

Deformation Estimation and Assessment of Its Accuracy in Ultrasound Images

Roozbeh Shams

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science at

Concordia University

Montréal, Québec, Canada

December 2017

© Roozbeh Shams, 2017

**CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES**

This is to certify that the thesis prepared

By: Mr. Roozbeh Shams

Entitled: Deformation Estimation and Assessment of Its Accuracy in Ultrasound Images

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____	Chair
Dr. R. Raut	
_____	Examiner, External
Dr. Catherine Laporte, Associate Prof (ETS)	To the Program
_____	Examiner
Dr. Krzysztof Skonieczny	
_____	Supervisor
Dr. H. Rivaz	
_____	Co-Supervisor
Dr. R. Brooks	

Approved by: _____
Dr. W. E. Lynch, Chair
Department of Electrical and Computer Engineering

_____ 20_____

_____ Dr. Amir Asif, Dean
Faculty of Engineering and Computer
Science

ABSTRACT

Deformation Estimation and Assessment of Its Accuracy in Ultrasound Images

Roozbeh Shams

This thesis aims to address two problems; one in ultrasound elastography and one in image registration. The first problem entails estimation of tissue displacement in Ultrasound Elastography (UE). UE is an emerging technique used to estimate mechanical properties of tissue. It involves calculating the displacement field between two ultrasound Radio Frequency (RF) frames taken before and after a tissue deformation. A common way to calculate the displacement is to use correlation based approaches. However, these approaches fail in the presence of signal decorrelation. To address this issue, Dynamic Programming was used to find the optimum displacement using all the information on the RF-line. Although taking this approach improved the results, some failures persisted. In this thesis, we have formulated the DP method on a tree. Doing so allows for more information to be used for estimating the displacement and therefore reducing the error. We evaluated our method on simulation, phantom and real patient data. Our results shows that the proposed method outperforms the previous method in terms of accuracy with small added computational cost.

In this work, we also address a problem in image registration. Although there is a vast literature in image registration, quality evaluation of registration is a field that has not received as much attention. This evaluation becomes even more crucial in medical imaging due to the

sensitive nature of the field. We have addressed the said problem in the context of ultrasound guided radiotherapy. Image guidance has become an important part of radiotherapy wherein image registration is a critical step. Therefore, an evaluation of this registration can play an important role in the outcome of the therapy. In this work, we propose using both bootstrapping and supervised learning methods to evaluate the registration. We test our methods on 2D and 3D data acquired from phantom and patients. According to our results, both methods perform well while having advantages and disadvantages over one another. Supervised learning methods offer more accuracy and less computation time. On the other hand, for bootstrapping, no training data is required and also offers more sensitivity.

*To my parents,
for their love. for their support.*

ACKNOWLEDGEMENTS

I'd like to thank my supervisor, Dr. Hassan Rivaz. It was through his support that I could present this work. I deeply and truly appreciate his care through the ups and downs of my studies.

I also would like to thank my co-supervisor, Dr. Rupert Brooks, whom not only guided me academically, but also was there for me if I needed any other advice.

I feel blessed to have them as my supervisors.

I would also like to thank Dr. Paul Martineau and Dr. Mathieu Boily for their discussions and their valuable input.

I'd also like to thank Dr. Martin Lachaine, Francois Hebert and others at Elekta for their discussions and making my stay at Elekta a memorable one.

I'm thankful to Dr. Catherine Laporte , Dr. Krzysztof Skonieczny and Dr. Rabin Raut who are the examiners of my thesis. I'm grateful for their guidance and valuable inputs.

Also, I'd like to express my gratitude towards my friends and colleagues at the IMPACT lab, for their discussions, their supports, the coffee breaks and for the memories that are left with me. Not forgetting the friends that were my family away from home, who supported me and were there for me when I needed it the most.

In the end, I'm forever indebted to my family, my parents, my sister, my brother-in-law, my uncle and his family and all the rest. I wouldn't be here without your support and love. I

owe you everything. Thank you.

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2015-0413, an NSERC Engage Grant, and by Richard and Edith Strauss Canada Foundation.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xii
1 Introduction	1
1.1 Ultrasound imaging	1
2 Ultrasound Elastography Using Dynamic Programming on a Tree	7
2.1 Introduction	7
2.2 Methods	11
2.2.1 Initial Displacement Calculation	13
2.2.2 Subsample Displacement Calculation	18
2.3 Results	19
2.3.1 Simulation	20
2.3.2 Phantom Experiments	22
2.3.3 Patients With Liver Cancer	22
2.3.4 Patellar Tendon	23
2.4 Discussions and Conclusion	24
3 Assessment of Rigid Registration Quality Measures in Ultrasound-Guided Radiotherapy	27

3.1	Introduction	27
3.2	Methods	31
3.2.1	Registration	31
3.2.2	Data Preparation	33
3.2.3	Supervised Learning Methods	35
	Feature Extraction	35
	Training and validation	41
3.2.4	Bootstrapping	42
3.2.5	Bootstrapping for registration evaluation	42
3.2.6	Experimental setup	44
3.2.7	Phantom study	45
3.2.8	Patient Trials	47
3.3	Results	50
3.3.1	Feature selection	50
3.3.2	Registration evaluation	51
3.4	Discussions	51
3.5	Conclusion	57
4	Conclusion and Future Work	58
4.1	Conclusion	58
4.2	Future work	60

LIST OF TABLES

2.1	The MSE and the standard deviation of the squared error for the simulation data with different noise levels.	21
2.2	The MSE and the standard deviation of the squared error for the Phantom, Patient 1, Patient 2 and Patellar Tendon data.	23
3.1	Comparison of approaches.	52

LIST OF FIGURES

1.1	Alpinion E-cube 15.	2
1.2	Ultrasound imaging using a linear transducer.	4
1.3	Types of ultrasound probes. (a) is a linear, (b) is a curvilinear and (c) is a sector probe.	5
1.4	Elekta Clarity ultrasound machine for ultrasound guided radiotherapy.	6
2.1	Calculating ID on an RF-line (left) and on a tree (right).	10
2.2	Strain images of a tissue mimicking phantom.	12
2.3	Tree structures used.	14
2.4	NCC values in phantom data.	17
2.5	The MSE error for the simulation data.	21
2.6	Results of the phantom experiment.	22
2.7	<i>In-vivo</i> images of human data.	24
2.8	Patellar tendon B-mode and displacement.	25
3.1	Generating poor and successful registrations.	36
3.2	Demonstration of good and bad registration results for the patient US data.	37
3.3	An overview of bootstrapping for registration evaluation.	43
3.4	The experiment setup.	46

3.5	Feature importance according to RF.	49
3.6	ROC curves of different registration assessment methods for the 3D patient data.	53
3.7	Figures showing the registration error.	56

Chapter 1

Introduction

In this chapter, we will first review ultrasound (US) imaging and outline its principles. We then overview image registration techniques and provide background and fundamental knowledge on the matter.

1.1 Ultrasound imaging

Ultrasound waves are acoustic waves with frequencies higher than 20 KHz, with medical ultrasound usually operating in the 1 MHz to 20 MHz. Ultrasound waves propagate into the tissue, and interact with it in several different ways including reflection, refraction, scattering and attenuation. Ultrasound imaging is a widely used modality in the medical field. It is the second most used modality of imaging after X-ray based imaging. This widespread use is mainly due to it being noninvasive, safe, realtime, easy to use for both patients and operators



Figure 1.1: Alpinion E-cube 15.

and relatively inexpensive. An ultrasound machine is depicted in Fig. 1.1.

Ultrasound gained practical usage during World War I for detecting submarines; however, it was first used in the medical field in 1950s. It was initially used in obstetrics and later in abdominal imaging, field of pelvis, cardiology and orthopedics. It is seeing growing applications in diagnosis of several pathologies, as well in guiding operations [1].

The first step of ultrasound imaging involves the generation of sound waves by a piezoelectric transducer which is contained in a probe. The piezoelectric property is the ability of a

material to generate electricity under mechanical stress and vice versa; meaning upon applying an electric field, this material will produce a mechanical wave. Electric pulses generated by the ultrasound machine produce acoustic waves through the transducer. The frequency of these waves typically vary between 1 MHz to 20 MHz. Lower frequencies are usually used for imaging deep into the tissue, whereas higher frequencies are used for superficial structures.

The generated sound wave travels through the body and is reflected at the tissue boundaries due to the difference in the acoustic impedance of the layers. The reflections are received by the transducer, converted to electrical pulses or Radio Frequency (RF) signals and processed to generate the image. The process mainly includes finding the envelope of the RF signal and reducing the dynamic range which produces brightness values. This image is called B-mode and is the image seen as the output of the ultrasound machine. Fig. 1.2 shows a patient's knee being imaged and the produced image.

In ultrasound images, a speckle "noise" is present. Although the term noise is used here, the speckles are not of random nature; they are created due to the micro structures in the tissue that are much smaller than the wavelength of the ultrasound wave. Although the speckle noise is considered visually degrading, it also contains information about the tissue and is of use in tissue tracking applications [2].

There are different types of probes developed for different applications. Linear, curvilinear and sector probes are among the most common types. In linear probes, as the name suggest, the piezoelectric crystals are positioned on a line. These probes scan the tissue with

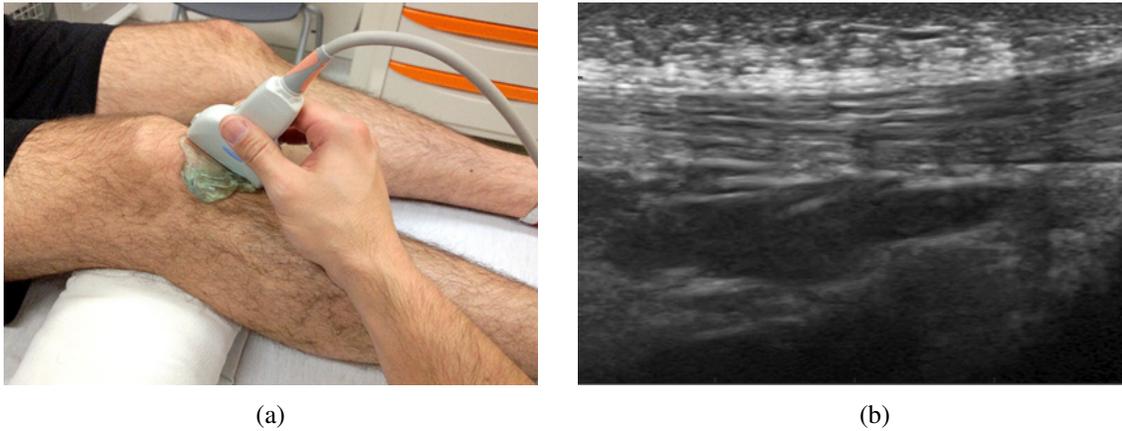


Figure 1.2: Ultrasound imaging using a linear transducer. (a) Imaging the patellar tendon (b) the image produced.

the same width as the probe and are used for low depth scanning applications, e.g. vascular, musculoskeletal, breast imaging. In a curved linear probe, the crystals are also placed on a line but on a curved surface. This allows for wider scanning angles but comes at the cost of reduced resolution at far field i.e. away from the probe. These probes are used for abdominal imaging. Sector probes are made with smaller footprints. These type of probes are used in situations where the imaging window is small, such as echocardiography and neuroimaging.

Ultrasound can be used to acquire both 2D images and 3D volumes. The 3D volumes can be acquired by a mechanically sweeping transducer or by electronically steered beams with a 2D phased array transducer. Although the former is slower, is much more price efficient than the latter.

There are many applications for ultrasound in medicine; in this work, we have focused on two applications: ultrasound elastography and ultrasound guided radiotherapy. In ultrasound elastography the goal is to infer mechanical properties of the tissue including elasticity.

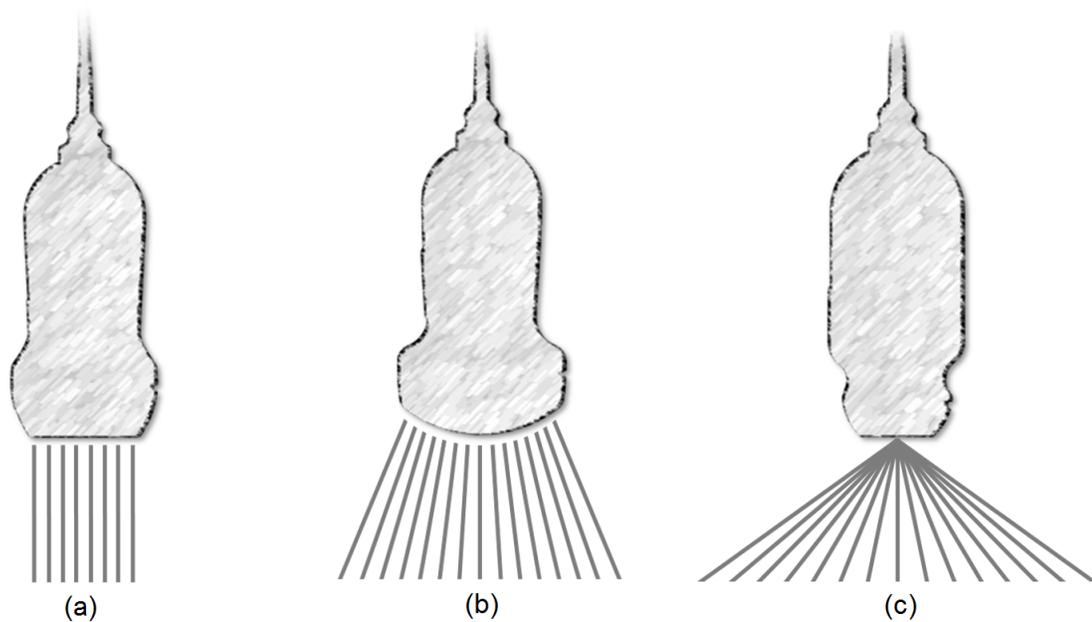


Figure 1.3: Types of ultrasound probes. (a) is a linear, (b) is a curvilinear and (c) is a sector probe.

It is based on the fact that pathologies have different elasticity than healthy tissues. It can be used for diagnosis and monitoring/guiding procedures. Ultrasound is also used during radiotherapy to track the target organ and ensure that intended dose is received. This also helps reduce the amount of unwanted radiation applied to healthy tissue and therefore minimize the side effects of radiotherapy. Fig. 1.4 shows Elekta Clarity ultrasound machine used for motion management during radiotherapy.

The contributions of others in the papers that formed this thesis is as follows. Yiming Xiao, Francois Hebert, Matthew Abramowitz have contributions in the paper included in Chapter 3. Y.X. contributed in the writing and data analysis, F.H. contributed in conception and data analysis and M.A. contributed in conception and data collection. The contribution



Figure 1.4: Elekta Clarity ultrasound machine for ultrasound guided radiotherapy. Used with permission from Elekta.

of the author of thesis were in conception of methods and approaches, implementation of the registration framework based on Elekta's tools compatible with bootstrapping; also, building on the previous machine learning framework, carrying out the experiment and data processing and lastly drafting the manuscript. Chapter 3 has been published in [3]. Parts of chapter 2 were previously published as [4] and the chapter as presented is lightly edited from a submission to BioMedical Engineering OnLine journal. In the said paper, the contributions of Hoda Sadat Hashemi, Paul Martineau and Mathieu Boily were in data collection.

Chapter 2

Ultrasound Elastography Using Dynamic Programming on a Tree

2.1 Introduction

Ultrasound elastography involves imaging mechanical properties of tissue by estimating tissue deformation due to external or internal sources of deformation [5]. Elastography has evolved into many different variations with promising results [6], [7], [8], [9]. The focus of this work is on palpation quasi-static elastography, where the probe is hand-held and tissue is compressed by manually pressing the ultrasound probe [10], [11]. Analyzing the pre- and post-compression ultrasound Radio Frequency (RF) data yields a deformation map, which is then used to estimate strain images. Palpation elastography does not require any special

equipment, and as such, can be conveniently used in both diagnosis and surgical planning [11], [12], [13], [14]. However, there are challenges to overcome before it can be widely used for clinical purposes. Two major issues that need to be addressed are real time constraints and effects of signal decorrelation between pre-compression and post-compression images.

Methods commonly used for displacement calculation fall into two main categories based on amplitude [5, 10, 15] or phase of the RF data [16], [17], [18]. In both approaches, an important parameter is the window size. Larger windows decrease the estimation variance but suffer from larger decorrelation especially for large compressions. Furthermore, there is more chance that few samples in a large window are outliers, which can render the window inadmissible. Finally, correlation methods can be computationally expensive and real-time performance is sometimes achieved by parallelization [19].

Given these challenges, accurate and robust estimation of the displacement field is an active field of research. Rao *et al.* [20] used beam steering to improve the quality of the displacement field, especially in the lateral direction. Rivaz *et al.* [14] proposed using three (or multiple) frames for generating strain images. This was done by deriving biomechanical constraints on the variation of the displacement field with time, and incorporating those constraints with displacement estimation using an Expectation Maximization (EM) framework. Kuzmin *et al.* proposed a method [21] wherein three RF data frames are acquired with a force-controlled ultrasound probe. The calculated displacement from the first two frames is used to improve the accuracy of the displacement calculation between the first and third frame. Jiang

& Hall [22] proposed to calculate the iso-contours of Normalized Cross Correlation (NCC) to find the accurate location of the maximum NCC [23].

An alternative approach to estimate the displacement field is by optimizing a regularized cost function, where the prior information of displacement continuity can be utilized to overcome signal decorrelation. It is shown that these methods generally outperform methods that only rely on RF data [24–26]. Dynamic Programming (DP) is an attractive optimization method, which is computationally efficient and can find the global optimum [24, 27, 28].

DP can be thought of as a "smart brute-force" method for optimization where the results of calculations are saved and reused. In other words the optimization is solved iteratively and the result of recurrent calculations are saved and later used in the next steps. More details on this method will be discussed later in the thesis and also can be found in [24].

All of the aforementioned methods focus on calculating a more accurate displacement image, but do not address the problem of preventing failure in the displacement map. Previous work has focused on selecting "good" pairs of ultrasound images to both reduce the chance of failure and improve the quality of strain images. Foroughi *et al.* [29] proposed to use an external tracking device to track the ultrasound probe, and use the tracking information to select two images that minimize the chance of failure in strain estimation. A different approach is taken in [30] whereby a one-prediction-one-correction method is developed for dynamically choosing the pre- and post-compression images in real-time. The goal of the said work is robust estimation of a strain image by optimizing a cost function, given a pair of

ultrasound images that possibly contain outlier data.

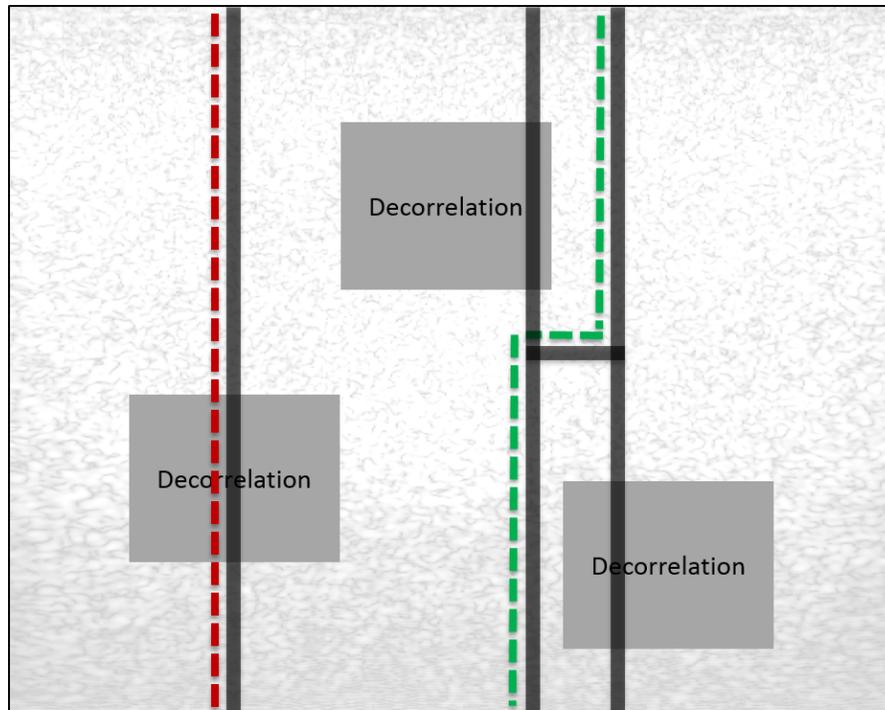


Figure 2.1: Calculating ID on an RF-line (left) and on a tree (right). Rectangles indicate regions with large signal decorrelation. DP will likely fail in left, but will find the correct path (in green) in right.

Optimization-based displacement estimation methods are usually iterative [28, 31, 32], and their success relies heavily on the Initial Displacement (ID) field. DP is an efficient optimization method for finding the *global* optimum, but gives only integer displacements. Therefore, DP is an ideal method for providing an ID field to a following sub-pixel displacement estimation technique. Previous work [14, 24, 27, 28] utilizes only a single RF-line for DP optimization, and as such, can fail if large segments of that RF-line contains large noise or outlier data. The starting RF-line can be called a seed-line, whose integer displacements are

the propagated to the entire image. Thereby, robust estimation of ID in the seed-line is critical. Fleming *et al.* [33] proposed to run DP on multiple lines and select the best outcome. While this work significantly improves the performance, it only uses single RF-lines for DP optimization. Therefore a portion of RF-line with large decorrelation can render the entire DP displacement inadmissible.

Herein, we propose a technique wherein DP is estimated on a tree instead of a single RF-line, as shown in Fig. 2.1. Optimizing DP on a tree instead of a line was first proposed by Veksler [34] in the field of computational stereo. This approach allows us to exploit data from multiple RF-lines to improve the reliability of ID field estimated from DP. In this figure, DP will likely fail due to the region with large decorrelation (left line). However, formulating DP on a tree can overcome this issue by selecting “good parts” of each RF-lines. We call our method Elastography using Dynamic Programming On a Tree (EDPOT). This chapter is summarized as follows. In the next section, we illustrate the technical details of our algorithm. We then show the results on simulation, phantom and *in-vivo* human data, and provide conclusions and avenues for future work.

2.2 Methods

Let \mathbf{I}_1 and \mathbf{I}_2 to be pre- and post-compression ultrasound images of size $m \times n$. The ultimate goal is to calculate the matrices \mathbf{A} and \mathbf{L} such that A_{ij} and L_{ij} are the axial and lateral displacements for pixel (i, j) of the ultrasound image.

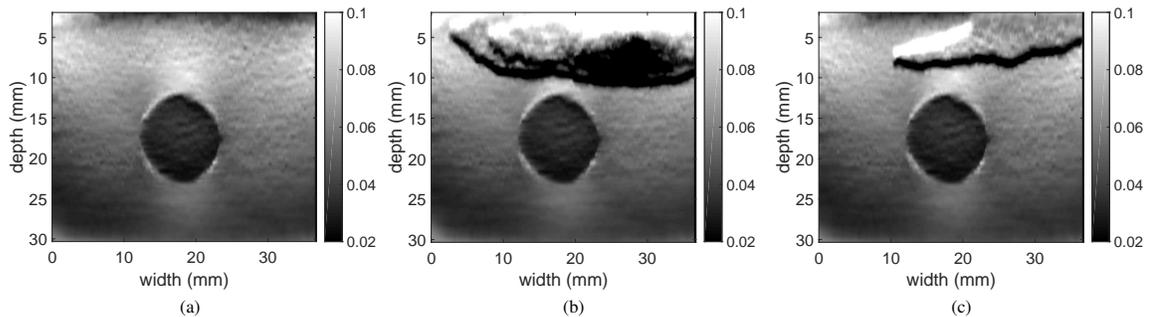


Figure 2.2: Strain images of a tissue mimicking phantom. A correct strain image is shown in (a), and two examples of incorrect strain images are shown in (b) and (c). The long dark and bright bands in the top part of (b) and (c) are artifacts and are caused by failure in DP. All results are generated using the method proposed in [28].

Erroneous ID results create distinct artifacts in the strain images as shown in Fig. 2.2. The outcome of DP for the seed-line itself very much depends on the RF-line chosen as the seed-line; for the seed-lines whose out-of-plane and lateral motion is large, DP will likely fail. To address this problem, our proposed algorithm has two steps: estimation of ID for a “seed-tree”, followed by estimation of subpixel displacement. For the second step, we use the technique proposed in to [28]. The focus of this work is on the first step so to improve the DP estimation of the ID. Our proposed algorithm for estimation of the strain image can be summarized as following:

1. Calculating ID

- (a) Designing a tree to calculate the ID
- (b) Constructing a recursive cost function for pixels on the said tree
- (c) Using DP to find the optimum displacements

- (d) Choosing a path on the tree with the most accurate displacement
2. Calculating sub-sample displacement
 - (a) Deriving the sub-sample displacement of the seed-line by means of AM [28]
 - (b) Using the sub-sample displacement of the seed-line as an estimate for calculating the displacement of the neighboring RF-lines and propagating the displacement
 3. Calculating the strain image using Kalman filter

Step 1 is the focus of this work and is described below.

2.2.1 Initial Displacement Calculation

Our underlying goal is to exploit more information in the RF data. To achieve this goal, the information in the neighboring lines of the seed-line is utilized. A general solution to discrete global optimization of a cost function that considers 4 neighbors of a pixel in the regularization term is NP-hard [34] and therefore is computationally intractable.

To overcome this issue, Veksler [34] proposed to formulate DP on a tree to take advantage of more information. We adopt a similar approach and calculate DP on a tree instead of a single RF-line. Fig. 2.1 shows DP on a single seed-line in left, and the method that estimates DP on a tree in right. Let $G_H(V, E)$ be the graph with vertices V and edges E . The structure of the tree and the key vertices can be seen in Fig. 2.3(a). The parameters involving this structure are the distances between some of the key vertices: $D_{tm} = dist(v_{t_1}, v_{m_1})$, $dist(v_{m_1}, v_{b_1})$ and

$dist(v_{m_1}, v_{m_2})$; Also $dist(v_{m_2}, v_{m_3}) = dist(v_{m_1}, v_{m_2})$. Indices t , m and b respectively refer to top, middle and bottom. The details regarding the structure are further explained in Section 3.3.

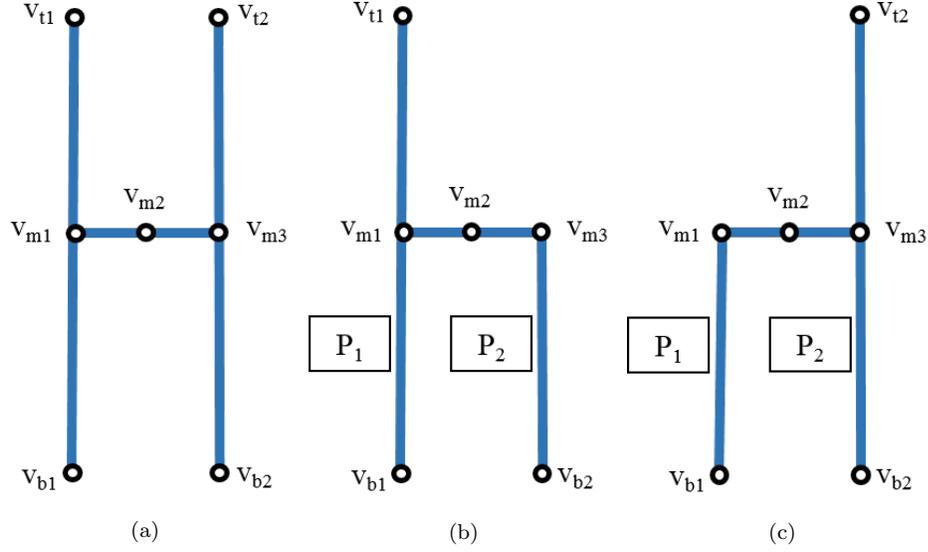


Figure 2.3: Tree structures used. ID is calculated for the pixels on G_H , the tree depicted in (a). G_H is broken down to G_1 and G_2 , shown in (b) and (c) for ID calculation.

The next step after deciding on the tree structure is to calculate the ID on the tree. In this regard, we construct a cost function:

$$C(a_{i,j}, l_{i,j}, i, j) = \Delta(i, j, a_{i,j}, l_{i,j}) + \min_{\delta_a, \delta_l} \{C(\delta_a, \delta_l, i_p, j_p) + wS(a_{i,j}, l_{i,j}, \delta_a, \delta_l)\}, \quad (2.1)$$

where

$$\Delta(i, j, a_{i,j}, l_{i,j}) = |I_1(i, j) - I_2(i + a_{i,j}, j + l_{i,j})|, \quad (2.2)$$

and

$$S(a_{i,j}, l_{i,j}, a_{i_p,j_p}, l_{i_p,j_p}) = |a_{i,j} - a_{i_p,j_p}| + |l_{i,j} - l_{i_p,j_p}|. \quad (2.3)$$

Δ is the data term and S is the regularization term of the cost function, C . i and j are integers from 1 to m and 1 to n respectively. (i_p, j_p) is the coordinate of the parent of the node at (i, j) . $a_{i,j}$ and $l_{i,j}$ are the axial and lateral displacements at pixel (i, j) which is on the tree. Also, w is the regularization weight which determines the smoothness of the calculated displacement function. We use DP to optimize this cost function on the tree structure and generate the ID estimates of the seed-tree.

In order to calculate ID on the tree depicted in Fig. 2.3(a), we break down G_H to the trees in Fig. 2.3(b) and 2.3(c) (G_1 and G_2). Thereafter, we aggregate the results of each of the trees to calculate the final displacement. For the G_1 , assume P_1 and P_2 to be two paths on the tree: P_1 is the path from v_{m_1} to v_{b_1} and P_2 is the path from v_{m_1} to v_{b_2} . The next step involves choosing the path wherein a more accurate ID can be calculated. Veksler [34] considers the cost value at v_{t_1} which is calculated based on the pixels on each path and chooses the path with smaller cost. However, our result showed that this approach does not necessarily select the best path in ultrasound images due to the following reason. The value of C heavily depends on intensity values of RF data, which are highly dependent on tissue echogeneity and wave attenuation. Intensity of RF data can, in fact, vary across an image by three order of magnitude. Therefore, we propose the following novel approach. The optimum ID is first calculated for the pixels on P_1 and P_2 . Let (i_1, j_{max}) and (i_2, j_{max}) be the coordinates of the

pixels on P_1 and P_2 which are on the same row and have the greatest difference in calculated ID:

$$j_{max} = \operatorname{argmax}_j |a_{P_1}(i_1, j) - a_{P_2}(i_2, j)| \quad (2.4)$$

where a_{P_1} and a_{P_2} are the calculated axial displacements on P_1 and P_2 and i_1 and i_2 are the column indexes for pixels on P_1 and P_2 . We then calculate NCC_1 and NCC_2 for the mentioned points on P_1 and P_2 respectively:

$$NCC_p = \frac{1}{N} \sum_{x,y} \frac{(w_1(x, y) - \bar{w}_1)(w_2(x, y) - \bar{w}_2)}{\sigma_{w_1} \sigma_{w_2}} \quad (2.5)$$

where w_1 and w_2 are 9×5 windows, centered at (i_p, j_{max}) , $(i_p + a_{i_p, j_{max}}, j_{max} + l_{i_p, j_{max}})$ in I_1 and I_2 respectively and N is the number of pixels in the window.

The path which contains the point with higher NCC will be chosen and the ID of this path will be used for the next steps. RF data is the result of modulation of a high-frequency carrier signal with an input signal, and therefore, NCC can change significantly even with a small shift of the window (Fig. 2.4). Moreover, presence of small errors in ID is inevitable due to it being integer. Therefore the changes in NCC with small shifts renders NCC of RF data ineffective, and therefore, we use envelope data in Equation 2.5. Also, to further reduce the result of these errors, in practice, we also calculate the NCC for w_2 centered at

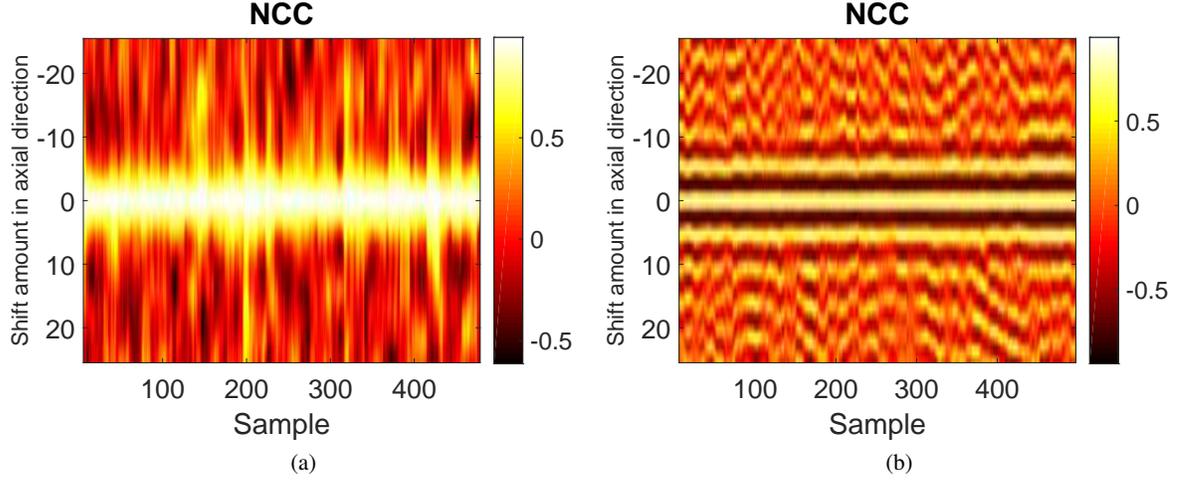


Figure 2.4: NCC values in phantom data. The x-axis is the sample number in the lateral direction (i.e. different RF-lines) and the y-axis is the shift from correct displacement (i.e. maximum NCC is expected at 0). (a) shows the NCC for a 9×5 window with vertical shifts for pixels of a single row of the envelope data, (b) shows the same for the raw RF-data. The effect of the carrier wave is clearly visible in (b).

$(i_p + a_{i_p, j_{max}} + 1, j_{max} + l_{i_p, j_{max}})$ and $(i_p + a_{i_p, j_{max}} - 1, j_{max} + l_{i_p, j_{max}})$. The maximum NCC value between the three will be used to compare the paths. In the next step, the ID on the core seed-line (the vertical line containing v_{m_2}), ID_1 , is estimated based on the displacement calculated on the chosen path.

We then proceed to calculating the ID for G_2 (ID_2). The final ID is selected by comparing ID_1 and ID_2 in the same manner that we chose either P_1 or P_2 : finding the point where ID differs the most and verify the accuracy using NCC for the ID of that point. This displacement is then used in the next step to find a subpixel displacement estimate.

2.2.2 Subsample Displacement Calculation

In this step, subsample displacement (SD) is first calculated for the core seed-line and propagated to the left and right using SD of the previous RF-line as the initial displacement, as proposed in [28]. Therefore, for one line at a time, the goal is to find the optimum Δa_i and Δl_i which make the duple $(a_i + \Delta a_i, l_i + \Delta l_i)$ the optimum solution for the following function:

$$\begin{aligned}
 C(\Delta a_1, \dots, \Delta a_m, \Delta l_1, \dots, \Delta l_m) = & \\
 \sum_{i=1}^m \{ [I_1(i, j) - I_2(i + a_i + \Delta a_i, j + l_j + \Delta l_j)]^2 \} + & \\
 \alpha (a_i + \Delta a_i - a_{i-1} - \Delta a_{i-1})^2 + & \\
 \beta_a (l_i + \Delta l_i - l_{i-1} - \Delta l_{i-1})^2 + \beta'_l (l_i + \Delta l_i - l_{i,j-1})^2 \}, & \tag{2.6}
 \end{aligned}$$

where $l_{i,j-1}$ is the lateral displacement of the previous line and α , β_a and β'_l are the regularization terms. Considering the cost is calculated for each RF-line separately, we have dropped the index j . Hence a_i , l_i , Δa_i and Δl_i are in fact $a_{i,j}$, $l_{i,j}$, $\Delta a_{i,j}$ and $\Delta l_{i,j}$.

In the final step, Kalman filtering [28] is used to estimate a low noise strain image from the displacement image(SD). The strain is piecewise smooth except on the boundaries of the tissues with different mechanical characteristics. This filter takes advantage of this prior and results in a strain that is smooth within the same tissue and sharp at boundaries.

2.3 Results

We test EDPOT using simulation, phantom and *in-vivo* human data. The human data is composed of RF data of the liver from patients with liver cancer. These datasets are further described in corresponding sections below.

The main part of the program, i.e. performing DP on the tree and calculating displacements, is written in C++ and used as a Matlab MEX function. The data processing is done on a 3.40 GHz Core i7 quad core computer. For a 1000×100 ultrasound image, EDPOT takes approximately 0.064 seconds to run and DP takes 0.026 seconds. Our implementation can be further optimized to reduce this time.

We have empirically chosen D_{tm} to be half the height of the ultrasound images; where a good balance is struck between overall improvement and computational complexity. In our tests, changing this parameter does not result in significant variation in the results. In an extreme case which D_{tm} is equal to the height of the image, the tree is reduced to one horizontal line and we get the same results as [28]. Also, we have chosen $dist(v_{m_2}, v_{m_3}) = dist(v_{m_1}, v_{m_2}) = 1$.

We compare the results of EDPOT with those of DP [28]. In order to measure the improvement of EDPOT over DP, a ground truth displacement field is required. As mentioned before, failure in DP primarily depends on the choice of seed-line: if shadowing artifact, large out-of-plane or lateral motion, blood vessels or cysts are present at the seed-line, DP will likely fail. Failure in DP, results in distinct errors in the displacement and strain images (Fig.

2.2), and as such, is easy to detect by visual inspection. Therefore, to generate the ground truth, we run DP on multiple seed-lines and visually select a correct strain image. We use this displacement image as a ground truth displacement estimate. Note that while this ground truth is not perfect, it provides sufficient accuracy for our purpose of finding large displacement errors (Fig. 2.2). In cases where visual inspection is not feasible (Section 2.3.4), another method calculating ground truth is used. Subsample displacement based on every possible seed-line is used and the median of these displacement field is used as the ground truth.

In the next step, for every RF-line, ID is calculated with that RF-line as the seed-line. With the ground truth at hand, we measure the error for both methods in terms of Mean Squared Error (MSE). We then report the mean and the standard deviation of the squared error for all seed-lines.

As stated in section 2.2, the impact of the regularization term on the cost function is governed by w . Thereby, we compare EDPOT and DP over a range of w . It is worth mentioning that due to the low lateral resolution of ultrasound images, we do not show lateral displacement results. Nevertheless, EDPOT estimates 2D displacement maps. Experimental results are provided below.

2.3.1 Simulation

For simulation evaluation we generate RF-data for a uniform tissue using Field II software [35, 36] with 4% strain. The MSE and the variance of the squared error are reported in

Table 2.1: The MSE and the standard deviation of the squared error for the simulation data with different noise levels. The minimum values are in bold font.

		$\sigma = 0.0$									
w		0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP		0.61 ± 2.67	0.60 ± 2.47	0.62 ± 2.55	0.63 ± 2.61	0.64 ± 2.60	0.67 ± 2.69	0.69 ± 2.78	0.71 ± 2.87	0.72 ± 2.91	
EDPOT		0.54 ± 2.14	0.52 ± 2.09	0.53 ± 2.12	0.53 ± 2.15	0.54 ± 2.19	0.55 ± 2.24	0.57 ± 2.33	0.59 ± 2.40	0.59 ± 2.43	
		$\sigma = 0.05$									
w		0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP		1.26 ± 8.03	0.91 ± 5.01	0.85 ± 4.43	0.83 ± 4.14	0.80 ± 3.84	0.81 ± 3.81	0.83 ± 3.84	0.82 ± 3.68	0.83 ± 3.71	
EDPOT		1.03 ± 6.83	0.82 ± 4.74	0.72 ± 3.63	0.69 ± 3.30	0.67 ± 3.03	0.66 ± 2.93	0.67 ± 2.92	0.66 ± 2.82	0.67 ± 2.84	
		$\sigma = 0.10$									
w		0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP		8.86 ± 53.78	4.38 ± 25.20	2.65 ± 16.64	2.05 ± 13.13	1.73 ± 11.01	1.74 ± 11.79	1.70 ± 11.84	1.70 ± 11.71	1.67 ± 11.15	
EDPOT		4.53 ± 24.19	2.81 ± 19.78	1.84 ± 11.23	1.49 ± 9.31	1.27 ± 7.40	1.24 ± 8.56	1.18 ± 6.85	1.17 ± 6.96	1.17 ± 6.85	
		$\sigma = 0.15$									
w		0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP		95.06 ± 435.93	45.51 ± 235.40	27.34 ± 163.93	18.28 ± 114.43	10.73 ± 80.10	7.13 ± 40.11	7.12 ± 44.48	7.35 ± 46.38	7.43 ± 47.52	
EDPOT		33.15 ± 206.17	19.17 ± 132.24	11.94 ± 72.41	7.47 ± 54.14	6.63 ± 52.31	5.83 ± 51.17	4.50 ± 44.02	3.31 ± 23.95	3.03 ± 22.73	

Fig. 2.5 and Table 2.1. In order to measure the robustness of our method against signal decorrelation, we also add Gaussian noise with σ in the range of 0 to 0.15 with 0.05 increments to the RF data. The results of Table 2.1 and Fig. 2.5 show that EDPOT consistently outperforms DP.

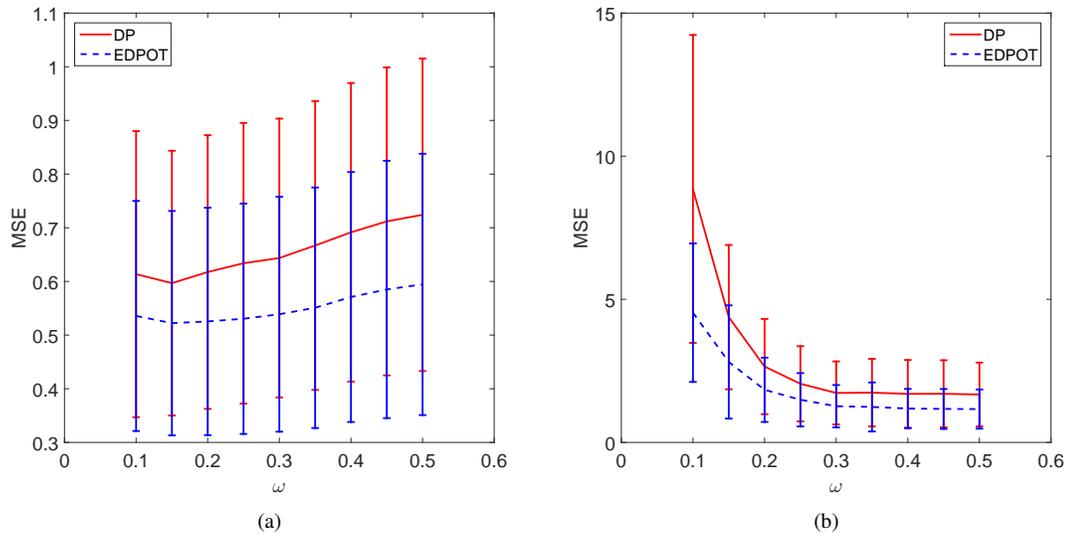


Figure 2.5: The MSE error for the simulation data. The figures also show the standard deviation of the error, divided by 10. (a) shows the result without noise and (b) with noise $\sigma = 0.1$

2.3.2 Phantom Experiments

The phantom data was acquired from a CIRS (Norfolk, VA) breast phantom. The data was collected with an Antares Siemens system (Issaquah, WA) at a center frequency of 6.67 MHz using A VF10-5 linear array with a 40MHz sampling rate. A B-mode image and a strain sample of the phantom data can be seen in Fig. 2.6. The MSE and standard deviation for a range of w is also depicted in Fig. 2.6, and the numerical values are reported in Table 2.2. Compared to DP, EDPOT gives substantially lower MSE over the entire range of w . In addition, EDPOT has much smaller variance, which means that it consistently estimates the correct displacement field.

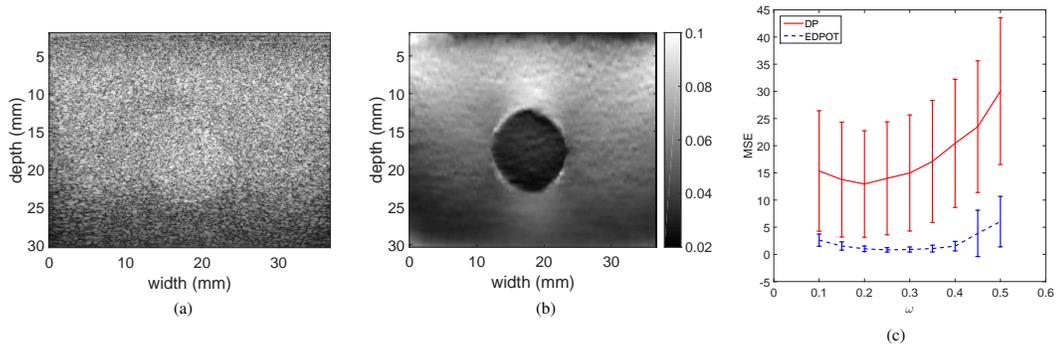


Figure 2.6: Results of the phantom experiment. (a) shows the B-mode ultrasound image of the Phantom. (b) shows the axial strain where the DP method has not failed and in (c), the MSE of DP and EDPOT are compared. $\sigma/10$ is used in error bars to ease comparison.

2.3.3 Patients With Liver Cancer

The data was collected from two patients with primary or secondary liver cancer who underwent open surgical radio-frequency thermal ablation. Data collection was performed at

Johns Hopkins Hospital and was approved by its ethics board. These patients had unresectable disease and were recommended for RF ablation after review from Johns Hopkins University multidisciplinary conference. The RF data was acquired from an Antares Siemens system (Issaquah, WA) at the center frequency of 6.67 MHz with a VF10-5 linear array at a sampling rate of 40 MHz. Further details of the data acquisition are available in [28]. B-mode images, strain images without any artifact and with artifact for Patient 1 and Patient 2 are depicted in Fig. 2.7. It is clear that EDPOT substantially outperforms DP, and can prevent large artifacts in the strain image.

Table 2.2: The MSE and the standard deviation of the squared error for the Phantom, Patient 1, Patient 2 and Patellar Tendon data.

Phantom										
w	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP	15.35 ± 110.81	13.77 ± 105.62	12.95 ± 97.90	14.00 ± 103.89	14.98 ± 106.77	17.11 ± 112.36	20.43 ± 117.92	23.51 ± 121.08	30.03 ± 134.84	
EDPOT	2.62 ± 11.14	1.54 ± 7.53	1.03 ± 5.19	0.83 ± 4.02	0.90 ± 4.60	1.07 ± 6.20	1.52 ± 8.72	3.87 ± 42.60	6.03 ± 46.39	
Patient 1										
w	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP	0.49 ± 3.37	0.43 ± 1.99	0.40 ± 1.75	0.40 ± 1.76	0.41 ± 1.80	0.42 ± 1.85	0.42 ± 1.89	0.43 ± 1.90	0.43 ± 1.91	
EDPOT	0.31 ± 1.14	0.29 ± 1.11	0.28 ± 1.11	0.26 ± 0.87	0.26 ± 0.88	0.26 ± 0.87	0.26 ± 0.88	0.26 ± 0.87	0.25 ± 0.85	
Patient 2										
w	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP	4.35 ± 11.23	4.84 ± 11.58	5.52 ± 12.61	5.95 ± 13.00	6.39 ± 13.44	6.67 ± 13.57	7.37 ± 14.66	8.11 ± 15.82	9.18 ± 17.84	
EDPOT	4.26 ± 11.71	4.54 ± 11.96	4.77 ± 12.38	4.95 ± 12.62	5.27 ± 13.33	5.29 ± 13.26	5.62 ± 13.93	5.98 ± 14.73	6.93 ± 20.10	
Patellar Tendon										
w	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	
DP	4.95 ± 70.33	4.34 ± 55.61	4.40 ± 60.07	3.94 ± 55.57	3.86 ± 52.45	3.77 ± 50.35	3.74 ± 50.23	3.57 ± 49.01	3.56 ± 48.84	
EDPOT	4.64 ± 63.13	3.66 ± 39.93	3.65 ± 38.27	3.40 ± 42.07	3.16 ± 38.36	3.17 ± 41.59	3.23 ± 42.60	3.37 ± 46.28	3.14 ± 43.43	

2.3.4 Patellar Tendon

This set of data was collected at the PERFORM Centre at Concordia University. The in-vivo patellar tendon data was collected at using an Alpinion E-Cube (Bothell, WA) ultrasound machine with an L3-12 linear transducer at the centre frequency of 11MHz and sampling

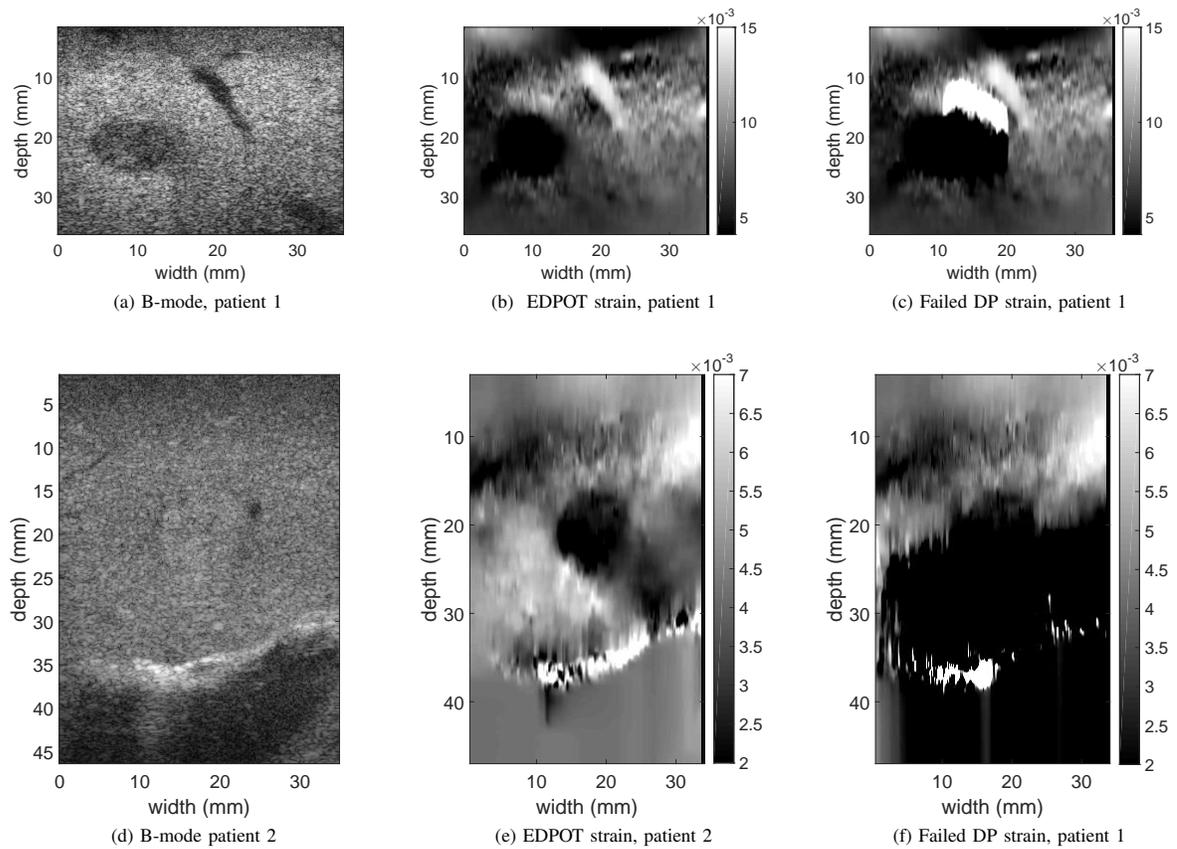


Figure 2.7: *In-vivo* images of human data. (a) and (d) show the B-mode ultrasound images of patient 1 and patient 2 respectively. (b) and (e) show the axial strains with EDPOT. (c) and (f) show cases where DP has failed.

frequency of 40MHz. The subjects were asked to flex their knees and their patellar tendon was imaged during this isometric flexion.

2.4 Discussions and Conclusion

Herein we focused on optimization-based methods of displacement estimation, which require an ID field. However, a reliable ID is also very useful for correlation-based methods to

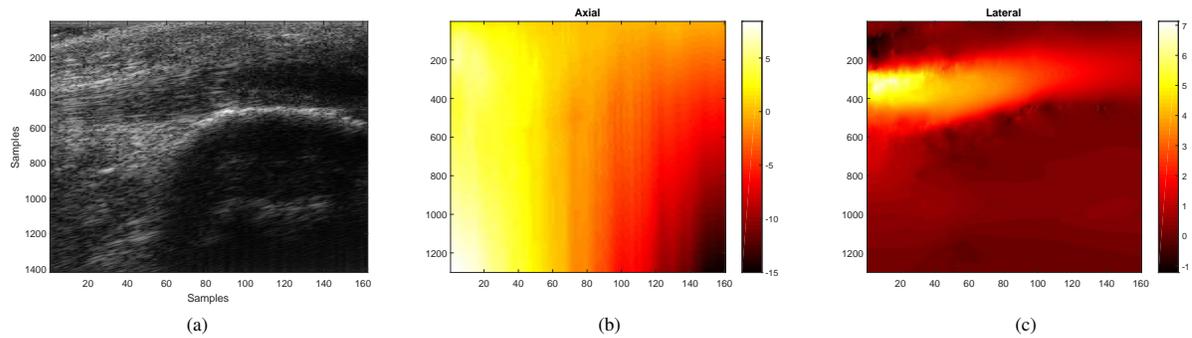


Figure 2.8: Figure (a) shows the B-mode of the patellar tendon which itself can be seen in the upper half of the image. Figure (b) and (c) show the axial and lateral displacement during isometric contraction. The tendon in (c) moves laterally by a large amount as expected.

both limit the search range and reduce the chance of peak-hopping. As such, EDPOT can be utilized in a wide range of displacement estimation techniques such as phase- and amplitude-based cross-correlation methods, which is a subject of future work.

Another topic for future work is to analyze which paths on the tree graph were chosen during ID calculation. This would help in better understanding how much efficiency is gained by using this method. We can also choose the best regularization coefficient (w) by utilizing the same approach mentioned in Section 3.3 for calculating ground truth. As for other ways of calculating ground truth, optical flow can be an independent method for this purpose. Although not real-time, it can be used for verification among other applications. Moreover, a use of calibrated phantom can be investigated which could provide a known ground truth.

In this work we proposed a new method wherein the ID is calculated for pixels on a tree, contrary to previous work which utilized only a single vertical line. This resulted in exploiting more information and thus improved the robustness of the ID. We validated the proposed

method using simulation, phantom and *in-vivo* human data, and showed that it substantially outperforms conventional DP.

Chapter 3

Assessment of Rigid Registration Quality

Measures in Ultrasound-Guided

Radiotherapy

3.1 Introduction

Accurately targeting the pathological loci during radiotherapy is crucial to ensure the treatment outcomes. However, patient motions limit the precision with which radiation can be applied, resulting in less effective treatment plans. In modern radiotherapy, image guidance is used to align and update the patient's anatomy with the treatment isocenter, providing better target coverage and in some cases reducing dose to surrounding healthy tissues. Such alignment

(i.e., patient positioning) can be achieved through widely used image registration algorithms based on a number of techniques, including external surface motion, implanted markers, X-ray imaging, and ultrasound imaging [37–39]. Compared with X-ray imaging, ultrasound is non-ionizing and provides good soft tissue contrast in real time [40], and thus it has become a popular imaging modality to track patient motions.

Radiotherapy frequently involves the delivery of radiation dose in multiple sessions, known as fractions. Two types of patient motions can occur, including interfraction motion (i.e., on each day of treatment, as the patient is positioned for that day), and intrafraction motion (i.e., short term during radiation delivery). Interfraction positioning affects the entire treatment fraction. Although it must be completed reasonably quickly, more time is available for calculation and review. However, intrafraction positioning, or monitoring, must be completed in near real time to be of use. An operator often has to rapidly verify the positioning quality during the entire duration of the treatment, which is challenging due to time limitations and 3D nature of the images. To help ensure the quality of patient positioning and mitigate the workload of the operator, who may not offer consistent quality assurance, a robust automatic method for assessing image registration quality is needed.

Most of the previous work in quality evaluation can be broadly categorized into Bayesian and supervised learning methods [41]. Typically in the former, a Bayesian framework for the registration problem is proposed and a posterior distribution over the model parameters is calculated. Next, using the posterior, a measure of uncertainty is given. For instance,

Risholm et al. [42] proposed a Bayesian non-rigid registration framework using Boltzmann's distribution for the prior and likelihood and Markov Chain Monte Carlo (MCMC) to estimate the most likely deformation and the uncertainty associated with it. Janoos et al. [43] proposed a similar framework which is used for image registration for multi-modal images. In [44], the authors introduce ways to summarize the uncertainty of an elastic registration framework which they proposed in [45]. Simpson et al. [46] try to solve the problem of choosing the regularization coefficient with a Bayesian approach which also can estimate the uncertainty in the form of a covariance matrix. As shown in [47], the uncertainty can also be used to construct a filter to smooth the areas with higher uncertainty.

A supervised classification method was first introduced by Wu and Samant [48] for automatic detection of unsuccessful registrations during radiotherapy. The authors used one feature (e.g. mutual information or cross correlation) as an input for the classifier and the classifier itself, used a threshold calculated based on the training data to classify different registrations. Wu and Murphy [49] then improved their previous work by extracting more features and also using a neural network as the classifier. Muenzing et al. [50] did a comprehensive study of different features and classifiers that could be of use for the task at hand and evaluated their method on lung CT images. Finally, Sokooti et al. [51] constructed a random regression forest to estimate the registration error of chest CT scans and also classify based on the estimated error.

An advantage of learning methods over the Bayesian approach is the lower computational complexity for classifying the registrations at runtime. This advantage makes these methods suitable for real-time applications. However, to train such a classifier, an appropriately sized training set is needed and acquiring such training set is not always feasible. An unsupervised method may prove to be useful in such cases.

In [52], the author has taken a frequentist approach to measuring uncertainty using bootstrapping. It is assumed that the input images are realizations of random processes. Given several realizations of the input images, the registration method could be run on these images and the uncertainty can be calculated based on the results of these registrations. Since only one realization of the random variable, the image at hand, is available, bootstrapping is used to simulate different realizations. This method does not require any training which makes it an attractive candidate for registration quality evaluation that can be readily applied to different ultrasound systems and even other modalities. We will therefore propose a technique for assessing the quality of ultrasound registration using bootstrapping, and validate it using phantom and *in-vivo* data.

In this work, we propose to use bootstrapping and supervised learning methods for assessing the quality of rigid ultrasound image registration in the context of ultrasound-guided

radiotherapy. More specifically for supervised learning methods, we employed Linear Discriminant Analysis (LDA) [53] and Random Forest (RF) [54] to classify the registration quality. All methods were compared using both phantom and *in-vivo* data for intrafractional prostate motion management. In this work, we have made three major novel contributions to the field. First, to the best of our knowledge, this is the first work that introduces automatic registration assessment techniques for ultrasound-guided radiotherapy, and more generally for registration of ultrasound images. Second, in the context of machine learning techniques, we introduced new features due to the unique characteristics of ultrasound images. Lastly, we compared the performance of bootstrap and machine learning techniques for the application, which has not been reported previously. Given that ultrasound has numerous applications in image-guided applications, this work can be further extended and utilized in several other applications. This chapter is organized as follows. In the next section the methodology is explained. In Section 3.3, the results are presented and are discussed in Section 3.4. The conclusions are provided in the final section.

3.2 Methods

3.2.1 Registration

Assume $f, g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ to be the fixed and moving images. Also, let $\Omega \in \mathbb{R}^m$ be a set of points from the domain of f . We aim to find a transformation, $T(x, \theta) : \mathbb{R}^m \rightarrow \mathbb{R}^m$,

with $\theta \in \Theta \subseteq \mathbb{R}^d$, such that $f(x)$ corresponds to $g(T(x, \theta))$. To calculate $\hat{\theta}$, the transform parameters, a cost function, $J(\theta)$, is constructed and $\hat{\theta}$ is estimated by:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} J(\theta) \quad (3.1)$$

$$J(\theta) = D(f(x), g(T(x, \theta))), \quad (3.2)$$

where $D(\cdot)$ is the dissimilarity function.

Both f and g can be considered outcomes of random processes and therefore $J(\theta)$ is a random process and $\hat{\theta}$ is a random variable.

In order to evaluate the registration results, it is necessary to measure how close these results are to the true value. A popular approach is to use mean Target Registration Error (mTRE) [55–57]. Since a rigid transformation model is used in our work, mTRE is calculated on 4 or 6 points for 2D and 3D data respectively. We define the distance between two transforms, T_1 and T_2 , as follows. Let $\{P_i\}$ be a set of N points in the fixed image near the center of the transformation, C , and the center itself. The points are selected by moving r millimeters away from C in each cardinal direction; therefore 4 points in the 2D images and 6 in 3D volumes are chosen. The distance is then defined as:

$$d(T_1, T_2) = \frac{1}{N} \sum_{i=1}^N \|T(P_i, \theta_1), T(P_i, \theta_2)\| \quad (3.3)$$

In other words, the distance between two transformations is the mean distance of the corresponding transformed points. Calculating the distance as explained, instead of doing so on a grid, reduces the computational complexity of the evaluation while keeping the evaluation valid because of the rigidness of the transform. Also by using the 4 or 6 point distance measurement method, the comparison between ROIs with different sizes will be equivalent.

Before we present the supervised learning and bootstrapping techniques, it is important to clearly state what is called a “successful” registration or “poor” registration. During registration, the optimizer either converges to an optimum or not. If it diverges, the result is a poor registration. If it converges, but converges to a local optima which is far from the true parameters, the result is again a poor registration. A successful registration is one that the optimizer converges to the correct optimum.

3.2.2 Data Preparation

To validate the registration assessment methods, a great number of both cases of poor and successful registrations were needed. Both phantom and *in vivo* patient data were utilized for validation. The following procedure was used to obtain poor and successful registrations. First, a reference registration was carried out to be used as the true registration. For the phantom data, this registration was known *a priori* with a robotic system. For the 3D patient data, as each session represented a tracking sequence, each sequence was made into a video showing anterior-posterior (A-P) and superior-inferior (S-I) cuts through the center of the original

prostate position. These videos were visually inspected by experts experienced in prostate radiotherapy to ensure that the reference registration was of high quality. To further evaluate the automatic registration quality for the ground truths, we selected 2 image pairs from each of the 7 treatment sessions for the patient, and asked an expert to manually align the image pairs based on visual inspection. In addition, for each image pair, 10 pairs of homologous anatomical landmarks were selected, and the mean target registration error (mTRE) was obtained with these landmarks for both manual and automatic registrations. The mTREs (mean \pm sd) from 14 image pairs for manual and automatic registrations are 1.91 ± 0.83 mm and 1.86 ± 1.09 mm, respectively. The difference in the results obtained from the two approaches is not statistically significant based on a Wilcoxon signed-rank test ($p = 0.358$).

Next, the true transform parameters were moved in the parameters space in a random direction and the registration was restarted from that point. If the result of the new registration was within a determined distance defined by Eq.3.3 (i.e., the smallest resolution of the images), the registration result was regarded as "good", and the initial parameters are moved further away from the true registration result. This was repeated until the new registration either diverges or converges to another point far from the true result, hence generating a "bad" registration. Instead of changing the registration parameters with equal step sizes along a direction in the registration parameter space, the parameter steps for each new starting point was defined as an increase of 2 mm by Eq.3.3. This way, the interpretation was more intuitive as the metric was the same as the measurement of the image resolution. Furthermore, these

incremented registrations were selectively inspected by clinical experts who are experienced in prostate radiotherapy. As such, a set of successful and poor registrations were generated from an initial limited set of inspected, good registrations. This procedure is depicted in Fig. 3.1, and instances of good and failed registration results for the patient data are demonstrated in Fig. 3.2. Here, the moving image was moved in the parameter space until a failed registration occurred while the fixed image was kept the same. The successful registration visibly improved the alignment of the walls of the bladder and prostate.

3.2.3 Supervised Learning Methods

There are numerous classifiers available in the literature; we chose two for our experiments: LDA [53], a simple classifier, and RF [54] as a state-of-the-art classifier. As with other supervised learning methods, this requires feature extraction, training and validation.

Feature Extraction

There may be a trade-off between calculation time and discriminative value of a feature. The ideal feature would cost no additional calculation. We selected a subset of 10 features from a pool of features for training and classification. This selection was done based on feature importances (Gini importance [54]) resulting from an RF classifier using all the features.

The registration is implemented by optimizing the negative Normalized Cross Correlation (NCC) between a selected set of pixels in the reference and target images. The resulting

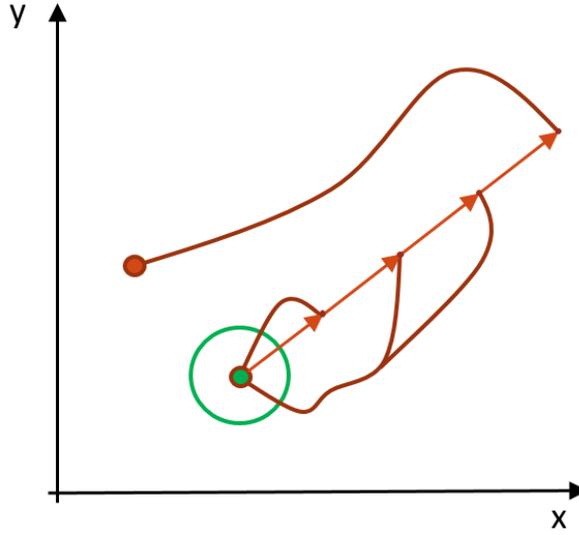
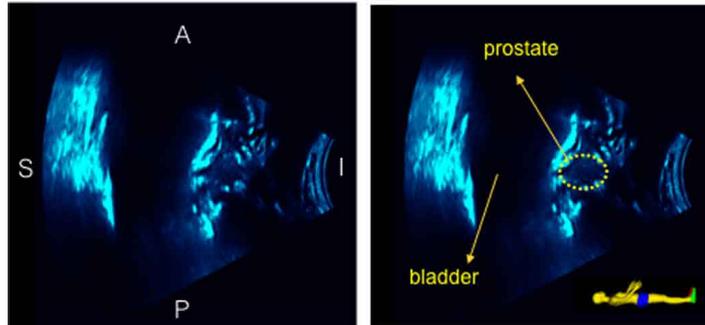


Figure 3.1: Generating poor and successful registrations. The green dot shows parameters of the correct registration. Each arrow shows the start of a new registration process. In this schematic example, three registrations converge to the correct result, and one converges to an incorrect result (red dot). The green circle shows the area wherein the registration is still considered successful.

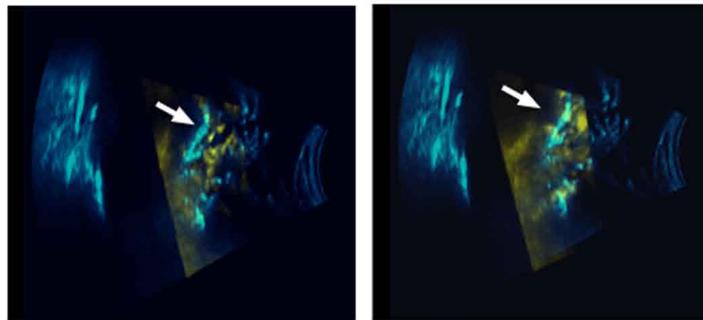
optimal NCC can be used as a criterion for distinguishing between successful and poor registrations. This measure costs no additional computation, as we are already computing it.

Let $f_i = f(x_i)$ and $g_i = g(T(x_i, \hat{\theta}))$ represent the fixed and moving image intensities where $\{x_i\}$ is the set of points used to calculate the NCC and $T(x, \theta) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the transformation. N is the number of pixels (i.e. the number of points in $\{x_i\}$). The NCC can be calculated in a single pass over the image using:

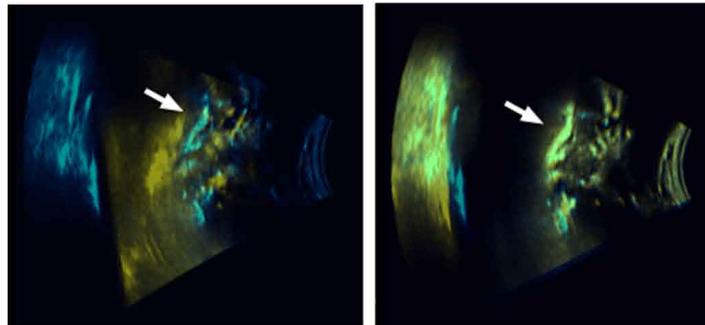
$$NCC = \frac{S_{fm} - S_f \cdot S_g / N}{C_f \cdot C_g} \quad (3.4)$$



(a) Fixed image and the annotation



(b) An example of a bad registration



(c) An example of a good registration

Figure 3.2: Demonstration of good and bad registration results for the patient US data. The fixed image (cyan) and moving images (yellow) are overlaid to show the quality of registration. (a) shows the fixed image along with the anatomical annotation and the orientation of the image with respect to the patient. (b) shows a failed registration (left: before registration; right: after registration). (c) shows a case of successful registration (left: before registration; right: after registration). Here, the white arrows point to the wall of the bladder.

where

$$\begin{aligned}
 S_f &= \sum_i f_i & S_g &= \sum_i g_i \\
 S_{ff} &= \sum_i f_i f_i & S_{fg} &= \sum_i f_i g_i \\
 S_{gg} &= \sum_i g_i g_i
 \end{aligned}$$

These sums can be accumulated during a loop over the pixels. Note that S_{ff} , S_f and N are not necessarily constant, as some of the reference pixels may map outside of the moving image and will therefore be excluded from the calculation. From these, we can compute the contrast of each image, C_x , using:

$$C_x = \sqrt{S_{xx} - S_x \cdot S_x / N} \quad (3.5)$$

and the NCC using Equation 3.4. An advantage of calculating the NCC this way is that each part can be used as a feature. It was conceivable that one or more of these measures were more distinctive than correlation alone. There is no additional cost to these measures as they are already computed.

The Distinctiveness of Optimum (DO) [58] was used together with Mirror Symmetry (MS) in [49]. It is an average descriptor of the shape of the dissimilarity function around the found solution. It requires $2U$ evaluations of the dissimilarity measure over the registration

cost function J with respect to each registration parameter in the positive and negative directions. Here, U is the number of registration parameters, and $U = 6$ for rigid registration. Therefore, the DO metric is defined as:

$$DO(s) = \frac{1}{2sU} \sum_u \left(\frac{J(\hat{\theta} - se_u) + J(\hat{\theta} + se_u)}{2} - J(\hat{\theta}) \right) \quad (3.6)$$

where s is the step size, $\hat{\theta}$ are the optimal parameters, J is the cost function to be optimized for registration and e_u is a unit parameter vector in direction u .

The Mirror Symmetry (MS) [49], [59] is a measure of the evenness of the shape of the similarity function around the found solution. Letting

$$\bar{J}_u = \frac{J(\hat{\theta} - se_u) + J(\hat{\theta} + se_u)}{2}, \quad (3.7)$$

MS can be calculated as:

$$MS(s) = \frac{1}{2P} \sum_u \left(\frac{(J(\hat{\theta} - se_u) - \bar{J}_u)^2 + (J(\hat{\theta} + se_u) - \bar{J}_u)^2}{(J(\hat{\theta}) - \bar{J}_u)^2} \right). \quad (3.8)$$

It can be generated from the same samples as the distinctiveness of optimum. It's worth mentioning that MS may not be an optimal feature to detect some local optimums such as saddle points.

An indication of a good registration is that the correlation score at one step size away in any direction from the found location is significantly worse. Therefore, we also include individual cost evaluations as features. For the convenience of annotation in the later sections, we name these evaluations as

$$\{DissimProbes[2k], DissimProbes[2k + 1]\}$$

where

$$DissimProbes[2k] = J(\hat{\theta} + se_u)$$

$$DissimProbes[2k + 1] = J(\hat{\theta} - se_u).$$

and

$$k = 0, 1, \dots, U - 1.$$

With $k = \{0, 1, 2, 3, 4, 5\}$ corresponding to the probing of J in the direction of each of the transformation parameters (3 rotations and 3 translations), we obtained the evaluations as $DissimProbes[0]$ to $DissimProbes[12]$. Here, $DissimProbes[.]$ is short for “Dissimilarity Probes”.

In a successful registration, it is expected for all the pixels in the ROI to be registered equally well. To quantify this, the ROI is divided into orthants and the correlation score

is calculated for each. In a poor registration, the correlation score varies between the orthants. Therefore, several measures of quality can be considered regarding this. The individual orthant scores (OrthantScores), the maximum and minimum score (MaxOrthantScores and MinOrthantScores) and finally, the score difference between the maximum and minimum (OrthantScoreRange).

Instead of treating the two intensity distributions as identical, we can instead examine the joint distribution of intensity. The Mutual Information (MI) of this joint intensity distribution is a commonly used similarity measure. We did not use it to construct the cost function because our work is focused on mono-modal registration. In addition, MI costs more to compute than the above, as it involves keeping track of a joint distribution.

If the images are correctly aligned, it is reasonable to presume that the corresponding set of pixels in the fixed and the moving images have similar intensity distribution. The Kullback-Leibler divergence [60] can be used to quantify the difference between the two distribution functions and therefore is used as a feature.

Training and validation

As mentioned before, we used LDA and RF to classify the registrations. Here, half of the total data were used as a testing set, and the other half was used to train the classifiers through a 4-fold cross-validation process (training set vs. validation set ratio = 3:1). The machine

learning algorithms and validations were implemented in scikit-learn package, version 0.17 [61].

3.2.4 Bootstrapping

Bootstrap resampling is a technique that can be used to estimate the properties of an estimator, such as mean, variance, etc. [62]. Assume a random variable X with N i.i.d samples $X = \{x_1, \dots, x_n\}$ drawn from it. A bootstrap resample, $X^{(b)}$, is a multiset constructed by selecting N points from X with replacement. This is repeated B times, thus leading to B multisets: $X^{(b)}, b = 1, \dots, B$.

Assume a statistic on X , ϑ , and its estimator $\hat{\vartheta} \approx \varphi(X)$. Our goal is to measure the reliability of this estimator. This can be done by finding the estimates of ϑ based on each bootstrap: $\hat{\vartheta}^{(b)} = \varphi(X^{(b)})$. These bootstrap values can be used to form a non parametric distribution on the estimates which can be used to express a measure of reliability, such as the covariance matrix.

3.2.5 Bootstrapping for registration evaluation

Image registration can be thought of as an estimator of the transformation parameters, θ . Therefore we can use bootstrapping to measure the reliability of this estimator similar to what was explained in the previous section [52]. In our case, we use the result of bootstrapping to classify the registration as reliable and unreliable.

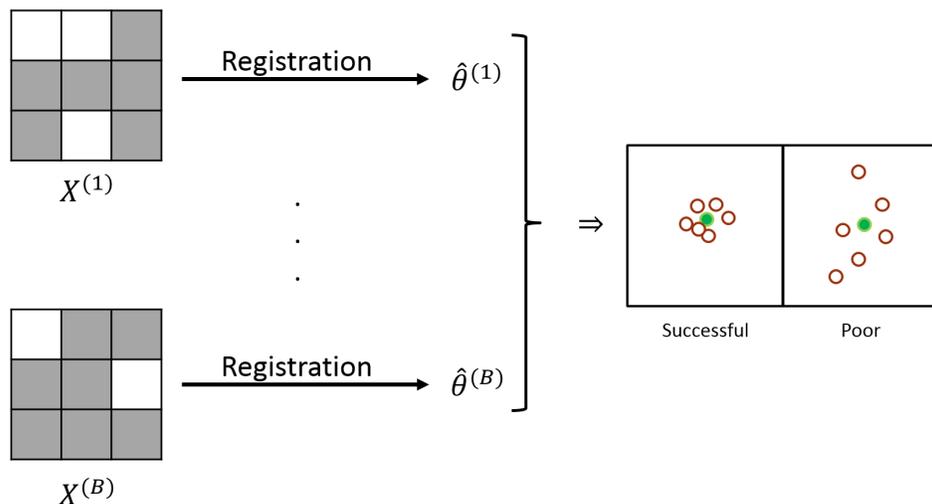


Figure 3.3: An overview of bootstrapping for registration evaluation. $X^{(b)}$ shows different multisets, and $\hat{\theta}^{(b)}$ denotes the results of the registration using each multiset. The grayed out pixels on the left show selected pixels for registration (possibly more than once). The green dots show the correct registration parameters, and the red dots represent registration results (i.e. $\hat{\theta}^{(b)}$). In a poor registration, $\hat{\theta}^{(b)}$ are expected to be more dispersed than a successful case.

To this end, it is needed to solve B registration problems, based on each bootstrap:

$$J^{(b)}(\boldsymbol{\theta}) = D(f(\mathbf{x}), g(T(\mathbf{x}, \boldsymbol{\theta}))), \mathbf{x} \in \Omega^{(b)} \quad (3.9)$$

$$\hat{\theta}^{(b)} = \underset{\theta}{\operatorname{argmin}} J^{(b)}(\theta) \quad (3.10)$$

where $\Omega^{(b)}$ is a bootstrap resample. From here, the measure of dispersion on $\hat{\theta}^{(b)}$ can be calculated.

$$\bar{d}_B = \frac{1}{B} \sum_{b=1}^B d(T(\mathbf{x}, \theta^{(b)}), T(\mathbf{x}, \bar{\theta}_B)) \quad (3.11)$$

where $\bar{\theta}_B$ is the mean of parameters resulting from bootstraps: $\text{mean } \theta^{(b)}$. In order to exclude outliers from the calculations, the trimmed mean [52] is used: the furthest 10% of the results from the mean bootstrap result are taken out and the mean is recalculated accordingly. If this dispersion measure is higher than a threshold τ , then the registration is poor and if not, successful. In other words, if we are not able to estimate the registration parameters with sufficient confidence through the sampling process, then a single registration is likely to provide a bad image alignment that is far off the optimum. Note that for each bootstrap sampling, only a portion of the pixels/voxels were randomly selected for registration. To facilitate easier interpretation of the dispersion measurement, instead of measuring the metric in the registration parameter space, we employed Eq3.3 to evaluate the transform distance. This way, the mean transform distance is in the same spacing unit as the images or volumes, the threshold can be set based on the resolution of the data and according to what accuracy is needed.

Figure 3.3 shows a general overview of the bootstrapping scheme for classifying registrations and Algorithm 1 describes an in depth implementation.

3.2.6 Experimental setup

We compare the two approaches using experiment data and patient data. For the experimental data, 2D images were acquired with a Clarity (Anticosti Research Version, Elekta AB, Stockholm, Sweden) monitoring system with a linear ultrasound probe. The patient data was collected with the Clarity system (Version 3.0) with a wobbler probe, providing a sequence of

Algorithm 1 Bootstrap resampling for image registration quality evaluation.

```
1: for  $b=1$  to  $B$  do
2:    $S \leftarrow$  empty multiset
3:   for  $i=1$  to  $N$  do
4:      $\Omega^{(b)} \leftarrow \Omega^{(b)} \cup \{x_k\}; k \sim_{i.i.d} \{1, \dots, N\}$ 
5:   end for
6:    $J^{(b)}(\theta) = D(f(x), g'(x)), x \in \Omega^{(b)}$ 
7:    $\hat{\theta}^{(b)} \leftarrow \underset{\theta}{\operatorname{argmin}} J^{(b)}(\theta)$ 
8: end for
9: Calculate  $\bar{d}_B$  from  $\{\hat{\theta}^{(b)}; b = 1 \dots N\}$ 
10: if  $\bar{d}_B > \tau$  then
11:   poor registration
12: else
13:   successful registration
14: end if
```

2D images in a sweeping pattern and thus forming a 3D volume. The 2D phantom data were collected using translational motions by a robotic arm. With better controlled ground truths, this approach is ideal for preliminarily testing the proposed techniques. Then, we further validated the methods with 3D patient data under full rigid body motions in order to reveal their performance for potential real clinical applications.

3.2.7 Phantom study

We imaged a Clarity QC phantom (Elekta AB), with the ultrasound probe attached to a Cartesian gantry robot (Velmex, Inc. Bloomfield, NY, USA) to control the probe movements. A laser level was used to set the orientation of the probe so that a) the image plane and the motion plane of the probe would be parallel and b) the probe would be perpendicular to the surface of the phantom. The former is to minimize any movement not on the image plane and

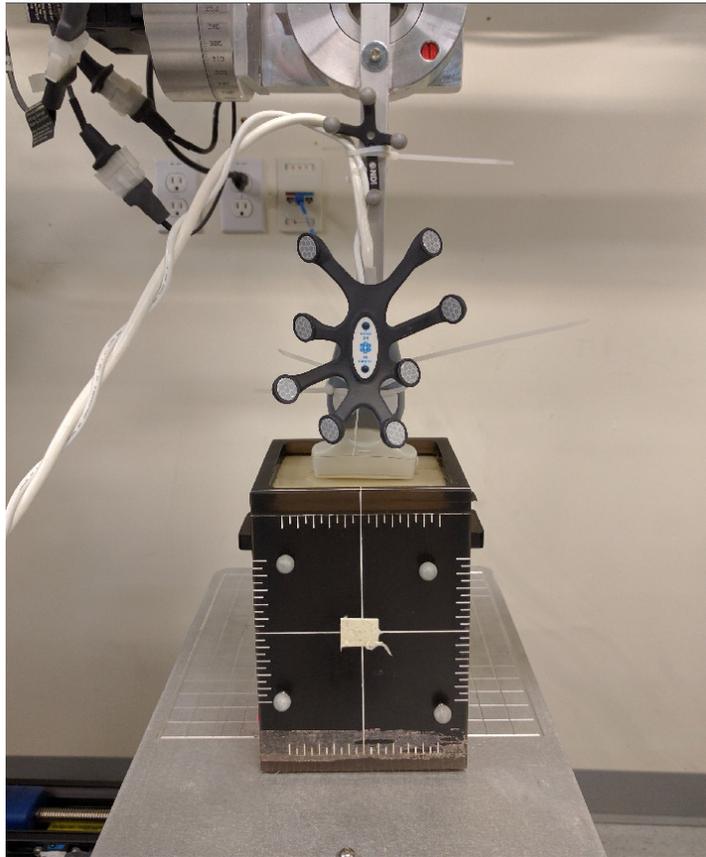


Figure 3.4: The experiment setup. The phantom, ultrasound probe, the tracked markers and the robot can be seen.

the latter, to assert only translation in one dimension of the images. The probe was also tracked with a Polaris Spectra optical tracking system (Northern Digital Inc, Waterloo, ON, Canada); this was used so that the true probe translation would be available (within the precision of the tracker). Moreover, the tracking information was used to ensure correct movement of the probe. Figure 3.4 shows the experimental setup.

The following procedure was used to acquire the images. First between 15 to 20 frames were captured. The probe was moved then in the lateral direction for 10, 15, or 20 mm. After

each translation, another 15 to 20 frames were captured and the two image sets, from before and after the translation, were registered.

This was repeated for 8 different runs. Between runs, the amount of probe movement, the settings of the ultrasound machine, the part of the phantom which was imaged and the medium (gel or water) were changed to produce a wide range of images with different qualities. To have more variety in registrations, image sets from different runs were also registered. These image sets were chosen so that they would be images from the same structure in the phantom, with the same orientation of the probe. The difference between them being the settings of the ultrasound machine. Good and bad registrations were generated with the procedure described in Section II.B. As a result, 1688 sets of registrations were used for supervised learning methods, and 3376 sets for testing bootstrapping. The good vs. bad ratio is about 4:1.

For the supervised learning methods, features were extracted and different classifiers were trained and evaluated. Bootstrapping was also carried out, with 20 bootstrap resamples and $\tau = 0.14$ mm. τ was chosen based on the pixel size, which was 0.14 mm in this experiment.

3.2.8 Patient Trials

For the experiment, ultrasound data were collected from one patient acquired during a previously scheduled and planned radiotherapy treatment session for the prostate [63]. The data were acquired using the same scanner mentioned in the previous section, and included

7 separate treatment sessions to help increase the variability among the images. Imaging in each session lasted about 4-10 min. The patient images were acquired in the context of an Institutional Review Board (IRB) approved clinical study, and were not used to make clinical decisions. The patients were undergoing radiation treatment and as such had bladders comfortably full and rectums empty, increasing internal patient anatomy uniformity with radiation planning CT studies. During each session, the patient was positioned supine, legs akimbo, with the probe imaging via the perineum. In this scan position, the prostate can be imaged between the pelvic bones. The probe position was adjusted to obtain a good image of the prostate with, and fixed in place. The patient was instructed not to voluntarily move during the procedure. The probe continuously sweeps the image plane, forming a continuously updated 3D dataset, and a total of 2193 images were acquired from all the sessions. Intrafractional target tracking is performed by registering the current 3D dataset to the first reference dataset and the quality of registration was visually inspected by a clinical expert. Using the same procedure, we generated a set of good and bad registrations, and used them to compare the learning methods against bootstrapping. For bootstrapping, 43328 registrations were used, and for supervised learning, 21664 were used. The ratio between good and bad registrations is about 4:1. Since the resolution of the volumes were different from that of the experiment images, the threshold, τ was set to 0.4 mm, which is the axial voxel resolution of the volumes.

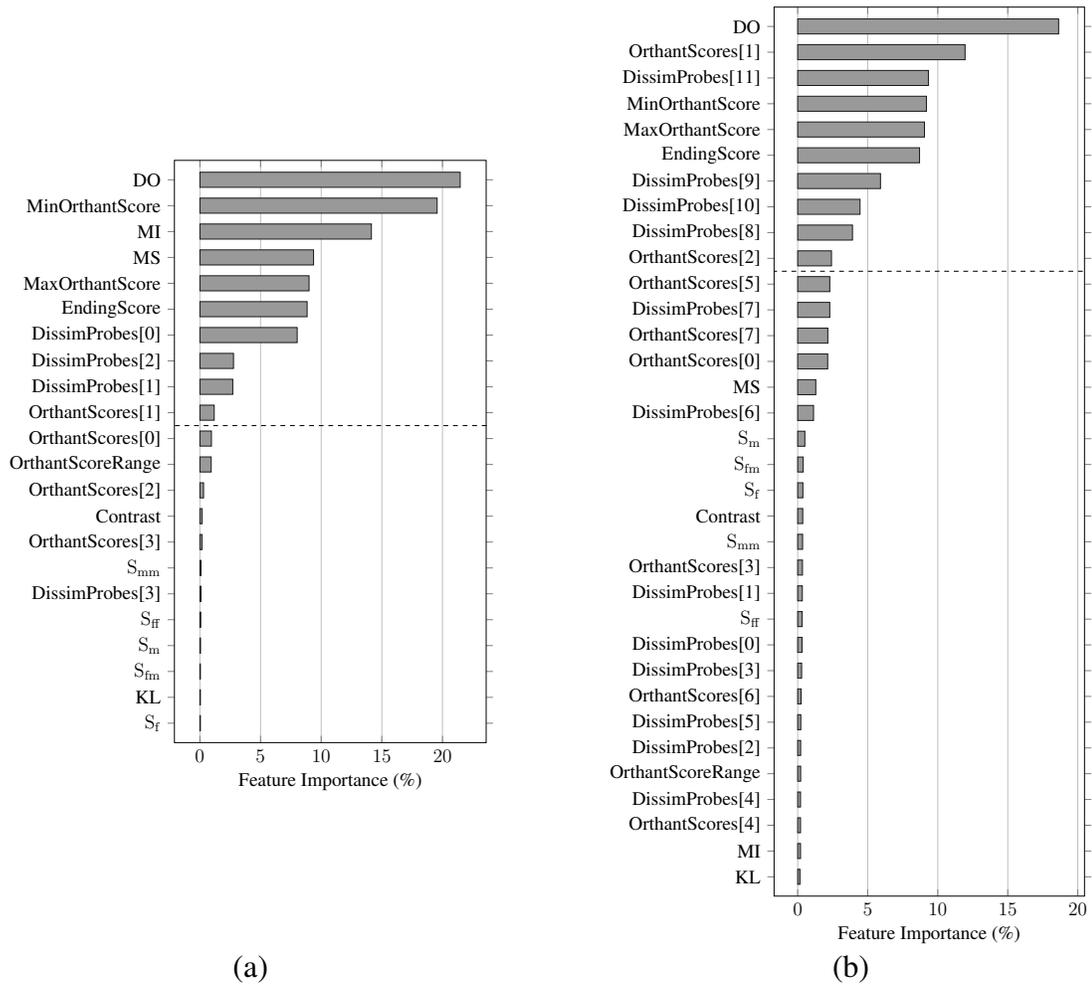


Figure 3.5: Feature importance according to RF. (a) shows the feature importance for the 2D experimental data and (b) for that of the patient data. The features above the dashed lines are those chosen for classification.

Registration was performed using the Insight Toolkit (ITK) [64] parametric registration framework by optimizing the Normalized Cross Correlation (NCC) computed over the selected pixels using gradient descent (specifically the regular step gradient descent optimizer in ITK was used) and linear interpolation. In the 3D case, the images are collected on a

fan-shaped geometry. To avoid unnecessary interpolation errors the images were directly registered as originally sampled (i.e., in the fan shape) using the methodology described in [65]. For the 2D experiment, 35% of the pixels in the ROI were randomly selected to build the cost function. Also, a translation transformation model was used since the movement was in one direction. For the 3D data, 20,000 voxels were used to construct the cost function. As for the transformation model, one was used to accommodate the non-rectilinear image frames which is the result of using a wobbler probe [65]. The supervised learning methods were implemented in Python using the scikit-learn library [61].

3.3 Results

3.3.1 Feature selection

The feature importances for the 2D and 3D data can be seen in Figures 3.5(a) and 3.5(b) respectively, and the features shown in the figures above the dashed lines are selected to be used for the classifiers. Although DO was ranked as the most prominent feature for both cases classifying good and bad registrations, the rest of the features differ.

3.3.2 Registration evaluation

The results for registration evaluation using bootstrapping and machine learning methods are shown in Table 3.1 for both 2D phantom and 3D patient data. In our experiments, supervised learning methods outperformed the bootstrapping method in terms of accuracy. An advantage of using bootstrapping is that it does not require training data for classifying at the expense of lower classification accuracy and higher computation complexity. Another advantage of the bootstrapping method is the higher sensitivity of 99.92% compared to 96.15% and 96.95% for LDA and RF respectively for the patient data. This makes bootstrapping a reliable method for ensuring the registration is yielding correct results. To further demonstrate the performance of the techniques, their receiver operating characteristics (ROC) curves for assessing the registrations of patient data are shown in Fig. 3.6.

3.4 Discussions

For supervised learning techniques, we have explored existing features and explored new ones for classifying good and bad registrations. For the 2D and 3D data, the selected features differ greatly, and this is likely a result of the differences in image dimensions, imaging contents, as well as the degrees of freedoms in registration. From Table 3.1, we have observed a superior registration assessment quality when using machine learning approaches than bootstrapping. Besides the inherent power of machine learning techniques, the phenomena may

	2D			3D		
	BOOT	LDA	RF	BOOT	LDA	RF
ACC	86.54%	99.35%	99.76%	87.45%	96.97%	97.73%
MCC	66.58%	97.63%	99.14%	75.14%	92.42%	94.29%
TP	16.19%	15.94%	16.47%	26.94%	25.92%	26.14%
TN	70.35%	83.41%	83.29%	60.51%	71.05%	71.59%
FP	12.80%	0.06%	0.18%	12.53%	1.99%	1.45%
FN	0.65%	0.59%	0.06%	0.02%	1.04%	0.82%
N	3376	1688	1688	43328	21664	21664
Sensitivity	96.12%	96.42%	99.64%	99.92%	96.15%	96.95%
Specificity	84.60%	99.93%	99.79%	82.84%	97.28%	98.02%

Table 3.1: ACC and MCC are accuracy and Mathews Correlation Coefficient (MCC) [66]; TP,TN,FP and FN are percentages of true positive, true negative, false positive and false negative relative to the total number of samples, N. Sensitivity (true positive rate) and Specificity (true negative rate) of the classifiers are also included

be partially explained by the employed data that lacked high variability. For the 2D data, the images were collected from the same phantom. Although for the 3D data, the images were obtained from seven different treatment sessions, the variabilities due to diseases and individual anatomy of the data are relatively low. Thus, we postulate that this limitation may cause the evaluation results to be slightly better. However, this does not mean that the classifiers were over-fitted in the scope of the data. We will examine our methods based on multiple patients' data in the future. In order to translate the proposed machine learning methods to clinical applications, data will be gathered from more patient cases and also from human volunteers, and the classifiers will be retrained to improve their generalizability. In the case of volunteers, a wider range of imaging settings and patient motions can be explored, as there is no risk of

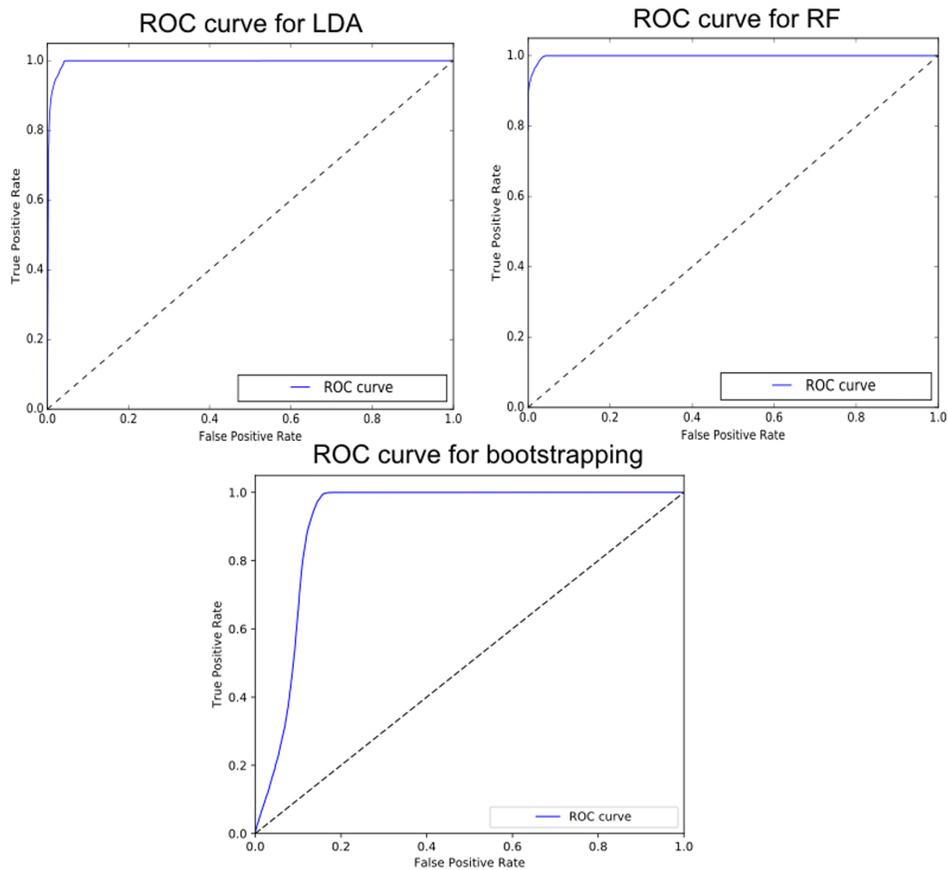


Figure 3.6: ROC curves of different registration assessment methods for the 3D patient data.

affecting the patient’s treatment. Therefore, it is not required to retrain the system for every new patient. It’s also worth mentioning that overall in practice , the samples were divided into two sections, one which the training data were chosen from and the other which training data were chosen from.

Aside from the distances from the ground truth image alignment, there can be other factors, such as image noise, that can influence the registration quality. In our experiments, we have attempted to incorporate variability in image quality in both phantom and patient

data through varying image settings and obtaining images from different treatment sessions, thus incorporating the factors in the experiments implicitly. Explicitly analyzing the effect of individual factors, which require meticulous control during data acquisition and are very difficult to isolate for patient data, is out of the scope of this work. However, the potential of the proposed techniques has been demonstrated with real clinical data.

Both approaches mentioned in the thesis have advantages and disadvantages and therefore both can be viable choices depending on the application. The supervised learning approach has higher classification accuracy and is faster. More specifically, for each registration to be assessed, the bootstrapping takes around 18 seconds while the machine learning methods take less than 1 second. Note that the algorithms were implemented on a Window7 desktop computer with a 16GB RAM and an Intel core i7-4770@3.40GHz processor. The bootstrap method does not require training data, is less accurate, and has a higher computational cost. As a result, supervised learning methods are a better fit for intrafraction motion management, where speed and accuracy are critical. Nonetheless, bootstrapping can still be considered for this application. Since the calculation of each bootstrap result is independent of the other, it is possible to run the registrations in parallel and reduce the runtime. Moreover, bootstrapping is also applicable in cases where the timing requirements for registration validation are not as strict as the target tracking itself, whereby it can be calculated independently from the registration. Bootstrapping is a more natural fit for interfraction registration wherein the algorithm does not need to run in real time. Also, as the patient is positioned, if a reliable registration

can not be performed, the registration setup could be modified until the registration can be reliably carried out. Due to the variation of the images from each day, acquiring the interfraction training data for supervised learning methods that generalize well would be a challenging task, which makes bootstrapping a better choice for this application.

A contributing factor in bootstrapping having lower accuracy is the relatively higher FP rate. Upon further investigation, we realized these FPs occur at steps close to where the registration fails when the initial parameter is far from the true registration. However, due to the randomness involved (the pixel selection), the registration result is successful. An example of this is given in Figure 3.7. In Figure 3.7(a), a case of TP can be seen. The initial registration parameter is far away from the true registration parameters (at step 4) and therefore the result is a poor registration. By inspecting at the mean bootstrap distance, the same can be deduced. In Figure 3.7(b), the registration is poor at step 4, however, judging by the bootstrap result, it seems that the registration was poor from step 3 onwards. Although the registration at step 3 was successful, it was not reliable and had a good chance of failing. Using bootstrapping enables us to detect these cases. From a technical perspective it adds to the number of false positives, but from a practical perspective it ensures that the registration result is reliable.

It should be acknowledged that bootstrapping cannot detect the bias as briefly mentioned in [52]. For instance if the minimum of the cost function is at a distance from the true parameters, caused by the interpolation for example, bootstrapping would fail to detect the bias. In this work, we have intentionally neglected the bias. As shown in [52] bias of image

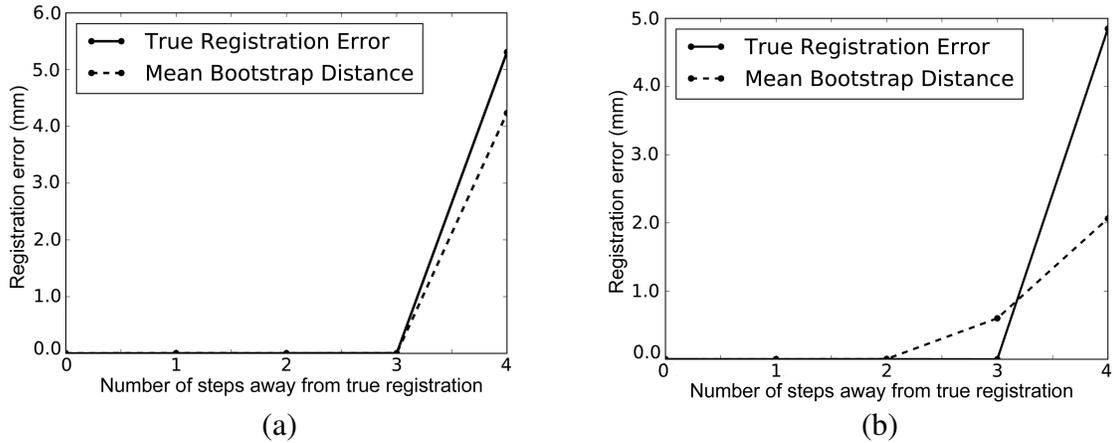


Figure 3.7: Figures showing the registration error as the initial parameters are moved further away from the true result. The x-axis shows the number of steps away from the true registration parameters. (a) and (b) were generated from different bootstrap sampling of the same image pair.

registration algorithms is often quite small, and in this work, we adopted the same assumption as in [52]. As for supervised learning methods, the bias is detected through the features that are independent of the optimization, such as orthant scores, MI, KL-divergence, etc.

We focused on rigid registration to follow the intrafraction motion of the prostate for two main reasons. First, in most current radiotherapy workflows in the clinic, only 3 dimensional (translational), or in certain cases, 6 dimensional (3 translations & 3 rotations) patient positioning is possible. Adaptation of the dose delivery plan to account for deformed anatomy remains an open research problem. Second, for the *in vivo* data collected, no rectal probe or balloon is used. Under these conditions, while some nonlinear deformation may occur, the majority of the motion can be described by a combination of rotations and translations [67]. Deformations of the prostate during treatment are considered relatively small with respect to the margins in

use. To help select features for supervised learning methods, we employed RF, which has been employed previously in image feature selection in medical image analysis [68,69] and offered satisfactory results. While there are also many other techniques for feature selection, a comprehensive comparison is out of the scope of this thesis and will be studied in a future work. Furthermore, the bootstrapping method introduced here is not dependent on the registration framework, it can be used in any other case where displacement is calculated with minimal changes. One area this could be of use is for verification of elastography which was presented in chapter 2.

3.5 Conclusion

In this work, we proposed to use bootstrapping and supervised learning methods (i.e., LDA and RF) to assess ultrasound registration quality. By using both phantom and real clinical data, the two categories of methods were evaluated and compared against each other. While both bootstrapping and supervised learning methods demonstrate good performance, the latter showed better accuracy. In addition, we explored existing features and devised new features that are essential given the unique characteristics of ultrasound images to robustly evaluate the registration quality using machine learning methods. To the best of our knowledge, it is the first time that automatic registration assessment techniques are proposed for ultrasound imaging, which is widely used in image-guided procedures.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

Ultrasound imaging is a popular modality due to it being safe, non-invasive, easy to use and relatively cheap compared to other modalities. This thesis was focused on two main topics. Firstly, estimating deformation in ultrasound images and secondly, verification of an estimated deformation. In chapter 2, a method for displacement estimation was introduced for ultrasound elastography. In chapter 3, methods for deformation verification of ultrasound images were introduced and discussed in the context of ultrasound radiotherapy.

Ultrasound elastography is a technique which is used to estimate the mechanical properties of the tissue. This is done by estimating the displacement between two RF frames acquired

before and after a deformation in the tissue. A common method for calculating the displacement is to divide the RF frame into windows and calculate the displacement with correlation based methods. This method can fail for cases with large signal decorrelation. For this reason, dynamic programming was used to calculate displacement on all the pixels on an RF-line. Although this method improved on the previous work, it still fails with signal decorrelation. Therefore, we proposed a method to calculate the displacement on a tree instead of a single RF-line. By using more information from the frames, we were able to reduce the displacement error. We verified this by testing our method on simulation, phantom and patient data. As a result, the deformation field was more accurate which results in better diagnostics.

Ultrasound is also used during radiotherapy. In case of prostate radiotherapy, the prostate is imaged transperineally using ultrasound and tracked in the course of the treatment. This tracking is done by registering the acquired 3D volumes with a reference volume. The quality of these registration have an important factor in the outcome of the treatment. In order to evaluate this quality, we proposed two methods based on supervised learning and bootstrapping. We evaluated our methods on 2D phantom data and 3D patient data. According to our results, supervised learning performed better in terms of accuracy and computational time; that is in cases where enough training data was available. However, availability of training data is not always guaranteed and in these cases, bootstrapping can be utilized. Bootstrapping also offered higher sensitivity which makes it a tool for ensuring the reliability of registration setups. The contributions in this chapter was threefold. First, we proposed new features to be used

with the supervised learning method. To the best of our knowledge, dissimilarity probes and the orthant scores have not been used as features for registration quality evaluation. Second, we introduced the utilization of bootstrapping for ultrasound guided radiotherapy. Third, we compared the two approaches and offered usages for different case scenarios. The combination of higher computational complexity of bootstrapping and its sensitivity makes it an ideal candidate for inter-fraction registration. The fact that acquiring training data for this case usage is cumbersome also adds to the value of using bootstrapping. However, for inter-fraction registration, supervised learning methods are more viable due to their better accuracy, runtime and availability of training data.

4.2 Future work

In chapter 2 we introduced a new method to calculate displacement on a tree. Although we saw improvement over the previous methods there is room for improvements in terms of error and computational time. In the proposed method, the tree used has a simple structure. A structure that covers all the RF frame would use more information from the ultrasound images and further reduce the error. Also this method can be parallelized to run on Graphical Processing Units for speedups.

In chapter 3, we introduced bootstrapping and compared it with supervised learning methods for registration quality evaluation. In the supervised learning methods, the feature ranking method used can be further improved to choose better features. For instance, by using

Minimum Redundancy Maximum Relevance (mRMR) technique [70], more relevant features can be chosen. Also in bootstrapping, the merit used to classify the registration is fairly simple. A simple classifier can be used to train on the bootstrap results to take advantage of more information from these results, which should increase the accuracy of the method.

Bibliography

- [1] K. A. Kaproth-Joslin, R. Nicola, and V. S. Dogra, “The history of US: from bats and boats to the bedside and beyond: RSNA centennial article,” *Radiographics*, vol. 35, no. 3, pp. 960–970, 2015.
- [2] H. S. Hashemi, M. Boily, P. A. Martineau, and H. Rivaz, “Ultrasound elastography: Efficient estimation of tissue displacement using an affine transformation model,” pp. 1 013 903–1 013 903, 2017.
- [3] R. Shams, Y. Xiao, F. Hebert, M. Abramowitz, R. Brooks, and H. Rivaz, “Assessment of rigid registration quality measures in ultrasound-guided radiotherapy,” *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1–1, 2017.
- [4] R. Shams, M. Boily, P. A. Martineau, and H. Rivaz, “Dynamic programming on a tree for ultrasound elastography,” *SPIE Medical Imaging, 97901F–97901F, International Society for Optics and Photonics*, 2016.

- [5] J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, and T. Varghese, "Elastography: ultrasonic estimation and imaging of the elastic properties of tissues." *Proceedings of the Institution of Mechanical Engineers. Part H, Journal of engineering in medicine*, vol. 213, pp. 203–233, 1999.
- [6] L. Gao, K. Parker, R. Lerner, and S. Levinson, "Imaging of the elastic properties of tissue - a review," *Ultrasound in medicine & biology*, vol. 22, no. 8, pp. 959–977, 1996.
- [7] K. J. Parker, L. S. Taylor, S. Gracewski, and D. J. Rubens, "A unified view of imaging the elastic properties of tissue," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 2705–2712, 2005.
- [8] J. F. Greenleaf, M. Fatemi, and M. Insana, "Selected methods for imaging elastic properties of biological tissues," *Annual review of biomedical engineering*, vol. 5, no. 1, pp. 57–78, 2003.
- [9] K. M. Hiltawsky, M. Krüger, C. Starke, L. Heuser, H. Ermert, and A. Jensen, "Freehand ultrasound elastography of breast lesions: clinical results," *Ultrasound in medicine & biology*, vol. 27, no. 11, pp. 1461–1469, 2001.
- [10] T. J. Hall, Y. Zhu, and C. S. Spalding, "In vivo real-time freehand palpation imaging," *Ultrasound in medicine & biology*, vol. 29, no. 3, pp. 427–435, 2003.

- [11] S. Goenezen, J.-F. Dord, Z. Sink, P. E. Barbone, J. Jiang, T. J. Hall, and A. A. Oberai, “Linear and nonlinear elastic modulus imaging: an application to breast cancer diagnosis,” *Medical Imaging, IEEE Transactions on*, vol. 31, no. 8, pp. 1628–1637, 2012.
- [12] E. E. Konofagou, J. D’hooge, and J. Ophir, “Myocardial elastography—a feasibility study in vivo,” *Ultrasound in medicine & biology*, vol. 28, no. 4, pp. 475–482, 2002.
- [13] H. Rivaz, P. Foroughi, I. Fleming, R. Zellars, E. Boctor, and G. Hager, “Tracked regularized ultrasound elastography for targeting breast radiotherapy,” pp. 507–515, 2009.
- [14] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, “Ultrasound elastography using multiple images,” *Medical Image Analysis*, vol. 18, no. 2, pp. 314–329, 2014.
- [15] F. Viola and W. F. Walker, “A comparison of the performance of time-delay estimators in medical ultrasound,” *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 50, no. 4, pp. 392–401, 2003.
- [16] M. Lu, Y. Tang, R. Sun, T. Wang, S. Chen, and R. Mao, “A real time displacement estimation algorithm for ultrasound elastography,” *Computers in Industry*, vol. 69, pp. 61–71, 2015.
- [17] A. Pesavento, C. Perrey, M. Krueger, and H. Ermert, “A time-efficient and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation,” *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 46, no. 5, pp. 1057–1067, 1999.

- [18] J. E. Lindop, G. M. Treece, A. H. Gee, and R. W. Prager, “Phase-based ultrasonic deformation estimation,” *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 55, no. 1, pp. 94–111, 2008.
- [19] S. Rosenzweig, M. Palmeri, and K. Nightingale, “GPU-based real-time small displacement estimation with ultrasound,” *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 58, no. 2, pp. 399–405, 2011.
- [20] M. Rao, Q. Chen, H. Shi, T. Varghese, E. Madsen, J. Zagzebski, and T. Wilson, “Normal and shear strain estimation using beam steering on linear-array transducers,” *Ultrasound in medicine & biology*, vol. 33, no. 1, pp. 57–66, 2007.
- [21] A. Kuzmin, A. Zakrzewski, B. Anthony, and V. Lempitsky, “Multi-frame elastography using a handheld force-controlled ultrasound probe,” *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 62, no. 8, pp. 1486–1500, August 2015.
- [22] J. Jiang and T. J. Hall, “A coupled subsample displacement estimation method for ultrasound-based strain elastography,” *Physics in medicine and biology*, vol. 60, no. 21, p. 8347, 2015.
- [23] H. S. Hashemi and H. Rivaz, “Global time-delay estimation in ultrasound elastography,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 10, pp. 1625–1636, 2017.

- [24] H. Rivaz, E. Boctor, P. Foroughi, R. Zellars, G. Fichtinger, and G. Hager, "Ultrasound elastography: A dynamic programming approach," *IEEE Transactions on Medical Imaging*, vol. 27, no. 10, pp. 1373–1377, 2008.
- [25] B. Byram, G. E. Trahey, and M. Palmeri, "Bayesian speckle tracking. part II: biased ultrasound displacement estimation," *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 60, no. 1, 2013.
- [26] T. J. Hall, P. Barbone, A. A. Oberai, J. Jiang, J. F. Dord, S. Goenezen, and T. G. Fisher, "Recent results in nonlinear strain and modulus imaging," *Current medical imaging reviews*, vol. 7, no. 4, p. 313, 2011.
- [27] J. Jiang and T. Hall, "6F-3 a regularized real-time motion tracking algorithm using dynamic programming for ultrasonic strain imaging," pp. 606–609, 2006.
- [28] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager, "Real-time regularized ultrasound elastography," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 928–945, April 2011.
- [29] P. Foroughi, H. Rivaz, I. N. Fleming, G. D. Hager, and E. M. Boctor, "Tracked Ultrasound Elastography (TrUE)," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, PT II*, vol. 6362, 2010, pp. 9–16.

- [30] R. Xia, G. Tao, and A. K. Thittai, "Dynamic frame pairing in real-time freehand elastography." *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 6, pp. 979–85, 2014.
- [31] C. Pellot-Barakat, F. Frouin, M. F. Insana, and A. Herment, "Ultrasound elastography based on multiscale estimations of regularized displacement fields," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 153–163, 2004.
- [32] E. Brusseau, J. Kybic, J.-F. Déprez, and O. Basset, "2-D locally regularized tissue strain estimation from radio-frequency ultrasound images: Theoretical developments and results on experimental data," *IEEE Transactions on Medical Imaging*, vol. 27, no. 2, pp. 145–160, 2008.
- [33] I. Fleming, H. Rivaz, E. Boctor, and G. Hager, "Robust dynamic programming method for ultrasound elastography," pp. 83 201–83 201, 2012.
- [34] O. Veksler, "Stereo correspondence by dynamic programming on a tree," pp. 384–390 vol. 2, 2005.
- [35] J. Jensen and N. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, vol. 39, no. 2, pp. 262–267, March 1992.
- [36] J. A. Jensen, "Simulation of advanced ultrasound systems using field II," pp. 636–639 Vol. 1, April 2004.

- [37] T. O’Shea, J. Bamber, D. Fontanarosa, S. van der Meer, F. Verhaegen, and E. Harris, “Review of ultrasound image guidance in external beam radiotherapy part II: intra-fraction motion management and novel applications,” *Physics in medicine and biology*, vol. 61, no. 8, p. R90, 2016.
- [38] D. A. Jaffray, “Image-guided radiotherapy: from current concept to future perspectives,” *Nature Reviews Clinical Oncology*, vol. 9, no. 12, pp. 688–699, 2012.
- [39] J. De Los Santos, R. Popple, N. Agazaryan, J. E. Bayouth, J.-P. Bissonnette, M. K. Bucci, S. Dieterich, L. Dong, K. M. Forster, D. Indelicato *et al.*, “Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery,” *Int J Radiat Oncol Biol Phys*, vol. 87, no. 1, pp. 33–45, 2013.
- [40] M. Lachaine and T. Falco, “Intrafractional prostate motion management with the Clarity autoscan system,” *Med. Phys. Int*, vol. 1, no. 1, pp. 72–80, 2013.
- [41] R. Datteri, Y. Liu, P.-F. D’Haese, and B. Dawant, “Validation of a Non-Rigid Registration Error Detection Algorithm using Clinical MRI Brain Data.” *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 1–11, 2014.
- [42] P. Risholm, F. Janoos, I. Norton, A. J. Golby, and W. M. Wells, “Bayesian characterization of uncertainty in intra-subject non-rigid registration,” *Medical Image Analysis*, vol. 17, no. 5, pp. 538–555, 2013.

- [43] F. Janoos, P. Risholm, and W. Wells, “Bayesian characterization of uncertainty in multi-modal image registration,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7359 LNCS, no. III, pp. 50–59, 2012.
- [44] P. Risholm, S. Pieper, E. Samset, and W. M. Wells, “Summarizing and visualizing uncertainty in non-rigid registration,” *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, vol. 6362 LNCS, no. PART 2, pp. 554–561, 2010.
- [45] P. Risholm, E. Samset, and W. Wells, “Bayesian estimation of deformation and elastic parameters in non-rigid registration,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6204 LNCS, pp. 104–115, 2010.
- [46] I. J. A. Simpson, J. A. Schnabel, A. R. Groves, J. L. R. Andersson, and M. W. Woolrich, “Probabilistic inference of regularisation in non-rigid registration,” *NeuroImage*, vol. 59, no. 3, pp. 2438–2451, 2012.
- [47] I. J. Simpson, M. Woolrich, A. R. Groves, and J. A. Schnabel, “Longitudinal brain MRI analysis with uncertain registration,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*. Springer, 2011, pp. 647–654.

- [48] J. Wu and S. S. Samant, “Novel image registration quality evaluator (RQE) with an implementation for automated patient positioning in cranial radiation therapy,” *Medical Physics*, vol. 34, no. 6, pp. 2099–2112, 2007.
- [49] J. Wu and M. J. Murphy, “A neural network based 3D/3D image registration quality evaluator for the head-and-neck patient setup in the absence of a ground truth,” *Medical Physics*, vol. 37, no. 11, pp. 5756–5764, 2010.
- [50] S. E. A. Muenzing, B. van Ginneken, K. Murphy, and J. P. W. Pluim, “Supervised quality assessment of medical image registration: Application to intra-patient CT lung registration,” *Medical Image Analysis*, vol. 16, no. 8, pp. 1521–1531, 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.media.2012.06.010>
- [51] H. Sokooti, G. Saygili, B. Glocker, B. P. Lelieveldt, and M. Staring, “Accuracy estimation for medical image registration using regression forests,” pp. 107–115, 2016.
- [52] J. Kybic, “Bootstrap resampling for image registration uncertainty estimation without ground truth,” *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 64–73, 2010.
- [53] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [54] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

- [55] P. Jannin, J. M. Fitzpatrick, D. J. Hawkes, X. Pennec, R. Shahidl, and M. W. Vannier, “Validation of medical image processing in image-guided therapy,” *IEEE Transactions on Medical Imaging*, vol. 21, no. 12, pp. 1445–1449, 2002.
- [56] E. B. van de Kraats, G. P. Penney, D. Tomazevic, T. Van Walsum, and W. J. Niessen, “Standardized evaluation methodology for 2-D-3-D registration,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1177–1189, 2005.
- [57] H. Rivaz, S. J.-S. Chen, and D. L. Collins, “Automatic deformable MR-ultrasound registration for image-guided neurosurgery,” *IEEE transactions on medical imaging*, vol. 34, no. 2, pp. 366–380, 2015.
- [58] D. Skerl, B. Likar, and F. Pernus, “A protocol for evaluation of similarity measures for rigid registration,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 6, pp. 779–791, 2006.
- [59] H. Zabrodsky, S. Peleg, and D. Avnir, “A measure of symmetry based on shape similarity,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 1992, pp. 703–706.
- [60] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

- M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [62] M. R. Chernick, *Bootstrap methods: A guide for practitioners and researchers*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2011, vol. 619.
- [63] M. Abramowitz, E. Bossart, L. Martin, R. Brooks, F. Lathuiliere, F. Laura, A. Iskanian, and A. Pollack, “Noninvasive real-time prostate tracking using a transperineal ultrasound: A clinical trial comparison to RF transponders with visual confirmation,” *International Journal of Radiation Oncology Biology Physics*, vol. 87, no. 2, p. S682, 2013.
- [64] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, “Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit,” *Studies in health technology and informatics*, pp. 586–592, 2002.
- [65] R. Brooks, “Intrafraction prostate motion correction using a non-rectilinear image frame,” in *International Workshop on Prostate Cancer Imaging*. Springer, 2011, pp. 57–59.
- [66] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, “Assessing the accuracy of prediction algorithms for classification: an overview,” *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000.

- [67] M. J. Ghilezan, D. A. Jaffray, J. H. Siewerdsen, M. Van Herk, A. Shetty, M. B. Sharpe, S. Zafar Jafri, F. A. Vicini, R. C. Matter, D. S. Brabbins, and A. A. Martinez, “Prostate gland motion assessed with cine-magnetic resonance imaging (cine-MRI),” *Int J Radiat Oncol Biol Phys*, vol. 62, no. 2, pp. 406–17, 2005. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/15890582>
- [68] F. Aalamifar, H. Rivaz, J. J. Cerrolaza, J. Jago, N. Safdar, E. M. Boctor, and M. Linguarraru, “Classification of kidney and liver tissue using ultrasound backscatter data,” *Medical Imaging 2015: Ultrasonic Imaging and Tomography*, vol. 9419, 2015.
- [69] N. Uniyal, H. Eskandari, P. Abolmaesumi, S. Sojoudi, P. Gordon, L. Warren, R. N. Rohling, S. E. Salcudean, and M. Moradi, “Ultrasound RF time series for classification of breast lesions,” *IEEE Trans Med Imaging*, vol. 34, no. 2, pp. 652–61, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/25350925>
- [70] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.