D imaging. Second, the radiation dose to both the patient and the doctor is reduced if only image slices are recorded rather than all the projections needed to reconstruct a 3D volume. Third, by using only 2D images, it is sufficient to have simpler imaging equipment in the operating suites.

Unfortunately, 2D images lack significant information that is present in the 3D modalities. Hence, in order to relate the changes recorded by the 2D modalities to the detailed 3D model, medical experts need to fuse the information from the pre-operative and intra-operative images mentally, which can be a challenging task. Therefore, it is useful to find a way to both automate that procedure and to make it reliable.

The combination of pre-operative and intra-operative images conveys the most information if the components are properly aligned in space. To achieve this it is necessary to determine their relative position and orientation. The procedure that identifies a geometrical transformation that aligns two datasets, or in other words locates one of them in the coordinate system of the other, is called *registration*. There already exist several techniques that can perform this task either semi- or fully-automatically. Matching, for example, different types of MRI with each other or with CT datasets is routinely done at numerous medical institutions. Most of these applications operate on images of the same dimensionality, aligning inputs from either 2D or 3D. Could we, nevertheless, align images with different dimensionality and complement the information from high-resolution pre-operative datasets with the more up-to-date, intra-procedural images? To achieve this goal, not only would we have to account for the different representations of a particular anatomical structure in the multimodal inputs, but we would also need to process information represented in different spaces. Additionally, as the registration results are expected during the medical procedure, the computation time would also be constrained. In a nutshell, these are the main challenges that one needs to address when solving the 2D-3D registration task. In our work, we present a solution to these problems and discuss the performance behavior of our registration framework.

1.2 Medical Applications

In this section, we give some specific examples of medical applications that could benefit from a reliable (and efficient) solution to the 2D-3D registration problem. They belong to the field of Image Guided Surgery (IGS). Their main objective is to introduce highly accurate pre-operative information about the examined anatomy into the operating room (where normally only lower dimensional images can be acquired) and help the execution of interventions carefully planned prior to the procedure by fusing the more detailed pre-operative with the more current intra-operative images.

1.2.1 *3D Roadmapping*

There exist a number of serious illnesses which can treated by the use of catheters that are maneuvered into the blood vessels of the brain. These include aneurysms and arteriovenous malformations.

Traditionally, X-ray fluoroscopy has been widely used in these cranio-catheter procedures. There is a currently existing procedure called *2D roadmapping* in which doctors follow the path of a catheter in the patient's body with the help of dynamic intra-operative 2D imaging. The procedure takes place in a special fluoroscopy suite. Prior to the intervention, opaque contrast material is injected into the patient, and a 2D acquisition is obtained. The resulting image shows vessels with high contrast because of the injected contrast agents. This type of data is used pre-operatively for diagnosis and planning, and it is also often acquired at the beginning of a procedure to serve as a reference set during the procedure. When the contrast agent is no longer present in the blood, dynamic fluoro images are acquired to follow the changes due to the intervention and to record the most current state of the treated body part. These are then subtracted from the pre-operative static image. As a result the vessels (of high contrast in the pre-operative data) and the catheter (not present at all in the pre-operative data) are the only structures highlighted. Continuing this process allows the physician to obtain information about the actual location and the movement of the catheter.

The main disadvantage of this method lies in having only a static 2D reference image highlighting the vessels. It is not rare that cranio-catheter procedures take more than 5 hours. During such a long time it is difficult to prevent any patient movement. Misalignment between the pre-operative image and the intra-procedural ones is inevitable. When that happens re-injection of the contrast agent is necessary for obtain-

ing another static reference image and the intervention is halted.

In the future, the drawbacks of the *2D roadmapping* method might be overcome by using a 3D dataset as the reference from which synthetic 2D images can be generated as needed[1].

Prior to the surgery, when the initial dose of contrast agent is injected, it requires a 3D volume rather than 2D images to be taken. During the procedure, when the dynamic fluoro images are obtained, they are compared to simulated projection images created from the 3D dataset. In this way, if the patient moves, it is only the parameters that describe the patient position and orientation in the imaging model that have to be modified in order to have the simulated and intra-procedural images line up again. These parameters are the ones that a 2D-3D registration algorithm would compute.

1.2.2 Orthopedics

Metastatic Bone Cancer

Another application is related to an orthopedics procedure, the treatment of metastatic cancer in the bones. The task here is to remove localized lesions from particular locations of the bones. Again, the treatment plan can be thoroughly designed prior to the operation using 3D CT volumes with high information content, but during the intervention, guidance and verification is only practical by making use of intra-operative images. Utilizing both of the two data sources requires the alignment of the intra-operative and pre-operative datasets.

Total Hip Replacement

Hip joint replacement surgery has several uses for 2D-3D registration. One is implanting an acetabular cup into the pelvic bone during total hip replacement procedures. In order to verify the correct position and orientation of the metal cup before the operation terminates 2D images are acquired. These need to be related to the 3D model of the anatomy. Another use concerns cases in *revision* surgery. Such a procedure is necessary if, following a total hip replacement procedure, the acetabular cup gets mislocated or gets deattached from the pelvis.

These orthopedics applications are currently pursued by the HipNav project at CMU and researchers at Johns Hopkins University.

[1] This project has been jointly proposed by Alexander M. Norbash, MD (Department of Radiology, Brigham and Women's Hospital and Prof. William Wells (Artificial Intelligence Laboratory, MIT).

Spine Procedures

Spine procedures are another very large application area for IGS, since back problems are very common, and the potential complications of damage to the spinal cord are devastating. Planning may effectively use pre-operative CT, while the interventions may be most practically guided by the use of C-arm X-ray equipment. One example procedure is *vertebroplasty*, which is the reinforcement of a failing vertebra by the placement of cement. Other applications include the placement of pedicle screws as components of stabilization hardware.

1.3 Problem Statement

The goal of the project described in this document is to register pre-operative volumetric data to intra-procedural 2D images. We are particularly interested in examining the problem of aligning 3D CT volumes to corresponding X-ray fluoroscopy. As a single 2D image, in practice, does not convey sufficient information about the spatial location of the imaged object, we require two projection images to achieve our task. We assume that the two imaging views are related by a known transformation, hence it is necessary to recover the required transformation with respect to only one of them. (This is a realistic assumption as biplanar images are often taken by rotating the imaging source by a pre-specified angle around one of the imaging axis. Also, biplanar acquisitions are considered to be standards in cranio-catheter applications.)

In solving the proposed problem, our main challenges lie in identifying a similarity measure, or objective function, that can quantify the quality of the alignment between the images and defining a procedure to modify and refine current estimates of the problem parameters in a way that the similarity score is optimized.

An additional primary focus of this effort is finding 2D-3D alignment methods which have computational complexity that is compatible with the time constraints implied by the interventional applications.

Experimentally, we aim to demonstrate the performance characteristics of our registration algorithm on a wide variety of datasets. The collection includes fluoroscopy and CT datasets of a plastic pelvis and a real skull and also a high-resolution CT-derived dataset of a real and plastic skull, a plastic pelvis and a plastic lumbar spine segment.

1.4 Thesis Outline

In Chapter 2, we introduce the problem of 2D-3D registration in a more thorough manner. We present the technical difficulties involved in the analysis and comparison of the multimodal and multidimensional datasets. We then summarize a handful of approaches that have already presented promising results in this area. We also introduce some frequently-used medical image modalities, describe some objective functions and some fast methods that simulate X-ray generation; which is a subtask of some registration methods. In Chapter 3, we focus on the computational details of our own approach. We describe the particular choices made when designing the components of our algorithm, we demonstrate the data structures used to encode the transformation variables and provide an in-depth derivation of the most important formulas used in the implementation. In Chapter 4, registration experiments are described using both synthetic and real datasets as well as detailed analysis of their results. The thesis concludes with Chapter 5, which summarizes the project and our contributions. Finally we describe some related future research ideas that we would like to investigate. In the Appendix, the reader may find a precise derivation of a particular mathematical formula and also a summary of the fascinating case of Phineas Gage, whose skull was used in our experiments.

Chapter 2

Background and Technical Issues

Introduction

In this chapter, we give a general introduction to the 2D-3D rigid-body registration problem applied specifically to medical modalities. We present a concise summary of research studies that have been applied to the problem while outlining a highly selective set of objective functions, optimization procedures and medical image modalities that are most frequently used in medical image processing. We also describe a fast technique that produces simulated projection images, called digitally reconstructed radiographs, as this technique was crucial in speeding up and monitoring our registration procedure. Then we introduce a new approach that we used to address the 2D-3D registration task.

2.1 2D-3D Rigid-Body Registration

Registering pre-operative datasets to images acquired intra-operatively can provide up-to-date information at the treatment site, guiding surgery or other interventions. When using different image modalities, information invisible in one of them can be incorporated into the other. Three-dimensional intra-procedural image acquisition is uncommon - typically only two-dimensional datasets can be obtained for such purposes. Although these images lack the spatial detail of volumetric data, they have the advantages of faster acquisition time and potentially reduced amount of radiation exposure to both patients and doctors. Ide-

ally, one can recover the advantages of the volumetric data by aligning the intra-operative 2D images with pre-operative volumes. However, not only do we have to focus on solving the multi-dimensional registration problem, but the algorithm running time should also be kept reasonable. If the alignment results cannot be produced well within the time-limits of an intervention, the algorithm cannot be used.

The majority of the medical applications for the proposed kind of registration task has emerged in the field of radiology. Alignment information is crucial in planning, guidance and treatment procedures. More specifically, the medical community has expressed interest in applying the 2D-3D alignment results in the following application areas: placement of pedicle screws in spine surgery [5, 6], aortic endoprostheses in transfemoral endovascular aneurysm management [7], verifying patient setup accuracy for radiotherapy and acetabular implant position in case of total hip replacement [1, 2, 11], displaying surgical instruments in the pre-operative CT volume [5], projecting important anatomical structures visible in CT onto 2D acquisitions and confirmation of depth electroencephalogram electrode position [33].

Our collaborators[1], in specific, are interested in applying the 2D-3D registration in the field of orthopedics and neuroradiology. Two of the major projects that of interest are head catheter tracking in case of cranio-catheter procedures and monitoring acetabular cup insertion during total hip replacement surgery. (A more detailed description of these and other procedures can be found in Chapter 1.) Therefore, the experimental dataset that we have acquired is mostly images of the skull and the pelvis.

2.1.1 Medical Image Modalities

The most commonly used 2D medical image modalities for the 2D-3D alignment task have been portal images and X-ray fluoroscopy (fluoro). Portal images are used in radiation treatment procedures. Their creation employs high-energy treatment radiation beams instead of low-energy imaging radiation, hence they could be considered byproducts of a procedure and their quality is extremely poor — they are of low resolution and they have low contrast. Research studies involving this modality use various segmentation techniques prior to or simultaneously with the registration procedure [1, 2, 30] in order to identify key structures in the portal images. Otherwise the individual intensity val-

[1] Alexander M. Norbash, MD (Department of Radiology, Brigham and Women's Hospital) and the Engineering Research Center (ERC) group including collaborators from CMU, Johns Hopkins University and MIT

ues have not been found to be sufficiently informative to describe the imaged anatomy.

Fluoroscopic images, on the other hand, reveal much more detail about the examined anatomy. They are taken by X-ray machines and are created by short wavelength energy. Fluoro images best visualize bony structures of the anatomy (Figure 2.1), as it is the bony tissues that absorb the most amount of radiation in the human body. The major disadvantage of this modality stems from the fact that without correction, its geometric accuracy degrades due to pincushion and radial distortion effects in current equipment. (Distortions of this sort are not a problem with the newer generation solid-state detectors.)

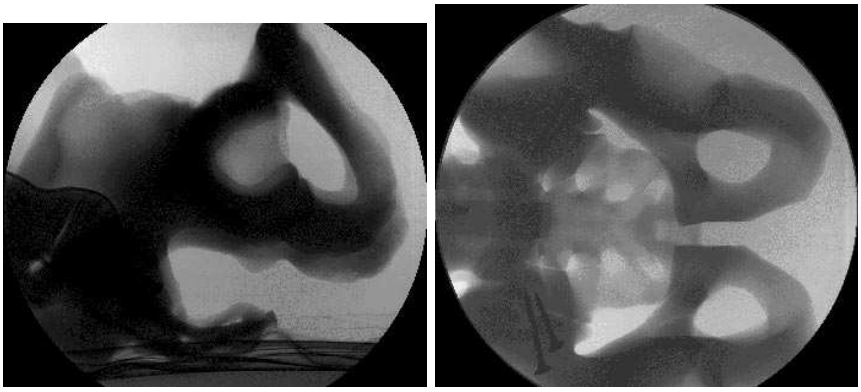


Figure 2.1: Lateral and AP acquisitions of X-ray fluoroscopic images of the pelvis phantom.

Among the 3D image modalities, Computed Tomography (CT) has been most widely considered for the registration task. CT images are created by assimilating multiple X-ray acquisitions. The X-ray machine rotates around the patient's body and at pre-specified angles shoots X-ray beams through the imaged object. The reconstructed images represent the absorption rate due to the intervening tissues called the Hounsfield number.

On the other end, the imaging plate records the absorption rate of different tissue types which quantities are referred to as Hounsfield numbers. The tomographic data acquisition is conventionally modeled by the Radon Transform and reconstructed according to the Filtered Backprojection algorithm. Distortion problems are usually not of major concern in case of this modality. Figure 2.2 shows three orthogonal slices of a real head CT acquisition.

Figure 2.2: Orthogonal slices of a head CT acquisition: axial, sagittal and coronal views

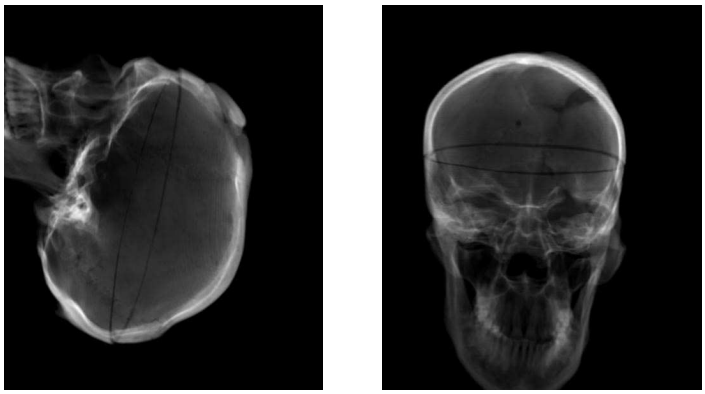
2.1.2 Digitally Reconstructed Radiographs

In our application, we focus on fusing CT and X-ray fluoroscopy images. One of the key challenges when attacking the 2D-3D registration problem is the need for an appropriate way to compare input images that are of different dimensionalities. The most common approach is to simulate one of the modalities given the other dataset and an estimate about their relative spatial relationship, so that the images can be compared in the same space. Then the transformation estimate can be updated to maximize an alignment score according to some similarity measure.

Reconstructing the 3D volume from 2D images is one alternative, but it requires numerous projection acquisitions and large computation time. It is more feasible to simulate 2D images from the 3D volume. Most existing applications follow this approach.

Ray-Casting

Simulated projection images, that are to model the production of X-ray acquisitions from volumetric CT are called Digitally Reconstructed Radiographs (DRRs). These images are traditionally formed by implementing the so-called *ray-casting* algorithm which we briefly summarize. Rays are first constructed between points of the imaging plane and the imaging source. Then the individual intensity values of the DRR images are computed by summing up the attenuation coefficients associated with each volume element (voxel) along a particular ray. An example of a DRR image created according to this algorithm is shown in Fig. 2.3.



(a) (b)

Figure 2.3: CT-derived DRR images produced by the ray-casting algorithm

Although producing high-quality results, this procedure can be quite inefficient for our purposes. As it must visit every voxel while computing the projection image, it tends to be extremely time-consuming. The creation of just one projection slice can take up to 100 seconds on a fast 1000 MHz machine. If we want to introduce a registration algorithm for interventional use, which task might require the creation of hundreds of DRRs as intermediate steps, we need to find alternative methods to approximate the 2D projections.

The speed limitations of the ray-casting algorithm are partly due to the size of the volumetric datasets. The majority of the CT volumes that we analyzed had dimensions of (512x512x200). (See a more detailed summary of the specifications of our datasets in Table 4.1 of Chapter 4). But the other part of the problem stems from the fact that if we closely follow the ray-casting algorithm, the data voxels are not accessed in an optimal way. As DRR-creation is a significant component of the registration application, several research studies have concentrated on defining more practical methods for their computation.

One way to address the problem of handling large volumes is to somehow restrict the size of the 3D datasets to be analyzed. In [3], the authors introduce a *focused* registration technique. The region of interest in the CT acquisition is segmented out prior to the intervention (e.g., the image of a vertebra) and the alignment algorithm is applied only with respect to that sub-entity. The same issue may also be effectively addressed by the application of a multiresolution approach,

where it is a downsampled and smoothed version of the input images that are first aligned[18, 15, 16]. (The hierarchical approach not only decreases the computational time, but also increases the robustness of the algorithm. A more detailed description of the hierarchical approach can be found in Chapter 4, where we present our experimental results.)

Voxel-Projection

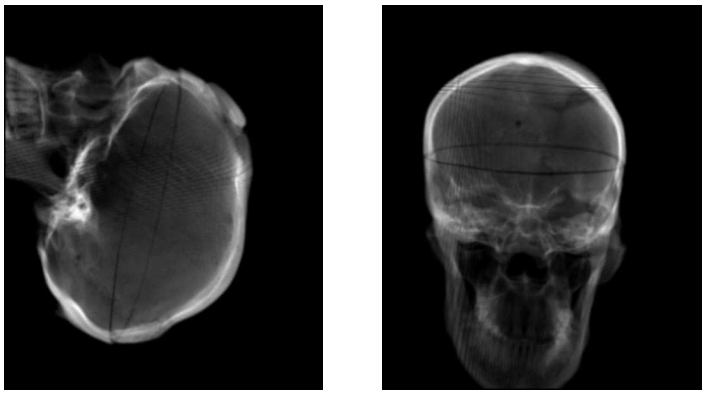
To approach the problem from an algorithmic development point of view, it is useful to invent new approximation methods for constructing the DRRs. One such procedure, which we used in our registration experiments, is called *voxel-projection* [14]. The main idea behind this new method is the attempt to maximally optimize memory accesses while processing the input datasets. Instead of carrying out the calculations following the layout of the DRR intensities to be determined in memory (and traversing the CT volume in a random manner), it accesses the volume elements in the order in which they are stored. First the algorithm estimates how much influence an individual volume element would contribute to elements of the DRR image. Then, after projecting the voxel centers onto the imaging plane, a smoothing function assures that the resulting image is not corrupted by banded intensities. That could happen due to lack of interpolation and due to ignoring the impact of a voxel on neighboring pixels. In our application, we achieved some improvement in the quality of the DRR images by increasing the minimal size of the smoothing kernel originally determined [14]. To compare the image quality of radiographs produced by the ray-casting method and the voxel-projection technique, compare Fig. 2.3 and Fig. 2.4, which display DRR images derived from the same CT volume with the two different algorithms.

Careful examination of Fig. 2.3 and Fig. 2.4 reveals that the two DRR-production algorithms result in images that are very similar. The main criticism against the outputs of the fast, voxel-projection technique could be that its images are not as smooth as that of the traditional procedure. Some intensity banding is visible on the more uniformly colored regions of its images.

The voxel-projection strategy has led to a speedup of factor 6, especially when relatively lower resolution projection images are sufficient.

Other DRR Techniques

Other approaches that also improve the computational burden of the ray-casting procedure include shear-warp factorization [32, 8] and the



(a) (b)

Figure 2.4: CT-derived DRR images produced by the voxel-projection algorithm

pre-computation of line integrals with the construction of a new data structure called Transgraph[2] [12]. The main idea behind the latter comes from the field of computer graphics, and is referred to as *view-based rendering*. It allows for fast computation of the DRR values and easy differentiation of the function generating them. Interpolating the densely sampled pre-computed line integrals proves to be more efficient than implementing the ray-casting technique. However, that strategy imposes a significant pre-computational/pre-processing step.

2.1.3 Similarity Measures

In many registration systems, the quality of alignment is scored by objective functions. Common registration methods can be grouped into two major categories based upon the nature of the similarity measure that they apply: they can be classified as feature- or intensity-based.

Feature-based Techniques

Feature-based approaches rely on the presence and identification of natural landmarks or fiducial markers in the input datasets in order to determine the best alignment. It is necessary to segment the most significant features in both of the input images and the matching criterion

[2]The name *Transgraph* is based on *Lumigraph* from the field of Computer Graphics.

is then optimized with respect to them. Contour- and point-based techniques [5, 6, 10, 41] are examples of this strategy, as well as registration methods that compare medialness properties of segmented anatomies [30]. Others carry out a minimax entropy strategy [1, 2] executing simultaneous registration and segmentation steps. Although the reduced number of features to be registered could provide great computational speedup (after the segmentation procedure is completed), major drawbacks of these methods lie in the need to carefully plan the image acquisition protocols in advance and the need for potentially re-scanning the patient if the diagnostic images do not contain the fiducials, the assumption that most of the fiducial markers can be located in all of the analyzed inputs, the inconvenience of planting artificial markers on the patient and the dependence on the segmentation procedure that can potentially introduce (additional) errors. These solutions might also require some level of user interaction, which generally is not desirable throughout medical procedures.

Intensity-based Measures

Intensity-based measures operate on the pixel or voxel intensities directly. They calculate various statistics using the raw intensity values of the inputs which are then compared in the images to be aligned. Though the number of points to be registered is much greater than in the case of the feature-based methods, no feature extraction step is required.

An extensive study of intensity-based similarity measures applied specifically to 2D-3D applications has evaluated the performance of six different objective functions in matching X-ray fluoroscopy and CT images [3]. The imaged organ was a phantom spine, and it was only a user-defined small region of interest (e.g., an individual vertebra) that was registered at a time. The objective functions considered by the authors were: normalized cross-correlation [33], entropy of the difference image [9], pattern intensity [6], mutual information [20, 15], gradient correlation [34, 33] and gradient difference [3]. After a careful registration study (using fiducial markers to ensure accuracy), the authors ranked these measures based upon their accuracy and robustness. They found that the best objective functions for the examined multi-modal registration task are pattern intensity and gradient difference. These measures proved to be the most robust with respect to the (simulated) presence of soft tissue and of a surgical instrument appearing only on one of the modalities. Both of these objective functions were implemented to use the whole input image in order to evaluate the current

quality of alignment.

The information theoretic measure of mutual information (MI) performed poorly in these experiments. It did not handle partial occlusions and truncations well and its performance further deteriorated when soft tissue was present. The study found two possible explanations for the failures of this similarity measure that has at the same time been very successful in the 3D-3D domain. First, MI is stated to require a large set of samples to obtain a good probability density estimate for the underlying entropy calculations. Although that is given in the 3D-3D registration problems, for the 2D-3D application that was not true. We say more about this aspect of their results later, in Chapter 3. Second, the authors claimed that as the search space of MI is much larger than what the problem requires, it is more difficult to recover the required parameters in it. (MI does not make the assumption that the two compared modalities are related via a linear function, it assumes a broader statistical relationship between the analyzed variables.)

Other intensity-based measures that have also been introduced for solving the CT-DRR registration task are absolute correlation coefficient [34], cross correlation and magnitude of scalar product of gradient [33] and a second order estimation to mutual information that aims to incorporate spatial information into its MI-measure [31]. The pattern intensity measure was also successfully applied in an MR-derived DRR and CT registration problem [14].

2.1.4 Optimization

Provided we have a suitable similarity function, the best alignment parameters can be located with the help of an optimization procedure. Such a protocol is responsible for modifying the current parameter estimates in a way that the similarity function eventually takes on its (local) extremum. In this work, we assume that the similarity measure is a reward and not a cost function. Hence the perfect/ideal alignment is assigned the highest score and an optimization procedure aims to maximize the objective function.

There are two major types of strategies that perform the maximization task: non-gradient and gradient methods. Non-gradient strategies execute a local search in the parameter space by evaluating the objective function at different locations according to a pattern, while gradient procedures use the gradient information to indicate the direction to the desired extremum. The former strategy might be easier to implement as it requires only the evaluation of the objective function and

no additional computations to derive the consecutive search directions. However, the latter could potentially be much faster as its search is specifically guided towards the extremum. Among the non-gradient methods, we found that the Powell method [34], the downhill simplex strategy [14] and an iterative optimization of individual transformation parameters (often called as "coordinate ascent" method) [5, 3] are the most popular. Among the gradient-based approaches, it is the Levenberg-Marquardt-type strategies [11, 29] and the hill-climbing (gradient ascent) approach [42, 15] that dominate.

2.1.5 Number of Views

In our experiments, examining only a single 2D image is not sufficient to robustly recover all registration parameters required to properly position the examined anatomy in the 3D world. While we can quite accurately recover in-plane rotation and displacement transformations, it is difficult to determine any out-of-plane transformations. In order to establish all of the transformation components with a desired level of certainty, it has proven advantageous to use two or more 2D acquisitions [2, 12, 14, 35] for the proposed alignment problem.

2.1.6 Transformation Representation

Our task when attacking the 2D-3D registration problem is to return a geometric transformation that best specifies the position and orientation of the examined anatomy at the time of obtaining the 2D projection images. In other words, we want to find a way to align the imaging and the world coordinate systems or to determine the correspondence between the intra-operative imaging environment and the coordinates of the pre-operative volumetric data (Fig. 2.5).

We focus on fusing CT and biplanar X-ray fluoroscopy images. In that specific case, the emphasis is on registering bony structures, since both modalities best visualize such information. Characterizing the rigid movement of bones implies six degrees of freedom. One 3D parameter specifies orientation, the other provides displacement information. No other transformation, such as shearing or scaling is allowed. If we also wished to align finer details, such as soft tissues, we would define higher-dimensional transformations.

Throughout this thesis, we denote the transformation that aligns the two coordinate systems by transformation T. In order to obtain a better intuition for what movement T represents, we decompose it into a collection of sub-transforms. When operating on data-points of

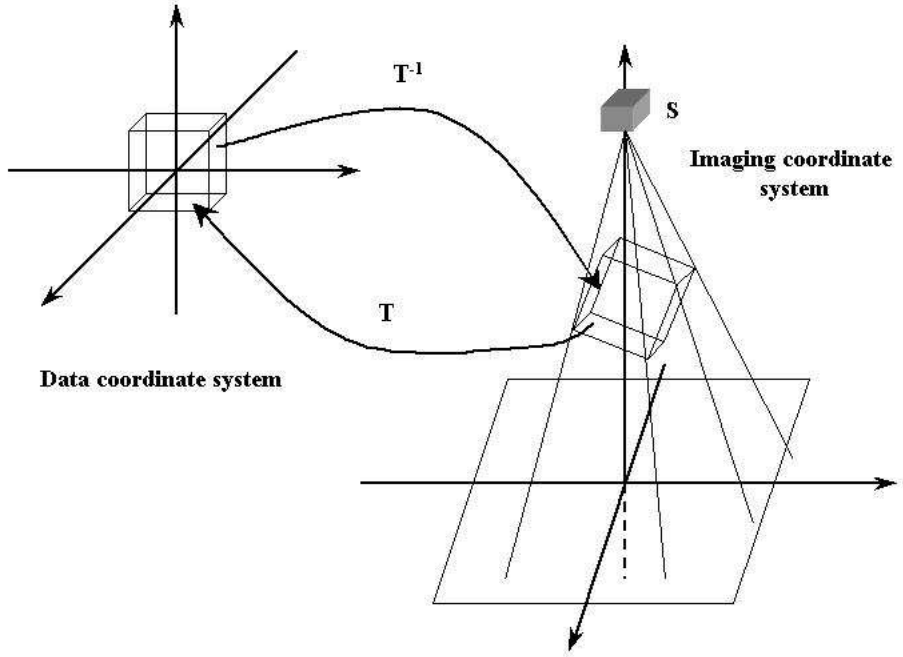


Figure 2.5: The transformation parameter T which relates the coordinate frames of the imaging environment and the data volume; $T = D_c \circ R \circ D_d$.

the 3D volume, it is most natural to have all rotations happen around the center of the volume. Hence, if the data is not centered in its own coordinate system, a displacement operator needs to be applied. This displacement operator is constant for a given registration task as it only depends on the specifications of the input volumetric dataset. Following the displacement, it is a rotational step and a translation in the oriented system that ensure the desired alignment. If we denote these operations by B_c, Q and B respectively (the underscore c notation emphasizes the fact that the associated variable refers to a constant), then a transformation G from data coordinates to the imaging environment could be composed as

$$G = B \circ Q \circ B_c.$$

As mentioned above, though, we are interested in computing the

inverse of this transform, G^{-1}, which converts image coordinates into data coordinates. Hence, we can write transformation T:

$$T = G^{-1} = B_c^{-1} \circ Q^{-1} \circ B^{-1}. \tag{2.1}$$

In order to simplify our notation, we introduce new transformation variables for the inverse operations

$$D_c \equiv B_c^{-1},\ R \equiv Q^{-1},\ \text{and}\ D_d \equiv B^{-1},$$

and thus modify the way we express T as:

$$T = G^{-1} = B_c^{-1} \circ Q^{-1} \circ B^{-1} = D_c \circ R \circ D_d. \tag{2.2}$$

The objective of the registration algorithm is to recover the non-constant components of T as accurately as possible. In Chapter 3, where we iteratively estimate the best parameters to provide the ideal alignment between the input images, the nature of the above decomposition plays an important role. (Note that we keep the same notation introduced for the decomposition of T throughout the rest of this document.)

2.1.7 Transformation Parameterization

For representing all six degrees of freedom of the rigid-body transformation, we use a new data structure. It is called *pose* and its name stems from the two notions that it describes: position and orientation. Given a pose parameter we can easily identify both its rotational and displacement components. As the rotational component is not linear in its parameters, the order of applying the transformation elements is essential; reversing them could produce a significantly different transformation. We use the usual convention of applying rotation first and then displacement. Therefore, if pose S were composed of rotational and displacement components (r, d), when applied to a coordinate point x, the resulting point could be written as

$$x' = S(r, d, x) = r(x) + d.$$

The composition of two pose transformations is not commutative. Given two poses $S_1(r_1, d_1)$ and $S_2(r_2, d_2)$, we have

$$\begin{aligned} S_3(r_3, d_3, x) &= S_2(r_2, d_2, S_1(r_1, d_1, x)) \\ &= S_2 \circ S_1(r_1, d_1, x) \\ &= r_2(r_1(x)) + r_2(d_1) + d_2, \\ \text{so} \quad & \quad r_3 = r_2 \circ r_1 \text{ and } d_3 = r_2(d_1) + d_2. \end{aligned}$$

That is to say, in the case of two consecutive transformations, the rotational elements are composed and the total displacement results from the rotated version of the first translation added to the second translation.

If the pose parameter only had a displacement component, we would write

$$x' = S(d, x) = x + d,$$

and if it only involved rotation, then the transformed point would become

$$x' = S(r, x) = r(x).$$

It is important to remember the above conventions, as in Chapter 3, when deriving the partial derivatives of the objective function with respect to the transformation parameters, we heavily rely on them.

There exists several ways to encode the transformation parameters that need to be recovered. The displacement part of T can be conveniently represented in a 3D vector format, however, the rotation parameter can be formulated in several different ways. Just to name a few of the options, we could use: roll-pitch-yaw; Z-Y-X Euler angles; Z-Y-Z Euler angles; equivalent angle-axis, orthonormal matrices and quaternions [23, 36]. We decided to represent our rotation operators as unit quaternions. This representation was appropriate for our needs as the quaternion encoding is easy to formulate and the composition of rotation operators (which occurs very frequently in our code) becomes a vector multiplication in that space. One way to define a quaternion is by a four-dimensional vector whose elements encode the rotational information as follows:

$$q = \left\{ cos\frac{\theta}{2}, \sin\frac{\theta}{2}\hat{\omega} \right\}. \tag{2.3}$$

In Definition (2.3), θ refers to the angle of rotation around the unit-length axis $\hat{\omega}$. Quaternions are appropriate measures if we want to define a metric on the space of rotations and they allow a uniform sampling of the rotation space [36].

We also use the equivalent angle-axis notation when illustrating the derivation of one of the update terms of the gradient ascent procedure in Chapter 3. In that case, if we represent the rotation transform with vector k, the magnitude of k determines the angle of rotation and its direction stands for the axis of rotation.

2.1.8 Other Notations

To follow the conventional notation in the medical imaging literature, we write U to denote the reference image and V to express the intensity values of the moving or floating images. In our case, U stands for the X-ray fluoroscopy acquisitions while V stands for the simulated radiographs. As the DRRs are constructed from the CT volume given a transformation estimate T, when we indicate the images that we compare, we use the notation $(U(x); V(T(x)))$ to explicitly emphasize that dependence.

2.2 Outline of Our Registration Approach

Goal

The aim of our study is the registration of biplanar 2D X-ray fluoroscopic images to a corresponding 3D CT dataset. The geometry of the imaging environment is assumed to be known, so the location of the two imaging sources for the 2D acquisitions is taken to be fixed. By updating our initial best estimate of the transformation components, we aim to make the CT-derived simulated projection images (DRRs) best approximate the observed fluoro acquisitions.

The Choice of Similarity Measure

Our choice of similarity measure depended on the examined image modalities, prior knowledge about features and possible distortions in the images to be registered, speed requirements (whether the registration needed to be completed in *real time* during a surgical intervention or the procedure was for treatment purposes and hence it could run for hours prior to or following the intervention) and implementation issues.

We decided to use the information theoretic notion, mutual information, to measure the quality of image alignment. While Penney et al. found the performance of pattern intensity to be superior to MI [3], we have chosen this particular objective function because of several reasons.

First, we have experienced robust performance and good accuracy in the past using MI, both in addressing the 3D-3D multi-modal rigid-body registration [15, 16] and the 2D-3D video-frame to model surface alignment [17].

Secondly, execution time played another critical factor in our decision. We did not intend to use any pre-segmentation techniques to reduce the size of the examined data volume to make the algorithm run faster. We made this choice partly because we wanted to eliminate user interaction from our procedure and partly because, even if desired, it could be quite difficult to segment out individual bone segments in the anatomies that we analyzed. For instance, in case of the pelvis, the ischium, ileum and sacrum are so uniformly and smoothly joined that it would be extremely difficult to distinguish clear boundaries between them. Also, in case of MI, it has been shown that it is possible to reliably maximize its value even without using all available intensity information provided by the inputs. We investigate a stochastic sampling approach, which was introduced in a 3D-3D multimodal registration problem [16]. The full input volume is considered in the registration task, but only a few randomly selected samples of it represent the dataset at each iteration. According to that scheme, we estimate probability distributions of image intensities by a sparse ray-casting method as opposed to by constructing full DRRs. It is not clear that pattern intensity could be implemented in this framework. That similarity measure is evaluated over the whole input image or at least on connected subregions of it. Hence, using pattern intensity in case of bigger datasets could become very computationally intensive and time-consuming.

Third, the generality of MI, the fact that it does not assume a linear relationship between the random variables being compared, allows for a potential reuse of the algorithm for image modalities other than the ones currently presented.

Maximization Strategy

In our study, to automatically locate the transformation that corresponds to the best alignment, we consider two optimization procedures: a stochastic gradient ascent procedure and the non-gradient Powell method. We preferred a gradient-guided search because of its computational efficiency, however, the Powell method was found to be extremely robust and was very helpful when designing experiments on the real X-ray datasets. The stochastic nature of the gradient-based optimization procedure is explained by using noisy approximations of partial derivatives instead of relying on true and accurate measures. The reason for applying such an estimate is to simplify computations, to speed up the overall registration process and to help escaping local

extrema of the similarity measure.

2.3 Summary

In this Chapter, we presented a high-level description of the 2D-3D registration problem and we provided some terminology and background information relevant to our proposed project. Additional details included specifics about medical image modalities, similarity functions, optimization techniques and about the transformation representation that we used to encode the searched pose parameters. We also gave a short summary of the motivation and the basic framework of the alignment approach that we investigated.

Chapter 3

The Registration Algorithm

Chapter Summary

In this Chapter, we give a detailed description of our registration procedure. First we remind the reader what transformation components we aim to recover as a result of our rigid-body registration algorithm. Then we introduce mutual information, the objective function we use, and describe its implementation details. We also compare two different optimization approaches, Powell's method and stochastic gradient ascent, which we have used to locate the extremum of the objective function. We derive in detail some of the update terms that are necessary for finding the desired alignment transformation. Lastly, we give a general overview of the registration algorithm. The description, results and performance evaluation of our experiments are presented in Chapter 4.

3.1 The Transformation Parameter

For the specific case of fusing CT and X-ray images, the primary focus is on registering bony structures, since both modalities best visualize such information. Characterizing the rigid movement of bones implies six degrees of freedom, three describing a rotational and three a displacement term. Our registration tool can also be thought of as a tool for aligning two different coordinate systems: that of the intra-operative imaging environment and that of the pre-operative image volume itself.

Transformation T is used to transform the imaging coordinates to their corresponding equivalent in world coordinates (Fig. 2.5).

As detailed in Chapter 2, T is a pose parameter. It is constructed from a rotational and a translational element. However, in order to distinguish constant and variable components of T, we decompose it into three individual sub-transforms. We write

$$T = D_c \circ R(r) \circ D_d(d). \tag{3.1}$$

In Eq. (3.1), D_c is a constant displacement term that is responsible for positioning the data volume into the center of its own coordinate system (so that rotation may be performed around its center). R encodes the rotational component required to perform the match, and translation D_d positions the object in the imaging coordinate system. As we specify T to be the transformation that expresses imaging coordinates in terms of data coordinates, the appropriate order of the sub-transforms is D_d followed by R and D_c. Decoupling the components of the transformation in such a way is useful because it makes the parameter space more directly searchable for the optimization procedures.

When we have access to multiple views of the same anatomy, we assume that the relationship between the various viewing sources is known. Hence, when we want to simulate projection images taken by other than the initial imaging source, we first apply a known, view-dependent transform to the coordinates and then apply the above introduced T. In the case of a biplanar application, where transformation N provides the relationship between the two imaging locations, we have $T_2 = T \circ N$. In more detail, we can write the expression of a point transformed by T and T_2 as

$$\begin{aligned} T(x) &= D_c \circ R \circ D_d(x) = D_c \circ R(r) \circ D_d(d, x) \\ &= D_c(R(r, x + d)) = D_c(r(x + d)) \end{aligned}$$

and

$$T_2(x) = D_c(r(N(x) + d)).$$

Given this formulation, it is only the variables R and D_d that we need to accurately recover. The rest of the components are known and constant; they are determined from a calibration procedure. D_c is purely dependent on the specifications of the input volume dataset and the imaging geometry is characterized by the non-varying transform N. Hence, when we investigate how the alignment quality changes with respect to infinitesimal changes in T (Section 3.4), we implicitly refer to modifications with respect to the operations R and D_d.

3.2 The Objective Function

We refer to measures that quantify the alignment quality of the input images as objective functions. From a broad range of candidates that have been used to assist in registration procedures, we decided to employ an information theoretic term called mutual information (MI). This similarity measure has quickly gained popularity in multi-modal medical image registration after it was first introduced [15, 20]. In case of 3D-3D rigid registration of head datasets, MI has proved to be a robust objective function, that could be applied with numerous image modalities.

Recently, there have been several extensions suggested to improve the general performance of MI. In most cases, it is gradient or some type of spatial information that is incorporated in the measure. One such example is the introduction of both the magnitude and the direction of the gradients into the mutual information formulation [18]. Although some robustness improvements can be demonstrated with these new methods, the altered objective functions did not preserve the information theoretical framework of the original formulation. We did not use these measures.

3.2.1 Definition of MI

In information theory, the mutual information of two discrete random variables expresses how much the knowledge about one of these variables increases the knowledge about the other. More informally, instead of assuming a linear relationship between the values of the random variables that are compared (as some of the widely used correlation functions do), it proposes that, in the registration problem, the intensity values from the corresponding images maximally explain each other if they are perfectly aligned. When mutual information is zero between two random variables, knowing one of them will convey no further information about the other and they are statistically independent. However, a non-zero mutual information term indicates that given one of the variables, the value of the other could be predicted with a given level of certainty.

There exists several definitions of mutual information. For example, according to the Kullback-Leibler distance interpretation, mutual information of two random variables, A and B, is defined to be the relative entropy between the joint probability distribution of the two variables and the product of their marginal distributions which would be the correct joint model if they were statistically independent. Thus

MI is a measure of the extent to which they are not statistically independent. (Note that the information theoretical notation, in the rest of this chapter adheres to the conventions of [43].)

$$\begin{aligned} I(A,B) &= D(p(A,B) \parallel p(A)p(B)) \\ &= E_{A,B}\left[\log \frac{p(A,B)}{p(A)p(B)}\right] \\ &= \sum_{a\in A}\sum_{b\in B} p(a,b)\log\frac{p(a,b)}{p(a)*p(b)}. \end{aligned}$$

In our computations, we will use another definition of MI. In order to introduce that formulation, we need to introduce another information theoretic term, entropy. The Shannon entropy of a random discrete variable A, $H(A)$, measures the uncertainty about that variable, or the amount of "randomness". It is formulated as the expected value of the negative log probability:

$$H(A) = E_A\left[-\log p(A)\right] = -\sum_{a\in A} p(a) * \log(p(a)). \tag{3.2}$$

Likewise, the joint entropy term between two random variables A and B is written as

$$\begin{aligned} H(A,B) &= E_{A,B}\left[-\log p(A,B)\right] \\ &= -\sum_{a\in A}\sum_{b\in B} p(a,b) * \log(p(a,b)). \end{aligned} \tag{3.3}$$

The formula that we apply for our registration calculations involves the sum of individual entropy terms less the joint entropy of the variables.

$$I(A,B) = H(A) + H(B) - H(A,B). \tag{3.4}$$

3.2.2 MI in the Registration Problem

In our biplanar registration procedure, we use two 2D projection images to guide the search for the best alignment parameters. Hence, we define our objective function g as the sum of mutual information terms,

$$g = I_1 + I_2,$$

where I_1 and I_2 stand for the mutual information quantities computed between the two observed fluoroscopy images and the CT-derived DRRs

that are to be registered. Hereafter, for sake of simplicity, when describing the computational details of the algorithm, we use only the first MI term of the sum, I_1, and refer to it as I (leaving the subscript off). All procedures, however, need to be carried out with respect to both of the image pairs.

Our 2D-3D registration strategy is based upon the comparison of the input X-ray fluoroscopic acquisitions to their simulated equivalents produced from the 3D volumetric dataset by applying the current estimate of the transformation parameter. These are treated as two discrete random variables whose mutual information needs to be evaluated. As noted in Chapter 2, we denote the observed 2D image by $U(X)$ and the transformation-dependent DRR by $V(T(X))$, where X is the set of sample points examined for comparison purposes. When writing our objective function with respect to these terms,

$$\begin{aligned} I(U(X), V(T(X))) &= \\ &= H(U(X)) + H(V(T(X))) - H(U(X), V(T(X))) \\ &= E_{U,V}[\log(p(U(x), V(T(x))))] - E_U[\log(p(U(x)))] - \\ &\qquad - E_V[\log(p(V(T(x))))]. \end{aligned} \tag{3.5}$$

3.3 Probability Density Estimation

One of the main challenges in evaluating the objective function expressed in Eq. (3.5) lies in accurately estimating the marginal and joint probability densities of the random variables. These quantities denote the probability distributions of the image intensities. We apply two different types of density estimators in our calculations. One of our approaches uses the non-parametric density estimator, called Parzen Windowing [38], and the other uses 1D and 2D histograms.

The available information presented by high resolution image volumes however is huge, and considering contributions from all pairs of corresponding image pixels is not always practical. It requires the generation of full DRR images for each iteration which creates an immense computational burden. Therefore, we experimented with both a dense and a sparse sampling approach. In the latter scenario, we base our probability estimates on a small number of random image intensity samples instead of using all the intensity values available from overlapping image regions. The smaller the sample size we use, the faster the estimations become. However, with each intensity value ignored we trade off accuracy (and possibly convergence of the algorithm). This

strategy provides noisy estimates at each individual step, however, if the samples are randomly selected and the estimation procedure occurs a sufficiently large amount of times, it can be shown that the estimates converge to the true values [26].

3.3.1 Parzen Windowing

Applying Parzen Windowing for probability density estimation is a standard technique in the computational community. With this method, the underlying probability density is estimated by the sum of symmetric kernels. The centers of the kernels are fit to individual sample points and most frequently, the kernel is defined to be Gaussian. This kernel choice significantly simplifies computations.

Given the Parzen Windowing formulation and a Gaussian kernel, we can write the probability density estimate of a random variable z as

$$p(z) = \frac{1}{n}\sum_{i=1}^{n} G_{\psi}(z - z_i), \tag{3.6}$$

$$\text{where} \qquad G_{\psi}(z) \equiv (2\pi)^{\frac{-n}{2}} |\psi|^{\frac{-n}{2}} \exp -\frac{1}{2} z^T \psi^{-1} z.$$

In Eq. (3.6), n signifies the number of points in the sample collection Z (where $\forall i \in N^+$ and $0 <= i < n,\ z_i \in Z$) based upon which our estimates rest, ψ indicates the covariance matrix and G stands for the Gaussian kernel.

3.3.2 Histogramming

As opposed to the continuous Parzen Windowing strategy, the histogramming approach uses a discrete approximation. Probability densities are calculated after the construction of 1D and 2D intensity histograms. Sample points from overlapping regions of the corresponding image pairs are used to fill the histograms and probability densities are estimated directly based upon those entries. *Dense histograms*, for which all the available corresponding intensity pairs are utilized in formulating these estimates have been widely used. Many use it for 3D-3D MI registration and one specific application that successfully applied this method is a 2D-3D registration algorithm aligning surface models to video [17]. We used dense histogramming when evaluating MI in the case of our Powell procedure.

Experimenting with sparsely sampled histograms is a new idea. It would be reasonable to expect that just a few random samples from a

large dataset cannot provide a valid approximation about the underlying probability density function of the image intensities. However, empirically, we found that in the case of the medical modalities that we have examined, the estimates can be useful and they can be used in the stochastic gradient ascent optimization procedure.

In general, building histograms from 100-150 random samples and using 32 intensity bins on the intensity range of 0-255 proved to be adequate in our stochastic optimization framework. Decreasing the bin sizes did not improve the registration results.

3.4 The Optimization Procedures

As a reminder, our task of finding the ideal transformation T is formulated as a maximization problem. We do an iterative search to locate the parameters that produce the highest score according to a reward function. At each iteration of the algorithm, we use the current estimate of transformation T to simulate the creation of the observed X-ray images by DRRs. We compute the quality of the alignment between these DRRs and the corresponding fluoro acquisitions. Then, to improve the matching score, we update the transformation estimate and start the registration loop over.

In order to identify in an efficient and/or reliable manner the set of updates, we need to select an optimization procedure. We compare the performance of a direction set and a gradient-based optimization strategy: Powell's method and the stochastic gradient ascent procedure [40].

3.4.1 Powell's Method

Powell's method is a direction set method. It optimizes the input function in a succession of one dimensional line maximization steps. Given an n dimensional search space, the function maximum could be located in just one pass of n line optimizations. That would, however, assume that linearly independent search directions are provided. In practice, it can be difficult to identify those. Hence, instead of aiming to work with mutually conjugate directions, a few *good* directions are selected that enable the localization of function extrema quickly.

The Powell procedure requires no calculation of the gradient. However, in order to evaluate the similarity measure in case of the individual line optimizations, full reconstruction of the DRRs is necessary. That

can easily cause a computational bottleneck in the algorithm. As we operate on huge datasets, applying the multiresolution approach was inevitable when using this strategy.

It has also been established that the Powell method is sensitive to the order in which the parameters are optimized during the line maximizations. One study concluded that the updates should happen in the order of in-plane followed by out-of-plane parameters [21]. We handle the translation components first and then the rotational elements.

3.4.2 Gradient Ascent Strategy

The gradient ascent technique is a maximization method whose local search for the optimal parameter settings is guided by calculations of the objective function's gradient. As opposed to the Powell method, whose search directions are either predetermined or continuously modified to approximate the state of being linearly independent, it explores the parameter space by making steps in the directions defined by the gradients. As a result, the objective function does not even need to be evaluated at each round, it is sufficient to only calculate the partial derivative terms. (Nevertheless, as explained in Chapter 4, in order to monitor the convergence behavior of the algorithm, we do compute the similarity measure at each step of the maximization phase.)

We use a probabilistic version of the gradient ascent procedure to find the local maximum of our objective function. The stochastic approximation [26, 27, 28] approach uses noisy estimates of the derivatives instead of the true ones in order to increase computational efficiency. The stochastic nature of our algorithm originates from two sources: we approximate the probability distributions of our variables by Parzen Windowing or sparse histogramming and we use various simplifications to compute the required derivatives. This approach has performed remarkably well in 3D-3D multi-modal medical image registration problems [15, 16].

3.4.3 Defining the Update Terms

In case of the Powell experiments, the optimization task is carried out almost as a black-box procedure. Mutual information is evaluated for each estimate of T and the transformation updates are calculated by the Brent line optimization method [40]. The optimization procedure finishes as soon as the gain from refining the estimate for T falls below a threshold tolerance measure.

When the gradient-based approach is used, we formulate a good estimate of the transformation update by computing the partial derivative of the objective function g with respect to the transformation parameters. We write:

$$T_{\text{updated}} = T_{\text{current}} + \lambda * \frac{\partial g}{\partial T_{\text{current}}}. \tag{3.7}$$

In Eq.(3.7), λ represents the learning rate (or step-size) of the algorithm; it constrains the maximal magnitude of individual update operations. Finding the appropriate range for λ forms a crucial part of the experiments. If its magnitude is too small, convergence might take a long time, however, if it is set to be too high, convergence to the searched optimum might not even occur.

One way to avoid fixing an ideal value for the learning rate is to vary it over time. This practice is called *annealing*, and it assigns decreasing values to λ as the iteration number increases. We eliminate the difficulty of selecting the ideal learning rate by using a hierarchical structure. Registration can be executed on several levels of resolution in order to make the algorithm run faster and to make it more robust. At the bottom of the pyramid, working with downsampled and smoothed versions of the input, we expect that it is easier to jump over local extrema and calculations can be executed in a smaller amount of time. At this level, the estimates might not be very accurate (they are indeed quite noisy), but that can be easily and swiftly refined on higher levels where smaller step sizes and more data samples can be used. As the resolution of the inputs increases, the transformation approximation can be made more precise. Details of this strategy with some experimental results are explained in Chapter 4.

We differentiate between the learning rates of rotational and displacement components. It is important to have both types of components contributing at the same rate to the overall transformation update. Further distinction could be made between components corresponding to in-plane and out-of-plane operations.

3.5 Gradient-based Update Calculations

As explained above, to improve our current transformation estimate according to the gradient ascent procedure, we require the computation of the partial derivative of our objective function with respect to the transformation parameters (Eq.(3.7)). Using Eq.(3.4) to express MI,

we want to compute

$$\frac{\partial I(U(X), V(T(X)))}{\partial T} = \\ = \frac{\partial H(U(X))}{\partial T} + \frac{\partial H(V(T(X)))}{\partial T} - \frac{\partial H(U(X), V(T(X)))}{\partial T}. \quad (3.8)$$

As the first term on the right of Eq.(3.8), the entropy of the observed image, does not depend on the transformation parameter, the equation can be simplified:

$$\frac{\partial I(U(X), V(T(X)))}{\partial T} = \frac{\partial H(V(T(X)))}{\partial T} - \frac{\partial H(U(X), V(T(X)))}{\partial T}. \quad (3.9)$$

The first approximation step in our algorithm results from the fact that we estimate statistical expectation terms with sample averages. In such a case, the entropy of a random variable A can be estimated as follows:

$$H(A) = E_A[-\log p(A)] \approx -\frac{1}{N} \sum_{a \in A} \log(p(a)), \quad (3.10)$$

where a is one of N observed samples drawn from sample set A.

Therefore, in the specific case of our registration problem, given M samples in our observation set X whose i^{th} sample point is x_i, we can write Eq. (3.9) as

$$\begin{aligned} \frac{\partial I(U(X), V(T(X)))}{\partial T} &= \quad (3.11) \\ &- \frac{1}{M} \sum_{i=1}^{M} \frac{\partial}{\partial T} \log(p(V(T(x_i)))) + \\ &+ \frac{1}{M} \sum_{i=1}^{M} \frac{\partial}{\partial T} \log(p(U(x_i), V(T(x_i)))). \end{aligned}$$

3.5.1 Partial Derivatives of Density Estimators

Parzen Windowing Approach

Given the definition of the Parzen Windowing probability density estimator in Def. (3.6) we can rewrite the entropy approximation in Eq.(3.10):

$$h(z) \approx -\frac{1}{N_A} \sum_{z_j \in A} \ln \frac{1}{N_B} \sum_{z_i \in B} G_\psi(z_j - z_i),$$

where B is another random variable, another set of sample points.

This expression is continuous, taking its partial derivative with respect to the transformation parameter produces

$$\frac{d}{dT}h(z(T)) \approx \tag{3.12}$$
$$\approx \frac{1}{N_A}\sum_{z_j\in A}\sum_{z_i\in B} W_z(z_j,z_i)(z_j-z_i)^T\psi^{-1}\frac{d}{dT}(z_j-z_i)$$
$$\text{where } W_z(z_j,z_i) \equiv \frac{G\psi(z_j-z_i)}{\sum_{z_k\in B} G_\psi(z_j-z_k)}. \tag{3.13}$$

Writing the partial derivative of mutual information from Eq. (3.9) and (3.11) then becomes:

$$\frac{dI}{dT} \approx \frac{1}{N_A}\sum_{x_j\in A}\sum_{x_i\in B}(v_j-v_i)^T K_{ij}\frac{d}{dT}(v_j-v_i), \tag{3.14}$$

where we use the following definitions

$$K_{ij} = \left[W_v(v_j,v_i){\psi_v}^{-1} - W_{uv}(w_j,w_i){\psi_{uv}}^{-1}\right],$$
$${\psi_{uv}}^{-1} = \text{DIAG}({\psi_{uu}}^{-1},{\psi_{vv}}^{-1}),$$
$$W_v(v_j,v_i) \equiv \frac{G_{\psi_v}(v_j-v_i)}{\sum_{x_k\in B} G_{\psi_v}(v_j-v_k)},$$
$$W_{uv}(w_j,w_i) \equiv \frac{G_{\psi_{uv}}(w_j-w_i)}{\sum_{x_k\in B} G_{\psi_{uv}}(w_j-w_k)}$$
$$\text{and } u_i \equiv U(x_i),\ v_i \equiv V(T(x_i)),\ w_i \equiv [u_i,v_i]^T.$$

This formulation of entropy manipulation and estimation is called EMMA[1] [42]. It provides an efficiently optimizable entropy measure, which is calculated from random samples of the available data points. Exhaustive sampling would be of quadratic cost in the sample size, hence only a few samples are selected. Although the less samples are used the more noise this approach introduces into the calculations, this allows it to effectively escape from local extrema. The convergence of this estimate to its true value was proved by Viola [42]. The EMMA estimate uses a Gaussian function for the Parzen kernel, but that could be replaced by any differentiable function.

The only unknown expression in Eq. (3.14) is the partial derivative of volume intensities with respect to the transformation components: $\frac{d}{dT}(v_i-v_j)$. It is computed in great detail in Section 3.5.2.

[1] The acronym stands for Empirical Entropy Manipulation and Analysis

Histogramming

For the histogramming approach we need to further manipulate Eq. (3.11). After some algebraic operations and expanding the partial derivatives,

$$\begin{aligned}
&\frac{\partial I(U(X), V(T(X)))}{\partial T} = \\
&= \tfrac{1}{M} \sum_{i=1}^{M} \tfrac{1}{p(u_i,v_i)} \tfrac{\partial p(u_i,v_i)}{\partial T} - \tfrac{1}{M} \sum_{i=1}^{M} \tfrac{1}{p(v_i)} \tfrac{\partial p(v_i)}{\partial T} \\
&= \tfrac{1}{M} \sum_{i=1}^{M} \left[\tfrac{1}{p(u_i,v_i)} \tfrac{\partial p(u_i,v_i)}{\partial T} - \tfrac{1}{p(v_i)} \tfrac{\partial p(v_i)}{\partial T} \right]. \qquad (3.15)
\end{aligned}$$

To complete the optimization task, the density estimator needs to be differentiated (Eq.(3.15)) with respect to the components of transformation T. We adopted ideas that were introduced for dense histograms [17]. Given a histogramming function f, approximating the probability density function of random variable A based upon a collection of sample points B, the probability of $a \in A$ is given by $p(a) \approx f(a, B)$ and the derivative of f with respect to variable s is estimated according to

$$\frac{d}{ds} f(a, B) = \frac{\partial}{\partial a} f(a, B) \frac{da}{ds} + \frac{\partial}{\partial B} f(a, B) \frac{dB}{ds}. \qquad (3.16)$$

The application of the chain rule in Eq.(3.16) makes an implicit assumption. It holds only for cases when the histogram estimator function f is not explicitly dependent on the variable s with respect to which the derivative is taken. Although this assumption is not quite valid in our scenario (the histograms do depend on transformation T with respect to which we take derivatives), empirically, it was established that small changes in the parameters of T are unlikely to (greatly) alter the nature of the density estimator. Hence, we apply the simplification. Furthermore, the last term on the right of Eq.(3.16) can be ignored if differential changes in the sample intensities in B result in vanishingly small changes in the density values estimated by f. Based on our experiments, that condition also holds for sparse sampling.

Utilizing the assumptions explained in case of equation (3.16) and after some algebraic manipulations, the terms in Eq.(3.15) can be expressed as:

$$\begin{aligned}
\frac{\partial p(u_i, v_i)}{\partial T} &\approx \frac{\partial p(u_i, v_i)}{\partial v_i} * \frac{\partial v_i}{\partial T}, \\
\text{and } \frac{\partial p(v_i)}{\partial T} &\approx \frac{\partial p(v_i)}{\partial v_i} * \frac{\partial v_i}{\partial T}. \qquad (3.17)
\end{aligned}$$

The terms in Eq. (3.17) correspond to changes in the DRR image intensity values resulting from modifications in the transformation parameters and to changes in the probability densities as a result of changes in sample intensities.

We approximate the derivatives of the probability densities by the use of finite differences calculated from their corresponding histogram estimates. Deriving the other unknown term, $\frac{\partial v_i}{\partial T}$, though is more complex, and the details of the related computations are explained below. This is the same term that we need for the Parzen approximation in Eq. (3.14).

3.5.2 Partial Derivatives of Volume Intensities *wrt* T

Each iteration of our iterative search corresponds to a small angle rotation and small displacement applied to the current transform estimate. As among the components of transformation T, only rotation R and displacement D_d need to be recovered, we only take partial derivatives with respect to these terms.

For calculations of the update elements, we introduce a new term, the *update* rotation R_u. This operator represents the small angle rotation r which adjusts the value of the current rotation estimate at the end of each iteration cycle. (Note that at the beginning of each iteration of the registration algorithm r is reset to be a zero angle rotation.) Hence we write the new rotation component of the transform as $(R_u \circ R)$ and the transformation itself as

$$T = D_c \circ R_u \circ R \circ D_d = D_c \circ R_u(r) \circ R \circ D_d(d).$$

A transformed point becomes

$$T(x) = T(r, d, x) = D_c(R_u(r, R(D_d(d, x)))). \tag{3.18}$$

From Eq.(3.17), we need to compute

$$\frac{\partial v_i}{\partial T} = \frac{\partial V(T(x_i))}{\partial T} = \left\{ \frac{\partial V(T(x_i))}{\partial r}; \frac{\partial V(T(x_i))}{\partial d} \right\}. \tag{3.19}$$

In the following calculations, the vector r encodes a rotation transform according to the equivalent angle-axis notation. The magnitude of vector r determines the angle of rotation and its direction stands for the axis of rotation (see Section 2.1.7).

In order to express the partial derivative terms, we use the ray-casting algorithm to model the formation of the fluoro image intensities.

(The ray-casting algorithm is used instead of a more efficient procedure, as we only sample a small fraction of the image intensities and the whole image is not constructed.) In particular, a sample of the simulated fluoroscopic image at location x_i on the image plane (or at $T(x_i)$ in data coordinates) is approximated as

$$V(T(x_i)) = \sum_{\mathrm{z}\in \mathrm{ray}(T(x_i),S)} \mathrm{Vol(z)},$$

where *ray* refers to the line segment which connects the imaging source S with $T(x_i)$ on the imaging plane and z indicates uniformly distributed steps along that ray within the volume. As the steps are located in the transformed coordinate space, we could write $z = T(y) = T(r, d, y)$. Therefore,

$$\frac{\partial V(T(x_i))}{\partial T} = \sum_{\mathrm{z}\in \mathrm{ray}(T(x_i),S)} \frac{\partial \mathrm{Vol}(T(r,d,y))}{\partial T}. \tag{3.20}$$

Update *wrt* Displacement

We first calculate the partial derivative of the volume intensity with respect to the i^{th} component of displacement d, denoted as d_i. In Eq.(3.26) and (3.21), e_i stands for a unit vector whose components are all zero except for the i^{th} one which equals one.

$$\begin{aligned}
&\left[\frac{\partial}{\partial d_i}\mathrm{Vol}(T(y))\right] = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet \tfrac{\partial (D_c(R_u(r,R(D_d(d,y)))))}{\partial d_i} = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet \tfrac{\partial (R_u(r,R(D_d(d,y))))}{\partial d_i} = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet \tfrac{\partial (R_u(r,R(y+d)))}{\partial d_i} = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet \tfrac{\partial (R_u(r,R(y))+R_u(r,R(d)))}{\partial d_i} = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet \tfrac{\partial (R_u(r,R(d)))}{\partial d_i} = \\
&\quad \nabla \mathrm{Vol}(T(y)) \bullet (R_u(r,R(e_i)))
\end{aligned} \tag{3.21}$$

The full expression is

$$\frac{\partial}{\partial d}\mathrm{Vol}(T(y)) = \begin{bmatrix} \nabla \mathrm{Vol}(T(y)) \bullet (R_u(r,R(e_1))) \\ \nabla \mathrm{Vol}(T(y)) \bullet (R_u(r,R(e_2))) \\ \nabla \mathrm{Vol}(T(y)) \bullet (R_u(r,R(e_3))) \end{bmatrix}. \tag{3.22}$$

Update *wrt* Small Angle Rotation

We derive the partial derivative terms of the volume intensities with respect to the rotation component of transformation T similarly to the above. First we only consider the i^{th} element of r, which we denote as r_i.

$$\left[\frac{\partial}{\partial r_i}\text{Vol}(T(y))\right] =$$

$$\nabla\text{Vol}(T(y)) \bullet \tfrac{\partial(D_c(R_u(r,R(D_d(d,y)))))}{\partial r_i} = \tag{3.23}$$

$$\nabla\text{Vol}(T(y)) \bullet \tfrac{\partial(R_u(r,R(D_d(d,y))))}{\partial r_i} = \tag{3.24}$$

$$\nabla\text{Vol}(T(y)) \bullet \tfrac{\partial(R(D_d(d,y))+r\times R(D_d(d,y)))}{\partial r_i} = \tag{3.25}$$

$$\nabla\text{Vol}(T(y)) \bullet \tfrac{\partial(r\times R(D_d(d,y)))}{\partial r_i} =$$

$$\nabla\text{Vol}(T(y)) \bullet (e_i \times R(D_d(d,y))) = \tag{3.26}$$

$$e_i \bullet (R(D_d(d,y)) \times \nabla\text{Vol}(T(y))) =$$

$$(R(D_d(d,y)) \times \nabla\text{Vol}(T(y)))_i$$

Hence, with respect to the full vector r,

$$\frac{\partial}{\partial r}\text{Vol}(T(y)) = R(D_d(d,y)) \times \nabla\text{Vol}(T(y)). \tag{3.27}$$

We note two of the steps in the above derivation. First, Eq.(3.24) is a result of a simplification to the formula in Eq.(3.23). As the constant displacement operation D_c only happens after the rotation, this has no effect on the partial derivatives that are being calculated. That term disappears from the numerator. Secondly, to arrive at Eq.(3.25), we use the fact that R_u is strictly defined to stand for a *small* angle rotation. In that case that we can make the assumption that a coordinate point p, after a rotation by r can be expressed in the form:

$$p' = R_u(r,p) = r(p) = p + r \times p. \tag{3.28}$$

For a more detailed explanation of why Eq. (3.28) holds, see the Appendix.

As a reminder, calculations in case of the second projection image and the corresponding fluoroscopy image are performed in the same manner. The only difference is that before T is applied to transform a coordinate, an additional transformation takes place which is responsible for expressing the second projection environment.

3.6 Summary

We provided a detailed description of the objective function that we selected to evaluate the estimated transformation parameters at intermediate stages of the registration algorithm. We use two different strategies to identify the parameters that would maximize that measure. One of them, Powell's method, only needs to know how to evaluate the matching score, while the more efficient gradient-based techniques rather compute the direction of the updates that could lead to the optimum. We derived, in great detail, the terms that are required for the latter strategy and presented two distinct ways of estimating probability densities which is also a required component of the maximization process.

Chapter 4

Experimental Results

Chapter Summary

This chapter introduces the experimental framework that we used in order to characterize the performance of our registration procedure. We also describe the 2D and 3D datasets that were available to us and provide quantitative and qualitative evaluation of the registration results. We present results not only from experiments with CT and fluoroscopic images, but also with CT and CT-derived simulated DRRs. The latter experiments were necessary as we did not obtain ground truth information along with the real projection images. In order to thoroughly explore the characteristics of our method, we provide accuracy results also with respect to simulated datasets. In that analysis, we address issues related to multiresolution techniques, speed criterion and robustness.

4.1 Probing Experiments

Before we started evaluating the performance of our registration approach, we intended to carefully and extensively investigate the robustness of our objective function. We also intended to compare the properties of mutual information to those of another widely used similarity function, *pattern intensity*. Therefore, we designed probing experiments that would quantitatively describe the behavior of an objective function with respect to its free variables. Given a ground-truth estimate of the searched parameters as the starting position and orientation, the matching qualities were computed while some/all of the free

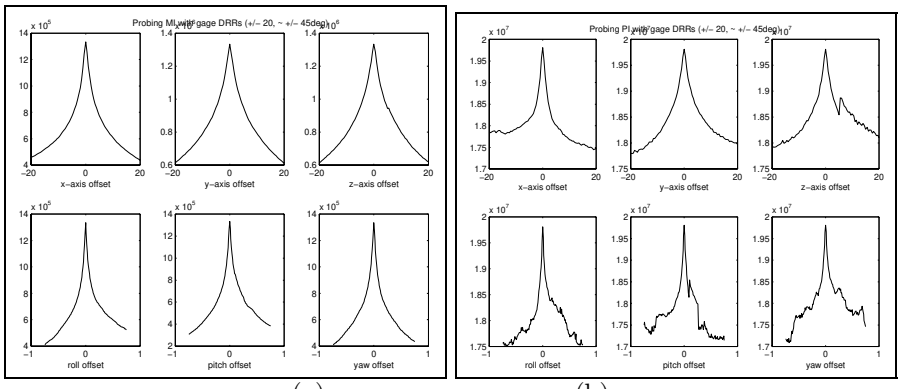
variables were slightly and iteratively modified. In the majority of the experiments, we only varied one of the variables at a time. Although that decision prohibited us from acquiring a more complete characterization of the similarity measure, it was a way to keep the computation time under reasonable limits. (Otherwise, thoroughly evaluating any kind of a matching criterion in a higher dimensional space could be a real challenge especially given the average size of our input volumes. For a detailed reference on the size of the datasets, see Table 4.1.)

With the help of the probing experiments, we were able to form a reasonable prediction about the major objective function characteristics. The capture range, the height and location of the function extremum were all useful for estimating the registration performance of the examined objective function given a specific dataset. We show two examples of outputs of such probing experiments in Fig. 4.1. The one on the left hand side, Fig. 4.1 (a), evaluates mutual information and the other, Fig. 4.1 (b), the pattern intensity on a CT-derived skull dataset.

Pattern intensity (PI) is an objective function, that some studies found to be quite robust when solving the 2D-3D rigid-body registration task [6, 3]. It operates on the difference image of its two inputs and computes the *structuredness* in small neighborhoods of each individual pixel. The more uniform the neighboring intensities are the higher the score that pattern intensity assigns at that particular pixel location. We provide the formula for calculating PI on the difference image (I_{diff}) of the two input images (I_{fluoro} and I_{drr}) in Eq. (4.1). The detailed definition of mutual information was provided in Section 3.2.1.

$$\begin{aligned} \text{PI}_{r,\sigma}(I_{\text{diff}}) &= \sum_{x,y}\sum_{u,v} \frac{\sigma^2}{\sigma^2 + (I_{\text{diff}}(x,y) - I_{\text{diff}}(u,v))^2}, \\ &\quad \text{s.t. } \sigma \text{ is a constant,} \\ &\quad (u-x)^2 + (v-y)^2 < r^2, \\ \text{and } I_{\text{diff}} &= I_{\text{fluoro}} - s * I_{\text{drr}}, \text{where } s \in \Re^+. \end{aligned} \tag{4.1}$$

A collection of probing experiments, similar to the ones displayed in Fig. 4.1 could verify that mutual information peaked when the correct alignment was evaluated. Although pattern intensity also took its maximum at the zero offset location (ground truth value, in this case), we found more local extrema in the vicinity of the ideal transformation parameters. In Fig.4.1, probing experiments are displayed with respect to all six of the free variables. Evaluation curves in the upper rows correspond to experiments with displacement modifications,

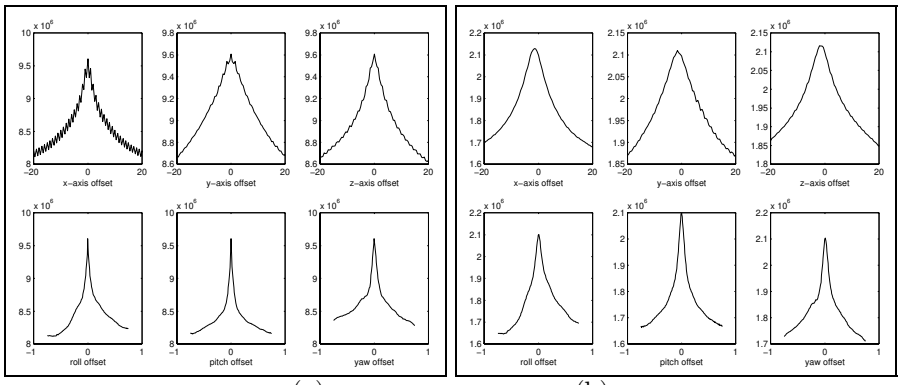


(a) (b)

Figure 4.1: Results of two probing experiments evaluating (a) mutual information and (b) pattern intensity on the skull dataset. Displacement range of $+/-20$ (mm) and rotational range of $\approx +/-45$ (deg) were specified.

while similar curves in the bottom rows contain the evaluation results due to perturbing rotation angles (roll, pitch and yaw respectively). We can see that especially in case of changes due to rotational changes, curves corresponding to pattern intensity are more jagged. That means that the optimization function could more easily get trapped in local extrema when using PI as opposed to MI. These and more extensive analysis of the same sort led us to decide that pattern intensity was a less preferable function to work with.

We also carried out numerous probing experiments at different levels of resolution. Our aim was to use their results to support our argument about increasing robustness with decreasing resolution level. In Fig. 4.2, we present probing curves from two identical experiments with the only difference being that the input CT volume was downsampled by a factor of 2 and smoothed (with a small Gaussian kernel) in case of the second one (Fig. 4.2 (b)). Analyzing the objective function curves, it is apparent that in case of the downsampled dataset the peaks of the curves are less pointy, all six curves are smoother and there are fewer local extrema encountered around the ground truth position. (The curves in the top row, just in case of Fig. 4.1, indicate evaluation results due to displacement changes and curves at the bottom represent changes with respect to rotational components.)



(a) (b)

Figure 4.2: Results of two probing experiments evaluating a cost function on (a) the original and (b) the downsampled and smoothed version of the same phantom pelvis dataset. Displacement range of $+/-20$ (mm) and rotational range of $\approx +/-45$ (deg) were specified.

4.2 Summary of the Registration Algorithm

We first provide a brief summary of our registration framework. The short, top-level outline is followed by more implementation details in the subsequent sections. The three major steps of our alignment algorithms are:

1. Preprocessing the input images and input volume
2. Initializing the imaging environment
3. Iterative optimization of the similarity measure

4.2.1 Step 1: Preprocessing

During the preprocessing step, we smooth the fluoro images to better match the resolution of the CT-derived DRR images during the alignment process. We also eliminate all artificial labels from the acquisitions that were placed there for patient identification and check whether the fluoroscopy data contains a black rim around the image margins. If it is present, that is the artifact of the image intensifier of the imaging X-ray machine (Figure 4.6 serves as a good example).

In case of the CT volume, the desirable *window* and *level* settings have to be defined. These entities determine the range and average

value of intensities which are usually set by radiologists following an image acquisition.

In a multiresolution approach, it is at this stage that additional downsampling and smoothing operators are applied to the volumetric dataset and the fluoro images. If the input volumetric dataset contains too much background (in case of head imaging that could easily happens), we can also crop the volume. In this way we do not have to spend time ignoring voxels with no useful data content during the registration procedure. (Note, however, that this step is not equal to a full segmentation task. We do not eliminate the whole background, which provides useful information during alignment procedures, we just roughly estimate the smallest bounding volume around the imaged anatomy.)

4.2.2 Step 2: Initialization

The initialization step involves reading in the parameters that are known about the imaging environment. That information is necessary in order to simulate the X-ray creation procedure as accurately as possible. Also, at this point, we roughly position the CT volume in the scene: we make an initial guess about the parameters of transformation T. (We make the assumption that a rough estimate about the required transformation is always available. That is a realistic/reasonable expectation as registration algorithms, solving the proposed problem, are not applied to find alignments greater than 30° and 30 (mm) in general, but instead to provide finer details.)

4.2.3 Step 3: Optimization Loop

Non-gradient Powell Approach

Until convergence is detected or, in other words, as long as the Powell tolerance measure is smaller than the individual improvements that are made after each iteration towards the function optimum, two steps alternate. First, the evaluation of the similarity measure given the CT volume, the observed projection images and the current transformation estimate T takes place. Second, the transformation estimate is updated in a way that increases the matching score. This method can be treated almost as a black box procedure. Besides the Powell tolerance measure, the order of linear optimizations and an upper limit for maximum optimization iterations, there are no other parameters that need to be fine-tuned.

Gradient-based Maximization

The current version of the algorithm executes the iterated part of the code a predetermined number of times. In our case, that number is 5000. (This number was experimentally determined.) Hence, for a fixed number of iterations and for all fluoro-DRR image pairs, we follow these steps.

1. *Fluoro sampling* Randomly sample image points from the observed image and extract their intensities ($U(X)$ where X denotes the sample collection).

2. *DRR sampling* Calculate the corresponding DRR values ($V(T(X))$) by applying the current transformation estimate to the CT volume and running the ray-casting algorithm. (In case we did not use the sparse sampling approach but utilized all available intensity information, we would apply one of the more efficient techniques for creating the DRR images (see Section 2.1.2), because the fast DRR-creating strategies usually achieve significant speedup only if the whole DRR image has to be calculated.)

3. **Objective function evaluation* This step is indicated with a (*) symbol as it is not an integral part of the algorithm using the gradient maximization method. This approach guides its search towards the target extremum considering only the gradient and not the value of the objective function value. We evaluate the MI estimate at each iteration for the purposes of monitoring the convergence behavior of the algorithm as a function of iterations, but it is not a required step in the optimization procedure. To compute our similarity measure, mutual information, we use Eq.(3.4).

4. *Transformation update* Compute transformation updates and assign a new transformation estimate according to Eq. (3.7), applying all the computations derived for probability density and partial derivative estimates in Chapter 3. The update is

$$T_{\text{updated}} = \tag{4.2}$$
$$T + \frac{\lambda}{N} \sum_{j=1}^{2} \sum_{i=1}^{N} \sum_{z \in \text{ray}} \frac{\partial \text{Vol}(z)}{\partial T} * \left(\frac{1}{p(u_i, v_{ji})} \frac{\partial p(u_i, v_{ji})}{\partial v_{ji}} - \frac{1}{p(v_{ji})} \frac{\partial p(v_{ji})}{\partial v_{ji}} \right)$$

for the histogramming approach and

$$T_{\text{updated}} = \tag{4.3}$$
$$T + \frac{\lambda}{N_A} \sum_{x_j \in A} \sum_{x_i \in B} (v_j - v_i)^T K_{ij} \frac{d}{dT} (v_j - v_i),$$

where

$$K_{ij} = \left[W_v(v_j, v_i) {\psi_v}^{-1} - W_{uv}(w_j, w_i) {\psi_{uv}}^{-1} \right]$$

for the method using Parzen estimation.

In Eq. (4.2) and (4.3), λ denotes the learning rate of the update variables which is experimentally determined for our application. The learning rates for rotational and translational components are significantly different but the influence of their unit update on the transformation should be approximately the same.

Many times, it can also be useful to distinguish between in- and out-of-plane transformations, as the latter transformations are usually more difficult to correctly estimate.

We found that the magnitude of the step size is more crucial in case of the histogramming approach and especially in case of the higher resolution computations. It took us some experimentation to find the best set of values (which can sometimes vary among the different datasets as well).

In most of our experiments, convergence took place much earlier than the 5000 iteration that were used. It usually took less than half as many steps as we required to reach the ideal settings. It also happened, though, that in certain cases the fixed number of iterations was not enough. Hence, instead of using the predetermined number of iterations as a stopping criterion, it would be more desirable to halt the registration procedure as soon as convergence is detected[1].

4.3 Registration Results

We would like to point out at the very beginning of this section that no special code optimization has been applied to our algorithm. All relative speed-ups demonstrated are purely the result of either the sampled or the multiresolution approach. We also carry out a great number of additional/superfluous similarity function evaluations and full DRR generations that significantly increase the execution time. Hence, our running times are not directly comparable to solutions geared towards minimal running time.

4.3.1 Registration Error Evaluation

In case of our controlled experiments, when we possess ground truth information about the searched transformation, we determine the qual-

[1] We did not investigate this problem.

ity of the registration results by calculating an *error* transformation, T_{error}. This variable is defined to be the transformation that takes the registration output pose to the ground truth one.

$$T_{\text{GT}} = T_{\text{error}} \circ T_{\text{output}}. \tag{4.4}$$

When referring to the magnitude of the registration error, we actually describe a tuple, (d_e, r_e). One component of that tuple, d_e, is the magnitude of the displacement component of T_{error} and the second element, r_e, is the magnitude of the rotation angle encoded by the unit quaternion component of the error pose. (See Chapter 2 for a detailed description of *pose* parameters.) We want the reader to notice that d_e and r_e do not directly represent errors in the searched components of the transformation. For instance, displacement errors in the estimate for sub-transform D_d are summed together to produce d_e and a non-zero d_e could also be the result of merely a noisy estimate for R, the rotational sub-transform of T, without any actual error in D_d itself. (As a reminder, transformation T, in Chapter 2, is defined to be $T = D_c \circ R \circ D_d$.) See an example for that case in Fig. 4.5 (b). Even though we only perturbed the ground truth pose by a rotation angle of 15° around the y-axis, there is a significant error in the displacement term d_e as well.

This unusual error interpretation is the result of the particular way we constructed transformation T. Results in the registration literature are often times presented with respect to the displacement and rotation angle error specific to displacement directions and rotation axes. Therefore, one should keep in mind this difference when comparing measures produced by the various approaches.

In order to determine whether we perform our task with high accuracy, we establish a range for d_e within which the results satisfy our interpretation of *sub-voxel accuracy* requirements. For r_e, no special bound needs to be defined as R is the only rotational component of T, so all rotation angle errors are directly related to R.

If (dx, dy, dz) denoted the size of the CT voxels, we formulate a criterion for sub-voxel accuracy in the displacement parameter by specifying the range

$$0 <= d_e <= \sqrt{(dx^2 + dy^2 + dz^2)}. \tag{4.5}$$

That is to say, we bound the displacement error term with the magnitude of the diagonal of a volume element. That upper limit represents a worst case scenario: having a translational offset of exactly the length of the volume element in all three directions. So for example, if the voxel

dimensions were $(0.5; 0.5; 1)$, then a displacement error in the range of $d_e <= \sqrt{1.5} \approx 1.224$ meets the criterion of being *highly accurate*. As this measure depends on the input data specifications, we calculated its value for all of the CT volumes that we used in our experiments. They are denoted by $d\star$ and appear in the last column of Table 4.1.

As an alternative, we could also look at error in the subcomponents themselves. That is, we could look at how much transformation D_d and R are different from their ground truth values. We look at these measurements more closely in case of the real X-ray - CT experiments in Section 4.5, where we obtain results with bigger variance than in case of the controlled setup.

4.3.2 Objective Function Evaluation

In order to closely follow the convergence pattern of our algorithm, we computed the mutual information measure at each iteration of the registration procedure. This is an optional step in case of the stochastic gradient ascent methods. Although the objective function curves are quite jagged in all cases (which is an expected result of the sparse sampling method), we can definitely observe an overall convergence pattern. Some example figures, to which we are going to refer in a later analysis, are displayed in Fig. 4.5.

In case of the Powell maximization method, we did not record intermediate evaluation results throughout the registration process.

4.4 CT-DRR Experiments

4.4.1 CT-DRR Registration

We designed several controlled experiments in order to obtain a thorough characterization of the algorithmic behavior of our registration methods given a known, ground truth transformation parameter. We wanted to test their accuracy under a wide range of circumstances: on different resolution levels, with different learning rates and with different sample sizes.

After an initial offset was specified, high-quality, CT-derived DRR datasets of a plastic pelvis, plastic skull, real skull, real head and plastic lumbar spine were registered to their volumetric counterparts.

The specifications for the CT datasets from which we created the simulated projection images are listed in Table 4.1. The second column contains row, column, and slice information in that order and the third column specifies voxel dimensions (dx, dy, dz). The last column, with

quantity $d\star$, represents the *sub-voxel accuracy* upper bound (Sec. 4.3.1) to which we compare our registration results. As a fascinating bit of information, one set of our experiments was run on Phineas Gage's skull dataset[2].

DATASET	VOLUME	VOXEL	$d\star$
Pl. Pelvis	265 x 455 x 107	[0.6621^2;2.0]	2.2083
Pl. P. (sm1)	132 x 227 x 107	[1.3242^2; 2.0]	2.7399
Pl. P. (sm2)	66 x 113 x 107	[2.6484^2; 2.0]	4.2460
Pl. P. (sm3)	33 x 56 x 53	[5.2969^2; 4.0]	8.4920
Real Skull	512 x 512 x 390	[0.4844^2; 0.5]	0.8481
Real S. (sm1)	256 x 256 x 195	[0.9688^2; 1.0]	1.6962
Real S. (sm2)	128 x 128 x 97	[1.9375^2; 2.0]	3.3923
Gage's Skull	512 x 512 x 424	[0.4473^2; 0.5]	0.8063
Gage's S. (sm1)	256 x 256 x 212	[0.8945^2; 1.0]	1.6126
Gage's S. (sm2)	128 x 128 x 106	[1.7891^2; 2.0]	2.8985
Pl. Skull	188 x 128 x 105	[1.0156^2; 2.0]	2.4623
Real Head	512 x 512 x 127	[0.4883^2; 1.0]	1.2153
Pl. L. Spine	512 x 512 x 103	[0.2344^2; 1.5]	1.5362

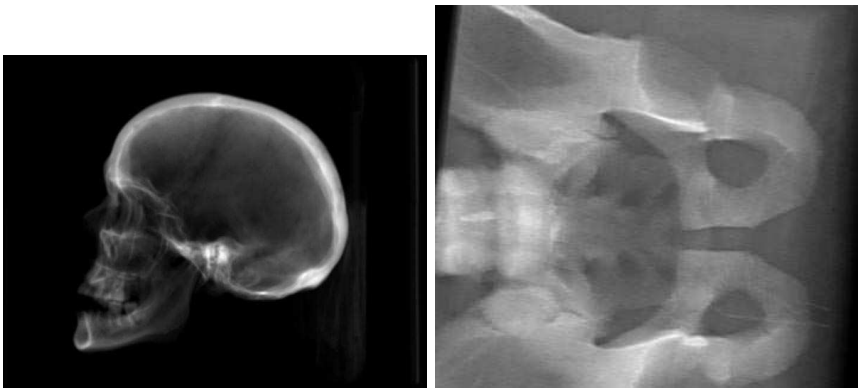
Table 4.1: CT dataset specifications (for Plastic Pelvis, Real Skull, Gage's Skull, Plastic Skull, Real Head and Plastic Lumbar Spine); sm1: smoothed volume on hierarchy level 2; sm2: smoothed volume on hierarchy level 3; sm3: smoothed volume on hierarchy level 4

Single views of the *observed* DRR images that were used as the simulated fluoro images in the registration experiments are displayed in Figure 4.3.

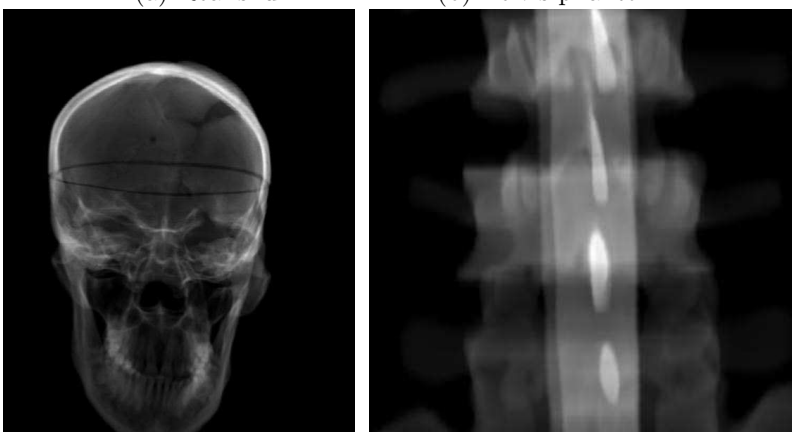
In the following, we provide a performance evaluation of our registration approaches on these datasets. We focus on three key features of the algorithms: benefits from the multiresolution hierarchy, capture range and accuracy. During the evaluation, we use the following naming convention to distinguish between our algorithms with different optimization methods:

- *Reg-Pow*: registration using the Powell maximization method
- *Reg-Hi*: registration using stochastic gradient ascent with sparse histogramming for density estimation

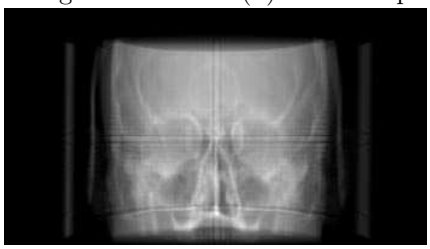
[2]We provide a brief description of his famous head injury in the Appendix.



(a) Real skull (b) Pelvis phantom



(c) Phineas Gage's skull (d) Lumbar spine segment



(e)Real head

Figure 4.3: Single-view simulated fluoroscopic images from the controlled experiments.

- *Reg-Pz*: registration using stochastic gradient ascent with Parzen Windowing for density estimation

Further abbreviations used in the succeeding sections are:

- *LEVELS sm0, sm1, sm2, sm3*: To indicate the various levels of resolution we use in the hierarchy. They denote, respectively, the highest level of resolution (the original dataset with no downsampling), 1st, 2nd and 3rd level of hierarchy of downsampled datasets[3].
- *No.*: number of experiments executed for the given task.

Machine Specifications

We had access to two types of computing resources and we indicate their characteristics in Table 4.2. The *name* section in the first column demonstrates the abbreviations by which we refer to them in our analysis.

NAME	MODEL NAME	CPU (MHz)	CACHE SIZE
M1	Pentium III	1004.542	256KB
M2	Pentium III	497.446	512KB

Table 4.2: Computing resources – machine specifications.

4.4.2 Multiresolution Approach

Motivation

As we have already mentioned in Chapter 2, we investigated a hierarchical approach to the 2D-3D registration problem. The idea behind this formulation stems from an approach originally offered in the field of image compression [37]. Essentially, as we descend to lower levels of the hierarchy we aim to eliminate superfluous information encoded in the dataset and attempt to represent it in a more compact manner.

Since the introduction of this strategy, it has been widely used in the computer vision community. In medical imaging applications, for instance, excellent results have been presented in multi-modal 3D-3D head registration applications [16, 18].

Our main motivation behind running our experiments on various levels of resolution was to increase the speed and the robustness of our

[3]The downsampling procedure always takes place with respect to a factor of two.

alignment procedure. Even if we did not want to find the fastest solution to the problem, we can see in Table 4.1 that most of the datasets available to us are extremely large. These were taken specifically to test research efforts, so their resolution is generally higher than that of an acquisition for medical/treatment purposes would be. For example, the skull dataset (indicated in the 5th row of Table 4.1) has 390 slices and a slice thickness of 0.5 mm which is 2 or 3 times more than it would have been if requested for ordinary diagnostic purposes. Handling such large datasets efficiently is a challenging task, especially when we have to traverse the volume several times in order to produce full projection images (in case of the Powell method and when displaying intermediate registration results). Hence, given the 3D and 2D input datasets, we downsampled and smoothed them (with a Gaussian kernel) to obtain versions of the original with lower resolution. Due to the high accuracy of our initial volumetric datasets, we used 3-4 levels of hierarchy. The data specifications about the lower resolution volumes are also included in Table 4.1.

Experiments

We demonstrated in Section 4.1 that the objective function can be made much smoother if we downsample the original input datasets, see Fig. 4.2, for an example, where we show probing experiments on Gage's skull dataset. It is apparent from the figures that local extrema can be avoided by downsampling the images.

To measure the registration speedup that we could gain from applying the hierarchy, we ran some experiments with the same initial offset applied to the ideal/ground truth transformation at each level of the pyramid using all three of the optimization methods. In all of the experiments, we recorded the running time it took for the algorithms to converge to the optimal parameters.

We summarize our results in Table 4.3. For these experiments, we used two different ways to initialize our transformation estimate. We either defined a fixed offset of 20 or 30 mm in displacement or 15-20° in rotation angle for one of the free variables or we specified a displacement and rotation range from which the offset value was randomly picked for all of the six unknown elements. The former initialization method, in general, allows testing for maximum possible offsets that could be recovered in individual dimensions of the space. Only a single parameter is offset and goal is to identify what the upper bound is for it. The latter initialization helps to evaluate the robustness of an algorithm in a given subspace. As all of the components are perturbed

from their optimal value, it is a more complex task to optimize for them simultaneously.

We recorded the computation time for all three of the alignment strategies: the Powell, the Parzen windowing and the histogramming methods. In all cases, but most significantly, for the Powell approach, the computational speedup achieved by the hierarchy was enormous. As Table 4.3 indicates, for the direction set method, computations on the pelvis dataset converged 8.357 times faster on the third level of the hierarchy than on the second and in case of the Gage dataset the algorithm was completed 6.812 and 19.413 times faster on the forth pyramid level than on the second and third respectively. (In our table, running time is indicated in seconds and the offset measures are given in *mm* and degrees.)

Method	Data	Level	Time (s)	CPU	No.	Offset (mm, deg)
Reg-Pow	pelvis	sm1	14992.50	M2	6	[20;15]
		sm2	1794.00	M2	6	[20;15]
	gage	sm1	29663.25	M1	6	[20;15]
		sm2	3327.00	M1	6	[20;15]
		sm3	1528.00	M1	6	[20;15]
Reg-Hi	pelvis	sm0	8619.50	M1	6	[20;15]
		sm1	4311.50	M1	6	[20;15]
		sm2	2156.33	M1	6	[20;15]
	skull	sm0	6613.00	M1	6	[30;20]
		sm1	3410.00	M1	6	[30;20]
		sm2	1592.00	M1	12	[30;20]
Reg-Pz	gage	sm1	12084.33	M2	6	[10;20]
		sm2	6337.33	M2	6	[10;20]
		sm3	3567.33	M2	6	[10;20]

Table 4.3: Timing measurements to contrast registration running time on different hierarchical levels.

In case of the histogramming approach (using the *Reg-Hi* method), running the alignment procedure on the pelvis dataset was 1.9992 and 3.9973 faster on the second and first pyramid levels than on the original input. The same experiments produced a 1.94 and a 4.154 speed-up in case of the skull experiments comparing results with the original and the second level of the hierarchy.

With the Parzen Windowing approach we achieved similar results. For example, a speedup of 1.907 and of 3.387 was noted when register-

ing images taken of the skull dataset.

It is important to point out that we carry out the downsampling procedure with a factor of two in all desired dimensions in a way that the voxels in the resulting volume approximate cubical elements. Therefore, with each level of hierarchy, the data volume size grows by 2^3 in the worst case scenario. For algorithms that fully traverse the 3D image and produce full-sized DRRs that increase directly appears in the running time. In the sampled approach, the slowdown is (approximately) at most by a factor of two if the number of samples used remains the same between the levels. That is because we only have to access more voxels when following rays going through the volumes.

We note that the execution times of the three different approaches should not be directly compared using data indicated in Table 4.3. That is especially true for the two different optimization strategies. The reason for that is the fact that the experiments using the Powell method ran until convergence was detected while the gradient methods were executed for a fixed number of iterations.

We are able to conclude, though, that the hierarchical approach is inevitable when using the non-gradient maximization method on large datasets. Otherwise the running time grows unreasonably high, exceeding several hours. In case of a sampled approach, using only a few data points for the computations keeps the execution time well-manageable. The relative gain in between hierarchy levels is smaller but still significant.

4.4.3 Robustness, Size of Attraction Basin

Given the ground-truth pose parameter T_{GT}, the capture range of the algorithm with respect to a particular dataset can be established by finding the greatest perturbation of individual components that could be consistently reset by our application.

First, we could get a good intuition for the extent of the capture range while running probing experiments (Section 4.1). As probing the six dimensional parameter space is computationally highly expensive, lower dimensional experiments can be carried out to characterize the objective function. However, the probing results can be misleading as they are not capable of representing the parameter space in its full complexity. So to evaluate the registration power of our optimization methods, we ran some initial experiments by only offsetting one of the free parameters from its ideal value. In these controlled experiments,

we found that a displacement offset of 35-40 mm and rotation angle offset of 20-25 degrees could generally be registered with all three of the methods. The alignment was more sensitive when the images were truncated. In these cases, the maximum displacement offsets had to be lowered. (That situation occurs in case of the pelvis, spine and the Gage's skull.) We then also ran some experiments where the initial offset was determined by randomly assigning offset values to all of the free variables given a pre-specified range of values. As all six parameters were offset at the same time, the individual perturbation ranges were specified to be smaller in this case.

4.4.4 Accuracy Testing

Preliminary Results

Without using the hierarchical approach and only offsetting one of the free variables at a time, we obtained good preliminary results. We were able to show that the algorithms could be applied to several types of datasets with good accuracy. We assigned offsets in the range of $35 - 40$ (mm) or $20 - 25$ (deg) and we ran the algorithms on four different datasets. These were a plastic pelvis, plastic skull, plastic lumbar spine and a real head. Most of the times, the quality of the convergence results only depended on finding the appropriate set of registration parameters (step size and sample size, mostly). We present our results for the new*Reg-Hi* method in Table 4.4.

DATASET	No.	d_e (mm)	r_e (deg)
Pelvis	31	1.9637	0.3100
Plastic Skull	26	1.1783	0.1564
Plastic Spine	18	0.5319	0.0430
Real Head	14	2.3513	0.3810

Table 4.4: Controlled, registration accuracy tests using the *Reg-Hi* method; No hierarchy;

Table 4.4 displays the number of experiments executed in case of each given dataset and the average magnitude of the displacement component and of the rotation angle of T_{error}. In the majority of the experiments, the displacement error terms fall under the $d\star$ value (Table 4.1) and the rotation angle errors are under 0.5 deg. We can notice a relative weakness in the accuracy results related to the real head experiments. This, most probably, stems from the fact that the dataset

is seriously truncated with the top and the bottom (below the nose) of the head completely missing (see Fig. 4.3).

With Hierarchy

After obtaining the above described preliminary results, we wanted to optimize both the accuracy and the running time of our methods. That is when we implemented the multiresolution approach.

In the following, we analyze performance results on two particular datasets comparing all three of our algorithms. We decided to run these more extensive experiments on the Gage skull dataset and the images of the phantom pelvis as our collaborators are particularly interested in seeing results related to two of these anatomies.

We constructed the multiresolution hierarchy in the following way. We started the experiments on the lowest level of the volume pyramid. We randomly offset the ground truth transformation variable in a particular range. The extent of this range was determined based upon some initial tests with respect to capture range calculations. More specifically, they were specified by providing an upper bound for all of the displacement components and by assigning an maximum value for the rotation angle, while the rotation axis was determined randomly. The four parameters were selected uniformly from their specified ranges. We continued the alignment process on higher resolution levels using the results of the lower stages as inputs until the top level of the pyramid was reached.

It was not only the resolution level that changed in-between the multiresolution steps. We also used lower step sizes (λ) and more sample points towards the top of the hierarchy. In case of the *Reg-Pow* experiments, we also added an additional small perturbation to the starting poses on higher levels, in order to prevent the optimization method to get trapped in local extrema. The stochastic approach introduces sufficient amount of noise into its estimates that this perturbation was not necessary in case of the gradient ascent procedures. The results of these controlled experiments are summarized in Table 4.5 and 4.6.
All timing results were measured on our $M1$-type machines.

Although, in general, the pre-specified 5000 iteration were more than sufficient for the gradient-based algorithms to converge, both in case of the *Reg-Pz* and *Reg-Hi* method we encountered exceptions. In those cases, the randomly assigned offset values were so large compared to the given step size that the registration process could not take place during the provided interval. When that happened we did not average in the results of those runs into our accuracy measures. When such an

Method	Dataset	Level	No.	(d_e (mm), r_e (deg))
Reg-Pz	Pelvis	sm0	9/10	(0.5199, 0.0947)
		sm1	9/10	(0.3580, 0.1146)
		sm2	9/10	(0.6604, 0.1337)
		sm3	8/10	(3.5453, 0.3173)
	Skull	sm0	10	(0.8551, 0.0938)
		sm1	10	(1.5816, 0.1919)
		sm2	10	(5.5943, 0.6463)
Reg-Hi	Pelvis	sm0	10/11	(0.5368, 0.1071)
		sm1	10/11	(0.3170, 0.1031)
		sm2	10/11	(1.4763, 0.1576)
		sm3	10/11	(1.9451, 0.2131)
	Skull	sm0	11	(0.8145, 0.0869)
		sm1	11	(1.0348, 0.1331)
		sm2	11	(11.5438, 1.1923)
Reg-Pow	Pelvis	sm0	12	(1.1164, 0.1472)
		sm1	12	(2.9557, 0.1495)
		sm2	12	(5.3905, 0.3676)
		sm3	12	(13.6580, 1.298)
	Skull	sm1	12	(2.7144, 0.1432)
		sm2	12	(5.2431, 0.4744)
		sm3	12	(10.9750, 1.1474)

Table 4.5: Registration results of methods *Reg-Pz*, *Reg-Hi* and *Reg-Pow* on controlled experiments of a phantom pelvis and a real skull

event happened, we specifically indicate in the fourth column of Table 4.5 the actual number of data samples out of the total with respect to which we present the accuracy measures. (In case of the *Reg-Pz* pelvis experiments, the considered number of experiments increases from 8 to 9 after the second hierarchical level, as even though the first set of iterations was not sufficient, during the second step the parameters did manage to converge to the optimal settings.)

The offset range for the registration algorithms was almost the same in all cases, with a slight variation in the displacement settings (and only once in the rotation bound). The capture range for the skull dataset was smaller in translation, but the rotational element did not have to be modified. The table entries nicely present how by ascending the multiresolution pyramid the registration errors decrease, both with respect to the rotational and the displacement components. When

Method	Dataset	Level	Offset	Time (sec)
Reg-Pz	Pelvis	sm0	from sm1	3608
		sm1	from sm2	2591
		sm2	from sm3	1558
		sm3	[15,15,15,20°]	843
	Skull	sm0	from sm1	4573
		sm1	from sm2	3516
		sm2	[10,10,10,20°]	2084
Reg-Hi	Pelvis	sm0	from sm1	3620
		sm1	from sm2	2422
		sm2	from sm3	1388
		sm3	[15,15,15,20°]	699
	Skull	sm0	from sm1	4666
		sm1	from sm2	3553
		sm2	[10,10,10,20°]	2820
Reg-Pow	Pelvis	sm0	from sm1	25358
		sm1	from sm2	4839
		sm2	from sm3	1478
		sm3	[15,15,15,20°]	240
	Skull	sm1	from sm2	13744
		sm2	from sm3	2583
		sm3	[12,12,12,15°]	561

Table 4.6: Timing results of registration methods *Reg-Pz*, *Reg-Hi* and *Reg-Pow* on controlled experiments of a phantom pelvis and a real skull

comparing the results of the controlled *Reg-Hi* and *Reg-Pz* experiments, we can say that their performance was very similar. Running on the pelvis and the skull datasets, they both completed the registration task even before getting to the top-most level of the hierarchy. (See Table 4.1 for upper bound on error terms to achieve high accuracy.) We have to admit though, that in case of the histogramming method, we had to experiment more with the appropriate parameter settings. That was especially true as we got close to the top of the registration pyramid. It proved to be more crucial to increase the sample size for more accurate density estimations, and *Reg-Hi* was also much more sensitive to the optimization step size.

The data also suggests that the *Reg-Pow* procedure might have got attracted to several local extrema. That would be the only explanation for the fact that, even on the original dataset, its error terms are larger

than that of the two other methods.

For each registration run, we also provide the running time. As we have already hinted it in earlier sections, in case of the gradient-based methods these represent only an upper bound as convergence might occur way before the pre-specified iteration number is completed. Although the Powell method converges very quickly during the first couple of stages, the additional refinement of the intermediate results approaching the top of the hierarchy takes an extremely long time. Such a delay prevents this approach to even be considered for interventional applications.

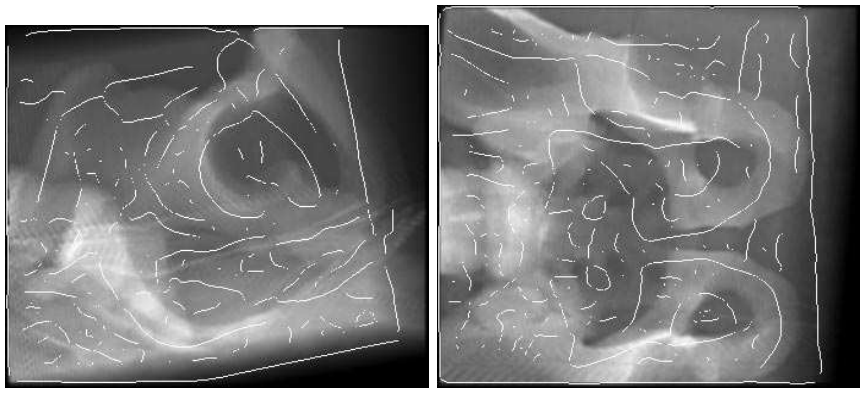
To demonstrate the extent of the registration errors in a way other than the numerical scores, we present the results of the registration experiments of the pelvis phantom images using the non-gradient technique. These examples contain the contours of projection images produced by the registration output parameters overlaid on the observed, input acquisitions. See Figure 4.4 for a qualitative evaluation of the performance of Powell's method. By visual inspection, only tiny misalignments are detectable on the images created with the output of the algorithm; the majority of the contours aligns well.

4.4.5 Convergence Pattern

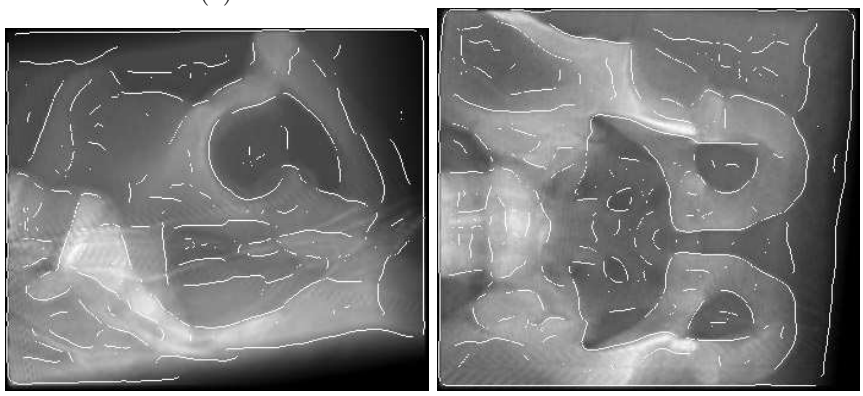
As part of the registration algorithm for the gradient-based optimization approaches, we evaluated the similarity measure at intermediate steps of the procedure. We recorded the MI estimates together with the error transformation components. Displaying these values as a function of the iteration number allowed us to monitor the convergence behavior of the alignment procedure closely. Two such convergence plots are displayed in Fig. 4.5. These present the results of a *Reg-Hi* plastic pelvis experiment. In case of the left-hand side, on Fig. 4.5 (a), the displacement parameter was perturbed by 20 (mm) in the direction of the y-axis and in case of 4.5 (b) the rotation angle around the y-axis was perturbed by 15°.

The MI curve, in both experiments, is very noisy. This can be explained by the fact that the objective function is only evaluated on a small-sized, random set of sample points. However, it is apparent that the closer the transformation estimate gets to its optimal value (as iteration counter increases), the higher the reward value is that we assign to the most current set of transformation variables. The optimization method in this particular example was *Reg-Hi* and the number of sample points used was 150.

In the first experiment, Fig. 4.5 (a), the initial error in the displace-

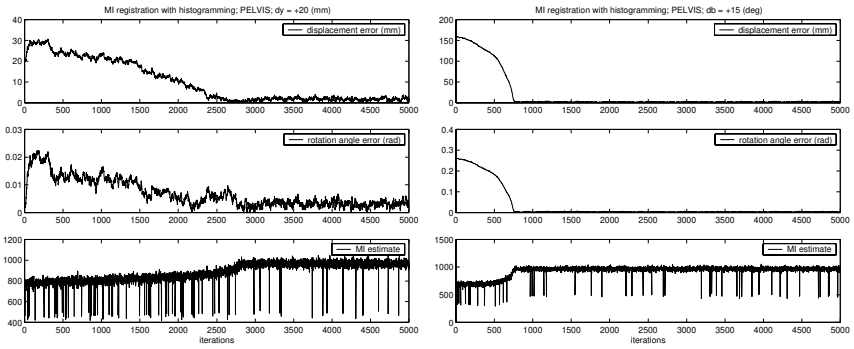


(a) With initial transformation estimate



(b) With transformation estimate resulting from registration

Figure 4.4: Registration results of a phantom pelvis controlled experiment with the *Reg-Pow* method: contours of registration results are overlaid on the observed DRR images



(a) (b)

Figure 4.5: Sample output from a controlled set of *Reg-Hi* experiments. Dataset: plastic pelvis. Initial offsets: (a) $y = 20$ (mm) and (b) $\beta = 15$ (deg). Plots display the magnitude of displacement error, rotation angle and the MI estimate at each iteration.

ment component quickly causes an offset in the rotational component as well. It is only then that they converge to the zero-offset solution simultaneously. The second run, Fig. 4.5 (b), is a good example of the discussion in Section 4.3.1. Even though it is only the rotational component of transformation T that is offset initially, our displacement error measure d_e is also non-zero at the outset of the alignment procedure. As the rotation error decreases, the displacement error vanishes, too.

4.4.6 Registration Parameter Settings

It was mainly in case of the registration methods using gradient-based optimization that we had to carefully set the operating parameters. For the Powell method the only setting that we tried to alter was the tolerance level. However, changing that variable did not produce significantly different results, so we did not invest time in quantifying its influence on the registration results.

In case of the *Reg-Hi* and *Reg-Pz* methods there are several settings that could be adjusted; for example: size of sample collection, step size, iteration number, standard deviation of Gaussian kernels (for *Reg-Pz*) and number of intensity buckets used for constructing histograms. Some of these values were treated as constants and were never changed throughout all our study. For example, the standard deviation of the Gaussian kernel was set to be 2.0 and the number of intensity bins to

32. We experimented more with the sample and the step size measures. For all our sampled experiments we used only 100-150 sample points with an increasing number towards the top of the hierarchy.

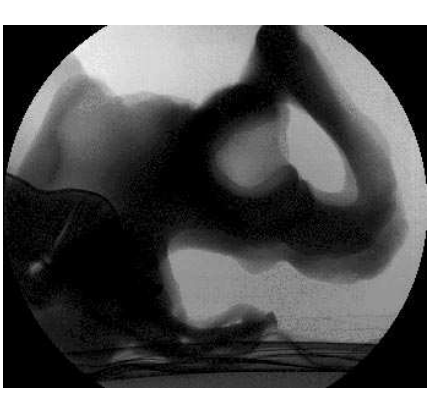
Most of the time was invested in finding appropriate step sizes (learning rates) for the optimization procedure. We found that, especially when the multiresolution approach was not used and the algorithm was run only on the top-most level, the maximization could be very sensitive to these values. If the learning rates were too high, convergence did not take place and when they were too small, the predetermined number of iterations were not sufficient to include the full convergence path. What is more, we differentiated between the update rate of the rotational and the translational components. It was crucial to set the ratio between them properly as their influence on the update had to be balanced. Neither of the update terms were to dominate the other.

4.5 CT-Fluoroscopy Experiments

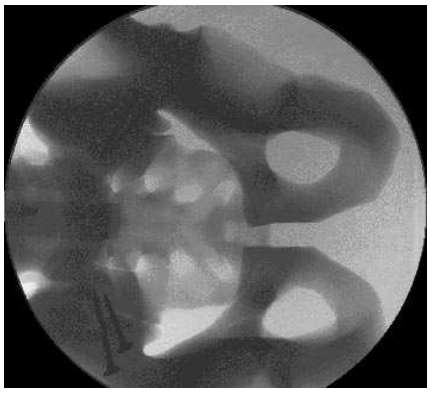
We obtained real X-ray fluoroscopy and X-ray images for two of the CT datasets that are listed in Table 4.1. We had corresponding biplanar 2D acquisitions of the phantom pelvis and Phineas Gage's skull. We present these four images on Fig. 4.6.

Unfortunately though, the ground truth specifications describing the imaging geometry at the time of image acquisition were not recorded. Hence, our best estimates about the ideal transformation T_{GT} were the results of a manual registration procedure. We attempted to reconstruct the imaging environment as accurately as possible while producing the best equivalents of the real fluoroscopy or X-ray images. However, it is highly probable that the ground truth approximations contain some (hopefully, slight) amount of error. Therefore, when registering the real fluoro/X-ray images to the CT volume, we allow for a bigger range of variance in the accuracy results compared to the standards established in the synthetic examples (Section 4.4).

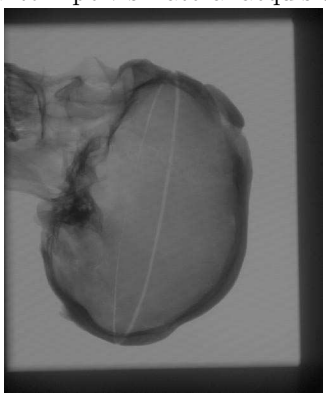
In order to make it more intuitive to interpret the extent of the registration errors, we also present our registration results in two additional ways for these datasets. It is in this section that we look at not only the error term provided by the computation of the error transformation T_{error}, but also errors with respect to the subcomponents of T themselves. We examine how precise our estimates are regarding the rotational component R and the displacement term D_d. We also



(a) Phantom pelvis: lateral acquisition



(b) Phantom pelvis: AP acquisition



(c) Phineas Gage's skull: sagittal view

(d) Phineas Gage's skull: coronal view

Figure 4.6: Real X-ray fluoroscopy of the phantom pelvis and real X-ray images of Phineas Gage's skull

provide a qualitative analysis of our results. We produce a DRR image both with the offset parameters and the results of the registration algorithm and overlay their main contours onto the observed (real) 2D images. In this way we are able to see how much of an error was recovered by our alignment procedure and also how well our final estimates match the original acquisitions. To extract the major outlines of the examined objects in these experiments, we used the Canny edge detector algorithm [4].

4.5.1 Experiments with X-Ray Images of Gage's Skull

We ran extensive experiments on the Gage dataset with all three of our registration methods. The initial offset range was specified to be a random combination of maximum [0,10,10] or [5,10,10][5] in displacement and maximum 20 degrees in rotation. We used at most three levels of the multiresolution registration pyramid. Below we present results obtained by the Parzen windowing and the Powell methods. We do not include a report on the histogramming technique as we did not find it to be robust and consistent enough in these experiments. With *Reg-Hi*, the confident range of convergence was much smaller than in case of the other two strategies (only [5,5,5,5°] as opposed to [5,10,10,20°]). It also seemed even more sensitive to code parameter settings than in case of the controlled experiments.

Reg-Pz Experiments

We ran 30 experiments with the Parzen windowing method starting from the third pyramid level (*sm2*). We obtained the following encouraging results. On the 3rd level of the hierarchy, with the input volume downsampled twice, 27 ($\approx 90\%$) of the experiments finished converging to an approximate wider neighborhood of the ground truth pose. In two cases, the initial offsets were too high to be recovered. These originated from a coupling of a large rotational and a large displacement offset. And in the remaining third experiment, convergence has started but the number of iterations was not sufficient to complete the procedure. On the second level, continuing the registration with the output from the lower hierarchy level, we could acquire even higher accuracy.

[4] Thanks to Lily Lee for providing the C++ code implementation of the Canny edge detector

[5] The range is smaller in the x-direction as the chin is missing form the datasets.

Although in all cases we got closer to the optimal settings, the three cases that did not get close enough in the first round of registration remained behind. That is explained by the fact that on higher hierarchy levels the step size decreases, hence even with an increased resolution the size of error that could be corrected decreases. (Also, we showed that the danger of being attracted by local extrema also increases.)

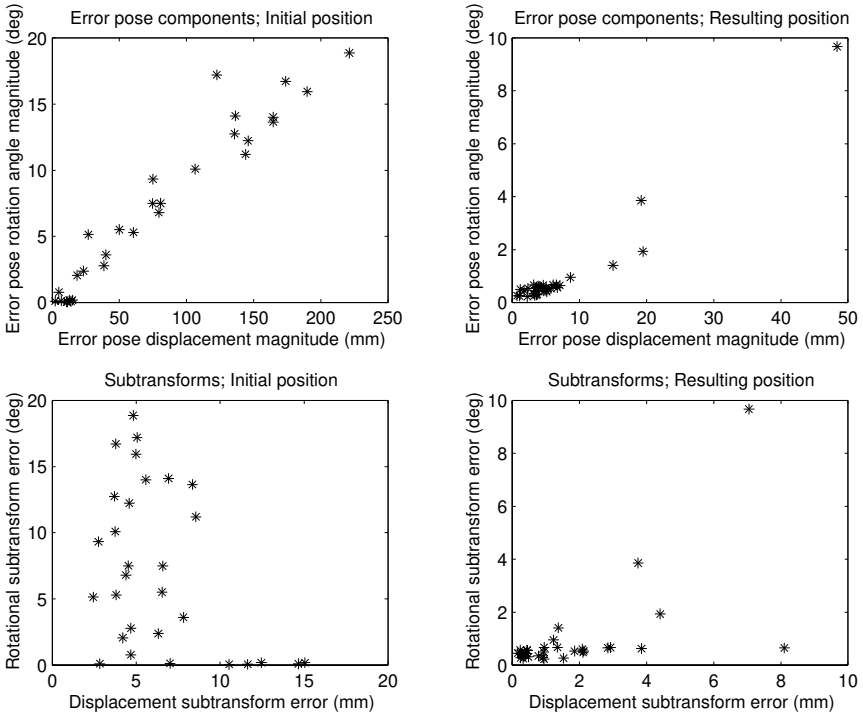
We first present our results quantitatively. We prepared two plot diagrams displaying the displacement and the rotational error terms both prior to and after the registration procedure was run. We calculated these error measures both with respect to the components of the error transformation T_{error} and the individual (variable) components of transformation T: D_d and R. The error terms, in all cases, are specified with respect to the manually determined ground truth pose. Figure 4.7 displays results obtained on the third pyramid level and Fig. 4.8 reports on the outcomes achieved on the second level. On each of the sub-plots, the x-axis stands for displacement error (measured in mm) and the y-axis represents rotation angle error (measured in degrees). The left hand side columns, on both of the main figures, represent settings from before the alignment procedure and the right hand side from after completing registration. The top row corresponds to measures with respect to the error transformation T_{error}, and in the bottom row indicates error measures computed with respect to the two varying subcomponents of transformation T.

On Fig. 4.7, we can see that with the exception of the few cases where the algorithm did not have enough time to complete convergence, the results cluster closely around the ideal settings. (The outlier data points are indicated by a circled cross-mark to distinguish them from the others.) These results are further improved on a higher level of the pyramid. These results are presented in Fig. 4.8.

We also ran experiments on the original dataset, trying to even refine the outputs from the 2nd pyramid level, but these experiments did not improve the results very much and they are very expensive computationally. Hence, we do not include those results here.

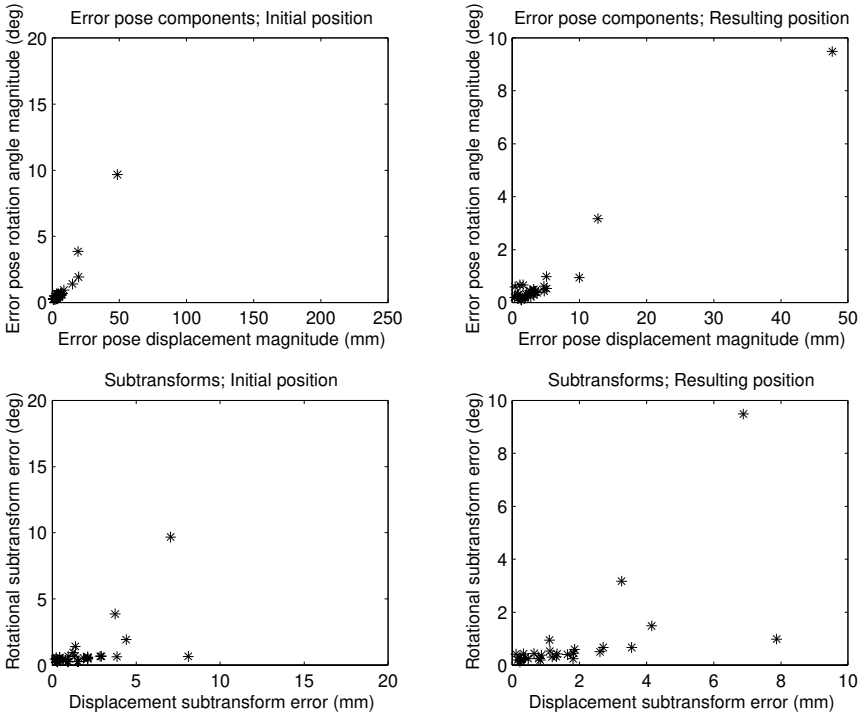
A summary of the error terms corresponding to T_{error} are also summarized in Table 4.7 and 4.8.

We also invite the reader to judge the registration results qualitatively. Images in the top row of Figure 4.9 display the two views created using the offset transformation estimates and in the bottom row they show projection images produced with the output of the registration. These images are suitable to demonstrate the extent of the recovered offset. Figure 4.10 is the one that helps to judge accuracy. There, the main outlines of the DRR images in both the offset and registered



(a) Prior to registration (b) After registration

Figure 4.7: Error distribution based upon the results of 30 experiments with random initial offset on a given interval. Row 1 displays plots with respect to error terms d_e and r_e while row 2 demonstrates errors in D_d and R



(a) Prior to registration (b) After registration

Figure 4.8: Error distribution based upon the results of 30 experiments with random initial offset on a given interval. Row 1 displays plots with respect to error terms d_e and r_e while row 2 demonstrates errors in D_d and R

Dataset	Method	Level	No.	Offset (mm, deg)
Gage's skull	*Reg-Pz*	sm1	27/30	from sm2
		sm2	27/30	[0,10,10,15°]
	Reg-Pow	sm1	9	from sm2
		sm2	9	from sm3
		sm3	9	[5,10,10,15°]

Table 4.7: Experiment description for the X-ray fluoroscopy and CT registration experiments on the Gage skull dataset

Dataset	Method	Level	(d_e (mm), r_e (deg))
Gage's skull	*Reg-Pz*	sm1	(2.4073, 0.2256)
		sm2	(4.1094, 0.4835)
	Reg-Pow	sm1	(8.5726, 0.9849)
		sm2	(9.1946, 0.8982)
		sm3	(17.6407, 1.627)

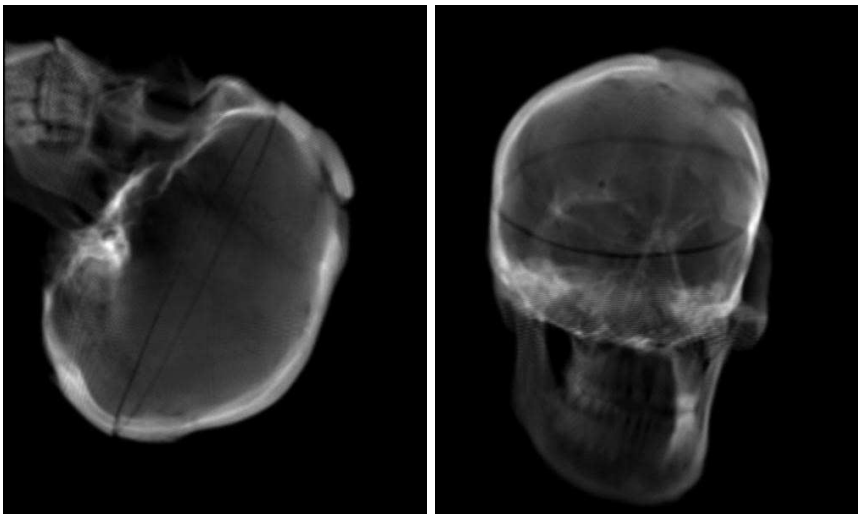
Table 4.8: Error measurements for the X-ray fluoroscopy and CT registration experiments on the Gage skull dataset

poses are displayed on the real X-ray images. We can see that the DRR boundaries closely follow the outlines appearing on the original acquisitions.

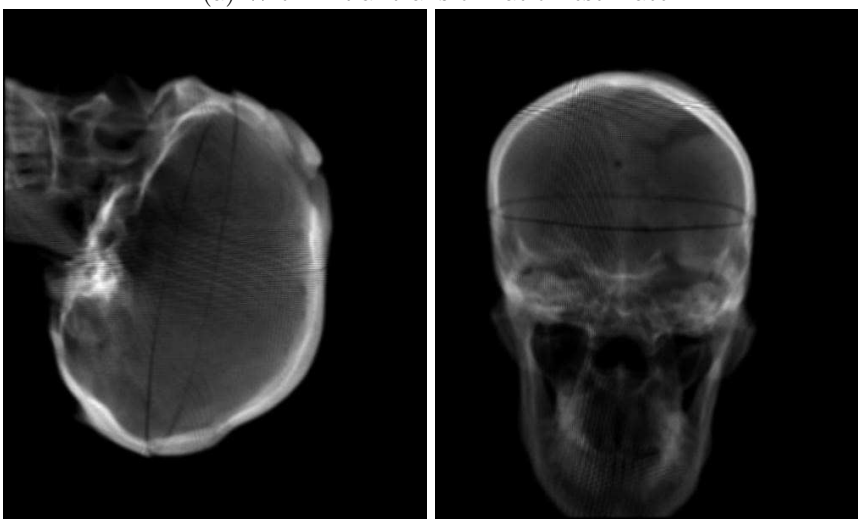
Reg-Pow Experiments

We ran a set of nine experiments on three levels of the registration hierarchy using the Powell optimization method. All nine of the experiments converged to the optimal transformation settings. Although there is a definite improvement in the accuracy results of the different stages, on average, these experiments could not produce the same accuracy results as the above presented stochastic gradient method. Table 4.7 and 4.8 present the relevant registration outcomes. It is possible, that we might have gained some error reduction on the top-most level of the pyramid, however, the running time was so high even in case of the second level (*sm1*), that these experiments were not conducted. Such a time-consuming solution could not be considered in the applications that we focus our attention on.

Figure 4.11 present results from an experiment where a fixed displacement offset of 30 mm was applied in the x-direction.

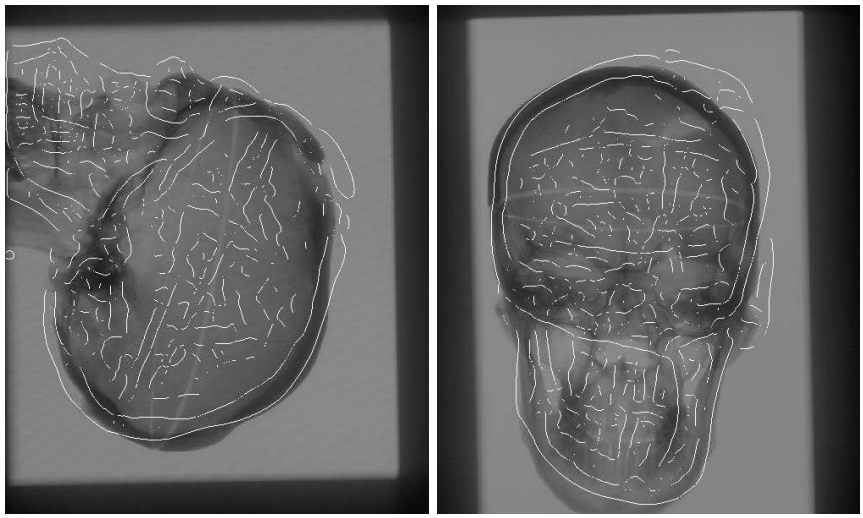


(a) With initial transformation estimate

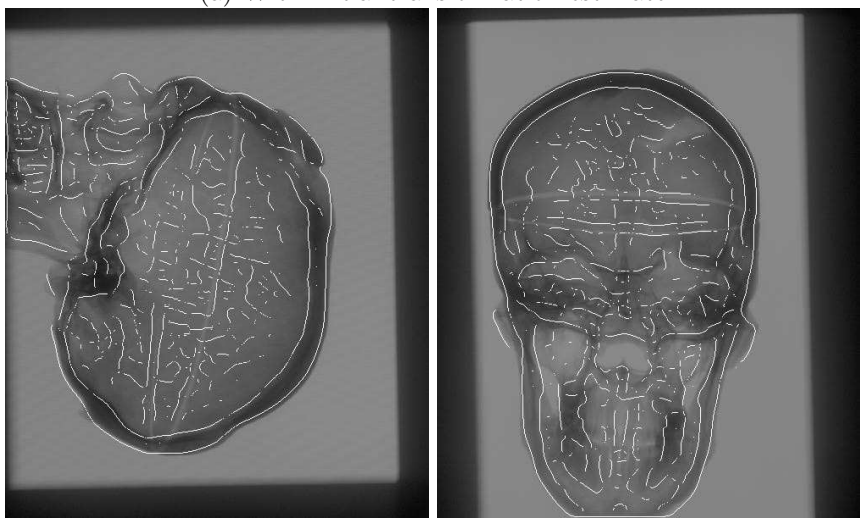


(b) With transformation estimate resulting from registration

Figure 4.9: Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pz* method.

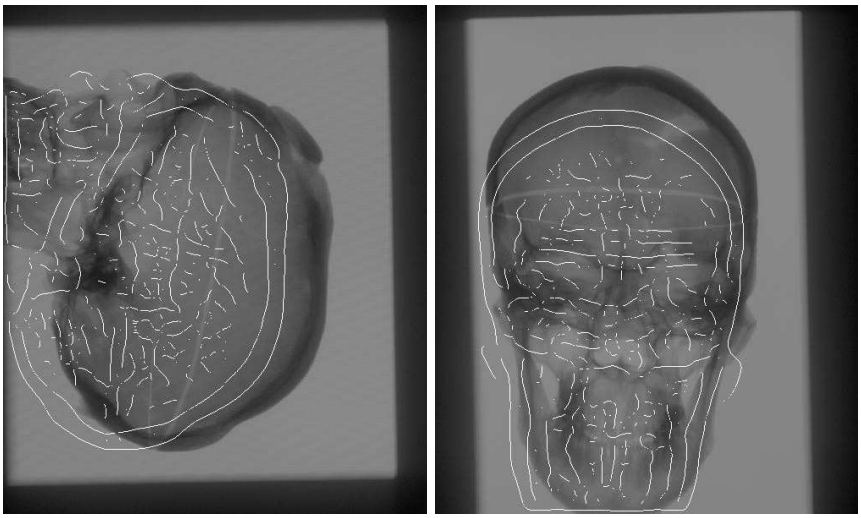


(a) With initial transformation estimate

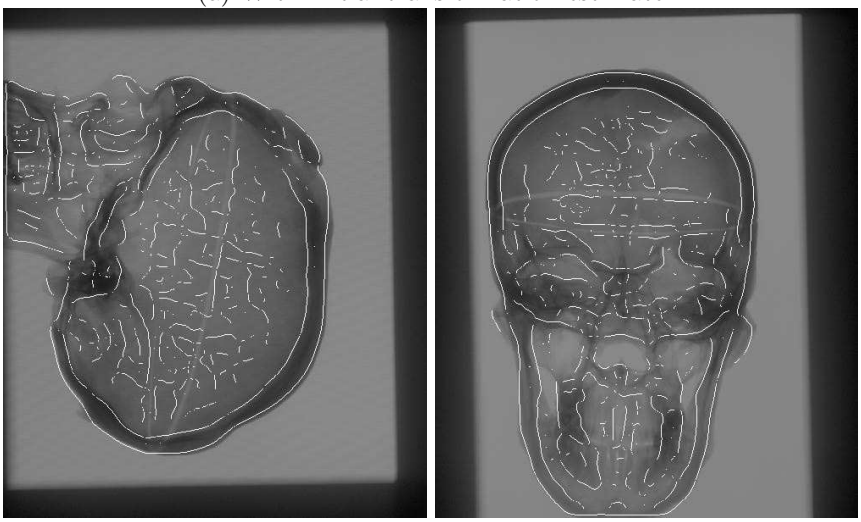


(b) With transformation estimate resulting from registration

Figure 4.10: Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pz* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images



(a) With initial transformation estimate



(b) With transformation estimate resulting from registration

Figure 4.11: Registration results of an experiment on real X-ray and CT of the Gage's skull dataset using the *Reg-Pow* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images

4.5.2 Experiments with Fluoroscopy of the Phantom Pelvis

In case of the experiments involving the phantom pelvis, we had a very difficult time finding the accurate ground truth pose even manually. The lateral images contain considerably less information about the position of the anatomy as the AP ones as images from the two opposite sides are almost identical.

Two other challenges involved were the fact that the fluoroscopic images of the phantom pelvis are greatly truncated (some parts of the ileum is missing and a black rim appears around the margin of the acquisition) and the pincushion distortion effects were not accounted for at the time of imaging.

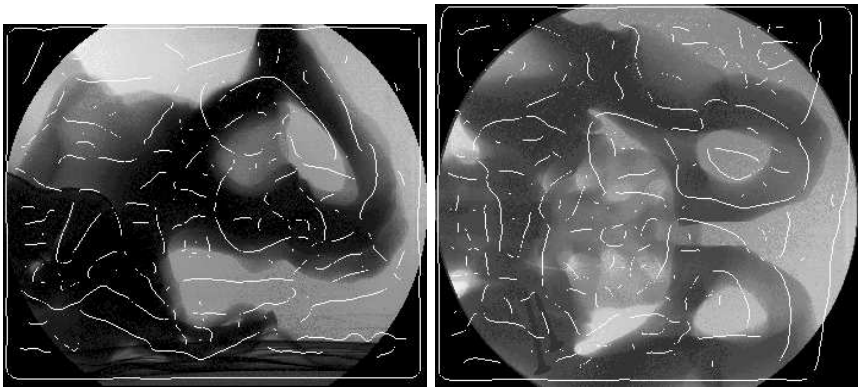
Hence, our results with respect to this anatomy are in very early stages. We present only qualitative results in this Section. Figure 4.12 presents results from an experiment with the *Reg-Pow* method and Fig. 4.13 shows the results of method *Reg-Hi*.

We can see that while at the outset of the algorithm the DRR outlines do not really fit the edges in the fluoro acquisitions, the edges at the final stage nicely match the boundaries of the observed images.

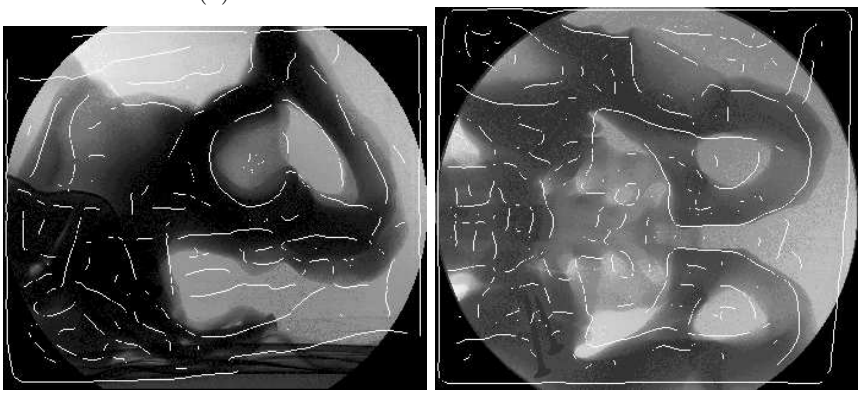
In case of the pelvis images, one should focus on matching object boundaries closer to the image centers as the warping effect is not as strong in that region as towards the image margins.

4.6 Summary

This chapter presented the experimental analysis of our newly proposed alignment procedures. We first characterized our objective function via probing experiments and gave an incentive for the use of a multiresolution registration framework. Then the performance of the various algorithms was tested both on CT-derived and real medical image datasets. In the controlled settings, the two gradient-based techniques produced very similar results both in terms of accuracy and running time. Their error range proved to be consistently lower than that of the standard Powell method. In case of the latter, running time was also a major handicap. When applied to high-resolution datasets the direction set method took several hours to converge to the optimal settings. We presented some preliminary results on real X-ray - CT registration, too. On Gage's skull dataset the *Reg-Pz* method performed the best. In case of the phantom pelvis dataset, we qualitatively evaluate our results. A quantitative analysis is missing because we have not yet included a distortion model in our registration techniques which would account for

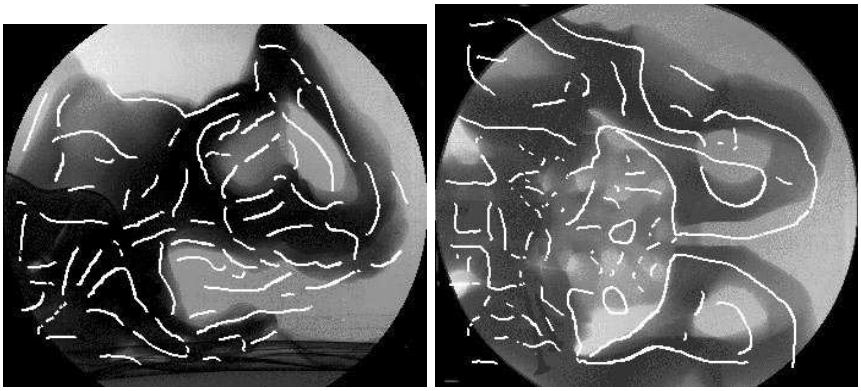


(a) With initial transformation estimate

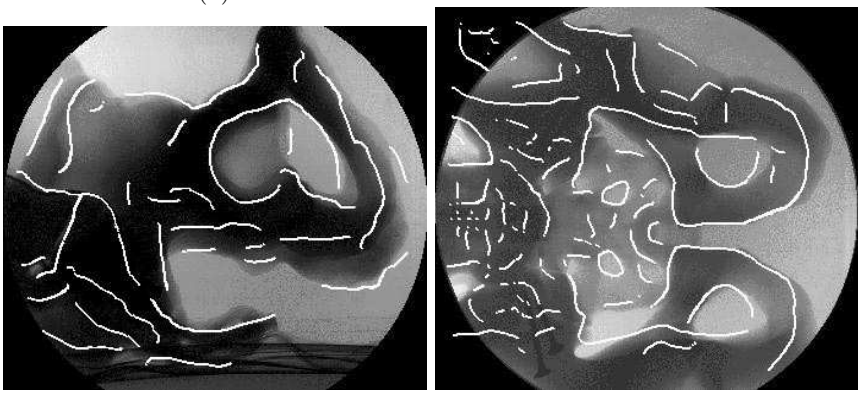


(b) With transformation estimate resulting from registration

Figure 4.12: Registration results of an experiment on real X-ray fluoroscopy and CT of the phantom pelvis dataset using the *Reg-Pow* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images.



(a) With initial transformation estimate



(b) With transformation estimate resulting from registration

Figure 4.13: Registration results of an experiment on real X-ray fluoroscopy and CT of the phantom pelvis dataset using the *Reg-Hi* method. Contours of the DRR images created by the output of the registration algorithm are overlaid on the original fluoro images.

the warping artifacts present in X-ray fluoroscopy.

Chapter 5

Concluding Remarks

5.1 Summary

We have introduced a novel framework to address the 2D-3D rigid-body registration problem between the medical image modalities of CT and X-ray fluoroscopy. The main objective of our task was to (efficiently) identify a geometrical transformation in 3D space that describes the relative position and orientation of coordinate systems of the imaging environment and the pre-operative data volume.

Our main contribution to the state of the art is demonstrating that the sparse-sampling approach of stochastic gradient ascent, which has been heavily used in 3D-3D registration, can also be applied to significant advantage in 2D-3D registration. Its usage obviates the need for the generation of full DRRs at each iteration, which typically requires a full traversal of the volumetric data. In contrast, our approach only samples approximately 100 ray traversals from the volume. We presented two such approaches and compared their performance to that of a standard method. All of the algorithms that we implemented apply the information theoretic measure called mutual information to evaluate the alignment quality of the input datasets given a transformation estimate. They only differ in the way they optimize that similarity value. The standard optimization method, Powell's method, carries out a series of one dimensional searches to determine the best settings, while the stochastic gradient ascent strategies guide the estimate-update procedure by computing partial derivatives of the objective function with respect to the transformation components. The difference between our two gradient-based strategies originates from the estimation of probability density functions. One of them uses Parzen windowing while

the other makes use of histogramming. In order to reduce the running time of these methods, they both operate on only a sparsely sampled version of the original inputs and approximate the true update terms.

We examined the registration characteristics of these algorithms by both running them on simulated and real datasets. The controlled experiments used images of a phantom spine, of a real and a plastic skull, of a real head and of a lumbar plastic spine segment, and the experiments using real X-ray images concerned datasets acquired of the pelvis and of the skull.

The experiments were run in a multiresolution setting, which not only proved to decrease execution time, but also increased the robustness of the algorithms.

5.1.1 Controlled Experiments

Gradient-based Techniques

In the controlled experiments the two gradient-based methods performed very similarly. They achieved highly accurate alignment results that were almost equivalent and their running times were not significantly different. We found, that on high resolution pyramid levels the histogramming method was more sensitive to code parameter settings and we had to experiment longer to identify proper values for such measures as the optimization step size and the sample size.

Powell's Method

The Powell method proved to be a robust optimizer, too. Although, the initial offset values were reduced swiftly on the lowest pyramid levels, the refinement of the transformation estimates on higher levels took an unreasonably long time. In some cases the experiments were even stopped before starting to operate on the original resolution of the input datasets as the running time exceeded several hours. The accuracy results for this method were also overruled by the outputs of the gradient-based techniques.

5.1.2 CT - X-ray Registration

For the experiments using real X-ray images, we showed promising, though somewhat preliminary, results. The Parzen windowing method

consistently aligned the input images according to the ground truth estimates. Powell's method also converged, but its resulting error terms were even higher in these experiments. The histogramming approach proved to be the least robust in this scenario. It only converged on a fairly small offset range. The full explanation for this phenomenon is still under investigation. We also demonstrated some initial results on aligning X-ray fluoroscopy and CT acquisitions of a phantom pelvis. We only presented a qualitative analysis of these experiments as the ground truth pose could not be determined with high accuracy and our framework did not model the radial/pincushion distortion effects present in the fluoroscopies. Comparing the major anatomy contours the registration results seem to be promising, but a quantitative evaluation was not prepared.

5.2 Future Research Questions and Ideas

In this final section, we describe several ideas that emerged while we worked on the presented 2D-3D rigid-body registration project. These are interesting thoughts or techniques that could either improve or complement the performance of the existing system.

5.2.1 Coupling Segmentation and Registration

Segmentation, along with registration, is another major area within medical image processing. Its goal is to identify and group together image components representing the same region or having similar properties based upon some specific criteria. Traditionally, registration and segmentation methods have been investigated in relative isolation, but many times the assumption is made that the solution to one is known in order to tackle the other. That statement clearly applies for feature-based registration algorithms. In their case (as discussed in Chapter 2), the alignment procedure only considers a few landmark points and identifying those depends on either some user interaction or some form of a segmentation technique.

Segmentation could benefit from registration when multiple acquisitions are processed at the same time. The relative correspondence between the targeted images contain useful information for completing the task. Executing the two procedures simultaneously is motivated by the hope that overall computation time could be saved by not having to complete one before the other and that the parallel registration and segmentation procedures can provide sufficient information to each

other in order to improve their individual performances.

We have started investigating[1] the problem of coupling intensity-based registration and segmentation. Our main motivation for involving an intensity-based method within such a framework is the aim to improve or extend the performance of already robust methods in more complex rigid-body or non-rigid registration problems. For example, the problem of registering spine datasets has long been a research challenge. Although the individual vertebrae behave as rigid bodies individually, the global movement of the spine cannot be described as that of a single rigid entity. A general approach to the problem is to piecewise align single vertebrae and then to describe the spine movement as some combination of these. That scenario could be closely related to the study by Penney et al., which reduced the computational burden of its 2D-3D registration algorithm by only considering a small portion of the original CT dataset [3] . The imaged anatomy (in their case, a single vertebra) was segmented out in the CT volume prior to any other processing. It was only then, that the alignment step took place with only the sub-volume considered.

Within our new approach the segmentation and registration stages would not have to be isolated, but they could operate dynamically in parallel. Even the hierarchical approach could be incorporated in that framework. This could allow the algorithm to focus more and more on the targeted (parts of the) anatomies as the resolution level is increasing. That application could be useful in addressing the 2D-3D registration problem as reducing the size of the image volume to be traversed when creating DRR images, could improve computation time significantly.

5.2.2 View and Number of Fluoroscopic Acquisitions

Another interesting question addresses the problem of how important the view is from which the fluoroscopic images are taken. It is an open question whether the registration results could be improved at all if these images were taken from a specific view and a given angle apart. Initial (synthesized) probing studies demonstrate that, in case of the skull, the location of the projection source has a very insignificant impact on the performance of the objective function. In case of the

[1] These ideas have been discussed with Tina Kapur (Chief Scientist, Visualization Technology Inc.) and Prof. Anthony Yezzi (Electrical and Computer Engineering Department, Georgia Institute of Technology).

pelvis, however, these results might be different as it has a much more irregular geometry.

We would also like to investigate the related problem of how many images are sufficient (and still practical) to accurately carry out the registration procedure. It would be very useful to know how much the robustness and the capture range of the algorithm would grow with a couple of more additional input projection images (without the computational costs overwhelming the potential benefits).

5.2.3 Defining Automatic Stopping Criterion for Gradient Optimization Protocols

An algorithm that would automatically identify convergence to the optimal settings would be of great value if incorporated into our current registration framework. The fact that we execute the alignment method a predetermined number of times could be quite inefficient if execution time is of key importance.

The complexity of this task originates from the fact that in case of our sampled stochastic approaches, the objective function evaluations are noisy estimates of the true values. Hence, any kind of stopping criterion would have to anticipate a large enough variance in the similarity values even in case of convergence.

5.2.4 Truncation/Limited Field of View

We would also like to carry out more experiments to study the robustness of our algorithms with respect to errors due to truncation and limited field of view. In case of mutual information, it has been known to be a weakness, but we would like to fully understand how much that problem surfaces in 2D-3D applications. This is an important question to be analyzed as in case of real procedures, intra-procedural images are likely to contain images of surgical/treatment tools that could partially obstruct the anatomy of interest.

5.2.5 Distortion Effects & Dewarping

We mentioned in Section 2.1.1 that one of the main disadvantages of using X-ray fluoroscopies is the geometrical distortion effects that are not corrected for at the time of image acquisition. Our current registration framework does not account for this problem, we have not implemented a dewarping algorithm. As the distortion effects can potentially be significant and qualitative evaluation of the algorithm on distorted images

is not truly meaningful, our preprocessing step could definitely benefit from this feature. The sensitive sparsely sampled histogramming approach could also be expected to have an improved performance with that addition.

5.2.6 Histogram Characteristics

We found that the histogramming approach worked quite well when the number of bins for both the 1D and 2D histograms was set to be 32. However, we did not experiment thoroughly with other histogram settings/specifications, so we do not really know how much influence the bin size would have on the registration results. We suspect that it is unlikely to obtain comparable results if the bin numbers were decreased (as that would make the probability density estimates even cruder) but increasing them might not increase the performance significantly either given the gray-scale image modalities. It also depends on the number of sample points that we use for the stochastic estimation.

As discussed in Section 4.5, the histogram approach, while promising in the controlled experiments, was not consistently successful in the experiments using real X-ray images. We would like to further characterize this behavior, both empirically and analytically, with an eye towards potentially useful modifications to the method.

5.2.7 Code Optimization

As of now, the C++ code that we are using for the 2D-3D registration task is not optimized. We initiate numerous IO operations to print out partial results, to create projection images at intermediate stages and we also do an additional function evaluation at each iteration of the gradient-based approach in order to be able to analyze the registration performance. It would be interesting to compare our algorithm's running time to the fastest state of the art solutions after eliminating all unnecessary computations and using aggressive implementations of inner loops.

5.2.8 Improving MI

Since mutual information has become a similarity measure applied in a wide range of applications, there has been several modifications suggested that would complement its performance in particular scenarios. Improved results have been reported by, for example, using normalized mutual information (NMI) [21, 44] and by another approach which

incorporates gradient information into the definition of its MI-based objective function [18]. In the latter, both magnitude and directional information are included. It has been also investigated how spatial information could be introduced in MI-based similarity measures. Such an attempt was to use higher level entropies in order to calculate mutual information [31].

We are also interested in improving the registration results achieved by mutual information. However, instead of modifying the objective function itself, we aim to investigate the independence properties of image pixels. In order to make the mathematical formulation of the registration problem manageable, we make the assumption that image components are independent and identically distributed (IID), or that there is no correlation in between them. Although that assumption is not valid, especially on the top of a multiresolution pyramid, a great portion of the image processing algorithms heavily rely on it. We would like to study the question of how to process our input images in order to more accurately satisfy the IID criterion, or else ways to relax the assumption.

Chapter 6

APPENDIX

6.1 Small Angle Rotation

This section provides the proof for why we could approximate the movement resulting from a small angle rotation as the vector sum when deriving the partial derivative of volume intensities with respect to the rotation parameters of transformation T. In Section 3.5.2, we assumed that

$$R_u(r, R(D(d, x))) \approx R(D(d, x)) + r \times R(D(d, x)).$$

In the following calculations, we use the *equivalent angle-axis* notation to parameterize rotation transformations. As opposed to the 4-element quaternion notation, here the rotation operator is encoded as a 3D vector. The length of the rotation vector equals to the rotation angle and its direction stands for the rotation axis (which is a vector of unit length).

Let's represent rotation in 3D with vector r, where rotation angle θ can be written as the magnitude of the vector $\theta = \sqrt{{r_x}^2 + {r_y}^2 + {r_z}^2}$ and the direction of r, $\hat{n}$ (a unit length vector), stands for the rotation axis. By Rodrigue's formula, the rotated version of a given 3D coordinate point can be written as

$$R(r, x) = x * \cos(\theta) + \hat{n}(\hat{n} \cdot x)(1 - \cos\theta) + \sin\theta(\hat{n} \times x)$$

$$\text{If } \theta \ll 1, \text{ then } R(r, x) \approx x + \theta(\hat{n} \times x) = x + r \times x$$

That result, if applied to the registration scenario of Chapter 3, does indeed produce

$$R_u(r, R(D(d, x))) \approx R(D(d, x)) + r \times R(D(d, x)).$$

6.2 The Story of Phineas Gage

In our experiments, we used the CT and X-ray fluoroscopic images of a real skull. As an interesting bit of information, it is the skull of Phineas Gage whose case has been studied during and well after his lifetime by neurologists. The following is an excerpt from an article revealing the accident that resulted in the profound change in Gage's life and explaining the intriguing questions that this event raised in neuroscience [39].

> *" On 13 September 1848, Phineas P. Gage, a 25-year-old construction foreman for the Rutland and Burlington Railroad in New England, became a victim of a bizarre accident....One the fateful day, a momentary distraction let Gage begin tamping directly over the powder before his assistant had had a chance to cover it with sand. The result was a powerful explosion away from the rock and toward Gage. The fine-pointed, 3-cm-thick, 109-cm-long tamping iron was hurled, rocket-like, through his face, skull, brain, and into the sky. Gage was momentarily stunned but regained full consciousness immediately thereafter. He was able to talk and even walk with the help of his men. The iron landed many yards away....Phineas Gage not only survived the momentous injury, in itself enough to earn him a place in the annals of medicine, but he survived as a different man, and therein lies the greater significance of this case."*
>
> (Science, Vol. 264; 20 May 1994)

Bibliography

[1] R. Bansal, L.H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath and J. Duncan A Novel Approach for the Registration of 2D Portal and 3D CT Images for Treatment Setup Verification in Radiotherapy *Proceedings of MICCAI 1998, p 1075-1086*

[2] R. Bansal, L.H. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath and J. Duncan Entropy-Based, Multiple-Portal-to-3D CT Registration for Prostate Radiotherapy Using Iteratively Estimated Segmentation *Proceedings of MICCAI 1999, p 567-578*

[3] G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, D.J. Hawkes A Comparison of Similarity Measures for Use in 2D-3D Medical Image Registration *IEEE Transactions on Medical Imaging* 17(4), 1998, pp. 586-595

[4] J. Weese, G.P. Penney, T.M. Buzug, C. Fassnacht, C. Lorenz 2D/3D registration of pre-operative CT images and intra-operative x-ray projections for image guided surgery *CAR '97 Computer Assisted Radiology and Surgery* Proceedings of the 11th International Symposium and Exhibition; Berlin, 25-28 June 1997

[5] J. Weese, G.P. Penney, P. Desmedt, T.M. Buzug, D.L.G. Hill, D.J. Hawkes Voxel-Based 2-D/3-D Registration of Fluoroscopy Images and CT Scans for Image-Guided Surgery *IEEE Transactions on Information Technology in Biomedicine* Vol. 1, No. 4, December 1997

[6] J. Weese, T.M. Buzug, C. Lorenz, C. Fassnacht An Approach to 2D/3D Registration of a Vertebra in 2D X-ray Fluoroscopies with 3D CT images *CVRMed-MRCAS'97, First Joint Conference Computer Vision, Virtual Reality and Robotics in Medicine*

*and Medical Robotics and Computer-Assisted Surgery*Grenoble, France, March 19-22, 1997

[7] J. Weese, T.M. Buzug, G.P. Penney, P. Desmedt 2D/3D Registration and Motion Tracking for Surgical Interventions *Philips Journal of Research* Vol 51, p. 299-316, 1998

[8] J. Weese, R. Gocke, G.P. Penney, P. Desmedt, T.M. Buzug, H. Schumann Fast voxel-based 2D/3D registration algorithm using a volume rendering method based on the shear-warp factorization

[9] T.M. Buzug,J. Weese,C. Fassnacht, and C. Lorenz Image registration: Convex weighting functions for histogram-based similarity measures *Proc. CVRMed/MRCAS* pp.203-212. Berlin, Germany: Springer-Verlag, 1997.

[10] T.S.Y. Tang, R.E. Ellis, G. Fichtinger Fiducial Registration from a Single X-ray Image; A New Technique for Fluoroscopic Guidance and Radiotherapy *Proceedings of MICCAI 2000, p 502-511*

[11] D. LaRose, L. Cassenti, B. Jaramaz, J. Moody, T. Kanade, A. DiGioia Post-operative Measurement of Acetabular Cup Position Using X-ray/CT Registration *Proceedings of MICCAI 2000, p 1104-1113*

[12] D. LaRose, J. Bayouth, T. Kanade Transgraph interactive intensity-based 2D/3D registration of X-ray and CT data

[13] D. LaRose Iterative X-Ray/CT Registration Using Accelerated Volume Rendering *Ph.D. Thesis* Carnegie Mellon University; May 2001

[14] E. Cosman,Jr. Rigid Registration of MR and Biplanar Fluoroscopy *Masters Thesis* Massachusetts Institute of Technology, 2000.

[15] P. Viola and W. Wells. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* 1997; 24:137-154.

[16] W. Wells, P. Viola, R. Kikinis. Multi-modal Volume Registration by Maximization of Mutual Information. *Medical robotics and computer assisted surgery 1995, Wiley-Liss, New York* pp. 155-162; 1995

[17] W. Wells.,M. Halle,R. Kikinis,P. Viola Alignment and Tracking using Graphics Hardware *Image Understanding Workshop, DARPA* pp. 837-842.; 1996.

[18] J.P.W. Pluim, J.B.A. Maintz a nd M.A. Viergever Image Registration by Maximization of Combined Mutual Information and Gradient Information *Proceedings of MICCAI 2000* p 452-461

[19] J.P.W. Pluim Mutual Information Based Registration of Medical Images *Ph.D. Thesis* University of Utrecht, 2000.

[20] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, G. Marchal Automated multi-modality image registration based on information theory *Proc. Information Processing in Medical Imaging Conf.; Kluwer Academic Publishers* pp. 263-274, 1995

[21] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens Multimodality Image Registration by Maximization of Mutual Information *IEEE Transactions on Medical Imaging, 16(2):187-198, 1997*

[22] F. Maes Segmentation and registration of multimodal medical images: from theory, implementation and validation to a useful tool in clinical practice *Ph.D. Thesis* K.U.Leuven, May 1998, Leuven, Belgium

[23] J.J. Craig Introduction to Robotics; Mechanics and Control *Addison-Wesley Publishing Company, 1986*

[24] J.A. Nelder, R. Meade *Computer Journal, 1965; vol. 7, pp.308-313*

[25] H. Goldstein Classical Mechanics *2nd ed. Reading, MA: Addison-Wesley, 1980*

[26] B. Widrow, M. Hoff Adaptive Switching Circuits *1960 IRE WESCON Convention Record* vol 4, p. 96-104; IRE, New York

[27] L. Ljung, T. Soderstrom Theory and Practice of Recursive Identification *MIT Press* 1983

[28] S. Haykin Neural Networks: A comprehensive Foundation *Macmillan College Publishing*

[29] P. Thevanaz, M. Unser Optimization of Mutual Information for Multiresolution Image Registration *IEEE Transactions on Image Processing* Vol. 9, No. 12, December 2000

[30] P. Yushkevich, D. Fritsch, S. Pizer, E. Chaney Towards Automatic , Model-Driven Determination of 3D Patient Setup Errors in Conformal Radiotherapy *Med. Phys.* 1999

[31] D. Rueckert, M.J. Clarkson, D.L.G. Hill, D.J. Hawkes Non-rigid registration using higher-order mutual information *Proc. SPIE Medical Imaging 2000: Image Processing* 2000, San Diego, CA

[32] P. Lacroute, M. Levoy Fast Volume Rendering Using a Shear-Warp Factorization of the Viewing Transformation *Proc. SIGGRAPH '94* Orlando, Florida, July, 1994, pp. 451-458

[33] L. Lemieux, R. Jagoe, D.R. FIsh, N.D. Kitchen, D.G.T. Thomas A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs *Med. Phys.* 21(11), November 1994

[34] L.M.G. Brown Registration of Planar Film Radiographs with Computed Tomography *IEEE; Proceedings of MMBIA* 1996

[35] M. Leventon, W.M. Wells III, W.E.L. Grimson Multiple View 2D-3D Mutual Information Registration *Image Understanding Workshop* 1997

[36] B.K.P. Horn Robot Vision *The MIT Press; Cambridge, MA* 1998

[37] E.H. Adelson, P.J. Burt Image Data Compression with the Laplacian Pyramid *Proceedings of the Pattern Recognition and Information Processing Conference* pp. 218-223, Dallas, TX 1981

[38] R.O. Duda, P.E. Hart, D.G. Stork Pattern Classification (2nd ed.) *John Wiley & Sons, Inc.* 1998

[39] H. Damasio,T. Grabowski,R. Frank,A.M. Galaburda,A.R.Damasio The Return of Phineas Gage: Clues About the Brain from the Skull of a Famous Patient *Science* New Series, Volume 264, Issue 5162 (May 20, 1994), 1102-1105

[40] W.T. Vetterling, S.A. Teukolsky, W.H. Press Numerical Recipes Example Book (C) *Cambridge Univ Press; 2nd edition* 1993

[41] Z.R. Yaniv Fluoroscopic X-ray Image Processing and Registration for Computer-Aided Orthopedic Surgery *Masters Thesis* The Hebrew University of Jerusalem, Israel; 1998.

[42] P. Viola Alignment by Maximization of Mutual Information *Ph.D. Thesis* Massachusetts Institute of Technology, 1995

[43] T.M. Cover, J.A. Thomas Elements of Information Theory *John Wiley* 1991.

[44] C. Studholme,D.L.G. Hill,D.J. Hawkes An overlap invariant entropy measure of 3D medical image alignment *Pattern Recognition* vol. 32, No. 1, pp.71-86, 1999.