

**Intensity based Image Registration using Robust Similarity  
Measure and Constrained Optimization: Applications for  
Radiation Therapy**

by

**Jeongtae Kim**

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Electrical Engineering: Systems)  
in The University of Michigan  
2004

Doctoral Committee:

Associate Professor Jeffrey A. Fessler, Chair  
Professor Alfred O. Hero  
Professor Charles R. Meyer  
Associate Professor Douglas C. Noll  
Professor Randall K. Ten Haken

## ABSTRACT

Intensity based Image Registration using Robust Similarity Measure and  
Constrained Optimization: Applications for Radiation Therapy

by

Jeongtae Kim

Chair: Jeffrey A. Fessler

In radiotherapy, an x-ray beam arrangement should be planned based on a correct patient model using a planning CT (Computed Tomography) and the x-ray beam should be focused accurately to implement the plan on actual patient. Since the patient model may not be accurate due to organ motion and there is patient set-up error, the actual delivery of the x-ray may differ from the optimal one intended by a physician, thus risking damage to normal tissues and possibly delivering a suboptimal radiation dose to the tumors.

Correct estimation of the patient set-up error and organ motion is important since one can retrospectively calculate the actual x-ray dose accumulation from the treatment using the estimated set-up error and organ motion. Moreover, if the set up estimate can be completed quickly before the treatment, then one can compensate for the set-up error by adjusting either the radiotherapy table or the x-ray beam position prior to treatment delivery. Also, if one can build a *dynamic model* of patient organ motions before treatment, the treatment plan can be established more accurately considering the motions.

Image registration is a very useful technique for estimating both patient set up and organ motion for radiation therapy. Patient set up may be estimated by 3D/2D image registration, which registers planning CT image onto radiograph images from the treatment room and organ motion from one time to another may be estimated using nonrigid image registration of two images from two time instances.

We investigated several rigid and nonrigid image registration methods that are useful for estimating patient set up positioning and organ motion. By conducting an experiment

with anthropomorphic chest phantom, we investigated the feasibility of 3D/2D registration methods for the set-up estimation. We achieved sub-voxel accuracy using two orthogonal projection images by the sample correlation coefficient based and the MI(Mutual Information) based methods.

We have proposed a novel robust image registration method based on a *robust correlation coefficient*, which is useful for registering images containing unexpected objects. Images from treatment rooms usually contain objects that are not present in the planning CT image, such as radiotherapy table. The statistical properties such as bias, variance and robustness of the proposed method in comparisons with the sample correlation and the MI based method have been analyzed.

We also investigated a novel nonrigid image registration method in which the estimated deformation obeys the physical constraint of positive Jacobian determinant. We derived a closed form expression of possible minimum and maximum Jacobian in terms of gradient bounds analytically. To enforce the gradient bounds of the deformation in optimization, we have introduced constraint sets in the parameter space. The optimization was accomplished using the gradient projection method with Dykstra's *cyclic projection* method.

*Key Words:* Radiotherapy, Set up error, Organ motion, Image registration, 3D/2D image registration, Robust correlation coefficient, Nonrigid image registration, Jacobian, Gradient projection method, Dykstra's cyclic projection.

© Jeongtae Kim 2004  
All Rights Reserved

*To my parents, my wife Youngju and my beloved daughter Grace Young Kim*

## ACKNOWLEDGEMENTS

During my Ph.D study in the University of Michigan, I have been privileged to have Prof. Jeffrey Fessler as my advisor. I am deeply grateful to Jeff for his insightful advice, warm encouragement, patience and kindness in discussing with me. He has shown me an exemplary scholar. Without him, my Ph.D study would have been much more difficult.

I also thank my colleagues from Radiation oncology department, Prof. Ten Haken, Prof. Balter and Prof. Lam. Not only they have funded me but also provided me priceless knowledges that make me understand clinical applications of medical imaging better. It has been a pleasant experience working with them.

I also thank my committee members, Prof. Hero, Prof. Meyer and Prof. Noll for their time and efforts to supervise my dissertation. Their valuable comments have made my dissertation better.

My gratitude is also with my friends who have shared difficult and joyful memories with me. Without their friendship, Ann Arbor would have been colder place to me.

The love, sacrifice and encouragement from my families has made it possible for me to complete the Ph.D study. My gratitude and love to my parents, my wife Youngju and my Grace is more than that can be expressed in words. I thank and praise God for all he has done for me.

This work was supported by NIH grant CA60711, PO1-CA59827 and RO1-CA81161.

# TABLE OF CONTENTS

<b>DEDICATION</b>		ii
<b>ACKNOWLEDGEMENTS</b>		iii
<b>LIST OF TABLES</b>		vi
<b>LIST OF FIGURES</b>		vii
<b>LIST OF APPENDICES</b>		ix
<b>CHAPTERS</b>		
1	INTRODUCTION	1
1.1	Radiation Therapy and Image Registration	1
1.2	Set up Estimation	3
1.3	Robust Image Registration	4
1.4	Building Patient Model using Nonrigid Registration	6
1.5	Outline of Dissertation	7
2	3D/2D Registration	8
2.1	Introduction	8
2.2	Materials and Methods	9
2.2.1	Experimental Methods	9
2.2.2	Technical Problem Formulation	10
2.2.3	MI-based Method	12
2.2.4	Correlation-based Method	13
2.2.5	Marker-based Method	14
2.3	Results	15
2.3.1	Radiographs	15
2.3.2	Estimated joint pdf	16
2.3.3	Position Estimation Results	17
2.4	Discussion	18
2.5	Conclusion	21
3	Robust Registration	27
3.1	Introduction	27
3.2	Theory: similarity measures	28
3.2.1	Correlation Coefficient Estimates	29

3.2.2	Robust Correlation Coefficient . . . . .	33
3.2.3	Mutual Information . . . . .	34
3.2.4	Analytical Comparisons . . . . .	35
3.3	Experimental Results . . . . .	37
3.3.1	1D Simulation . . . . .	37
3.3.2	2D MRI Registration . . . . .	44
3.3.3	3D/2D Phantom Study . . . . .	47
3.4	Discussion . . . . .	52
3.5	Conclusion . . . . .	55
4	Nonrigid Registration . . . . .	56
4.1	Introduction . . . . .	56
4.2	Theory . . . . .	58
4.2.1	Image Model . . . . .	58
4.2.2	Cubic B-spline based Deformation . . . . .	59
4.2.3	Invertibility . . . . .	60
4.2.4	Constraints in Parameter Space . . . . .	61
4.2.5	Constrained Optimization using Dykstra’s Algorithm . . . . .	63
4.2.6	Penalty Function . . . . .	65
4.2.7	Multi-resolution . . . . .	66
4.3	Experimental Results . . . . .	67
4.3.1	Image Data . . . . .	67
4.3.2	Synthetic Transformation . . . . .	67
4.3.3	Inhale and Exhale Image Registration . . . . .	75
4.4	Discussion . . . . .	78
4.5	Conclusion . . . . .	80
5	Conclusion and Future Research . . . . .	81
5.1	Conclusion . . . . .	81
5.2	Future Research . . . . .	82
	<b>APPENDICES . . . . .</b>	<b>84</b>
	<b>BIBLIOGRAPHY . . . . .</b>	<b>100</b>

## LIST OF TABLES

### Table

2.1	Estimated set-up error for Experiment A . . . . .	22
2.2	Estimated set-up error for Experiment B . . . . .	22
2.3	Estimated set-up error for Experiment C . . . . .	22
3.1	Mean (and STD) of estimated registration parameters for 2D-2D MRI registration. . . . .	48
3.2	Mean (and STD) of estimated set-up parameters for 3D-2D registration. . .	51
4.1	Characteristics of the synthetic transforms and estimated transforms . . .	74
4.2	Lung CT registration results . . . . .	76

## LIST OF FIGURES

Figure		
2.1	Radiographs . . . . .	23
2.2	Slices with markers . . . . .	24
2.3	Estimated mutual information with respect to the translation errors . . . . .	24
2.4	Non-registered and registered radiographs and DRRs . . . . .	25
2.5	Joint histograms from registered and nonregistered images (Numbers in X,Y axis represent histogram bin numer. 0 corresponds to the lowest intensity and 128 corresponds to the highest intensity. . . . .	26
3.1	Reference signal $s_1(\vec{t})$ (signal being translated to achieve registration). . . . .	38
3.2	Observed signal $s_2(\vec{t})$ with outliers and Gaussian noise (STD = 0.3). . . . .	39
3.3	Sample and robust correlation coefficients vs 1D translation. . . . .	40
3.4	Joint pdf at registered position (computed by $64 \times 64$ interpolated joint histogram). . . . .	40
3.5	Weighting function at registered position (evaluated at $64 \times 64$ joint histogram domain). . . . .	41
3.6	MI similarity measure ( $32 \times 32$ interpolated joint histogram) vs 1D translation. . . . .	41
3.7	Bias of the translation estimators vs Gaussian noise levels. . . . .	42
3.8	Standard deviations of the translation estimators vs Gaussian noise levels. . . . .	43
3.9	Standard deviations vs bias of the translation estimators (upper part for noise STD=0.4, lower part for noise STD=0.2). . . . .	43
3.10	Reference MRI image (with field inhomogeneity correction). . . . .	44
3.11	Target MRI image (without field inhomogeneity correction). . . . .	45
3.12	Contour plot of sample correlation coefficient vs vertical scaling and translation. . . . .	46
3.13	Contour plot of robust correlation coefficient vs vertical scaling and translation. . . . .	46
3.14	Contour plot of estimated MI ( $32 \times 32$ interpolated joint histogram) vs vertical scaling and translation. . . . .	47
3.15	Measured lateral radiograph for 3D-2D registration. . . . .	49
3.16	Lateral DRR computed from 3D CT volume. . . . .	50
3.17	Estimated joint pdf at the registered position ( $64 \times 64$ interpolated joint histogram). . . . .	50
3.18	Weighting function at the registered position (evaluated at $64 \times 64$ joint histogram domain). . . . .	51
4.1	Penalty function for Jacobian determinant . . . . .	66
4.2	A slice of inhale CT image. . . . .	67

4.3	A slice of exhale CT image. . . . .	68
4.4	Synthetic deformation $f(x, y, z)$ evaluated at slice 3. . . . .	69
4.5	Synthetic deformation $g(x, y, z)$ evaluated at slice 3. . . . .	69
4.6	Synthetic deformation $h(x, y, z)$ evaluated at slice 3. . . . .	70
4.7	Estimated deformation $\hat{f}(x, y, z)$ evaluated at slice 3 . . . . .	71
4.8	Estimated deformation $\hat{g}(x, y, z)$ evaluated at slice 3. . . . .	71
4.9	Estimated deformation $\hat{h}(x, y, z)$ evaluated at slice 3. . . . .	72
4.10	Average error vs iteration (Synth.1). . . . .	73
4.11	Average error vs iteration (Synth.2). . . . .	73
4.12	Minimum Jacobian determinant vs iteration (Synth.2). . . . .	74
4.13	Estimated deformation $\hat{f}(x, y, z)$ . . . . .	75
4.14	Estimated deformation $\hat{g}(x, y, z)$ . . . . .	76
4.15	Estimated deformation $\hat{h}(x, y, z)$ . . . . .	77

## LIST OF APPENDICES

### APPENDIX

A	Fisher Information . . . . .	85
B	Mean and Covariance Approximation . . . . .	86
C	Bias and Robustness of Correlation based Methods . . . . .	88
D	Efficiency . . . . .	90
E	Gradient of Estimated joint pdf . . . . .	92
F	Proof of Proposition 1 . . . . .	94
G	Proof of Proposition 2 . . . . .	98

# CHAPTER 1

## INTRODUCTION

### 1.1 Radiation Therapy and Image Registration

The goal of external beam radiotherapy is to irradiate a tumor to a high x-ray dose while sparing normal tissues as much as possible. To achieve this goal, the x-ray beam arrangement and the dose distribution are carefully planned based on the target tumor position within an individualized patient model. Each 3D patient model is constructed using a 3D planning x-ray CT (Computed Tomography) scan or 2D x-ray scan that is acquired several days before treatment.

Accurate radiotherapy requires the patient to be set up for treatment in a coordinate system that is consistent with the treatment plan. In practice, however, patient set-up errors occur despite the use of laser alignment. Furthermore, patient anatomy may change relative to the treatment plan due to breathing, organ movements, etc. Also, since the patient model is updated sporadically, the treatment plan may not be optimal [1]. Practically, it is difficult to update patient model frequently since it requires extra CT scan. Such errors continue to be a concern in radiotherapy not only due to the unnecessary irradiation of normal tissues but also due to the sub-optimal irradiation of the target tumor [2–4].

If the patient set-up error and anatomy changes can be correctly estimated after each treatment session, one can retrospectively calculate the actual x-ray dose accumulation from the treatment and review the patient set-up procedure. Moreover, if the set up estimate can be completed quickly before the treatment, then one can compensate for the set-up error by adjusting either the radiotherapy table or the x-ray beam position prior to the treatment delivery.

Many studies have been conducted to quantify the statistics of the patient set-up error [5], to consider the effect of the set-up error in planning [1], to estimate the set-up error accurately, to reduce the set-up error and/or to compensate for the set-up error by adjusting the x-ray beam position or the patient position [2–4, 6–11].

Image registration is a very useful technique for estimating both patient set up and anatomy change for radio therapy [8, 11, 12]. Patient set up may be estimated by 3D/2D image registration, which registers planning CT image onto radiograph images from the treatment room. In general, the patient set up estimation is accomplished by a rigid image registration method.

Compared with the patient set up estimation problem, the patient anatomy change estimation problem requires a nonrigid image registration method since the anatomy change occurs in nonrigid fashion. The patient anatomy change from one time to the other can be estimated using nonrigid registration of two images from two different time instances. As mentioned before, if the patient anatomy change can be estimated accurately after the treatment, it can be used effectively to compute accumulated dose distribution from the treatment. Furthermore, if one can build a *dynamic model* of patient organ motion before treatment, treatment plan can be established more accurately with the consideration of the motion [13].

For example, a simple *dynamic* patient model of lung motion can be built by the nonrigid registration of inhale and exhale CT images. Using the estimated deformation by the registration, one may build a deformed organ model at any moment during the breathing cycle by interpolating the estimated deformation field.

In this dissertation, we investigate several rigid and nonrigid image registration methods that are useful for estimating patient set up and organ motion. We investigate 3D/2D image registration methods for the set up estimation and propose a novel robust image registration method, which is useful to register the images containing unexpected objects. We also investigate nonrigid registration methods and propose a novel nonrigid image registration method in which estimated deformation conforms the physical constraint of invertibility [14]. We apply the proposed nonrigid image registration method for building a lung motion model.

## 1.2 Set up Estimation

Technically, the patient setup error estimation problem is the estimation of patient pose difference in the treatment room from the planning CT. Similar applications arise in the area of image-guided surgery that requires the mapping of positions in pre-operative data to a coordinate system defined in the operating room [11, 15].

The patient setup is usually estimated by an image registration method that registers 2D radiographs acquired during patient set-up to 2D simulated images or 3D planning CT image at the desired position. In most cases, the 2D simulator images or the 3D planning CT image is geometrically transformed to achieve the registration, and the patient set-up is estimated as the geometrical transformation that accomplishes the registration.

Although approaches that use 2D simulator images have the advantage of fast computation time, those are not very accurate since the 3D transformation is estimated by registering 2D images in a 2D plane [8]. It has been reported that these methods have problems in estimating rotations in planes non-parallel to the 2D radiographs planes [16]. Approaches that register 3D planning CT image onto the 2D radiographs overcome this limitation at the expense of greater computation.

Several groups have investigated 3D/2D registration methods [8] [15–20]. Those methods can be classified into feature-based methods [16, 17, 19] and intensity-based methods using DRRs (Digitally Reconstructed Radiographs) [15, 18, 20]. Feature-based methods use anatomical or artificial landmarks segmented from the images to perform the registration. Because of the segmentation, the feature based methods have several drawbacks. For example, segmentation procedures often require skilled human interaction, thereby making the procedure difficult to automate. Automatic procedures for segmentation have been studied [17]. Even if one designs an automatic segmentation algorithm, segmentation errors may occur and cause setup estimation errors. However, since feature-based methods have the advantage of fast computation, those are widely investigated for setup error estimation [16, 17, 19].

Unlike feature-based methods, intensity-based methods using DRRs (Digitally Reconstructed Radiographs) do not require segmentation. The DRRs, which are computed 2D projection images of the planning CT volume, are computed at several angles and compared to the radiographs from the same angles. The registration is achieved by maximizing

a similarity measure based on the intensity values of such DRRs and the radiographs. Since intensity-based methods do not require segmentation, those can be easily automated. In addition, the segmentation error can be avoided. However, these methods require much more computation than feature-based methods, thus posing a significant challenge for clinical application.

The investigations in this dissertation are focused on the intensity based image registration methods since we prefer to develop an automatic and accurate method. We believe that the computation time will be eventually reduced by faster computer hardware, storing pre-computed DRRs with appropriate interpolation, etc.

We have investigated set up error estimators experimentally first by conducting an experiment with an anthropomorphic chest phantom since we preferred to investigate the feasibility of existing image registration methods to the 3D/2D registration problem and to identify practical technical issues through the experiment. Particularly, we were interested in the statistical properties of the registration methods and the number of views required to achieve registrations within acceptable error bounds.

Similarity measures such as sample correlation coefficient [15, 18, 20], MI(Mutual Information) [21–23] have been investigated. The detail methods and results are summarized in Chapter 2.

### 1.3 Robust Image Registration

We have investigated a novel similarity measure motivated by the experimental results. Since the images being registered for patient set up estimation often have *outlier* image samples due to the presence of unexpected objects such as a radiation therapy table, the registration accuracy can be degraded by the outliers.

To design a method that is robust to such outlier samples, we have investigated a novel similarity measure. One fundamental design criterion is that the similarity measure should be maximized at the true registered position in the absence of noise. Establishing this characteristic analytically is challenging since the behavior of the objective function depends on the nature of the images being registered. Another important criteria is the statistical *efficiency* of the registration method, *i.e.*, the variability that results from repeating the registration with identical images except for noise. In addition, registration methods can

differ in their *robustness* to the presence of unexpected objects in images.

Many intensity-based image registration methods implicitly treat the intensity pairs taken from corresponding spatial locations in two images as i.i.d. (independent and identically distributed) samples of two random variables. With that assumption, statistical concepts such as correlation, joint entropy and mutual information (MI) are used as similarity measures by estimating those statistical properties from the i.i.d. samples.

The correlation coefficient is a particularly popular similarity measure, and is a natural choice when registering two images from the same modality [18, 20]. Although correlation is poor similarity measure for multi-modality image registration, in terms of statistical efficiency and computational efficiency, the correlation coefficient is one of the best similarity measures for intra-modality image registration. Since image registration for set-up estimation in radiation therapy and image-guided surgery often involves images from the same (or similar) modality, the correlation coefficient can be useful for those applications.

The sample correlation coefficient has been used widely to estimate the correlation coefficient due to its simplicity. However, a drawback of the sample correlation coefficient is its sensitivity to outliers [24, p. 199]. Even a few outliers can affect the sample correlation coefficient greatly and thus degrade image registration performance. A significant number of “outliers” may be present in the image-guided surgery application due to the presence of operational instruments and in the radiation therapy application due to the effect of radiotherapy table [12]. For X-ray CT images, differences in contrast agents also occur. Although a bias in estimating the correlation coefficient need not directly imply a bias in image registration, we have observed such biases empirically when outliers are present [25].

Another widely used similarity measure is the estimated MI. The MI is a particularly useful similarity measure for multi-modality image registration since it does not assume any functional relationship between the two image values [21–23]. In this sense, the MI method has an inherent degree of robustness. However, as illustrated by our empirical results in Section 3.3 and analyses in the Appendices for intra-modality image registration, the robustness of the MI method depends on images being registered. Moreover, the MI method may not be statistically efficient, *i.e.*, the registration variability due to noise can be larger than the sample correlation coefficient.

To overcome the drawbacks of the sample correlation method and the MI method, we have investigated an image registration method that uses robust correlation coefficients [24,

p. 204] as a similarity measure, thereby improving the robustness without compromising the statistical efficiency much.

Detail algorithm, analytical and experimental comparisons of the proposed method with the sample correlation and the MI based method are summarized in Chapter 3.

Also, the analyses of the statistical properties of the correlation-based and MI-based image registration methods by approximating the mean and the variance using first-order Taylor series expansions [26] are presented in appendices. Since image registration is highly nonlinear and the objective function is an implicit function of the images, it is challenging to obtain concise and insightful results from such approximations. Nevertheless, we summarize some theoretical arguments that complement the empirical results.

## 1.4 Building Patient Model using Nonrigid Registration

Compared with the set up estimation problem, the organ motion estimation has not been studied intensively. Recently, estimating patient deformation to compute more accurate accumulated dose distribution, building a dynamic model and considering the deformation in planning have been investigated [13,27]. To achieve those goals, nonrigid image registration is a useful technique.

In general, nonrigid registration is achieved by deforming one image using a model based deformation and the *true* deformation is estimated as the model based deformation that achieves the registration. One important physical constraint of the estimated deformation is that the Jacobian determinant should not be zero due to the *inverse function theorem* [14] since physical deformations are invertible. Moreover, since the Jacobian determinant is positive (*i.e.* unity) where the transformation is identity, if we assume that there is a region with the identity transformation in the images, the Jacobian determinants should be positive everywhere since the determinant is continuous in the spatial domain.

To prevent the Jacobian determinant from being negative, a regularization penalty function has often been introduced in nonrigid registration. For example, penalizing roughness penalty [28], bending energy [29], and negative Jacobian [30] have been proposed and investigated.

There are drawbacks in penalty function based methods. First, even though the penalty function based methods can reduce the risk of negative Jacobian determinant significantly,

there is no guarantee that the resulting deformation always has a positive Jacobian determinant, at least theoretically. Second, regularization parameter may need to be tuned differently for different images. Finally, since the penalty function is usually computed using the deformation values only at grid points, there is no guarantee that the estimated deformation is invertible in between the grids even if the Jacobian determinants are positive at the grids.

To remedy those problems, we propose sufficient conditions that ensure positive Jacobian determinants of the estimated deformation. We also propose a constrained optimization method to maximize a similarity measure subject to the conditions.

Chapter 4 presents the theory and the experimental results of the proposed constrained optimization method. Two propositions, which are sufficient conditions for positive Jacobian determinant and conditions in the parameter space to ensure gradient bound everywhere, are also presented in Chapter 4. Proofs of the propositions can be found in appendices.

## 1.5 Outline of Dissertation

The outline of this dissertation is summarized as follows. In Chapter 2, the methods, results and discussions for 3D/2D image registration using an anthropomorphic chest phantom are presented. Chapter 3 reviews the theoretical backgrounds of intensity based image registration methods and the robust estimation of correlation coefficients. Experimental results of the sample correlation, the robust correlation and the MI based image registration methods using 1D synthetic signals, 2D MRI images and 3D CT/2D radiographs are presented. Chapter 4 reviews the technical issues of nonrigid registration methods and presents a novel nonrigid image registration method. Experimental results of the proposed method using inhale and exhale lung CT images are also presented. Sufficient conditions to ensure positive Jacobian determinant are also claimed in Chapter 4. Discussions and future research directions are presented in Chapter 5. Mathematical derivations of Fisher information from Chapter 2, of mean and variance from Chapter 3 are presented in Appendix A, B, C and D. The proofs of two claims from Chapter 4 are included in Appendix F and G.

## CHAPTER 2

### 3D/2D Registration

#### 2.1 Introduction

We investigate an approach for estimating the set-up error based on the 3D/2D, intensity-based registration method using DRRs. We have chosen a MI (Mutual Information) registration criterion since it is robust to the intensity differences between two images. In image registration for radiotherapy, the radiographs and the DRRs have different intensities since the radiographs and the CT are generated by x-ray sources with different spectra. Moreover, other effects such as different scalings between two imaging devices, beam hardening, scattering and the radiotherapy table also cause intensity differences. The MI-based image registration method has been successful in 3D/3D multi-modality image registration [21–23]. By adopting the MI-based image registration technique, we design a fully automatic and accurate estimator.

Despite the computation time issues, we have chosen the intensity-based method using DRRs since we want to design an automatic and accurate estimation method. Moreover, we believe that the computation time will be eventually reduced by faster computer hardware, storing pre-computed DRRs with appropriate interpolation, etc.

To evaluate the performance of the MI-based estimator, we conducted an experiment with an anthropomorphic chest phantom. We placed 11 radiopaque markers on the phantom, enabling determination of the “ground truth” set-up error by registering the positions of the markers in the DRRs and the radiographs. We evaluated the performance of the MI-based estimator by comparisons with the marker-based method since we expected the latter to be more accurate in this experiment.

The structure of this chapter is as follows. In Section 2.2, we formulate the technical problem of the set-up error estimation. We also report the materials that were used for the experiment. In Section 2.3, we present experimental results of the proposed MI-based method in comparison with a well-known correlation-based method [20]. In Section 2.4, we discuss the proposed method and suggest future research directions.

## 2.2 Materials and Methods

### 2.2.1 Experimental Methods

We attached eleven 1mm diameter lead markers to the exterior surface of an anthropomorphic phantom to help establish the “ground truth” set-up error. By placing markers on locations that would be imaged around the boundaries of the projection views, we could apply the MI-based method using just the center portions of the projection images excluding the markers.

A  $512 \times 512 \times 398$  voxel planning CT image with  $0.9375 \times 0.9375 \times 1$  mm spacing was acquired on a GE CT/i scanner with a 140 kV x-ray source. Tattoos were drawn on the phantom where three alignment laser planes crossed the phantom to facilitate consistent set-up in a treatment room.

Next, the phantom was moved to the treatment room and it was set up at the isocenter by manually aligning tattoos to three laser planes in the treatment room. Four radiographs were obtained from different angles by rotating the x-ray source and Varian Portal Vision amorphous silicon active matrix flat panel image detector in  $30^\circ$  increments around the Z-axis as in Figure 2.1. For each of the  $0^\circ$  and  $90^\circ$  views, we acquired 10 repeated radiographs without realignment for evaluating the effect of noise on the estimator. The x-ray source voltage was 6 MV and the detector size was  $512 \times 384$  pixels with  $0.78\text{mm} \times 0.78\text{mm}$  spacing. We used only two radiographs from  $0^\circ$ ,  $90^\circ$  (i.e, AP and lateral images) for the MI-based and the correlation-based method. However, to enhance the accuracy of the “ground truth”, we used all four radiographs for the fiducial marker-based method.

We acquired additional radiographs in the treatment room after manually moving the phantom to each of two other positions for testing the robustness of the proposed method to different set-up errors. The experiments for the three different phantom positions are

called Experiment A, B and C, respectively in subsequent sections.

We applied three methods for estimating the set-up error: the MI-based method, the correlation-based method, and the marker-based method. For the MI-based method and the correlation-based method, the planning CT image was down-sampled by four along each axis to reduce computation time and memory usage. Since our planning CT had finer axial sampling than typical CTs in clinical use, the down sampling yields more representative CT quality in axial sampling. However, it also caused slightly coarser in-plane sampling.

The down-sampling was implemented by averaging the nearest  $4 \times 4 \times 4$  voxel values. As a result, the planning CT that was used to compute DRRs had  $3.750 \times 3.750 \times 4$  mm spatial spacing. However to preserve the accuracy of the “ground truth”, we did not down-sample the CT for the marker based method.

For the MI-based method and the correlation-based method, we used only the central  $400 \times 300$  sub-image of each of the DRRs and the radiographs to avoid the effect of the markers which are not usually used in clinical practice.

We have established the geometry of the EPID imaging systems by determining radiation field edges using simple thresholding method [31]. We assumed that the distance from source to detector known during calibration is correct.

For numerical search, we used the *Nelder-Mead simplex* method for all three methods [32]. We started the search from nominal unrotated and untranslated position. Termination criteria for the simplex method were that both the sum of the six parameter changes was less than 0.1, 0.1,  $10^{-12}$  and the maximum difference of the objective function within the simplex was less than 0.0001, 0.0001,  $10^{-12}$  for the MI-based, the correlation-based, and the marker-based method, respectively. Units for the rotation parameters were degrees and the translational parameters were mm.

### 2.2.2 Technical Problem Formulation

For the phantom experiments, we assume that the set-up errors are generated by rigid body motion and there is no change in the attenuation coefficients from the time of the planning CT scan. With those assumptions, we can model each voxel intensity value  $U_k$  of the planning CT with a mono-energetic source approximation as follows:

$$U_k = \alpha \mu(\vec{x}, \varepsilon_{kV}), \quad k = 1, 2, \dots, M, \quad (2.1)$$

where,  $\mu(\vec{x}, \varepsilon)$  is 3D attenuation coefficients distribution,  $\vec{x}$  is 3D spatial index,  $\varepsilon$  is the photon energy,  $\varepsilon_{kV}$  is the effective energy of the x-ray source for the planning CT,  $\alpha$  is the constant scaling of the x-ray CT, and  $M$  is the number of CT voxels.

The radiographs acquired in the treatment room correspond to the projection images of the rotated and translated attenuation coefficients based on MV spectrum source since the MV treatment beam is also used for the imaging. Making a mono-energetic source approximation and neglecting scattered radiation, we model the measured pixel intensity values of the radiographs as follows:

$$\tilde{Y}_k = I_o e^{-\int_{L_k} \mu(T_{\tilde{\theta}}(\vec{x}), \varepsilon_{MV}) d\vec{l}} + N_k, \quad k = 1, 2, \dots, N \quad (2.2)$$

$$\tilde{\theta} = [\phi_x \phi_y \phi_z t_x t_y t_z] \quad (2.3)$$

$$T_{\tilde{\theta}}(\vec{x}) = R(\phi_x)R(\phi_y)R(\phi_z)\vec{x} + [t_x t_y t_z]', \quad (2.4)$$

where  $\tilde{Y}_k$  is the measured intensity value of  $k$ th detector pixel,  $I_o$  is the MV x-ray source intensity,  $L_k$  is the x-ray path from source to  $k$ th detector pixel,  $T_{\tilde{\theta}}$  is the translate-rotation transform with parameters  $\tilde{\theta}$ ,  $\varepsilon_{MV}$  is the effective source energy for the radiographs,  $N_k$  is measurement noise,  $N$  is the number of detector pixels,  $R(\cdot)$  is a  $3 \times 3$  rotation matrix,  $\phi_x, \phi_y, \phi_z$  are the unknown rotations around X, Y, Z axes, and  $t_x, t_y, t_z$  are the unknown translations along X, Y, Z axes from the planning CT scan to treatment.

To estimate  $\tilde{\theta}$ , we compute the DRRs of the planning CT transformed according to any given guess for the parameter  $\theta$  using the following formula:

$$X_k(\theta) = \int_{L_k} \mu(T_{\theta}(\vec{x}), \varepsilon_{kV}) d\vec{l}, \quad k = 1, 2, \dots, N, \quad (2.5)$$

where  $X_k$  is the value of  $k$ th pixel in DRR.

In practice, the DRRs are computed by approximate summations for the line integral using the values  $U_k$ . We have implemented the line integral by computing the ray crossing lengths

within each voxel and summing over the lengths multiplied by voxel values [33]. The performance of the MI-based method and the correlation-based method are unaffected by global constant scaling differences.

The technical problem of set-up error estimation is estimating the six elements of the registration parameters  $\tilde{\theta}$  using the intensity values of the radiographs and the planning CT image. If we define  $Y_k = -\log(I_o/\tilde{Y}_k)$  and ignore the measurement noise,  $Y_k$  is the measured line integral of the attenuation coefficient based on mega-voltage spectra as following:

$$Y_k = \int_{L_k} \mu(T_{\tilde{\theta}}(\vec{x}), \varepsilon_{MV}) d\vec{l}, \quad k = 1, 2, \dots, N. \quad (2.6)$$

Then, the problem becomes estimating  $\tilde{\theta}$  using  $Y_k$  and  $X_k$ , which is the computed line integral of the attenuation coefficients with kilo-voltage source spectra, as defined in (2.5).

If there were a functional relationship between  $Y_k$  and  $X_k$ , we could use the MLE (Maximum Likelihood Estimator) which has many desirable properties [34]. Moreover, if there were a linear relationship, we could use successfully the well-known correlation-based method [20]. However, since the DRRs and the radiographs do not have a exact functional relationship, we would like to use a similarity measure that is robust to the intensity differences between the two images. Note that if noise is not ignored,  $Y_k$  is a Poisson type random variable. The Fisher information matrix for estimating  $\tilde{\theta}$  from the noisy observation is presented in Appendix A.

### 2.2.3 MI-based Method

Like many other intensity-based image registration methods using statistical analysis, the MI-based method conceptually considers the values of pixels in an image as samples of some random variable. Likewise, we can define a joint pdf (probability density function) of two random variables based on the corresponding pixel values in two images.

The idea of the MI-based method is that the two random variables are less jointly random if two images are more registered. When registered, pixels that have the same intensity value in one image will correspond to a more clustered distribution of the intensity values in the other image. This observation need not to be limited to the case that the intensities of the corresponding pixels are clustered around single value. There can be two or more clusters around different values. If we estimate the joint pdf from corresponding pixels in two images,

the conditional pdf becomes more clustered as the two images become better registered.

Based upon this observation, the MI-based method achieves registration by minimizing the estimated conditional entropy given the random variable which represents the transformed image or equivalently maximizing the estimated mutual information between two images [21–23].

We treat the pixel intensity values of the DRRs  $X_k(\theta)$  in (2.5) from AP image as samples of a random variable  $X_{AP}$  and those from lateral image as samples of random variable  $X_{LA}$ . We also consider the pixel values of the radiographs  $Y_k$  from AP and lateral images as samples of a random variable  $Y_{AP}$  and  $Y_{LA}$ .

To estimate the mutual information, we first estimate the joint pdf of  $(X_{AP}, Y_{AP})$  and  $(X_{LA}, Y_{LA})$  by using the samples  $X_k$  and  $Y_k$ . We used a  $128 \times 128$  bin joint histogram to estimate the joint pdf. The estimated mutual information is then computed from the estimated joint pdf by the definitions of the entropy and the mutual information [22, 23, 35]. The set-up error is estimated by seeking  $\theta$  which maximizes the sum of the two estimated mutual information as follows,

$$\hat{\theta}_{\text{MI}} = \arg \max_{\theta} \hat{I}_{\theta}(X_{AP}, Y_{AP}) + \hat{I}_{\theta}(X_{LA}, Y_{LA}) = \arg \min_{\theta} \hat{H}_{\theta}(Y_{AP}|X_{AP}) + \hat{H}_{\theta}(Y_{LA}|X_{LA}) \quad (2.7)$$

where  $\hat{I}_{\theta}(X_{AP}, Y_{AP})$ ,  $\hat{I}_{\theta}(X_{LA}, Y_{LA})$  are the estimated mutual information between random variable  $(X_{AP}, Y_{AP})$ ,  $(X_{LA}, Y_{LA})$ , and  $\hat{H}_{\theta}(Y_{AP}|X_{AP})$ ,  $\hat{H}_{\theta}(Y_{LA}|X_{LA})$  are the estimated conditional entropy of  $Y_{AP}$  given  $X_{AP}$ ,  $Y_{LA}$  given  $X_{LA}$ .

To solve this nonlinear optimization problem numerically, we applied the *Nelder-Mead simplex method*.

#### 2.2.4 Correlation-based Method

For comparison purposes, we also applied a correlation-based method. Although the intensity scales of the MV and the kV X-ray images are not exactly linearly related in theory, we may expect this method to work well since if MV attenuation is large, so is kV attenuation.

Moreover, one may try to improve the performance of the correlation-based method by

computing more MV-like DRR. The MV attenuation coefficients may be computed from kV attenuation coefficient by classifying voxels into different compounds and referencing typical MV attenuation coefficients of those compounds [20].

Since generating MV-like DRR not only requires such additional procedures but also may require periodic calibration of the procedures due to changes of the CT scanner characteristics, this approach may not be practical. Moreover, the performance improvement may be limited by other factors such as beam hardening effects, scattering effects, presence of the unmodeled radiotherapy table, etc.

We compute the correlation between a DRR and a radiograph of an AP image as follows:

$$\hat{\rho}_\theta(X_{AP}, Y_{AP}) = \frac{\sum_{k=1}^N (X_k(\theta) - \bar{X}_k(\theta))(Y_k - \bar{Y}_k)}{[(\sum_{k=1}^N (X_k(\theta) - \bar{X}_k(\theta))^2)(\sum_{k=1}^N (Y_k - \bar{Y}_k)^2)]^{1/2}}, \quad (2.8)$$

where  $X_k(\theta)$  is the  $k_{\text{th}}$  pixel value of AP DRR,  $\bar{X}_k(\theta)$  is the mean of  $X_k(\theta)$ ,  $Y_k(\theta)$  is the value of  $k_{\text{th}}$  pixel in AP radiograph,  $\bar{Y}_k$  is the mean of  $Y_k$ , and  $N$  is the number of pixels.

Similarly, we can compute the correlation between a DRR and a radiograph from lateral images. The set-up error is estimated by maximizing the sum of the estimated correlation coefficients from AP and lateral images as follows,

$$\hat{\theta}_{\text{co}} = \arg \max_{\theta} \hat{\rho}_\theta(X_{AP}, Y_{AP}) + \hat{\rho}_\theta(X_{LA}, Y_{LA}) \quad (2.9)$$

where  $\hat{\rho}_\theta(X_{AP}, Y_{AP})$  is the estimated correlation between a DRR and a radiograph from the AP direction and  $\hat{\rho}_\theta(X_{LA}, Y_{LA})$  is that from lateral images.

We also applied the *Nelder-Mead simplex method* to solve this optimization problem.

### 2.2.5 Marker-based Method

To establish the “ground truth”, we estimated the set-up error using the positions of the radiopaque markers. The results from the marker-based method were assumed to be the true phantom position. As a result, the performance of the MI-based estimator and the correlation-based method were evaluated by referencing the position estimated by the marker based method.

The procedures for estimating the set-up error by the marker-based method were as follows. First, we identified the slices of the CT which contain the markers. Figure 2.2

shows example two slices among slices that contain markers. We manually identified  $8 \times 8 \times 4$  voxel region around each marker and estimated each center position using the centroid method [36]. We also manually identified  $7 \times 7$  pixels around each marker in the four different radiographs and identified each center position as  $p_i$  using the centroid method. Next, while transforming the coordinates of the center position of each marker in 3D space, we projected marker center positions onto 2D planes which are the same planes as the radiograph planes, and identified the 2D indices of the projected center positions as  $d_i(\theta)$ .

We estimated the set-up error by minimizing the mean square Euclidean distance between the  $p_i$  and  $d_i(\theta)$  as follows:

$$\hat{\theta} = \arg \min_{\theta} \sum_i \|p_i - d_i(\theta)\|^2. \quad (2.10)$$

There were 22, 19, and 21 clearly identifiable markers in the four radiographs for Experiment A, B, and C respectively. Although the projection of 11 markers on the phantom to four different projection views resulted in 44 markers in the radiographic planes, some markers were not visible since those markers projected beyond the bounds of the detector.

## 2.3 Results

### 2.3.1 Radiographs

Figure 2.3 (a) shows the estimated MI from radiograph/DRR 1 and Figure 2.3 (b) shows the average of two estimated MI from radiograph/DRR 1,4 as the planning CT is translated along the X, Y and Z axes from the registered position respectively while other five parameters were kept at the registered position. The registered position was defined as the mean of the marker-based estimated positions from 10 radiographs.

For the radiograph/DRR 1 case, the estimated MI changes only slightly with respect to the translation along the Y-axis, and the point of the maximal estimated MI is far from the registered point. This large error could be due to the fact that the movement along the Y-axis does not cause much change in DRR 1. Thus 3D/2D registration based on a single-view similarity measure would be sensitive to noise. The insensitivity of the single-view estimated MI with respect to changes in  $t_y$  is clear from Figure 2.1, because the Y-axis translation causes only a small change of magnification in the projection image.

Similar phenomena occur if only radiograph/DRR 4 is used. Thus, using only one projection view may cause a significant error in estimating the translation that is orthogonal to the detector plane.

This problem is alleviated by using two radiograph/DRR images. Figure 2.3 (b) shows the estimated MI using radiograph/DRR 1 and 4. The maximal mutual information position is close to the true position for all six parameters. Based on this observation, we used radiograph/DRR 1,4 for evaluating the proposed MI-based estimator and the correlation-based estimator.

### 2.3.2 Estimated joint pdf

Figure 2.4 shows an example of non-registered and registered radiograph/DRR 1 and 4. The non-registered DRR was generated by translating the planning CT 12.5mm along the Z-direction from the registered position. For non-registered images, the bottom rows of DRR 1 contain bright pixels, unlike the corresponding pixels in radiograph 1. Also the top left parts of radiograph 4 are imaged dark, since those areas are air, while some of the corresponding pixels in DRR 4 have bright pixels. These phenomena diffuse the joint histogram and decrease the estimated mutual information.

Figure 2.5 shows the joint histograms that are estimated from the registered and the non-registered DRR/Radiograph 1 and 4. The joint histogram from the registered images shows interesting characteristics. It was expected that the joint histogram from the registered images would show more clustered shape along increasing functional curve, since the MV attenuation coefficients tend to be high if the kV attenuation coefficients are high. However, a range of DRR intensity values corresponded to high radiograph intensity values in Figure 2.5 (b), the estimated joint histogram from lateral images. This can be explained as the effect of the radiotherapy table. The radiotherapy table is slightly visible in the rightmost parts of radiograph 4 in Figure 2.1 (e). Although most of the radiotherapy table parts were not used for the MI-based and the correlation-based method, it still made the upper right parts of the radiograph 4 brighter than the DRR 4 since the corresponding pixels in the DRR 4 only represents air as we can see in Figure 2.4. For increased intensity values in the upper right parts due to the radiotherapy table in Figure 2.4 (b), even if the DRR and the radiographs are registered, corresponding pixels of DRR in Figure 2.4 (f) are darker than radiograph 4.

### 2.3.3 Position Estimation Results

In Experiment A, we used laser alignment to set up the phantom without error, while in Experiment B and C, we generated set-up errors deliberately to test the robustness of the proposed method to different set-up errors.

In Experiment A, radiographs 1 and 4 were acquired 10 times. Table 2.1 summarizes the empirical means and the standard deviations of each method in Experiment A. The estimated set-up errors by the proposed MI-based method were close to the set-up errors determined by the marker-based method. The differences between the means of the MI-based method and the marker-based method did not exceed 1.0 mm for translation parameters and  $0.8^\circ$  for rotational parameters. Considering that the voxel spacing of the sub-sampled planning CT that was used for computing DRR was  $3.75 \times 3.75 \times 4$  mm, estimation errors for every parameter were sub-voxel.

The sample STD (Standard Deviation) of the proposed estimator was very small. This was because the EPID has very low noise and we used an automatic method without human interaction.

The results of the correlation-based method were also fairly good. This indicates that the relationship between DRR and radiograph is approximately linear. Compared with the proposed method, the result of the correlation-based method shows relatively larger variance to noise. However, since the noise level was quite low, the noise-induced variability was insignificant compared to the mean errors.

Table 2.2 shows the results from Experiment B. In this experiment, we tried to generate patient set-up error of  $t_x=12\text{mm}$ ,  $t_y=-8$  mm  $t_z=-7\text{mm}$  and  $\phi_x = \phi_y = \phi_z = 0^\circ$ . Compared with the results from Experiment A, the correlation-based method performed worse while the MI-based method still worked well.

Table 2.3 shows the results from Experiment C. For this experiment, we tried to generate rotational set-up error. The planned set-up error was  $\phi_x=0^\circ$ ,  $\phi_y=2.2^\circ$ ,  $\phi_z=1.2^\circ$ ,  $t_x = t_y = t_z = 0\text{mm}$ .

In summary, for three different phantom positions, we estimated the set-up error 12 times including 10 repeated estimates using 10 different acquisitions. The means of the estimated set-up error differences between the marker-based method and the MI-based estimator did not exceed 1mm for translation parameters and did not exceed  $0.8^\circ$  for rotation parameters.

## 2.4 Discussion

By using two orthogonal radiograph/DRR pairs, we achieved average accuracies of better than 1mm for translational movement parameter and better than  $0.8^\circ$  for rotational movement parameters in estimating the set-up error.

We established the “ground truth” positions using the marker-based method. The accuracy of the marker-based method was expected to be the best among the three methods that were tested because the higher resolution CT was used to identify the marker position in the CT and four radiographs/DRRs were used in the marker-based method while only two radiographs/DRRs were used for the other methods. Although this method also contains error due to segmentation errors and noise, we used the marker-based method as “ground truth” since it is expected to be more accurate than the other methods.

There was approximately 3mm set-up variation from the positioning of the phantom at its proper reference location. Factors that contribute to this value include the limits of human operators in positioning relative to laser marks, as well as differences in laser calibration between the CT scanner and the treatment room. A retrospective review indicated a 1mm offset of the CT lasers from the center of the image matrix. Such errors in transferring a phantom or patient from one system to another have been previously reported, and are unlikely to be dramatically reduced in routine radiotherapy quality assurance.

For rigid body set-up error estimation using the chest phantom, two radiographs/DRRs were adequate for sub-voxel accuracy. We suspect that the performance of the estimator would be improved only modestly if more DRRs and radiographs from different angles were added. Practically, using fewer radiographs/DRRs is strongly preferable because of acquisition time and computation time.

The correlation-based method also achieved sub-voxel accuracy. Even though many factors could cause the intensity relation between DRR and radiograph to be nonlinear, we found the relationship is approximately linear so the correlation-based method worked fairly well.

We have found that the standard deviations of the correlation-based method and the MI-based method were very small in Experiment A. This implies that both methods may be statistically efficient. We are not sure if this phenomena is just for our experiments or general. One may investigate this problem by theoretically approximating the variances of

the estimator [26] as well as experimentally.

We also tried to compute the MV DRRs for the correlation-based method. However, results were no better than using kV DRRs. We think that was because we did not have CT data from test phantoms for voxel classification [20] nor information about scattering, etc. Since we were unable to implement the algorithm thoroughly, we have excluded the results.

One may try to compute better MV DRRs by following thorough procedures, however we suspect that the result will not be improved dramatically since unmodeled effects such as the presence of the radiotherapy table, difficulty of correctly compensating for scatter and beam hardening may limit the performance.

Compared with methods that require an *a priori* model, we think that the MI-based method may perform better in the presence of unexpected objects. As presented in the experimental results section, even though there exist non-modeled effects of the radiotherapy table, the results of the MI-based estimator show good performance. This robustness of the MI-based estimator to non-modeled effects partly supports the advantages of the method over *a priori* model-based methods such as MLE [34]. Therefore, the MI-based method may be useful in applications in which non-modeled objects may present, such as image-guided surgery.

For estimating patient set-up error, around 225 evaluations of the estimated mutual information were required. Each evaluation of the estimated mutual information requires computation of two DRRs. The joint histogram and the mutual information also must be computed for each mutual information evaluation. The most time consuming part was generating DRRs; it took around 16 sec to compute one  $400 \times 300$  DRR from the  $128 \times 128 \times 85$  planning CT on Pentium II 600 MHz machine. As a result, estimating one patient set-up error using two orthogonal radiographs/DRRs required about 2 hours.

This long computation time is not due to the MI criteria but because of DRR computation. Compared with other intensity-based method using DRR, our computation time was longer since we implemented a more accurate line integral instead of a trilinear interpolation approximation [18]. One might use trilinear interpolation approximations for computing DRR with proposed MI criteria for faster estimation.

Due to long computation time, it is difficult to estimate the patient set-up error in real time with the proposed method. As a result, adjusting the position of the x-ray source

or the position of the radiotherapy table to compensate for the patient set-up error is presently impractical. Instead, as a first application, we expect our method to be applied to review the patient set-up procedure. However, in near future, we expect that the proposed method can be used for estimating the set-up error in real time using pre-computed DRRs stored in large memory with appropriate interpolation and perhaps parallel processing. Furthermore, methods such as multi-resolution optimization technique may further reduce computing time [37].

Our investigation can be discussed in comparisons with other investigations. Lemieux *et al.* applied the correlation-based similarity measure to skull phantom experiments [18]. They used kV X-ray source for both CT and radiographs so that the correlation measure may work better.

Dong *et al.* also investigated the correlation-based method by testing the method using a head phantom [20]. Although they used MV radiographs and kV CT, they generated MV DRRs so that the DRRs and the radiographs have linear intensity relationship. MV DRRs were generated by classifying voxels into several categories such as bone, muscles, etc. based on CT numbers and seeking typical MV attenuation coefficients of those.

There also have been many studies of set-up estimation [8, 15–20]. Most investigations reported set-up estimation methods of less than a few mm estimation error for translation parameters and less than a few degree for rotation parameters. It is hard to compare the performance of each method directly since different types of CT images and radiographs were used. We believe that the performance of an estimator can depend greatly on the image characteristics. For example, the presence of high contrast objects can improve performance as implied by the Fisher information matrix in Appendix A. Moreover, different image resolution and different optimization stopping criteria may also affect the performance of the estimators.

Penney *et al.* investigated the performance of several similarity measures by applying those similarity measures to the image registration of a 3D CT to a fluoroscopy image [38]. However, since only one fluoroscopy image was used for estimating six parameters, the results may not be directly applicable to our study that used two radiographs for registration.

The MI-based method was investigated for the set-up error estimation problem by Hadley *et al.* [39]. In their investigation, the MI-based method worked well for simulated

images but did not show good performance for clinical images. Since the investigation was for 2D/2D image registration and there was no comparison to other similarity measures for clinical images, it does not necessarily imply that the MI-based method perform worse in 3D/2D registration problems than other similarity measures.

Our future work on set-up error estimation includes the application of the MI-based estimator to clinical data. Although we have achieved good registration results in a phantom experiment, we believe that much more verification with clinical data is crucial for potential application of the MI-based method to the clinical practice.

We plan to investigate non-rigid body motion set-up error. Since the human body is not rigid and radiotherapy may change the size and shape of the tumor, an accurate non-rigid body motion set-up error estimation is required for practical application.

We also plan to investigate other intensity-based similarity measures. For example, Rényi entropy may be a more flexible and computationally efficient similarity measure to accomplish image registration [40, 41] since it is more generalized definition of entropy. Investigating better similarity measures such as Rényi entropy may improve both speed and performance.

## 2.5 Conclusion

We have investigated a set-up error estimation method using 3D/2D, intensity-based image registration. To achieve 3D to 2D image registration, the radiographs and the DRRs of the planning CT were registered by maximizing the MI between DRRs and radiographs. In the experiment with an anthropomorphic chest phantom, we achieved accuracies better than 1mm for estimating the translational parameters and 0.8 degree for estimating the rotational parameters using two orthogonal pairs of the MV radiographs and DRRs. The true set-up error was established by the fiducial marker-based method. Based upon theoretical background and the experimental results, we believe that MI has significant potential as an effective similarity measure for 3D/2D intensity-based registration.

(\*In each table, the units for rotational parameters are degrees and for translational parameters are mm.)

Table 2.1: Estimated set-up error for Experiment A

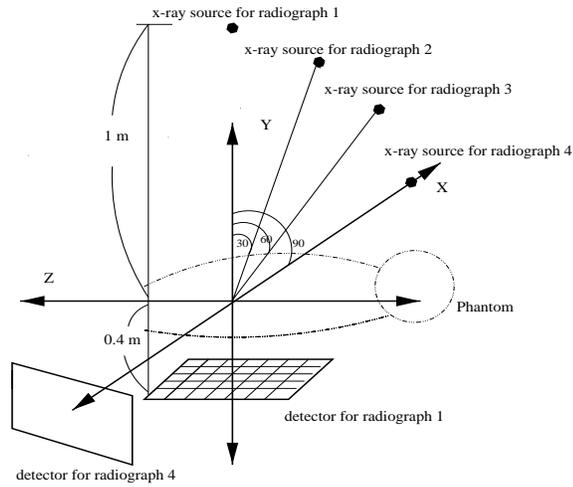
	$\phi_x$	$\phi_y$	$\phi_z$	$t_x$	$t_y$	$t_z$
Marker (mean)	0.601	-0.114	-1.892	-1.968	-3.568	-2.995
MI error (mean)	-0.064	0.035	-0.729	-0.635	0.692	-0.151
Corr. error (mean)	0.336	0.143	-1.308	-0.939	-0.597	0.102
Marker STD	0.004	0.002	0.002	0.012	0.009	0.004
MI STD	0.045	0.008	0.071	0.088	0.060	0.071
Corr. STD	0.318	0.080	0.158	0.254	0.114	0.248

Table 2.2: Estimated set-up error for Experiment B

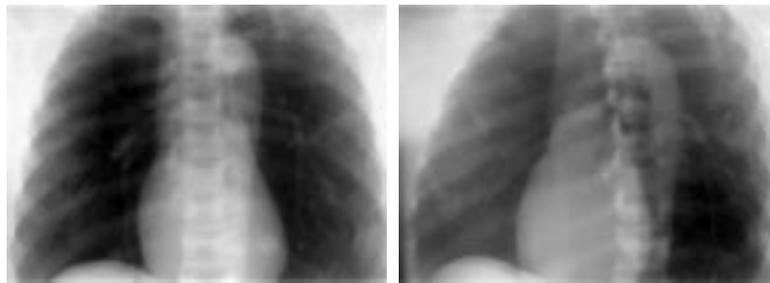
	$\phi_x$	$\phi_y$	$\phi_z$	$t_x$	$t_y$	$t_z$
Marker	0.519	0.005	-1.879	9.703	-11.32	-9.58
MI error	-0.153	-0.044	-0.594	-0.597	1.030	0.668
Corr. error	0.286	-0.001	-2.534	-0.924	-3.092	1.451

Table 2.3: Estimated set-up error for Experiment C

	$\phi_x$	$\phi_y$	$\phi_z$	$t_x$	$t_y$	$t_z$
Marker	0.755	1.864	-0.698	-1.515	-3.201	-2.947
MI error	0.067	-0.121	-0.667	-1.043	0.992	-0.453
Corr. error	0.72	0.069	-1.172	-1.050	-0.391	-0.010

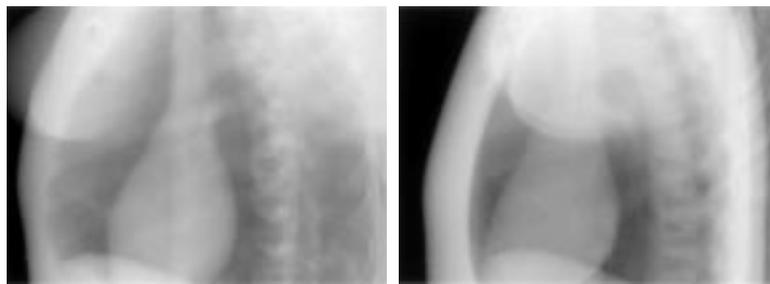


(a) Radiographs and coordinate system



(b) Radiograph 1: angle 0

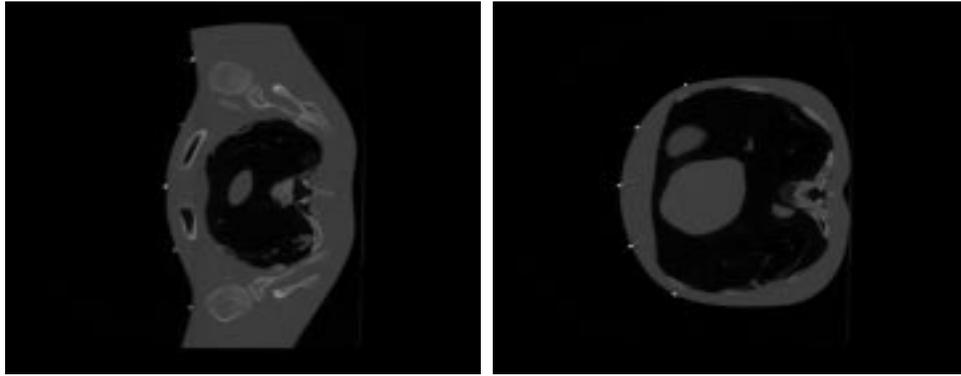
(c) Radiograph 2: angle 30



(d) Radiograph 3: angle 60

(e) Radiograph 4: angle 90

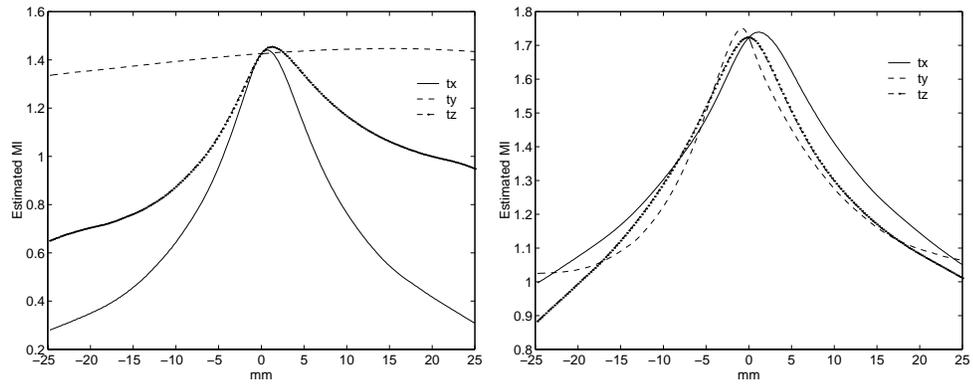
Figure 2.1: Radiographs



(a) 86th slice

(b) 251th slice

Figure 2.2: Slices with markers



(a) Radiograph/DRR 1

(b) Radiograph/DRR 1, 4

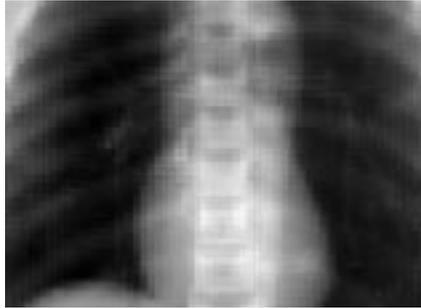
Figure 2.3: Estimated mutual information with respect to the translation errors



(a) Radiograph 1



(b) Radiograph 4



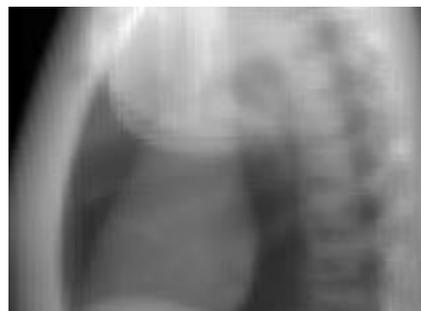
(c) Non-registered DRR 1



(d) Non-registered DRR 4

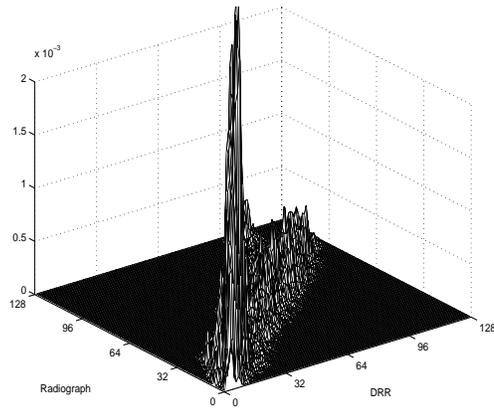


(e) Registered DRR 1

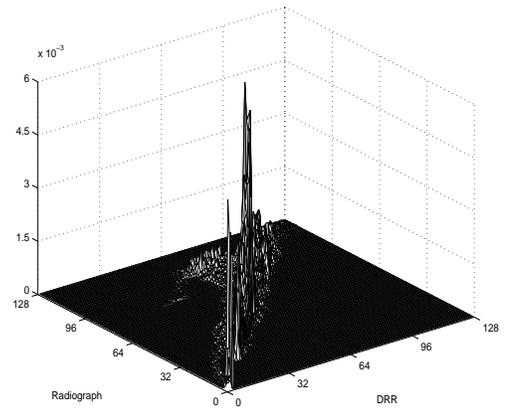


(f) Registered DRR 4

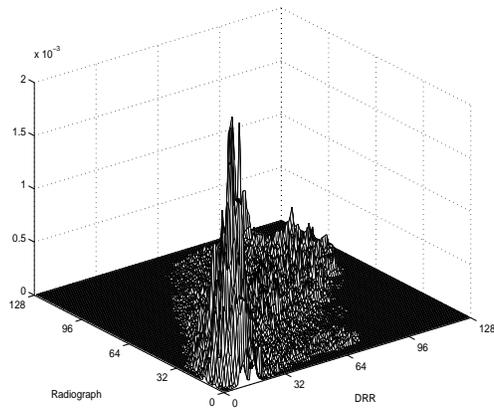
Figure 2.4: Non-registered and registered radiographs and DRRs



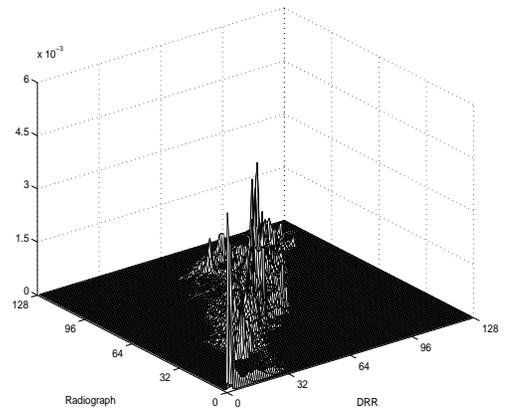
(a) Registered (DRR/radiograph 1)



(b) Registered (DRR/radiograph 4)



(c) Nonregistered (DRR/radiograph 1)



(d) Nonregistered (DRR/radiograph 4)

Figure 2.5: Joint histograms from registered and nonregistered images (Numbers in X,Y axis represent histogram bin number. 0 corresponds to the lowest intensity and 128 corresponds to the highest intensity).

## CHAPTER 3

### Robust Registration

#### 3.1 Introduction

As explained in Chapter 1, two most well known similarity measures, the sample correlation coefficient and the MI, have drawbacks such as lack of robustness and statistical inefficiency. To overcome the drawbacks, we have investigated an image registration method that uses robust correlation coefficients [24, p. 204] as a similarity measure, thereby improving the robustness without compromising the statistical efficiency much.

Robust estimation of mean and covariance has been extensively studied in statistics [24, 42–44]. The basic idea of robust estimation is to weight the measured samples in a way that reduces the effect of outlier samples, or even trim out the outlier samples completely.

For example, one may compute the statistical distance of each sample value from the mean, the *Mahalanobis distance*, evaluate a weighting function based on that distance, and determine a new weighted mean and covariance and iterate until convergence [43]. Alternatively, one may estimate the pdf after trimming out the outliers by determining the minimum volume ellipsoidal pdf [44] or minimum covering ellipsoidal pdf [42]. A robust mean and covariance may then be estimated from the estimated pdf.

Another possibility is *R*-estimation methods that estimates parameters based on a statistical rank [45, 46]. For example, the mean can be estimated using the concept of *depth* [47, 48], which is a generalization of median to the multidimension. These principles could also be applied to form a kind of “trimmed” estimate, based on removing samples that have extreme ranks.

In this study, we focused on *M*-estimation methods for robust correlation estimation [24,

p. 211]. This framework helps to explain why the sample correlation coefficient is sensitive to outliers and provides insight into how to design a method with improved robustness. Moreover, we can explain some properties of the MI method within the same framework.

Investigating the advantages and disadvantages of using alternative robust correlation estimation methods for image registration in the presence of outliers is deferred to future research. Such a study should consider the sample distributions of representative images in a particular context.

This chapter is organized as follows. Section 3.2 reviews the image registration problem and describes the proposed similarity measure. Section 3.3 compares the image registration accuracies of the proposed robust correlation similarity measure, the conventional sample correlation coefficient and the mutual information similarity measure. Three comparisons are reported: 1D simulation, 2D fMRI image registration, and 3D/2D registration of an X-ray CT volume to orthogonal radiographs of an anthropomorphic chest phantom.

The appendices present analyses of the statistical properties of the correlation-based and MI-based image registration methods by approximating the mean and the variance using first-order Taylor series expansions [26]. Since image registration is highly nonlinear and the objective function is an implicit function of the images, it is challenging to obtain concise and insightful results from such approximations. Nevertheless, we summarize some theoretical arguments that complement the empirical results.

## 3.2 Theory: similarity measures

The goal of image registration is to find a geometric transformation (rigid or non-rigid), denoted  $T$ , that aligns two given images, denoted  $s_1(\vec{t})$  and  $s_2(\vec{t})$ , where  $\vec{t}$  denotes the spatial coordinates<sup>1</sup>. Intensity-based image registration methods achieve this goal by maximizing a similarity measure based on the image intensity values. If we parameterize the transformation  $T$  using  $\theta$  (*e.g.*, three translation and three rotation parameters for rigid transformation), the image registration becomes a parameter estimation problem:

$$\hat{\theta} = \arg \max_{\theta} \Phi(s_1(T_{\theta}(\cdot)), s_2(\cdot)), \quad (3.1)$$

---

<sup>1</sup>We focus on image-to-image or volume-to-volume registration, but the general approach applies equally to volume-to-projection registration, *e.g.*, [12]. Also, we treat  $s_1(\vec{t})$  and  $s_2(\vec{t})$  as continuous-space functions in the presentation; in practice, sampling and interpolation are essential [22].

where  $\Phi(s_1, s_2)$  is some measure of the *similarity* between the images  $s_1$  and  $s_2$ . In practice, registration is performed using finite number of samples  $X_i$  and  $Y_i$  as follows:

$$\begin{aligned} X_i &= s_1(T_\theta(\vec{t}_i)) \\ Y_i &= s_2(\vec{t}_i), \quad i = 1, \dots, N, \end{aligned} \tag{3.2}$$

where  $\{\vec{t}_i\}$  denotes the sample locations, and where  $s_1(T_\theta(\cdot))$  denotes a spatially transformed (and interpolated) version of  $s_1(\vec{t})$ . Since  $X_i$  depends on the parameter  $\theta$ , all statistical quantities computed using  $X_i$  are functions of  $\theta$ . However, for simplicity of notation usually we leave this dependence implicit.

Considering the sampling, a more precise expression for the registration problem is

$$\hat{\theta} = \arg \max_{\theta} \Phi(\mathbf{X}(\theta), \mathbf{Y}), \tag{3.3}$$

where  $\mathbf{X} = (X_1, \dots, X_N)$  and  $\mathbf{Y} = (Y_1, \dots, Y_N)$ .

A variety of similarity measures  $\Phi$  have been proposed for image registration. Many of these are statistical quantities such as the correlation coefficient, joint entropy and mutual information. For such metrics, there is an underlying assumption that the  $(X_i, Y_i)$  pairs are i.i.d. samples of jointly distributed random variables with some (unknown) joint probability density function (pdf)  $f_{XY}(x, y)$ . This i.i.d. assumption is somewhat artificial, but nevertheless leads to useful similarity measures so we continue in this tradition in this paper.

Next we review the usual correlation coefficient similarity measure and contrast it with our proposed robust correlation coefficient approach. Note that we want robustness to the outliers in the “artificial” pdf  $f_{XY}(x, y)$  of the two images, as opposed to the pdf of noises that are present in the images.

### 3.2.1 Correlation Coefficient Estimates

For two random variables  $X$  and  $Y$  with joint pdf  $f_{XY}(x, y)$ , the correlation coefficient  $\rho(X, Y)$  is defined as follows:

$$\rho(X, Y) \triangleq \frac{C(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}, \tag{3.4}$$

where the covariance is

$$C(X, Y) \triangleq \int (x - E[X])(y - E[Y]) f_{XY}(x, y) \, dx \, dy \tag{3.5}$$

and where  $E[X]$  and  $E[Y]$  denote the means, and  $\sigma_X^2$  and  $\sigma_Y^2$  denote the variances of the random variables. Note that (under mild regularity conditions on  $f_{XY}(x, y)$ ), the correlation coefficient  $\rho$  is well-defined even if the pdf  $f_{XY}(x, y)$  is not parameterized in terms of it.

Given  $N$  i.i.d. sample pairs  $(X_i, Y_i)$ , there are several ways to estimate the correlation coefficient  $\rho$ . For example, one could first use the samples to compute an estimated joint pdf  $\hat{f}_{XY}(x, y)$ , and then substitute  $\hat{f}_{XY}(x, y)$  into (3.4) to estimate  $\rho$ . The following subsections describe two other approaches.

### Sample correlation coefficient

Perhaps the most popular way to estimate  $\rho$  is the sample correlation coefficient, defined as follows:

$$\hat{\rho}_s(\mathbf{X}, \mathbf{Y}) = \frac{\hat{C}_s(\mathbf{X}, \mathbf{Y})}{\sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}}, \quad (3.6)$$

where the sample means, sample variances, and sample covariance are defined in the usual way as follows:

$$\begin{aligned} \hat{C}_s(\mathbf{X}, \mathbf{Y}) &\triangleq \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \\ \bar{X} &\triangleq \frac{1}{N} \sum_{i=1}^N X_i \\ \bar{Y} &\triangleq \frac{1}{N} \sum_{i=1}^N Y_i \\ \hat{\sigma}_X^2 &\triangleq \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \\ \hat{\sigma}_Y^2 &\triangleq \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2. \end{aligned} \quad (3.7)$$

Relative to alternative methods for estimating the correlation coefficient, the sample correlation method has the advantage of simplicity since  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is an *explicit* function of the data samples  $(\mathbf{X}, \mathbf{Y})$ . Furthermore, on the surface it appears not to require any specific model for the joint pdf  $f_{XY}(x, y)$ . A minor drawback is that  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is not unbiased, even for normal distributions although it is asymptotically unbiased in that case [49, p. 90]. More importantly,  $\hat{\rho}_s(\mathbf{X}, \mathbf{Y})$  is not robust to outliers [24, p. 199], as explained intuitively in the next sub-section.

### Maximum likelihood estimates of $\rho$

An alternative approach to estimating  $\rho$  is the following: (i) hypothesize a parametric form for the joint distribution of  $X$  and  $Y$  that depends on  $\rho$ , (ii) find the maximum

likelihood (ML) estimate of the parameters given the data  $\mathbf{X} = \mathbf{X}(\theta)$  and  $\mathbf{Y}$ , and (iii) determine  $\rho = \rho(\theta)$  from the ML parameter estimates. Usually we will have to compute the ML estimates numerically, so the estimator  $\hat{\rho}$  will be an *implicit* function of  $\mathbf{X}$  and  $\mathbf{Y}$ . To our knowledge, previous similarity measures used in image registration have all been *explicit* functions of the two images, so the proposed approach departs from that convention.

Let  $\mathbf{Z}_i = (X_i, Y_i)$  denote the pairs of corresponding image intensity values. As usual, we assume the  $\mathbf{Z}_i$ 's are i.i.d. samples of a two dimensional random vector  $\mathbf{Z} = (X, Y)$ . Following Huber [24, p. 211], we model the pdf of  $\mathbf{Z}$  by an elliptic density that is constructed (by the algorithm designer) as follows. First we choose a nonnegative function  $f_0$  for which the corresponding 2D circularly symmetric density  $f_0(\|\mathbf{z}\|)$  integrates to unity over  $\mathbb{R}^2$ , where  $\|\mathbf{z}\| = \sqrt{x^2 + y^2}$ . Then we consider an (unknown) non-degenerate transformation  $\mathbf{z} \mapsto \mathbf{V}(\mathbf{z} - \boldsymbol{\mu})$  that leads to the following density:

$$f_{\mathbf{z}}(\mathbf{z}; \boldsymbol{\mu}, \mathbf{V}) = |\det \mathbf{V}| f_0(\|\mathbf{V}(\mathbf{z} - \boldsymbol{\mu})\|), \quad (3.8)$$

which has elliptical contours. Under this parametric model,  $\boldsymbol{\mu}$  denotes the mean of  $\mathbf{Z}$  and the  $2 \times 2$  covariance matrix of  $\mathbf{Z}$  is  $(\mathbf{V}^T \mathbf{V})^{-1}$ . In other words,  $\mathbf{V} = \text{Cov}\{\mathbf{Z}\}^{-1/2}$ . For example, if one were to choose

$$f_0(r) = \frac{1}{2\pi} e^{-r^2/2}, \quad (3.9)$$

then (3.8) would become the standard bivariate normal distribution.

Having chosen some  $f_0$ , one may estimate the mean  $\boldsymbol{\mu}$  and the covariance term  $\mathbf{V}$  from the sample pairs  $\mathbf{Z}_i$  by ML estimation as follows:

$$(\hat{\boldsymbol{\mu}}, \hat{\mathbf{V}}) = \arg \max_{(\boldsymbol{\mu}, \mathbf{V})} \prod_{i=1}^N |\det \mathbf{V}| f_0(\|\mathbf{V}(\mathbf{Z}_i - \boldsymbol{\mu})\|). \quad (3.10)$$

Usually there is no closed-form expression for the estimates  $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{V}}$ , so (3.10) is an *implicit* definition.

To help understand these ML estimates, we differentiate the log-likelihood for (3.10) with respect to  $\boldsymbol{\mu}$  and  $\mathbf{V}$ . Equating these expressions to zero yields the following two necessary conditions for the ML estimates [24, p. 212]:

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N w \left( \left\| \hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) \right\| \right) \mathbf{Z}_i}{\sum_{i=1}^N w \left( \left\| \hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) \right\| \right)} \quad (3.11)$$

$$(\hat{\mathbf{V}}^T \hat{\mathbf{V}})^{-1} = \sum_{i=1}^N w \left( \left\| \hat{\mathbf{V}}(\mathbf{Z}_i - \hat{\boldsymbol{\mu}}) \right\| \right) (\mathbf{Z}_i - \hat{\boldsymbol{\mu}})(\mathbf{Z}_i - \hat{\boldsymbol{\mu}})^T, \quad (3.12)$$

where  $w(\cdot)$  denotes the following *weighting function*:

$$w(r) \triangleq -\frac{f_0'(r)}{rf_0(r)}. \quad (3.13)$$

Huber proposed a simple iterative procedure for solving these two nonlinear equations to obtain the ML estimates [24, p. 215].

After estimating the covariance term  $\hat{\mathbf{V}}$ , we can extract the ML estimate  $\hat{\rho}$  of the correlation coefficient as follows:

$$\begin{aligned} (\hat{\mathbf{V}}^T \hat{\mathbf{V}})^{-1} &= \begin{bmatrix} \hat{\sigma}_X^2 & \hat{C}_{XY} \\ \hat{C}_{XY} & \hat{\sigma}_Y^2 \end{bmatrix} \\ \hat{\rho} &= \frac{\hat{C}_{XY}}{\sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}}, \end{aligned} \quad (3.14)$$

where  $\hat{\sigma}_X^2$  and  $\hat{\sigma}_Y^2$  denote the ML estimated variances of  $X$  and  $Y$  under the assumed model (3.8).

After finding  $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{V}}$ , one can compute the weighting function  $w\left(\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\|\right)$  to examine the relative influence of different data values  $\mathbf{z}$  on the estimates. (See Fig. 3.5 for an example.)

In classical estimation theory, ideally  $f_0$  would be chosen so that the pdf  $f_Z$  in (3.8) agrees with the actual distribution of the  $\mathbf{Z}_i$ 's. However, as noted above, in the context of image registration the notion that the  $\mathbf{Z}_i$ 's are i.i.d. is somewhat artificial. Instead, it is more useful to think of  $f_0$  as a user-selectable function that one should design to impart desirable properties in the context of image registration, such as robustness to outliers.

If we were to choose (3.9) for  $f_0$ , then the weighting function in (3.13) simplifies to  $w(r) = 1$ . In this special case, there is an explicit solution for the ML estimates:  $\hat{\boldsymbol{\mu}}$  is simply the sample mean of the  $\mathbf{Z}_i$ 's, and  $\hat{\mathbf{V}}$  is the square root of the inverse of the sample covariance of the  $\mathbf{Z}_i$ 's. This corresponds to the well-known result that the sample correlation coefficient is the ML estimator for  $\rho$  under a bivariate normal density. However, using constant weighting  $w(r) = 1$  means that *all* data points are weighted equally, even outliers. As a result, the sample mean, sample covariance, and sample correlation coefficient are all sensitive to outliers [24].

### 3.2.2 Robust Correlation Coefficient

The non-robustness of the sample correlation coefficient can be explained by considering that it is the ML estimator for an assumed normal distribution, which has “light tails” so outliers are extremely unlikely. Conversely, if an ML estimate is based on a model distribution that has “heavy tails,” then even those data values that are far from the mean still belong within the tail portion of the pdf so have less effect on the likelihood function. Thus, to design a robust estimator, we choose a model pdf  $f_Z$  that has “heavy tails.” One choice would be the Laplacian distribution, which would correspond to using  $f_0(r) = ce^{-|r|}$  for some constant  $c$ . However, that  $f_0$  is not differentiable at  $r = 0$ , so the expression (3.13) is inapplicable. Instead, we have chosen the following function:

$$f_0(r) = ce^{-\delta\sqrt{1+r^2/\delta^2}-1}, \quad (3.15)$$

where  $c$  is the constant that ensures that (3.8) integrates to unity. The constant  $\delta > 0$  is a design parameter. For small  $\delta$  this model approaches the Laplacian distribution, and for large  $\delta$  it approaches the normal distribution.

For the choice (3.15), the weighting function  $w(r)$  in (3.13) becomes the following;

$$w(r) = \frac{1}{\sqrt{1 + \frac{r^2}{\delta^2}}}. \quad (3.16)$$

Unlike with the normal choice (3.9), for this model the weighting of a given data point  $w\left(\left\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\right\|\right)$  will decrease with increasing distance from the mean  $\hat{\boldsymbol{\mu}}$ . This has the desirable effect of reducing the influence of outliers that are, by definition, data points that are far from the mean.

We can make no claim of optimality of the choice (3.15). Indeed the optimal choice would depend on the actual “distribution” of the  $\mathbf{Z}_i$ 's, which is unknown in practice. The function  $f_0$  is simply a design parameter for our robust estimator. For example, one could try to increase robustness relative to (3.16) by using following weighting:

$$w(r) = \frac{1}{\sqrt{1 + \frac{r^4}{\delta^4}}}. \quad (3.17)$$

A pdf with heavier tails than a normal distribution should improve robustness relative to the conventional sample correlation coefficient.

Unfortunately, Huber’s algorithm for solving (3.11) and (3.12) has been proven to converge only when estimating *one* of the two parameters, *i.e.*,  $\boldsymbol{\mu}$  or  $\mathbf{V}$ , but not necessarily both [24, p. 237]. However, in practice, the algorithm converged every time in our simulations and experiments. We initialize  $\boldsymbol{\mu}$  with the sample median and  $\mathbf{V}$  with the square root of the inverse of the sample covariance matrix.

In summary, our robust registration method works as follows. For the similarity measure  $\Phi$  described in (3.3), we propose to use the robust correlation coefficient:

$$\Phi(\mathbf{X}(\theta), \mathbf{Y}) = \hat{\rho}(\mathbf{X}(\theta), \mathbf{Y}), \quad (3.18)$$

here  $\hat{\rho}$  was defined in (4.6) for the weighting function defined in (3.16) or (3.17). To maximize  $\Phi$  with respect to  $\theta$ , one must use some type of search algorithm such as the simplex method [37]. For each trial value of the registration parameter  $\theta$ , one must compute  $\mathbf{X}(\theta)$  by interpolation, and then compute  $\hat{\rho}$  by applying Huber’s algorithm. So there are “iterations within iterations” in this approach; fortunately, the inner iteration converges quite quickly.

### 3.2.3 Mutual Information

Another similarity measure that has robust characteristics is *mutual information* (MI). MI is a measure of the statistical dependence between two random variables. The MI  $I(X, Y)$  is defined in terms of marginal and joint entropies as follows:

$$\begin{aligned} h(X) &= - \int f_X(x) \log f_X(x) dx \\ h(Y) &= - \int f_Y(y) \log f_Y(y) dy \\ h(X, Y) &= - \int f_{XY}(x, y) \log f_{XY}(x, y) dx dy \\ I(X, Y) &= h(X) + h(Y) - h(X, Y). \end{aligned} \quad (3.19)$$

Usually MI is estimated by first estimating the joint pdf  $f_{XY}(x, y)$ , and then computing the MI using (3.19). Two popular methods for estimating joint pdf are the kernel density approach<sup>2</sup> [50] and the histogram approach. Kernel density estimates are smooth and differentiable but can require considerable computation. Histogram estimates are usually faster to compute but yield pdfs that are discontinuous functions of the registration parameter  $\theta$ . As a practical compromise, we have used the interpolated joint histogram method in which

---

<sup>2</sup>This is often called the *Parzen window* method.

the effect of a sample is distributed to four adjacent histogram bins using bilinear interpolation; this approach remedies the discontinuity problem of the histogram with modest computation. Methods for estimating MI directly from the samples without first estimating a pdf are also under development [51, 52].

### 3.2.4 Analytical Comparisons

Most previous studies of image registration methods have focused on empirical comparisons. The appendices of this paper describe approximate analyses of the statistical properties of image registration methods using mean and variance approximations presented in Appendix B. We used these approximations because exact analytical expressions are unavailable since the estimator for image registration is defined implicitly as the maximizer of an objective function. Even if an analytical expression were available, finding exact expressions for the mean and variance would still remain difficult since the estimator  $\hat{\theta}$  is a nonlinear function of the images.

Our approximation method uses a first-order Taylor series expansion of the estimator about the mean data, an approach that has been used successfully for image reconstruction problems [26]. By comparison, the estimators used for image registration are more nonlinear, but we proceed with linearization nevertheless, hoping for insights.

For simplicity we focus on the asymptotic case as the number of image samples increases to infinity, *i.e.*, the images are continuous-space functions. And we assume that two images are exactly the same images except for the geometric transformation and additive white Gaussian noise. In other words, we analyze the case of registering a reference image  $s_1(\theta, \vec{t})$  to noisy image  $s_2(\vec{t}) = s_1(\vec{t}) + n(\vec{t})$ , where  $n(\vec{t})$  is white Gaussian noise.

The main points of the analyses can be summarized as follows. First, the sample correlation coefficient method is unbiased and the most efficient in the presence of Gaussian noise (*i.e.*, the smallest variance). The unbiasedness is argued since the sample correlation coefficient without noise is maximized at the true position, *i.e.*, gradient of the sample correlation coefficient is zero at the true position. In our approximation, if the gradient of the objective function without noise is zero at true position, it is unbiased estimator (See Appendix B). Regarding the efficiency, the sample correlation method is the most efficient estimator among unbiased estimators since it is the MLE [34] in our problem setting, in the absence of outliers. One can easily show this fact by using the variance approximation since

the approximation equals the *Cramer-Rao* bound (See Appendix D). However, the sample correlation method may be significantly biased in the presence of unexpected objects as argued in the Appendix C. In summary, the sample correlation method has good properties such as the unbiasedness and efficiency but poor robustness to unexpected objects.

Appendix C argues that the robust correlation coefficient method is unbiased by similar approximations. Moreover, in the presence of outlier samples, the robust correlation method is more robust than the sample correlation method since the effect of the outlier samples is reduced by smaller weighting (See Appendix C). However, for data without outliers, the variance of the robust correlation method is larger than the sample correlation method as argued by Cauchy-Shwarz inequality [53] (See Appendix D).

For the MI method, many investigators have reported that the estimated MI from noiseless data is maximized at the true registered position since the estimated joint pdf is the most clustered along the  $Y = X$  line in the joint pdf domain. Moreover, one may argue that the values of the joint pdf reach maxima around the  $Y = X$  line and minima elsewhere by the same observation. In that case, the gradient of the estimated joint pdf is zero almost everywhere at true registered position. As a result, the gradient of the estimated MI is zero at true registered position. Nevertheless, due to the smoothing effect of the kernel function, it is difficult to prove the unbiasedness analytically except for rigid registration with constant backgrounds (See Appendix E).

One may also argue that the MI method has inherent robustness. Suppose that the portion of the estimated pdf from inliers is the most clustered at the true registered position but that from outliers is not the most clustered at true position<sup>3</sup>. In that case, the estimated MI is maximized at a position where the entire pdf is the most clustered in average sense. Since the portion of the pdf from inliers is usually much larger than that from outliers, the entire pdf is likely to be the most clustered around the true position. Nevertheless, since the robustness of the MI method is due to the behavior of the joint pdf rather than by explicit reducing the influence of outliers, the robustness can depend greatly on images being registered.

Another interesting perspective is to express the MI as a generalized weighted correlation

---

<sup>3</sup>Apparently, if pdf from outliers is also the most clustered at true position, there should be no bias due to outliers.

coefficient as follows:

$$I_\theta(X, Y) = \int (x - E[X])(y - E[Y])w_I(x, y)\hat{f}_{XY}(x, y), \quad (3.20)$$

where  $w_I(x, y) = \frac{\log \hat{f}_{Y|X}(y|x)}{(x-E[X])(y-E[Y])}$ .

One may think that  $w_I(x, y)$  as a very generalized weighting. Depending on the images to be registered, the robustness characteristics as well as the statistical efficiency can vary since the weighting depends on the estimated joint pdf. For example, if the estimated pdf is normally distributed, the weighting is almost constant and the MI method is almost the same as the sample correlation method. In that case, the MI method can be very efficient like the sample correlation method. In fact, the joint entropy has one-to-one relationship with correlation coefficient for bivariate normal distribution [35]. For most intra-modality images with non-constant  $w_I(x, y)$ , the MI method is less efficient than the sample correlation coefficient for Gaussian noise.

In summary, both the robust correlation method and the MI method have improved robustness but larger variance compared to the sample correlation method, when noise is Gaussian. We do not have direct comparison of the statistical properties between the robust correlation method and the MI method since the properties depend not only on design parameters (such as underlying pdf for robust correlation method and the pdf estimation method for MI method) but also on the images being registered.

### 3.3 Experimental Results

To evaluate the statistical properties of the proposed image registration method, we performed three studies: simulations using a synthetic 1D signal, affine 2D-to-2D registration of fMRI images, and rigid 3D-to-2D registration of a torso phantom.

#### 3.3.1 1D Simulation

We first consider a 1D “registration” problem where the only unknown parameter is the translation of the signal. (This is called *delay estimation* in communications.) This study illustrates the statistical properties of the registration methods in the simplest possible setting. Fig. 3.1 shows the reference signal  $s_1(\vec{t})$  that must be translated for registration with the signal  $s_2(\vec{t})$  shown in Fig. 3.2 that includes both additive Gaussian noise and a

segment of “outlier” signal values.

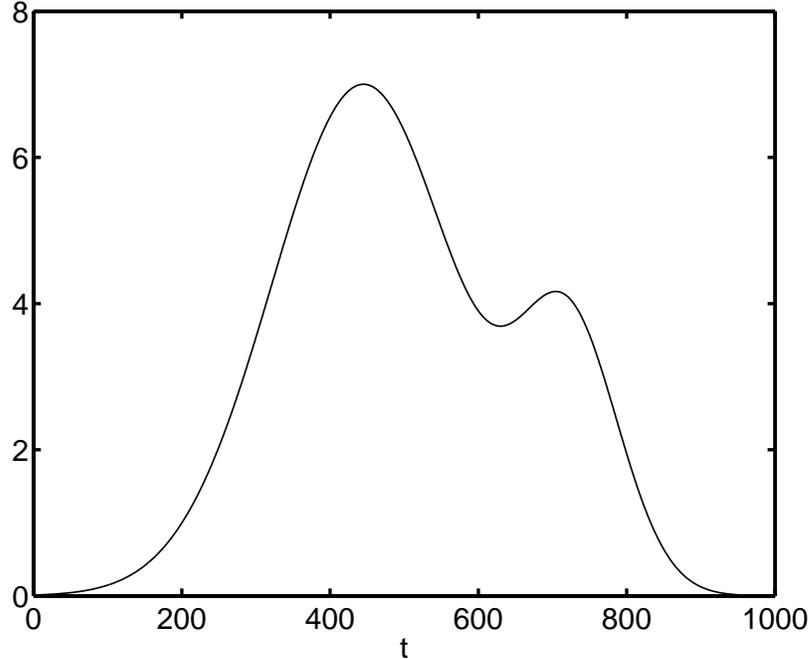


Figure 3.1: Reference signal  $s_1(\vec{t})$  (signal being translated to achieve registration).

Using these signals, we computed three similarity measures as a function of the translation parameter  $\theta$ : the sample correlation coefficient, the robust correlation coefficient using weighting (3.17) with  $\delta = 2$  and the estimated MI using a  $32 \times 32$  interpolated joint histogram.

Fig. 3.3 shows the two correlation coefficients versus  $\theta$ , where the true value is  $\theta = 0$ . Due to the presence of the outliers, the sample correlation coefficient is maximized at an incorrect translation ( $\approx -4.8$  pixel). This type of systematic offset is observed for most noise realizations. In contrast, the robust correlation coefficient is maximized near the true translation ( $\approx -0.1$  pixel), illustrating the robustness of this similarity measure. This robustness can be understood by considering the joint histogram shown in Fig. 3.4 and the weighting function  $w\left(\left\|\hat{\mathbf{V}}(\mathbf{z} - \hat{\boldsymbol{\mu}})\right\|\right)$  shown in Fig. 3.5 (at the registered position where  $\theta = 0$ ). Although most of the histogram mass lies along a line, there is a group of outliers that degrade the conventional correlation coefficient estimate. Fig. 3.5 shows that the weighting function decreases the influence of those outliers, particularly those that are far from the primary linear ridge, thereby providing robustness.

Note that Fig. 3.4 and Fig. 3.5 are only for an illustration. In practice, the robust corre-

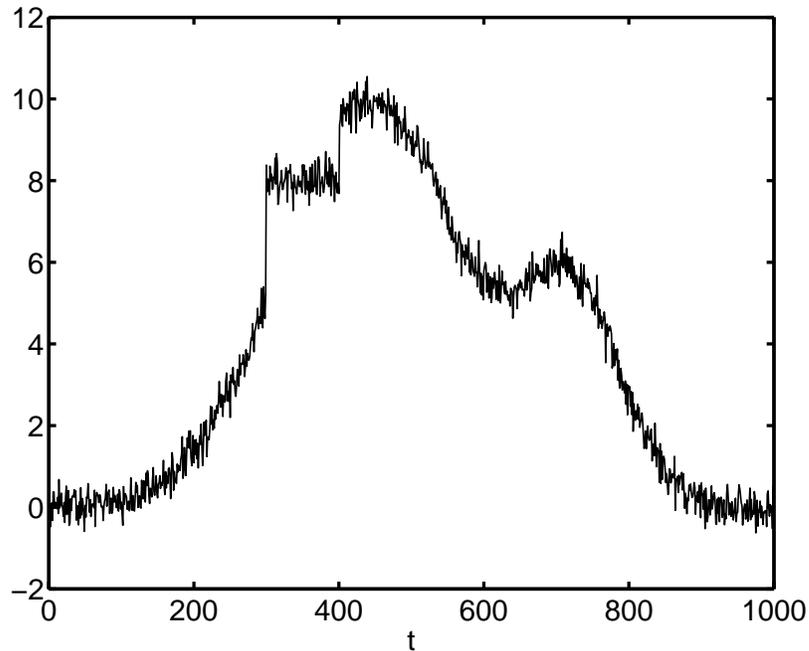


Figure 3.2: Observed signal  $s_2(\vec{t})$  with outliers and Gaussian noise (STD = 0.3).

lation coefficient is determined by solving (3.11) and (3.12) iteratively without computing pdf or weighting.

Fig. 3.6 shows that the estimated MI was maximized at which is more accurate ( $\approx -0.8$  pixel) than the sample correlation coefficient results. This robustness of MI is expected as argued in the previous section.

The preceding results were for a single noise realization, so they do not fully characterize the bias and variance of the translation estimates. We performed 1000 noise realizations at each of several noise levels and computed translation estimates  $\hat{\theta}$  using each of the three similarity measures for each realization.

Fig. 3.7 shows the empirical translation estimation biases caused by the presence of the outliers for all three methods. As expected, the sample correlation coefficient method based registration technique was the most sensitive to outliers, as argued in Appendix C.

The robust correlation based method reduced the bias almost completely for low noise levels but showed increased bias for high noise levels. Although the MI based method also reduced bias comparing to the sample correlation based method, the bias was larger than the robust correlation based method.

Fig. 3.8 shows the standard deviations of the three estimators. As expected, the sample

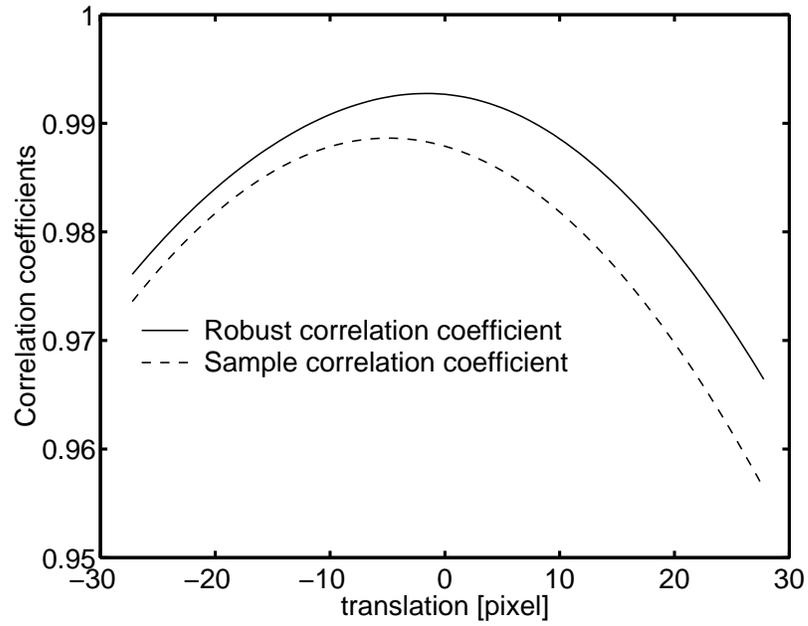


Figure 3.3: Sample and robust correlation coefficients vs 1D translation.

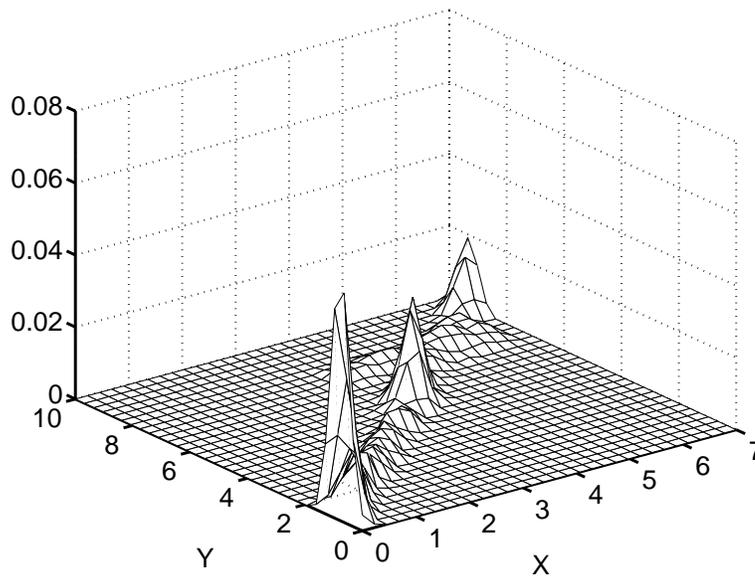


Figure 3.4: Joint pdf at registered position (computed by  $64 \times 64$  interpolated joint histogram).

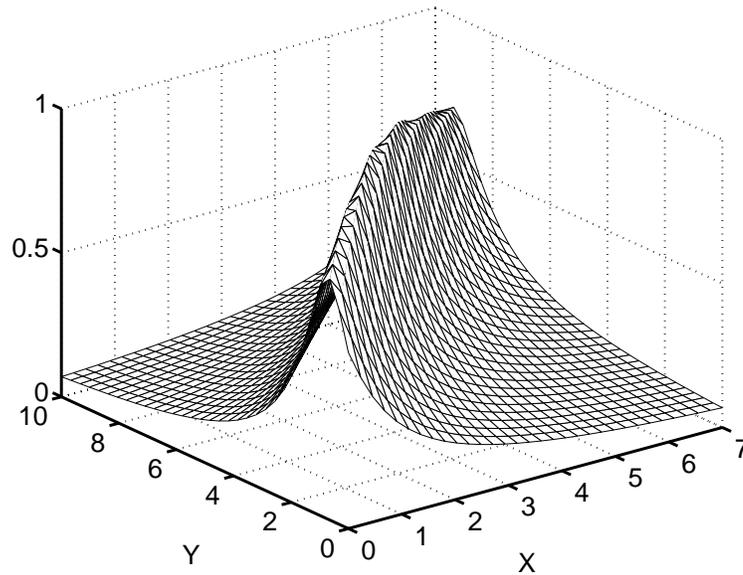


Figure 3.5: Weighting function at registered position (evaluated at  $64 \times 64$  joint histogram domain).

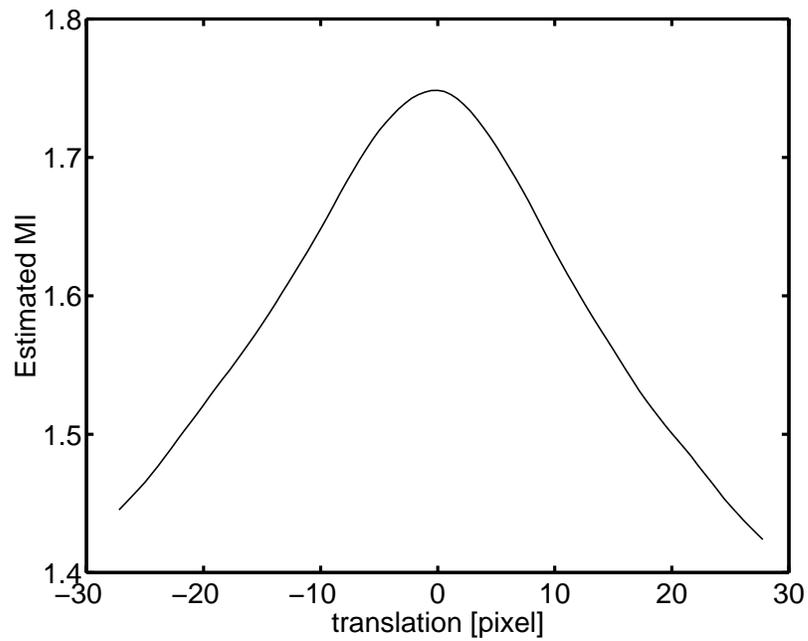


Figure 3.6: MI similarity measure ( $32 \times 32$  interpolated joint histogram) vs 1D translation.

correlation coefficient method had the smallest variance as argued in Appendix D. Also, shown for reference is the *Cramer-Rao* bound computed for a “no outlier” model. In general, one would not expect the sample correlation method to match this bound for data containing outliers. However, as explained in Appendix D, for the particular signals in Fig. 3.1, Fig. 3.2, the standard deviation of the sample correlation method happened to match the *Cramer-Rao* bound. The standard deviation of the robust correlation method was almost the same as the MI method.

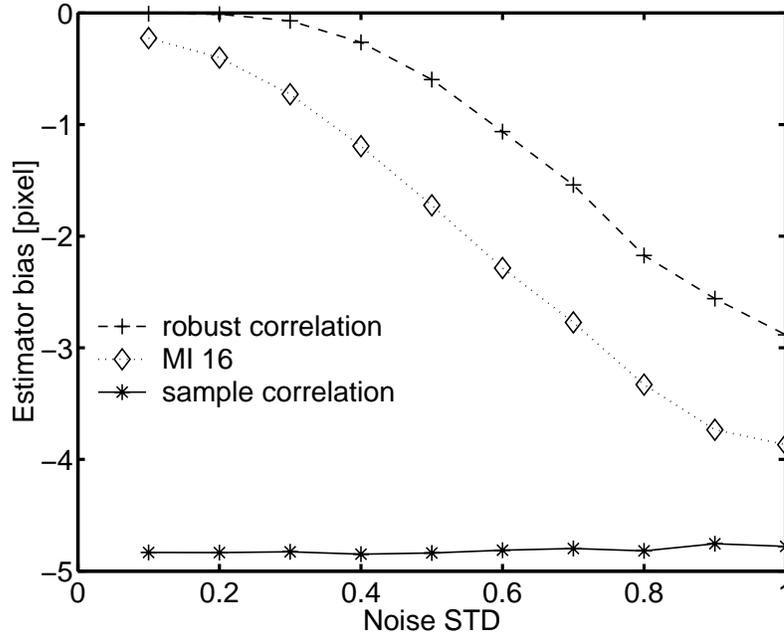


Figure 3.7: Bias of the translation estimators vs Gaussian noise levels.

Since one may trade off robustness and efficiency by changing design parameter (*i.e.*, underlying pdf for robust correlation coefficient and the number of bins for MI method), we plotted bias-variance plot of the estimators to evaluate the performance of each method.

Fig. 3.9 shows bias-variance plots of low noise case (lower part) with noise STD=0.2 and high noise case (upper part) with noise STD=0.4. Design parameters for the robust correlation method were  $\delta = 2, 3, 4, 5, 6, 8, 10$  and for the MI method were bin size  $8 \times 8, 10 \times 10, 12 \times 12, 16 \times 16, 20 \times 20, 24 \times 24, 28 \times 28$ .

For the robust correlation method, the bias is decreased as  $\delta$  increased while the variance increased. This is consistent with analytical results since smaller  $\delta$  implies “heavier tail” underlying pdf. Note that the robust correlation coefficient was almost the same as the

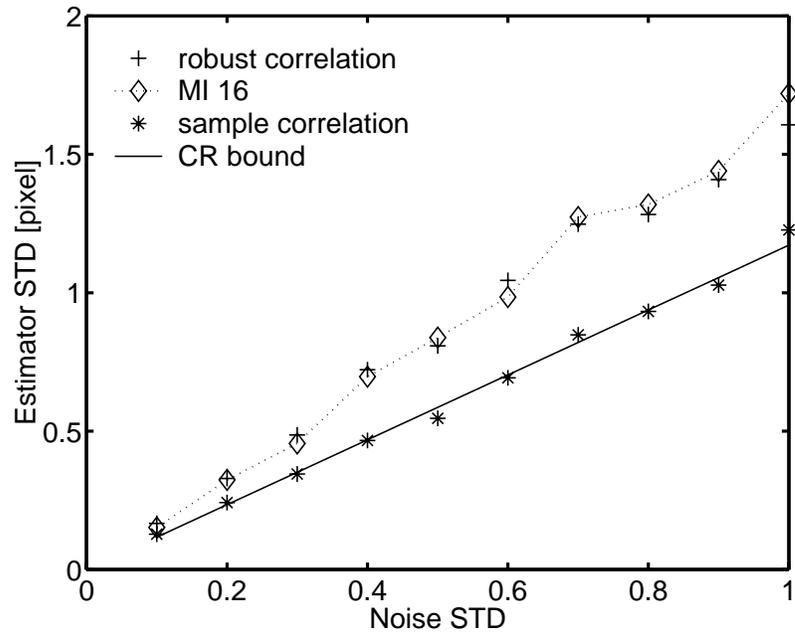


Figure 3.8: Standard deviations of the translation estimators vs Gaussian noise levels.

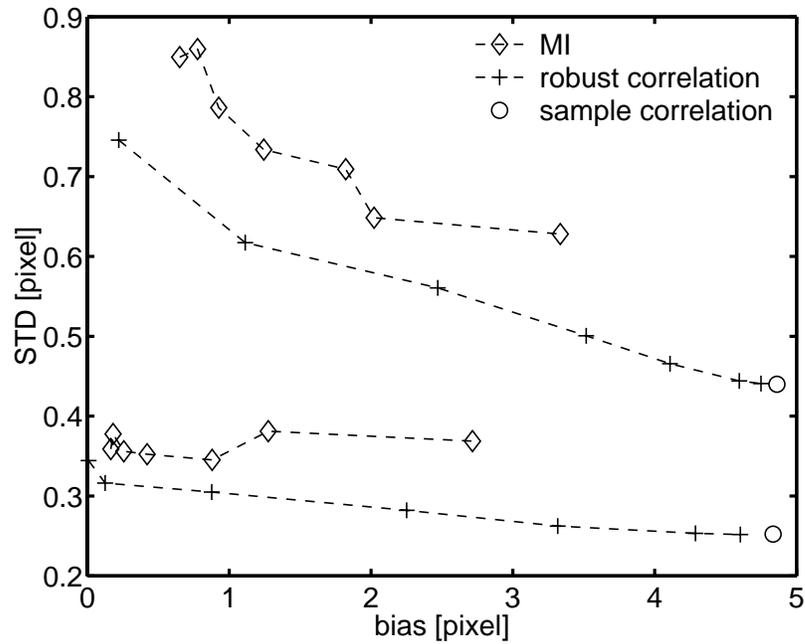


Figure 3.9: Standard deviations vs bias of the translation estimators (upper part for noise STD=0.4, lower part for noise STD=0.2).

sample correlation coefficient when  $\delta = 10$ . For the MI method, roughly speaking, the bias is decreased as the number of bin increased (*i.e.*, less smoothing) while variance increased. However, for low noise signal, the bias-variance characteristic of the MI method was very irregular. The performance of the robust correlation method was better than the MI method since the bias of the robust correlation method was smaller at the same variance or the variance was smaller at the same bias.

### 3.3.2 2D MRI Registration

We applied the sample correlation, the robust correlation and the MI based image registration method for registering two functional MRI images acquired with a spiral  $k$ -space trajectory. Both images were reconstructed from the same raw data but one image was reconstructed with field inhomogeneity correction [54] while the other was without correction, so the true registered geometric transformation is identity.

Because of field inhomogeneity, some part of the uncorrected images are occluded. We have chosen these images to test the registration methods since the true registered position (*i.e.*, identity transformation) is known, yet the occlusion can be considered as outliers. The robustness of MI method for occluded images has been investigated previously [22].

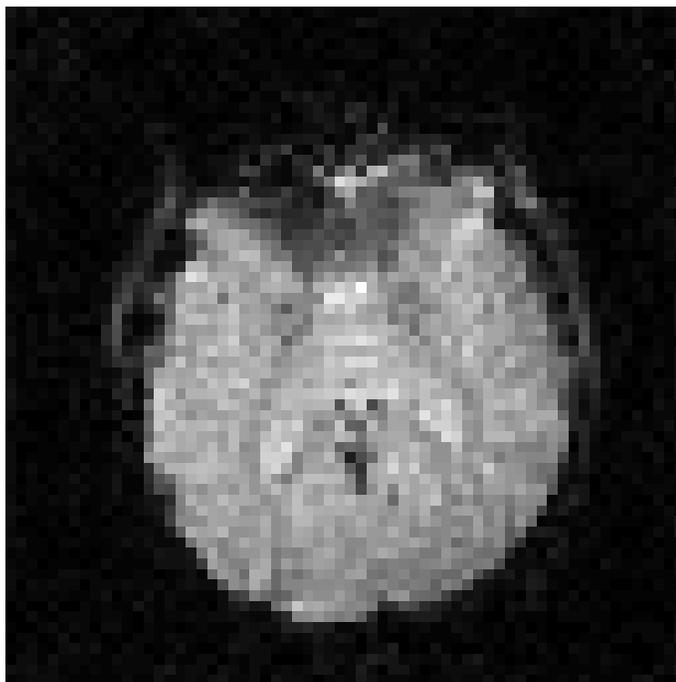


Figure 3.10: Reference MRI image (with field inhomogeneity correction).

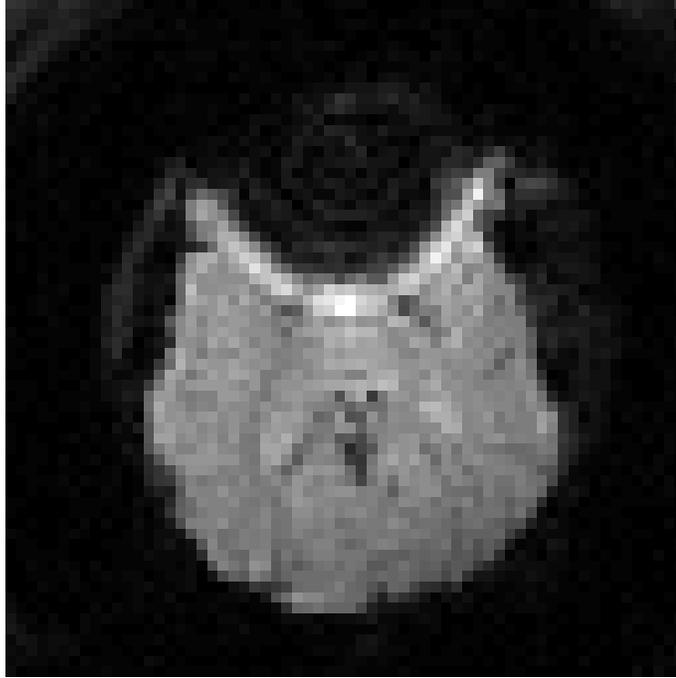


Figure 3.11: Target MRI image (without field inhomogeneity correction).

Fig. 3.10 and Fig. 3.11 show the reference image and the target image. The anterior of the brain shows signal void in target image that is corrected in the reference image.

We plotted the change of each similarity measure as a function of vertical translation and scale. Fig. 3.12 shows that the sample correlation coefficient is maximized at incorrect position since if we translate the reference image down, brighter pixels correspond the brighter pixels in target image better. Moreover, if we shrink the image, the correlation coefficient increases more since the brighter pixel region in Fig. 3.10 is larger than Fig. 3.11. As a result, the sample correlation coefficient is maximized around 5% scaling down and -0.9 pixel translation. As expected from the analysis, the sample correlation based method is easily biased due to the outliers.

Fig. 3.13 and Fig. 3.14 show that the bias is reduced in both MI method and robust correlation method.

We tested the bias and variance of each estimator by adding Rician noise. The noise in MRI raw data is complex Gaussian, whereas the noise in reconstructed magnitude images is Rician [55]. Table 3.1 shows the empirical means and standard deviations based on 100 registration trials of the three cost functions, for the case case of a horizontal ( $t_x$ ) and vertical ( $t_y$ ) translation and vertical scaling ( $k_y$ ). As expected, the sample correlation method had

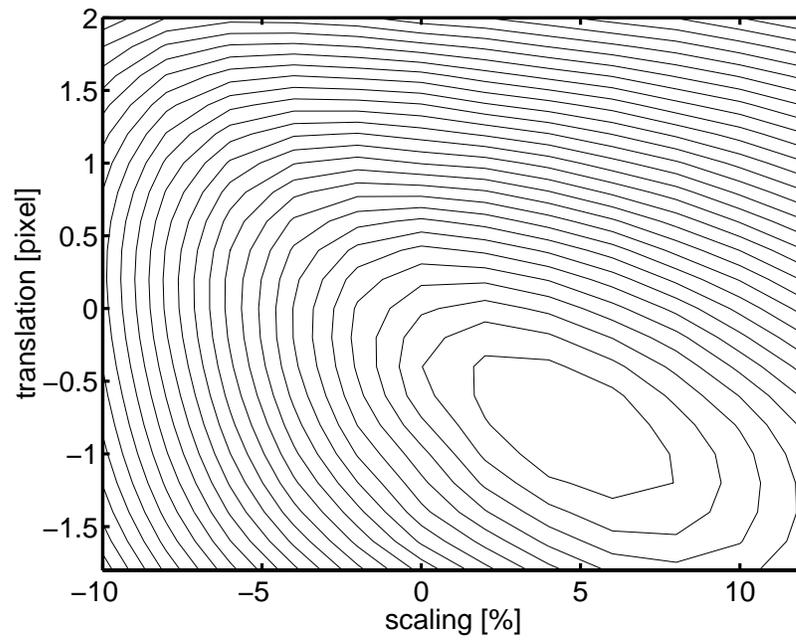


Figure 3.12: Contour plot of sample correlation coefficient vs vertical scaling and translation.

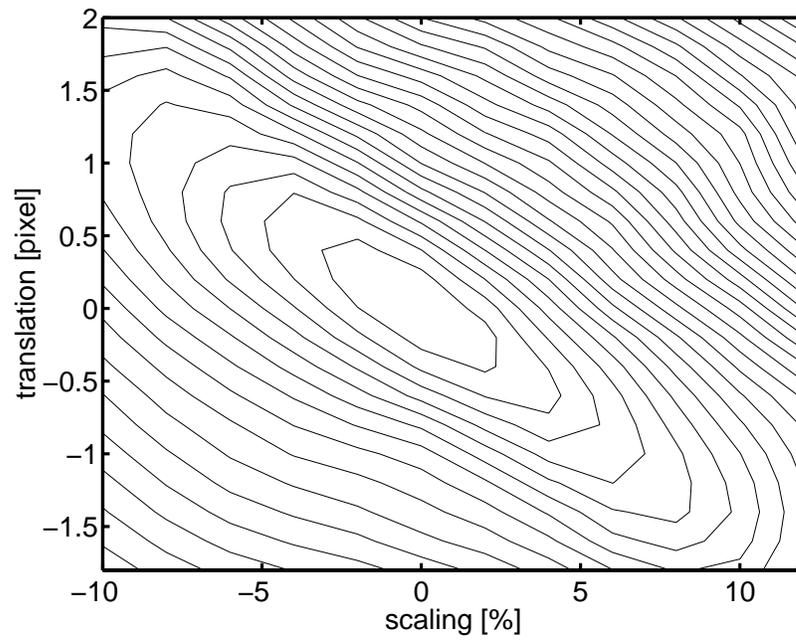


Figure 3.13: Contour plot of robust correlation coefficient vs vertical scaling and translation.

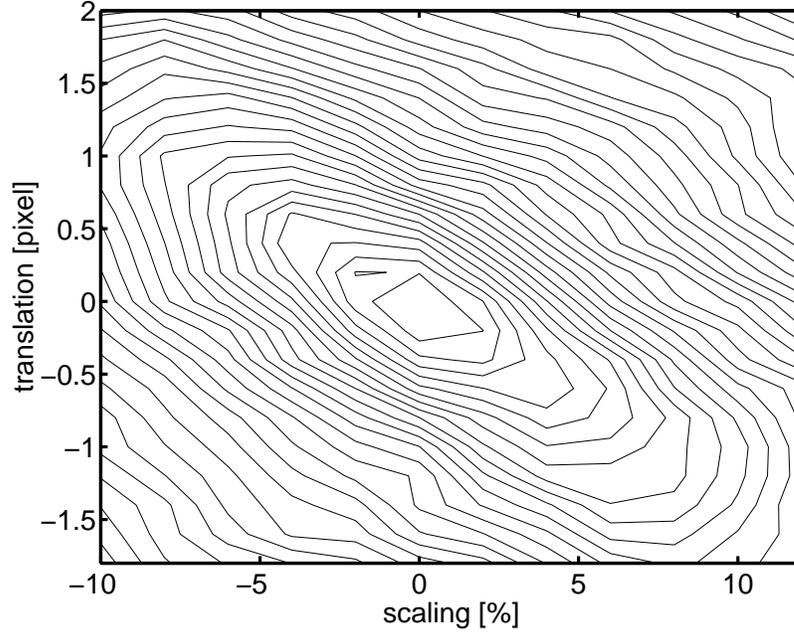


Figure 3.14: Contour plot of estimated MI ( $32 \times 32$  interpolated joint histogram) vs vertical scaling and translation.

the smallest variance but the largest bias due to the outliers. The robust correlation method was the most robust and had smaller variance than the MI method. When we increased the number of histogram bins for the MI method, the robustness was improved but the efficiency was degraded. These trade-offs are consistent with the 1D simulations.

### 3.3.3 3D/2D Phantom Study

We previously conducted an anthropomorphic phantom experiment to evaluate the performance of the set-up estimators by 3D/2D image registration [12]. The estimation problem was estimating six parameters that were rotations and translations along the X,Y,Z axis using two orthogonal images. We used the same data set for this study but used only one lateral image to estimate one rotation parameter and two translation parameters. We chose the lateral image that has outliers generated by the effect of the radiotherapy table. For this research, the other three parameters were kept fixed at the “ground truth” position that were established by the most accurate marker-based method using eleven 1mm diameter lead markers attached on the surface of the phantom [12].

A  $512 \times 512 \times 398$  voxel CT image with  $0.9375 \times 0.9375 \times 1$  mm spacing was acquired on

Table 3.1: Mean (and STD) of estimated registration parameters for 2D-2D MRI registration.

(The unit for translation parameter is pixel and for scaling parameter is unitless.)

similarity measure	$t_x$	$t_y$	$k_y$	SNR
sample	0.41 (0.07)	-2.82 (0.10)	0.05 (0.01)	27.7 dB
MI (16 × 16 histogram)	0.14 (0.14)	-1.10 (0.17)	0.01 (0.01)	
MI (24 × 24 histogram)	0.07 (0.21)	-0.10 (0.24)	0.01 (0.01)	
robust ( $\delta = 2$ )	0.03 (0.10)	0.03 (0.12)	0.01 (0.01)	
sample	0.12 (0.05)	-2.89 (0.28)	0.05 (0.01)	13.8 dB
MI (16 × 16 histogram)	0.41 (0.21)	-1.90 (0.69)	0.02 (0.01)	
MI (24 × 24 histogram)	0.07 (0.34)	-1.41 (0.90)	0.02 (0.02)	
robust ( $\delta = 2$ )	0.10 (0.21)	-0.10 (0.52)	0.01 (0.01)	

a GE CT/i scanner with a 140 kv x-ray source. Tattoos were drawn on the phantom where three alignment laser planes crossed the phantom to facilitate consistent set-up in a treatment room. Next, the phantom was moved to the treatment room and it was set up at the isocenter by manually aligning tattoos to three laser planes in the treatment room. Four radiographs were obtained from different angles by rotating the x-ray source and Varian Portal Vision amorphous silicon active matrix flat panel image detector in 30° increments. For 90° view, we acquired 10 repeated radiographs without realignment for evaluating the effect of noise on the estimator. The x-ray source voltage was 6 MV and the detector size was 512×384 pixels with 0.78mm×0.78mm spacing. We used only radiograph from 90° (*i.e.*, lateral image) for the correlation-based methods and the MI-based method. However, to enhance the accuracy of the “ground truth”, we used all four radiographs for the fiducial marker-based method. For all other methods except for the marker-based method, the planning CT image was down-sampled by four along each axis to reduce computation time and memory usage.

For image registration, while geometrically transforming the CT image, we computed DRR(Digitally Reconstructed Radiograph) of the transformed CT from the same angle as the radiograph. The registration is achieved by maximizing the similarity measure between such DRR and radiograph. We used only the central 400×300 sub-image of the DRR and

the radiograph to avoid the effect of the markers which are not usually used in clinical practice. We have established the geometry of the EPID imaging systems by determining radiation field edges using simple thresholding method [31].

Fig. 3.15 shows the radiograph and Fig. 3.16 shows the DRR at the registered position. We can see the effect of the radiotherapy table around the rightmost parts of the radiograph. Pixels around the right most parts of the radiograph are brighter than those from the DRR due to the radiotherapy table.



Figure 3.15: Measured lateral radiograph for 3D-2D registration.

Fig. 3.17 shows the estimated joint histogram from the registered DRR and radiograph. The histogram has a dominant linear distribution and an outlier distribution due to the presence of the radiotherapy table. Fig. 3.18 shows the weighting function that clearly reduces the influence of the outliers.

We repeated 10 estimations using 10 acquisitions of the radiograph. Table 3.2 summarizes the experimental results. The experimental results were consistent with the previous simulation results. The robust correlation coefficient using (3.17) with  $\delta = 2$  was the most robust one. The sample correlation based method was the worst in terms of the robustness but the best in terms of the variance. Interestingly, the MI method showed small variance and small bias as well. We think this was because the shape of the estimated joint pdf was



Figure 3.16: Lateral DRR computed from 3D CT volume.

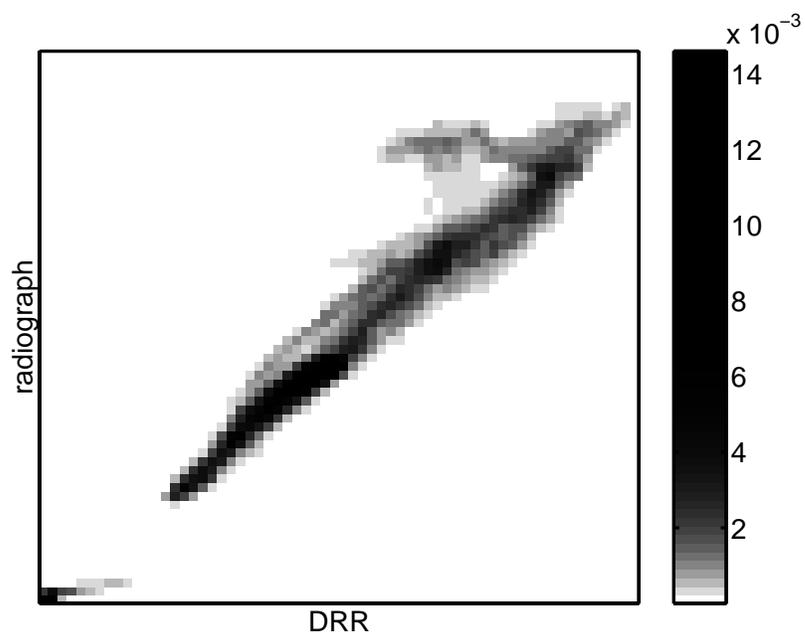


Figure 3.17: Estimated joint pdf at the registered position ( $64 \times 64$  interpolated joint histogram).

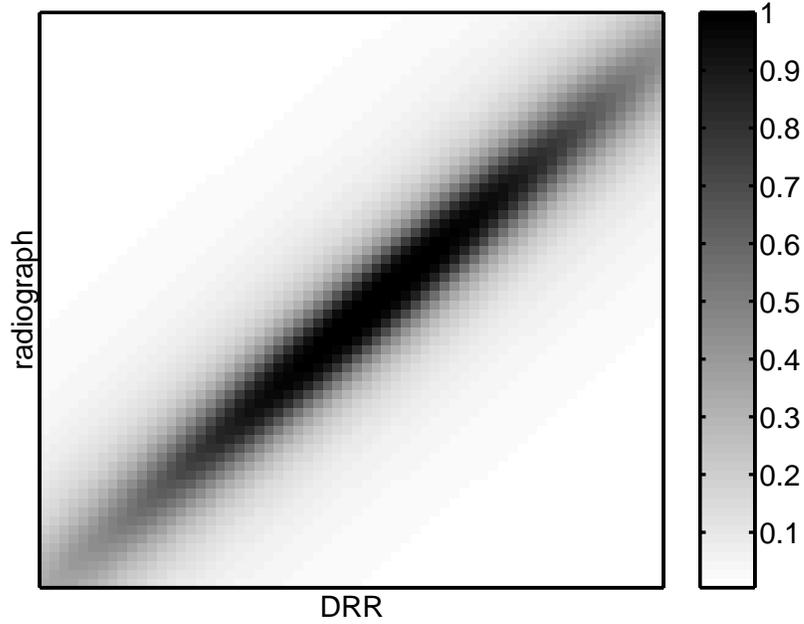


Figure 3.18: Weighting function at the registered position (evaluated at  $64 \times 64$  joint histogram domain).

Table 3.2: Mean (and STD) of estimated set-up parameters for 3D-2D registration.  
(The unit for rotation parameter is degree and for translation parameter is mm.)

Similarity measure	$\phi_x$	$t_y$	$t_z$
sample	1.39 (0.02)	-2.06 (0.03)	2.11 (0.03)
robust ( $\delta = 2$ )	0.95 (0.08)	-0.09 (0.09)	0.58 (0.23)
robust ( $\delta = 4$ )	1.19 (0.10)	-1.52 (0.07)	1.74 (0.13)
MI ( $32 \times 32$ histogram)	0.98 (0.08)	-0.55 (0.09)	0.81 (0.14)
MI ( $64 \times 64$ histogram)	0.86 (0.07)	-0.44 (0.07)	0.73 (0.13)
MI ( $64 \times 64$ histogram w/o interpolation)	0.90 (0.36)	-0.37 (0.16)	0.63 (0.38)

close to a normal distribution and the number of samples from outlier was small.

### 3.4 Discussion

Statistical properties such as bias, robustness, efficiency are very important in designing image registration methods. In previous investigations, the bias of the sample correlation method for intra-modality image registration and that of the MI method for multi-modality registration have been studied extensively empirically<sup>4</sup> [12, 18, 20–22]. Some authors also reported that the MI method is more robust than the sample correlation coefficient method [12, 56].

It has been well known that the sample correlation is a natural similarity measure for intra-modality image registration [18, 20, 56] and the MI method performs well for both the intra-modality [55, 56] and the multi-modality image registrations [21–23]. If we consider only the intra-modality image registration, we point out that those two most well known similarity measures have drawbacks such as the lack of robustness for the sample correlation method and the statistical inefficiency for the MI method. Moreover, even though the robustness of the MI method can be explained qualitatively, the MI method may not be robust for some images since the MI method does not reject nor reduce the effect of outliers.

We have proposed the robust correlation coefficient method to improve the robustness of the sample correlation method without compromising efficiency very much. By mean and variance approximations, we were able to show analytically that the robust correlation method has improved robustness but larger variance comparing to the sample correlation method.

There is a direct relationship between the statistical properties of the robust correlation method and design parameters. One may improve the robustness by defining “heavier tail” underlying pdf model  $f_0$  in (3.10) of the expense of the efficiency. In our study, we presented the results with different parameters. However, it is desirable to determine the parameters automatically in practice. Investigating such an automatic method is one of our future research topics.

Compared to the robust correlation method, it is more difficult to relate the design parameters of the MI method to the statistical properties. In our experiments, the MI

---

<sup>4</sup>Since many previous studies ignored the effect of noise, bias was called *registration error*.

method was more robust when larger number of bins was used while the variance was larger. Roughly speaking, the larger number of bins is equivalent to the narrower Parzen window, *i.e.*, less smoothing. The variance when using histogram without interpolation was generally larger than interpolated histogram. We also observed that it may not be a good approach to increase the number of bins too much for better robustness since many local maxima were generated as the number of bins increased.

More investigation is required to design better kernel function for MI method. To our knowledge, designing kernel function for MI method in terms of the statistical properties has not been investigated much although one method has been proposed in a different point of view [21].

It is challenging to compare image registration methods since the properties of the methods depend on both design parameters and the images being registered. In spite of the difficulties, if we compare the performance of the robust correlation method and the MI method based on our simulations and experiments, the robust correlation method performed better than the MI method in our 1D simulation and 2D MRI simulation, *i.e.* the variance of the robust correlation method was smaller at the same bias or the bias was smaller at the same variance. Interestingly, the MI method was very efficient in 3D/2D phantom experiments. We suspect this was because the estimated pdf shape was close to the normal distribution as argued in Section 3.2.4.

We think that there exist some advantages of the robust correlation method over the MI method. First of all, the robust correlation method always has robustness since it reduces or rejects the effect of outlier samples based on the statistical distances. As a result, the registration is performed mostly by the inlier samples. In contrast, the MI method relies on the behavior of estimated pdf without explicit excluding the effect of outlier samples. Therefore, the robustness is very dependent on image characteristics. We suspect that the MI method may not be very robust for some images. Similar observation can be made in terms of variance, too. The variance of the MI method depends greatly on the image characteristics. Another advantage of the robust correlation method is that design parameters are directly related to the properties while the relationship of the MI method is not very clear analytically.

The disadvantage of the robust correlation method is computation time since the robust correlation coefficient is determined by another numerical optimization routine. The com-

putation time is dependent on images and underlying pdf. It may be possible to reduce the computation time by using a robust correlation coefficient estimator that does not require an iterative procedure.

One may argue that the performance of the MI method can be improved by designing different MI estimation method and/or using more generalized Rényi entropy [22,35]. Also, one may improve the performance of the robust correlation method using different robust techniques to estimate the correlation coefficient. There can be many other ways to design the robust correlation coefficient including  $M$ -estimates and  $R$ -estimates [42–44, 46–48, 57]. More thorough investigations and comparisons between the robust correlation methods and the MI methods are deferred to future study. We think that such comparisons should be made considering images being registered. For comparing the robustness of each method, *influence function* [24] of each estimator with given images might be a useful tool.

We have argued the efficiency of the sample correlation method under the i.i.d. Gaussian noise assumption without outliers. Even though the noise was not Gaussian in practice (Rician for MRI simulation, Poisson for 3D/2D experiment) and outliers were present, the sample correlation method was the most efficient one in every simulation and experiment. We suspect this is because the Rician and Poisson noise are approximately Gaussian although the variance at each pixel is different.

Analysis using mean and variance approximation provided qualitative arguments about the statistical properties of the intensity based image registration methods. The results for the sample correlation and the robust correlation method were concise and insightful. Moreover, since we have chosen the  $M$ -estimation method for robust correlation method, not only we were able to analyze the robust correlation and the sample correlation within the same framework but also to represent the MI method as a type of weighted correlation method. It would be desirable to describe the statistical properties in terms of design parameters. For example, if one can approximate the bias and the variance of the MI method in terms of the kernel function, that approximation may be very useful for design. Since the statistical properties of the image registration methods have not been investigated analytically much, we think that our analyses can be a first step for further investigations.

### 3.5 Conclusion

We have introduced robust correlation coefficient as a novel intensity based similarity measure to improve the robustness of the sample correlation coefficient based image registration while not degrading the statistical efficiency much. In 1D simulation, 2D and 3D experiments, the proposed method reduced the bias of the sample correlation method caused by outliers. The robust correlation coefficient may be an effective similarity measure for intra-modality image registration task where the presence of the outliers is unavoidable such as set-up estimation for radiotherapy and image-guided surgery.

## CHAPTER 4

### Nonrigid Registration

#### 4.1 Introduction

Nonrigid image registration is a useful technique for estimating patient anatomy change, aiding diagnosis, atlas based segmentation, etc. In general, nonrigid registration is achieved by deforming one image using a deformation model. Several different models of geometric transformations using TPS(Thin Plate Spline) [23, 58], B-splines [28, 30, 59] and sinusoidal functions [60] have been proposed and investigated.

In this study, we have chosen cubic B-spline based deformation model since compactly supported basis functions can be useful for representing local deformation. Moreover cubic B-splines have desirable properties such as twice differentiability and ability to represent rigid translation while computation time requirement is modest.

As explained in Chapter 1, Jacobian determinants of the estimated deformation must be positive. To prevent the Jacobian determinant from being negative, regularizing penalty functions have often been introduced in nonrigid registration. For example, penalizing roughness penalty [28], bending energy [29], and small Jacobian determinant [30] have been proposed and investigated. In nature, most penalty functions rely on the fact that small gradient values are helpful to prevent the negative Jacobian determinant.

The penalty based methods have several drawbacks as explained in Chapter 1. To remedy those problems, one previous study investigated a method that bounds the magnitude of gradient components by bounding the parameters (*i.e.*, coefficients) [61]. Sufficient condition to guarantee the positive Jacobian determinants was derived by Neuman series arguments and the conditions were achieved using a constrained optimization subject to

a box constraint in the parameter space. The problem of this method is that the search space is too much restricted. For example, large deformations with small gradients are not allowed.

In this dissertation, we extend the results of the previous investigation. First, we derive closed form expressions for possible minimum and maximum Jacobian determinants in terms of  $x, y, z$  axis gradient bounds analytically. Next, we introduce a constraint set in the parameter space that allows large deformations but achieves the gradient bounds. Since the constraint set is not a simple box constraint, we solve the optimization problem using the gradient projection method with Dykstra’s cyclic projection algorithm.

We have chosen the gradient projection method to solve the constrained optimization problem since we wanted to remove *line search* procedure that requires much computation time. The gradient projection method first computes unconstrained next step based on the gradient of the cost function and then projects the unconstrained next step onto the constraint set to compute actual next step. To implement the projection step, we have used Dykstra’s *cyclic projection* algorithm [62], which is a useful method to compute the orthogonal projection onto the intersection of many convex sets. Since the constraint set to bound the gradients was a convex set defined by the intersection of convex half-spaces, on which the orthogonal projections are easily computed, we were able to use Dykstra’s algorithm effectively.

Besides the gradient projection method, many different optimization methods can be applied for solving the constrained optimization problem. For example, constrained versions of the *Quasi Newton* or *conjugate gradient* method may be applied effectively [63]. Other methods such as barrier function methods and interior point methods can be also used [64]. Investigating such optimization methods in comparison with the gradient projection method is deferred to future study.

This chapter is organized as follows. Section 4.2 formulates the nonrigid registration problem using cubic B-splines and describes conditions to ensure positive Jacobian determinants. The constraints set in the parameter space to ensure positive Jacobian determinant is also introduced in Section 4.2. Section 4.2 also presents the gradient projection method and Dykstra’s algorithm to solve the constrained optimization problem. Section 4.3 presents the results of simulation with known deformation and experiment for inhale/exhale lung CT registration. We compare our proposed method with existing methods in Section 4.4.

The appendices present proofs for the two propositions claimed in Section 4.2.

## 4.2 Theory

The goal of nonrigid image registration is to determine a nonrigid geometric transformation, that aligns two given images, denoted  $A(x, y, z)$  and  $B(x, y, z)$ , where  $x, y, z$  denote the spatial coordinates. If we parameterize the geometric transformation using  $\theta$ , the registration problem becomes a parameter estimation problem. This problem is usually solved by maximizing a similarity measure and *a priori* information in a MAP (Maximum *A posteriori*) estimator framework as following:

$$\hat{\theta} = \arg \max_{\theta} \Phi(A(T_{\theta}(\cdot)), B(\cdot)) + \beta \mathcal{R}(\theta), \quad (4.1)$$

where,  $T_{\theta} : R^3 \rightarrow R^3$  is a nonrigid deformation model,  $\Phi(A(\cdot), B(\cdot))$  is a similarity measure,  $\beta$  is a regularization parameter and  $\mathcal{R}(\theta)$  is *a priori information*.

Note that the similarity measure can be either feature based or anatomy based. In this dissertation, we focus on intensity based image registration methods that use a similarity measure based on image intensity values.

### 4.2.1 Image Model

In practice, registration is performed using finite number of samples  $A(x_i, y_j, z_k)$  and  $B(x_i, y_j, z_k)$ ,  $i = 1, \dots, N_x, j = 1, \dots, N_y, k = 1, \dots, N_z$ , where  $N_x, N_y, N_z$  are the number of voxels in  $x, y, z$  direction. Since the image intensity values of the deforming image need to be evaluated in between grid points frequently during optimization, interpolation is essential. To address this issue, we have adopted a cubic spline based continuous model of the image that showed good performance in a previous investigation [30]. In addition to the interpolation, one can compute the gradient and Hessian of the objective function effectively by adopting the image model since closed form expressions of the gradient and Hessian can be determined based on the image model [65].

### 4.2.2 Cubic B-spline based Deformation

A variety of nonrigid deformation models have been proposed and investigated [23, 28, 30, 58–60]. In general, the nonrigid deformation is modeled with finite sum of basis functions multiplied by coefficients. During optimization, the coefficients are changed to maximize a similarity measure. Therefore, one may consider the nonrigid registration problem as a parameter estimation problem. Not only the coefficients but also the locations and the number of basis functions can be also considered as parameters that are changed during the optimization.

The basis functions can be classified into infinite support basis function and finite support basis function. The infinite support basis function such as TPS or sinusoidal basis function is a natural choice for representing global deformations while the finite support basis function is more appropriate for representing local deformations. In addition to the capability of representing local deformations, it has been reported that the finite basis function based deformation may achieve registration faster than the infinite basis function based deformation [29].

In this study, we focus on the B-spline based deformation model that is one of the most well known finite basis function based models.

We model the deformation  $T_\theta(\cdot)$  with  $\theta = (\theta^x, \theta^y, \theta^z)$  as follows:

$$T_\theta(x, y, z) = (x + f_{\theta^x}(x, y, z), y + g_{\theta^y}(x, y, z), z + h_{\theta^z}(x, y, z)), \quad (4.2)$$

$$f_{\theta^x}(x, y, z) = \sum_{ijk \in K_x} \theta_{ijk}^x \beta^3\left(\frac{x}{T_x} - i\right) \beta^3\left(\frac{y}{T_y} - j\right) \beta^3\left(\frac{z}{T_z} - k\right), \quad (4.3)$$

$$g_{\theta^y}(x, y, z) = \sum_{ijk \in K_y} \theta_{ijk}^y \beta^3\left(\frac{x}{T_x} - i\right) \beta^3\left(\frac{y}{T_y} - j\right) \beta^3\left(\frac{z}{T_z} - k\right), \quad (4.4)$$

$$h_{\theta^z}(x, y, z) = \sum_{ijk \in K_z} \theta_{ijk}^z \beta^3\left(\frac{x}{T_x} - i\right) \beta^3\left(\frac{y}{T_y} - j\right) \beta^3\left(\frac{z}{T_z} - k\right). \quad (4.5)$$

where,  $K_x, K_y, K_z$  are the sets of locations for  $\theta^x, \theta^y, \theta^z$ , and  $T_x, T_y, T_z$  are expansion parameters.

In this model, the B-spline basis function has wider region of support as expansion parameters become larger, which implies more global deformation. For many cases, the choice of the expansion parameters is related to image resolution and often combined with *multi-resolution* optimization strategy. For example, one may register low resolution images

first using small number of wide basis functions and advance to high resolution images using large number of narrow basis functions.

### 4.2.3 Invertibility

One important physical constraint for estimated deformation is that the estimated deformation should be invertible since physical deformations are invertible. By *inverse function theorem* [14], the invertibility is guaranteed if Jacobian determinant is nonzero. Moreover, since the determinant is continuous in spatial domain, Jacobian determinants should be either negative or positive to make the deformation invertible everywhere. In general, the Jacobian determinants must be positive since it is assumed that there are regions with identity transformations (*i.e.* Jacobian determinant is unity).

The Jacobian matrix of nonrigid transformation can be written as follows:

$$J(x, y, z) = \begin{bmatrix} 1 + \frac{\partial f(x,y,z)}{\partial x} & \frac{\partial f(x,y,z)}{\partial y} & \frac{\partial f(x,y,z)}{\partial z} \\ \frac{\partial g(x,y,z)}{\partial x} & 1 + \frac{\partial g(x,y,z)}{\partial y} & \frac{\partial g(x,y,z)}{\partial z} \\ \frac{\partial h(x,y,z)}{\partial x} & \frac{\partial h(x,y,z)}{\partial y} & 1 + \frac{\partial h(x,y,z)}{\partial z} \end{bmatrix}. \quad (4.6)$$

In one previous investigation [61], it has been shown that the possible minimum Jacobian determinant is greater than zero if the magnitude of each gradient is bounded by  $\frac{1}{3}$ .

We would like to extend the previous result by deriving closed form expressions for possible minimum and maximum Jacobian determinants in terms of gradient bounds for  $f(x, y, z)$ ,  $g(x, y, z)$  and  $h(x, y, z)$ , respectively. By doing this, one may bound the minimum Jacobian determinant by other quantities than zero by bounding the gradients. We think that it may be useful for some applications to bound the minimum Jacobian determinant by a specified quantity. Moreover, we think that understanding the possible maximum Jacobian determinant is also important since it implies the possible maximum local volume expansion.

The following Proposition 1 relates gradient bounds to the possible minimum and maximum Jacobian determinants.

*Proposition 1.* Suppose that  $\left| \frac{\partial f(x,y,z)}{\partial x} \right| \leq k_f$ ,  $\left| \frac{\partial f(x,y,z)}{\partial y} \right| \leq k_f$ ,  $\left| \frac{\partial f(x,y,z)}{\partial z} \right| \leq k_f$ ,  $\left| \frac{\partial g(x,y,z)}{\partial x} \right| \leq k_g$ ,  $\left| \frac{\partial g(x,y,z)}{\partial y} \right| \leq k_g$ ,  $\left| \frac{\partial g(x,y,z)}{\partial z} \right| \leq k_g$  and  $\left| \frac{\partial h(x,y,z)}{\partial x} \right| \leq k_h$ ,  $\left| \frac{\partial h(x,y,z)}{\partial y} \right| \leq k_h$ ,  $\left| \frac{\partial h(x,y,z)}{\partial z} \right| \leq k_h$ , for

$\forall x, y, z$ . If  $0 \leq k_f, k_g, k_h \leq \frac{1}{2}$ , then the possible minimum  $\det J(x, y, z)$  is  $1 - (k_f + k_g + k_h)$  and the possible maximum is  $(1+k_f)(1+k_g)(1+k_h) + (1+k_f)k_gk_h + (1+k_g)k_fk_h + (1+k_h)k_fk_g$ .

*Proof:* See Appendix F.

By Proposition 1, one may bound the minimum and maximum Jacobian determinants by bounding the magnitude of each gradient component. For example, if one set  $k = k_f = k_g = k_h$ , the minimum is  $1 - 3k$  and the positive Jacobian determinant is guaranteed if  $k \leq \frac{1}{3}$  as argued in [61] by Neuman series. For that case, one can understand the possible maximum Jacobian determinant is approximately 2.8148 by the proposition.

If *a priori* information such as minimum/maximum gradient of *true* deformation is available, one can effectively select the search space using Proposition 1. Note that Proposition 1 does not state that the minimum/maximum Jacobian determinants are defined by the expressions but the *possible* minimum/maximum Jacobian determinants do. In other words, even if some of the gradient components exceed the bounds, the resulting Jacobian determinant can be still positive since the minimum occurs only when every gradient component has boundary values.

#### 4.2.4 Constraints in Parameter Space

One may enforce the gradient bounds using unconstrained optimization methods with proper penalization of gradient values exceeding the bounds such as in *interior point method* or *barrier function methods* without converting the gradient bound conditions into constraints in the parameter space. However, there are several drawbacks for this method.

First, regularization parameter may need to be readjusted if images are changed. Considering the fact that parameter tuning is usually done manually, regularization parameter tuning is cumbersome. Another problem is even if the gradients are effectively bounded at grid points where the gradients are evaluated, there is no guarantee that gradients are bounded in between the grids. In the previous investigation using a unconstrained optimization method, it was assumed that the Jacobian determinant was positive everywhere if it was positive at each grid point [30]. There is another investigation to ensure positive Jacobian determinants everywhere [66].

The problems of the penalty based methods may be overcome by introducing a constrained optimization. For example, if the coefficients of the deformation are bounded, the

gradients are bounded everywhere [61]. Therefore, if a similarity measure is maximized subject to the coefficient bound constraint, the gradient of the resulting deformation is bounded. However, this method restricts the search space too much since large deformations with small gradient components are excluded from the search space.

To address this problem, we derive sufficient conditions in the parameter space to bound the gradient components while allowing large deformations. Our approach is based on not bounding the coefficients but bounding the differences of two consecutive coefficients in  $x, y, z$  direction. It is natural to bound the differences to bound the gradients. Since we do not bound the coefficients, large deformations are included in our search space as long as their gradient components do not exceed the gradient bounds.

To derive gradient bound condition, first consider  $\left| \frac{\partial f(x,y,z)}{\partial x} \right|$  for the cubic B-spline deformation model and an inequality to bound  $\left| \frac{\partial f(x,y,z)}{\partial x} \right|$  by  $b$  represented as follows:

$$\begin{aligned} \left| \frac{\partial f(x_l, y_l, z_l)}{\partial x} \right| &= \left| \sum_{ijk \in K_x} \theta_{ijk}^x \frac{\partial \beta^3(\frac{x_l}{T_x} - i)}{\partial x} \beta^3(\frac{y_l}{T_y} - j) \beta^3(\frac{z_l}{T_z} - k) \right| \\ &\leq \sum_{ijk \in K_x} \left| \theta_{ijk}^x \frac{\partial \beta^3(\frac{x_l}{T_x} - i)}{\partial x} \beta^3(\frac{y_l}{T_y} - j) \beta^3(\frac{z_l}{T_z} - k) \right| \leq b. \end{aligned} \quad (4.7)$$

Note that a parameter set that satisfies condition (4.7) at a given location  $(x_l, y_l, z_l)$  is a polyhedral convex set. Therefore, one may optimize the similarity measure subject to (4.7) to bound the  $\frac{\partial f(x,y,z)}{\partial x}$  at  $(x_l, y_l, z_l)$ . However, this approach may require significant amount of computation time if one would like to bound gradients at many different locations. Since we desire to bound the gradient component everywhere, it is more practical to derive a more sufficient condition than (4.7) to bound  $\frac{\partial f(x,y,z)}{\partial x}$  everywhere.

The following proposition ensures that if the maximum differences between two consecutive coefficients in the  $x, y, z$  directions are bounded, the gradients are bounded.

*Proposition 2.* If  $\left| \theta_{i+1,j,k}^x - \theta_{i,j,k}^x \right| \leq b, \forall_{ijk} \in K_x$ , then  $\left| \frac{\partial f(x,y,z)}{\partial x} \right| \leq \frac{b}{T_x}$ . Similarly, if  $\left| \theta_{i,j+1,k}^x - \theta_{i,j,k}^x \right| \leq b, \forall_{ijk} \in K_x$ , then  $\left| \frac{\partial f(x,y,z)}{\partial y} \right| \leq \frac{b}{T_x}$  and  $|\theta_{i,j,k+1} - \theta_{i,j,k}| \leq b, \forall_{ijk} \in K_z$  implies  $\left| \frac{\partial f(x,y,z)}{\partial y} \right| \leq \frac{b}{T_x}$ . Similar bounds for  $\theta_{i,j,k}^y, \theta_{i,j,k}^z$  imply bounds on the gradients of  $h(x, y, z)$  and  $g(x, y, z)$ .

*Proof:* See Appendix G.

Using Proposition 2, one can design parameter sets that achieve the gradient bounds by specified quantities. Note that the coefficient difference of longer basis function is larger than shorter basis function for the same bound since the distance between two knots is larger.

#### 4.2.5 Constrained Optimization using Dykstra’s Algorithm

To ensure the conditions in Proposition 2 are not violated, one may use unconstrained optimization with penalty functions that have very large values for coefficient differences that exceed the bounds. Or one can use constrained optimization techniques subject to the conditions. The drawbacks of the unconstrained methods include that tuning of regularization parameter may be required and that convergence may be slow due to very steep slope of the penalty function. Considering those problems, constrained optimization methods may be attractive. There are many constrained optimization methods that can be applied to optimize (4.1) subject to the constraints in Proposition 2. For example, one may apply constrained version of *Conjugate Gradient*, *Quasi Newton* method, etc [63]. The drawback of the constrained methods is that it may require more computation time.

In this investigation, we propose the gradient projection method to solve the optimization problem. This method is guaranteed to converge with an appropriate selection of step size, if the objective function is convex [63]. We have chosen this method since it does not require *line search* of which computation requirement is intensive. For nonrigid registration problem with a large degree of freedom deformation model, the computation time for line search can be significant.

The gradient projection method with step size  $\alpha$  is defined as follows:

$$\theta^{n+1} = P_{\mathcal{K}}(\theta^n - \alpha \nabla_{\theta} \Phi(A, B; \theta^n)), \quad (4.8)$$

where,  $\mathcal{K}$  is the convex constraint set and  $P_{\mathcal{K}}$  denotes the orthogonal projection onto the convex set  $\mathcal{K}$ . To implement the gradient projection method, the orthogonal projection onto the convex constraint set must be computed. The convex set  $\mathcal{K}$  for the constraints developed in Proposition 2 can be written as follows:

$$\mathcal{H}_i = \{\theta \in X \mid \langle \theta, f_i \rangle \leq c_i\}, \quad i = 1, \dots, r \quad (4.9)$$

$$\mathcal{K} = \bigcap_{i=1}^r \mathcal{H}_i, \quad (4.10)$$

where,  $X$  is the parameter space,  $r$  is the number of constraints,  $f_i$  and  $c_i$  are appropriate vectors and scalars.

In general, determining an orthogonal projection onto a convex set is challenging except for some special cases. If the convex set  $\mathcal{K}$  is the intersection of convex sets on which the orthogonal projection can be computed easily such as (4.10), Dykstra's *cyclic projection* method is an effective tool to compute the orthogonal projection onto  $\mathcal{K}$  [62].

Dykstra's algorithm iterates sequential projections of a vector onto convex sets whose intersection is  $\mathcal{K}$ . Note that it is not difficult to compute an orthogonal projection onto a half space defined as (4.9). For polyhedral case, to determine the orthogonal projection of a given vector  $\theta$  onto the set  $\mathcal{K}$ , Dykstra's algorithm generates the sequence  $\{\theta_n\}$  as follows:

$$\begin{aligned} \theta_0 &= \theta, & \theta_n &= 0 \quad \text{when } m \leq 0, \\ \theta_m &= P_{\mathcal{H}_{[m]}}(\theta_{m-1} + e_{m-r}), & e_m &= \theta_{m-1} + e_{m-r} - \theta_m. \end{aligned}$$

Then, the Boyle-Dykstra convergence theorem shows that

$$\|\theta_m - P_{\mathcal{K}}(\theta)\| \rightarrow 0 \quad \text{as } m \rightarrow \infty. \quad (4.11)$$

For the special case of the polyhedral convex set such as (4.10), it has been shown that the Dykstra's algorithm converges geometrically [62].

In the gradient projection method using Dykstra's algorithm, there are two iterations for optimization. In each iteration for maximizing the similarity measure, the orthogonal projection onto  $\mathcal{K}$  should be computed iteratively using Dykstra's algorithm. Fortunately, we have found that Dykstra's algorithm converges within small number of iterations. However, since the number of the half spaces increases as the number of the coefficients increases, the computation time can be significant for large number of coefficients. One may reduce the computation time using a parallel version of Dykstra's method [67].

### 4.2.6 Penalty Function

Penalty functions have been often used to prevent Jacobian determinants from being negative. Such penalty functions include smoothness penalty [28], bending energy [29] and direct penalty for small Jacobian determinant [30]. In this study, we have investigated two penalty based methods for comparison purposes: penalizing the bending energy and penalizing small Jacobian determinants.

The penalty function plays a role of  $\mathcal{R}(\theta)$  in (4.1). Since weighting between the similarity term and the penalty term is determined by regularization parameter  $\beta$ , selection of  $\beta$  may greatly affect registration results.

#### Bending Energy

One can penalize the bending energy to regularize the estimated deformation. Minimizing bending energy may offer a flavor of TPS to the estimated deformation since the TPS interpolation function has minimum bending energy for 2D case [58]. The 3D bending energy  $E_b$  is defined as following [29]:

$$\begin{aligned}
E_b = & \iiint \left( \frac{\partial^2 f(x, y, z)}{\partial x^2} \right)^2 + \left( \frac{\partial^2 f(x, y, z)}{\partial y^2} \right)^2 + \left( \frac{\partial^2 f(x, y, z)}{\partial z^2} \right)^2 + \left( \frac{\partial^2 g(x, y, z)}{\partial x^2} \right)^2 \\
& + \left( \frac{\partial^2 g(x, y, z)}{\partial y^2} \right)^2 + \left( \frac{\partial^2 g(x, y, z)}{\partial z^2} \right)^2 + \left( \frac{\partial^2 h(x, y, z)}{\partial x^2} \right)^2 + \left( \frac{\partial^2 h(x, y, z)}{\partial y^2} \right)^2 \\
& + \left( \frac{\partial^2 h(x, y, z)}{\partial z^2} \right)^2 + 2 \left( \frac{\partial^2 f(x, y, z)}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f(x, y, z)}{\partial y \partial z} \right)^2 + \left( \frac{\partial^2 f(x, y, z)}{\partial z \partial x} \right)^2 \\
& + \left( \frac{\partial^2 g(x, y, z)}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 g(x, y, z)}{\partial y \partial z} \right)^2 + \left( \frac{\partial^2 g(x, y, z)}{\partial z \partial x} \right)^2 + \left( \frac{\partial^2 h(x, y, z)}{\partial x \partial y} \right)^2 \\
& + \left( \frac{\partial^2 h(x, y, z)}{\partial y \partial z} \right)^2 + \left( \frac{\partial^2 h(x, y, z)}{\partial z \partial x} \right)^2 dx dy dz. \tag{4.12}
\end{aligned}$$

Note that  $E_b$  is computed using finite sum at grid points instead of the integral, in practice.

#### Small Jacobian Determinants

A previous study investigated a penalty function that penalizes Jacobian determinant using exponential function [30]. Unlike the previous method, we designed a new penalty function that penalizes Jacobian determinants that are smaller than given threshold. We do not think that it is necessary to penalize local transformations that are invertible.

The penalty function  $E_J$  based on Jacobian determinant values is defined as following:

$$E_J = \iiint e_J(x, y, z) dx dy dz, \quad (4.13)$$

where,

$$e_J(x, y, z) = \begin{cases} 0 & \text{if } \det J(x, y, z) > J_t \\ (\det J(x, y, z) - J_t)^2 & \text{otherwise,} \end{cases}$$

where  $J_t$  is a threshold. Fig. 4.1 shows the shape of the penalty function  $e_J$  that penalizes Jacobian determinants smaller than 0.2 quadratically.

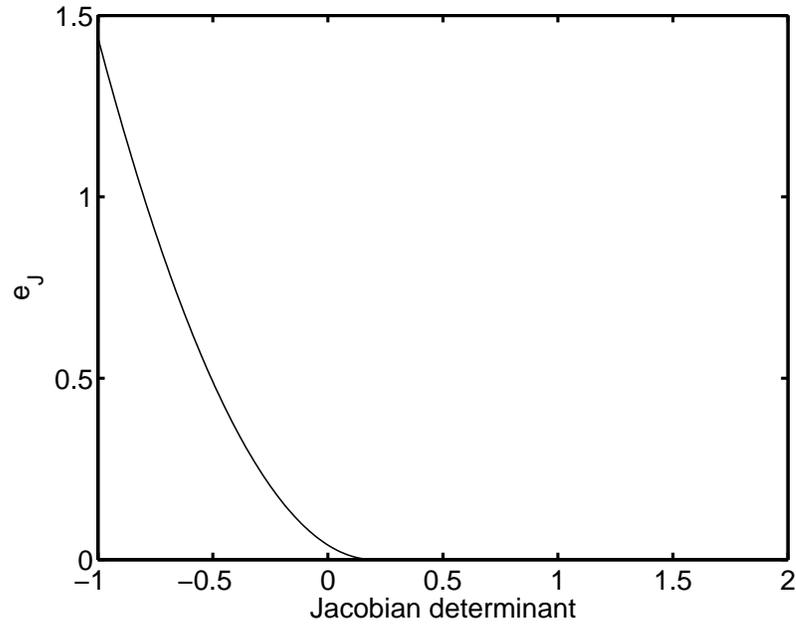


Figure 4.1: Penalty function for Jacobian determinant

#### 4.2.7 Multi-resolution

The proposed method can be incorporated with the multi-resolution strategy to avoid local minima and expedite the optimization. One may use the multi-resolution strategy for both images and geometric transformations. For example, one may first register low resolution images with wide support spline functions and proceed to high resolution images with narrow support spline functions.

## 4.3 Experimental Results

To investigate the performance of the proposed method in comparison with the penalty function based methods, we applied those methods for registering inhale and exhale CT images. Since the inhale and exhale images are from the same patient and the same scanner, we have used MSE (Mean Square Error) as the similarity measure. It has been shown that the MSE based method is the MLE (Maximum-likelihood Estimator) for this problem setting under i.i.d. Gaussian noise assumption [68]. Therefore, one can expect good statistical properties such as the unbiasedness and efficiency.

### 4.3.1 Image Data

Inhale and exhale CT images from 8 patients were scanned using GE CT/i scanner. The image size was  $512 \times 512 \times 60$  with pixel spacing of  $0.9375 \times 0.9375 \times 5$  mm. We have down sampled the images by 2 and 4 for multi-resolution strategy. Fig. 4.2 and Fig. 4.3 show slices of inhale and exhale image from one patient.

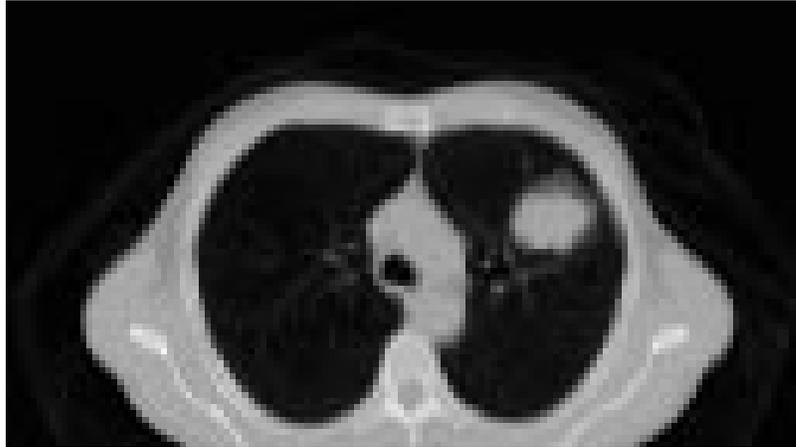


Figure 4.2: A slice of inhale CT image.

### 4.3.2 Synthetic Transformation

Since the *ground truth* deformations of nonrigid registration problems are usually not available, it is difficult to evaluate the performance of image registration methods. To alleviate this difficulty, we conducted registration experiments using deformed images by known synthetic deformations. We generated the synthetic deformations by tensor products

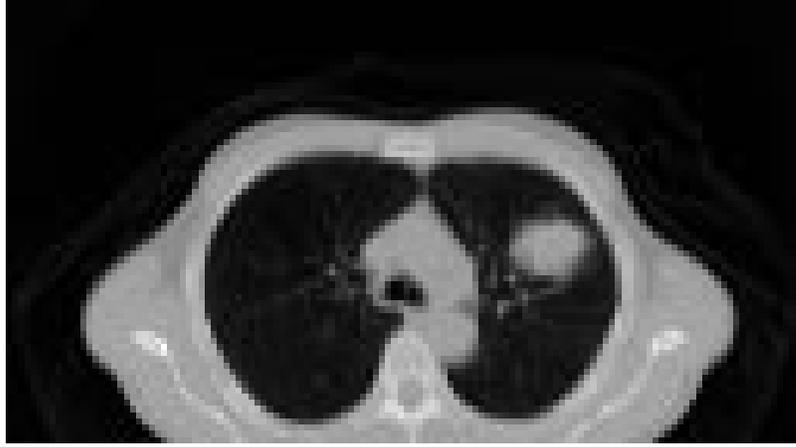


Figure 4.3: A slice of exhale CT image.

of sum of sinusoidal functions in  $x, y, z$  direction. We deformed an inhale CT image from one patient using the synthetic transformation and registered original inhale image onto the deformed image.

For this simulations, we down sampled the CT image by  $4 \times 4 \times 4$  and removed constant backgrounds to reduce computation time. We used a cubic spline based deformation with the expansion parameters 2, 2, 1 for  $f(x, y, z), g(x, y, z), h(x, y, z)$ , respectively.

We investigated the bending energy penalty method, the small Jacobian determinant penalty method and the constrained optimization method with gradient bound of  $\frac{1}{3}$ . We have applied the methods for estimating two synthetic deformations: Synth.1 and Synth.2. We tested two different deformations since we would like to investigate the performance of the proposed method for the cases that the gradient bounds were consistent with true deformation and not consistent with the truth. For Synth.1, every gradient component was smaller than  $\frac{1}{3}$  but for Synth.2, some gradient components were larger than  $\frac{1}{3}$ . Fig. 4.4, Fig. 4.5 and Fig. 4.6 show  $f(x, y, z), g(x, y, z)$  and  $h(x, y, z)$  of Synth.1 evaluated at slice location 3.

We evaluated the performance of each method using an average error index in estimating the true deformation, which is defined as following:

$$\epsilon = \frac{1}{N_x N_y N_z} \sum_{i,j,k} ((\hat{f}(x_i, y_j, z_k) - f(x_i, y_j, z_k))^2 + (\hat{g}(x_i, y_j, z_k) - g(x_i, y_j, z_k))^2 + (\hat{h}(x_i, y_j, z_k) - h(x_i, y_j, z_k))^2)^{\frac{1}{2}}. \quad (4.14)$$

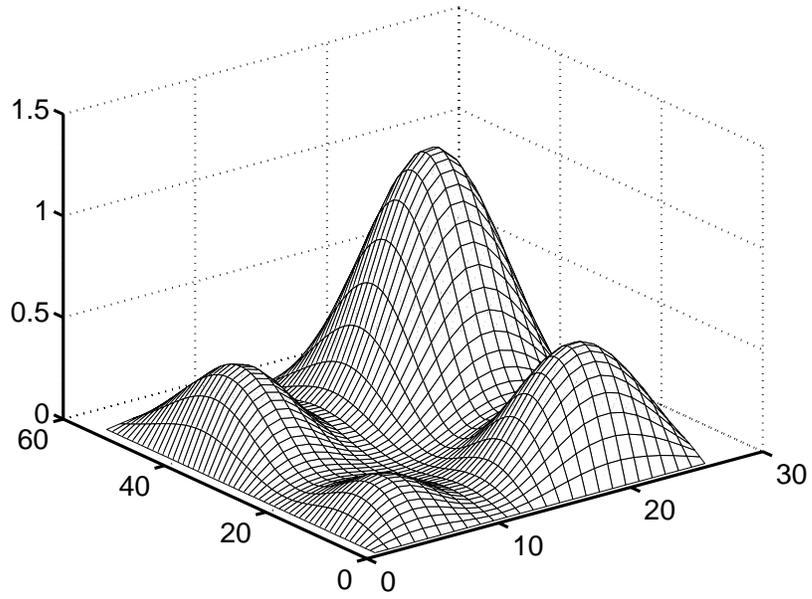


Figure 4.4: Synthetic deformation  $f(x, y, z)$  evaluated at slice 3.

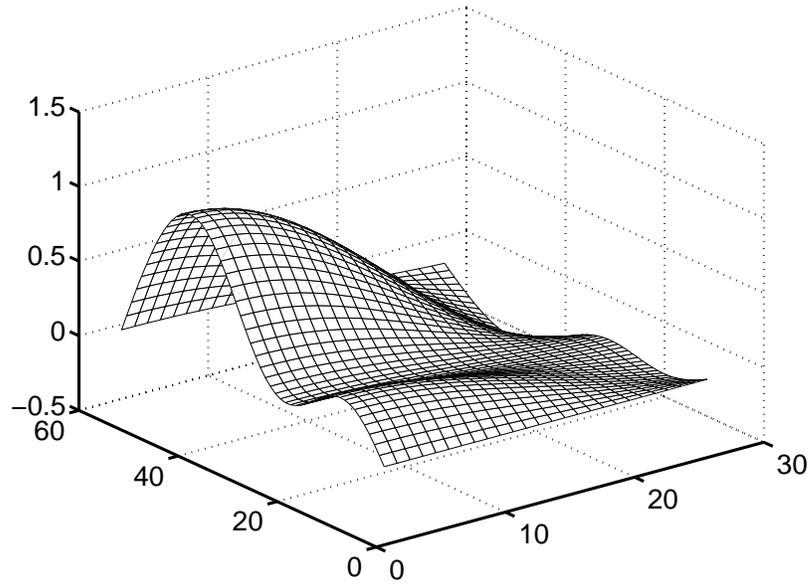


Figure 4.5: Synthetic deformation  $g(x, y, z)$  evaluated at slice 3.

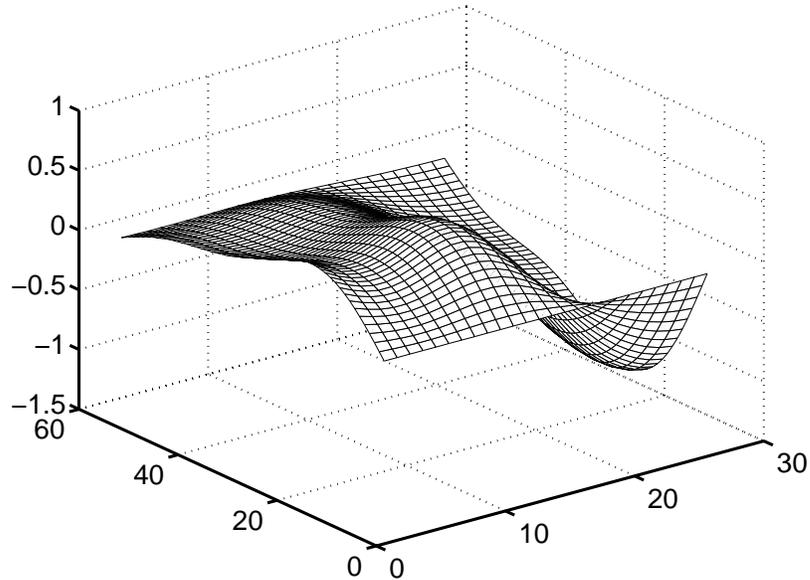


Figure 4.6: Synthetic deformation  $h(x, y, z)$  evaluated at slice 3.

Fig. 4.7, Fig. 4.8 and Fig. 4.9 show the estimated deformations using the constrained optimization method. It was not possible to recover true deformation exactly by image registration due to several reasons such as limited warp space, local minima and insufficient convergence of optimization as discussed in [30]. In addition to those, deformation applied to constant image regions is not recoverable since it does not change image values.

Fig. 4.10 and Fig. 4.11 show changes of the performance index versus optimization iteration for Synth.1 and Synth.2, respectively. In Synth.1, both the penalty based methods and the constrained method performed well. For this simulation, it was not necessary to impose penalty terms to make the deformation invertible since the minimum Jacobian determinant was positive even without using penalty term. The performance of the constrained method was slightly better than the penalty methods.

In the simulation using Synth.2, we do not report the results of bending energy penalty method since it performed very poor. To make estimated deformation invertible using the bending energy penalty, we were required to increase regularization parameter too much. As a result, it yielded very inaccurate estimation.

We expected that the penalty based methods might outperform the constrained method since the true deformation had larger gradients than the gradient bound for the constrained method. On the contrary to the expectation, the constrained method performed better than

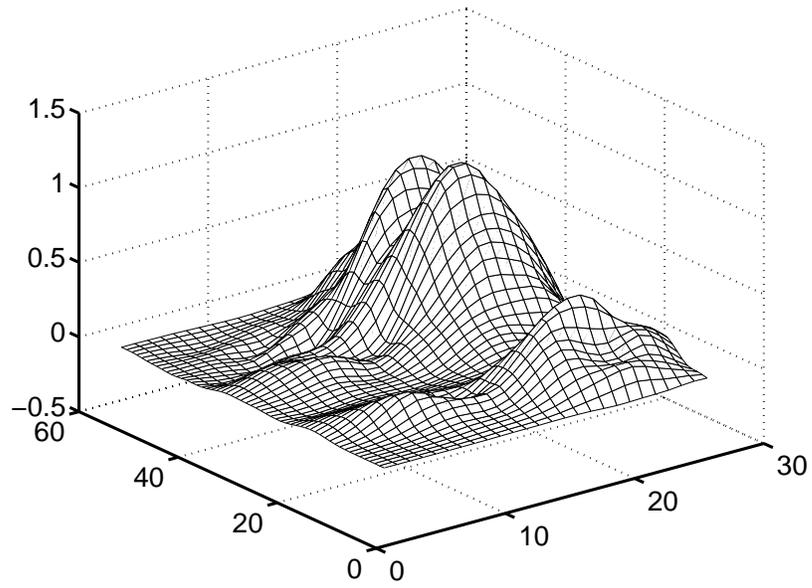


Figure 4.7: Estimated deformation  $\hat{f}(x, y, z)$  evaluated at slice 3

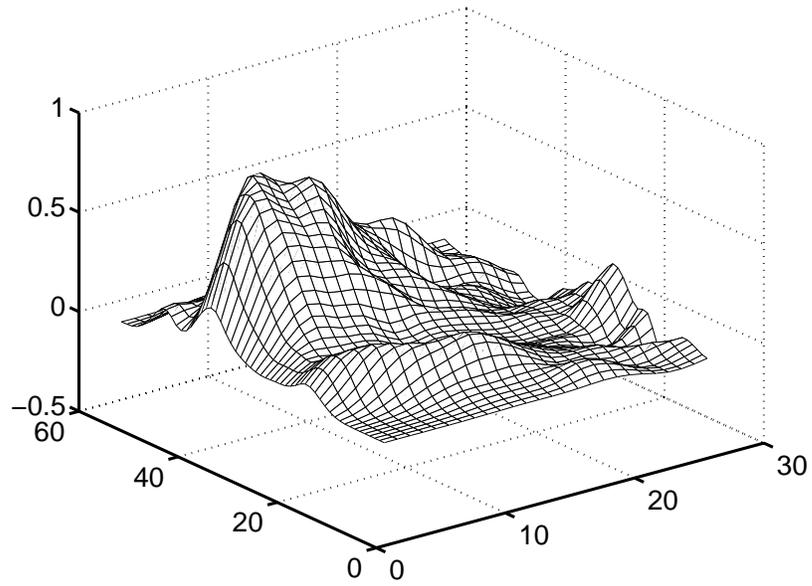


Figure 4.8: Estimated deformation  $\hat{g}(x, y, z)$  evaluated at slice 3.

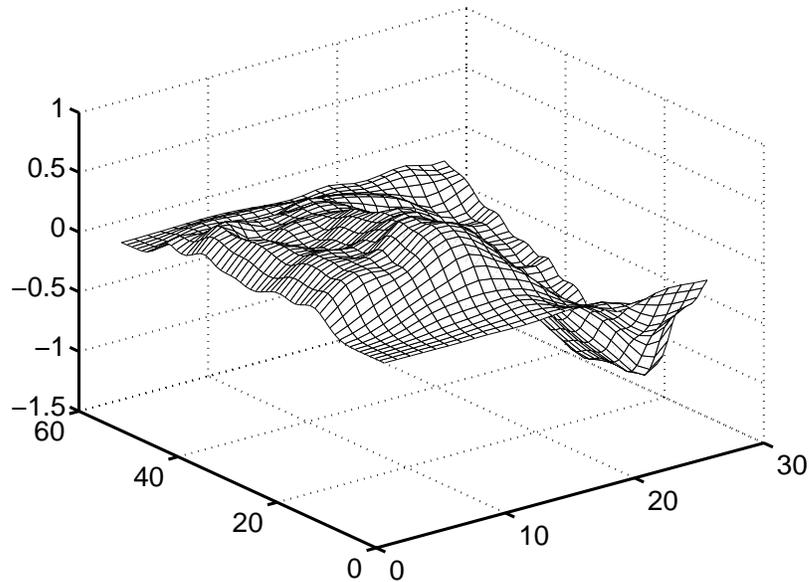


Figure 4.9: Estimated deformation  $\hat{h}(x, y, z)$  evaluated at slice 3.

the small Jacobian determinant penalty method as shown in Fig. 4.11. The small Jacobian determinant penalty method failed to penalize large gradients. As a result, it yielded a rough deformation. This phenomenon is due to the fact that the Jacobian determinant may not be small even if the gradient components are large. We can identify these facts in Table 4.1 that summarizes the minimum/maximum Jacobian determinants and the maximum gradient components from the synthetic deformations, the estimated deformations using the small Jacobian penalty method with threshold value of 0.1 and the estimated deformation using the constrained method. As shown in the table, the small Jacobian penalty method tended to generate larger gradients than true ones.

Fig. 4.12 shows the changes of the minimum Jacobian determinants during the optimization. The minimum Jacobian determinants for the Jacobian penalty methods converged to the values around the Jacobian determinant thresholds. Meanwhile, the minimum Jacobian determinant of the proposed method converged to 0.433, which is larger than the true one. This phenomena is due to the fact that some gradient components of the true deformation are larger than the gradient bound for the constrained optimization.

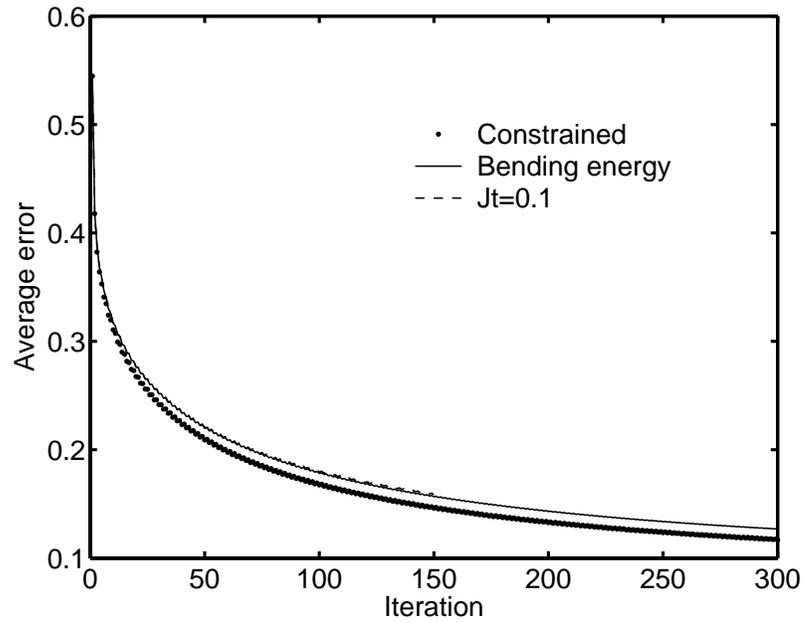


Figure 4.10: Average error vs iteration (Synth.1).

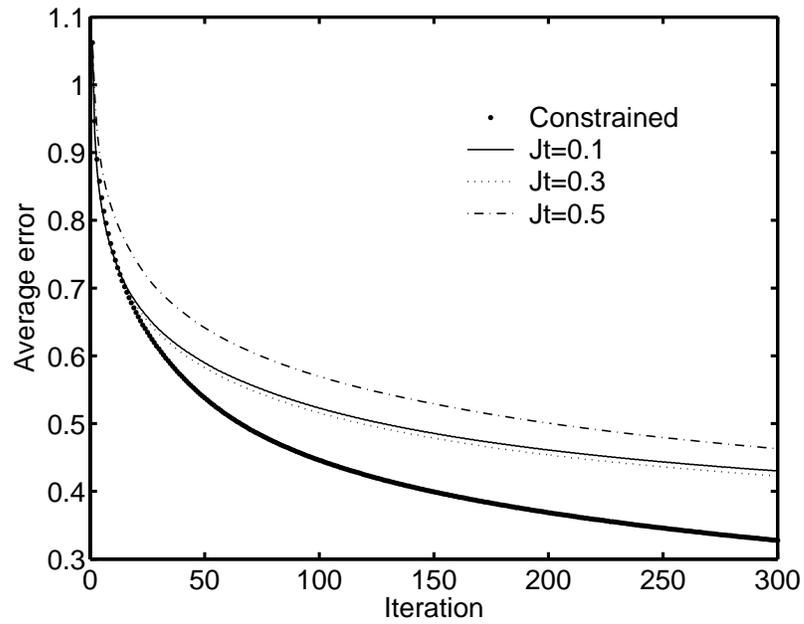


Figure 4.11: Average error vs iteration (Synth.2).

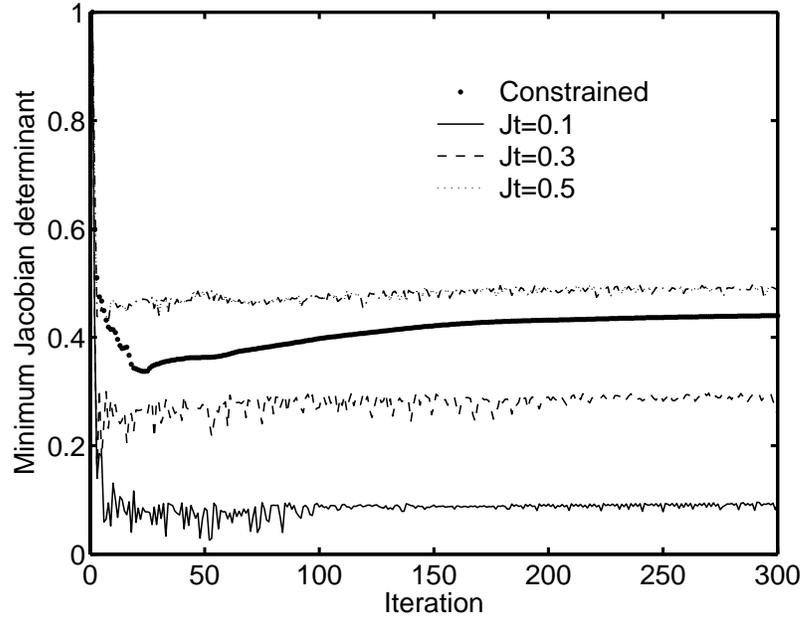


Figure 4.12: Minimum Jacobian determinant vs iteration (Synth.2).

Table 4.1: Characteristics of the synthetic transforms and estimated transforms

	Synth.1	Const. 1	$E_J$ 1	Synth. 2	Const.2	$E_J$ 2
$\min J $	0.807	0.648	0.231	0.201	0.433	0.029
$\max J $	1.324	1.398	1.939	1.457	1.912	3.887
$\max \left  \frac{\partial f(x,y,z)}{\partial x} \right $	0.251	0.196	0.446	0.199	0.303	0.893
$\max \left  \frac{\partial f(x,y,z)}{\partial y} \right $	0.229	0.204	0.264	0.551	0.328	1.547
$\max \left  \frac{\partial f(x,y,z)}{\partial z} \right $	0.269	0.318	0.473	0.319	0.319	2.034
$\max \left  \frac{\partial g(x,y,z)}{\partial x} \right $	0.192	0.174	0.321	0.445	0.331	3.002
$\max \left  \frac{\partial g(x,y,z)}{\partial y} \right $	0.224	0.224	0.389	0.317	0.242	3.616
$\max \left  \frac{\partial g(x,y,z)}{\partial z} \right $	0.329	0.323	0.586	0.609	0.333	2.442
$\max \left  \frac{\partial h(x,y,z)}{\partial x} \right $	0.109	0.163	0.162	0.233	0.241	0.598
$\max \left  \frac{\partial h(x,y,z)}{\partial y} \right $	0.201	0.191	0.212	0.204	0.276	0.889
$\max \left  \frac{\partial h(x,y,z)}{\partial z} \right $	0.287	0.310	0.672	0.786	0.333	0.867

### 4.3.3 Inhale and Exhale Image Registration

We have applied the constrained optimization method for registering inhale/exhale lung CT images from 8 patients. We registered the exhale images onto the inhale images using cubic spline based deformation with expansion factor of  $4 \times 4 \times 2$ . As a result, the support region of a B-spline function is approximately  $3.75 \times 3.75 \times 10$ mm. We used  $84 \times 56 \times 20$  central sub-images after removing constant backgrounds from the images.

Fig. 4.13, Fig. 4.14 and Fig. 4.15 show estimated deformations  $\hat{f}(x, y, z)$ ,  $\hat{g}(x, y, z)$  and  $\hat{h}(x, y, z)$  from one patient data. Table 4.2 summarizes correlation coefficients before and after registration, minimum and maximum Jacobian determinants. As shown in Table 4.2, we were able to achieve high correlation coefficients while bounding each gradient component by  $\frac{1}{3}$ . The resulting Jacobian determinants were in the range between 0.18 to 2.48.

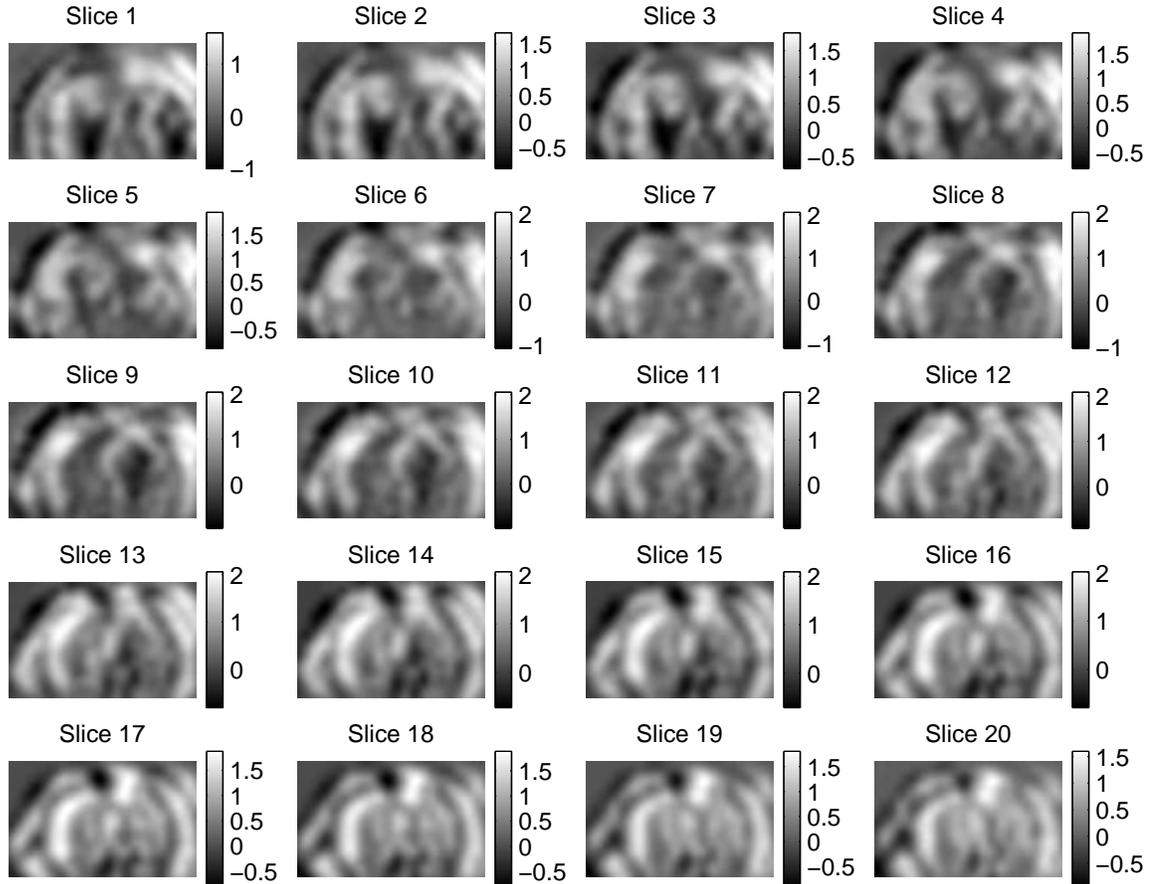


Figure 4.13: Estimated deformation  $\hat{f}(x, y, z)$ .

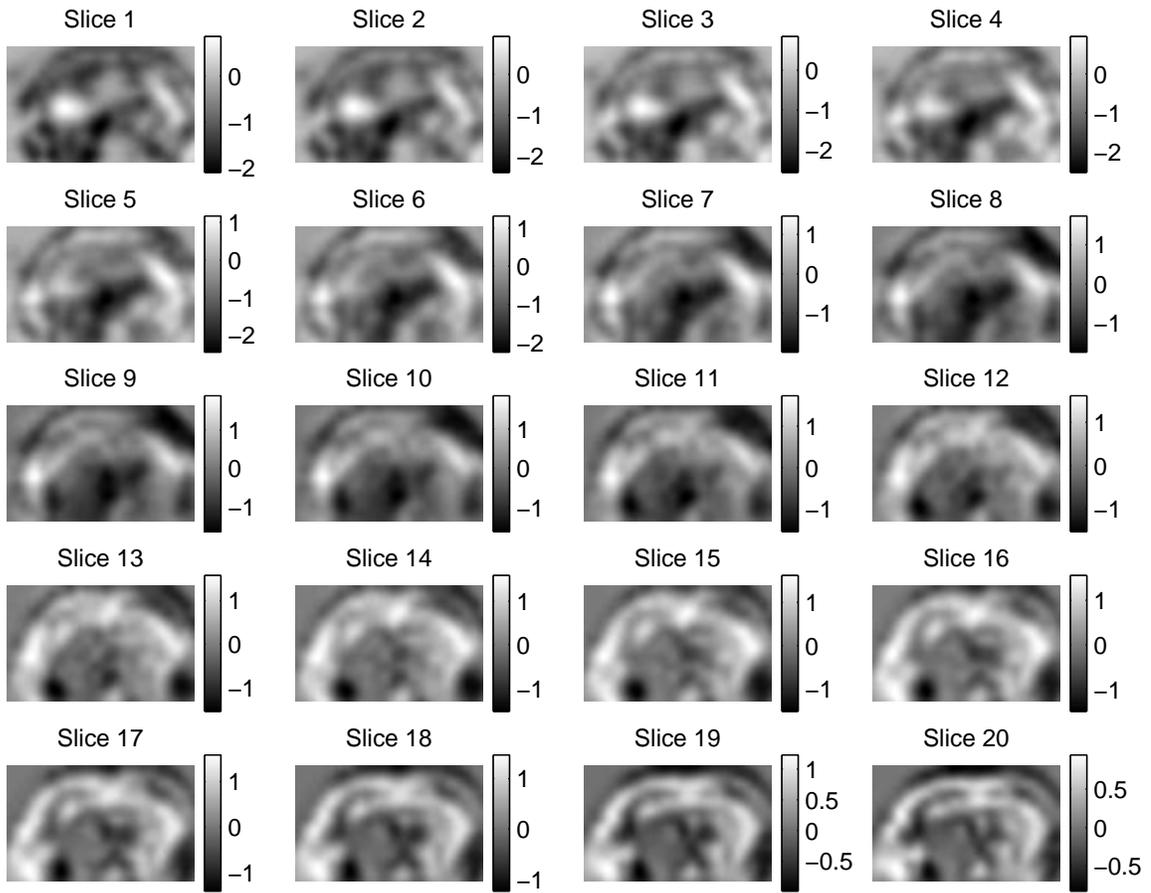


Figure 4.14: Estimated deformation  $\hat{g}(x, y, z)$ .

Table 4.2: Lung CT registration results

( $\rho$  is correlation coefficient between images)

	PT01	PT02	PT03	PT04	PT05	PT06	PT07	PT08
$\rho$ before registration	0.701	0.678	0.852	0.722	0.888	0.755	0.956	0.930
$\rho$ after registration	0.981	0.964	0.978	0.970	0.979	0.935	0.970	0.963
$\min J $	0.332	0.277	0.444	0.295	0.337	0.180	0.428	0.413
$\max J $	2.323	2.477	2.089	2.176	2.269	2.395	2.103	2.023

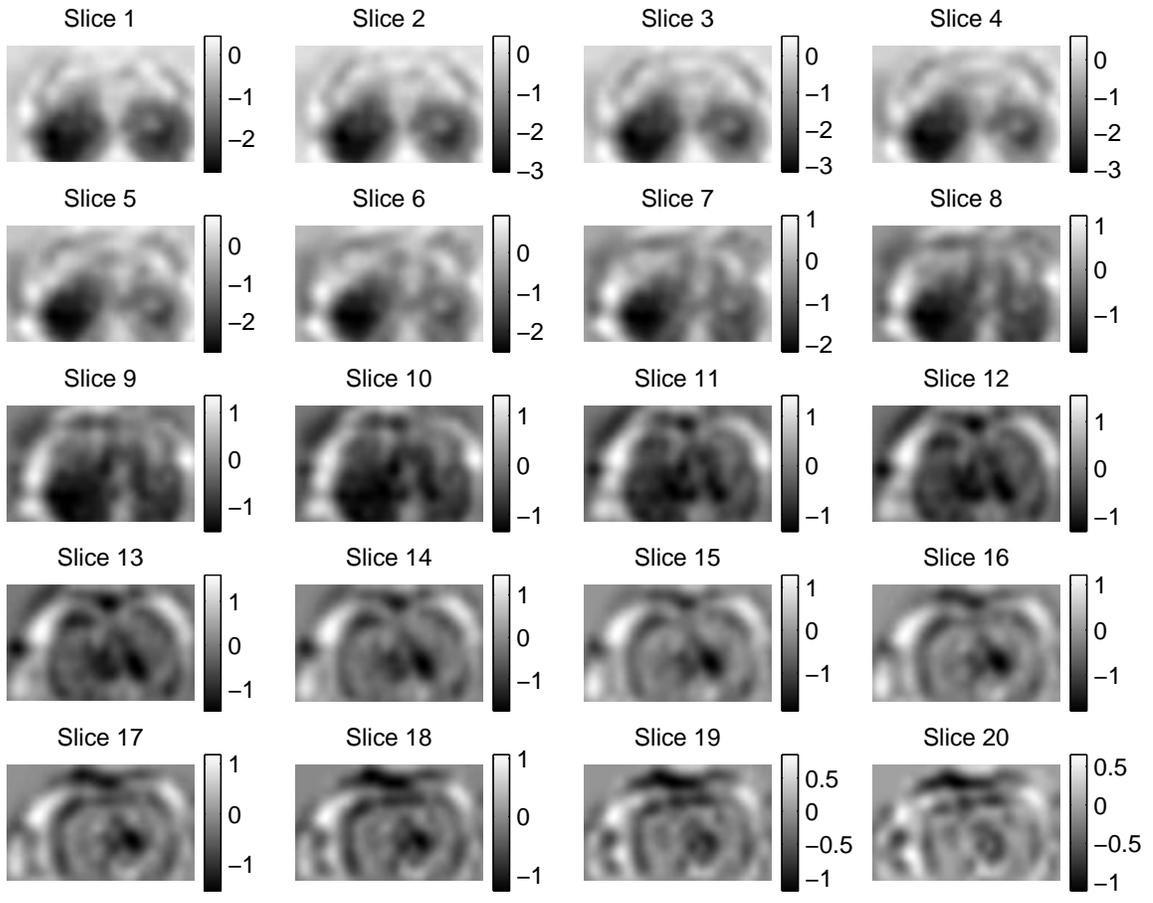


Figure 4.15: Estimated deformation  $\hat{h}(x, y, z)$ .

## 4.4 Discussion

We have bounded gradient components of estimated deformation to enforce positive Jacobian determinants. We also related the gradient bounds to possible minimum and maximum Jacobian determinants (*i.e.*, maximum local volume shrinks and expansions) using Kuhn-Tucker conditions. Compared with conventional methods based on penalty functions, the proposed method has an advantage that the estimated transformation is invertible everywhere as argued in Proposition 2.

In the simulations, the bending energy penalty function method did not yield good results since we were required to smooth deformation excessively to acquire positive Jacobian determinants. We suspect that this phenomenon is due to the fact that small gradients are also penalized as well as large gradients. Since there was a region where Jacobian determinant was negative due to large gradients, the regularization parameter had to be increased to make the deformation invertible. The increased regularization parameter affected the entire deformation and resulted in over-smoothed deformation.

For the penalty based method that penalizes small Jacobian determinants, we penalized only smaller Jacobian determinants than a given threshold since we did not want to penalize deformations that are invertible. This method often yielded deformations that have large gradients and large local volume expansions. The large local volume expansions were due to the fact that those were not penalized. Although a transformation that has large Jacobian determinant is invertible, it may not be a good estimation of true deformation, if we assume that the true deformation does not have too large volume expansion. Therefore, one may want to penalize both small and large Jacobian determinants to acquire more realistic transformation. Even if both small and large Jacobian determinants are avoided by proper penalty, we think that it may not be sufficient enough to exclude the possibilities of unrealistic estimations for some cases, at least theoretically. For example, the gradient component  $k$  in the following  $2 \times 2$  Jacobian matrix can be very large while its determinant is unity,

$$J = \begin{bmatrix} 1+k & k \\ -k & 1-k \end{bmatrix}. \tag{4.15}$$

We conjecture that this phenomenon partly explains why the estimated deformations using this method had large gradients in the experiments.

In addition to those problems, there is a difficulty in optimization if one uses Jacobian determinant in penalty function since it is not a convex function of parameters. Moreover, the constraint set to bound Jacobian determinants in the parameter space is not a convex set either. Therefore, it may be challenging to design an effective optimization method.

We believe that bounding each gradient component can be a useful method to acquire a realistic deformation since it yields both bounded gradients and bounded Jacobian determinants. Moreover, since a parameter set that bounds gradients is a convex set, it is possible to design an effective optimization method.

One may implement the gradient bounds using either a constrained optimization that maximizes similarity measure subject to the constraints or a unconstrained optimization method that has large penalty values outside the bounds. Compared with the unconstrained optimization methods, the constrained optimization has an advantage since tuning of regularization parameter is not required. In our experiences, tuning of regularization parameters was cumbersome since it often needed to be readjusted if images were changed.

Thanks to the convexity of the constraint set, we were able to use the gradient projection method with Dykstra’s algorithm. This method has an advantage that *line search* is not required. However, it requires additional computation time since orthogonal projection onto the constraint set must be computed at each iteration.

It is our future work to investigate both unconstrained and constrained optimization methods to bound gradients in detail. We are interested in not only improving optimization speeds and accuracies but also removing manual tuning procedures. Note that the step size  $\alpha$  for the gradient projection method was adjusted manually. We plan to investigate several methods for that purpose including optimization transfer method [69].

One can use Proposition 1 and Proposition 2 to relate possible maximum volume changes to the constraints sets in the parameter space. Note that the Propositions does not define minimum and maximum Jacobian determinants but *possible* minimum and maximum. In Proposition 1, the possible minimum or maximum Jacobian determinant occurs only when every gradient component has the boundary value. In our simulations and experiments, such phenomenon was not observed. Even though the possible minimum Jacobian was zero in the experiments, the estimated minimum Jacobian determinants were around 0.3.

Therefore, it may be possible to relax the gradient bounds while achieving positive Jacobian determinants.

In inhale and exhale image registration experiments, we were able to achieve high correlation coefficients between inhale images and deformed exhale images after registration using the gradient bounds of  $1/3$ . Although it is difficult to argue the accuracies of the estimated deformations based on only the similarity measure values since the ground truths are unknown, we conjecture that true physical deformations do not have very large gradients. More investigations are required to support this conjecture.

It would be very useful if *a priori* informations about physical deformations are available. For example, if population based distributions of minimum and maximum volume changes or minimum and maximum gradient components of deformations are available, those can be effectively applied to improve nonrigid registration methods. Analytical results and optimization methods investigated in this study can be used to incorporate such statistical data for more accurate estimation of the deformations.

## 4.5 Conclusion

We have proposed a novel constrained optimization method that bounds each gradient component of estimated deformation to achieve positive Jacobian determinants. We related the gradient bounds to possible maximum volume shrinks and expansions and derived sufficient conditions in the parameter space to achieve the gradient bounds in the entire spatial domain. The gradient bounds were implemented using the gradient projection method with Dykstra's cyclic projection algorithm. In the experiments of registering inhale and exhale CT images, we were able to achieve high similarity measures after registration while bounding gradient components of the estimated deformation.

## CHAPTER 5

### Conclusion and Future Research

#### 5.1 Conclusion

We have investigated a set-up error estimation method using a 3D/2D, intensity-based method. To achieve 3D to 2D image registration, the radiographs and the DRRs of the planning CT were registered by maximizing the MI between DRRs and radiographs. In the experiment with an anthropomorphic chest phantom, we achieved accuracies better than 1mm for estimating the translational parameters and 0.8 degree for estimating the rotational parameters using two orthogonal pairs of the MV radiographs and DRRs. The true set-up error was established by the fiducial marker-based method. Based upon theoretical background and the experimental results, we believe that MI has significant potential as an effective similarity measure for 3D/2D intensity-based registration.

We also introduced robust correlation coefficient as a novel intensity based similarity measure to improve the robustness of the sample correlation coefficient based image registration while not degrading the statistical efficiency much. In 1D simulation, 2D and 3D experiments, the proposed method reduced the bias of the sample correlation method caused by outliers. The robust correlation coefficient may be an effective similarity measure for intra-modality image registration task where the presence of the outliers is unavoidable such as set-up estimation for radiotherapy and image-guided surgery.

For nonrigid image registration, we have proposed a novel constrained optimization method that bounds each gradient component to enforce positive Jacobian determinants. We related the gradient bounds to possible minimum and maximum Jacobian determinants and derived sufficient conditions in parameter space to bound gradients everywhere. We

also implemented the gradient bounds using the gradient projection method with Dykstra’s cyclic projection algorithm. In experiments of registering inhale and exhale CT images, we were able to achieve high similarity measure while bounding each gradient component. Using the registration results, one can build a preliminary *dynamic* model for lung motion.

## 5.2 Future Research

Although we have addressed some technical issues in image registration methods for radiation therapy and image-guided surgery, there are still many technical issues to be investigated. For example, 3D/2D nonrigid registration of planning CT onto projection images is important to estimate organ motion of a patient in each treatment session. Accurate estimation of the organ motion in each treatment session may have a significant potential for improving radiation therapy since X-ray dose calculation can be accomplished more accurately by accounting for the organ motion.

Other technical issues such as effective modeling of geometric deformation and fast optimization methods also need to be investigated to accomplish more accurate and fast registration. Those research results can be applied for not only radiation therapy but also other fields such as atlas based segmentation using nonrigid registration of a template onto images.

In addition to image registration, research on 3D CT reconstruction methods using limited number of projections can be useful for radiation therapy. In consecutive radiotherapy treatments, the anatomy of a patient may change from that of planning. It is natural because not only the anatomy of the patient changes in time but also the treatments affect the patient body. To monitor this change, frequent CT scan is required. However, since conventional 3D CT scan requires extra x-ray irradiation, frequent CT scan is not desirable. Instead, it would be beneficial if 3D CT scan could be reconstructed using a few numbers of radiographs that are acquired in the treatment room.

Although this problem is challenging since the image reconstruction problem is ill-posed, it might be possible to reconstruct useful CT images by designing an effective regularization method. Compared with conventional regularization methods that are not based on individual patient model but based on generic *a priori* information such as smoothness, a method based on *a priori* information of the same patient such as the CT image taken before

can be effectively applied for addressing the problem. For example, one may reconstruct 3D image by maximizing a likelihood term and a similarity measure between reconstructing image and *a priori* image if two images are already registered. Moreover, one may combine reconstruction and registration procedures together to implement joint reconstruction and registration. Since the previous patient CT may have better information about the patient anatomy than generic smoothness assumption, we conjecture that good reconstruction may be achieved using significantly small number of projections. Moreover, one may challenge to the problem of image reconstruction using population based CT image as *a priori* information if the previous CT of the same patient is not available.

Dynamic 3D imaging is also a valuable research topic since radiation therapy often deals with moving organs such as lungs. Since most CT reconstruction methods assume that anatomy does not change during acquisitions, the dynamic image reconstruction requires different reconstruction methods than normal CT reconstruction algorithms. Several interesting methods for reconstructing dynamic images have been proposed. For example, a Kalman filter type approach can be used to estimate the moving 3D images if the motion can be modeled by dynamic equations. Also image registration techniques may be applied for dynamic reconstruction since motion estimation and compensation can be accomplished by image registration. Moreover, registering reconstructing dynamic CT image to dynamic ultrasound image may be useful since ultrasound image has already dynamic motion information. Such a 4D image registration method can be a valuable future research topic.

## **APPENDICES**

## APPENDIX A

### Fisher Information

Consider a set of the independent Poisson Random Variables which are defined by (A.1)

$$Y_k \sim \text{Poisson}\{\lambda_k(\tilde{\theta})\}, \quad \lambda_k(\tilde{\theta}) = I_0 e^{-\int_{L_k} \mu_2(T_{\tilde{\theta}}(\tilde{t})) d\tilde{t}}, \quad k = 1, \dots, N. \quad (\text{A.1})$$

Because the random variables are independent, the logarithm of the joint PMF of the random variables  $Y_k$  is computed by (A.2):

$$\log P[Y_1 = y_1, Y_2 = y_2, \dots, Y_k = y_k | \tilde{\theta}] = \sum_{k=1}^N y_k \log \lambda_k(\tilde{\theta}) - \lambda_k(\tilde{\theta}) - \log y_k!. \quad (\text{A.2})$$

The partial derivative of the log joint PMF with respect to parameter  $\theta_i$  is calculated by (A.3):

$$\frac{\partial \log P[Y = y | \tilde{\theta}]}{\partial \tilde{\theta}_i} = \sum_{k=1}^N y_k \frac{1}{\lambda_k(\tilde{\theta})} \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_i} - \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_i}. \quad (\text{A.3})$$

If we take the derivative of (A.3) with respect to parameter  $\theta_j$ , (A.4) is acquired,

$$\frac{\partial^2 \log P[Y = y | \tilde{\theta}]}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j} = \sum_{k=1}^N y_k \left\{ -\frac{1}{\lambda_k(\tilde{\theta})^2} \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_i} \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_j} + \frac{1}{\lambda_k(\tilde{\theta})} \frac{\partial^2 \lambda_k(\tilde{\theta}_i)}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j} \right\} - \frac{\partial^2 \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j}. \quad (\text{A.4})$$

Finally, the  $ij_{th}$  component of the Fisher information matrix  $J$  for is computed by (A.5):

$$J_{ij} = -E \left\{ \frac{\partial^2 \log P[Y = y | \tilde{\theta}]}{\partial \tilde{\theta}_i \partial \tilde{\theta}_j} \right\} = \sum_{k=1}^N \frac{1}{\lambda_k(\tilde{\theta})} \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_i} \frac{\partial \lambda_k(\tilde{\theta})}{\partial \tilde{\theta}_j}. \quad (\text{A.5})$$

## APPENDIX B

### Mean and Covariance Approximation

This appendix presents approximations for the mean and variance of implicitly defined estimators such as (3.1). Let  $\Phi(\theta, Z)$  be an objective function depending on unknown parameter  $\theta$  and noisy measurement  $Z$ . We assume that the objective function has a unique maximum for any  $Z$  and that one can find the maximum by zeroing the partial derivatives of  $\Phi(\cdot, Z)$  to determine the estimates  $\hat{\theta}$ :

$$0 = \frac{\partial}{\partial \theta_j} \Phi(\theta, Z) \Big|_{\theta=\hat{\theta}}, \quad j = 1, \dots, p. \quad (\text{B.1})$$

Since  $\hat{\theta}$  is an implicit function of  $Z$ , it is difficult to determine its exact mean and variance. To approximate its mean and variance, we linearize the estimator by first order Taylor series expansion. First, we define the column gradient of the objective function as follows:

$$\Psi(\theta, Z) \triangleq \nabla^{10} \Phi(\theta, Z), \quad (\text{B.2})$$

where, the  $j$ th element of  $p \times 1$  operator  $\nabla^{10}$  is  $\frac{\partial}{\partial \theta_j}$ .

Next, we rewrite (B.1) as  $\Psi(\hat{\theta}, Z) = 0$ . If we linearize  $\Psi(\hat{\theta}, Z)$  around the *true* parameter  $\tilde{\theta}$  by the first-order Taylor series expansion, the estimator can be approximated as following:

$$\Psi(\hat{\theta}, Z) \approx \Psi(\tilde{\theta}, Z) + \nabla^{20} \Phi(\tilde{\theta}, Z)(\hat{\theta} - \tilde{\theta}), \quad (\text{B.3})$$

where, the  $(j, k)$ th element of  $p \times p$  operator  $\nabla^{20}$  is  $\frac{\partial^2}{\partial \theta_j \partial \theta_k}$ .

We assume that  $[-\nabla^{20} \Phi(\tilde{\theta}, Z)]$  is a positive definite symmetric matrix so that its inverse is well defined. As a result, one may linearize the estimator as follows:

$$\hat{\theta} \approx \tilde{\theta} + [-\nabla^{20} \Phi(\tilde{\theta}, Z)]^{-1} \Psi(\tilde{\theta}, Z). \quad (\text{B.4})$$

Therefore, the bias of the estimator may be approximated as following:

$$E\{\hat{\theta}\} - \tilde{\theta} \approx E \left\{ [-\nabla^{20}\Phi(\tilde{\theta}, Z)]^{-1} \Psi(\tilde{\theta}, Z) \right\}. \quad (\text{B.5})$$

Moreover, the covariance of the estimator may be approximated as follows:

$$\begin{aligned} \text{Cov}\{\hat{\theta}\} &\approx \text{Cov}\left\{[-\nabla^{20}\Phi(\tilde{\theta}, Z)]^{-1}\Psi(\tilde{\theta}, Z)\right\} \\ &\approx H^{-1} \text{Cov}\left\{\Psi(\tilde{\theta}, Z)\right\} H^{-1}, \end{aligned} \quad (\text{B.6})$$

where possible approximations for  $H$  matrix include:

$$H = -\nabla^{20}\Phi(\tilde{\theta}, E[Z]). \quad (\text{B.7})$$

One may approximate further by linearizing the  $\Psi(\theta, Z)$  with respect to the measurements  $Z$  as developed previously [26].

## APPENDIX C

### Bias and Robustness of Correlation based Methods

This appendix uses the general results from the preceding appendix to analyze the approximated bias and robustness of correlation-based registration methods. Let us define our noisy measurement  $Z = [s_1(\cdot), s_2(\cdot)]$ . We represent  $s_1(T_\theta(\vec{t})) = s_1(\theta, \vec{t})$  for notational convenience and  $\mathbf{z}_\theta(\vec{t}) = [s_1(\theta, \vec{t}), s_2(\vec{t})]^T$ . Without loss of generality, we assume that the means of  $s_1(\vec{t}), s_2(\vec{t})$  are zero and the volume of the images integrates to unity. We consider the robust correlation method with fixed mean for mathematical simplicity.

First, the objective function is represented as follows:

$$\Phi(\theta, Z) = \frac{C_{XY}(\theta)}{\sigma_X(\theta)\sigma_Y(\theta)}, \quad (\text{C.1})$$

where  $C_{XY}(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|) s_1(\theta, \vec{t}) s_2(\vec{t}) d\vec{t}$ ,  $\sigma_X^2(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|) s_1^2(\theta, \vec{t}) d\vec{t}$  and  $\sigma_Y^2(\theta) = \int w(\|V_\theta \mathbf{z}_\theta(\vec{t})\|) s_2^2(\vec{t}) d\vec{t}$ .

The gradient of the objective function is evaluated as followings:

$$\Psi(\theta, Z) = \frac{1}{\sigma_X^2(\theta)\sigma_Y^2(\theta)} [C'_{XY}(\theta)\sigma_X(\theta)\sigma_Y(\theta) - C_{XY}(\theta)(\sigma_X(\theta)\sigma'_Y(\theta) + \sigma'_X(\theta)\sigma_Y(\theta))], \quad (\text{C.2})$$

where  $C'_{XY}(\theta) = \nabla_\theta C_{XY}(\theta)$ ,  $\sigma'_X(\theta) = \nabla_\theta \sigma_X(\theta)$  and  $\sigma'_Y(\theta) = \nabla_\theta \sigma_Y(\theta)$ .

If  $s_1(\tilde{\theta}, \vec{t}) = k s_2(\vec{t})$  with a constant  $k$ , (*i.e.* two images are linearly related at true registered position), then  $\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) = C_{XY}(\tilde{\theta})$  and  $C'_{XY}(\tilde{\theta}) = \sigma'_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) + \sigma_X(\tilde{\theta})\sigma'_Y(\tilde{\theta})$ . Therefore,  $\Psi(\tilde{\theta}, Z) = 0$ , which implies the unbiasedness. Note that this is true not only for the constant weighting function but also for the arbitrary  $w(\cdot)$ .

Suppose that there are additive outlier images  $o(\vec{t})$  such that  $\tilde{s}_2(\vec{t}) = s_1(\tilde{\theta}, \vec{t}) + o(\vec{t})$ . For the sample correlation method, the Hessian at the true registered position may be approximated as following:

$$\begin{aligned} \nabla^{20}\Phi(\tilde{\theta}, Z) = & \frac{1}{\sigma_X^2(\tilde{\theta})\sigma_Y^2(\tilde{\theta})} [C''_{XY}(\tilde{\theta})\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) - C_{XY}(\tilde{\theta}) \\ & \cdot (\sigma_X''(\tilde{\theta})\sigma_X(\tilde{\theta}) + 2\sigma_X'(\tilde{\theta})\sigma_Y'(\tilde{\theta}) + \sigma_X(\tilde{\theta})\sigma_Y''(\tilde{\theta}))], \end{aligned} \quad (\text{C.3})$$

where,  $C''_{XY}(\theta) = \nabla_{\theta}^2 C_{XY}(\theta)$ ,  $\sigma_X''(\theta) = \nabla_{\theta}^2 \sigma_X(\theta)$  and  $\sigma_Y''(\theta) = \nabla_{\theta}^2 \sigma_Y(\theta)$ .

If we assume that the images have constant backgrounds,  $\sigma_X'(\theta)$ ,  $\sigma_Y'(\theta)$ ,  $\sigma_X''(\theta)$  and  $\sigma_Y''(\theta)$  are all zero for a constant weighting. In that case, the bias is approximated as follows:

$$\begin{aligned} E \{ \hat{\theta} \} \approx & \tilde{\theta} + [C'_{XY}(\tilde{\theta})][C''_{XY}(\tilde{\theta})]^{-1} \\ & = \tilde{\theta} + \left[ \int [\nabla^{10}s_1(\tilde{\theta}, \vec{t})]o(\vec{t}) d\vec{t} \right] \left[ \int [\nabla^{10}s_1(\tilde{\theta}, \vec{t})][\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T - \nabla^{20}s_1(\tilde{\theta}, \vec{t})o(\vec{t})d\vec{t} \right]^{-1}. \end{aligned} \quad (\text{C.4})$$

As expected, if weighting is a constant such as in the sample correlation method, in general, the estimator is biased by outliers since  $\int [\nabla^{10}s_1(\tilde{\theta}, \vec{t})]o(\vec{t}) d\vec{t}$  does not equal to zero for many cases. For the robust correlation method, the bias can be significantly by proper weighting. If weighting for outlier samples is very small,  $\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta}) \approx C_{XY}(\tilde{\theta})$  and  $C'_{XY}(\tilde{\theta}) \approx \sigma_X'(\tilde{\theta})\sigma_Y(\tilde{\theta}) + \sigma_X(\tilde{\theta})\sigma_Y'(\tilde{\theta})$  and the robust correlation coefficient at  $\tilde{\theta}$  becomes unity, which implies the robust correlation coefficient is maximized at true position.

## APPENDIX D

### Efficiency

It is well known that the sample correlation coefficient based method is MLE for this problem setting [70] with constant background and i.i.d. Gaussian noise assumption. Therefore the covariance of the sample correlation based estimator equals to the inverse of the Fisher information matrix [34]. The Fisher information matrix  $I_{\tilde{\theta}}$  is computed as follows:

$$\begin{aligned} I_{\tilde{\theta}} &= \frac{1}{2\sigma_n^2} E \left\{ -\nabla_{\tilde{\theta}}^2 \int (s_2(\vec{t}) - s_1(\tilde{\theta}, \vec{t}))^2 d\vec{t} \right\} \\ &= \frac{1}{\sigma_n^2} \int [\nabla^{10} s_1(\tilde{\theta}, \vec{t})][\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}, \end{aligned} \quad (\text{D.1})$$

where  $\sigma_n^2$  is noise power.

We approximate the covariance matrix of the correlation based methods by the covariance approximation developed in Appendix A. First, we approximate the covariance matrix of  $\Psi(\tilde{\theta}, Z)$  as follows:

$$\begin{aligned} \text{Cov} \left\{ \Psi(\tilde{\theta}, Z) \right\} &\approx \frac{1}{\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})} \text{Cov} \left\{ C'_{XY}(\tilde{\theta}) - \left( \frac{\sigma'_X(\tilde{\theta})}{\sigma_X(\tilde{\theta})} + \frac{\sigma'_Y(\tilde{\theta})}{\sigma_Y(\tilde{\theta})} \right) \right\} \\ &\approx \frac{1}{\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})} E \left\{ \int w(\|V_{\tilde{\theta}} \mathbf{z}_{\tilde{\theta}}(\vec{t})\|) n(\vec{t}) [\nabla^{10} s_1(\tilde{\theta}, \vec{t})] d\vec{t} \right. \\ &\quad \left. \cdot \int w(\|V_{\tilde{\theta}} \mathbf{z}_{\tilde{\theta}}(\vec{\tau})\|) n(\vec{\tau}) [\nabla^{10} s_1(\tilde{\theta}, \vec{\tau})]^T d\vec{\tau} \right\} \\ &= \frac{\sigma_n^2}{\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})} \int w(\|V_{\tilde{\theta}} \mathbf{z}_{\tilde{\theta}}(\vec{t})\|)^2 [\nabla^{10} s_1(\tilde{\theta}, \vec{t})][\nabla^{10} s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}. \end{aligned} \quad (\text{D.2})$$

We also approximate the Hessian as follows:

$$\begin{aligned} H &= -\nabla^{20}\Phi(\tilde{\theta}, E[Z]) \\ &= \frac{\sigma_n^2}{\sigma_X(\tilde{\theta})\sigma_Y(\tilde{\theta})} \int w(\|V_{\tilde{\theta}}z_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10}s_1(\tilde{\theta}, \vec{t})][\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}. \end{aligned} \quad (\text{D.3})$$

Finally, the covariance may be approximated by plugging (D.2), (D.3) into (B.6) as follows:

$$\text{Cov}\{\hat{\theta}\} \approx \sigma_n^2 C_1^{-1} C_2 C_1^{-1}, \quad (\text{D.4})$$

where, the matrix  $C_1 = \int w(\|V_{\tilde{\theta}}z_{\tilde{\theta}}(\vec{t})\|) [\nabla^{10}s_1(\tilde{\theta}, \vec{t})][\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}$  and the matrix  $C_2 = \int w(\|V_{\tilde{\theta}}z_{\tilde{\theta}}(\vec{t})\|)^2 [\nabla^{10}s_1(\tilde{\theta}, \vec{t})][\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T d\vec{t}$ .

By vector Cauchy-Schwarz inequality [53], one can prove following inequality:

$$\text{Cov}\{\hat{\theta}\} \geq I_{\tilde{\theta}}^{-1}, \quad (\text{D.5})$$

where,  $A \geq B$  means the positive semi-definiteness of matrix  $A - B$  and the equality holds iff  $w(\cdot)[\nabla^{10}s_1(\tilde{\theta}, \vec{t})] = k[\nabla^{10}s_1(\tilde{\theta}, \vec{t})]$  with a constant  $k$ .

Therefore any weighting function other than a constant will make the covariance larger while the constant weighting such as in the ordinary sample correlation coefficient method achieves the *Cramer-Rao* lower bound. However, note that we have assumed  $s_2(\vec{t}) = s_1(\tilde{\theta}, \vec{t}) + n(\vec{t})$  that is not a reasonable assumption in the presence of outliers. If there exist outliers in a certain location, the noise variance is not uniform. In that case, the  $C_1$  matrix in (D.4) becomes as following:

$$C_1 = \int w(\|V_{\tilde{\theta}}z_{\tilde{\theta}}(\vec{t})\|) \left\{ [\nabla^{10}s_1(\tilde{\theta}, \vec{t})][\nabla^{10}s_1(\tilde{\theta}, \vec{t})]^T - o(\vec{t})[\nabla^{20}s_1(\tilde{\theta}, \vec{t})] \right\} d\vec{t}. \quad (\text{D.6})$$

Note that if  $\int w(\|V_{\tilde{\theta}}z_{\tilde{\theta}}(\vec{t})\|) o(\vec{t})[\nabla^{20}s_1(\tilde{\theta}, \vec{t})] d\vec{t} \approx 0$ , the covariance approximation formula in the presence of outlier is the same as without outlier. In that case, one may argue that the variance of the sample correlation method is smaller than the robust correlation method, based on (D.5).

## APPENDIX E

### Gradient of Estimated joint pdf

We consider 2-D rigid image registration for simplicity. Extension to 3D is straightforward. Consider asymptotic estimated joint pdf represented as following:

$$\hat{f}_\theta(x, y) = \frac{1}{T_1 T_2} \int_{t_{1o}}^{t_{1f}} \int_{t_{2o}}^{t_{2f}} K(x - s_1(T_1(t_1, t_2, \phi, d_1), T_2(t_1, t_2, \phi, d_2))) K(y - s_2(t_1, t_2)) dt_1 dt_2, \quad (\text{E.1})$$

where,  $s_1(\cdot)$ ,  $s_2(\cdot)$  are two images,  $K(\cdot)$  is a kernel function,  $t_{1o}$ ,  $t_{1f}$  is the starting and the final point of the continuous images in horizontal direction,  $t_{2o}$ ,  $t_{2f}$  are those in vertical direction,  $T_1 = t_{1f} - t_{1o}$  and  $T_2 = t_{2f} - t_{2o}$ .

Taking partial derivative of the estimated joint pdf w.r.t.  $\phi$  and evaluating at *true* registered position yields following:

$$\begin{aligned} \frac{\partial f_\theta(x, y)}{\partial \phi} \Big|_{\theta=\bar{\theta}} &= \frac{1}{T_1 T_2} \int_{t_{1o}}^{t_{1f}} \int_{t_{2o}}^{t_{2f}} \dot{K}(x - s_1(t_1, t_2)) [D_1 s_1(t_1, t_2) \frac{\partial T_1(t_1, t_2, \phi, d_1)}{\partial \phi} \Big|_{\theta=\bar{\theta}} \\ &\quad + D_2 s_1(t_1, t_2) \frac{\partial T_2(t_1, t_2, \phi, d_2)}{\partial \phi} \Big|_{\theta=\bar{\theta}}] \cdot K(y - s_2(t_1, t_2)) dt_1 dt_2 \\ &= \frac{1}{T_1 T_2} \int_{t_{1o}}^{t_{1f}} \int_{t_{2o}}^{t_{2f}} \dot{K}(x - s_1(t_1, t_2)) [-t_2 D_1 s_1(t_1, t_2) + t_1 D_2 s_1(t_1, t_2)] \\ &\quad \cdot K(y - s_2(t_1, t_2)) dt_1 dt_2. \end{aligned} \quad (\text{E.2})$$

Suppose that the target image intensity values are some function  $g$  of the geometrically transformed reference image. If we define new variable  $u_{t_2} = s(t_1, t_2)$  with fixed  $t_2$  and  $v_{t_1} = s(t_1, t_2)$  with fixed  $t_1$ , then  $D_1 s(t_1, t_2) dt_1 = du_{t_2}$  and  $D_2 s(t_1, t_2) dt_2 = dv_{t_1}$ . Therefore,

$$\begin{aligned}
\frac{\partial f_{\theta}(x, y)}{\partial \phi} \Big|_{\theta=\bar{\theta}} &= \int_{t_{1o}}^{t_{1f}} -t_2 \int_{s(t_{1o}, t_2)}^{s(t_{1f}, t_2)} \dot{K}(x - u_{t_2}) K(y - g(u_{t_2})) du_{t_2} dt_2 \\
&+ \int_{t_o}^{t_f} t_1 \int_{s(t_1, t_o)}^{s(t_1, t_f)} \dot{K}(x - v_{t_1}) K(y - g(v_{t_1})) dv_{t_1} dt_1. \tag{E.3}
\end{aligned}$$

Therefore, if  $s(t_{1f}, t_2) = s(t_{1o}, t_2)$  for every  $t_2$  and  $s(t_1, t_{2o}) = s(t_1, t_{2f})$  for every  $t_1$ , which is true by the constant background assumption, the partial derivative of the estimated pdf is equal to zero at true registered position. By the similar arguments, one can show that the partial derivative of the estimated pdf with respect to translation parameters are also equal to zero.

## APPENDIX F

### Proof of Proposition 1

First, consider a 2D Jacobian matrix defined as following:

$$J(k_1, k_2, k_3, k_4) = \begin{bmatrix} 1 + k_1 & k_2 \\ k_3 & 1 + k_4 \end{bmatrix},$$

where,  $|k_1| \leq b_1$ ,  $|k_2| \leq b_2$ ,  $|k_3| \leq b_3$ ,  $|k_4| \leq b_4$  and  $|b_i| > 0$ . It is clear that the possible maximum  $\det J(k_1, k_2, k_3, k_4)$  is  $(1+b_1)(1+b_4)+b_2b_3$ . The possible minimum  $\det J(k_1, k_2, k_3, k_4)$  is  $(1-b_1)(1-b_4)-b_2b_3$ , if  $|b_i| \leq 1$ . Note that the possible maximum and minimum occur when each gradient component has a boundary value.

For 3D Jacobian matrix, we develop the similar argument as for 2D matrix using Kuhn-Tucker multiplier. Consider 3D Jacobian matrix  $J(\underline{\mathbf{x}})$  and Jacobian determinant  $\det(J(\underline{\mathbf{x}}))$  whose elements are  $\underline{\mathbf{x}} = (x_1, \dots, x_9)$  defined as follows [71]:

$$J(\underline{\mathbf{x}}) = \begin{bmatrix} 1 + x_1 & x_2 & x_3 \\ x_4 & 1 + x_5 & x_6 \\ x_7 & x_8 & 1 + x_9 \end{bmatrix},$$

$$\begin{aligned} \det(J(\underline{\mathbf{x}})) &= (1 + x_1)(1 + x_5)(1 + x_9) + x_2x_6x_7 + x_3x_4x_8 \\ &\quad - (1 + x_1)x_6x_8 - (1 + x_5)x_3x_7 - (1 + x_9)x_2x_4. \end{aligned}$$

(F.1)

By Kuhn-Tucker Theorem [64], the necessary conditions for the minimizer  $\underline{\mathbf{x}}^*$  of  $\det(J(\underline{\mathbf{x}}))$  subject to the constraints stated in the Proposition 1 are as follows:

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_1} + \lambda_1 - \lambda_2 = (1 + x_5^*)(1 + x_9^*) - x_6^*x_8^* + \lambda_1 - \lambda_2 = 0, \quad (\text{F.2})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_2} + \lambda_3 - \lambda_4 = -x_4^*(1 + x_9^*) + x_6^*x_7^* + \lambda_3 - \lambda_4 = 0, \quad (\text{F.3})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_3} + \lambda_5 - \lambda_6 = -x_7^*(1 + x_5^*) + x_4^*x_8^* + \lambda_5 - \lambda_6 = 0, \quad (\text{F.4})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_4} + \lambda_7 - \lambda_8 = -x_2^*(1 + x_9^*) + x_3^*x_8^* + \lambda_7 - \lambda_8 = 0, \quad (\text{F.5})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_5} + \lambda_9 - \lambda_{10} = (1 + x_1^*)(1 + x_9^*) - x_3^*x_7^* + \lambda_9 - \lambda_{10} = 0, \quad (\text{F.6})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_6} + \lambda_{11} - \lambda_{12} = -x_8^*(1 + x_1^*) + x_2^*x_7^* + \lambda_{11} - \lambda_{12} = 0, \quad (\text{F.7})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_7} + \lambda_{13} - \lambda_{14} = -x_3^*(1 + x_5^*) + x_2^*x_6^* + \lambda_{13} - \lambda_{14} = 0, \quad (\text{F.8})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_8} + \lambda_{15} - \lambda_{16} = -x_6^*(1 + x_1^*) + x_3^*x_4^* + \lambda_{15} - \lambda_{16} = 0, \quad (\text{F.9})$$

$$\frac{\partial \det(J(\underline{\mathbf{x}}^*))}{\partial x_9} + \lambda_{17} - \lambda_{18} = (1 + x_1^*)(1 + x_5^*) - x_3^*x_7^* + \lambda_{17} - \lambda_{18} = 0, \quad (\text{F.10})$$

$$\lambda_1(x_1^* - k_f) = 0, \quad \lambda_2(-x_1^* - k_f) = 0, \quad (\text{F.11})$$

$$\lambda_3(x_2^* - k_f) = 0, \quad \lambda_4(-x_2^* - k_f) = 0, \quad (\text{F.12})$$

$$\lambda_5(x_3^* - k_f) = 0, \quad \lambda_6(-x_3^* - k_f) = 0, \quad (\text{F.13})$$

$$\lambda_7(x_4^* - k_g) = 0, \quad \lambda_8(-x_4^* - k_g) = 0, \quad (\text{F.14})$$

$$\lambda_9(x_5^* - k_g) = 0, \quad \lambda_{10}(-x_5^* - k_g) = 0, \quad (\text{F.15})$$

$$\lambda_{11}(x_6^* - k_g) = 0, \quad \lambda_{12}(-x_6^* - k_g) = 0, \quad (\text{F.16})$$

$$\lambda_{13}(x_7^* - k_h) = 0, \quad \lambda_{14}(-x_7^* - k_h) = 0, \quad (\text{F.17})$$

$$\lambda_{15}(x_8^* - k_h) = 0, \quad \lambda_{16}(-x_8^* - k_h) = 0, \quad (\text{F.18})$$

$$\lambda_{17}(x_9^* - k_h) = 0, \quad \lambda_{18}(-x_9^* - k_h) = 0, \quad (\text{F.19})$$

where  $\lambda_i \geq 0$  for  $i = 1, \dots, 18$ .

In (F.2), if we assume  $(1 + x_5^*)(1 + x_9^*) - x_6^*x_8^* > 0$ , then  $\lambda_2 > \lambda_1 \geq 0$ . Therefore,  $x_1^*$  is  $-k_f$  by (F.12). With similar assumptions and arguments,  $x_5^* = -k_g$  and  $x_9^* = -k_h$ . The assumption can be satisfied if  $|x_i^*| < \frac{1}{2}$ ,  $i = 1 \dots 9$  as stated in the Proposition 1. We suspect that it may be possible to remove this assumption since it is only required to argue that the diagonal components of the Jacobian matrix have boundary values.

With those conditions, it is clear from (F.1) that  $\det(J(\underline{\mathbf{x}}))$  achieves the minimum where  $x_6^*x_8^* = k_gk_h$ ,  $x_2^*x_4^* = k_fk_g$  and  $x_3^*x_7^* = k_fk_h$ . Therefore, the minimum Jacobian determinant can be represented as following:

$$\min \det(J(\underline{\mathbf{x}})) = 1 - (k_f + k_g + k_h). \quad (\text{F.20})$$

If  $k_f = k_g = k_h = k$ , then the possible minimum is  $1 - 3k$ . Therefore, the Jacobian determinant does not have a negative value if  $k \leq \frac{1}{3}$ , which is the same conclusion from the previous investigation [61].

The conditions for maximum determinant are obtained by multiplying the conditions (F.2)-(F.10) by  $-1$ . By similar arguments as for the minimum case, one can show  $x_1^* = k_f, x_5^* = k_g, x_9^* = k_h$ .

For the rest of  $x_i^*$ , we first assume that each  $x_i^*$  has the boundary value. In that case, if  $x_4^* = k_g$ , then  $\lambda_3 > \lambda_4$  by (F.3), which implies  $x_2^* = -k_f$ . Similarly,  $x_3^*, x_4^*$  have opposite sign and so  $x_6^*, x_8^*$  have. As a result, the maximum Jacobian determinant for this case is represented as following:

$$\begin{aligned} \max \det(J(\underline{\mathbf{x}})) &= (1 + k_f)(1 + k_g)(1 + k_h) + (1 + k_h)k_fk_g \\ &\quad + (1 + k_f)x_gx_h + (1 + k_g)k_fx_h. \end{aligned} \quad (\text{F.21})$$

Next we show that Jacobian determinant when  $x_2^*$  does not have the boundary value is smaller than  $\max \det(J(\underline{\mathbf{x}}))$ . For this case,  $\lambda_3 = \lambda_4 = 0$  to satisfy  $\lambda_3(x_2^* - k_f) = 0$  and  $\lambda_4(-x_2^* - k_f) = 0$ . Therefore,  $x_4^*$  is computed from (F.3) as following:

$$x_4^* = \frac{x_6^*x_7^*}{1 + k_h}. \quad (\text{F.22})$$

It is clear from (F.22) that  $x_4^*$  is neither  $k_g$  nor  $-k_g$  since the maximum possible absolute value of  $x_4^*$  is  $\frac{k_gk_h}{1+k_h}$ , which is less than  $k_g$ . Therefore,  $\lambda_7 = \lambda_8 = 0$  and  $x_2^*$  is computed from (F.5) as following:

$$x_2^* = \frac{x_3^*x_8^*}{1 + k_h}. \quad (\text{F.23})$$

By (F.22) and (F.23),  $x_2^*x_6^*x_7^* = x_3^*x_4^*x_8^* = (1 + k_h)x_2^*x_4^*$ . Denote a vector that satisfies this condition as  $\hat{\underline{\mathbf{x}}}$ . Substituting the condition into (F.1) yields,

$$\begin{aligned}
\det(J(\hat{\underline{\boldsymbol{x}}})) &= (1 + k_f)(1 + k_g)(1 + k_h) + (1 + k_h)x_2^*x_4^* \\
&\quad - (1 + k_f)x_6^*x_8^* - (1 + k_g)x_3^*x_7^* \\
&\leq (1 + k_f)(1 + k_g)(1 + k_h) + (1 + k_h)k_fk_g \\
&\quad + (1 + k_f)k_gk_h + (1 + k_g)k_fx_h \\
&= \max \det(J(\underline{\boldsymbol{x}})). \tag{F.24}
\end{aligned}$$

One can verify (F.24) for the case that  $x_3^*, x_6^*$  or  $x_7^*, x_8^*$  are not on the boundaries by using the same arguments. Therefore, we conclude that the maximum determinant value of the Jacobian matrix is represented by (F.21).

## APPENDIX G

### Proof of Proposition 2

First, consider the derivative of the 1D cubic spline defined by

$$\frac{\partial f(x)}{\partial x} = \sum_i \theta_i \frac{\partial \beta^3(\frac{x}{T} - i)}{\partial x}. \quad (\text{G.1})$$

Note that every  $x$  can be represented by  $x = \tilde{x} + w$  where  $0 \leq w \leq T$  and  $\tilde{x}$  is the nearest knot location that is smaller than  $x$ . Suppose that  $\theta_0, \theta_1, \theta_2, \theta_3$  are the coefficients of the B-spline functions that contribute to determine the gradient value at  $x$  where  $\theta_0$  is the coefficient of the B-spline function at leftmost knot in the coordinates. By using the explicit formula of 3rd order B-spline, the gradient at  $x$  is computed as follows:

$$\begin{aligned} \frac{\partial f(x)}{\partial x} \Big|_{x=\tilde{x}+w} &= \frac{1}{T} \left[ -\theta_0 \frac{1}{2} \left(1 - \frac{w}{T}\right)^2 + \theta_1 \left(-2\frac{w}{T} + \frac{3}{2}\frac{w^2}{T^2}\right) \right. \\ &\quad \left. + \theta_2 \left(2\left(1 - \frac{w}{T}\right) - \frac{3}{2}\left(1 - \frac{w}{T}\right)^2\right) + \theta_3 \frac{w^2}{2T^2} \right]. \end{aligned} \quad (\text{G.2})$$

It is clear that  $\left| \frac{\partial f(x)}{\partial x} \right|$  is bounded by  $\frac{b}{T}$  where  $w = 0$  or  $w = T$ . When  $0 < w < T$ , we determine the minimum and maximum value using Kuhn-Tucker theorem [64]. The necessary conditions for the minimizer  $\theta_0^*, \theta_1^*, \theta_2^*, \theta_3^*$  of (G.1) is determined as follows:

$$-\frac{1}{2}\left(1 - \frac{w}{T}\right)^2 + (\lambda_1 - \lambda_2) = 0 \quad (\text{G.3})$$

$$-2\frac{w}{T} + \frac{3w^2}{2T^2} + (\lambda_2 - \lambda_1) + (\lambda_3 - \lambda_4) = 0 \quad (\text{G.4})$$

$$2\left(1 - \frac{w}{T}\right) - \frac{3}{2}\left(1 - \frac{w}{T}\right)^2 + (\lambda_4 - \lambda_3) + (\lambda_5 - \lambda_6) = 0 \quad (\text{G.5})$$

$$\frac{w^2}{2T^2} + (\lambda_6 - \lambda_5) = 0 \quad (\text{G.6})$$

$$\lambda_1(\theta_0^* - \theta_1^* - b) = 0, \lambda_2(\theta_1^* - \theta_0^* - b) = 0 \quad (\text{G.7})$$

$$\lambda_3(\theta_1^* - \theta_2^* - b) = 0, \lambda_4(\theta_2^* - \theta_1^* - b) = 0 \quad (\text{G.8})$$

$$\lambda_5(\theta_2^* - \theta_3^* - b) = 0, \lambda_6(\theta_3^* - \theta_2^* - b) = 0, \quad (\text{G.9})$$

where  $\lambda_i \geq 0$ ,  $i = 1, \dots, 6$ .

Since  $\lambda_1 - \lambda_2 > 0$  by (G.3),  $\theta_1 = \theta_0 - b$ . Also,  $\lambda_6 - \lambda_5 < 0$  by (G.6) and hence  $\theta_3 = \theta_2 - b$ .

Combining (G.3) and (G.4) yields following condition:

$$-\frac{2w}{T} + \frac{3w^2}{2T^2} - \frac{1}{2} + \frac{w}{T} - \frac{w^2}{2T^2} + \lambda_3 - \lambda_4 = \frac{w^2}{T^2} - \frac{w}{T} - \frac{1}{2} + \lambda_3 - \lambda_4 = 0 \quad (\text{G.10})$$

Therefore,  $\lambda_3 > \lambda_4 \geq 0$  and  $\theta_2 = \theta_1 - b$ . Similarly, by (G.5) and (G.6), one can check that this does not violate the condition.

By plugging  $\theta_0, \theta_1, \theta_2, \theta_3$  into (G.2), the minimum gradient is computed as following:

$$\begin{aligned} \min \frac{\partial f(x)}{\partial x} &= -\theta_0 \frac{1}{2} \left(1 - \frac{w}{T}\right)^2 + \left(-2\frac{w}{T} + \frac{3w^2}{2T}\right)(\theta_0 - b) \\ &\quad + (\theta_0 - 2b) \left(2\left(1 - \frac{w}{T}\right) - \frac{3}{2}\left(1 - \frac{w}{T}\right)^2\right) + \frac{w^2(\theta_0 - 3b)}{2T^2} \\ &= -\frac{b}{T}. \end{aligned}$$

Similarly, one can show that the maximum gradient is bounded by  $\frac{b}{T}$ .

Now consider 3D case. By following inequalities, the magnitude of the gradient at everywhere is bounded by  $b$ .

$$\begin{aligned} \left| \frac{\partial f(x, y, z)}{\partial x} \right| &= \left| \sum_{ijk \in K_x} \theta_{ijk}^x \beta^n(x-i) \beta^n(y-j) \beta^n(z-k) \right| \\ &\leq \sum_{ijk \in K_x} |\theta_{ijk}^x \beta^n(x-i) \beta^n(y-j) \beta^n(z-k)| \\ &= \sum_{j,k} \sum_i |\theta_{ijk}^x \beta^n(x-i) \beta^n(y-j) \beta^n(z-k)| \\ &\leq \frac{b}{T} \sum_{j,k} \beta^n(y-j) \beta^n(z-k) \\ &= \frac{b}{T}. \end{aligned}$$

Using the similar arguments, one can show that the other gradient components are bounded by the differences of consecutive coefficients. This proves the Proposition 2.

## **BIBLIOGRAPHY**

## BIBLIOGRAPHY

- [1] M M Urie, M Gotein, K Doppke, J G Kutcher, T LoSasso, R Mohan, J E Muzenrider, M Sontag, and J W Wong, “The role of uncertainty analysis in treatment planning,” *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.
- [2] G J Kutcher, G S Mageras, and S A Leibel, “Control, correction and modeling of setup errors and organ motion,” *Seminars in Radiation Oncology*, vol. 5, no. 2, pp. 134–145, 1995.
- [3] D A Jaffray, D Yan, and J W Wong, “Managing geometric uncertainty in conformal intensity-modulated radiation therapy,” *Seminars in Radiation Oncology*, vol. 9, no. 1, pp. 4–19, 1999.
- [4] J M Balter, R K Ten Haken, and K L Lam, “Treatment setup verification,” in *Teletherapy: present and future*, T R Mackie and J R Palta, Eds., pp. 471–493. Advanced Medical Publishing, Inc., WI, 1996.
- [5] J E Schewe, J M Balter, K L Lam, and R K Ten Haken, “Measurement of patient setup errors using port films and a computer-aided graphical alignment tool,” *Med. Dosim.*, vol. 21, no. 2, pp. 97–104, 1996.
- [6] S Shalev, “Treatment verification using digital imaging,” in *Radiation therapy physics*, A R Smith, Ed., pp. 155–173. Springer-Verlag, New York, 1995.
- [7] M van Herk, A Bel, KGA Gilhuijs, JV Lebesque, A van Dalen, P van der Ven, and R Vijlbrief, “Electronic portal imaging,” *Bull Cancer*, vol. Suppl 5, no. 82, pp. 601s–605s, 1995.
- [8] K G A Gilhuijs, P J H van de Ven, and M van Herk, “Automatic three-dimensional inspection of patient setup in radiation therapy using port images, simulator images, and computed tomography data,” *Med. Phys.*, vol. 23, no. 3, pp. 389–399, 1996.
- [9] L J Verhey, “Patient immobilization for 3-D RTP and virtual simulation,” in *A practical guide to 3-D planning and conformal radiation therapy*, J A Purdy and G Starkschall, Eds., pp. 151–174. Advanced Medical Publishing, Inc., WI, 1999.
- [10] R K Ten Haken and B A Fraass, “Components of a 3-D CRT quality assurance program,” in *A practical guide to 3-D planning and conformal radiation therapy*, J A Purdy and G Starkschall, Eds., pp. 309–322. Advanced Medical Publishing, Inc., WI, 1999.

- [11] M J Murphy, “An automatic six-degree-of-freedom image registration algorithm for image-guided frameless stereotactic radiosurgery,” *Med. Phys.*, vol. 24, no. 6, pp. 857, June 1997.
- [12] J Kim, J A Fessler, K L Lam, J M Balter, and R K Ten Haken, “A feasibility study of mutual information based set-up error estimator for radiotherapy,” *Medical Physics*, vol. 28, no. 12, pp. 2391–2593, 2001.
- [13] K K Brock, D L McShan, R K Ten Haken, S J Hollister, L A Dawson, and J M Balter, “Inclusion of organ deformation in dose calculation,” *Medical Physics*, vol. 30, no. 3, pp. 290–95, 2003.
- [14] Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1976.
- [15] J Weese, G P Penney, P Desmedt, T M Buzug, D L G Hill, and D J Hawkes, “Voxel-based 2-d/3-d registration of fluoroscopy images and CT scans for image-guided surgery,” *IEEE Trans. on Info. Tech. in Biomedicine*, vol. 1, no. 4, pp. 284–293, 1997.
- [16] K G A Gilhuijs, K Drukker, A Touw, P J H van de Ven, , and M van Herk, “Interactive three dimensional inspection of patient setup in radiation therapy using digital portal images and computed tomography data,” *Int. J. Radiation Oncology Biol. Phys.*, vol. 34, no. 4, pp. 873–885, 1996.
- [17] K G A Gilhuijs and M van Herk, “Automatic on-line inspection of patient setup in radiation therapy using digital portal images,” *Med. Phys.*, vol. 20, no. 3, pp. 667–677, 1993.
- [18] L Lemieux, R Jagoe, D R Fish, N D Kitchen, and D G T Thomas, “A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs,” *Med. Phys.*, vol. 21, no. 11, pp. 1749–1760, 1994.
- [19] J Bijhold, “Three-dimensional verification of patient placement during radiotherapy using portal images,” *Med. Phys.*, vol. 20, no. 2, pp. 347–356, 1993.
- [20] L Dong and A L Boyer, “An image correlation procedure for digitally reconstructed radiographs and electronic portal images,” *Int J. Radiation Oncology Biol. Phys.*, vol. 33, no. 5, pp. 1053–1060, 1995.
- [21] W M Wells, P Viola, H Atsumi, S Nakajima, and R Kikinis, “Multi-modal volume registration by maximization of mutual information,” *Med. Im. Anal.*, vol. 1, no. 1, pp. 35–51, Mar. 1996.
- [22] F Maes, A Collignon, D Vandermeulen, G Marchal, and P Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 187–98, Apr. 1997.
- [23] C R Meyer, J L Boes, B Kim, P H Bland, et al., “Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations,” *Med. Im. Anal.*, vol. 1, no. 3, pp. 195–206, 1997.
- [24] P J Huber, *Robust statistics*, Wiley, New York, 1981.

- [25] J. Kim and J. A. Fessler, "Image registration using robust correlation," in *Proc. IEEE Intl. Symp. Biomedical Imaging*, 2002, pp. 353–6.
- [26] J A Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Tr. Im. Proc.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.
- [27] K M Brock, J M Balter, L A Dawson, M L Kessler, and C R Meyer, "Automated generation of a four-dimensional model of liver using warping and mutual information," *Medical Physics*, vol. 30, no. 6, pp. 1128–33, 2003.
- [28] T Rohlfing, C R Maurer Jr., D A Bluemke, and M A Jacobs, "Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint," *IEEE Tr. Med. Im.*, vol. 22, no. 6, pp. 730–741, June 2003.
- [29] D Rueckert, L I Sonoda, C Hayes, D L G Hill, M O Leach, and D J Hawkes, "Nonrigid registration using free-form deformations: Application to breast mr images," *IEEE Tr. Med. Im.*, vol. 18, no. 8, pp. 712–20, Aug. 1999.
- [30] J Kybic, P Thévenaz, A Nirkko, and M Unser, "Unwarping of unidirectionally distorted epi images," *IEEE Tr. Med. Im.*, vol. 19, no. 2, Feb. 2000.
- [31] L Dong and A L Boyer, "A portal image alignment and patient setup verification procedure using moments and correlation techniques," *Phys. Med. Biol.*, vol. 41, pp. 697–723, 1996.
- [32] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling, *Numerical recipes in C*, Cambridge Univ. Press, 1988.
- [33] R L Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, Mar. 1985.
- [34] M Cramer, *Mathematical methods of statistics*, Princeton University Press, 1946.
- [35] T M Cover and J A Thomas, *Elements of information theory*, John Wiley and Sons, New York, 1991.
- [36] J M Fitzpatrick, D L G Hill, and C R Maurer, "Image registration," in *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, M. Sonka J. Michael Fitzpatrick, Ed., pp. 477–513. SPIE, Bellingham, 2000.
- [37] F Maes, D Vandermeulen, and P Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Med. Im. Anal.*, vol. 3, no. 4, pp. 373–86, 1999.
- [38] G P Penney, J Weese, J A Little, P Desmedt, D L G Hill, and D J Hawkes, "A comparison of similarity measures for use in 2-d-3-d medical image registration," *IEEE Trans. Medical Imaging*, vol. 17, no. 4, pp. 284–293, August 1998.
- [39] S W Hadley, C A Pelizzari, L S Johnson, and G T Y Chen, "Automated registration of portal and simulation films by mutual information," in *Proceedings of XIIth ICCR*, May 1997, pp. 243–244.

- [40] A Rényi, “On measures of entropy and information,” in *Proc. of the Fourth Berkeley Symposium on Math. Stat. and Prob.*, 1961, vol. 1, pp. 547–61.
- [41] A O Hero and O Michel, “Asymptotic theory of greedy approximations to minimal  $k$ -point random graphs,” *IEEE Tr. Info. Theory*, vol. 45, pp. 1921–1939, Sept. 1999.
- [42] D M Titterton, “Estimation of correlation coefficients by ellipsoidal trimming,” *Applied Statistics*, vol. 23, no. 3, pp. 227–234, 1978.
- [43] N A Campbell, “Robust procedure in multivariate analysis 1: Robust covariance estimation,” *Applied Statistics*, vol. 29, no. 3, pp. 231–237, 1980.
- [44] P J Rousseeuw and B C van Zomeren, “Unmasking multivariate outliers and leverage points,” *Journal of the American Statistical Association*, Sep. 1990.
- [45] J W Tukey, *Exploratory data analysis*, Addison-Wesley, 1977.
- [46] W L Hays, *Statistics*, Holt, Rinehart and Winston, 1981.
- [47] E Carrizosa, “A characterization of halfspace depth,” *Journal of Multivariate Analysis*, vol. 58, pp. 21–6, 1996.
- [48] X He and G Wang, “Convergence of depth contours for multivariate datasets,” *The Annals of Statistics*, vol. 25, no. 2, pp. 495–504, 1997.
- [49] E L Lehmann, *Theory of point estimation*, Wiley, New York, 1983.
- [50] B W Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, New York, 1986.
- [51] B Ma, A Hero, J Gorman, and O Michel, “Image registration with minimum spanning tree algorithm,” in *Proc. IEEE Intl. Conf. on Image Processing*, 2000.
- [52] H. F. Neemuchwala, A. O. Hero, and P. L. Carson, “Image registration using entropic graph-matching criteria,” in *36th Asilomar Conf. on Signals Systems and Computers, Pacific Grove, CA, Nov. 2002*, 2002.
- [53] J D Gorman and A O Hero, “Lower bounds for parametric estimators with constraints,” *IEEE Tr. Info. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.
- [54] D C Noll, C H Meyer, J M Pauly, D G Nishimura, and A Macovski, “A homogeneity correction method for magnetic resonance imaging with time-varying gradients,” *IEEE Tr. Med. Im.*, vol. 10, no. 4, pp. 629, Dec. 1991.
- [55] M Holden, D L G Hill, E R E Denton, J M Jarosz, T C S Cox, T Rohlfing, J Goodey, and D J Hawkes, “Voxel similarity measures for 3D serial MR brain image registration,” *IEEE Tr. Med. Im.*, vol. 19, no. 2, pp. 94–102, Feb. 2000.
- [56] P A Viola, *Alignment by maximization of mutual information*, Ph.D. thesis, Massachusetts Institution of Technology, 1995.
- [57] C V Stewart, “Robust parameter estimation in computer vision,” *SIAM Review*, vol. 41, pp. 511–37, 1999.

- [58] F L Bookstein, “Principal warps: thin-plate splines and the decomposition of deformations,” *IEEE Tr. Patt. Anal. Mach. Int.*, vol. 11, no. 6, pp. 567–87, June 1989.
- [59] D Mattes, D R Haynor, H Vesselle, T K Lewellen, and W Eubank, “Pet-ct image registration in the chest using free-form deformations,” *IEEE Tr. Med. Im.*, vol. 22, no. 1, pp. 120–28, Jan. 2003.
- [60] G E Christensen and H J Johnson, “Consistent image registration,” *IEEE Tr. Med. Im.*, vol. 20, no. 7, pp. 568–82, July 2001.
- [61] G K Rhode, A Aldroubi, and B M Dawant, “The adaptive bases algorithms for nonrigid image registration,” in *Proc. SPIE*, 2002, vol. 4684, pp. 933–43.
- [62] F Deutch and H Hundal, “The rate of convergence of dykstra’s cyclic projections algorithm:the polyhedral case,” *Numer. Funct. Anal. and Optimiz.*, vol. 15, pp. 537–65, 1994.
- [63] B T Polyak, *Introduction to Optimization*, Optimization Software Inc., New York, 1987.
- [64] D G Luenberger, *Optimization by vector space methods*, Wiley, New York, 1969.
- [65] M Unser, “Splines - a perfect fit for signal and image processing,” *IEEE Signal processing magazine*, pp. 22–38, Nov 1999.
- [66] B Karacali and C Davatzikos, “Topology preservation and regularity in estimated deformation fields,” in *Information Processing in Medical Imaging*, 2003.
- [67] A Iusem and A De Pierro, “On the convergence of han’s method for convex programming with quadratic objective,” *Mathematical Programming*, vol. 52, pp. 265–84, 1991.
- [68] J Kim and J A Fessler, “Intensity-based image registration using robust correlation coefficients,” *Submitted to IEEE Tr. Med. Im.*, 2003.
- [69] A Booker, John E Dennis, P Frank, D Serafini, V Torczon, and M Trosset, “A rigorous framework for optimization of expensive functions by surrogates,” *Structural Optimization*, vol. 17, no. 1, pp. 1–13, Feb. 1999.
- [70] M D Srinath, P K Rajasekaran, and R Viswanathan, *Introduction to statistical signal processing with applications*, Prentice-Hall, 1996.
- [71] G Strang, *Linear algebra and its applications*, Saunders, 3rd edition, 1986.