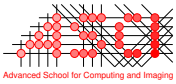
# Fast 4D Ultrasound Registration for Image Guided Liver Interventions

Jyotirmoy Banerjee

Fast 4D Ultrasound Registration for Image Guided Liver Interventions
Jyotirmoy Banerjee
PhD thesis, Erasmus University Rotterdam, the Netherlands.

The work in this thesis was conducted at the Departments of Radiology & Nuclear Medicine and Medical Informatics at the Erasmus MC University Medical Center, Rotterdam, The Netherlands. This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (project number 10482).


This work was carried out in the ASCI graduate school.
ASCI dissertation series number 353.

Additional financial support for the publication of this thesis was provided by Department of Radiology & Nuclear Medicine at the Erasmus MC and Philips.

Cover design by Indraneel Chatterjee & Jyotirmoy Banerjee
Printed by GVO printers & designers B.V.
ISBN: 978-94-6332-038-2

Copyright © 2016 by Jyotirmoy Banerjee
No part of these publications (neither scientific nor professional) may be reproduced, stored in a retrieval system or transmitted in any form or by any means including electronic, mechanical, photocopying, or otherwise, without the prior written permission of the author.

# Fast 4D Ultrasound Registration for Image Guided Liver Interventions

## Snelle 4D echografie registratie voor beeldgeleide interventies in de lever

**Thesis**

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
Rector Magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on
Thursday, June 30, 2016 at 11:30 hours

by

**Jyotirmoy Banerjee**
born in Bongaon, India

Erasmus University Rotterdam

## Doctoral committee

| | |
|---|---|
| Supervisor | : Prof.dr. W.J. Niessen |
| Other members | : Prof.dr. D. Hawkes |
| | : Prof.dr.ir. N. de Jong |
| | : Dr. T. Langø |
| Cosupervisors | : Dr.ir. T. van Walsum |
| | : Dr. A. Moelker |

*This is for you, Ma & Baba.*

# Contents

# Chapter 1

## Introduction

## 1.1 The liver and common liver problems

The liver is the largest glandular organ of the human body, see Figure 1.1. It weighs about three pounds in a full grown adult. It is located in the upper abdominal cavity of the human body, under the diaphragm. We cannot generally feel the liver from outside as it is protected inside the rib cage.

The liver plays an important role in our body. All of the blood leaving the digestive tract passes through the liver. The liver breaks down the blood, regulates most chemical levels and creates nutrients for the body. The liver is considered a gland because it secretes a product called bile. The bile breaks down the fat, preparing for further digestion and absorption by the body. Apart from that the liver also has other functions like getting rid of the alcohol and drugs present in the blood, storing glucose and some essential vitamins and iron. It also aids in the blood clotting mechanism.

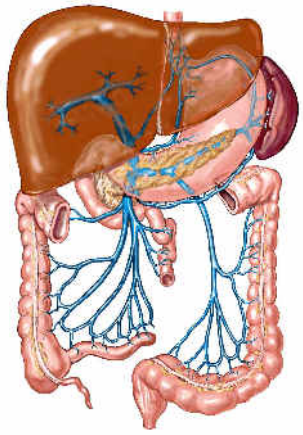


**Figure 1.1:** *Liver and the portal venous system.*

Liver problems are a serious health issue. According to a survey around 29 million persons alone in Europe suffer from chronic liver conditions [9]. The common liver problems are hepatitis, fatty liver, liver cancer and liver damage caused by alcohol abuse or the pain reliever acetaminophen [35]. Continuous, long term disease may cause a condition of the liver known as the *Liver Cirrhosis*. Liver cirrhosis makes the liver scarred and hardened up, see Figure 1.2. In such a situation the liver cannot perform its functions in the required manner. The common causes of liver cirrhosis are long term alcohol abuse and hepatitis infection. According to the 2010 disease study [81], per year around 1030,800 people die of liver cirrhosis and around 752,100 people die of liver cancer. Liver cirrhosis is responsible for approximately 170,000 deaths and liver cancer is responsible for approximately 47,000 deaths per year in the Europe [9].

Portal hypertension, increase of blood pressure in the portal venous system (Figure 1.1) is present in most of the liver cirrhosis patients. The portal vein is fed by the veins which come from the stomach, pancreases or intestines. It enters the liver and then branches into veins which travel throughout the liver. In case of cirrhotic liver the blood flow is blocked causing portal hypertension. Other major causes of portal hypertension include blood clots or blockages in the vein, focal nodular hyperplasia and parasitic infection (schistosomiasis). Portal hypertension causes collateral vessels to bypass the liver. These vessels are fragile and may rupture; causing

bleedings in esophagus and stomach. In addition, portal hypertension causes accumulation of fluid in the abdominal cavity, called ascites. One of the treatments of portal hypertension is Transjugular intrahepatic portosystemic shunt (TIPS).

Liver cancer, also known as hepatic cancer, can be divided into two types - primary (*hepatocellular carcinoma*) and secondary (*liver metastases*) liver cancer. Liver metastases are cancerous tumors that have spread to the liver from somewhere else in the body; hepatocellular carcinoma are cancers that grow in the liver from liver cells. Radiofrequency ablation (RFA) is a therapeutic treatment procedure for liver metastases and hepatocellular carcinoma [86]. In general, surgical resection is the preferred treatment for liver cancer, especially when the tumor is confined to one area. However, it is often not possible in medical conditions like cirrhosis, in cases where the tumor has spread to nearby blood vessels or in advanced stages of the cancer. In such cases alternate treatment options like ablation are considered [104].

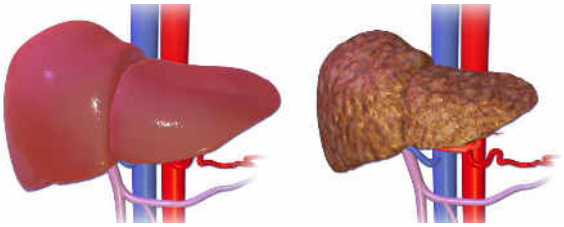


**Figure 1.2:** *Left - normal liver, right - cirrhotic liver (Courtesy - wikipedia).*

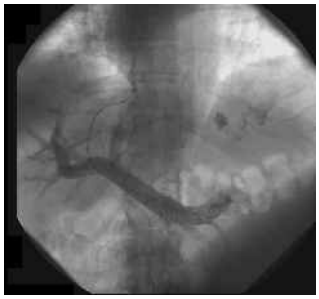
In this thesis, the focus is on treatment of portal hypertension and liver cancer.

## 1.2 Minimally invasive interventions

Minimally invasive procedures, where small incisions are made to perform the surgery are an alternative to open surgery. Replacing classical surgical interventions by minimally invasive alternatives is beneficial for the patient and the health care system, as it has large potential for reducing complication rates, minimizing surgical trauma and reducing hospital stay [73][38]. From the socio-economic perspective, minimally invasive interventions can help in reducing medical expenditures and increasing worker productivity [1]. However, in these procedures imaging is required due to lack to direct eyesight. Using imaging techniques, physicians can navigate interventional instruments inside the body without making large incisions.

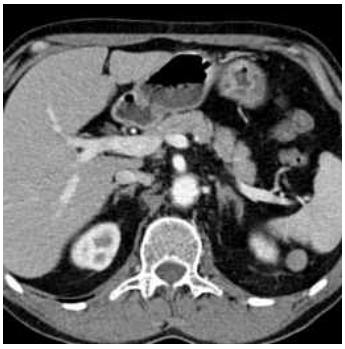
### 1.2.1 Imaging modalities

Imaging has traditionally been used for preoperative diagnosis and planning. Intraoperative imaging provides further information during interventions. Below, some commonly used interventional imaging modalities are discussed [24].

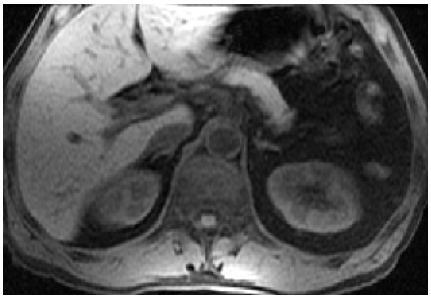


(a) X-ray angiography

(b) Ultrasound



(c) CT



(d) MR

**Figure 1.3:** *Liver images*

*X-rays :* X-ray imaging is 2D technique used to plan therapy and to perform surgery. Major disadvantages of x-rays imaging is that it produces projection images and uses ionizing radiation. It also requires contrast agent to highlight the vessels. However, as it is real-time, it is one of the key imaging techniques used in interventional procedures.

*Ultrasound :* Ultrasound imaging uses high-frequency sound waves transmitted longitudinally to image the organs and structures inside the body. Ultrasound is a safe modality, as it does not use any harmful radiation in the imaging process. A combination of good hand-eye coordination and correct positioning of the probe is required to acquire good image. For acquiring good contrast images, the ultrasound wave must be reflected off a tissue interface at right angles. The B or brightness mode provides a greyscale image that is noisy because of loss of sound waves due to scatter and refraction. Ultrasound imaging is real-time, safe and low-cost. Compared with other imaging modalities, such as computed tomography and magnetic resonance imaging, ultrasound is a very mobile device. It is frequently

used in interventional radiology. However, ultrasound images have limited depth and the image quality is not as clear and crisp as those obtained with computed tomography and magnetic resonance imaging [24].

*Computed tomography :* Computed tomography combines multiple x-ray images taken from different angles to create cross-sectional slice of the body. Series of cross-sectional images acquired during a computed tomography scan can be used to generate three-dimensional images. 3D imaging will help in exact localization of instruments during intervention. For interventional use, a major disadvantage of Computed tomography is the lack of real-time imaging. Additionally, similar to the x-ray imaging it uses ionizing radiation [25][45].

*Magnetic resonance imaging :* Magnetic resonance shows excellent tissue discrimination and is free of ionizing radiation. It has the ability to characterize functional and physiological parameters of the treated tissues during minimally invasive diagnostic and therapeutic radiologic procedures. The limitations of using interventional magnetic resonance imaging are restricted access to the patient, the associated expenses, limited real-time possibilities and the need for MR compatible devices [75].

### 1.2.2 Minimally invasive liver interventions

Minimally invasive interventional techniques can address a large variety of problems in the liver including both diagnostic and therapeutic interventions [99].

*TIPS procedure:* In case of portal hypertension and with appropriate clinical condition, TIPS may be the therapeutic option of choice [99] [116]. In a TIPS procedure a tract in the liver is created that shortcuts two veins in the liver, the portal vein and the hepatic vein [112][113][66]. The purpose is to decompress the systemic venous system, which reduces the portal hypertension. Interventional radiologists generally perform the procedure under x-ray guidance, see Figure 1.4. Access to the liver is achieved via the jugular vein in the neck. A guidewire followed by a sheath is advanced, travelling from superior vena cava into the inferior vena cava to gain access to the hepatic vein. A catheter is advanced over the wire into the hepatic vein to measure the pressure. Once the wire and the catheter are removed a long curved needle is advanced through the sheath, pushed through the wall of the hepatic vein and liver parenchyma to connect the hepatic vein to the large portal vein. A balloon is used to inflate the tract created by the needle, see Figure 1.4. The balloon is removed and the sheath catheter is advanced for the stent placement. All the instruments except the stent are removed, allowing proper blood flow through the liver. To cross check, pressure in the portal and hepatic vein is measured.

TIPS is one of the most challenging interventional procedures [107]. There are several procedural complications that may occur during the various stages of the procedure (vascular access, portal vein localization and access and stent placement). The most challenging step during the procedure is obtaining access to the

**Figure 1.4:** *Fluoroscopic image of TIPS in progress (Courtesy - wikipedia).*

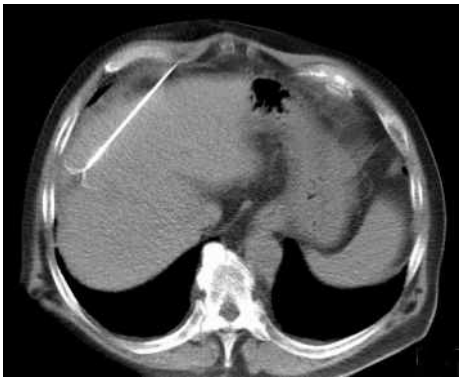


**Figure 1.5:** *CT scan showing radiofrequency ablation of a liver lesion (Courtesy - wikipedia).*

portal vein [107]. Direct puncture into the main portal vein can cause profuse bleeding with fatal outcome. Injuries to the hepatic artery and the biliary duct are also possible during the procedure.

The main problem while doing this procedure is the lack of three-dimensional information about the anatomy and instruments during the intervention. Multiple needle insertions are often needed during the critical phase of accessing the portal vein, with a risk of causing complications and lengthening the procedure time [26]. X-ray imaging (fluoroscopy), see Figure 1.4, lacks depth perception [107] and can result in high radiation doses to patients and staff [140]. According to the guidelines mentioned in Krajina et al. [63] ultrasound should be used to navigate puncture of the jugular and portal veins. However, ultrasound guidance suffer from difficult manual coordination and poor target visibility [68].

*RFA procedure:* RFA is a minimally invasive procedure used in the treatment of hepatocellular carcinoma and liver metastases when resection cannot be performed [86]. RFA is an image guided intervention where a needle electrode is placed through the skin into the liver tumor, see Figure 1.5. High-frequency electrical currents are passed through the electrode, creating heat that destroys the cancer cells, without damaging the surrounding liver tissues. A temperature range of 50 to 100°C is maintained throughout the target region for protein coagulation with irreversible damage to the cells [86]. A 1-cm-thick tumor-free margin around the tumor should be obtained to ablate all viable tumor tissues. Imaging follow-up after the ablation is necessary as there is little visual feedback on the ablated zone during the treatment [59]. Other ablation techniques commonly used in the treatment of liver include cryoablation and microwave ablation.

Hemorrhage is one of the major concerns during thermal ablation. To prevent hemorrhage the practitioner should try to minimize the number of punctures of the

liver capsule, avoid crossing major vessels and by cauterizing the needle tract after ablation [59]. Ablation may also cause thrombosis of the portal vein branch; especially in cirrhotic liver with already reduced liver function. The most difficult part of this procedure is the interpretation and selection of oblique views for needle insertion and target visualization. During the procedure computed tomography images are acquired to check the position and trajectory of the needle, and to mentally compare these to the planned trajectory. If the needle deviates from the planned trajectory, then the needle is readjusted. This process is repeated until the target is reached [127]. Apart from computed tomography, targeting of the lesion is also performed using ultrasound or magnetic resonance imaging, which is generally based on the operator preference and the availability of the imaging device [86].

### 1.2.3 Navigation tools

Minimally invasive procedures are challenging for the clinicians as there is no direct eyesight and conventional imaging modalities have limited capabilities. Navigation tools are useful in guiding these interventions. They provide surgical assistance in terms of the location of the devices and other necessary information required to successfully drive the procedure. These tools increase the ability to introduce or advance instruments like catheter or needle to a desired target inside the body. They provide assistance to physicians during all the phases of both diagnostic and therapeutic procedures.

**Figure 1.6:** *Electromagnetic and Optical tracking systems.*

Commercial software solutions like the Virtual Navigator from Esaote and PercuNav from Philips are focused on interventional imaging, where magnetic resonance imaging/computed tomography modalities are registered with real-time ultrasound. These systems typically use external tracking systems (e.g. electromagnetic or optical) which report the position

**Figure 1.7:** *Philips CT scanner.*

**Figure 1.8:** *iU22 xMATRIX Ultrasound System from Philips.*

and orientation of specially designed tracking objects with respect to a fixed base. By rigidly mounting the tracking objects on the devices, various images and instruments can be co-registered for precise overlay of the instrument and multimodal images. Optical tracking requires clear line-of-sight between cameras and targets; hence objects inside the body cannot be tracked. Electromagnetic tracking does not require line-of-sight to operate. However it is sensitive to ferromagnetic objects in the operating room, which can cause significant distortion to the system.

In literature a great deal of work has been published dealing with image-based needle or instrument navigation using x-rays, computed tomography, and magnetic resonance imaging [102]. For real-time guidance, x-rays and ultrasound are two frequently used imaging modalities. With its ease of use, improved image quality over the years and nonionizing energy for imaging, ultrasound is an essential part of modern diagnostic and therapeutic navigation.

**Figure 1.9:** *X6-1 ultrasound 3D probe from Philips.*

Navigation systems using the tracking devices (e.g., electromagnetic, optical) are well advanced and cater to clinical routines of largely rigid anatomical regions e.g. in neurosurgery [102]. For interventions of the soft tissue regions like the abdomen, the support by the navigation systems are limited, because the displacement of the organs caused by factors like the breathing may result in a large misalignment of the preoperative and the intraoperative data [73][60].

## 1.3 4D ultrasound for image guided interventions

Real-time 3D Ultrasound using a 2D matrix array transducer is a relatively novel imaging modality, see Figure 1.9. It is currently mainly used for diagnosis. It has a large potential to be used for image guidance, but a 2D slice display or a 3D rendering, which currently are the standard visualizations, do not provide support for guidance. Further during these procedures clinicians often hold the ultrasound probe steady to visualize a localized area in the liver ultrasound volume. Breathing shifts the region of interest and makes it difficult to constantly focus on a region of interest. To overcome this problem an approach to correct for the motion due to breathing is imperative.

**Figure 1.10:** *Demo of the real-time 3D ultrasound based navigation system for the TIPS procedure developed during the project.*

Image fusion of ultrasound and cross-sectional imaging modalities such as computed tomography or magnetic resonance imaging is important in cases where the lesions are visible in computed tomography or magnetic resonance imaging but not visible in ultrasound [25].

## 1.4 This thesis

Purpose of the research presented in this thesis is to develop and evaluate image processing techniques that enable the effective use of 3D ultrasound for image guided interventions of the liver, see Figure 1.10. To this end, we worked on the following four subjects:

1. In Chapter 2 we propose a fast 3D ultrasound registration, in a novel framework that is based on block-matching and game-theory-based outlier rejection methods, which was implemented on a graphics processing unit. In Chapter 3 this approach is extended to streaming ultrasound data.

2. A two-stage 3D ultrasound tracking approach based on the fast 3D ultrasound registration is discussed in Chapter 4. This approach, which was able to

align pre-operative 3D ultrasound planning data with live 3D ultrasound images in real-time, was integrated in a system for ultrasound-guided TIPS interventions. The approach was also used for and evaluated in 3D ultrasound tracking challenge at MICCAI 2015 as discussed in Chapter 5.

3. The pose of the imaging data in a clinical scenario may vary considerably depending on the modality and other factors like the patient position. In Chapter 6 we present a method for rotationally invariant description of landmarks or regions in 3D using local binary patterns (LBP). The LBP in 3D requires a spherical sampling, which is represented in a spherical harmonics framework. The framework helps in obtaining rotation invariant representation.

4. In Chapter 7 a fast multi-modal (computed tomography-3D ultrasound) registration approach, adapting the previously developed block-matching and outlier rejection strategy to multi-model image registration, proposing a novel similarity metric and improved outlier rejection is presented.

# Part I

# Ultrasound for guidance

# Chapter 2

## Fast and Robust 3D Ultrasound Registration

Based on:

Jyotirmoy Banerjee, Camiel Klink, Edward D. Peters, Wiro J. Niessen, Adriaan Moelker and Theo van Walsum, Fast and Robust 3D Ultrasound Registration - Block and Game Theoretic Matching, *Medical Image Analysis*, 2015.

**Abstract**

Real-time 3D US has potential for image guidance in minimally invasive liver interventions. However, motion caused by patient breathing makes it hard to visualize a localized area, and to maintain alignment with pre-operative information. In this work we develop a fast affine registration framework to compensate in real-time for liver motion/displacement due to breathing. The affine registration of two consecutive ultrasound volumes in time is performed using block-matching. For a set of evenly distributed points in one volume and their correspondences in the other volume, we propose a robust outlier rejection method to reject false matches. The inliers are then used to determine the affine transformation. The approach is evaluated on 13 4D ultrasound sequences acquired from 8 subjects. For 91 pairs of 3D ultrasound volumes selected from these sequences, a mean registration error of 1.8 mm is achieved. A graphics processing unit (GPU) implementation runs the 3D US registration at 8 Hz.

## 2.1 Introduction

### 2.1.1 Clinical motivation

Replacing classical surgical interventions by minimally invasive alternatives is beneficial for the patient and the health care system, as it has large potential for reducing complication rates, minimizing surgical trauma, and reducing hospital stay. The minimally invasive character, however, makes these interventions challenging for the clinician. There is no direct eyesight on the target region and conventional interventional imaging modalities have limited capabilities. Furthermore the user interfacing and interaction with the equipment involved often is not ergonomically well-designed, and does not match the interventional work flow well. Image guidance is crucial in minimally invasive interventions. Image guidance can be based on preoperative imaging data (mostly magnetic resonance imaging and computed tomography) or intraoperative imaging data (X-ray, ultrasound). Hybrid approaches can also be useful, in which the diagnostic quality of preoperative images can be combined with the real-time nature of intraoperative images (also known as fusion imaging). Four dimensional (4D) ultrasound (US) is a relatively novel imaging modality that currently is mainly used for diagnosis.

Radiofrequency ablation (RFA) and the transjugular intrahepatic portosystemic shunt (TIPS) procedure are examples of percutaneous minimally invasive image-guided interventions which are used more and more as alternative to surgical procedures. 4D US has large potential in assisting the clinicians in these procedures, as it provides real-time three dimensional (3D) vision. During these procedures the clinician often holds the US probe steady to visualize a localized area in the liver US volume. Breathing shifts the region of interest and makes it difficult to constantly focus on a region of interest, more so in the presence of a catheter. The purpose of our work is to develop a technique suited for fast 3D ultrasound registration during image guided minimally invasive intervention to compensate breathing motion. In addition, our approach would help in keeping the registration up to date in US fusion imaging.

### 2.1.2 Related work

Image registration is the process of aligning two or more frames of the same or similar scene. The basic input data to the registration process are two images: the fixed image and the moving image. Several methods for US to US registration have been proposed in literature. These approaches are either feature-based or intensity-based. Intensity-based methods compare intensity patterns in images via similarity metrics, registering either images or sub images. If sub images are registered, centers of corresponding sub images are treated as corresponding feature points.

An important ingredient in a registration framework is the choice of similarity metric. Sum of squared differences (SSD), sum of absolute differences (SAD) and

normalized cross-correlation (NCC), see [84], are simple and easy to implement. The mutual information (MI) metric by [85] has been suggested to be the most suitable metric for US to US registration in [54]. Mutual information is a powerful similarity metric, often used in multi-modal image registration, but it has high computational complexity, see [46]. In [118] a mutual information-based 3D registration was used for echocardiographs. [92] used an entropic similarity measure which is considered to be a generalization of the mutual information measure. Volume mosaicing was proposed in [65], to generate large US volumes from several acquired 3D US datasets. An offline method that registers the entire 4D sequence in a group-wise fashion was discussed in [129].

Ultrasound image acquisition is known to be affected by various factors, such as acoustic shadowing due to loss of probe contact (inadequate amount of gel) and a gamut of panel settings such as gain, time gain compensation, focus etc. While US images, because of these factors, may be of relatively poor quality, they are highly textured. A texture-based similarity measure was investigated in [34]. [16] used a combination of texture and edge features.

Feature correspondence based methods find correspondences between image features. The concept of attribute vectors has been used to define corresponding voxels in fixed and moving images in [32]. Information theoretic based feature detection was discussed in [134]. Popular 3D Scale Invariant Feature Transform (SIFT) feature descriptors were used for feature extraction in [93] and [117]. A local phase based method was employed by [41]. A hybrid feature-based registration approach using combination of local forces in a variational framework and block-matching correspondences was reported in [20].

Knowing the correspondence between a number of feature points in images, the transformation can be determined. Efficient algorithms for the correspondence problem have been an active research topic in the computational geometry and pattern recognition communities. Establishing correspondences between sparse image features in natural images using Markov random fields was discussed by [124]. One popular approach to the correspondence problem involves applying the algorithms for the more general graph matching problem, see [23]. Spectral matching is an eigenvector-based method for graph matching, see [56]. A game theoretic approach to correspondence estimation formulated as a quadratic assignment problem was presented in [3] and [111].

Intensity based registration has gained popularity over feature-based approaches in recent years. In feature-based approaches feature images are generated by extracting features from images, which are then used in subsequent analysis. From an information theoretical perspective, there is generally information loss in the process of creating a feature. Any feature constructed from a set of voxel data cannot contain more information than the same set of voxel data. Nevertheless feature based approaches are helpful when the noise in the data is significant or an invariant representation of the data is required (e.g. rotational invariance). Feature based

approaches also help in encoding image content in a way that makes it more distinctive to be used in classification or registration algorithms. A rigid registration approach that combines the properties of both intensity based and feature based approaches and is suitable for parallel implementation is the block-matching approach, see [22].

There are a few reports on real-time US registration. To address the issue of missing anatomical structures, near real-time image fusion of multiple single-view real-time 3D echocardiography using wavelet features was proposed in [106]. Real-time image based registration of 3D US images from a 4D US sequence is a challenging task. 3D frames from a 4D sequence have much lower resolution than a static 3D US image. In [117] and [98] a graphics processing unit (GPU) was employed to develop a real-time 3D US registration system. Both these methods register consecutive volumes in a 4D sequence and assume that small probe displacements occur between frames. [98] use SSD as image similarity measure and assume that the anatomical regions have echoes of similar intensity. [117] employed a feature correspondence based technique to have a real-time US to US registration.

### 2.1.3 Overview and contributions

Our ultrasound registration approach is motivated by methods described in ultrasound speckle tracking literature, see [47]. A speckle pattern contains densely positioned targets created by the interaction of ultrasonic beams and the tissue. The speckle pattern allows highly accurate (sub millimeter level) tracking of tissue motion, see [114]. Automatic speckle tracking can be performed using block-matching. In an ultrasound image speckle patterns are found in large collections. The tracking information from multiple such speckle patterns can be used to perform registration.

In an ideal scenario, if all the speckle patterns are tracked accurately, the speckle pattern correspondences can be used to estimate the transformation. However in practice not all of the speckle patterns will be tracked well, e.g. because of acoustic shadowing or due to motion decorrelation, see [76]. To address these issues and to remove false matches, we employ a matching approach, inspired by game theory, to retain only pairs that have been matched correctly. This outlier rejection is formulated in a game theoretic framework. Speed is an important aspect of our application, and a matching strategy reduces over reliance on selecting the speckle patterns.

In this work, we propose a novel, fast solution to the 3D ultrasound liver registration problem. Our contributions are fourfold: first, we integrate a fast outlier rejection approach to improve the result of a block-matching approach, second, we develop a method to use both the geometric consistency and the appearance information from block-matching to reject outliers, third, the (non-homogeneous quadratic) optimization function of the outlier rejection module is mapped to a ho-

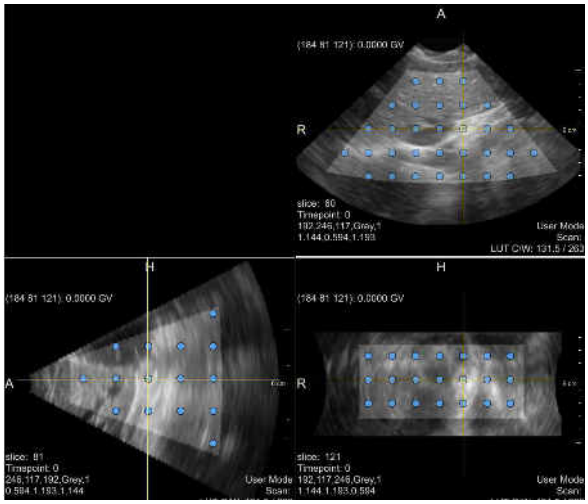



**Figure 2.1:** *Region of interest (ROI) and grid points.*

mogeneous quadratic function to be solved efficiently using replicator dynamics, and fourth, we perform an extensive evaluation on real 3D imaging data. Finally, we demonstrate that the method is able to perform registrations at 8 Hz.

## 2.2 Method

Our registration method consists of three basic steps a) naïve point selection, b) block-matching, and c) outliers rejection followed by an affine transformation using the inliers. The block-matching step uses a similarity metric to establish correspondences between the selected points in the fixed image and the moving image. The true displacements are retained and the false displacements are rejected in the outlier rejection step. The inliers are used to estimate the transformation using [122] for the affine case and [4] for the rigid case.

### 2.2.1 Naïve point selection

We sample the points used in our registration pipeline from a regular 3D grid structure. The grid structure is made up of a series of intersecting axes which are parallel to the $x$, $y$ and $z$ axis. The vertices or the junction location of the axes form the set of points used in the block-matching scheme. Before we sample points from the US volume based on the grid structure, we define a region of interest (ROI) inside the cone of the 3D US image. The region outside the ROI, see Figure 2.1, is not used in the registration process. Instead of a regular grid structure, other choices of

sampling schemes are equally applicable.

### 2.2.2 Block-matching

We employ a straightforward block-matching scheme to find matching homologous points across volumes. This is achieved by taking a block of a certain size around the voxel of interest in the fixed image and finding the homologous pixel within the block in the moving image, while searching in the corresponding neighborhood in the moving image. The size of the block in both the images is given by $B = (B_x, B_y, B_z)$. The neighborhood in the moving image is defined by the search window $\Omega = (\Omega_x, \Omega_y, \Omega_z)$. The goal is to find the corresponding (correlated) pixel that maximizes the similarity. The correspondences for all the input points from the fixed image to the points in the moving image form the input to our outlier rejection scheme.

NCC is used as a similarity metric for block-matching. Assuming an affine transformation and small rotation between the two consecutive volumes, no rotation of the block during matching is applied.

### 2.2.3 Outlier rejection

A game theoretic perspective on the matching problem is popular in literature, see [2]. These matching approaches are deduced from clustering approaches presented in [125], [100]. In these approaches an adjacency matrix (payoff matrix in game theory) is built from a graph, the vertices of the graph represent the potential correspondences and the edges embed the pairwise constraints between candidate assignments. A function of this adjacency matrix is optimized to find the cluster. The concept of dominant set as discussed in [100], is a game theoretic way of partitioning the graph. The optimization function of a dominant set formulation consists of a homogeneous quadratic term. Similar formulations are also found in spectral clustering literature, see [74]. In our work, we discuss the application of game theory in establishing potential correspondences based on (pairwise) geometric constraints and appearance information.

#### 2.2.3.1 Geometric constraint

Let $P = \{p_h\}$ and $Q = \{q_h\}$ where $0 < h \leq m$, be the set of locations from fixed volume and moving volume, respectively, having the highest degree of similarity. We have a one to one correspondence between the points in the point set $P$ and $Q$. As both the points are from volumes representing the same anatomical structure, the geometric distance between the points ought to be preserved. A point $q_h$ in the moving volume that preserves the geometric distances with most of the other points in the same set $Q$ has a higher chance of being an inlier, and vice versa. This criterion of preserving the geometric distances forms the core of our outlier

rejection scheme, similar to [124]. This information is embedded in a graph structure and is represented as an *adjacency matrix*. An adjacency matrix (**A**) represents a fully-connected undirected graph whose edges express the relative relationships, or affinities, between each pair of points in the point set, and is defined as follows:

$$f^{geo}(\mathbf{x}) = \sum_{u,v \in O} A_{u,v} x_u x_v \,, \tag{2.1}$$

where

$$A_{u,v} = e^{-\delta^2_{u,v}/2{\sigma_A}^2} \text{ and } \delta_{u,v} = \frac{|\|q_u - q_v\| - \|p_u - p_v\||}{\|q_u - q_v\| + \|p_u - p_v\|} \,.$$

$O = \{1, \cdots, m\}$ is a set of elements enumerating the bijective association between the point sets $P$ and $Q$. For $h \in O$, $x_h$ is the probability of $q_h$ being an inlier. The disparity between the pairwise distances is normalized by dividing the pairwise distance by the sum of the distance between the points. The parameter $\sigma_A$ moderates the strength of the term $\delta_{u,v}$. It is apparent from the pairwise distances that the inliers have high similarity with each other and poor similarity with the outliers.

#### 2.2.3.2 Appearance constraint

The appearance constraint is derived from the block-matching scores. This constraint favours locations that have high block-matching scores. The appearance term is defined for each point as follows:

$$f^{app}(\mathbf{x}) = \sum B_u x_u \,, \tag{2.2}$$

where

$$B_u = e^{-\gamma_u^2/2{\sigma_B}^2} \,,$$

$\gamma_u$ is the appearance term from the $u$th block, $\gamma_u = |1 - \mathrm{BM}_u|$, where $\mathrm{BM}_u$ is the block-matching score of the $u$th block. The parameter $\sigma_B$ moderates the strength of the term $\gamma_u$. Appearance and geometric constraints are combined to remove the false matches and retain true matches.

#### 2.2.3.3 Optimization

Combining the quadratic geometric constraint and the linear appearance constraint, the function to optimize for our case is:

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{A}\mathbf{x} + \lambda \mathbf{B} \cdot \mathbf{x} \,, \tag{2.3}$$

where each element $x_i$ of $\mathbf{x}$ states whether point $i$ is an inlier or not. $\mathbf{x} \in \Delta$, which is the unit simplex defined as:

$$\Delta = \{\mathbf{x} \in \mathbb{R}^m_+ : \mathbf{e}^\intercal \mathbf{x} = 1\} \,, \tag{2.4}$$

where $\mathbf{e} = [1, \ldots, 1]^\intercal \in \mathbb{R}^m$. The elements of the stochastic vector $\mathbf{x}$ are nonnegative and sum up to one.

By substituting $\mathbf{K} = \mathbf{A} + \frac{\lambda}{2}(\mathbf{e}\mathbf{B}^\intercal + \mathbf{B}\mathbf{e}^\intercal)$, Equation 2.3 can be rewritten as:

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{K}\mathbf{x}\,, \tag{2.5}$$

where $\mathbf{x} \in \Delta$, see [10]. Hence a non-homogeneous quadratic optimization function of the form of Equation 2.3 is mapped to a homogeneous quadratic optimization function of the form of Equation 2.5.

**Algorithm 1** : $\mathbf{x} \Leftarrow$ Optimize $(\mathbf{K}, \mathbf{x})$

**Require:** $\epsilon$ {Given the initial state $f_0$, $\epsilon$ determines whether the function increase sufficiently. For our application we choose $\epsilon =$ 1.0e-06.}

**Ensure:**

$\mathbf{x} \Leftarrow \mathbf{1}/m$
$f_0 \Leftarrow \mathbf{x}^T\mathbf{K}\mathbf{x}$
$f_1 \Leftarrow 0$
$f_2 \Leftarrow f_0$
**while** $\frac{(f_2 - f_1)}{f_0} > \epsilon$ **do**
  $f_1 \Leftarrow f_2$
  **for all** elements of $\mathbf{x}$ **do**
    $x_i \Leftarrow x_i \frac{(\mathbf{K}\mathbf{x})_i}{\mathbf{x}^T\mathbf{K}\mathbf{x}}$
  **end for**
  $f_2 \Leftarrow \mathbf{x}^T\mathbf{K}\mathbf{x}$
**end while**
$\mathbf{x} \Leftarrow$ Set Inliers $(\mathbf{x})$

The replicator dynamics update equation to maximize the above energy term is:

$$x_h(t+1) = x_h(t)\frac{(\mathbf{K}\mathbf{x}(t))_h}{\mathbf{x}(t) \cdot \mathbf{K}\mathbf{x}(t)}\,, \tag{2.6}$$

where $x_h$ is the $h$th term of $\mathbf{x}$, see [100]. The equation ensures that $\forall\, t, \mathbf{x}(t) \in \Delta$. The trajectory of Equation 2.6 moves towards an *asymptotically stable* state, see [135]. A fixed point of the replicator dynamics (or any dynamical system) is said to be asymptotically stable if any small deviations from that state are eliminated by the dynamics as $t \to \infty$. As matrix $\mathbf{K}$ is symmetric in Equation 2.5, the function $\mathbf{x} \cdot \mathbf{K}\mathbf{x}$ is strictly increasing along the trajectory of Equation 2.6, see [100]. Hence, we reach the (local) maxima at the location where the trajectory of Equation 2.6 does not increase any further. We initialize each element of $\mathbf{x}$ to $1/m$. The steps are presented in Algorithm 1. After convergence, the inliers are determined by thresholding the elements of $\mathbf{x}$, see Algorithm 2.

**Algorithm 2** : $\mathbf{x} \Leftarrow$ Set Inliers $(\mathbf{x})$

**Require:** $\epsilon$ {The elements of vector $\mathbf{x}$ that are outliers should ideally $\rightarrow 0$, thus we choose $\epsilon = 1.19209\text{e-}007$, which is the machine epsilon value for the floating types.}
**Ensure:**
  **for all** elements of $\mathbf{x}$ **do**
    **if** $x_i < \epsilon$ **then**
      $x_i \Leftarrow 0$ {This is an outlier}
    **else**
      $x_i \Leftarrow 1$ {This is an inlier}
    **end if**
  **end for**

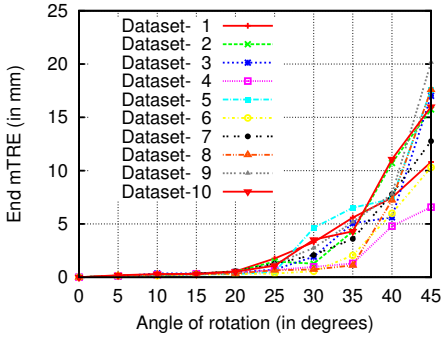



**Figure 2.2:** *Registration results on manually rotated 3D US datasets. Parameters are* $(\sigma_A, \lambda, \sigma_B) = (0.1, 0.1, 0.1)$.

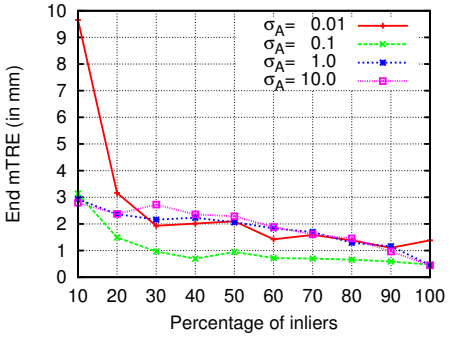



**Figure 2.3:** *Registration results on simulated datasets. The x-axis presents the percentage of inliers in the dataset.*

## 2.3 Experimental setup

### 2.3.1 Data and resources

4D US data were acquired with a Philips iU22 xMATRIX US system using the X6-1 PureWave xMATRIX transducer. A proprietary Philips protocol was implemented for digital streaming via an ethernet link, enabling high quality digital images and most scan parameters to be transferred from the iU22 xMATRIX system to our computer. The acquisition of the 4D image involves steering the US beam across the area of interest. Once the area of interest is identified, the operator keeps the probe steady to acquire continuous volumes (in this case liver volumes) over time. As the probe was kept steady while acquiring each dataset, the dataset contains a series of liver volumes capturing the liver motion due to breathing. Fifteen 4D US vol-

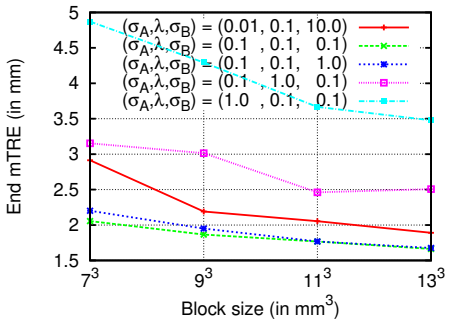



**Figure 2.4:** *Registration error for different block sizes. The number of grid points is* 194.

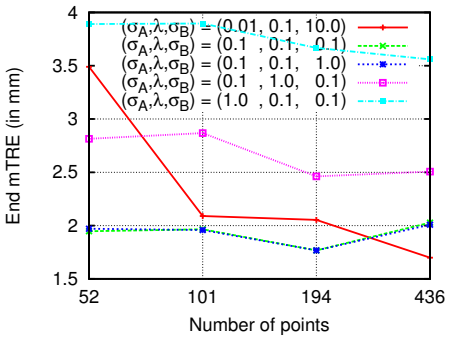



**Figure 2.5:** *Registration error for different number of grid points. The block size is* $11^3$ *mm*$^3$.

ume sequences were acquired from eight healthy volunteers. No specific breathing protocol (i.e. free breathing) was used during image acquisition. During manual annotations for ground truth, we found that two of the 4D ultrasound sequences were hard to annotate, because the ultrasound volume contained a small part or corner of the liver. It was difficult to find locations in the ultrasound volume that would form reliable landmarks. We discard these two 4D ultrasound sequences from the evaluation. The transducer was placed axial mid and coronal while acquiring the scans. The 3D+t US volumes were $192 \times 246 \times 117$ voxels with voxel size $1.144 \times 0.594 \times 1.193$ mm and acquired at 6 Hz using the proprietary protocol. From the 4D US sequences, pairs of frames were selected in a systematic way such that they are representative of the whole breathing cycle. The following pairs: $(3, 33)$, $(4, 29)$, $(5, 25)$, $(6, 21)$, $(7, 17)$, $(8, 13)$ and $(9, 10)$ were chosen from each of the thirteen 4D US sequences. Hence in total 91 pairs of US volumes were used for evaluation.

The code was implemented in C++ and MeVisLab. The block-matching and outlier rejection were additionally implemented in OpenCL, to be run on a GPU. The GPU implementation was run on a NVIDIA GTX 780 Ti GPU.

### 2.3.2 Parameter setting

To have an experimental setup invariant to the input data pixel spacing, we work in the world co-ordinate space (in mm).

*Naïve point selection :* It is obvious that with an increase in the number of sample points the number of true correspondences will increase. However, for practical purpose the number of sample points should be kept low. We restrict the number of sample points by fixing the grid spacing for the sample points grid. Notice that with the ROI the number of sample points will be less than if there was no ROI.

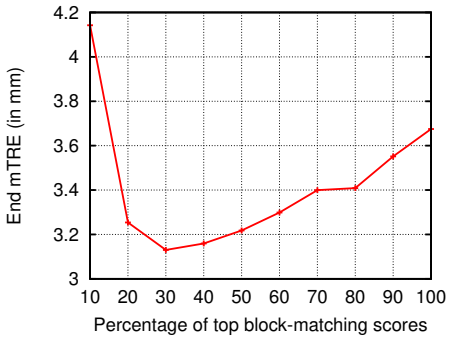



**Figure 2.6:** *Similarity-based algorithm. The number of grid points is* 194 *and the block size is* $11^3$ *mm*$^3$.

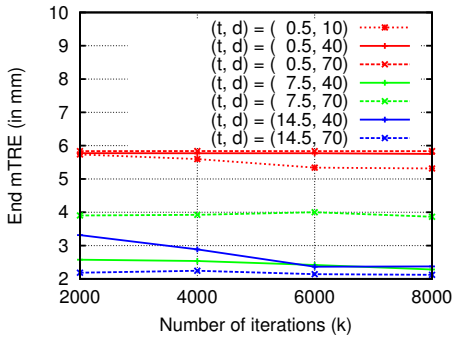



**Figure 2.7:** *RANSAC performance. The number of grid points is* 194 *and the block size is* $11^3$ *mm*$^3$.

Grid spacings of 22 mm, 18 mm, 14 mm and 11 mm are explored. Grid spacing is inversely related to the number of grid sample points. The corresponding number of grid points for the mentioned grid spacing are 52, 101, 194 and 436, respectively.

*Block-matching :* We choose the number of voxels in each dimension such that the block size is isotropic in the world co-ordinate space. Block sizes of $7^3$ mm$^3$, $9^3$ mm$^3$, $11^3$ mm$^3$ and $13^3$ mm$^3$ are investigated. The search region is selected such that it is large enough to capture the shift in the anatomy. For our registration experiments the search region is $\Omega = 40^3$ mm$^3$, i.e. 20 mm in all directions. We use normalized cross-correlation similarity metric for block-matching.

*Outlier rejection :* The respective parameters for our method are $\sigma_A$, $\lambda$ and $\sigma_B$. In the experiments, the evaluation range for each of the parameters is $\{0.01, 0.1, 1.0, 10.0\}$.

### 2.3.3 Evaluation metric and ground truth

On the 91 pairs of US volumes, ground truth registration was established using manual annotations. Let $g_i^F$ and $g_i^M$ be an annotation pair for the fixed and the moving image, respectively. We define a mean target registration error (mTRE) to determine the 3D registration error, where the "targets" in the mTRE calculation are the annotations in the moving image. The average distance between the points defines the mTRE:

$$mTRE = \frac{1}{n}\sum_{i=1}^{n} \| g_i^M - \mathcal{T}_{\text{reg}} g_i^F \|,$$

where $\mathcal{T}_{\text{reg}}$ is the transformation determined by a registration algorithm and $n$ being the number of annotations.

### 2.3.4 Experiments

We perform a series of experiments to evaluate the proposed method.

#### 2.3.4.1 Block-matching and outlier rejection

Block-matching may suffer from (large) rotational motion. To this end, we first evaluate the robustness of the block-matching (with outlier rejection) approach under controlled rotational transform of ultrasound data. We used ten 3D ultrasound volumes and rotated the dataset in the range from 0 to 45 degrees with a step size of 5 degrees. The axis of rotation was $(x = 1, y = 1, z = 1)$. All the rotated datasets paired with their unrotated counterpart, form the input to our experiment. Block size of $11^3$ mm$^3$ and 194 grid points was used for this experiment.

In the next experiment we evaluate the outlier rejection based on geometric consistency using simulated data. We study the performance of the module given certain percentages of inliers that are present in the correspondences from the block-matching scheme. To simulate this scenario, we select a set of points (200 sample points) using the grid. A known translation $(x = 3, y = 9, z = 5)$ and rotation angle $r = 45$ degrees along the axis $(x = 1, y = 1, z = 1)$, was applied to the point set. The transformed point set is our ground truth. The point set is then partitioned randomly into two subsets, namely the inliers and the outliers. As the inliers ought to be close to the ground truth, small random displacements are introduced to the set of inliers. The displacement interval of the co-ordinates of the inliers is given by $\delta^{inliers} = [0, 3]$ mm. Similarly, large random displacements are introduced to the outliers set. This simulates the situation in which outliers are randomly distributed with respect to the target position. The displacement interval of the co-ordinates of the outliers is given by $\delta^{outliers} = [3, 20]$ mm. We thus simulate a realistic situation with inliers and outliers representative for the output of the block-matching step. As we evaluate only the geometric consistency term of the outlier rejection module, $\lambda$ value is set to zero in Equation 2.3. The adjacency matrix is constructed and evaluated for $\sigma_A$ values of $\{0.01, 0.1, 1.0, 10.0\}$, see Equation 2.1. The percentage of inliers is gradually varied from 0% to 100%.

#### 2.3.4.2 Parameter tuning

We investigate the effect of changing block-size and grid spacing. We experiment with various block sizes keeping the grid spacing fixed. Next we analyze whether increasing the number of sample points improves the registration results, given a fixed block-size.

As we narrow down on the appropriate block size and the grid spacing for our application, we are interested in estimating the optimal $\sigma_A$, $\lambda$ and $\sigma_B$ values, see Equation 2.1, 2.2 and 2.3. This would complete our parameter optimization step. To estimate the optimal parameters we first evaluate our registration approach.

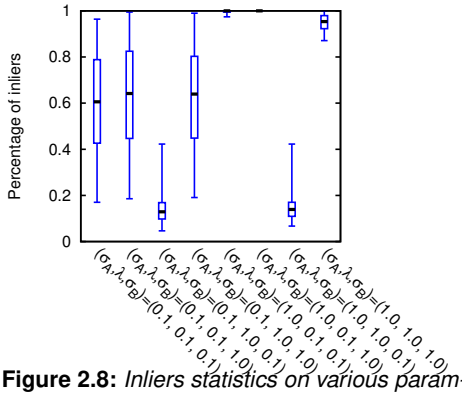



**Figure 2.8:** *Inliers statistics on various parameter settings of our method. The number of grid points is* 194 *and the block size is* $11^3$ *mm*$^3$*. Optimal setting is* $(\sigma_A, \lambda, \sigma_B) = (0.1, 0.1, 0.1)$.

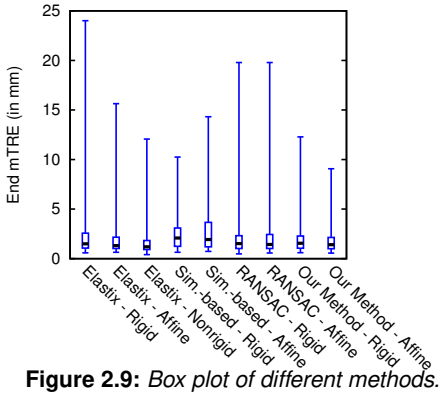



**Figure 2.9:** *Box plot of different methods.*

We use a leave-one-out cross-validation scheme to assess how the results will generalize to an independent data set. One round of leave-one-out strategy involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). As the 4D US sequences were acquired from eight different subjects, our testing set consists of image pairs from a single subject, and the remaining datasets for the training set. The cross-validation scheme is iterated over all subjects. In each of the iterations, optimal parameters are estimated over the training set (N-1 subjects) and evaluated on the testing set (1 subject). The validation results are averaged over all rounds to estimate the overall error. We report the performance of our approach using this leave-one-out cross-validation scheme. This provides a conservative estimate of the performance of our method. The parameter set that appears most often (mode) across the leave-one-out cross-validation rounds is chosen as the optimal parameter. Given the optimal parameter setting we compare the registration results of our method to the initial displacement between the fixed and moving image based on the ground truth. We report the registration error.

#### 2.3.4.3 Comparison with other methods

We compare the registration results of the proposed method with three other methods: Elastix, similarity-based outlier rejection and RANSAC.

1. *Elastix* (rigid, affine and B-spline transformation models), see [61]: is an intensity based medical image registration approach. The used parameter settings can be downloaded from the Elastix parameter file database, see [62]. The pa-

rameter file uses MI similarity metric. We evaluate the performance of Elastix for all the three transformation models.

2. *Similarity-based outlier rejection* (rigid and affine transformations): uses the same block-matching results, but selects as inliers the block-matching results with the best similarity scores. The number of inliers is a percentage of the total number of points. We follow the same leave-one-out strategy as for the proposed method to determine the optimal parameters of the algorithm. We report the resulting registration error.

3. *Geometry-based outlier rejection via RANdom SAmple Consensus (RANSAC)* (rigid and affine transformations), see [31], [33]: is an iterative method to estimate parameters of a mathematical model which describes the observed data containing outliers. RANSAC has the following parameters: $t$ (in mm) - a distance threshold value for determining when a data point fits a model, $d$ - the number of data values required to assert that a model fits well to data, $r$ - the minimum number of data values required to fit the model and $k$ - the maximum number of iterations allowed in the algorithm. We follow the same leave-one-out strategy as for the proposed method to determine the optimal parameters of the algorithm. We report the resulting registration error.

We also evaluate and report the GPU execution time of the OpenCL code for the block-matching and the outlier rejection modules.

### 2.3.5 Results

The presented results of the above listed experiments are enlisted below.

#### 2.3.5.1 Block-matching and outlier rejection

Figure 2.2 shows the registration error on the controlled dataset with respect to the ground truth. The mTRE value for the ten datasets are below 2 mm for rotations up to 25 degrees of the US volumes. For rotations larger than 25 degrees the error significantly increases.

Figure 2.3 shows the performance of the outlier rejection module for various percentage of inliers and for various parameter settings. For $\sigma_A = 0.1$ the mTRE is small in case of 20% or more of inliers.

#### 2.3.5.2 Parameter tuning

Figure 2.4 shows the performance of the method for various block sizes, ranging from $7^3$ mm$^3$ to $13^3$ mm$^3$. Figure 2.5 shows the performance of the method with various grid spacings keeping the block size fixed. The registration results improve with the increase in the number of sample points. A block size of $11^3$ mm$^3$ and 194 grid points provides good results. Decreasing the grid spacing or increasing the

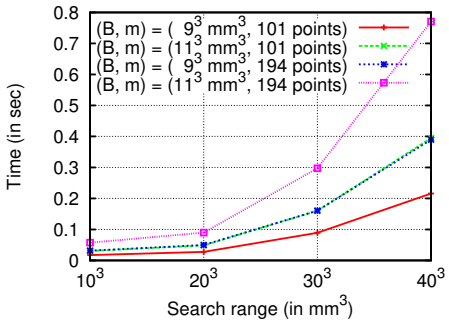



**Figure 2.10:** *GPU timings for block-matching (B - Block Size, m - Number of points).*

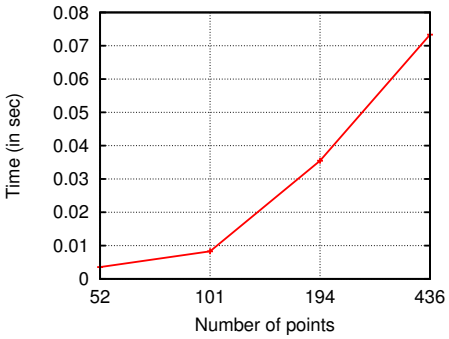



**Figure 2.11:** *GPU timings for outlier rejection.*

block size does not result in a significant gain in performance. Accordingly, we use a block size of $11^3$ mm$^3$ and 194 grid points for the remaining studies.

Figure 2.4 and Figure 2.5 show that the registration results are stable across the selected parameter range. Table 2.1 shows the parameter optimization for the function in Equation 2.3, with and without appearance term, for both rigid and affine registrations. In Table 2.1 each row corresponds to a leave-one-out cross-validation round per subject. We choose the parameter $\sigma_A$, $\lambda$ and $\sigma_B$ values that appears most consistently across the subjects in Table 2.1 as our optimum parameter value. For our method-*Affine* the chosen parameter setting is $(\sigma_A, \lambda, \sigma_B) = (0.1, 0.1, 0.1)$. Averaging over all the subjects gives a leave-one-out registration error of 1.83 mm, for 194 grid points and block size of $11^3$ mm$^3$, see Table 2.3. Figure 2.8 shows the statistics on the number of inliers generated by the outlier rejection module.

#### 2.3.5.3 Comparison with other methods

1. *Elastix* (rigid, affine and B-spline transformation models): Figure 2.9 shows the box plot of the registration errors for Elastix. Table 2.3 shows the performance of the Elastix registration method for rigid, affine and B-spline transformation models.

2. *Similarity-based outlier rejection* (rigid and affine transformations): Figure 2.6 shows the performance of the approach for various percentages of the block-matching results for affine transform. In Table 2.2 we report the leave-one-out cross-validation estimating the optimal parameters. The registration results of the method for rigid and affine transforms in Table 2.3. The overall optimal parameter setting for similarity-based-*Affine* approach is the top 30%, see Table 2.2.

3. *Geometry-based outlier rejection via RANdom SAmple Consensus (RANSAC)* (rigid

and affine transformations): As the number of unknowns of the transformation to estimate for the affine case is $(n^2 + n) = 12$, where $n$ is the dimension of the data point, we require greater than or equal to four 3D points to fit an affine model. Accordingly, we use a $r$ value of four 3D points. $t$ was evaluated for distances from 0.5 mm to 14.5 mm. The parameter $d$ which determines the number of data values required to assert whether a model fits well to data, is dependent on how many correspondences are found in the image and the anatomy being imaged. Parameter $d$ was varied between 10 to 70 points. Figure 2.7 shows the performance of the approach for various settings of $t$, $d$ and $k$ for affine transform. We fix the maximum number of iterations to $k = 6000$, as increasing the number of iterations does not result in a significant gain in performance, see Figure 2.7. In Table 2.2 we report the leave-one-out cross-validation estimating the optimal setting for the rest of the parameters. We then report the performance of the method for rigid and affine transforms in Table 2.3. Parameters $(t, d) = (13.5\text{ mm}, 60\text{ points})$ is the optimal setting for RANSAC-*Affine*, see Table 2.2.

Figure 2.9 shows the box plot of the registration errors for different methods. The leave-one-out error of the different methods are shown in Table 2.3. For the given dataset, the Elastix approach with a B-spline transformation model has the smallest registration error, and next best is the proposed method with an affine transformation model. We also tested if the performance of our method-*Affine* is statistically significantly different from other registration methods using a two-sided paired t-test. For determining statistically significant different results we use a threshold of 0.05 for the p-value of the t-test. The paired t-test shows that our method-*Affine* is statistically significantly different from all the other methods except for the RANSAC-*Rigid*, Elastix-*Affine* and Elastix-*B-spline* methods, see Table 2.3.

The GPU execution time of the OpenCL code for the block-matching and the outlier rejection modules is shown in Figure 2.10 and Figure 2.11, respectively. For NCC similarity metric, block size of $11^3$ mm$^3$, 194 grid points and search range of $20^3$ mm$^3$, the block-matching and outlier rejection take 0.08 and 0.04 seconds, respectively. Hence with GPU implementation the registration runs at 8 Hz, for the given parameter settings. Some representative registration results are shown in Figure 2.12.

## 2.4 Discussion and conclusion

In this work we presented a 3D US affine registration approach. The approach consists of three steps namely a) naïve point selection, b) block-matching and c) outlier rejection followed by a least square affine registration of the inliers. The point selection is performed by sampling points from a 3D grid. A one-to-one correspondence between the points in the fixed and the moving image is established in the

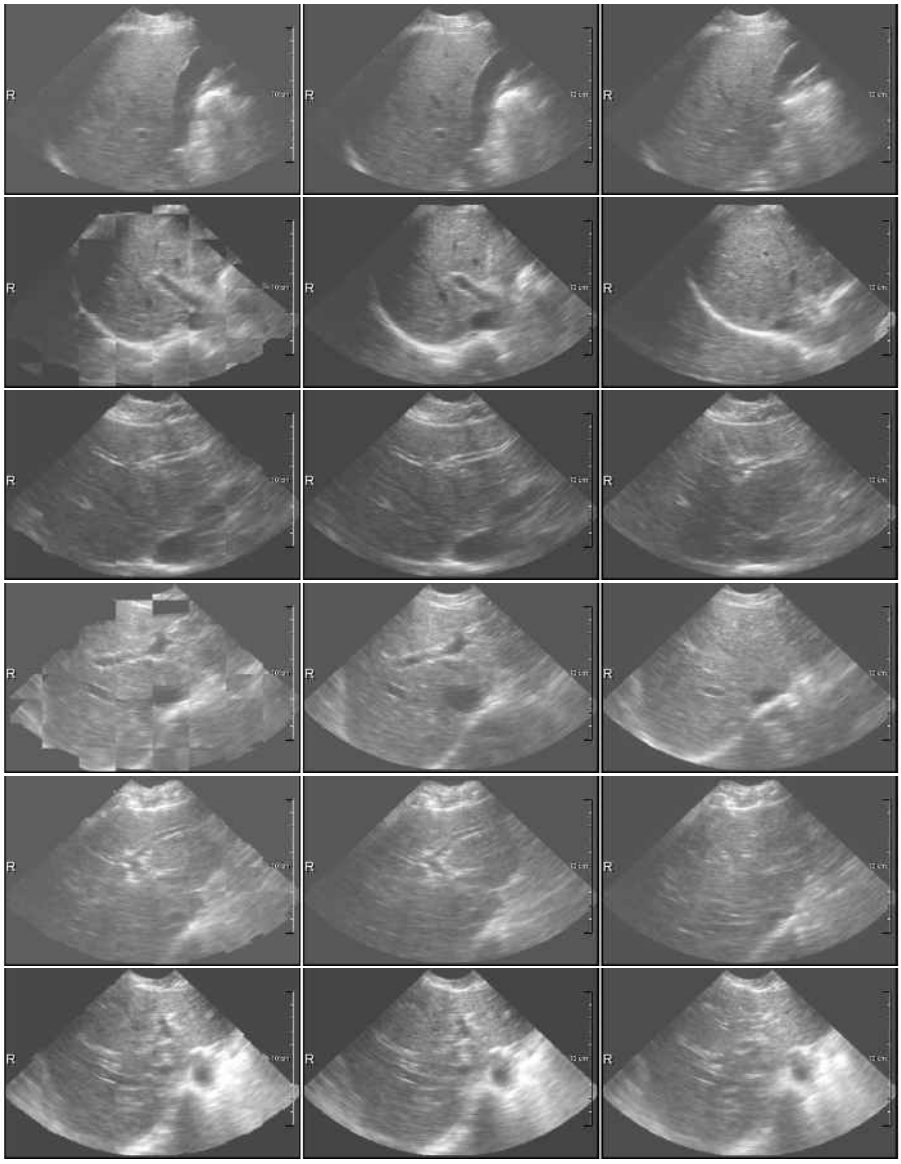


**Figure 2.12:** *Registered Volumes: Left - Registration result in checker box view, Middle - Fixed image, Right - Moving Image.*

block-matching step. The block-matching step potentially yields many false correspondences. An outlier rejection module is hence applied to reject the false matches

and retain the true matches. The outlier rejection module employs geometric consistency and appearance information to find the inliers. The framework maps a non-homogeneous quadratic optimization problem into a homogeneous quadratic optimization problem and solves it efficiently using replicator dynamics. The outlier rejection module which solves a clustering problem is a generic method and can be applied to various other similar problems. The GPU execution time as shown in Figure 2.10 and Figure 2.11, implies that for our given application and with an US frame rate of 6 Hz, our registration approach can operate in real-time.

Compared with other imaging modalities, such as computed tomography (CT) and magnetic resonance imaging (MRI), ultrasound transducers are mobile. They offer high user flexibility, however the field of view is small and usually a large organ like the liver is only partly visible. Hence for our application we assume a affine transformation between subsequent frames. Further, as we focus on a fast implementation approach, in a affine registration model, interpolation and resampling operations are computationally less intensive compared to a non-rigid registration.

In our application, the probe position is fixed. We demonstrated that our approach is robust to motion induced liver rotation. Our registration approach is tolerant to rotation up to 25 degrees, as shown in Figure 2.2. However, during complicated surgical procedures there could be large rotation between consecutive frames due to large probe displacement or change in patient position. Such motion of the probe or the patient can be accommodated for using an optical or electromagnetic tracker, see [11]. Furthermore the block-matching step is not designed to handle large rotations. To handle large rotations, block-matching can be performed on a feature space image which is rotationally invariant, see [120].

In order to show the robustness and practical applicability of our method, we systematically select two frames from the complete sequence and show the performance of our method. A leave-one-out experiment was performed on 91 pairs of US datasets with a performance of 1.8 mm, see Table 2.1. With the voxel size of the 4D US dataset being $1.144 \times 0.594 \times 1.193$ mm the voxel diagonal was estimated to be 1.8 mm, hence the error of 1.8 mm is around single voxel diagonal length.

Comparisons with other reports are hard to interpret, as different data and evaluation criteria have been used. In [54], for datasets with a voxel resolution of 0.6 mm in each dimension, they show a registration error of $2.99 \pm 1.54$ mm using the mutual information metric by Mattes. For ultrasound guided prostate biopsy application [57] found mTREs of $2.13 \pm 0.80$ and $2.09 \pm 0.77$ mm for rigid and non-rigid techniques, respectively.

[117] and [93] used a feature-based approach. The similarity between matched features is estimated by finding the similarity between the underlying transformations. The matching problem is posed as a robust parameter estimation problem and is solved using RANSAC, see [31]. RANSAC differs from our approach as it requires the setting of problem specific thresholds. Apart from geometric consistency criteria, our graph based approach utilizes appearance information to find

the outliers.

We also compared our method with other approaches, see Figure 2.9 and Table 2.3. The registration error of our method-*Affine* is lower than other methods except Elastix-*B-spline*, see Table 2.3, though the difference with RANSAC-*Rigid* and Elastix-*Affine* was not statistically significant in a paired t-test. Additionally, Figure 2.9 shows that, compared to other approaches, our method-*Affine* has lower variability. This indicates the robustness of our approach. The simulation experiment results shown in Figure 2.3 indicate that the outlier rejection module works efficiently for percentage of inliers $\geq 20\%$.

It is interesting to note that for the similarity-based algorithm and the RANSAC algorithm the affine versions perform poorer than the rigid versions of the methods, see Table 2.3. This shows that RANSAC which is purely based on the geometric locations of the points and does not employ image intensity information explicitly in its formulation, does not extend well to estimate transformations like affine. However in our method and in Elastix, the affine versions perform better than the rigid versions of the algorithm, see Table 2.3. In both Elastix and our method, intensity information (or appearance information in our case) is used in the formulation which helps in estimating the deformation. The Elastix-*B-spline* performs better than its affine version, suggesting that there is some non-rigid deformation. Figure 2.9 shows that Elastix-*Rigid* has large variations in the registration outcomes. Intensity based registrations may sometime perceive acoustic shadows in US as large structures and try to align them resulting in false registration results.

Ultrasound image acquisition is known to be operator dependent, and to be affected by various factors, such as acoustic shadowing due to loss of probe contact (inadequate amount of gel) and a gamut of panel settings such as gain, time gain compensation, focus etc. These factors lead to rapid changes in image quality. Hence image similarity metrics are a critical element of an ultrasound registration problem. A large number of similarity metrics have been proposed in the medical image community. Unfortunately there are no clear rules about how to select the most suitable metric, other than trying some of them in different conditions. US specific similarity metrics by modeling the speckle statistics in US are discussed in [131], [21], but they are much more computationally intensive. In order to have a fast implementation we choose to use normalized cross-correlation (NCC) similarity metrics. However, any other choices of similarity metric like the one mentioned by [131], can be used as one of the similarity metric in the block-matching module.

Figure 2.10 shows the runtime of the block-matching module on a GPU. For the block size of $11^3$ mm$^3$ and 194 grid points, the runtime reduces by half if either the block size is reduced from $11^3$ mm$^3$ to $9^3$ mm$^3$ or the number of grid points is reduced from 194 points to (approximately half) 101 points. The runtime increases exponentially with increase in search range. In the outlier rejection module the runtime increase approximately four times on GPU, when the number of grid points is increased from 101 points to (approximately double) 194 points, as shown in Fig-

ure 2.11.

There are several avenues of future work that would potentially improve our method. Incorporating a motion model for better estimating the position and temporal consistency of the US volume could be useful. The work could also be extended to a non-rigid registration technique. We intend to address both issues in future work.

In conclusion, we proposed and evaluated a US to US registration approach for a robust registration of liver volumes. A mean error of 1.8 mm is achieved for 91 non-consecutive pairs of 3D ultrasound volumes acquired from 13 4D ultrasound sequences. The registration approach is modular and each module has very few parameters. Registration results are stable with respect to changes in these parameters. Additionally we demonstrate that a GPU implementation of our registration approach can be used in real-time.

**Table 2.1:** *Leave-one-out cross-validation for our method. Seven pairs of 3D US volumes are extracted from each of the 4D US sequences in a systematic way such that they are representative of the whole breathing cycle. The time points of the pairs are:* $(3,33), (4,29), (5,25), (6,21), (7,17), (8,13)$ *and* $(9,10)$*. The number of grid points is* $194$ *and the block size is* $11^3$ *mm*$^3$*.*

| | | | $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ (*without appearance term*) | | | | $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{B}\mathbf{x}$ (*with appearance term*) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rigid | | Affine | | Rigid | | Affine | |
| Subject | No. of 4D US seq. | No. of 3D US pairs | Param ($\sigma_A$) | Error (in mm) | Param ($\sigma_A$) | Error (in mm) | Params ($\sigma_A, \lambda, \sigma_B$) | Error (in mm) | Params ($\sigma_A, \lambda, \sigma_B$) | Error (in mm) |
| 1 | 1 | 7 | 0.1 | 3.61 | 0.1 | 3.66 | (0.1 , 1.0 , 0.01) | 3.14 | (0.1, 0.1, 0.1) | 3.07 |
| 2 | 2 | 14 | 0.1 | 1.58 | 0.1 | 1.56 | (0.1 , 0.01, 1.0 ) | 1.58 | (0.1, 0.1, 1.0) | 1.55 |
| 3 | 2 | 14 | 0.1 | 2.25 | 0.1 | 1.98 | (0.1 , 0.01, 1.0 ) | 2.25 | (0.1, 0.1, 1.0) | 1.97 |
| 4 | 2 | 14 | 0.1 | 2.32 | 0.1 | 2.29 | (0.01, 0.1 ,10.0 ) | 2.96 | (0.1, 1.0,10.0) | 2.39 |
| 5 | 2 | 14 | 0.1 | 1.43 | 0.1 | 1.33 | (0.1 , 0.1 , 0.1 ) | 1.48 | (0.1, 0.1, 0.1) | 1.37 |
| 6 | 2 | 14 | 0.1 | 1.66 | 0.1 | 1.56 | (0.01, 1.0 ,10.0 ) | 1.71 | (0.1, 0.1, 0.1) | 1.58 |
| 7 | 1 | 7 | 0.1 | 1.83 | 0.1 | 1.79 | (0.1 , 0.01, 1.0 ) | 1.83 | (0.1, 0.1, 0.1) | 1.79 |
| 8 | 1 | 7 | 0.1 | 0.95 | 0.1 | 0.86 | (0.01, 1.0 ,10.0 ) | 1.16 | (0.1, 0.1, 0.1) | 0.91 |
| Leave-one-out error (*mean*) | | | | 1.95 | | 1.88 | | 2.01 | | 1.83 |
| Optimal Parameter (*mode*) | | | 0.1 | | 0.1 | | (0.1 , 0.01, 1.0 ) | | (0.1, 0.1, 0.1) | |

**Table 2.2:** *Leave-one-out cross-validation for Similarity-based Algorithm and RANSAC. Seven pairs of 3D US volumes are extracted from each of the 4D US sequences in a systematic way such that they are representative of the whole breathing cycle. The time points of the pairs are:* $(3,33)$, $(4,29)$, $(5,25)$, $(6,21)$, $(7,17)$, $(8,13)$ *and* $(9,10)$*. The number of grid points is* $194$ *and the block size is* $11^3$ *mm*$^3$.

| | | | Similarity-based Algorithm | | | | RANSAC | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rigid | | Affine | | Rigid | | Affine | |
| Subject | No. of 4D US seq. | No. of 3D US pairs | Param (%) | Error (in mm) | Param (%) | Error (in mm) | Params $(t,d)$ | Error (in mm) | Params $(t,d)$ | Error (in mm) |
| 1 | 1 | 7 | 10 | 3.73 | 30 | 5.26 | ( 9.5, 30) | 2.47 | (13.5, 60) | 3.23 |
| 2 | 2 | 14 | 10 | 2.19 | 30 | 2.24 | ( 9.5, 30) | 1.32 | (13.5, 60) | 1.41 |
| 3 | 2 | 14 | 10 | 2.98 | 40 | 4.99 | ( 9.5, 30) | 2.34 | (12.5, 50) | 2.68 |
| 4 | 2 | 14 | 10 | 3.32 | 30 | 4.34 | ( 8.5, 40) | 4.14 | ( 9.5, 50) | 4.84 |
| 5 | 2 | 14 | 10 | 2.11 | 30 | 2.14 | ( 9.5, 30) | 1.45 | (13.5, 60) | 1.60 |
| 6 | 2 | 14 | 10 | 2.41 | 30 | 2.69 | (10.5, 30) | 1.98 | (13.5, 60) | 1.77 |
| 7 | 1 | 7 | 10 | 1.99 | 30 | 1.90 | ( 9.5, 30) | 1.83 | (13.5, 60) | 1.76 |
| 8 | 1 | 7 | 10 | 1.36 | 30 | 1.44 | ( 9.5, 30) | 0.93 | (13.5, 60) | 1.01 |
| Leave-one-out error (*mean*) | | | | 2.51 | | 3.13 | | 2.06 | | 2.29 |
| Optimal Parameter (*mode*) | | | 10 | | 30 | | ( 9.5, 30) | | (13.5, 60) | |

**Table 2.3:** *Comparison with different methods. The number of grid points is* 194 *and the block size is* $11^3$ $mm^3$.

| Method | Sim. Metric | Error (in mm) | Paired t-test w.r.t. our method |
|---|---|---|---|
| Elastix-*Rigid* | MI | 2.30 | 0.0231 |
| Elastix-*Affine* | MI | 2.18 | 0.0646 |
| Elastix-*B-spline* | MI | 1.80 | 0.9799 |
| Sim.-based-*Rigid* | NCC | 2.51 | 0.0000 |
| Sim.-based-*Affine* | NCC | 3.13 | 0.0000 |
| RANSAC-*Rigid* | NCC | 2.06 | 0.1202 |
| RANSAC-*Affine* | NCC | 2.29 | 0.0168 |
| Our Method-*Rigid* | NCC | 2.01 | 0.0029 |
| Our Method-*Affine* | NCC | 1.83 | NA |

# Chapter 3

## 4D Liver Ultrasound Registration

Based on:

Jyotirmoy Banerjee, Camiel Klink, Edward D. Peters, Wiro J. Niessen, Adriaan Moelker and Theo van Walsum, 4D Liver Ultrasound Registration, *WBIR*, 2014.

**Abstract**

In this paper we present a rigid registration approach for 4D ultrasound (US) datasets, where images are registered over time. The 3D registration approach preceding the 4D registration consists of two main steps - block-matching and outlier rejection. The outlier rejection step removes the spurious matchings' from the block-matching module and ensures inverse consistency. For 4D registration, we perform registration of consecutive US volumes over the time series. Transformation between any two frames is estimated by taking the product of all the intermediate transforms. To avoid accumulation of error over the series of transformations, a long range feedback mechanism is proposed. A mean total registration error of 1 mm is achieved across six 4D ultrasound sequences of human liver with an execution speed of 10 Hz.

## 3.1 Introduction

*Motivation :* Ultrasound (US) is a unique imaging modality. Unlike computed tomography (CT) and magnetic resonance imaging (MRI), it is mobile and real-time. This is a desired combination in diagnostic and interventional setup. With the advent of 4D ultrasound, volumes of human anatomy can be visualized in real-time. Interoperative imaging using 4D ultrasound has huge potential in minimally invasive surgery of the liver. Image registration is a basic requirement in these applications and they aid in image stabilization for better visualization. A group wise 4D registration approach takes a stack of US volumes to perform the registration over the time series [129]. This approach is benefited from looking at US volumes in hindsight and is suited for offline processes as they have high computational and storage cost. A more dynamic approach would be to register images in streaming 4D US data. In this scenario, registrations are required to be up to date until the current time point, appending the registration results of the subsequent US frame. Given the registration results for the left half of the time axis the challenge is to move forward in time, keeping the registration up to date. A typical 3D US registration method when extended in the time domain is likely to face the following challenges -

- Due to motion (probe, patient or breathing), the region of interest might undergo large displacements, resulting in small overlap between the US frames.
- Over time small errors in the 3D registration could accumulate, yielding widely diverging outcomes.
- Continuous input stream of US volumes induces heavy computational and storage burden on the registration approach.

*Related work :* Image registration is the process of determining the geometrical transformation that aligns the moving image to the fixed image. 4D registration extends this notion in the temporal domain. Group wise 4D US registration was addressed by Vijayan et al. [129], where spatial and temporal smoothness of the transformations are enforced by using a temporal free-form deformation (TFFD) model. Shi et al. [119] extend the TFFD model with a sparse representation and use it to recover smooth motion from time sequence of cardiac US images. Øye et al. [98] propose a method to perform real time image registration on streaming 4D ultrasound data, and use it to deduce the positioning of each ultrasound frame in a global coordinate system. In this chapter, the 3D registration framework which is precursor to our 4D registration approach is part of an existing work submitted to a journal. In this work we extend the previous registration framework, to address the issues related to the 4D US registration problem.

*Our Contributions :* In this chapter we present a 4D registration approach which unlike the group wise approaches performs registration dynamically. First, to max-

imize the chance of overlap between the contents of the frames, we register consecutive frames in time. In order to have a robust registration an inverse consistency criteria is enforced. This ensures consistency between the forward backward transforms. Second, the inverse consistency criteria helps the preprocessing (which in our case is the block-matching scheme) done in a parallel fashion. Third, to neutralize or reduce the accumulation of registration errors over the time series, we propose a feedback mechanism over a time gap.

## 3.2 Method

### 3.2.1 3D registration

The 3D registration approach is based on block-matching [22]. For a collection of points from the fixed images the block-matching gives a set of correspondences in the moving image. As US images are poor in quality, the correspondences may have lots of spurious matches. We remove the false matches using an outlier rejection module. A game-theoretic matching approach is employed to reject the false matches [6]. Our contribution in this chapter is to extend the method to forward-backward (or inverse consistency based) registration. The true matches are further used to estimate the rigid transformation using Arun's et al. least-squares registration algorithm [4].

### 3.2.2 4D registration

Let $P = \{p_i\}$ and $Q = \{q_i\}$ where $0 < i \leq n$, be the set of locations from fixed volume and moving volume, respectively. We have a one to one correspondence between the point sets from the block-matching. Let a mapping $M \subseteq P \times Q$ represent potential correspondence from the point set $P$ to $Q$; and similarly in the reverse direction a mapping $N \subseteq Q' \times P'$ represent potential correspondence from the point set $Q' = \{{q'}_i\}$ to $P' = \{{p'}_i\}$. Ideally if all the points are tracked well both in the forward and the backward directions then $\forall i \; {p'}_i = p_i \Rightarrow {q'}_i = q_i$ and vice-versa (*Inverse consistency*). Note that we choose the point sets $P$ and $Q'$, independently. This has two advantages, first, inverse consistency is not enforced explicitly, and second, the block-matching can be executed in parallel. Later we show how we incorporate the inverse consistency in an implicit way. Further as the point sets are from volumes representing the same anatomical structure, the geometric distance between the points should be preserved (*Geometric consistency*). A point $q_i$ in the moving volume that preserves the geometric distances with most of the other points in the same set $Q$ and their corresponding $p_i$ and ${p'}_i$ are in close proximity, have a better chance of being an inlier. The criteria of preserving the geometric distances, similar to [124] and the inverse consistency criteria forms the basis of our outlier rejection scheme.

The geometric consistency information can be embedded in a graph structure and can be represented as two affinity matrices (*forward*):

$$A_{i,j} = \begin{cases} e^{-\delta_{ij}^2/2\sigma^2} & \text{if } i \neq j \\ 0 & \text{else} \end{cases}, \tag{3.1}$$

where $\delta_{ij} = (\|q_i - q_j\| - \|p_i - p_j\|)$; and (*backward*),

$$B_{i,j} = \begin{cases} e^{-\delta'^2_{ij}/2\sigma^2} & \text{if } i \neq j \\ 0 & \text{else} \end{cases}, \tag{3.2}$$

where $\delta'_{ij} = (\|p'_i - p'_j\| - \|q'_i - q'_j\|)$.

For ensuring an inverse consistency in an implicit way, we combine the forward and backward block-matching information into a single graph. This is done by combining the matrix $A$ and $B$ into a symmetric matrix $G$ as follows (*forward-backward*):

$$G = \begin{bmatrix} A & | & C^T \\ - & & - \\ C & | & B \end{bmatrix}, \tag{3.3}$$

where the matrix $C$ is given as:

$$C_{i,j} = e^{-\delta''^2_{ij}/2\sigma^2}, \tag{3.4}$$

where $\delta''_{ij} = (\|q_i - q'_j\| - \|p_i - p'_j\|)$.

Given a vector $\mathbf{x}$ defining the probability of points in $\{P, Q'\}$ being an inlier, we try to find a vector $\mathbf{x}$ that maximizes $\mathcal{F}$:

$$\max \mathcal{F}(\mathbf{x}) = \mathbf{x} \cdot G\mathbf{x} \qquad \text{subject to} \quad \mathbf{x} \in \Delta, \tag{3.5}$$

where $\Delta = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ and } \sum_{i=1}^n x_i = 1\}$. Finding the internal nodes or the inliers corresponds to the notion of a dominant set [100].

Equation 3.5 can be optimized (local optimum) using replicator dynamics [100][135]. The replicator dynamics update equation to maximize a energy term of the form $\mathbf{x} \cdot G\mathbf{x}$, subject to $\mathbf{x} \in \Delta$ is:

$$x_i(t+1) = x_i(t) \frac{(G\mathbf{x}(t))_i}{\mathbf{x}(t) \cdot G\mathbf{x}(t)}, \tag{3.6}$$

where $x_i$ is the $i$th term of $\mathbf{x}$. The equation ensures that $\forall t, \mathbf{x}(t) \in \Delta$. Equation 3.1, Equation 3.2 and Equation 3.6 are part from our previous work (submitted).

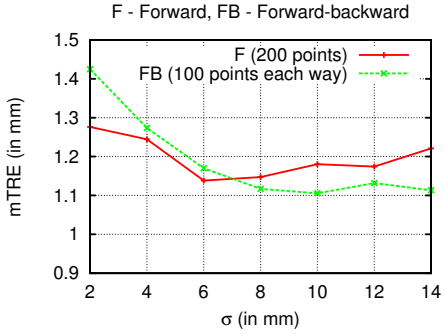



**Figure 3.1:** *Pairwise 3D Registration for various values of $\sigma$ (See equation 3.1, 3.4).*

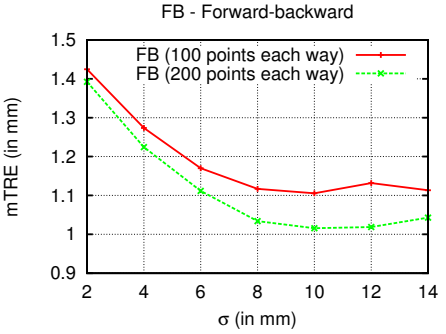



**Figure 3.2:** *Pairwise Forward-Backward 3D Registration for two grid spacings.*

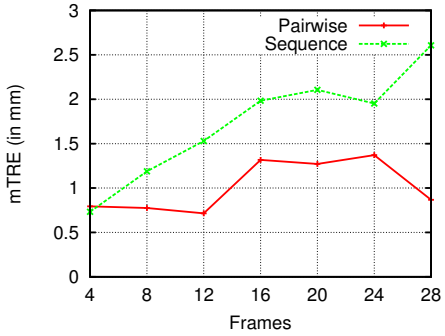



**Figure 3.3:** *Pairwise 3D vs. 4D Registration error.*

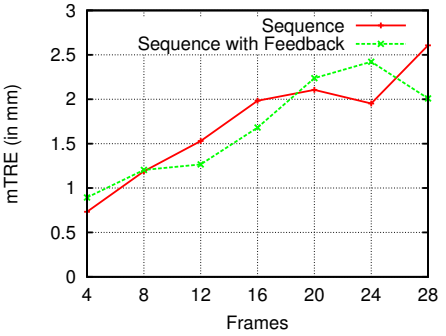



**Figure 3.4:** *4D vs. 4D with feedback.*

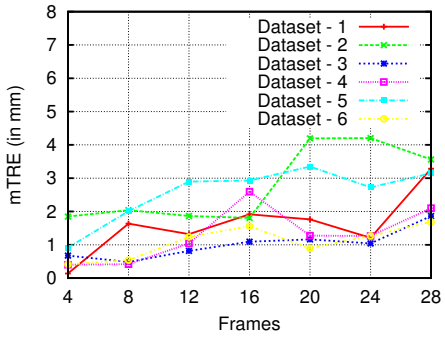



**Figure 3.5:** *4D Registration error per dataset over the time series.*

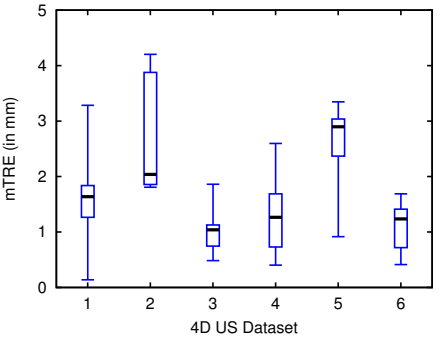



**Figure 3.6:** *Boxplot: 4D Registration error per dataset.*

### 3.2.3 Penalizing Drift

Registration over consecutive frames in a time series is likely to accumulate errors. To address this issue we bring a strategy that would help in reducing the drift. Consider $\mathcal{T}_{\tau_0}^{\tau_n}$ to be the transformation from frame $\tau_0 \rightarrow \tau_n$. It is estimated by taking product of all the transformations between successive frames in the sequence, i.e. $\mathcal{T}_{\tau_0}^{\tau_n} = \mathcal{T}_{\tau_{n-1}}^{\tau_n} \times \mathcal{T}_{\tau_{n-2}}^{\tau_{n-1}} \times \cdots \times \mathcal{T}_{\tau_0}^{\tau_1}$. To counter the drift we propose occasional long range interaction between the current frame $\tau_n$ and some previous frame in time, say $\tau_{n-d}$, where $d$ is the *feedback gap*. The steps of the strategy are as follows:

1. Register frames $\tau_{n-d}$ and $\tau_n$. Let the resulting transform be $\mathcal{F}_{\tau_{n-d}}^{\tau_n}$.
2. Let the pointsets $P_{\tau_{n-d}}$ and $P_{\tau_n}$ be pointsets in the frames $\tau_{n-d}$ and $\tau_n$, respectively. These pointsets are related by the transform $\mathcal{F}_{\tau_{n-d}}^{\tau_n}$, i.e. $P_{\tau_n} = \mathcal{F}_{\tau_{n-d}}^{\tau_n} * P_{\tau_{n-d}}$.
3. The projection of the pointsets $P_{\tau_{n-d}}$ over the frame $\tau_{n-1}$ is $P_{\tau_{n-1}}$, and is estimated as $P_{\tau_{n-1}} = \mathcal{T}_{\tau_{n-d}}^{\tau_{n-1}} * P_{\tau_{n-d}}$.
4. Include the additional points $P_{\tau_{n-1}}$ and $P_{\tau_n}$, in their respective frames, during the registration (outlier rejection module) of consecutive frames $\tau_{n-1}$ and $\tau_n$.

## 3.3 Experiments

The code was implemented in C++ and MeVisLab. A laptop with Intel(R) Core(TM) i7-2720QM CPU @ 2.20 G Hz, 4 Core(s) processor using 64-bit Windows 7 operating system and 8 GB of RAM is used for processing the code.

The 4D US data is acquired at 6 Hz from iU22 Philips machine. Three volunteers were used. From each of the volunteers, two (axial and coronal) sequences of 4D US were captured. The probe was kept steady during the acquisition. We use Elastix registration toolbox [61] to generate the reference standard. Different grids were used in the method and the evaluation. From the six 4D US datasets, systematically pairs of US volumes were selected to evaluate the performance. For all the experiments, we use the same set of parameters: block-size of $(11, 11, 11)$ mm and grid spacing of 14 mm (which translates to around 200 points). The sum of square distance (SSD) is the similarity metric using in block-matching. The first two experiments are performed to evaluate the registration approach. The third experiment is performed to evaluate the feedback mechanism.

1. The purpose of the experiment is to study the parameter $\sigma$ (given in equation 3.4) and its effect on pairwise 3D registration. The two pairwise registrations evaluated are a) Forward registration (using affinity matrix in equation 3.1) and b) Froward Backward registration (using affinity matrix in equation 3.4). The range of $\sigma$ values evaluated are $\{2, 4, 6, 8, 10, 12, 14\}$. We system-

atically select pairs of 3D US volumes from the six 4D US dataset. The pairs correspond to the following time points $\{(0,4), (0,8), (0,12), (0,16), (0,20), (0,24), (0,28)\}$ are used in the evaluation. Based on the registration results, we choose the best $\sigma$ value for our application and report the corresponding registration error. We use mean total registration error (mTRE) as the registration error metric. The mTRE is given as:

$$mTRE(\widehat{\mathcal{T}}_{\tau_1}^{\tau_2}, \mathcal{T}_{\tau_1}^{\tau_2}) = \frac{1}{n}\sum_{i=1}^{n} \|\widehat{\mathcal{T}}_{\tau_1}^{\tau_2} p_i - \mathcal{T}_{\tau_1}^{\tau_2} p_i\|\,, \tag{3.7}$$

where $\widehat{\mathcal{T}}_{\tau_1}^{\tau_2}$ is the reference standard transformation from time point $\tau_1$ to $\tau_2$.

2. Next we evaluate the performance of our registration approach over a sequence of US volumes. The 4D registration between the time points $(2,5)$, for example is estimated by registering $2 \rightarrow 3$, $3 \rightarrow 4$, $4 \rightarrow 5$ time points and then multiplying the respective transformation sequentially to derive the final transform. 4D registration and the pairwise 3D registration are compared with the reference standard, for the time points $\{(0,4), (0,8), (0,12), (0,16), (0,20), (0,24), (0,28)\}$.

3. In the third experiment we evaluate the feedback mechanism. Sequential registration similar to the previous experiment is performed for the time points $\{(0,4), (0,8), (0,12), (0,16), (0,20), (0,24), (0,28)\}$. Additionally, a feedback is used to counter the drift. The feedback gap $d$ is set to 4.

*Results :*

1. Figure 3.1 shows the forward and forward-backward registration results for various $\sigma$ values and two different grid space settings. The grid spacing of 14 mm, 18 mm correspond to 200, 100 sample points, respectively. For a fair comparison between the forward and forward-backward approach, the total number of points used for registration should be equal. In terms of the number of points, the forward approach with 200 points is equivalent to the forward-backward approach with 100 points each way. Figure 3.1 shows that the forward-backward registration approach performs better than the forward registration approach, given the number of points are same in the two methods. For grid spacing of 14 mm, the forward-backward approach performs best at $\sigma = 10$ mm with mTRE of 1 mm. Figure 3.2 shows the forward-backward registration results for various $\sigma$ values and two different grid space settings. Increase in sample points improves the registration results.

2. In Figure 3.3 we study the performance of the registration approach when applied between consecutive frames over the time series. For the first consecutive eight frames the mTRE is below 1.2 mm. Beyond these initial frames

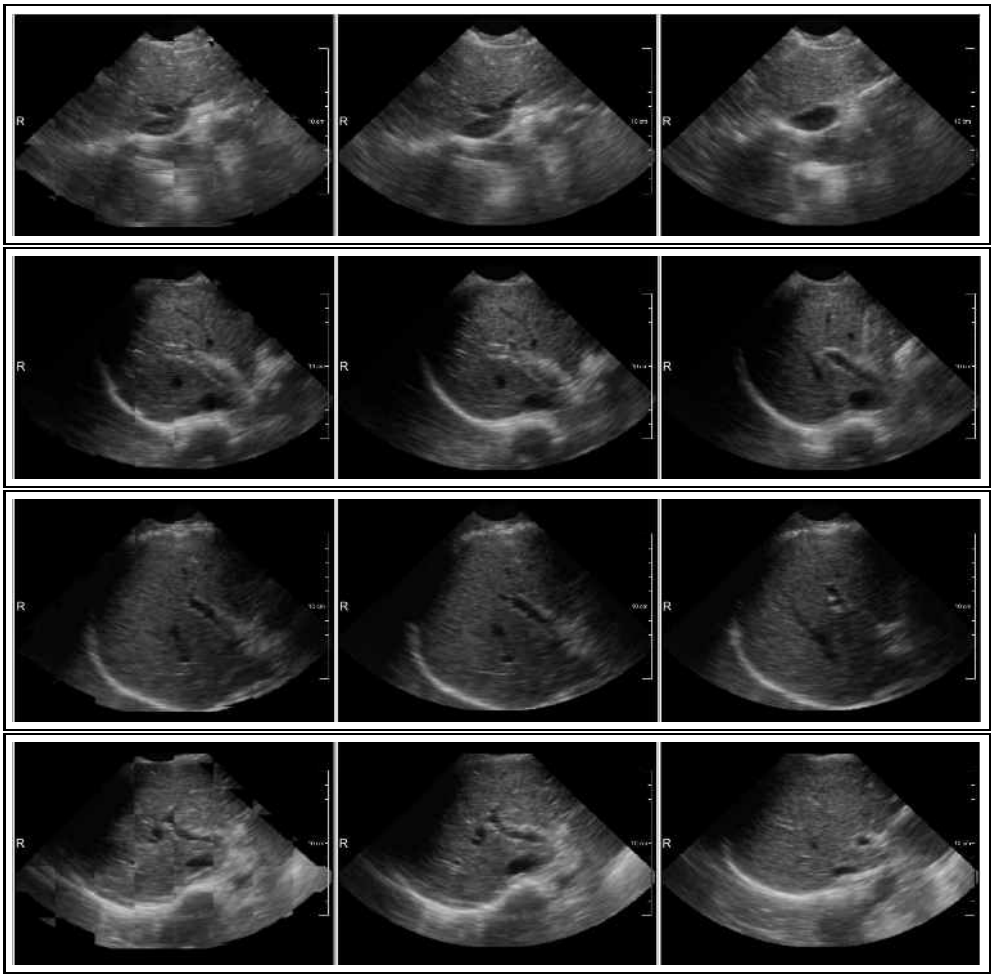


**Figure 3.7:** *4D Registration results for four datasets: Left - Registration result in checker box view, Middle - Fixed image, Right - Moving Image. First Row - Registration between time points 0 and 8, Second Row - Registration between time points 0 and 12, Third Row - Registration between time points 0 and 20, Last Row - Registration between time points 0 and 28.*

the registration performance diverge gradually. The results for all the six 4D US datasets are plotted in Figure 3.6. These results are also compared to the pairwise 3D registration results and are shown in Figure 3.3. Figure 3.5 and Figure 3.6 show the registration results on all the datasets.

3. In Figure 3.4 the 4D registration is evaluated with and without the feedback

mechanism. The 4D registration without feedback graph diverges. The 4D registration with feedback curve descends and moves closer to the reference standard. Some representative registration results are shown in Figure 3.7.

## 3.4 Discussion and Conclusion

In this work we present a registration framework to perform registration over a time series, and demonstrate its application on 4D US liver dataset. The method consists of two parts - the first part performs a 3D registration between subsequent volumes and ensures inverse consistency, the second part uses an feedback mechanism to counter accumulation of registration error.

Figure 3.3 shows that the pairwise registration performs well with mTRE of 1 mm. When applied sequentially, the registration results in Figure 3.3 show that the sequential registration works well for the first eight to ten frames. However after that the registration diverges and performance gradually deteriorates. Hence it is advisable to register back with the original frame after every tenth frame to maintain good registration accuracy. We further apply a feedback criterion to counter the accumulation of registration error over a time series. The initial results as shown in Figure 3.4 are encouraging. The feedback gap parameter $d$ is data dependent. In our experiments we select $d = 4$ as for $d - 1$ consecutive frames the accumulated registration error is below (1 mm) a tolerable limit, see Figure 3.3. More in-depth analysis is part of future work.

The block-matching and forward-backward outlier rejection was additionally implemented in OpenCL. The implementation was run on a NVIDIA GTX 780 Ti graphics processing unit. For block-size of $(11, 11, 11)$ mm, grid spacing of 18 mm (i.e. 100 points) and search range of $(20, 20, 20)$ mm the block-matching algorithm takes 0.045 seconds. The outlier rejection module with the forward-backward condition takes 0.05 seconds for grid spacing of 18 mm each way. Adding both the modules results in an execution speed of 10 Hz.

To conclude, we present a forward-backward transformation based registration approach for 4D US data. We evaluate a strategy to counter the accumulation of registration error using a feedback mechanism applying long range interaction. The approach is evaluated using six 4D US sequences with satisfactory results.

# Chapter 4

## 4D Ultrasound Tracking of Liver and its Verification for TIPS guidance

Based on:

Jyotirmoy Banerjee, Camiel Klink, Wiro J. Niessen, Adriaan Moelker and Theo van Walsum, 4D Ultrasound Tracking of Liver and its Verification for TIPS guidance, *IEEE Trans Med Imaging*, 2016.

**Abstract**

In this work we describe a 4D registration method for on the fly stabilization of ultrasound volumes for improving image guidance for transjugular intrahepatic portosystemic shunt (TIPS) interventions. The purpose of the method is to enable a continuous visualization of the relevant anatomical planes (determined in a planning stage) in a free breathing patient during the intervention. This requires registration of the planning information to the interventional images, which is achieved in two steps. In the first step tracking is performed across the streaming input. An approximate transformation between the reference image and the incoming image is estimated by composing the intermediate transformations obtained from the tracking. In the second step a subsequent registration is performed between the reference image and the approximately transformed incoming image to account for the accumulation of error. The two step approach helps in reducing the search range and is robust under rotation. We additionally present an approach to initialize and verify the registration. Verification is required when the reference image (containing planning information) is acquired in the past and is not part of the (interventional) 4D ultrasound sequence. The verification score will help in invalidating the registration outcome, for instance, in the case of insufficient overlap or information between the registering images due to probe motion or loss of contact, respectively. We evaluate the method over thirteen 4D US sequences acquired from eight subjects. A graphics processing unit implementation runs the 4D tracking at 9 Hz with a mean registration error of 1.7 mm.

# 4.1 Introduction

## 4.1.1 Clinical motivation

Ultrasound (US) is a unique medical imaging modality as it is non-invasive, affordable, portable and real-time. Ultrasonography is widely used in diagnosis. It is also used for therapy, for instance to guide biopsies and can be potentially used in percutaneous minimally invasive procedures such as radiofrequency ablation (RFA) and transjugular intrahepatic portosystemic shunt (TIPS) placement. Radiologists typically place the hand-held probe (called a transducer) on the surface of the body such that it is directed towards the anatomy of interest. US technology has recently significantly advanced with the introduction of matrix array transducers (3D probe). The 3D visualization using a US 3D probe has great potential, compared to conventional 2D US imaging, to further aid radiologists in planning and interventions. 3D US can create viewing planes which are not possible with a 2D US probe.

In our work we focus on improving image guidance for TIPS applications. In a TIPS procedure a bypass is created (stent is placed) from the hepatic vein to the portal vein, see Figure 4.1. This is an endovascular procedure, where the liver tissue is punctured with a needle. Such interventions are complex, mostly because imaging is inadequate (2D US, or X-ray), failing to adequately visualize the needle and target structure in 3D [26]. 4D US has the potential to improve image guidance as it enables continuous tracking of the target structure in 3D.

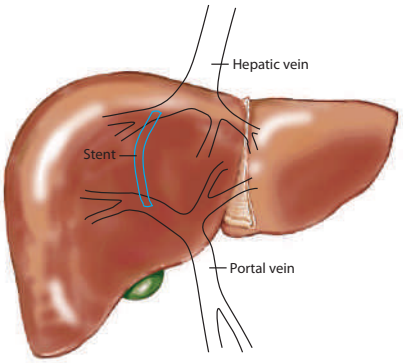



**Figure 4.1:** *TIPS Procedure - Stent Placement.*

Direct visualization of US volumes does not aid in image guidance for TIPS: there are no good direct volume rendering approaches that deal with the speckled nature of the US image and there is no way of automatically focusing the rendering to the anatomy of interest. Our proposed solution is to render specific planes (cross-sections) from the US volumes, i.e. by using planes from a planning US volume, in which the relevant vessels have been annotated and the planes are chosen such that they show the anatomy of interest, see Figure 4.2 [26]. By using a registration approach, 4D US can be used while keeping the area of interest in the visualized planes. Such an approach has two challenges: first, we need an initialization of the registration of the planning

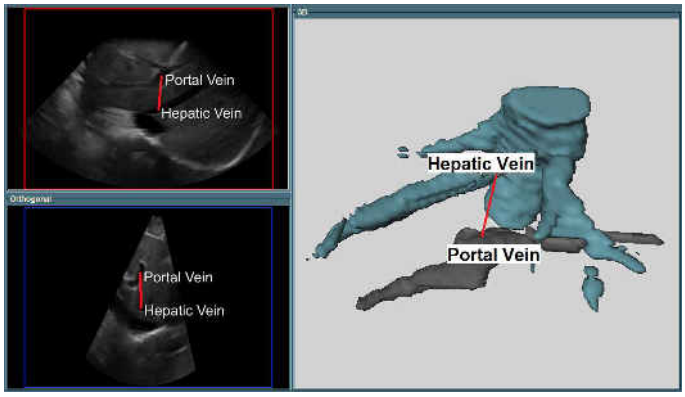



**Figure 4.2:** *Planning information consisting of appropriate scan planes and the necessary annotations required for easy navigation.*

US volume to the live 4D US images. And second, the alignment of the planning US volume should be maintained over time.

The purpose of this work is to develop and evaluate techniques that will assist in integrating planning information from the planning US volume to the interventional 4D US sequence in image guided minimally invasive interventions. To this end, we will address both the initial alignment of the planning information with the streaming 4D US, as well as compensating the liver motion for continuous visualization of the region of interest (ROI). The planning information in our clinical application is selecting the appropriate scan planes and the necessary annotations required for easy navigation as shown in Figure 4.2.

### 4.1.2 Related work

Our work is based on image registration, which is the process of spatially aligning images. Recently, there has been much interest in 4D registration techniques. Wachinger et al. [133] proposed a method of aligning multiple images by first registering consecutive images and then correcting the accumulated error using Lie normalization for mosaicing of US images. Group-wise 4D registration methods perform joint alignment of a collection of images. Group wise 4D US registration for motion compensation of liver was addressed by Vijayan et al. [130]. In their work, the spatial and temporal smoothness of the transformations are enforced using a temporal free-form deformation (TFFD) model originally proposed by Metz et al. [87]. Intensity variance over time is minimized using the group wise similarity metric mentioned in [8]. Group-wise methods generally use a global cost function and often do not require an explicit reference image. A *maximum likelihood estimation* (MLE) [133][132] and *congealing* [139][69] are other frameworks that have been

proposed in the literature for simultaneous registration of multiple images. In a congealing framework a set of images is aligned by minimizing the voxel-wise entropies. Its relationship with the MLE framework was established in [132]. Though these methods are reported to perform well [132], they are computationally intensive and not suitable for real-time applications.

Speckle tracking and optical flow are alternative approaches that have been used to register volumetric image data: both result in a (dense) deformation field. In the context of liver motion, speckle tracking based approaches have been presented recently by Harris et al. [47] and Bell et al. [7]. Harris et al. [47] developed speckle tracking-based motion tracking for abdominal organs during therapy. In their method tracking was performed with respect to a fixed reference frame and consecutive volumes. In Bell et al. [7] speckle tracking of consecutive volumes and cumulative addition of displacements was proposed for monitoring liver motion due to respiration and cardiac activity. Hjertaas et al. [51] used speckle tracking in 4D echocardiography for strain assessments. A 3D optical flow based method for myocardial motion tracking in echocardiography was presented by Mukherjee et al. [88]. A spatiotemporal smoothness constraint was incorporated by penalizing large gradients in the flow vectors at each voxel. The tracking locations are either manually initialized [82] or a coarse-to-fine multiscale approach is employed [88]. In Brattain et al. [11], real-time 4D registration was performed by tracking the probe using an electromagnetic (EM) tracking system.

In real-time 4D registration, the continuous input stream of US volumes puts a heavy computational and storage burden on the registration approach. A pair-wise solution will be needed to perform direct registration of the current incoming image (moving volume) to the reference image (fixed volume) [6], referred as Register-to-Reference (RTR) strategy. However this approach requires a large search range to cover not only the displacement induced by the breathing, but also due to probe motion.

Øye et al. [98] used the *dead reckoning* approach from navigation to perform real-time image registration on streaming 4D ultrasound data. Instead of registering the current incoming image to a single reference image, they register to a compound reference volume generated from a collection of images. This approach uses a graphics processing unit (GPU) to achieve real-time performance. Unlike the group-wise methods, real-time approaches use a simple transformation model like rigid or affine [98][117]. Schneider et al. [117] used a feature-based approach for real-time 4D US registration. They used SIFT-based [80] features followed by RANSAC [31][33] to establish correspondences between image pairs. To limit accumulation of error they maintained a global set of features and integrate them in their registration framework.

Sufficient proximity or overlap between the fixed and the moving images is required for a good registration; images that are far apart are harder to register, and registrations may be more computationally intensive. A detection approach may

be used to determine if the moving image (placement of the US probe) is well localized/initialized such that it images the same region of interest as the reference image. Ni et al. [94] proposed an automatic solution for localizing a fetal abdominal standard plane using an image feature based classification method. Liu et al. proposed an active appearance model [79] which mimics the visual cues for scan plane detection in US fetal head images. In radiation therapy Bell et al. [71] address the problem of tissue deformation caused by a US probe, resulting in a mismatch between planning and delivery, by using robotic probe placement.

### 4.1.3 Overview and contributions

In this work we present the following contributions. First, we propose a Register-to-Reference by Tracking (RTRT) strategy for *4D US tracking* to compensate for breathing motion in real-time. The RTRT approach helps in reducing the search range and is robust under rotation. Second, in a phantom experiment we show that the RTRT approach is robust under external devices such as needles. Third, we present an approach for registration *verification*, that would indicate whether the planning US volume and the interventional 4D US sequence can be adequately registered. This criterion will additionally help the registration application to switch from the RTR mode in the verification to the RTRT mode in the 4D US tracking, whenever required, automatically. Fourth, we thoroughly evaluate the technique on thirteen 4D US sequences. Additionally, a GPU implementation shows that the approach runs in real-time, and we demonstrate the application on US datasets.

The rest of the chapter is structured as follows. In Section 4.2 we discuss the RTRT 4D US tracking approach. In Section 4.3 we discuss the registration verification approaches. The experiments and results are presented in Section 4.4, which are discussed in Section 4.5.

## 4.2 4D US tracking

The outline of the RTRT approach is described in Figure 4.3. In our approach every incoming image is aligned with the reference image. The reference image is the US volume containing the planning information, and could be a US volume that has been acquired before the intervention. The approach consists of two steps. In the *tracking step* (step-I), the most recent consecutive images, i.e. the $t_{n-1}$ image and the current incoming image $t_n$ are rigidly aligned. The transformation from this step is combined with the previously estimated transformation between the reference image $t_{\text{ref}}$ and the $t_{n-1}$ image to approximate the transformation between the reference image $t_{\text{ref}}$ and the current incoming image $t_n$. In the subsequent *refine step* (step-II), a registration between the reference image $t_{\text{ref}}$ and the current incoming image $t_n$ is performed to estimate the final transformation, see Figure 4.3. The

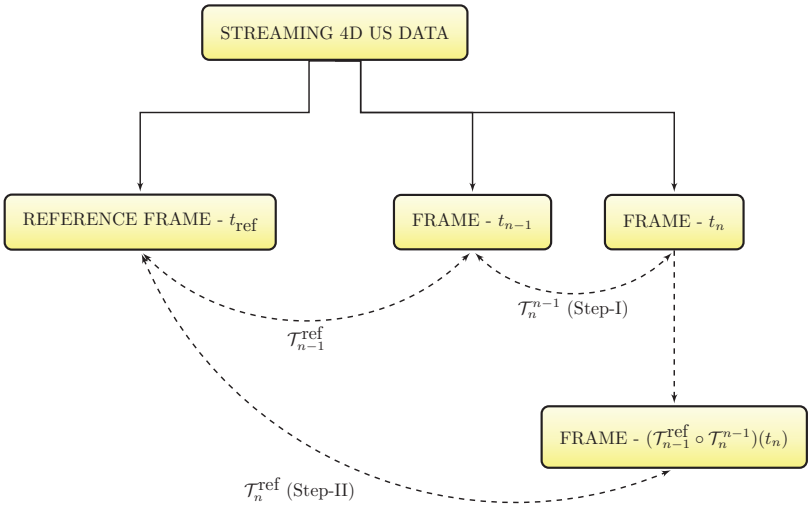



**Figure 4.3:** *Register-to-Reference by Tracking (RTRT). In step-I consecutive images are rigidly aligned using tracked points. The step-I information is propagated to the step-II. However, composition of transforms causes accumulation of error. Hence, in step-II the reference image and the transformed incoming image are realigned.*

motivation for this two-step registration is two-fold. First, compared to a direct registration of the incoming image to the reference image, which needs a large search range to be able to deal with liver and transducer motion, we can apply two registrations that only require a small search range. This reduces computation time and the registration is more likely to succeed. Second, compared to a concatenation of the transforms as in a tracking approach, our approach accumulates minimal error.

In the next subsections, we will first briefly review block-matching and outlier rejection, which is an essential building block of our 4D US tracking approach (Section 4.2.1). We subsequently provide details on the two steps of the 4D US tracking procedure (Section 4.2.2 and Section 4.2.3).

### 4.2.1 Block-matching and outlier rejection

We briefly summarize the block-matching and outlier rejection which are important components of our 4D tracking process and which were presented earlier; for complete review and details refer to the Chapter 2 [6]. Given a set of input points $\mathcal{X} := \{x_i\}_{i=0}^{m-1}$ in the fixed image, we employ a block-matching strategy to determine the corresponding target points $\mathcal{Y} := \{y_i\}_{i=0}^{m-1}$ in the moving image, and

find the best (rigid or affine) transform that matches the displacements from $\mathcal{X}$ to $\mathcal{Y}$. As block-matching in general, and in ultrasound in particular, is bound to fail in some cases (i.e. for some points), we need to first estimate the true correspondences, and discriminate those from erroneous block-matching results. For this, we employ a graph-based clustering approach, based on geometric consistency of the transformation: we use the fact that the distance between two points in $\mathcal{X}$ should not change after block-matching. Therefore, we fill an adjacency matrix $\mathcal{A}$, where each element $a_{ij}$ encodes this geometric consistency ($a_{ij}$ is small when the distance between point $i$ and $j$ before block-matching differs much from the distance between point $i$ and $j$ after block-matching, and $a_{ij}$ is large if the distances are equal). We also add the appearance information derived from the block-matching scores in to the adjacency matrix. Subsequently, to find the inliers and outliers of the block-matching, we maximize the following energy function of the adjacency matrix:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} \ \mathbf{f}^{\mathbf{T}} \mathcal{A} \mathbf{f}\,, \tag{4.1}$$

where $\mathbf{f}$ is a stochastic vector and $\mathbf{f} \in \Delta$, which is the unit simplex defined as:

$$\Delta = \{\mathbf{f} \in \mathbb{R}_{+}^{m} : \mathbf{e}^{\mathbf{T}}\mathbf{f} = 1\}\,, \tag{4.2}$$

where $\mathbf{e} = (1, \ldots, 1)^{\mathbf{T}} \in \mathbb{R}^m$. The $i$th element of $\mathbf{f}$, denoted by $f_i$, indicates whether the $i$th point is a true correspondence. A threshold on $\hat{f}_i$'s determines ($<$) the outliers and ($\geq$) the inliers [6].

### 4.2.2 Estimate transform by tracking

In step-I we track points across the streaming 4D US data. Figure 4.4 provides an overview of the tracking based registration approach. The inputs are two successive images $t_{n-1}$ and $t_n$. Block-matching is used to find correspondences between $S$ points randomly selected in image $t_{n-1}$ and image $t_n$. The block-matching may not always result in true correspondences. The outlier rejection scheme from the previous section is applied to retain the true matches and remove the false matches. The inliers are used to determine the rigid transformation between the images using the method described by Arun et al. [4]. Inliers also form input to the block-matching in the next cycle, i.e. registration between images $t_n$ and $t_{n+1}$. If the number of inliers, say $N$, is less than $S$, the minimum number of points required for robust tracking, then $(S - N)$ random points are added. Thus there are always $S$ number of points used for block-matching. By reusing inliers, the fraction of inliers in the total number of points is larger than when not reusing inliers. As a certain number of inliers are required to obtain a good registration, this allows us to reduce the total number of points to track while keeping a good registration.

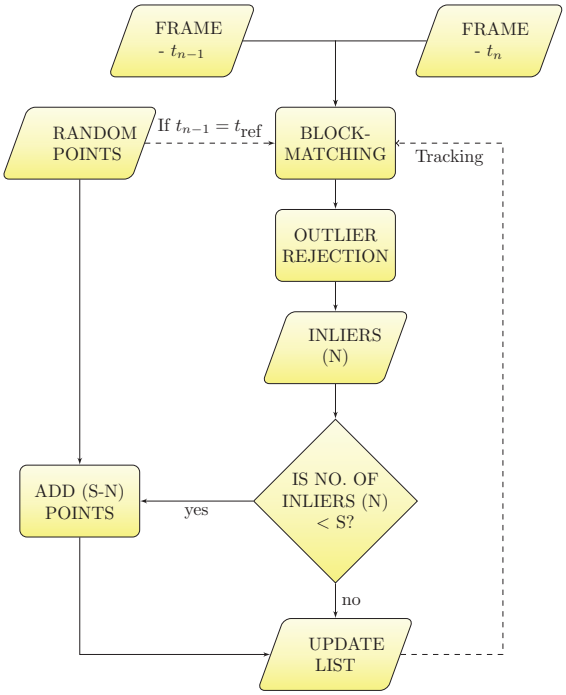



**Figure 4.4:** *Step-I: US Tracking. The flowchart shows the strategy to track points across the 4D streaming dataset. Inliers from the previous steps are used for the block-matching. If the number of inliers decreases below a certain threshold (S), the sample points are replenished.*

### 4.2.3 Refine transform

In step-II, given the tracking result from step-I, we determine the registration of reference image $t_{\text{ref}}$ to image $t_n$, see Figure 4.3. Consider we know the transformation between the reference image $t_{\text{ref}}$ and the image $t_{n-1}$, written as $\mathbf{T}^{\text{ref}}_{n-1}$. From step-I we have the transformation $\mathbf{T}^{n-1}_{n}$ between images $t_{n-1}$ and $t_n$. Composing transformation $\mathbf{T}^{\text{ref}}_{n-1}$ and $\mathbf{T}^{n-1}_{n}$ gives us the approximate estimate of the transformation $\mathbf{T}^{\text{ref}}_{n}$, written as $\tilde{\mathbf{T}}^{\text{ref}}_{n} = (\mathbf{T}^{\text{ref}}_{n-1} \circ \mathbf{T}^{n-1}_{n})$. The images $t_{\text{ref}}$ and $t_n$ are then roughly aligned using the transformation $\tilde{\mathbf{T}}^{\text{ref}}_{n}$. The alignment is not optimal, as the composed transformations may contain errors introduced by the participating transformations, which will be multiplied by combining the transformations. Hence these images are registered again to negate the accumulation of error. Note this is equivalent to registering the reference image $t_{\text{ref}}$ and the image $t_n$, where the sampling in the block-matching is altered by the transformation $\tilde{\mathbf{T}}^{\text{ref}}_{n}$. Implementation wise the latter approach is more efficient. It does not require explicit rotation and trans-

lation of the image $t_n$, and such sampling schemes are well supported by the GPU hardware. The block-matching is followed again by outlier rejection to estimate the inliers, similar to step-I in Figure 4.4. The inliers are used to estimate the transformation $\mathbf{T}_n^{\text{ref}}$ using the approach described by Arun et al. [4] for the rigid case and the approach described by Späth [122] for the affine case. The newly estimated transformation $\mathbf{T}_n^{\text{ref}}$ feeds into the next cycle of registration between the reference image $t_{\text{ref}}$ and the image $t_{n+1}$. In both the steps I and II the initial transformation of the images used in the registration is close to the transformation to be determined, which helps in limiting the search range. Finally the motion is modeled by the transformation $\mathbf{T}_{\text{ref}}^n$ between the reference image $t_{\text{ref}}$ and the input image at time point $t_n$.

## 4.3 Registration verification

The two-step registration approach from the previous section requires a previously known registration, and thus cannot be used at the start of a procedure, or at any moment when the registration between the planning volume and the 4D US volumes is not known. Therefore, in clinical practice, registration will start with an RTR approach, and a verification step is required to detect the moment from which the registration between the planning US volume and the 4D US sequence can be trusted. We use the RTR approach mentioned in [6]. Consider registering a reference US volume $t_{\text{ref}}$ with planning information to streaming in 4D US volumes $\mathbf{t} = \{t_0, t_1, \cdots\}$ from an ultrasound machine. The reference volume was acquired prior to the streaming 4D sequence, i.e. $t_{\text{ref}} \neq t_0$, where $t_0$ is the starting image of the sequence. In clinical practice, the first image of the sequence will not overlap sufficiently with the reference image to warrant a good registration. We are therefore interested in automatically detecting the first $t_i$ for which the registration to the reference image succeeded. Once $t_{\text{ref}}$ and $t_i$ are registered all the subsequent images in the 4D sequence can be registered to $t_{\text{ref}}$ using the approach described in the previous section, see Figure 4.3.

The purpose of the registration verification therefore is to automatically detect a (sufficiently) good registration while registering according to the RTR approach. Such detection thus enables to automatically switch to the accurate and fast RTRT registration, where the only requirement for the human operator is to position the transducer such that the image is approximately aligned with the reference image (i.e. within the search range of the RTR approach), which is not a difficult task, as it is the same as the normal search for the target anatomical structures.

Let the two possible output states of the registration verification be written as $\{\Omega_S, \Omega_F\}$, where $\Omega_S$ is when the images $\{t_{\text{ref}}, t_i\}$ are registered successfully and $\Omega_F$ otherwise, i.e. when the registration failed. We propose two methods to automatically distinguish between the states $\{\Omega_S, \Omega_F\}$ from the registration results. The underlying assumption of the methods is that, under a slowly moving transducer

(and slowly moving anatomy), good registrations are consistent over time. This leads to the following two criteria for deciding whether a registration is successful: a) consistency of the points that are selected as inliers, b) consistency of the transformation irrespective of the locations of the points in the block-matching. Below, we describe both methods, based on these two criteria.

### 4.3.1 Method 1 - Inlier consistency based verification

Let $\hat{\mathbf{f}}_{(\text{ref},i)}$ be the distribution of the inliers/outliers from the registration of images $t_{\text{ref}}$ and $t_i$, see Equation 4.1. We analyze $l$ consecutive pair-wise registration outcomes $\{\hat{\mathbf{f}}_{(\text{ref},i-l+1)}, \hat{\mathbf{f}}_{(\text{ref},i-l+2)}, \cdots, \hat{\mathbf{f}}_{(\text{ref},i)}\}$ over time. If the reference image $t_{\text{ref}}$ has sufficient overlap with the images $\mathbf{t}^i_{i-l+1} = \{t_{i-l+1}, t_{i-l+2}, \cdots, t_i\}$, and if the images are of good quality, then the descriptive locations in image $t_{\text{ref}}$ and its corresponding locations in images $\mathbf{t}^i_{i-l+1}$ are consistently picked up as inliers during the outliers rejection optimization process, see Equation 4.1. The vector $\hat{\mathbf{f}}$ represents the descriptiveness of the different locations of the US volume. Hence it is reasonable to assume that the vector $\hat{\mathbf{f}}$ does not change significantly during registrations of the image $t_{\text{ref}}$ and the sequence $\mathbf{t}^i_{i-l+1}$. On the other hand during improper registrations the vector $\hat{\mathbf{f}}$ may significantly change between the registrations of the image $t_{\text{ref}}$ and the sequence $\mathbf{t}^i_{i-l+1}$. Hence the variance of the set of vectors $\hat{\mathbf{f}}$ can be used as a surrogate to determine the output states $\Omega_S$ or $\Omega_F$ of the verification approach.

Principal component analysis (PCA) can be used to estimate the modes of variations of the random vector $\hat{\mathbf{f}}$. We analyze this by stacking a set of $\hat{\mathbf{f}}$ vectors into a $m \times l$ matrix represented as:

$$\mathbf{F} = \begin{bmatrix} \hat{\mathbf{f}}_{(\text{ref},i-l+1)} & \hat{\mathbf{f}}_{(\text{ref},i-l+2)} & \cdots & \hat{\mathbf{f}}_{(\text{ref},i)} \end{bmatrix}, \tag{4.3}$$

where $l$ is the number of images taken into account. The variance of $\mathbf{F}$ is given as:

$$\text{Var}[\mathbf{F}] = \begin{bmatrix} \sigma_0^2 \\ \sigma_1^2 \\ \vdots \\ \sigma_{m-1}^2 \end{bmatrix}, \tag{4.4}$$

where $\{\sigma_0^2 \geq \sigma_1^2 \geq \cdots \geq \sigma_{m-1}^2\}$ are the Eigen values of the covariance matrix $\frac{1}{l-1}\hat{\mathbf{F}}\hat{\mathbf{F}}^{\mathbf{T}}$. $\hat{\mathbf{F}}$ is the mean substracted data given as $\hat{\mathbf{F}} = \mathbf{F} - \mathbf{U}\mathbf{e}^{\mathbf{T}}$, where $\mathbf{e} = (1, \ldots, 1)^{\mathbf{T}} \in \mathbb{R}^l$ and $\mathbf{U}(i) = \frac{1}{l}\sum_{j=0}^{l-1}\mathbf{F}(i,j)$. The metric to quantify the variations in $\hat{\mathbf{f}}$ is given by the sum of the Eigen values $S_I = \sum_{i=0}^{m-1}{\sigma_i}^2$. A small $S_I$ value implies $\hat{\mathbf{f}}$ values are similar over a time series, which indicates a successful verification $\Omega_S$. Accordingly a large $S_I$ value implies bad verification $\Omega_F$. As trace of a matrix is the

sum of its Eigen values, $S_I$ can be efficiently estimated from the covariance matrix without any matrix decomposition.

### 4.3.2 Method 2 - Transform consistency based verification

In the second method to detect a good registration, we analyze transformations from pair-wise registrations over time to determine the verification results $\{\Omega_S, \Omega_F\}$. For two different set of input points $\mathcal{X}_1$ and $\mathcal{X}_2$ in the block-matching, let the transformation from the image $t_{\text{ref}}$ to image $t_i$ be given as $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_1)$ and $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_2)$, respectively. Given sufficient and evenly distributed input points, the transformation should be invariant to the location of points $\mathcal{X}$. Let $\delta\mathbf{x}^i$ be the difference between two point sets in image $t_i$, where $\delta\mathbf{x}^i \in \mathbb{R}^{3k}$, $k$ is the number of points. The two point sets are obtained by transforming an arbitrary point set $\mathbf{x}^{\text{ref}}$ in the reference image $t_{\text{ref}}$ using the transformations $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_1)$ and $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_2)$. $\delta\mathbf{x}^i$, which thus is a metric for the similarity of the transforms $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_1)$ and $\mathbf{T}^i_{\text{ref}}(.;\mathcal{X}_2)$, is given as:

$$\delta\mathbf{x}^i = \left(\mathbf{T}^i_{\text{ref}}(\mathbf{x}^{\text{ref}};\mathcal{X}_1) - \mathbf{T}^i_{\text{ref}}(\mathbf{x}^{\text{ref}};\mathcal{X}_2)\right). \tag{4.5}$$

For successful verification $\Omega_S$ we expect $\delta\mathbf{x}^i$ to be small. We analyze the random vector $\delta\mathbf{x}^i$ over few images by stacking them into a $3k\times l$ matrix represented as:

$$\mathbf{X} = \begin{bmatrix}\delta\mathbf{x}^{i-l+1} & \delta\mathbf{x}^{i-l+2} & \dots & \delta\mathbf{x}^{i}\end{bmatrix}, \tag{4.6}$$

where $l$ is the length of the time series. We again assume that the vector $\delta\mathbf{x}^i$ has small magnitude in case of proper registrations of the image $t_{\text{ref}}$ and the sequence $\mathbf{t}^i_{i-l+1}$, implies $\mathbf{X} \to \mathbf{0}$. On the other hand in case of improper verification the vector $\delta\mathbf{x}^i$ may have significantly large magnitude between the registrations of the image $t_{\text{ref}}$ and the sequence $\mathbf{t}^i_{i-l+1}$.

Similar to $S_I$ value in Subsection 4.3.1, the metric to quantify the variations in $\delta\mathbf{x}^i$ is given by a sum of the Eigen values of the covariance matrix of $\mathbf{X}$, $S_T$. A small $S_T$ indicates a successful verification $\Omega_S$, and similarly a large $S_T$ value implies failed verification $\Omega_F$. Note $\mathbf{X} \to \mathbf{0}$ implies $\text{Var}[\mathbf{X}] \to \mathbf{0}$.

## 4.4 Experimental Setup

### 4.4.1 Data and resources

4D US data were acquired with a Philips iU22 Ultrasound Imaging System using an X6-1 xMatrix Array Transducer. The data is transferred from the ultrasound machine to a computer via Ethernet. The US volumes were $192\times246\times117$ voxels with voxel size of $1.144\times0.594\times1.193$ mm$^3$ and were acquired at a frequency of 6 Hz. Thirteen 4D US sequences from eight volunteers previously acquired for [6] were used. Additionally four long 4D US sequences ($\sim$3 min) were acquired from three

**Table 4.1:** *Datasets - 4D US Sequences.*

| No. of seq.,subj. | Patient or volunteer | Experiment No. | Probe steady or motion | Length (in sec) |
|---|---|---|---|---|
| 13,8 | Volunteer | 4.4.4-1 | Steady | 6 |
| | | 4.4.4-3 | | |
| | | 4.4.4-4 | | |
| 4,3 | Volunteer | 4.4.4-2 | Steady | 150 |
| | | 4.4.4-4 | | |
| 5,2 | Patient | 4.4.4-5 | Motion | 7 |

volunteers. The volunteers were breathing freely during the acquisition. The operator manually kept the probe steady while acquiring the real-time 3D volumes of the liver.

To evaluate the registration verification approaches five patient datasets were acquired. In these datasets the probe was moved from 'out' to 'in' of the ROI while acquiring the real-time 3D volumes. The lengths of the 4D US sequences were approx. 40 images.

Table 4.1 shows the 4D US datasets and their corresponding experiments. US data was anonymized before processing and for the patient data informed consent was obtained.

The code was implemented in C++, OpenCL and MeVisLab. The OpenCL GPU implementation was run on a NVIDIA GTX 780 Ti GPU.

### 4.4.2 Parameter setting

In the 4D tracking method mentioned in Section 4.2 we used a block-size of $11^3$ $mm^3$, and for the outlier rejection parameters $(\sigma_A, \lambda, \sigma_B)$ we used $(0.1, 0.1, 0.1)$; these values were optimized in the previous work [6]. We used normalized cross correlation (NCC) as similarity metric. The number of points and search range is evaluated in the experimental section below.

In the verification approaches mentioned in Section 4.3 the optimized registration parameter settings from [6] were used. In the block-matching module we use a block-size of $11^3$ $mm^3$ and 200 points, and for the outlier rejection parameters $(\sigma_A, \lambda, \sigma_B)$ we used $(0.1, 0.1, 0.1)$. The number of frames $l$ and search range is evaluated in the experimental section below.

### 4.4.3 Evaluation metric

Ground truth registrations were obtained using manual annotations. The annotations establish correspondences between the fixed and the moving volumes in a limited number of images. Three to four points are annotated per volume. The registration error is quantified using the mean target registration error (mTRE) metric,

**Table 4.2:** *Notations.*

| Symbol | Description |
|---|---|
| $l$ | Number of frames |
| $S_I$ | Method 1 score |
| $S_T$ | Method 2 score |
| $S_I^\tau$ | Threshold for $S_I$ values |
| $S_T^\tau$ | Threshold for $S_T$ values |
| $\Omega_S$ | If $S_I \leq S_I^\tau$ or $S_T \leq S_T^\tau$ (Successful verification) |
| $\Omega_F$ | If $S_I > S_I^\tau$ or $S_T > S_T^\tau$ (Failed verification) |

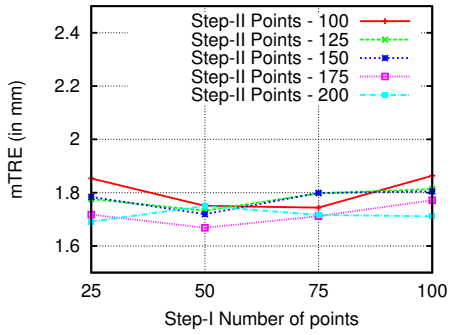



**Figure 4.5:** *Evaluating number of points. Step-I and Step-II search ranges are $30^3$ $mm^3$ and $20^3$ $mm^3$, respectively.*

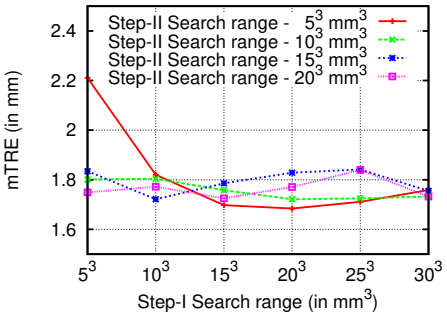



**Figure 4.6:** *Evaluating search range. Step-I and Step-II number of points are 50 and 125, respectively.*

measured in mm. mTRE is defined as the average distance between the points estimated by transforming the ground truths from the fixed to the moving volume and the corresponding ground truths in the moving volume.

In the verification experiments the ground truth was determined by visually comparing the registered images. For a stack of $l$ images the outcome was considered positive if and only if all the registrations with the reference image $t_{\text{ref}}$ were good, otherwise the outcome was considered negative.

### 4.4.4 Experiments

We performed a series of experiments to evaluate the 4D registration (1, 2 and 3) and verification schemes (4 and 5). Refer to Table 4.2 for quick reference of notations.

1. We investigated the accuracy of the RTRT 4D tracking method. From thirteen 4D ultrasound sequences over eight subjects the following 4D US image pairs

(3, 33), (4, 29), (5, 25), (6, 21), (7, 17), (8, 13) and (9, 10) were extracted and annotated. We ran the method over the first 34 images of the sequence. To find the optimum parameter values, we evaluated the performance of the method for various settings of number of points and search range in the *tracking step* (step-I) and the *refine step* (step-II) of the RTRT registration approach. The evaluation range of the number of points in the block-matching for the tracking step and the refine step are {25, 50, 75, 100} and {100, 125, 150, 175, 200}, respectively. The evaluation range of the search range in the block-matching for the tracking step and the refine step are $\{5^3, 10^3, 15^3, 20^3\}$ $mm^3$ and $\{5^3, 10^3, 15^3, 20^3, 25^3, 30^3\}$ $mm^3$, respectively. To determine the number of points we initially set the search range to the maximum value of the evaluation range. Hence the search range of the tracking step and the refine step were set to $30^3$ $mm^3$ and $20^3$ $mm^3$, respectively. Once the number of points for the block-matching were determined we varied the search range. Using the optimized settings we investigated inter-observer variability using a second observer annotating all pairs of seven of the thirteen 4D ultrasound sequences. We calculated the root mean square (RMS) error between the annotations in the fixed image and the moving image and the mTRE for both the observers. Additionally, the dependency between the observers is evaluated using a two-sided Wilcoxon signed-rank test between the registration errors over all image pairs of these seven 4D ultrasound sequences.

2. We next evaluated the robustness of the 4D registration method over long US sequences. The experiment is designed to evaluate whether the RTRT strategy, see Figures 4.3, performs consistently over long US sequences and does not accumulate registration error in the process. To demonstrate this we use four long 4D US sequences. We evaluate the registration after every 100th image, upto 900 images.

3. The proposed RTRT registration strategy with its two step approach should be able to compensate for large transformations. To evaluate this, we applied the RTRT approach under controlled rotational transform of an ultrasound image. We selected randomly ten different US volumes and rotated each dataset from 0 to 180 degrees with a step size of 10 degrees to form a 4D US sequence per volume. The axis of rotation was $(x = 1, y = 1, z = 1)$.

4. We investigated the optimum parameter values for the *location consistency based verification* and the *transform consistency based verification* approaches proposed in this work. Thereto we quantified the goodness of registration between the reference image $t_{\text{ref}}$ and a 4D US sequence $\mathbf{t}^i_{i-l+1}$=($t_{i-l+1}$, $t_{i-l+2}$, $\cdots$, $t_i$). A successful verification ($\Omega_S$) is simulated by choosing the reference image such that it has significant overlap with the volumes of the 4D US sequence, $t_{\text{ref}} \in \mathbf{t}^i_{i-l+1}$. A failed verification ($\Omega_F$) is simulated by choosing the

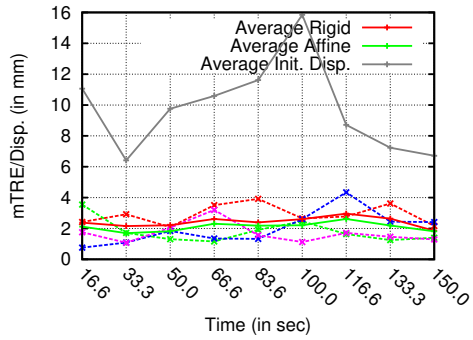



**Figure 4.7:** *Long ultrasound sequences - Affine registration results on four datasets (dotted lines); average rigid, average affine registration results and average initial displacement over the four datasets (solid lines).*

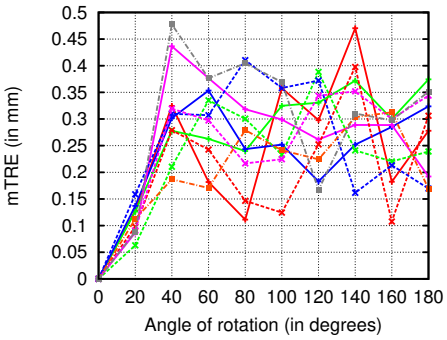



**Figure 4.8:** *Manually rotated sequences. Each line/color/shape represents one dataset.*

reference image from the same subject but acquired at a different location than the 4D US sequence, $t_{\text{ref}} \notin \mathbf{t}^{i}_{i-l+1}$. The evaluation range of the number of frames $l$ in Equation 4.3 is $\{2, 4, 6, 8\}$. We use five 4D US sequences. For each US sequence we evaluate at 20 different time points, where two consecutive time points are ten images apart. Hence we have 100 evaluations per parameter setting. The evaluation range of the search range for the block-matching is $\{10^3, 20^3, 30^3, 40^3\}$ mm$^3$. Sensitivity and specificity analysis is performed to find the optimum threshold values $S_I^\tau$ and $S_T^\tau$ of the two methods. In Equation 4.5, $\mathbf{x}^{\text{ref}}$ should be a set of at least four non-coplanar points. We select $\mathbf{x}^{\text{ref}}$ as the eight corners of the US volume.

5. Using the optimized threshold values $S_I^\tau$ and $S_T^\tau$ we evaluate their accuracy on five 4D US sequences where the probe was moved from 'out' to 'in' of the ROI. The accuracy is defined as the ratio between the number of correct assessments and the number of all assessments, i.e. Accuracy $= (\text{TN} + \text{TP})/(\text{TN} + \text{TP} + \text{FN} + \text{FP})$, where TP is true positive, TN is true negative, FN is false negative and FP is false positive.

### 4.4.5 Results

1. Figure 4.5 shows that for a fixed number of points in the refine step (step-II) of the RTRT approach, the mTRE values do not vary much for various number of points in the tracking step (step-I). For example, for 125 points in the refine step (step-II), the mTRE values for the number of points in the tracking step (step-II) 25, 50, 75 and 100 are 1.75 mm, 1.73 mm, 1.79 mm and 1.81

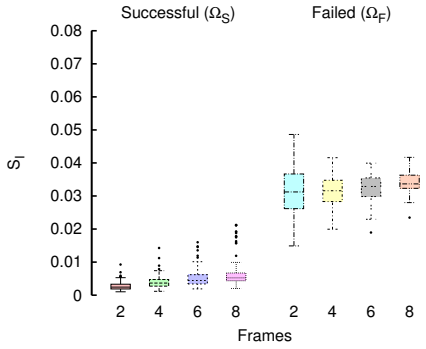



**Figure 4.9:** *Inlier consistency based verification - Various number of frames (l). Search range = $20^3$ $mm^3$, No. of points = 200.*

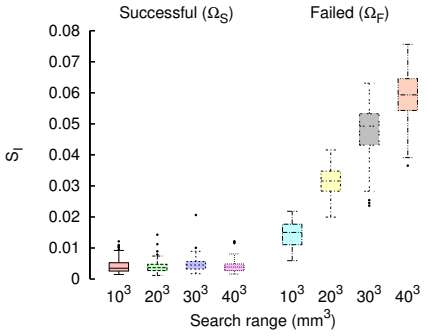



**Figure 4.10:** *Inlier consistency based verification - Various search range. Number of frames = 4, Number of points = 200.*

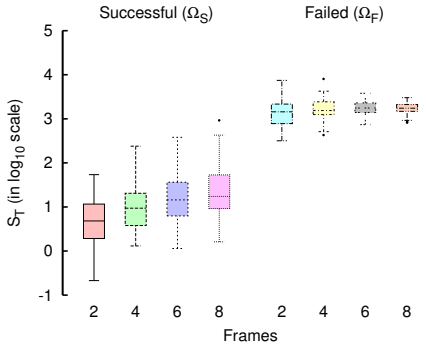



**Figure 4.11:** *Transform consistency based verification - Various number of frames (l). Search range = $20^3$ $mm^3$, Number of points = 200.*

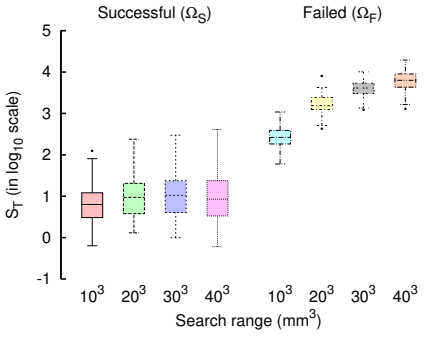



**Figure 4.12:** *Transform consistency based verification - Various search range. Number of frames = 4, Number of points = 200.*

mm, respectively. The greater the number of points, the better is the chance of getting an accurate registration [6], measured using mTRE. However, an increase in the number of points would increase the time required for the registration, see Figure 4.15. Considering the speed and the accuracy of the approach we choose the number of points in the tracking step (step-I) and the refine step (step-II) as 50 and 125 points, respectively. Given the number of points in tracking step and the refine step are 50 and 125, respectively, Figure 4.6 shows that for a fixed search range in the refine step (step-II) of the RTRT approach, the mTRE values do not vary much for different search ranges in the tracking step (step-I), except when the search range in the track-

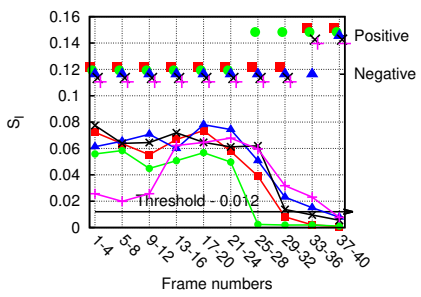



**Figure 4.13:** *Inlier consistency based verification - Score (left Y-axis and the lines) vs. Predicted outcome (right Y-axis and the different shaped points). Each line/color/shape represents one dataset.*

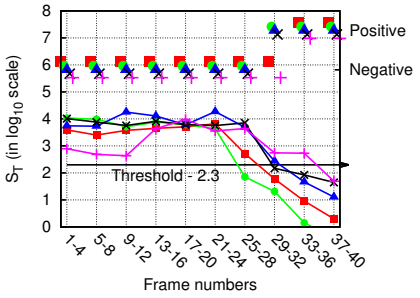



**Figure 4.14:** *Transform consistency based verification - Score (left Y-axis and the lines) vs. Predicted outcome (right Y-axis and the different shaped points). Each line/color/shape represents one dataset.*

**Table 4.3:** *Affine registration error - RTR [6] vs. RTRT method.*

| Subject | No. of 4D US sequence | RTR [6] Error (in mm) | RTRT Error (in mm) |
|---|---|---|---|
| 1 | 1 | 2.75 | 2.74 |
| 2 | 2 | 1.55 | 1.52 |
| 3 | 2 | 2.31 | 2.09 |
| 4 | 2 | 2.08 | 1.90 |
| 5 | 2 | 1.41 | 1.31 |
| 6 | 2 | 1.71 | 1.67 |
| 7 | 1 | 1.83 | 1.77 |
| 8 | 1 | 0.96 | 0.88 |
| Mean error | | 1.83 | 1.74 |

ing step (step-I) and the refine step (step-II) is $5^3$ mm$^3$. The search range used in the block-matching is directly related to the execution speed. We would like the search range in the tracking step (step-I) to compensate for the motion between two successive frames and the search range in the refine step (step-II) to compensate for the accumulation of error. Considering the speed and the accuracy of the approach we choose the search range in the tracking step (step-I) and the refine step (step-II) as $25^3$ mm$^3$ and $10^3$ mm$^3$, respectively. Table 4.3 shows the mean registration error of 1.74 mm over thirteen 4D US sequences. Table 4.3 shows the registration error per-subject and is compared to Banerjee et al. [6]. Table 4.3 show that the RTRT approach has slightly better accuracy compared to the RTR registration approach mentioned in Banerjee

et al. [6]. However a two-sided Wilcoxon signed-rank test over all image pairs shows that the differences in both the approaches are not statistically significant. The two tailed Wilcoxon signed-rank test was done for 5% level of significance. The inter-observer variability statistics over seven 4D US sequences is: first observer mTRE 1.59$\pm$0.94 mm, RMS error 0.85$\pm$0.49 mm; second observer mTRE 1.56$\pm$0.86 mm, RMS error 0.99$\pm$0.73 mm. A two sided Wilcoxon signed-rank test shows that the differences in the registrations results, evaluated using the annotations of the two observers, are not statistically significant.

2. Figure 4.7 shows the registration error of four US sequences over a long period. The registration error does not increase over time, which indicates that in our approach there is minimal accumulation of registration error over time. Figure 4.7 shows that the affine version slightly outperforms the rigid version of the RTRT method. The average initial displacement curve in the Figure 4.7 shows the average displacement between the fixed reference image and the moving image prior to registration.

3. Figure 4.8 shows the registration error on the controlled dataset with respect to the ground truth. The mTRE value for the ten datasets are below 0.5 mm for rotations up to 180 degrees of the US volumes.

4. The results of the registration verification methods based on *location consistency based verification* and *transform consistency based verification* approaches are shown in Figure 4.9, Figure 4.10, Figure 4.11 and Figure 4.12. In Figure 4.9 and Figure 4.11 we used a fixed search range ($20^3$ mm$^3$) and varied the number of images used in the verification. In Figure 4.10 and Figure 4.12 we used a fixed number of images (4) and vary the search range. For a search range of $40^3$ mm$^3$ and number of images $l = 4$, both the methods are able to distinguish between successful verification ($\Omega_S$) and failed verification ($\Omega_F$). We would like the RTRT method to have a high specificity i.e. low false positive rate. From Table 4.4 we choose the value of $S_I^\tau = 0.012$ and the value of $S_T^\tau = 10^{2.3}$ for the two methods.

5. Figure 4.13 and Figure 4.14 show the verification score vs. the predicted outcome for the two initialization methods. In the two plots, the transition from negative to positive registration (approximately) coincides with the dip in the verification score from greater than to less than the threshold value. In Figure 4.13, for one dataset (pink line/plus shape points) the *location consistency based verification* method delays the prediction of a positive outcome and for another dataset (red line/square shape points) the method advances the prediction of a positive outcome. In Figure 4.14, for one dataset (pink line/plus shape points) the *transform consistency based verification* method delays the prediction of a positive outcome and for another two datasets (green line/circle

**Table 4.4:** *Sensitivity (SE) and Specificity (SP) of the Verification Methods - search range* $= 40^3$ *mm*$^3$ *and number of images* $= 4$.

| | Method 1 | | | Method 2 | |
|---|---|---|---|---|---|
| $S_I$ | SE | SP | $S_T$ | SE | SP |
| 0 | 0 | 1 | $10^{-0.1}$ | 0.02 | 1 |
| 0.012 | 1 | 1 | $10^{0.5}$ | 0.23 | 1 |
| 0.024 | 1 | 0.99 | $10^{1.1}$ | 0.61 | 1 |
| 0.036 | 1 | 0.99 | $10^{1.7}$ | 0.88 | 1 |
| 0.048 | 1 | 0.88 | $10^{2.3}$ | 0.97 | 1 |
| 0.06 | 1 | 0.44 | $10^{2.9}$ | 1 | 0.98 |
| 0.072 | 1 | 0.01 | $10^{3.5}$ | 1 | 0.85 |
| 0.084 | 1 | 0 | $10^{4.1}$ | 1 | 0.06 |
| 0.096 | 1 | 0 | $10^{4.7}$ | 1 | 0 |
| $S_I^\tau \rightarrow$ | | 0.012 | $S_T^\tau \rightarrow$ | | $10^{2.3}$ |

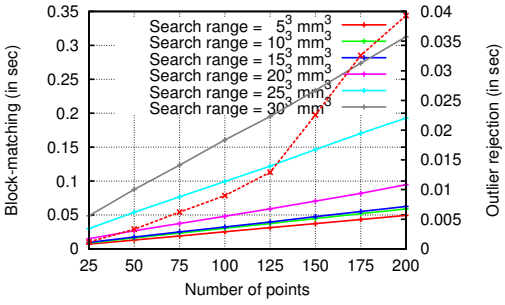



**Figure 4.15:** *GPU timings - Block-matching (left Y-axis and the solid lines) and Outliers rejection (right Y-axis and the dotted line).*

shape points and red line/square shape points) the method advances the prediction of a positive outcome. The accuracy of the registration verification approaches based on *location consistency based verification* method with $S_I^\tau$ = 0.012 is 96% and *transform consistency based verification* method with $S_T^\tau = 10^{2.3}$ is 92%.

### 4.4.6 GPU timings

The GPU execution time of the OpenCL code for the block-matching and the outlier rejection modules is shown in Figure 4.15. For NCC similarity metric, block size of $11^3$ mm$^3$, the number of points in the tracking step and the refine step of 50 points and 125 points, search range in the tracking step and the refine step of $25^3$ mm$^3$

and $10^3$ mm$^3$, the block-matching and the outlier rejection in the tracking step and the refine step combined takes 0.11 seconds. Hence with GPU implementation the registration runs at 9 Hz, for the given parameter settings.

### 4.4.7 Video Recordings

In the supplementary material we uploaded five videos showing 4D ultrasound registrations over long US sequences. The videos contain on the left the registered image, in the middle the fixed image and on the right the moving image. An additional video is added showing the registration result on a 4D US sequence where a TIPS needle was inserted into a phantom and to introduce motion the probe was displaced slowly.

## 4.5 Discussion and Conclusion

We presented a 4D US tracking approach (RTRT) and techniques to verify and initialize registration. These techniques enable 4D US image guidance in minimally invasive interventions such as TIPS. The RTRT registration approach helps to compensate for the liver motion and it works in real-time using GPU hardware with mean registration error of 1.74 mm.

The RTRT approach is performed in two steps. The tracking step (step-I) uses a rigid transformation as this step helps in only finding an approximate transformation between the fixed reference frame and the streaming input frame. In the refine step (step-II) the final transformation between the above two frame is determined. The choice of the final transformation could either be affine or rigid transformation as shown in Figure 4.7.

We also propose approaches which verify and score the quality of registration, facilitating the verification and the initial registration of the planning US volume with the interventional 4D US sequence. Whereas current approaches for integrating planning data often are cumbersome, requiring external equipment [71] [123], the proposed verification approaches enable user-friendly integrating the planning data in the interventional setup. The clinician only needs to move the probe to search the target region and generate images similar to the planning image. The verification approaches automatically indicate whether the current images can be adequately registered to the image on which the planning was performed. Table 4.4 shows that for $S_I \leq 0.012$ in *location consistency based verification* and for $S_T \leq 10^{2.3}$ in *transform consistency based verification*, the methods suggest successful verification $\Omega_S$. The accuracy of *location consistency based verification* is 96% and *transform consistency based verification* is 92%. As shown in Figure 4.13, Figure 4.14 and discussed in the results subsection 4.4.5-5, there are few cases of disparity (of one time point) between the transition from negative to positive vs. the verification score

from greater than to less than the threshold value. Considering that we are evaluating an initialization method, the cases where the verification is one time point too late in detecting a good registration does not have consequences for the use. For the other cases, where the verification is one time point too early in detecting a good registration, we observe that the probe enters the capture/search range resulting in a low verification score as part of the $l$ frames are registered well. However, the protocol for the human observer was to only label a registration good if all of the $l$ frames are registered well, which explains the difference in outcome. Again, for use in practice this is not a problem, as the last frames are registered well and these are used to start the RTRT approach.

A registration algorithm requires sufficient overlap between the participating volumes. In practice the probe may be displaced between planning and treatment. It may be easy to detect when the probe moves away from the ROI using the resulting transformation as a cue. However it is much harder the other way round, i.e. to ensure the probe is brought back such that the planning and treatment images are adequately aligned. Bell et al. [71] address the problem in radiation therapy using robotic probe placement. Our verification approaches do not involve external tracking equipment and are image-based. In the verification approaches, as shown in Figure 4.10 and Figure 4.12, increase in the search range corresponds to increase in gap of the $S_I$, $S_T$ values for successful ($\Omega_S$) and failed ($\Omega_F$) verifications. The verification methods infer from a stack of images, $l$, hence it introduces a delay as it evaluates all the registrations between the reference image $t_{\text{ref}}$ and the images $t_{i-l+1}, t_{i-l+2}, \cdots, t_i$ before scoring the result. Given US frequency of 6 Hz and $l = 4$, the delay of 0.67 seconds is not significantly large.

The RTRT registration approach as presented has three advantages over the RTR registration approach discussed in Banerjee et al. [6]. First, points that are consistently tracked well are inherently more descriptive and lesser number of such points are required for tracking. For the number of points in the tracking step (step-I) as low as 25 points, Figure 4.5 shows that the mean registration errors are $< 2$ mm. Second, at any instance of time, tracking successive images helps in reducing the search range. For the search range in the tracking step (step-I) as low as $10^3$ mm$^3$, Figure 4.6 shows that the mean registration errors are $< 2$ mm. Thus RTRT registration strategy typically requires less number of points and smaller search range than RTR registration thus is computationally more efficient. For the number of points in the tracking step and the refine step of 25 points and 100 points, search range in the tracking step and the refine step of $10^3$ mm$^3$ and $10^3$ mm$^3$, Figure 4.15 show that the RTRT 4D US tracking runs in 20 Hz. Such settings may be used in cases where a probe holder is used to fixate the probe, and thus only breathing motion needs to be corrected. Such a setup also reduces the personnel for the intervention. It allows fixating the transducer and selecting the visualization planes afterwards, independent of the probe orientation. Third, unlike the RTR registration approach the RTRT registration approach is robust (invariant) to large rotations, see Figure 4.8. In Banerjee

et al. [6] for rotations larger than 25 degrees the registration error significantly increases because the block-matching cannot handle large rotations. Schneider et al. [117] mention that their method is robust to registration up to 20 degrees.

The RTRT 4D US tracking strategy is conceptually closest to Donoghue et al. [27]. They propose a global registration framework where affine registrations are pre-computed in advance for neighbouring images only. When a pairwise image registration is required the shortest path across the graph is used to estimate the approximate transformation by composing the intermediate transformations. Lastly a refinement step is used to further improve the registration accuracy.

In 3D US mosaic applications, non-rigid registration is required for artefact free stitching. It is however challenging to build a non-rigid real-time 4D US registration. The optimization and interpolation schemes required for estimating non-rigid deformation fields are often computationally intensive. Our work uses an affine transformation model implemented in real-time using GPU hardware and is sufficient for the task of stabilizing the planes in an US volume required for image guidance in TIPS minimally invasive procedure.

Motion due to respiration also constitutes a major problem in applications such as radiotherapy. In Kaar et al. [54] registration error was found to be $2.99 \pm 1.54$ mm for US volume dimensions of $243 \times 227 \times 135$ voxels at a voxel resolution of 0.6 mm in each dimension for radiotherapy of the prostrate. Schneider et al. [117] for US volume dimensions of $144 \times 112 \times 112$ voxels and a resolution of $0.55 \times 0.54 \times 0.63$ mm/voxel report registration accuracy of $1.1 \pm 0.8$ mm for cardiac applications. For therapy using focused US Vijayan et al. [130] found registration error of 1 mm for liver US volume dimensions of $227 \times 229 \times 227$ with isotropic voxels of $0.7 \times 0.7 \times 0.7$ mm/voxel using non-rigid registration approach. Our mean registration error of 1.74 mm for voxel size of $1.144 \times 0.594 \times 1.193$ mm/voxel on liver US volumes is comparable to the existing approaches.

The RTRT 4D US tracking combined with registration verification permits 4D US guidance of TIPS interventions, and allows seamless integration of planning information in the intervention, and the registration accuracies obtained may help in stabilizing the scan planes required for image guidance. We are currently integrating our work in a prototype system for 4D US guidance. Whereas our work was developed for TIPS interventions, these methods may be relevant for other image guided interventions as well, e.g. motion tracking for radiotherapy, as discussed above. When combined with registration of US to other 3D modalities, such as CT or MRI [67], the RTRT registration would enable multimodal image guidance, which would be relevant e.g. for percutaneous tumour treatments.

The current approach has been evaluated mainly on healthy volunteers. For clinical use the test bed should be extended to more patient imaging data. We also did not test the RTRT method during interventions and in the presence of external devices such as catheters and needles. However, given the character of our registration method, we do not expect difficulties when instruments are inserted, as

demonstrated in the supplementary video where a TIPS needle was inserted into a phantom and motion was simulated by slowly displacing the probe.

To conclude, we present a 4D US tracking method to compensate breathing motion, enabling continuous visualization of planning planes in minimally invasive procedures, e.g. TIPS interventions. The method can register 4D US volumes at 9 Hz, with an accuracy of 1.7 mm. Additionally we proposed methods which facilitate the verification and the registration between the planning US volume and the interventional 4D US sequence. The method runs in real-time and hence can improve image guidance in TIPS and other minimally invasive interventions.

# Chapter 5

## Tracking anatomical landmarks in 4D ultrasound of the liver

Based on:

Jyotirmoy Banerjee, Camiel Klink, Erwin Vast, Wiro J. Niessen, Adriaan Moelker and Theo van Walsum, A combined tracking and registration approach for tracking anatomical landmarks in 4D ultrasound of the liver, *MICCAI Challenge - CLUST*, 2015.

**Abstract**

In this paper we present a method for tracking of anatomical landmarks in the liver. Our 4D ultrasound tracking method is based on global and local rigid registration schemes. We evaluate our method on the dataset that was presented in the MICCAI 2015 Challenge on Liver Ultrasound Tracking (CLUST 2015). On the test set a mean tracking error of $1.62 \pm 0.94$ mm is achieved.

## 5.1 Introduction

Ultrasound (US) is used by clinicians to image the human anatomy. It is an inexpensive, non-invasive and portable imaging modality. It is widely used in diagnostics. As US imaging is real-time it can be used for interventions and therapy. The anatomy can be tracked real-time. Some of the applications are tissue motion analysis and image guidance during interventions. One of the main purpose of an US tracking approach is to incorporate (pre-operative) planning information (to guide visualization), or to integrate preoperative imaging data during interventions. Tracking or motion compensation algorithms helps to negate the motion caused by the probe or the patient and the breathing motion in particular.

Several methods for tracking of anatomical landmarks [20, 82, 83] and motion tracking of liver [7, 70, 98, 105, 117, 130] in US have been proposed in literature. Our method is based on the previous work described in Banerjee et al. [5] and [6]. The previous methods were developed to track/register US liver volumes. In this work these methods are used to perform the specific task of tracking anatomical landmarks in the liver. The method is evaluated on the CLUST 2015 challenge datasets.

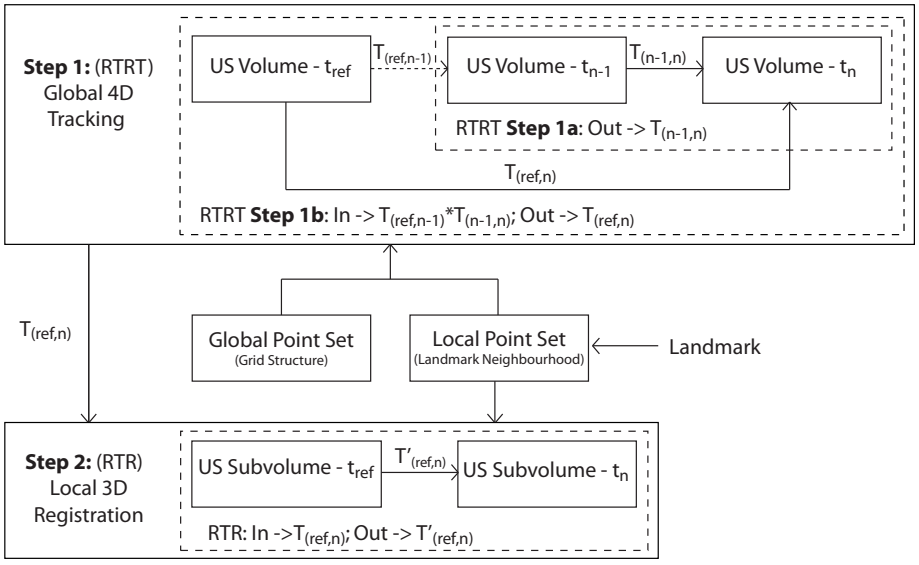



**Figure 5.1:** *Block diagram - tracking of anatomical landmarks.*

## 5.2 Tracking anatomical landmarks

We briefly review the register to reference strategy (RTR) [6] in Subsection 5.2.1 and the register to reference by tracking strategy (RTRT) [5] in Subsection 5.2.2 which are the core components of our landmark tracking approach. In Subsection 5.2.3 we discuss the landmark tracking approach, see the block diagram in Figure 5.1.

### 5.2.1 Step 1: RTR

The RTR approach [6] is a 3D to 3D US registration technique where the streaming input frame ($t_n$) is directly registered to the reference frame ($t_{ref}$). It is based on a block-matching scheme followed by an outlier-rejection scheme. For a set of points (generated using a grid structure or a Gaussian distribution, see the global point set and the local point set in Figure 5.2) located in the fixed image, block-matching is used to find corresponding locations in the moving image. The correspondences from the block-matching are inputs to the outlier rejection scheme. The outlier rejection scheme uses geometric and appearance consistency criteria to determine the block-matching results that can be trusted. The method then uses only the selected block-matching results from the outlier rejection scheme to estimate a rigid transformation using the approach described by Arun et al. [4]. For details refer [6].

### 5.2.2 Step 2: RTRT

The RTRT approach [5] is a 4D US registration/tracking technique, where the registration is performed in two steps. In the first step (Step 1a), the streaming input frame ($t_n$) is registered to the previous frame in the temporal domain ($t_{n-1}$). In the second step (Step 1b), the previously estimated transformation ($T_{(ref,n-1)}$) and the transformation from the first step ($T_{(n-1,n)}$) are used to initialize the registration between the streaming input frame and the reference frame, by composing the transformations as $T_{(ref,n-1)} * T_{(n-1,n)}$. To reduce the accumulation of error the reference frame is re-registered to the streaming input frame which was earlier transformed using the transformation $T_{(ref,n-1)} * T_{(n-1,n)}$, resulting in the final transformation, written as $T_{(ref,n)}$, see Figure 5.1. The RTRT approach additionally performs efficient tracking of points in the temporal domain. The tracking starts with a set of $\ell$ points (generated using a grid structure or a Gaussian distribution, see the global point set and the local point set in Figure 5.2) located in the fixed/reference image. Points that are consistently tracked are retained and the rest of the points are rejected. Additional points are introduced from a distribution (could be the same distribution as used earlier) if the number of points for tracking is less than $\ell$. For details refer [5]. Note that in the next cycle of the RTRT approach, the current estimated transformation $T_{(ref,n)}$ is used in determining the transformation between the reference volume ($t_{ref}$) and the next US volume $(t_{n+1})$, written as $T_{(ref,n+1)}$.

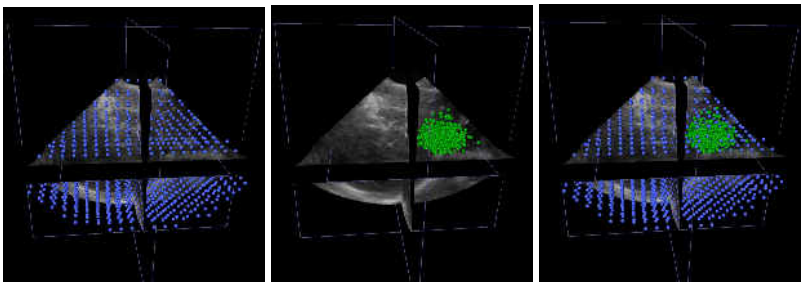


**Figure 5.2:** *Global and local point set: Left - Global point set generated using a grid structure (blue color), Middle - Local point set (green color), Right - Global cum local point set generated using a Gaussian distribution ($\sigma$=10 mm).*

### 5.2.3 Tracking landmarks

The anatomical landmark tracking approach, see Figure 5.1, consists of the following two rigid registration steps. First, in the *global* 4D registration/tracking step, the RTRT strategy is used track the whole (liver) US volume ($T_{(ref,n)}$). Second, in the *local* 3D registration step, we refine the tracking result by performing registration using the neighborhood region close to the anatomical landmark ($T'_{(ref,n)}$).

Both the RTR and RTRT strategies use block-matching followed by an outlier rejection scheme to find correspondences between the US volumes. Input to the block-matching module is a point set. The portion/region of the image used for the registration/tracking is determined by the locations of the points in the US volume. As shown in the block diagram in Figure 5.1, a combination of a global and a local point set is used to perform a global 4D tracking/registration and only a local point set is used to perform a local 3D registration. The global point set is generated using a grid structure spread over the entire US volume, the local point set is a collection of points in the neighborhood of the anatomical landmark (see Figure 5.2).

## 5.3 Experiment and results

The CLUST 2015 challenge dataset is used to evaluate the performance of the method. The challenge contained 16 4D sequences from multiple sources. The summary of the data is shown in Table 5.1. The data was divided into a training set of 8 4D sequences and a test set of 8 4D sequences. For tuning the algorithm, annotations (i.e. landmarks) across multiple frames per 4D sequence were provided for the training set. For the test set, one or more annotations in the first frame were provided. These annotated landmarks were tracked over time. The tracking performance of the test set was evaluated by the organizers of the challenge. The Euclidean distance between the tracked points and manual annotations was calculated. The error was

**Table 5.1:** *Summary of the data*

| Source | Traning Sequences | Test Sequences | Image size [voxels] | Image res. [mm] | Frame rate [Hz] | Scanner | Probe |
|---|---|---|---|---|---|---|---|
| EMC | 3 | 2 | 192x246x117 | 1.14x0.59x1.19 | 6 | Philips iU22 | X6-1 |
| ICR | 1 | 1 | 480x120x120 | 0.31x0.51x0.67 | 24 | Siemens SC2000 | 4Z1c |
| SMT | 4 | 5 | 227x227x229 | 0.70x0.70x0.70 | 8 | GE E9 | 4V-D |

summarized by the following statistics: mean, standard deviation, 95 percentile, minimum and maximum distances.

MeVisLab, OpenCL and C++ are used for software development. The OpenCL code was run on a NVIDIA GTX 780 Ti GPU.

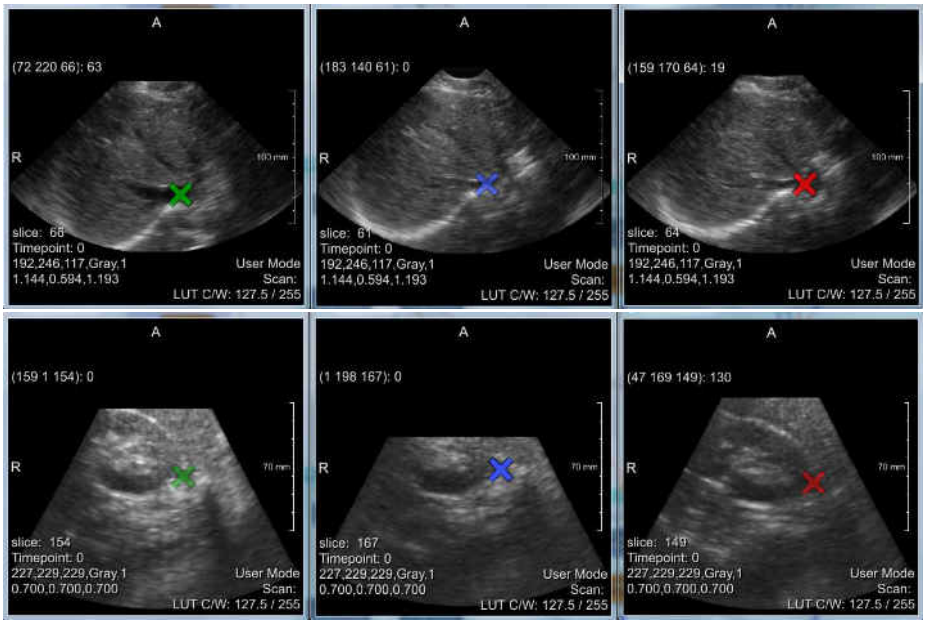



**Figure 5.3:** *Training set tracking results: Left - Reference image manual annotation, Middle - Moving image automatic annotation, Right - Moving image manual annotation. Row 1 - For the dataset EMC-03_1 at time point 42, the tracking error is 8.54 mm; Row 2 - For the dataset SMT-04_1 at time point 76, the tracking error is 15.14 mm.*

*Parameter setting :* We used a block-size of $11^3$ mm$^3$ for the block-matching. The block is evenly sampled 18x18x18 times. The similarity metric used is normalized cross correlation (NCC), $(\sigma_A, \lambda, \sigma_B) = (0.1, 0.1, 0.1)$ is used as the outlier rejection parameters. These values were optimized in the previous work [6]. The number of points for the block-matching (step one and step two) of the RTRT approach and the RTR approach are set to 100, 200 and 400 points, respectively. The search range for the block-matching (step one and step two) of the RTRT approach and the RTR approach is set to $40^3$ mm$^3$, $10^3$ mm$^3$ and $20^3$ mm$^3$, respectively. The search range (step one and step two) of the RTRT approach and the RTR approach are evenly sampled 60x60x60 times, 15x15x15 times and 30x30x30 times, respectively. The sampling determines the step size for the block-matching. A local point set of 1000 points is generated using a Gaussian distribution with mean located at the anatomical land-

mark and standard deviation of 10 mm. The adjacent horizontal/vertical nodes of the grid structure used to generate the global point set are 10 mm apart, see Figure 5.2. The points required for the block-matching in the RTR and the RTRT approaches are sampled from the global and the local point sets.

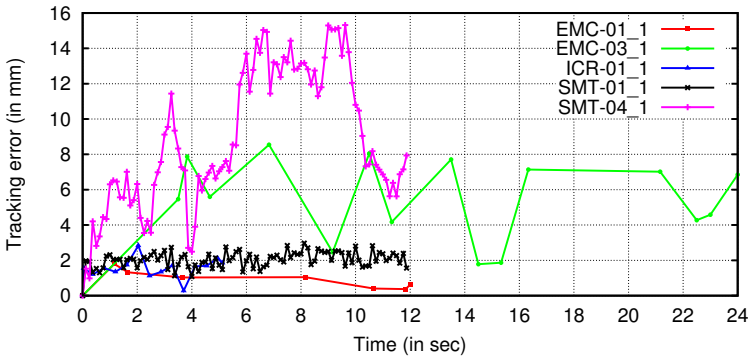



**Figure 5.4:** *Tracking examples from the training set.*

The training set and the test set results are presented in Table 5.2 and Table 5.3, respectively. The mean tracking error for the training set and the test set are 3.26 $\pm$ 2.62 mm and 1.62 $\pm$ 0.94 mm, respectively. The average run time of the Step 1 (RTRT) and the Step 2 (RTR) of our approach as shown in Figure 5.1 are 6.68 seconds and 4.18 seconds, respectively. Hence for the given parameter settings the GPU implementation runs at 11 seconds per frame.

## 5.4 Discussion and conclusions

In this chapter we perform the task of tracking anatomical landmarks using a combination of previous methods by Banerjee et al. [5] and [6]. A mean tracking error of 1.62 $\pm$ 0.94 mm is achieved on the test set. In the first step, the point set used for the global 4D tracking step is a combination of a global point set generated from a grid structure and the local point set generated randomly in the neighborhood of the anatomical landmark. This combination of point set ensures a high percentage of points close to the landmark position during the global 4D tracking step. The local point set is intended to track a specific landmark well, whereas the global point set helps in increasing robustness in tracking. In the second step, the local point set is again used in the local 3D registration step. This step is designed to track the landmark in the presence of local deformations.

The mean tracking error for the training set is 3.26 $\pm$ 2.62 mm. Two of the datasets (EMC-03_1, SMT-04_1) from the training set have large tracking errors, see

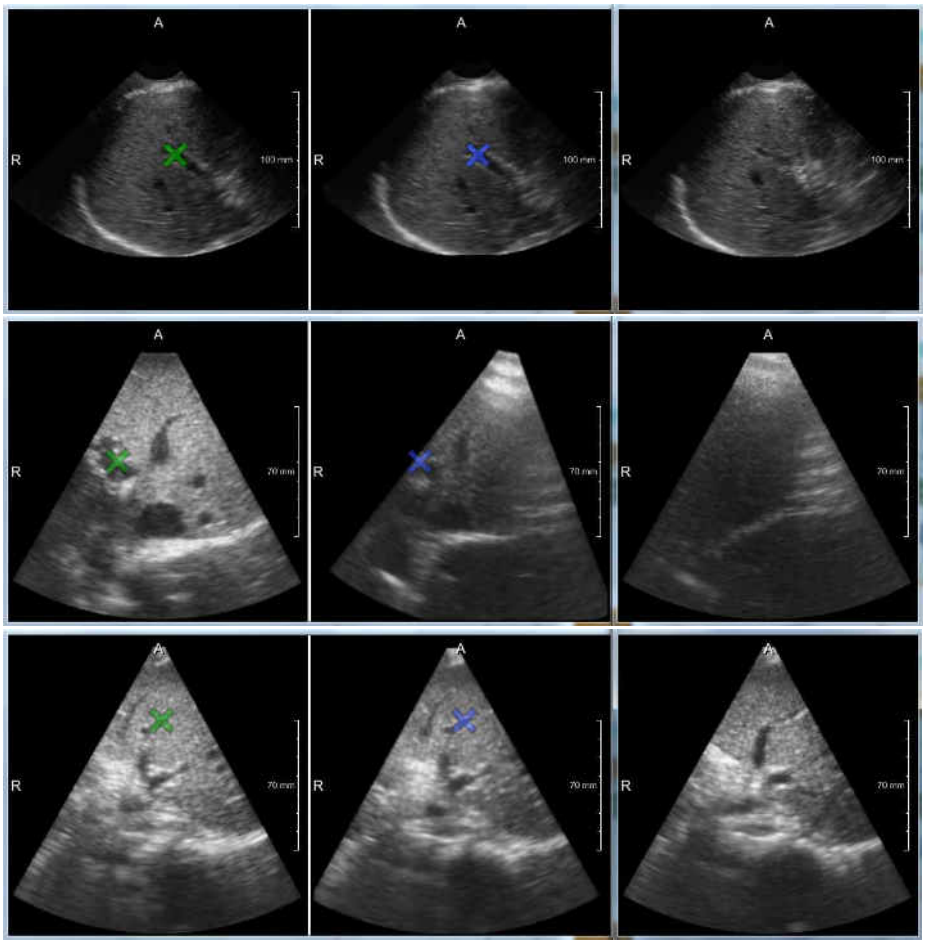



**Figure 5.5:** *Test set registered Volumes: Left - Reference image manual annotation, Middle - Registration result automatic annotation, Right - Moving Image. Row 1 - dataset EMC-04_1, time point 124; Row 2 - dataset SMT-05_1, time point 66; Row 3 - dataset SMT-07_2, time point 74.*

Figure 5.3. In the EMC-03_1 4D US sequence the anatomical landmark is located on a vessel which undergoes large deformations due to blood flow and in the SMT-04_1 4D US sequence the anatomical landmark is located outside the liver. Some of the tracking results from the training set are shown in Figure 5.4. In the test set the SMT-05_1 4D sequence has large tracking error. In rest of the dataset the tracking performance is satisfactory. Some representative test set registration results are shown in Figure 5.5.

The speed depends on the number of points, search range size, number of sam-

**Table 5.2:** *Training set results*

| Landmarks | Mean (in mm) | Std (in mm) | 95th% (in mm) | Min (in mm) | Max (in mm) |
|---|---|---|---|---|---|
| EMC-01_1 | 0.94 | 0.51 | 1.65 | 0.36 | 1.79 |
| EMC-02_1 | 1.19 | 0.47 | 1.83 | 0.80 | 2.01 |
| EMC-02_2 | 2.28 | 1.10 | 3.62 | 1.02 | 3.80 |
| EMC-02_3 | 2.05 | 0.58 | 2.80 | 1.48 | 2.96 |
| EMC-02_4 | 1.80 | 0.54 | 2.39 | 1.14 | 2.41 |
| EMC-03_1 | 5.55 | 2.28 | 8.20 | 1.78 | 8.54 |
| EMC | 3.01 | 2.42 | 7.85 | 0.36 | 8.54 |
| ICR-01_1 | 1.57 | 0.56 | 2.36 | 0.27 | 2.83 |
| SMT-01_1 | 2.06 | 0.41 | 2.74 | 1.08 | 2.97 |
| SMT-01_2 | 3.46 | 0.55 | 4.43 | 2.42 | 4.61 |
| SMT-01_3 | 3.00 | 0.42 | 3.75 | 1.91 | 3.89 |
| SMT-02_1 | 1.65 | 1.60 | 2.20 | 0.6 | 16.49 |
| SMT-02_2 | 1.92 | 0.47 | 2.74 | 0.91 | 3.27 |
| SMT-02_3 | 3.72 | 0.70 | 4.79 | 2.30 | 5.59 |
| SMT-03_1 | 2.29 | 0.72 | 3.43 | 1.19 | 3.62 |
| SMT-03_2 | 2.09 | 0.60 | 3.14 | 0.68 | 3.54 |
| SMT-04_1 | 8.88 | 3.82 | 15.04 | 0.97 | 15.31 |
| SMT | 3.30 | 2.64 | 9.20 | 0.60 | 16.49 |
| Tracking Results | 3.26 | 2.62 | 8.55 | 0.27 | 16.49 |

ples in the search range (step size), block size and number of samples in the block. The current approach runs at 11 seconds per frame. For tracking of liver, real-time (faster than image temporal resolution) speed is achieved by Banerjee et al. [5] by selecting appropriate parameters for the US data acquired from Philips iU22 machine with X6-1 probe.

To conclude, we extended our current registration approaches for 3D and 4D US volumes such that it enables tracking of anatomical landmarks in 4D US sequences. The method is evaluated using CLUST 2015 challenge datasets. For a test set of eight 4D US sequences, an accuracy of $1.62 \pm 0.94$ mm is achieved.

After the challenge, accuracy of $1.80 \pm 1.64$ mm was reported by the challenge organizer on the entire test set.

**Table 5.3:** *Test set results*

| Landmarks | Mean (in mm) | Std (in mm) | 95th% (in mm) | Min (in mm) | Max (in mm) |
|---|---|---|---|---|---|
| EMC-04_1 | 1.10 | 0.63 | 2.28 | 0.26 | 2.31 |
| EMC-05_1 | 1.79 | 0.36 | 2.16 | 1.15 | 2.17 |
| EMC | 1.45 | 0.61 | 2.18 | 0.26 | 2.31 |
| ICR-02_1 | 1.65 | 0.37 | 2.14 | 0.80 | 2.15 |
| SMT-05_1 | 3.39 | 2.53 | 10.13 | 0.90 | 10.24 |
| SMT-05_2 | 0.97 | 0.36 | 1.58 | 0.21 | 1.91 |
| SMT-06_1 | 1.56 | 0.37 | 2.11 | 0.57 | 2.49 |
| SMT-06_2 | 2.01 | 0.52 | 2.77 | 0.99 | 3.69 |
| SMT-06_3 | 1.37 | 0.35 | 1.95 | 0.43 | 2.18 |
| SMT-07_1 | 1.83 | 0.42 | 2.49 | 1.02 | 2.92 |
| SMT-07_2 | 1.79 | 0.39 | 2.46 | 1.04 | 2.69 |
| SMT-08_1 | 1.48 | 0.41 | 2.26 | 0.22 | 2.44 |
| SMT-08_2 | 1.09 | 0.29 | 1.52 | 0.37 | 1.77 |
| SMT-08_3 | 2.10 | 0.73 | 3.37 | 0.87 | 3.91 |
| SMT-09_1 | 1.10 | 0.35 | 1.70 | 0.14 | 1.89 |
| SMT-09_2 | 0.96 | 0.38 | 1.66 | 0.10 | 1.87 |
| SMT-09_3 | 2.25 | 0.60 | 3.16 | 0.18 | 3.73 |
| SMT | 1.63 | 0.94 | 2.86 | 0.10 | 10.24 |
| Tracking Results | 1.62 | 0.93 | 2.84 | 0.10 | 10.24 |

# Part II

# Ultrasound and computed tomography fusion for guidance

# Chapter 6

## 3D LBP-based Rotationally Invariant Region Description

Based on:

Jyotirmoy Banerjee, Adriaan Moelker, Wiro J. Niessen and Theo van Walsum, 3D LBP-based Rotationally Invariant Region Description, *ACCV Workshops*, 2012.

**Abstract**

Local binary patterns [LBP] [96] are popular texture descriptors in many image analysis tasks. One of the important aspects of this texture descriptor is their rotational invariance. Most work in LBP has focused on 2D images. Here, we present a three dimensional LBP with a rotational invariant operator using spherical harmonics. Unlike Fehr and Burkhardt [30], the invariance is constructed implicitly, without considering all possible combinations of the pattern. We demonstrate the 3D LBP on phantom data and a clinical CTA dataset.

Visual tasks such as detection, localization, categorization, and recognition are important subjects of study in computer vision and image analysis. These tasks are often difficult due to apparent within-class inhomogeneity or variability. Part of this within-class variability may be due to the image formation process. Invariant image descriptors extract information from images which is invariant to the variability introduced due to the imaging process, such as noise, distortions, illumination, scale changes, occlusion, etc. One class of such descriptors is texture patterns. Texture have received considerable attention [128] [126] [40] with application in areas of medical imaging [91] [121], image retrieval, remote sensing and object recognition [18]. The local binary patterns [LBP], introduced by Ojala et al. [95], is an efficient method for texture description in 2D. The aim of our work is to extend the conventional LBP and its rotational invariant property mentioned in [96], to a 3D paradigm.

*LBP* - LBP is a simple and computationally efficient way to describe local image content, with impressive texture discriminative properties. Applications of LBP descriptors are evident in texture classification and face analysis [53]. Though it encapsulates textural information, the conventional LBP operator has a number of limitations which are discussed by Liu et al. [78]. The prominent disadvantages are: weak spatial support and sensitivity to noise. Ojala et al. [96] addressed the first issue by introducing a multi-resolution framework. The sensitivity to noise was addressed by grouping the noisy patterns into one bin and defining the remainder of the patterns as "uniform", corresponding to binary label sequence that has no more than two transitions between "0" and "1" among all pairs of the adjacent binary labels. However, in practice this is an oversimplifying assumption. The uniform LBPs extracted from texture images having more complicated shapes may not necessarily be the patterns dominating the texture. Lioa et al. [77] proposed a method that makes use of the most frequently occurring patterns to capture textural information. The frequently occurring or dominant patterns are estimated from training examples. An adaptive framework was proposed by Guo et al. [44] to obtain most discriminative patterns.

*Complementary measures* - To boost the descriptive power of LBP, several complementary measures were proposed. Ojala et al. [96] included local contrast. Guo et al. [43] and Liu et al. [78] incorporated intensity information, considering the intensities of the center pixel and those of its neighbors. Orientation information was incorporated by Chen et al. [19]. Nanni et al. [91] considered different shapes for neighborhood calculations.

*Rotational Invariance* - Rotational invariance was originally described by Ojala et al. [96], where the pixel pattern is circularly bit-wise right shifted and the unique identifier is minimum of the generated patterns. Guo et al. [44] consider a rotation-invariant strategy from nonrotation-invariant histograms of LBPs. The method keeps the original rotation-variant features but finds a matching strategy to deal with the rotation. Invariance is globally constructed in Zhao et al. [138] for the whole region

by histogramming noninvariant LBPs. Unlike Ojala et al. [96], they achieve rotational invariance implicitly in the Fourier domain, without considering all possible combinations of the patterns. Their frequency domain representation of LBP histograms is a band-limited representation which ignore higher frequencies. This is shown to be robust to other histogram-based invariant texture descriptors, which normalize rotation locally. However, smaller footprints or regions would lead to sparse histograms. Fourier representation of sparse signals are not conducive to similarity measures.

*3D LBP* - Recently there has been interest in dynamic texture analysis. LBP descriptors were proposed to deal with rotations and view variations in video. They essentially analyze dynamic texture in 2D time series. Zhao et al. [138] have designed invariance for 2D images and extended to 2D time series using bi-planes. Extending LBP to full 3D volume presents few challenges. A circle in 2D would translate to a sphere in 3D. Equidistant sampling on a sphere is not as trivial as on a circle. The notion of ordering is lost in 3D because of the dimensionality, which was an essential step in calculating rotational invariance in 2D. Fehr and Burkhardt [30] proposed a rotationally invariant LBP on volume data. For each LBP computation, correlation between the gray values of all points on the neighborhood sphere with radius $R$ and the weight factor which is a volume representation in an arbitrary but fixed order binomial factors $\{2^0, \ldots, 2^{P-1}\}$, is performed in the spherical harmonic domain. Similar to Ojala et al. [96], rotational invariance is achieved from the computation of the minimum over all angles.

*Our method* - In this work, we present a rotationally invariant 3D LBP, where unlike Fehr and Burkhardt [30], the invariance is constructed implicitly, without considering all possible combinations of the pattern. Spherical harmonics is the mathematical foundation behind our computation [42]. The theory of spherical harmonics states that any rotation of a spherical function does not change its $L_2$-norm [58]. These features capture invariance to rotation, however with some ambiguity. There will be different signals having similar $L_2$-norm [29]. The ambiguity is due to loss of phase information. Fehr [29] additionally used bispectrum to address the issue. We choose a simple statistical measure, which encodes the phase angle of a signal. Gluckman [39] has shown that the phase information in an image has relationship with the non-Gaussian statistics, such as kurtosis.

Similar to the Fourier representation of Zhao et al. [138], our harmonic representation of LBP, increases spatial support. However, unlike Zhao et al. [138], where invariance is calculated globally, our method estimates it locally. This is useful in describing regions with small footprints. Our method is a complete three dimensional rotationally invariant modeling of LBP.

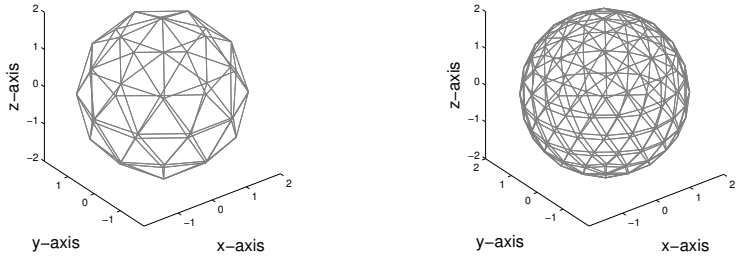



**Figure 6.1:** *Icosahedron Spherical Sampling. Left Fig.: 42 Sample points ($P_1$), Right Fig.: 162 Sample points ($P_2$).*

# 6.1 Method

We present a method for rotationally invariant description of landmarks or regions in 3D using LBP. The LBP in 3D requires a spherical sampling, which is represented in a spherical harmonics framework [42]. The framework helps in obtaining rotation invariant representation. Further, the region information is collected to a set of histograms that are invariant to rotation. The similarity between any two regions can be computed using the Chi-square distance measure [17] between the corresponding set of histograms.

## 6.1.1 Spherical harmonics

Spherical harmonics (SH) is a mathematical framework, generally used to describe a function on a sphere [42]. They are essentially a spherical analog to the Fourier basis. Spherical harmonic functions are defined on imaginary numbers. We are interested in approximating real functions over the sphere, so we will use the real basis of spherical harmonics. The real spherical harmonic function $Y_\ell^m(\theta,\phi)$ of degree $\ell$ and order $m$ is given by

$$Y_\ell^m(\theta,\phi) = \begin{cases} \sqrt{2}K_\ell^m \cos(m\phi) P_\ell^m(\cos\theta) & m > 0 \\ \sqrt{2}K_\ell^m \sin(-m\phi) P_\ell^{-m}(\cos\theta) & m < 0 \\ K_\ell^0 P_\ell^0(\cos\theta) & m = 0 \end{cases} \tag{6.1}$$

where $P$ is the Associated Legendre polynomials and $K$ is

$$K_\ell^m = \sqrt{\frac{(2l+1)}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}$$

and $\theta, \phi$ are azimuthal, polar angles respectively.

Projecting spherical harmonic functions into spherical harmonic coefficients is straight forward. To calculate a coefficient for a band we integrate the product of the function $f$ and the spherical harmonic function $Y$, in effect projecting how much the function is like the basis function:

$$c_\ell^m = \oint f(\theta,\phi) Y_\ell^m(\theta,\phi) d\Omega$$

where $\Omega$ represents the sphere and $(\theta,\phi) \in \Omega$.

The function can be reconstructed to a band-limited approximation ($n$-th order), by reversing the above step

$$\tilde{f}(\theta,\phi) = \sum_{\ell=0}^{n-1} \sum_{m=-\ell}^{m=\ell} c_\ell^m Y_\ell^m(\theta,\phi)$$

### 6.1.2 Three Dimensional Rotational Invariant LBP

For convenience we use the similar notation as used by Ojala et al. [96]. Texture representation $f^T$ in a local neighborhood of a monochrome volume is defined as the joint distribution of the binary values of $P$ voxels:

$$f^T \approx t(s(g_0 - g_c), s(g_1 - g_c), \ldots, s(g_{P-1} - g_c))$$

where

$$s(x) = \begin{cases} 1 & \text{if } x \geqslant 0, \\ 0 & \text{if } x < 0. \end{cases}$$

and gray value $g_c$ corresponds to the gray value of the center voxel of the local neighborhood and $g_p (p = 0, \ldots, P-1)$ correspond to the gray values of $P$ equally spaced sample points on a sphere of radius $R$, around the center voxel.

Spherical harmonics can be used to obtain a rotation invariant representations [58] in 3D. As any rotation of a spherical function does not change the $L_2$-norm, a set of equally spaced pixels on a sphere of radius $R$ can be represented using an index invariant to rotation.

We define the rotationally invariant local binary pattern per voxel as

$$LBP_{P,R}^{ri3D} = \{\|f_0\|, \|f_1\|, \ldots, \|f_\ell\|, \ldots\} \tag{6.2}$$

where $f_\ell$ are the frequency components [58] of a function $f^T$, given by

$$f_\ell(\theta,\phi) = \sum_{m=-\ell}^{m=\ell} c_\ell^m Y_\ell^m(\theta,\phi)$$

and

$$\|f_\ell\| = \Big( \sum_\theta \sum_\phi |f_\ell(\theta,\phi)|^2 \Big)^{1/2}$$

where $(\theta, \phi)$ is an element of the spherical sampling scheme $\Omega_S$.

However, as mentioned by Fehr [29], the power spectrum features are ambiguous, i.e. two different signals may have the same power spectrum. The ambiguity is due to loss of phase information. Variance cannot be used for discrimination, as same power spectrum shares the same variance. We propose the use of kurtosis to address the ambiguity. Gluckman [39] in his work has reported relationship between the phase angle of a signal and the non-Gaussian statistics, kurtosis. It is shown that both local and global correlations in the phase angle lead to many of the statistical regularities, such as kurtosis.

Kurtosis measures how sharply peaked a distribution is, relative to its width, and is defined as

$$\kappa = \frac{\mu_4}{{\mu_2}^2} - 3$$

where $\mu_i$ denotes the $ith$ central moment and in particular, $\mu_2$ is the variance. The kurtosis is normalized to zero for a Gaussian distribution.

The 3D rotationally invariant texture feature per voxel is then described as

$$\kappa LBP^{ri3D}_{P,R} = \{LBP^{ri3D}_{P,R}, \kappa\}. \tag{6.3}$$

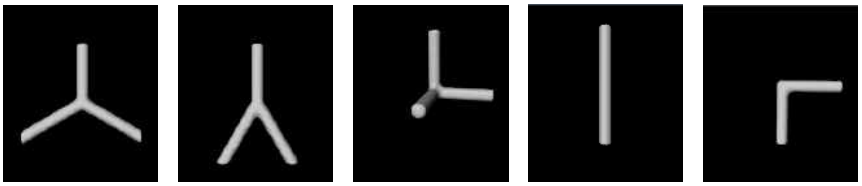
Note that the spherical harmonics are performed on the binary texture pattern $f^T$ and kurtosis $\kappa$ is estimated over the gray level intensity distribution obtained from the spherical sampling.

### 6.1.3 Spherical sampling in 3D

Three dimensional LBP construction requires sampling over a sphere of radius $R$. It is non-trivial to have an equidistant sampling over a sphere. To approximate this we use the icosahedron. Icosahedron structure is used to sample the surface of the sphere. To make an icosahedron approximate a sphere more closely, the triangles making up the icosahedron can intuitively be subdivided by splitting the edges of the triangle and then making the new split edges into more triangles [37]. The frequency component $f$ represents how many times the struts of the base icosahedron have been subdivided. Icosahedrons of frequency 2 and 4 as shown in Figure 6.1 have $10f^2 + 2$ number of vertices, i.e. 42 and 162 sample points, respectively. In the voxel grid, trilinear interpolation is used to estimate the gray value at the vertices of the icosahedron.

### 6.1.4 Histogram matching

The 3D rotationally invariant texture feature is a set of real numbers (see Equation 6.3) per voxel, where the number of elements in the set is equivalent to the number of spherical harmonic bands plus one. The additional term is due to the inclusion of kurtosis. After the rotationally invariant texture feature for each voxel



(a) (b) (c) (d) (e)

**Figure 6.2:** *Surface rendered images of 3D vessel-like structures - (a) All angles are* $120^\circ$*, (b) One of the angle is* $60^\circ$*, other angles are* $150^\circ$ *each, (c) All angles are* $90^\circ$*, (d) Straight and (e) Angle is* $90^\circ$*. The volume has binary intensity values. The vessel shapes have voxel intensity 1 and the rest of region in the volume have voxel intensity 0. The shapes are constructed out of cylindrical branches with diameter* $D$ *each. Shapes in (a), (b) and (c) are made of three cylindrical arms, shape in (d) is made of one cylindrical arm and shape in (e) is made of two cylindrical arms.*

is estimated according to Equation 6.3, a set of histograms can be built to represent the texture region. Each histogram corresponds to an element in the rotationally invariant texture feature. Similarity/dissimilarity between two regions can be estimated using the distance measure between the histograms. The Bhattacharyya measure (or coefficient) and Chi-square measure [17] are two popular measures of similarity between two distributions. The final matching score is derived by adding scores from all the histogram pairs.

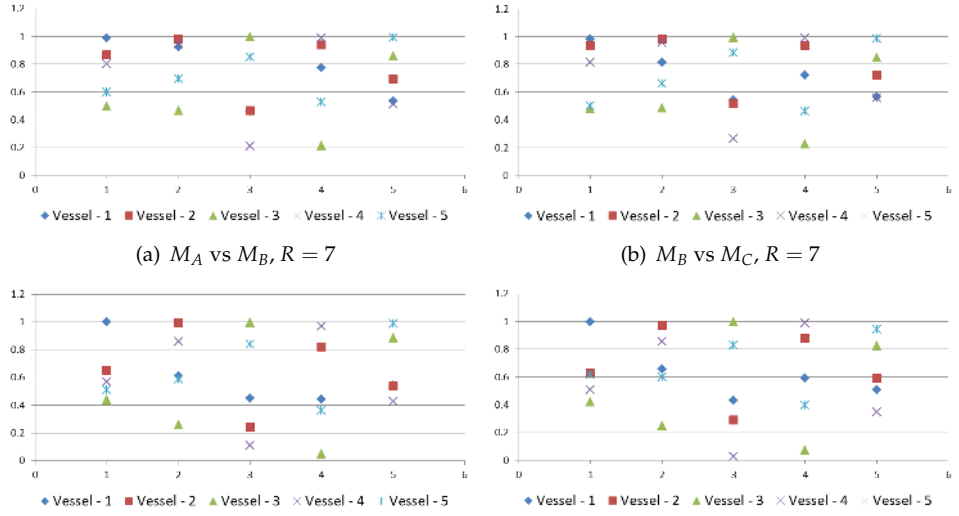
## 6.2 Experiments and results

We investigated the properties of both the rotationally invariant features and their application in histogram-based region descriptors, using phantom data. Additionally, we demonstrate the application of our method to the localization of landmarks in medical imaging data.

### 6.2.1 Evaluation of rotational invariant features

Purpose of the first experiment is to investigate to what extent our propose features are indeed rotationally invariant, and can be used to discriminate between various 3D patterns. To this end, we compute the correlation of the rotationally invariant features (Equation 6.3) on several rotated versions of phantom volumes with different embedded vessel-like structures (see Figure 6.2).

*Phantom volumes with different embedded vessel-like structures* - The images shown in Figure 6.2 are the surface rendered images of five cubic digital phantoms with different embedded vessel-like structures. Vessels shown in Figure 6.2 constitute foreground while the rest is background. The volumes have binary intensity levels, with foreground and background intensity levels as 1 and 0, respectively. The cen-




(a) $M_A$ vs $M_B$, $R = 7$ (b) $M_B$ vs $M_C$, $R = 7$

(c) $M_A$ vs $M_B$, $R = 9$ (d) $M_B$ vs $M_C$, $R = 9$

**Figure 6.3:** *Voxel level texture feature correlation between different vessel-like structures in Figure 6.2; Parameters - Diameter of the cylindrical arms $D = 11$ (in Figure 6.2); $P_2 = 162$ sample points; Sampling Radius $R$.*

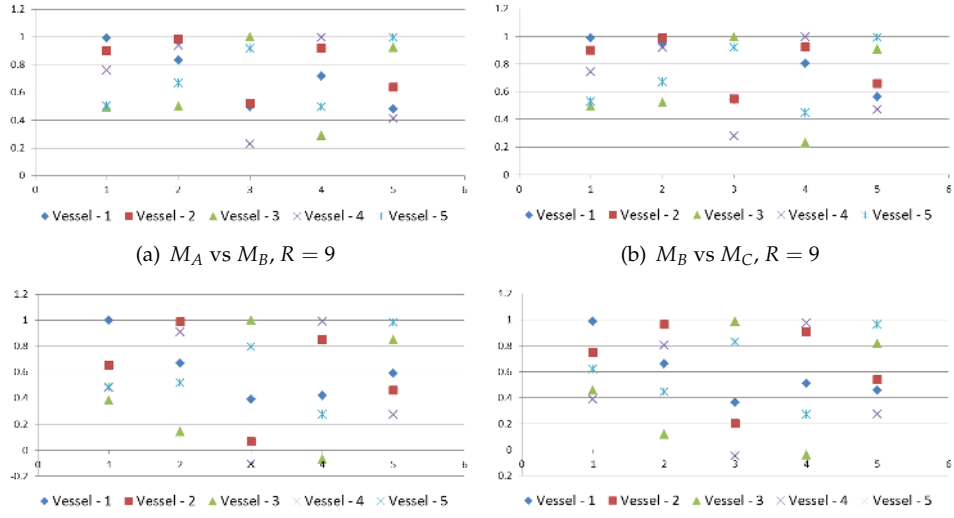
| Rotation Matrix | Center | Axis (x, y, z) | Angle (in radians) |
|---|---|---|---|
| $M_A$ | $(63.5, 63.5, 63.5)$ | - | - |
| $M_B$ | $(63.5, 63.5, 63.5)$ | $(0, 0, 1)$ | $\pi/2$ |
| $M_C$ | $(63.5, 63.5, 63.5)$ | $(1, 1, 1)$ | $\pi/3$ |

**Table 6.1:** *Rotation Matrix*

troids of the different shapes are located at $(63, 63, 63)$, in their respective volumes of size $[128 \times 128 \times 128]$. The vessel-like structures are constructed from cylindrical branches with diameter $D$ each.

In Figure 6.2(a) all angles between the vessel branches are 120°. In Figure 6.2(b) one of the angle is 60°, other angles are 150° each. In Figure 6.2(c) all angles between the vessel branches are 90°. Figure 6.2(d) vessel structure has no branches and is linearly oriented. Figure 6.2(e) the vessel structure is bent at 90°. Shapes in Figures 6.2(a), 6.2(b) and 6.2(c) are made of three cylindrical arms, shape in Figure 6.2(d) is made of one cylindrical arm and shape in Figure 6.2(e) is made of two cylindrical arms.

*Rotation matrix* - The rotation matrices used to rotate the phantoms shown in Figure 6.2 are described in Table 6.1. The rotation center coincides with the center



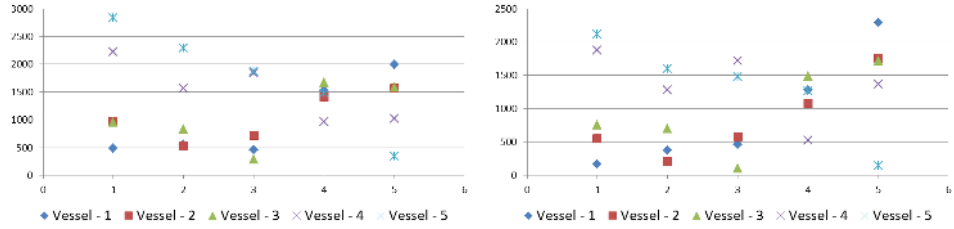

(a) $M_A$ vs $M_B$, $R = 9$ (b) $M_B$ vs $M_C$, $R = 9$

(c) $M_A$ vs $M_B$, $R = 13$ (d) $M_B$ vs $M_C$, $R = 13$

**Figure 6.4:** *Voxel level texture feature correlation between different vessel-like structures in Figure 6.2; Parameters - Diameter of the cylindrical arms $D = 15$ (in Figure 6.2); $P_2 = 162$ sample points; Sampling Radius $R$.*

of the vessel structure. Rotation matrix $M_A$ denotes no rotation. Matrix $M_B$ represents rotation along $z$-axis by $\pi/2$ radian angles. Matrix $M_C$ represents rotation along $(1,1,1)$ axis by $\pi/3$ radian angles. The rotation matrices when applied to the five volumes shown in Figure 6.2, rotate the shapes around the center of the vessel structure. The voxel location and the rotation center in our evaluation are $(63,63,63)$ and $(63.5,63.5,63.5)$, respectively.

We evaluate the rotational invariant property of the texture feature (see Equation 6.3), evaluating it per voxel location. Various intensity patterns can be obtained by sampling around any given voxel from the different volumes in Figure 6.2. For evaluation purpose we choose the centroids of the vessel structures appearing in these volumes. The results are shown in Figure 6.3 and Figure 6.4. They show the correlation between the $\kappa LPB$ index per voxel location obtained from the five different volumes in Figure 6.2, and their rotated versions. Various intensity patterns can be generated from Figure 6.2 by varying the sampling radius $R$ or the diameter $D$ of the cylindrical arms of the vessel-like structures. In Figure 6.3 we use a sampling radius $R$ of 7 and 9, keeping the diameter $D$ at 11. In Figure 6.4 we use a sampling radius $R$ of 9 and 13, changing the diameter $D$ to 15. To ensure different intensity patterns with each sampling, the sampling radius is chosen greater than the radius of the cylindrical arms in Figure 6.2. In Figure 6.3 and Figure 6.4 the




(a) $P_1 = 42$ sample points. (b) $P_2 = 162$ sample points.

**Figure 6.5:** *Chi-square distance measure between five regions from Figure 6.2, and their rotated versions (Rotation Matrix $M_C$). Parameters - Diameter of the cylindrical arms $D$ (in Figure 6.2) = 11; Sampling Radius $R = 9$; Descriptor Radius $K = 5$.*

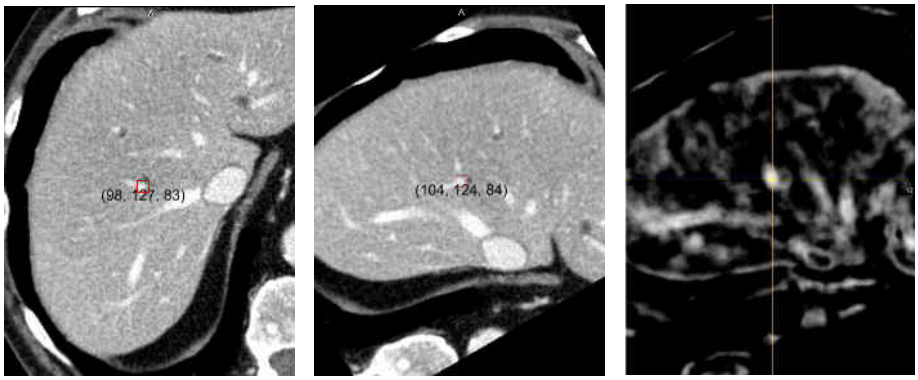
$\kappa LPB$ feature of a voxel from the volumes in Figure 6.2, correlates very well with its rotated version. Correlation ranges from $-1.00$ to $+1.00$. A correlation of 1.00, is a perfect correlation, while a correlation of 0 means there's no relationship between the two variables.

### 6.2.2 Evaluation of region descriptors

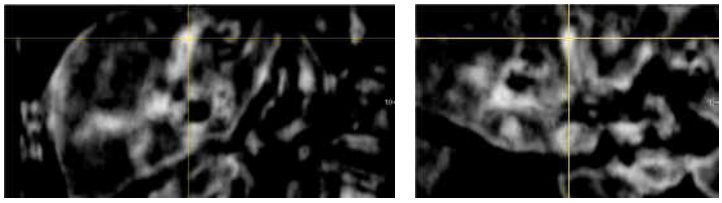
In the next experiment, we investigate the rotationally invariant property of the histogram-based region descriptors. To evaluate the rotational invariance property of the histogram-based region descriptors, the region ought to be spherical. The radius of spherical region is the descriptor radius $K$. We select five spherical regions from the volumes in Figure 6.2. The center of the regions coincides with the centroid of the vessel-like structures. Since the rotationally invariant texture feature (see Equation 6.3) is a set of real numbers, the region description is accumulated to multiple histograms, each histogram corresponding to one of the texture features. In Figure 6.5(a) and Figure 6.5(b) the number of sample points $P$ are 42 and 162, respectively. The descriptor radius $K$ is set to 5. Increase in sampling rate, improves the spherical function approximation, which translates to improved discriminative ability of the descriptor (see Figure 6.5). We use the Chi-square distance [17] to measure the difference between regions from all the volumes and their rotated version, as shown in Figure 6.5. The lower the Chi-square score, the more similar are the histograms. Figure 6.5 shows that the regions correlate very well with its rotated version, and weekly correlates with other regions.

### 6.2.3 Clinical examples / evaluation

To investigate the use of LBP as a 3D texture descriptor in medical image analysis (e.g. [91], [121]), we test our approach to describe a region or landmark in liver CTA volume. The liver CTA volume shown in Figure 6.6(a), is rotated on axis $(0, 0, 1)$




(a) Volume A. (b) Volume B. (c) Axial Dist Map.



(d) Sagittal Dist Map. (e) Coronal Dist Map.

**Figure 6.6:** *Liver CTA volume: A landmark location is selected in Volume A. The same landmark location is then searched in the Volume B; a) Volume A, b) Volume B: Obtained by rotating Volume A on axis* $(0, 0, 1)$ *by* $\frac{\pi}{3}$ *radians, c) The axial Chi-square distance map, d) The sagittal Chi-square distance map and e) The coronal Chi-square distance map. Parameters - Sampling Radius* $R = 5$*; Descriptor Radius* $K = 5$*; Sample points* $P_1 = 42$*.*

by $\pi/3$ angle to obtain the new volume, shown in Figure 6.6(b). The parameters, sampling radius is set to $R = 5$; descriptor radius is set to $K = 5$; and sample points is set to $P_1 = 42$. A landmark location is selected in the left-hand side volume in Figure 6.6(a). The same landmark is then searched in the rotated volume, and as shows in the Figure 6.6(b) we are able to retrieve back the landmark location. The Chi-square distance map for the axial, sagittal and coronal planes are shown in Figure 6.6(c), Figure 6.6(d) and Figure 6.6(e), respectively. The distance maps have a peak at the selected landmark location. Figure 6.7 show the histograms of the descriptor at various locations in the liver CTA volume like thin vessel structure, large vessel structure, liver tissue and liver boundary. The histograms show the first three elements of the rotationally invariant feature from the Equation 6.2. The histogram-based region descriptor has different signature for different structural location in the liver CTA volumes, which show its discriminative ability in visual classification tasks.

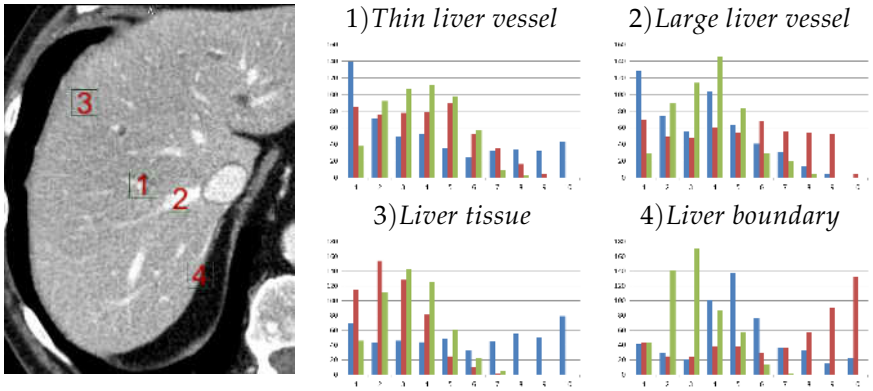



**Figure 6.7:** *Histogram descriptors of various locations in liver CTA volume. Histograms show the first three features from the Equation 6.2. Parameters - Sampling Radius $R = 5$; Descriptor Radius $K = 5$; Sample points $P_1 = 42$.*

## 6.3 Discussion and Conclusion

Volumetric data, like their 2D counterpart have an inherent textural property. The textural property of a 3D region can aid in its region description. LBP which was conventionally designed for image texture description is extended in this work to region description of volumetric data. The rotational invariance property helps in view invariant region detection. We show the application of our histogram-based region descriptors in distinguishing various vessel-like structures in phantom data and landmark detection in medical imaging data.

The method presented has several parameters that need to be determined appropriately. The LBP parameters, spherical sample points $P$ and sampling radius $R$ are related as the spherical neighborhood corresponding to a given $R$ contains a limited number of non-redundant sample points. To capture the vessel-like structures well (Figure 6.2) in the phantom experiments, the sampling radius $R$ is chosen greater than the radius of the cylindrical arms. The 3D region descriptor covers a spherical region of radius $K$ and is described in terms of multiple histograms. Histogram bins provide an estimate of the number of corresponding texture patterns. It is relevant to choose the appropriate number of bins and their range that appear in the histogram carefully, as they directly affect the distance measure. If we use too few bins, the histogram doesn't really portray the data very well. If we have too many bins, we get a broken comb look, which also doesn't give a sense of the distribution. Care should be taken so that there are enough bins and are utilized well.

To measure the similarity between two regions we use the Chi-square histogram distance metric, since it is popular in previous works [138] [44]. During experimen-

tations we considered the Bhattacharya distance; however from our pilot experiments we found Chi-square metric to be more discriminative.

In our work we focus on a rotationally invariant representation of the 3D LBP. Many of the proposed complementary measures [43] [78] [19] are independent of the LBP representation, and thus could easily be integrated in our 3D approach.

Concluding, we presented a method for rotationally invariant 3D LBP, using spherical harmonic decomposition. We applied the method on vessel-like phantom data and a clinical dataset, with encouraging results. More in-depth analysis and integration of complementary measures is part of future work.

# Chapter 7

## Multiple-correlation similarity for fast CT and ultrasound fusion

Based on:

Jyotirmoy Banerjee, Camiel Klink, Renske Gahrmann, Wiro J. Niessen, Adriaan Moelker and Theo van Walsum, Multiple-correlation similarity for block-matching based fast CT and ultrasound fusion in liver interventions, *submitted*

**Abstract**

In this work we present a fast approach to perform registration of CT and ultrasound volumes for image guided intervention applications. The registration method is based on a novel similarity metric, where the intensity and the gradient of the CT images along with the US volumes are the input images. The multimodal similarity metric uses multiple correlation coefficient in a block-matching framework, to find correspondences between regions in the CT and the US volumes. Geometric consistency and smoothness criteria are imposed to refine the block-matching results. The improved block-matching results are used to determine the affine transformation between the CT-US volumes. The training dataset consists of 8 CT and US pair from 5 patients and the testing dataset consists of 21 CT and US pairs acquired from 12 patients. A mean registration error of 3.5 and 4.3 mm is achieved over the training and testing dataset, respectively.

## 7.1 Introduction

### 7.1.1 Clinical motivation

Computed tomography (CT) is a popular diagnostic imaging technique with high image resolution and, due to fast imaging speed, few motion artifacts. Vessels are not always clearly visible in CT. To image blood vessels, scans are performed after the introduction of the contrast material. Contrast enhanced CT scans are frequently used for tumor diagnosis and treatment planning of the tumors (cancer).

Unlike CT, ultrasound (US) does not use ionizing radiations. It uses high frequency sound waves for imaging and is used to image soft tissues, e.g. in the abdomen. With recent matrix array probes US provides real-time 3D imaging. It can be potentially used as an intraoperative imaging modality, e.g. in thermal ablation tumor therapy. However, tumors are not always clearly visible in US. Preoperative CT can be combined with 3D US during interventions to incorporate complementary information. Additionally, incorporating CT data during intervention will add greater definition and precision to the US based navigation system.

The purpose of our work is to enable live overlay of CT in US guided tumor ablations. Towards this end we propose a fast registration technique between (preoperative) CT and (intraoperative) real-time 3D US data. Such a method would provide real-time virtual image overlay of CT on 3D US and thus help the clinicians to accurately visualize the target during the intervention. For example, with CTA overlay the tumor location can be precisely visualized on the US images during the intervention, which is very relevant as the tumor is often difficult to visualize using only US.

### 7.1.2 Related work

In literature there have been attempts to address the registration of CT and US volumes. Lee et al. [72] used an electromagnetic tracking based fusion imaging of real-time US and CT/MR images for percutaneous hepatic intervention of the liver. Crocetti et al. [25] performed a feasibility study of CT-US fusion imaging system on ex vivo (calf) livers using a commercially available multimodality fusion imaging system. Recent advancements in treatment of liver tumors using image guided ablation techniques are discussed in Kang et al. [55].

There exist commercial interventional software solutions that perform CT-US registration. These systems typically use external tracking systems (e.g., electromagnetic, optical) to co-register multimodal images. Optical tracking requires clear line-of-sight between cameras and targets; hence objects inside the body cannot be tracked. Electromagnetic tracking does not require line-of-sight to operate, however are sensitive to ferromagnetic objects. These external tracking based systems provide limited support for soft tissue regions like the abdomen.

Registration methods using image information do not suffer from the limitations of tracking based approaches. Lange et al. [67] proposed an algorithm that uses intensity and few additional landmark positions to perform non-rigid registration of preoperative CT and intraoperative US volumes. They used the normalized gradient field (NGF) to measure the similarity. To automate the estimation of the initial transformation a graph based edge matching algorithm was proposed by Nam et al. [90], where the edges are extracted from the segmentation of vessels. Their approach performs registration of US and CT volumes of the liver regardless of the patient or probe motion. However, robust segmentation of vessels in US is a challenging task. Nagpal et al. proposed a combination of intensity and point based registration methods to perform multi-vertebrae CT to US registration of the lumbar spine [89]. They used a spring model to constrain the movement of the individual vertebraes. In Penney et al. the ultrasound and the MR images are converted into vessel probability maps, which are then registered using normalized cross-correlation similarity metric [101]. Apart from the fact that accurate detection of vessels in ultrasound is a challenging task, the method does incorporate information like liver surface and liver tissue characteristics in the registration task.

Others have tried to directly register US images to CT images. As the physics behind US image formation is different from CT, the choice of the similarity metric for the registration algorithm is non-trivial and critical. Mutual information (MI) is the most generic similarity metric used in multimodal registration where statistical dependency between the modalities is used [52, 103].

The correlation-ratio (CR) similarity metric uses the functional relationship $y = f(x)$ between the modalities [109]. To address the spatial intensity inhomogeneity in the US images, Rivaz et al. used a local estimate of the CR similarity metric using patch based registration approach [108]. A bivariate version of the CR was first introduced by Roche et al. [110]. It assumes a polynomial functional relationship of the form $y = f(x_1, x_2)$, where $x_1$ is the intensity image and $x_2$ is the gradient image. The gradient of a CT or Magnetic resonance (MR) image is considered to approximately replicate the US imaging as the impedance mismatch at tissue borders results in a bright signature between tissues in the US images. Wein et al. [136] and Fuerst et al. [36] assume a linear function of the form $y = ax_1 + bx_2 + c$, where the coefficients of the function ($a$, $b$ and $c$) are estimated locally during the similarity metric evaluation.

In Wein et al. [136] simulation of US type physical phenomena is used to generate pseudo US images from CT data. The simulated US images are then registered to the real-time US images. In MR imaging, there is a less direct relationship between the voxel intensities and tissue characteristics relevant for US imaging. To that end, Kuklisova-Murgasova et al. [64] pre-segments MR volumes before converting them into pseudo US volumes, which are further registered to US fetal brain volumes.

Self-similarity based methods exploit neighborhood relationships. These mea-

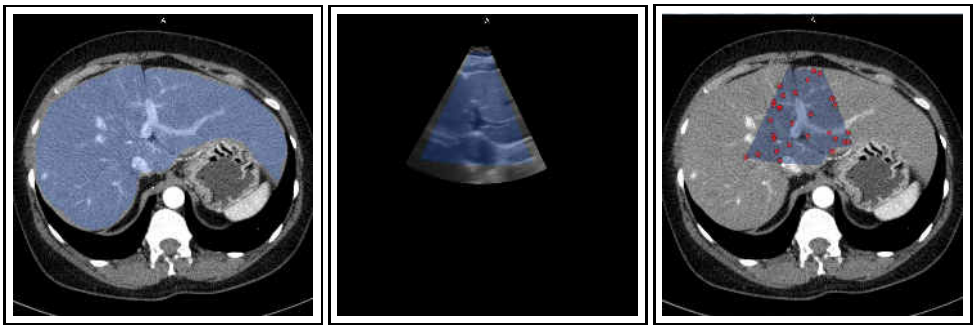


**Figure 7.1:** *Input points for block-matching: Left - CTA volume and overlayed liver mask, Middle - US volume and overlayed US mask, Right - Overlayed intersection of US and liver mask and input points for the block-matching.*

sures are relative and independent of the imaging modality. They are inspired by the success of the image denoising methods like non-local means [13, 14, 15]. As corresponding regions in different imaging modalities have a similar neighborhood relationship, they share similar self-similarity descriptors/signatures. The descriptors are extracted and compared across modalities using monomodal similarity metrics like sum of squared distances (SSD) or sum of absolute differences (SAD). The modality independent neighborhood descriptor (MIND) was used in Heinrich et al. [49] for the registration of CT and MR scans. A similar self-similarity based method was later used by Heinrich et al. [50] to register 3D US and MRI scans for neurosurgery.

### 7.1.3 Our contributions

This work has four main contributions: First, we propose a fast multiple correlation coefficient based similarity metric to perform multimodal registration of CT and US volumes. Second, we discuss an efficient implementation of this new similarity metric, which unlike MI and CR and similar to SAD and SSD, can be computed in a single pass. To the best of our knowledge this similarity metric is for the first time used in multimodal registration. Third, to improve the block-matching results, we extend our previously presented outlier rejection method based on a geometric consistency [6] with smoothness criteria. Fourth, we evaluate the approach on a large set of patient data.

The remaining of the chapter is structured as follows. In Section 7.2 we discuss the CT and real-time 3D US registration approach. The experiments are described and the results are presented in Section 7.3. The results are analyzed and discussed in Section 7.4.

## 7.2 Method

The method consists of three steps - selecting a set of input points for the block-matching, followed by block-matching based on a similarity metric and an outlier rejection step to remove the false correspondences from the block-matching. These steps are similar to the earlier work presented in Banerjee et al. [6] for the US-US registration. For multimodal CT-US registration, we propose the following two changes to the previous framework - first, we propose a fast multimodal similarity metric for the multimodal block-matching and second, the outlier rejection step is adapted and improved over the earlier work presented in Banerjee et al. [6].

In our approach, for a set of points located in the reference/fixed image, block-matching is used to find corresponding locations in the moving image using a *multiple correlation coefficient* based multimodal similarity metric. The correspondences from the block-matching are inputs to the outlier rejection scheme. The outlier rejection scheme uses geometric consistency and smoothness criteria to determine the block-matching results that can be trusted. The method then uses only the selected block-matching results from the outlier rejection scheme to estimate a rigid transformation using Arun et al. [4] and an affine transformation using Späth [122].

### 7.2.1 Selecting point set for block-matching

The registration method is based on block-matching strategy [22, 97]. Block-matching helps in finding a set of corresponding location form the reference image $\{p_i\}$ to the moving image $\{q_i\}$, where $1 \leq i \leq n$ and $n$ is the number of elements in the set. Input point set locations for the block-matching are selected randomly from the region of interest (ROI). The ROI is the intersection of the following regions - liver mask and (initialized/roughly aligned) US volume mask, see Figure 7.1. Manual segmentation was performed to obtain the liver mask from the CTA volume, but (semi-)automated methods [48] could be applied as well.

### 7.2.2 Multimodal block-matching

We use a multiple correlation coefficient (MCC) similarity metric for the multimodal block-matching. The multiple correlation coefficient (MCC) finds the correlation between the target variable $y$ and the multiple predictor variables $x_1, x_2, \cdots, x_m$ using a linear function. The square of the coefficient of multiple correlation $\mathcal{R}$ is given as:

$$\mathcal{R}^2 = c^\intercal R_{xx}^{-1} c\,, \tag{7.1}$$

where

$$R_{xx} = \begin{bmatrix} r_{x_1x_1} & \cdots & r_{x_1x_m} \\ \vdots & \ddots & \vdots \\ r_{x_mx_1} & \cdots & r_{x_mx_m} \end{bmatrix},$$

$c = \begin{bmatrix} r_{x_1 y} & \cdots & r_{x_m y} \end{bmatrix}^\mathsf{T}$ and $r_{uv}$ is the correlation between the variables $u$ and $v$.

Similar to Wein et al. [136] and Fuerst et al. [36] we assume a linear functional relationship between the CTA intensity volume $x_1$, the CTA gradient volume $x_2$ and the US volume $y$, where $y = ax_1 + bx_2 + c$. However, unlike the regression based method described in [36, 136] we determine the strength of the association between the variables in an implicit way using a correlation based method. Similar to the normalized correlation coefficient (NCC) an estimate of the coefficients of the linear function is not required to estimate the correlation coefficient in MCC. For MCC with two predictor variables $x_1$ and $x_2$, Equation 7.1 can be rewritten as:

$$\mathcal{R}^2 = \frac{r_{x_1 y}^2 + r_{x_2 y}^2 - 2 r_{x_1 y} r_{x_2 y} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}, \tag{7.2}$$

where $\mathcal{R} \in [0, 1]$. $\mathcal{R}^2$ is also called coefficient of determination.

The correlation coefficient $r_{uv}$ is calculated using a single-pass approach [28]:

$$r_{uv} = \frac{s \sum u_i v_i - \sum u_i \sum v_i}{\sqrt{s \sum {u_i}^2 - \left(\sum u_i\right)^2} \sqrt{s \sum {v_i}^2 - \left(\sum v_i\right)^2}}, \tag{7.3}$$

where $s$ is the number of samples. Substituting Equation 7.3 in Equation 7.2, results in a single-pass MCC expression for two predictor variables, which can be efficiently implemented on a graphics processing unit (GPU).

### 7.2.3 Outlier rejection

Here we first briefly review the geometric constraint used in the outlier rejection step, originally proposed for the US-US registration [6]. Subsequently, we propose an additional spatio-temporal smoothness constraint to improve the outlier rejection step, required for a robust CT-US registration.

The outlier rejection method that detects the false correspondences from the previous multimodal block-matching steps is based on the following two criteria:

*Geometric criterion:* The geometric consistency criterion preserves pairwise distance between locations in the reference image $P = \{p_i\}$ and their corresponding locations in the moving image $Q = \{q_i\}$, where $1 \leq i \leq n$. The pairwise information is encoded in an adjacency matrix $\mathbf{G}$, where an element of the matrix is given as:

$$G_{i,j} = \begin{cases} e^{-{\delta_{i,j}}^2 / 2{\sigma_\mathrm{G}}^2} & \text{if } i \neq j \\ 0 & \text{otherwise}, \end{cases} \tag{7.4}$$

where

$$\delta_{i,j} = \frac{|\|q_i - q_j\| - \|p_i - p_j\||}{\|q_i - q_j\| + \|p_i - p_j\|}.$$

*Smoothness criterion:* As the geometric consistency criterion does not involve orientation consistency, we added a spatio-temporal smoothness constraint that ensures that the displacement vectors from the block-matching are directionally consistent. Conceptually, this is similar to the piecewise smooth flow field criteria described in the optical flow based method in Brox et al. [12]. However, unlike the optical flow based approach in [12], which tries to find the displacement field between two nearby/consecutive images using a local approach, we address the problem of finding correspondences between two images using a graph-based global approach. The smoothness information is encoded in an adjacency matrix $\mathbf{S}$, where an element of the matrix is given as:

$$S_{i,j} = \begin{cases} e^{-{\zeta_{i,j}}^2/2{\sigma_S}^2} & \text{if } i \neq j \\ 0 & \text{otherwise}\,, \end{cases} \tag{7.5}$$

where

$$\zeta_{i,j} = \frac{|\overrightarrow{(q_i - p_i)} - \overrightarrow{(q_j - p_j)}|}{\|q_i - q_j\|}\,.$$

Notice that for sufficiently close points $q_i$ and $q_j$ in space, $\zeta_{i,j}$ is the magnitude of the gradient of the flow/deformation vector. Combining the geometric consistency and the smoothness constraint, the optimization function to be maximized is given as:

$$(1-\eta)\,\mathbf{w}\cdot\mathbf{G}\mathbf{w} + \eta\,\mathbf{w}\cdot\mathbf{S}\mathbf{w}\,, \tag{7.6}$$

where $0 \leq \eta \leq 1$ and $\mathbf{w} \in \Delta$. $\Delta$ is the unit simplex defined as:

$$\Delta = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w} \geq \mathbf{0} \text{ and } \mathbf{e}^\intercal\mathbf{w} = 1\}\,, \tag{7.7}$$

where $\mathbf{e} = [1, \ldots, 1]^\intercal \in \mathbb{R}^n$. The elements of the stochastic vector $\mathbf{w}$ are nonnegative and sum up to one.

The replicator dynamics update equation to maximize the equation 7.6 is:

$$w_i(t+1) = w_i(t)\frac{(\mathbf{K}\mathbf{w}(t))_i}{\mathbf{w}(t)\cdot\mathbf{K}\mathbf{w}(t)}\,, \tag{7.8}$$

where $\mathbf{K} = (1-\eta)\,\mathbf{G} + \eta\,\mathbf{S}$. The update Eq. 7.8 ensures that $\forall t, \mathbf{w}(t) \in \Delta$, see [135]. A local maximum is reached at the location where the trajectory of Eq. 7.8 does not increase any further, see [100]. Each element $w_i$ of the vector $\mathbf{w}$ states whether the $i$th point is an inlier or not. If $w_i < \epsilon$, then $w_i$ is considered as an outlier and the rest of the points are considered as inliers, where $\epsilon$ is a small positive real number.

**Table 7.1:** *Notations and parameters to optimize in the training experiment.*

| Symbol | Description |
|---|---|
| *Block-matching -* | |
| $n$ | Number of points |
| Block-size | Block-size |
| $\sigma_{CT}$ | Standard deviation of Gaussian derivative, required to calculate CT gradient |
| *Outlier rejection -* | |
| $\sigma_G$ | Standard deviation of geometric term |
| $\sigma_S$ | Standard deviation of smoothness term |
| $\eta$ | Constant in the outlier rejection equation determines the relative weight of the geometric term vs. the smoothness term |

## 7.3 Experiments

### 7.3.1 Data and resources

CT and 3D US were acquired from 17 patients. Informed consent was obtained from all patients prior to the data acquisition. 3D US data was acquired using a Philips iU22 US system with an X6-1 probe. Liver US data was acquired from the intercostal and the sub-xiphoidal regions. The subjects were asked to breathe freely during the image acquisition. The data from 17 patients were divided into training and testing set. The training set consists of 8 CT-US pairs from 5 patients which were randomly picked from the overall dataset and the rest of the data was the testing set. The testing set consists of 21 CT-US pairs from 12 patients. The US volumes were $512\times378\times222$ voxels with voxel size of $0.420\text{mm}\times0.387\text{mm}\times0.629\text{mm}$. The contrast-enhanced CT was acquired in the portal venous phase. The CT slice thickness was 3 mm, slice spacing was 2 mm except for one case, where slice spacing was 3 mm, and pixel size varied from 0.39 mm to 0.75 mm.

MeVisLab, OpenCL and C++ are used for software development. The OpenCL code was run on a NVIDIA GeForce GTX 780 Ti GPU.

### 7.3.2 Evaluation metric

Manual annotations were performed to obtain ground truth registrations. Vessel bifurcations were predominantly used for annotations as corresponding locations could uniquely be identified in CT and US data. The registration error is estimated using the mean target registration error (mTRE), measured in mm. mTRE is the average distance between the ground truth annotations from the fixed image to the moving image.

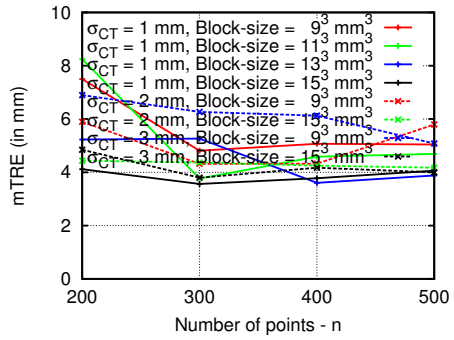



**Figure 7.2:** *Block-matching parameters: Number of points, block-size and $\sigma_{CT}$. The outlier rejection parameters are set as follows: $\sigma_G$ = 0.3, $\sigma_S$ = 0.5 and $\eta$ = 0.3.*

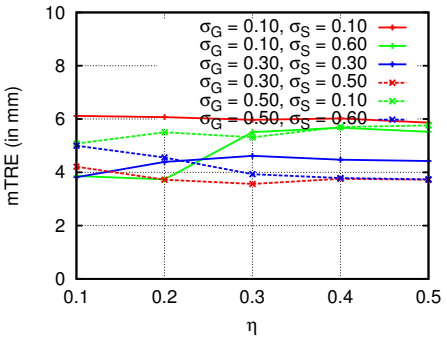



**Figure 7.3:** *Outliers rejection parameters: $\sigma_G$, $\sigma_S$ and $\eta$. The block-matching parameters are set as follows: number of points to 300, $\sigma_{CT}$ to 1 mm and block-size to $15^3$ $mm^3$.*

### 7.3.3 Experiments

We perform the following experiments to evaluate the proposed registration method.

- Parameter tuning on training set: The training set is used to estimate the optimal parameters of the algorithm, see Table 7.1. We first evaluate the block-matching parameters of the algorithm. We investigate the effect of change in a) sigma ($\sigma_{CT}$) used in calculating the gradient of the CT volume, b) block-size and c) number of points used in the block-matching. The parameter ranges for the sigma of the gradient, block-size and number of points were {1 mm, 2 mm, 3 mm}, {$9^3$ $mm^3$, $11^3$ $mm^3$, $13^3$ $mm^3$, $15^3$ $mm^3$} and {200, 300, 400, 500}, respectively. The block-matching search range and search grid spacing are set to $40^3$ $mm^3$ and 1 mm, respectively. Next we evaluate the outlier rejection parameters. We investigate the effect of change in a) $\sigma_G$, b) $\sigma_S$ and c) $\eta$. We apply a known translation (x, y, z) = (10 mm, 10 mm, 10 mm) to the US image which was earlier aligned using the ground truth annotations. The evaluation range of $\sigma_G$ is {0.1, 0.2, 0.3, 0.4, 0.5}, of $\sigma_S$ is {0.1, 0.2, 0.3, 0.4, 0.5, 0.6} and $\eta$ is {0.1, 0.2, 0.3, 0.4, 0.5}. The $\epsilon$ value is set to 1.19209e-07 (machine epsilon value for the floating type). From the parameter evaluation range we decide the optimal parameter settings on the training set data based on the mTRE metric.

- Rotation: In the next experiment we evaluate the robustness of the similarity metric and the block-matching under rotation. We used five CT-US pairs randomly selected from the test set and aligned them based on the ground truth annotation. We rotated the US volume in the range from 0 to 30 degrees with a step size of five degrees, where the axis of rotation was ($x = 0, y = 0, z = 1$). The CT and the corresponding rotated US volumes are then registered.

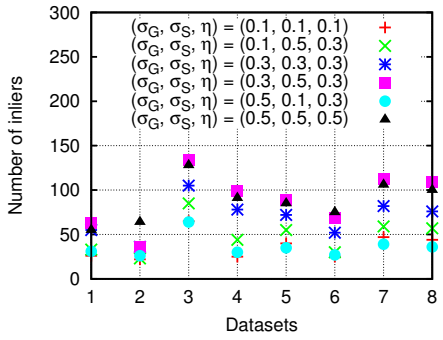



**Figure 7.4:** *Number of inliers for various outlier rejection parameter settings over the eight training set data. Block-matching parameters - search grid spacing is 1 mm, block-size is 15 mm, number of points is 300 and $\sigma_{CT}$ is 1 mm.*

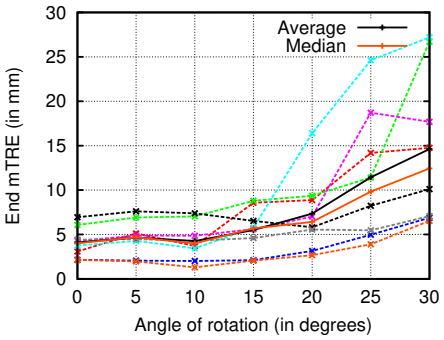



**Figure 7.5:** *Registration results on manually rotated US datasets. Block-matching parameters - search grid spacing is 1 mm, block-size is 15 mm, number of points is 300 and $\sigma_{CT}$ is 1 mm. Outlier rejection parameters - $\sigma_G$ = 0.3, $\sigma_S$ = 0.5 and $\eta$ = 0.3. Each of the dotted lines represents one dataset.*

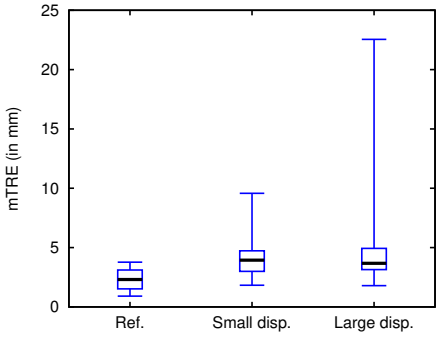



**Figure 7.6:** *Registration results on test set for small (median error of 3.9 mm) and large displacements (median error of 3.6 mm). Ref. - error in ground truth (median error of 2.3 mm).*

- Evaluation on testing set - small displacement: Given the optimal parameter setting we study the robustness of the registration approach for a set of small displacements with respect to the ground truth alignment on the test set. We apply a set of small displacements: (0.7 mm, 0.7 mm, 0.7 mm), (0.7 mm, 0.7 mm, 1.0 mm), (0.7 mm, 1.0 mm, 0.7 mm), (0.7 mm, 1.0 mm, 1.0 mm), (1.0 mm, 0.7 mm, 0.7 mm), (1.0 mm, 0.7 mm, 1.0 mm), (1.0 mm, 1.0 mm, 0.7 mm) and (1.0 mm, 1.0 mm, 1.0 mm).

**Table 7.2:** *Testing set error statistics for small and large displacements: Ref. - error in ground truth, Init. - error before registration and Reg. - error after registration.*

| | | Small displacement | | Large displacement | |
|---|---|---|---|---|---|
| | Ref. (in mm) | Init. (in mm) | Reg. (in mm) | Init. (in mm) | Reg. (in mm) |
| Mean | 2.3 | 2.7 | 4.1 | 14.9 | 4.5 |
| Median | 2.3 | 2.7 | 3.9 | 15.1 | 3.6 |

- Evaluation on testing set - large displacement: Given the optimal parameter setting we study the robustness of the registration approach for a set of large displacements with respect to the ground truth alignment on the test set. We apply a set of large displacements: (7 mm, 7 mm, 7 mm), (7 mm, 7 mm, 10 mm), (7 mm, 10 mm, 7 mm), (7 mm, 10 mm, 10 mm), (10 mm, 7 mm, 7 mm), (10 mm, 7 mm, 10 mm), (10 mm, 10 mm, 7 mm) and (10 mm, 10 mm, 10 mm).

### 7.3.4 Results

The results of the above listed experiments are presented below.

- Parameter tuning on training set: Figure 7.2 shows the registration results on various block-matching parameter settings for the training set. It shows that increasing block-size improves registration performance. Registration result over the training set is best when the $\sigma_{CT}$ is 1 mm, the block-size is $15^3$ mm$^3$ and the number of points is 300. Reduced block-size of $13^3$ mm$^3$, increased number of points of 400 and $\sigma_{CT}$ of 1 mm have similar registration error as the previous parameter setting. Figure 7.3 shows the registration results on various outlier rejection parameter settings for the training set. Registration result over the training set is best when $\sigma_G$ is 0.3, $\sigma_S$ is 0.5 and $\eta$ is 0.3. For the optimized parameter settings mean and median registration errors of 3.5 mm and 3.1 mm, respectively is achieved over the training set. The number of inliers for various outlier rejection parameter settings over the training set are shown in Figure 7.4.

- Rotation: Figure 7.5 shows that the registration results on the rotated dataset. The mTRE value is consistent until 20 degrees of rotation. The registration results deteriorate for rotation angle of 25 degrees or more.

- Evaluation on testing set - small displacement: Figure 7.6 shows registration results on small displacement. The mean and median registration errors are 4.1 mm and 3.9 mm, respectively is achieved over the testing set. In Table 7.2, we show the error in the ground truth, given small displacement error before the registration and error after the registration.

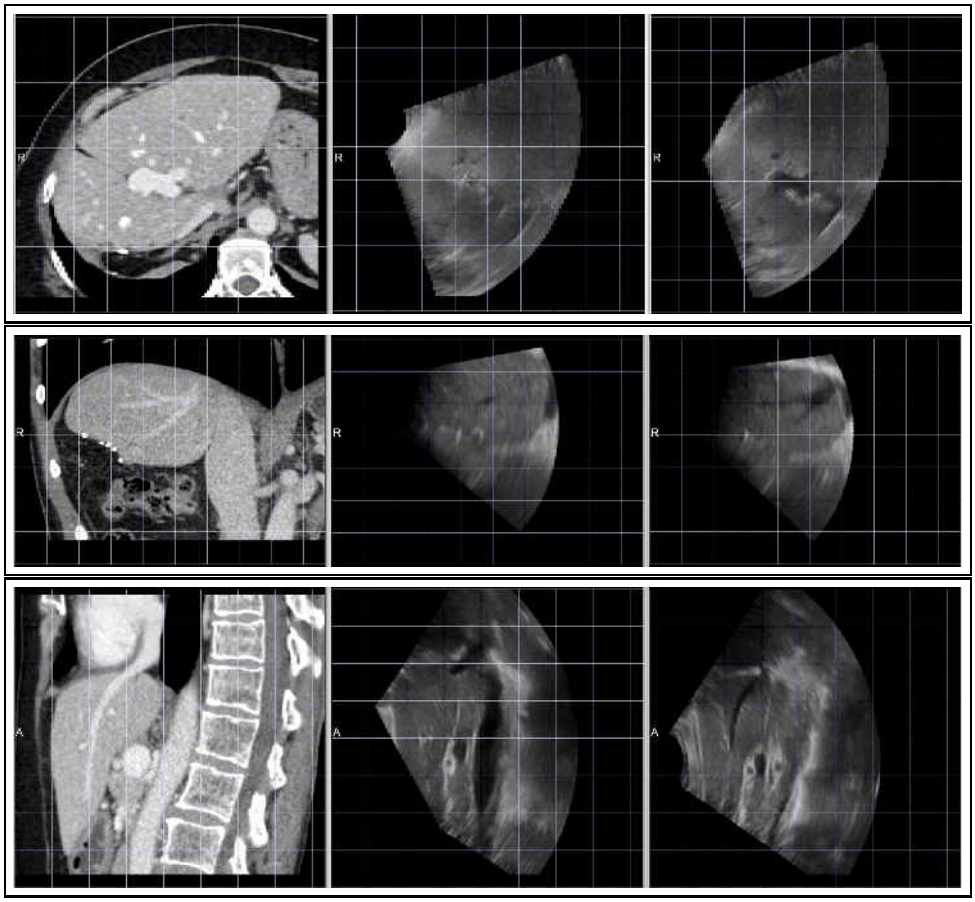


**Figure 7.7:** *Training set registration results: Left - Fixed volume, Middle - Moving volume and Right - Registered volume. Initial displacement for all the volumes is (x, y, z) = (10, 10, 10). Row 1 - mTRE = 2.6 mm, Row 2 - mTRE = 3.7 mm and Row 3 - mTRE = 3.8 mm.*

- Evaluation on testing set - large displacement: Figure 7.6 shows registration results on large displacement. The mean and median registration errors are 4.5 mm and 3.6 mm, respectively is achieved over the testing set. In Table 7.2, we show the error in the ground truth, given large displacement error before the registration and error after the registration. Considering both small and large displacement the mean and median registration errors over the testing set are 4.3 mm and 3.8 mm, respectively.

Examples of the training set registration results are shown in Figure 7.7 and examples of testing set registration results are shown in Figure 7.8 and Figure 7.9. The

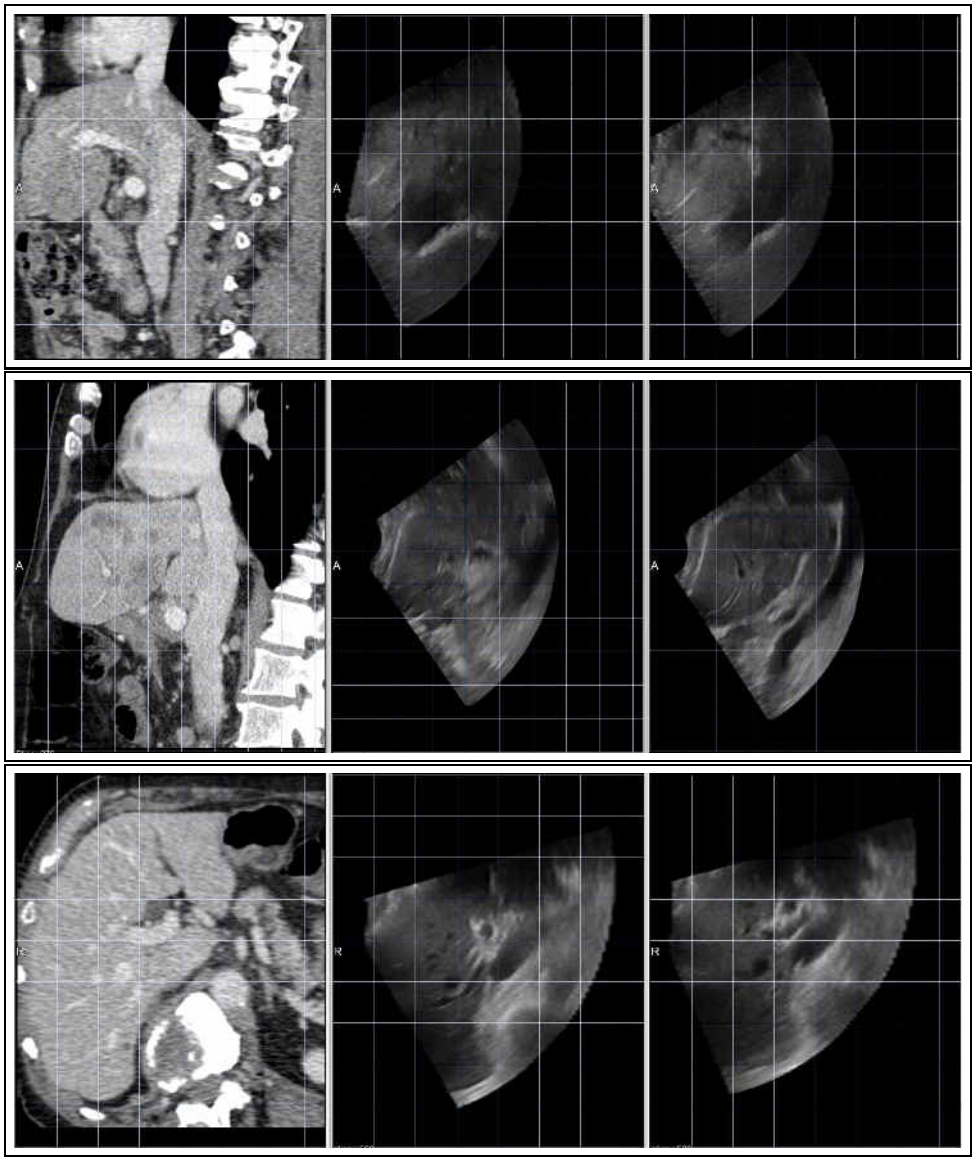


**Figure 7.8:** *Testing set registration results: Left - Fixed volume, Middle - Moving volume and Right - Registered volume. Initial displacement for all the volumes is (x, y, z) = (10, 10, 10). Row 1 - mTRE = 5.2 mm, Row 2 - mTRE = 2.1 mm and Row 3 - mTRE = 3.2 mm*

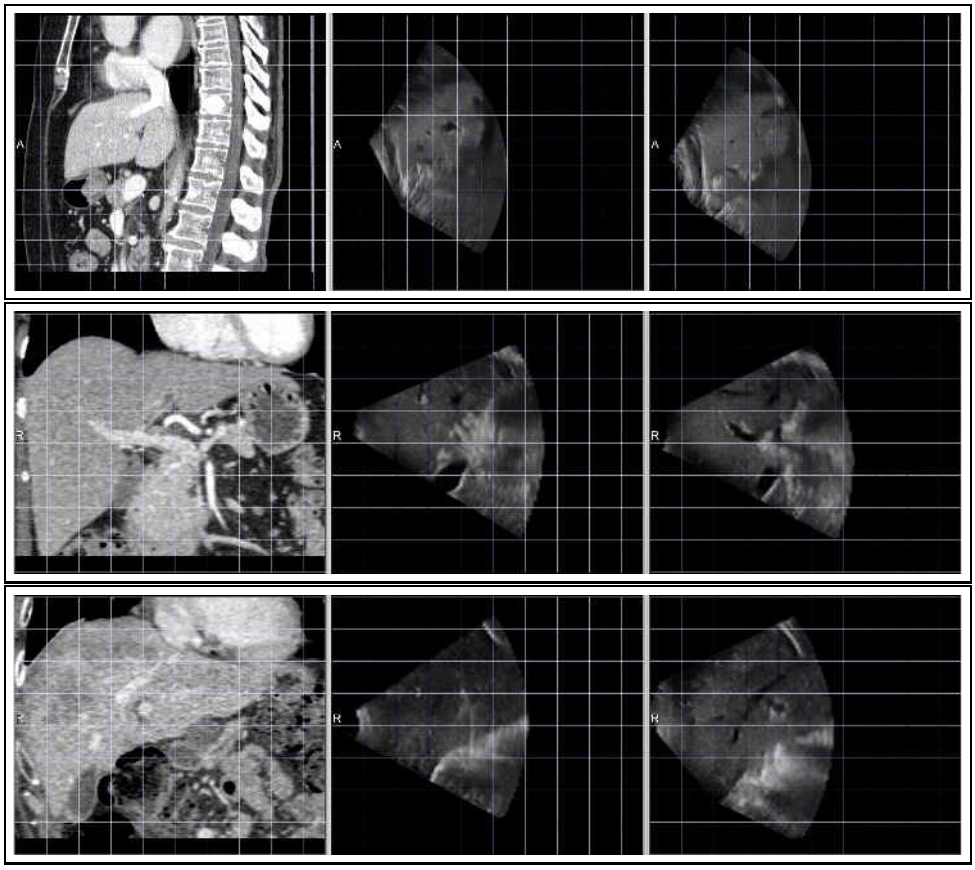


**Figure 7.9:** *Testing set registration results: Left - Fixed volume, Middle - Moving volume and Right - Registered volume. Initial displacement for all the volumes is (x, y, z) = (10, 10, 10). Row 1 - mTRE = 3.8 mm, Row 2 - mTRE = 3.9 mm and Row 3 - mTRE = 3.0 mm.*

GPU timings of the block-matching and the outliers rejection module are shown in Figure 7.10 and Figure 7.11, respectively. For 300 points, block-size of $15^3$ $mm^3$ and search range of $20^3$ $mm^3$, the block-matching and outlier rejection modules take 0.72 seconds and 0.13 seconds, respectively. Hence the GPU implementation runs the CT and US registration at 1.2 Hz.

## 7.4 Discussion and conclusion

The fusion of CT and US volumes would provide complementary information from CT during US based image guidance in interventions. In this work we presented

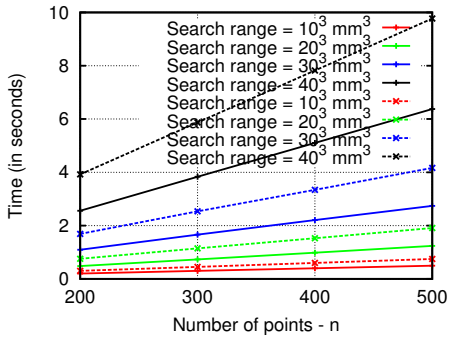



**Figure 7.10:** *Block-matching timings. Solid and dotted lines have block-size of $13^3$ $mm^3$ and $15^3$ $mm^3$, respectively.*

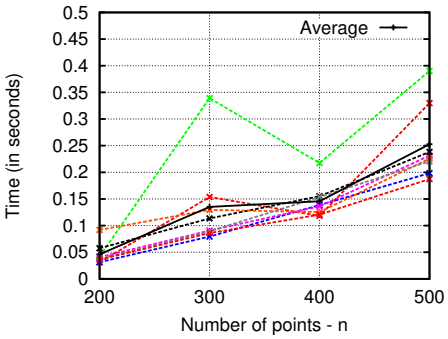



**Figure 7.11:** *Outliers rejection timings. The block-matching parameters are set as follows: block-size is $15^3$ $mm^3$ and search range is $40^3$ $mm^3$. Each of the dotted lines represents one dataset.*

a fast CT to 3D US affine registration approach, based on a previously proposed fast US-US registration framework [6], that enables multimodal registration. The approach consists of three steps namely a) point selection, b) block-matching and c) outlier rejection. The block-matching step uses MCC as similarity metric to find correspondences between the CT and the US volumes. We proposed a multimodal similarity metric, which can be efficiently implemented using the GPUs. The block-matching may yield many false correspondences; an outlier rejection step was used to remove the false correspondences. For the optimum outlier rejection parameter setting of $\sigma_G = 0.3$, $\sigma_S = 0.5$ and $\eta = 0.3$, figure 7.4 shows the number of inliers on the testing set. The inliers from the outlier rejection are used to estimate the affine transformation. The method is generic and maybe used in other multimodal registration.

Increase in the $\sigma_G$ value would reduce the number of fine edges from the CT gradient volume. Figure 7.2 shows, increase in the $\sigma_G$ value would deteriorate the accuracy of the method. Figure 7.4 shows, small values of $\sigma_G$ and $\sigma_S$ would result in insufficient number of inliers for a good registration, as shown in Figure 7.3. Similarly large values of $\sigma_G$ and $\sigma_S$ would result in high number of inliers with many false-positives, which would increase the registration error. Figure 7.3 also shows that at the optimum outlier rejection parameter setting and its vicinity the performance of the method is stable.

In the case of large initial displacement, as shown in Table 7.2 the registration error is smaller than the initial error, indicating that the method improves the alignment. However, in the case of small initial displacement, as shown in Table 7.2 the registration error is greater than the initial error, indicating that the method

marginally deteriorates the alignment. This could be due to factors like error in the ground truth annotations and the non-rigid deformation in the liver.

Wein et al. [136] performed affine registration of simulated US from CT and US to achieve a mean registration error of 8.1 mm on 25 liver dataset. They report an average computation time of 28 seconds (C++ implementation) using the $LC^2$ similarity measure. The $LC^2$ similarity measure uses the same functional modal as ours $y = ax_1 + bx_2 + c$, but unlike our correlation based method it uses a regression based approach to estimate the similarity. A GPU (NVIDIA GeForce GTX TITAN with 2688 cores) implementation of the $LC^2$ similarity measure applied to MR and US registration by Fuerst et al. [36] report an average time of 2.32 seconds for a capture range of 15 mm. The average run-time of our registration method on a GPU (NVIDIA GeForce GTX 780 Ti with 2880 cores) is 0.85 seconds for a search range of $20^3$ $mm^3$. Lange et al. [67] performed non-rigid registration on clinical liver data sets of three different patients and achieve a mean registration error of 3.7 mm. Nam et al. [90] performed affine registration with the joint use of vessels and the liver surface to achieve a fiducial registration error of 3 mm over 20 clinical dataset. Rucker et al. [115] using sparse surface data performed non-rigid registration of CT and US and achieved a mean registration error of 3.3 mm over 5 phantom deformation dataset. The method we presented has a mean registration error of 4.3 mm, which is comparable to other existing registration approaches.

US and CT fusion guided RFA has been shown to be a safe and effective treatment for liver cancer [137]. In RFA of the liver, interventional radiologist ablates an additional 1 cm thick tumor-free margin around the tumor to avoid reoccurrence of the tumor [86]. The registration error of the method presented is within this ballpark. Compared to existing, tracker-based approaches, an image-based fusion approach as presented would logistically be simpler, and would permit automatic updating of the fusion in case of patient motion (such as respiration). For a fast continuous registration of 4D US with CT, the registration to reference by tracking (RTRT) [5] strategy could be used. The RTRT strategy would help in further reducing the number of points and the search range used in the registration, and would help in improving the run-time of the method. However, real-time 3D US on our system was of inferior image quality (especially the resolution) compared to a static 3D US (3D sweep, obtained in $< 1$ sec using the 2D matrix array transducer). Newer systems may have superior image quality, and permit a live update of the fused images.

Obtaining accurate CT-US ground truth annotation was a challenging task, especially because the initial orientation of the US data with respect to the CT was not accurately known, as we did not use any external tracking device. To address this we performed annotations in two steps - in the first step we perform rough alignment using an approximate of a few landmark annotations and in the next step we refine the annotation to obtain accurate ground truth annotations.

The current approach has been evaluated on data obtained from a single cen-

ter and a single US device. Validation on a larger test bed should be performed before applying it to clinic. Other avenues of future work include incorporating a respiratory motion model and extension to fast non-rigid registration approaches.

In conclusion, we proposed and evaluated a fast CT to US affine registration which is based on a novel similarity metric. A mTRE of 3.5 is achieved over 8 CT and 3D US training pairs acquired from 5 patients and 4.3 mm is achieved over 21 CT and 3D US testing pairs acquired from 12 patients. For a search range of $20^3$ $mm^3$, the method performs CT and US registration at 1.2 Hz. The proposed fast CT and US fusion method can be potentially used to improve image guidance in liver interventions.

# Chapter 8

## Summary and Future Perspectives

## 8.1 Summary

In this thesis we proposed and evaluated algorithms for image guidance in liver interventions. We presented rotationally invariant features computation and image based motion compensation of the real-time 3D ultrasound (US) images of the liver. In addition a method to fuse computed tomography (CT) and US is proposed and evaluated. In this chapter we summarize the main findings of the work presented in this thesis and discuss future research directions.

Real-time 3D US can be potentially used for image guidance in minimally invasive liver interventions. Motion caused by patient breathing shifts the liver and makes it hard to continuously visualize a specific target area. Respiratory motion also hampers continuous accurate alignment with pre-operative data. In Chapter 2 we developed a fast affine registration framework to compensate liver motion or displacement due to breathing. The affine registration of two consecutive US volumes was performed using a fast combined block-matching and outlier rejection approach. For a set of evenly distributed points in one volume, their correspondences in the other volume were determined via block-matching. Next, a robust outlier rejection method was used to detect and reject false block-matching matches. The inliers were then used to determine the affine transformation. The registration approach was evaluated on 13 4D US sequences acquired from 8 subjects. For 91 pairs of 3D US volumes selected from these sequences, a mean registration error of 1.8 mm was achieved. We implemented the method in GPU harware. The GPU implementation runs the 3D US registration at 8 Hz.

In Chapter 3 we presented an extension of the work of Chapter 2 to 4D US image sequences. For 4D registration, we performed registration of consecutive US volumes over the time series. Transformation between any two frames was estimated by taking the product of all the intermediate transforms. To avoid accumulation of error over the series of transformations, a long range feedback mechanism is proposed. A mean total registration error of 1 mm was achieved across six 4D US sequences of human liver. The GPU implementation runs the registration at 10 Hz. The limitations of this approach are that it requires large search range and cannot handle all rotations.

In Chapter 4 work we improved on the method in Chapter 3, which was earlier tested on a limited dataset and for shorter time sequences. Here we described a fast and robust 4D registration method for on the fly stabilization of US volumes for improving image guidance for transjugular intrahepatic portosystemic shunt (TIPS) interventions. This enables continuous visualization of relevant anatomical planes determined in a planning stage at the start of the intervention. It requires mapping the planning information to the interventional images, which was achieved in two steps. In the first step tracking was performed across the streaming US volumes. An approximate transformation between the reference image and the incoming image was estimated by composing the intermediate transformations obtained from the

tracking. In the second step a subsequent registration was performed between the reference image and the approximately transformed incoming image to account for the accumulation of error. The two step approach helped in reducing the search range and was robust under rotation. We evaluated the method over thirteen 4D US sequences acquired from eight subjects. A graphics processing unit implementation runs the 4D tracking at 9 Hz. A mean registration error of 1.7 mm was achieved. Tests over long US sequences showed that the registration error did not increase over time. We additionally presented an approach to initialize and verify the registration. The verification method helped to detect and validate the registration outcome, and was used to automatically initialize the 2-step registration described above. It indicated whether the reference image (containing planning information) was aligned well with the (interventional) 4D US sequence.

In Chapter 5 we presented a method for tracking of anatomical landmarks in the liver. This 4D US tracking method was based on global and local rigid registration schemes. Our method was based on the previous work described in Banerjee et al. [5] and [6], which are in turn based on the work described in the Chapters 2 and 4. We evaluated our method on the dataset that was presented in the MICCAI 2015 Challenge on Liver US Tracking (CLUST 2015). On the test set a mean tracking error of $1.80 \pm 1.64$ mm was achieved.

Rotationally invariant feature may help in matching images acquired from different orientations, which is relevant in analysis of medical images acquired from different scanners. Local binary patterns [LBP] is a popular texture description method used mainly in 2D image analysis tasks. The 3D LBP-based rotationally invariant region description described in Chapter 6 extracts the texture features from volumetric data which are view/pose invariant. The 3D rotationally invariant texture feature generates a set of variables per voxel, where the number of variables are equivalent to the number of spherical harmonic bands plus the kurtosis. A set of histograms from each of the variables was used to describe a region. Similarity/dissimilarity between two regions can be estimated using histogram based distance measures like the Bhattacharyya measure (or coefficient) and the Chi-square measure. We applied the method for vessel detection on vessel phantom data and a clinical dataset to show its discriminative ability in visual classification tasks with encouraging results.

In Chapter 7 we presented an approach to perform registration of CT and US volumes for image guided liver intervention applications. The registration method was based on our previous block-matching approaches, and includes a novel multimodal similarity metric that fits our framework. The intensity and the gradient of the CT images along with the US volumes are the input images to the block-matching. The multimodal block-matching approach used multiple correlation coefficient similarity metric to find correlation between the CT and the US volumes. The geometric consistency criteria for outlier rejection were combined with smoothness criteria to further adapt the outlier rejection to the multi-modal registration

cases. The block-matching results after outlier rejection were used to determine the affine transformation between the CT and the US volumes. The training dataset consisted of 8 CT and US pairs from 5 patients and the testing dataset consisted of 21 CT and US pairs acquired from 12 patients. A mean registration error of 3.5 and 4.3 mm was achieved over the training and testing dataset, respectively.

## 8.2 Future Perspectives

In this work we developed a fast (real-time) registration approach for liver interventions. We addressed the issue of stabilizing the liver motion due to breathing. Liver tissue is not rigid, and the liver may deform during respiratory motion. However, in our work we have considered an affine transformation model for the following reasons:

- We focused on a fast implementation approach using graphics processing unit hardware which facilitates such real-time applications. In an affine registration model, interpolation or resampling operations are computationally less intensive compared to a non-rigid registration.
- The optimization schemes associated with estimating non-rigid deformation fields are often computationally intensive.
- The field of view in US is not large and usually a large organ like the liver is only partly visible. Because of this, an affine transform may be a sufficiently good approximation.
- An affine transformation model implemented in real-time is sufficient for the task of stabilizing the planes in an US volume required for image guidance in minimally invasive procedure.

Having mentioned the above points for the choice of transformation, we believe that there is merit in exploring and extending the method to a real-time non-rigid registration framework. This would require further research and development, and it may further improve the accuracy and enable extension to other interventions. Further, as our 4D US registration work mentioned in Chapter 4 was primarily developed for TIPS interventions, these methods may be relevant for other image guided interventions as well, e.g. motion tracking for radiotherapy.

In Chapter 7, we present US and CT registration method. It would be interesting to study the relationship between the CT-US registration and the phase of the contrast enhanced CT acquisition. Another interesting study would be performing the multimodal registration approach in a 4D registration framework which would potentially enable multimodal image guidance. This will be relevant for percutaneous treatments like tumors in the liver and other organs.

Incorporating a motion model for better estimating the position of the liver would be useful. The work could also be extended to a non-rigid registration technique. The CT-US registration method can be potentially used in other multi-modal registrations like magnetic resonance imaging and US registration. From a clinical perspective the technological advancements should be augmented with ergonomically well-designed user interface. The orthogonal planes to the plane parallel to the 4D US acquisition (and the plane containing the long axis of the transducer), typically suffers from low resolution. Hence, improving the 4D US image quality is required, as oblique planes are required for image guidance during interventions.

**Figure 8.1:** *The real-time 3D US based navigation system for the TIPS procedure at Erasmus MC.*

Concluding, in this thesis, we presented methods and evaluated their accuracies that demonstrate the use of real-time 3D US and its fusion with CT in potentially improving image guidance in minimally invasive US guided liver interventions. The fast US based registration method has been implemented in a 4D US guidance system for TIPS, see Figure8.1. Further extensions to multi-modal guidance and non-rigid registration methods are relevant directions for future research. From a clinical perspective, a larger clinical validation is required before using these systems in the clinic.

# References

[1] AJ, E., PW, G., MO, H., F, Y., D, P.; Impact of minimally invasive surgery on medical spending and employee absenteeism. JAMA Surgery 2013;148(7):641–647.

[2] Albarelli, A., Bulò, S.R., Torsello, A., Pelillo, M.; Matching as a non-cooperative game. In: ICCV. 2009. p. 1319–1326.

[3] Albarelli, A., Rodolà, E., Torsello, A.; Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective. International Journal of Computer Vision 2012;97(1):36–53.

[4] Arun, K.S., Huang, T.S., Blostein, S.D.; Least-squares fitting of two 3-D point sets. IEEE Trans Pattern Anal Mach Intell 1987;9(5):698–700.

[5] Banerjee, J., Klink, C., Niessen, W.J., Moelker, A., van Walsum, T.; 4d ultrasound tracking of liver and its verification for TIPS guidance. IEEE Trans Med Imaging 2016;35(1):52–62.

[6] Banerjee, J., Klink, C., Peters, E.D., Niessen, W.J., Moelker, A., van Walsum, T.; Fast and robust 3D ultrasound registration - block and game theoretic matching. Medical Image Analysis 2015;20(1):173–183.

[7] Bell, M.A.L., Byram, B.C., Harris, E.J., Evans, P.M., Bamber, J.C.; In vivo liver tracking with a high volume rate 4D ultrasound scanner and a 2D matrix array probe. Physics in Medicine and Biology 2012;57(5):1359.

[8] Bhatia, K.K., Hajnal, J.V., Hammers, A., Rueckert, D.; Similarity metrics for groupwise non-rigid registration. In: MICCAI (2). 2007. p. 544–552.

[9] Blachier, M., Leleu, H., Peck-Radosavljevic, M., Valla, D.C., Roudot-Thoraval, F.; The burden of liver disease in europe: A review of available epidemiological data. Journal of Hepatology 2013;58(3):593 – 608.

[10] Bomze, I.M., Locatelli, M., Tardella, F.; New and old bounds for standard quadratic optimization: dominance, equivalence and incomparability. Math Program 2008;115(1):31–64.

[11] Brattain, L.J., Howe, R.D.; Real-time 4D ultrasound mosaicing and visualization. In: MICCAI (1). 2011. p. 105–112.

[12] Brox, T., Bruhn, A., Papenberg, N., Weickert, J.; High accuracy optical flow estimation based on a theory for warping. In: Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV. 2004. p. 25–36.

[13] Buades, A., Coll, B., Morel, J.; A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA. 2005. p. 60–65.

[14] Buades, A., Coll, B., Morel, J.; Non-local means denoising. IPOL Journal 2011;1.

[15] Buades, A., Coll, B., Morel, J.; Self-similarity-based image denoising. Commun ACM 2011;54(5):109–117.

[16] Cen, F., Jiang, Y., Zhang, Z., Tsui, H.T., Lau, T.K., Xie, H.; Robust registration of 3-D ultrasound images based on gabor filter and mean-shift method. In: ECCV Workshops CVAMIA and MMBIA. 2004. p. 304–316.

[17] Cha, S.H.; Comprehensive survey on distance/similarity measures between probability density functions. International Journal of Mathematical Models and Methods in Applied Sciences 2007;1(4):300–307.

[18] Chen, C.H., Pau, L.F., Wang, P.S.P., editors; Handbook of Pattern Recognition and Computer Vision. 2nd ed. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2000.

[19] Chen, J., Shan, S., He, C., Zhao, G., Pietikäinen, M., Chen, X., Gao, W.; Wld: A robust local image descriptor. IEEE Trans Pattern Anal Mach Intell 2010;32(9):1705–1720.

[20] Cifor, A., Risser, L., Chung, D., Anderson, E., Schnabel, J.; Hybrid feature-based diffeomorphic registration for tumor tracking in 2-D liver ultrasound images. Medical Imaging, IEEE Transactions on 2013;32(9):1647–1656.

[21] Cohen, B., Dinstein, I.; New maximum likelihood motion estimation schemes for noisy ultrasound images. Pattern Recognition 2002;35(2):455–463.

[22] Commowick, O., Wiest-Daesslé, N., Prima, S.; Block-matching strategies for rigid registration of multimodal medical images. In: ISBI. 2012. p. 700–703.

[23] Conte, D., Foggia, P., Sansone, C., Vento, M.; Thirty years of graph matching in pattern recognition. IJPRAI 2004;18(3):265–298.

[24] Council, N.R.; Mathematics and Physics of Emerging Biomedical Imaging. The National Academies Press, 1996.

[25] Crocetti, L., Lencioni, R., Debeni, S., See, T.C., Pina, C.D., Bartolozzi, C.; Targeting liver lesions for radiofrequency ablation: an experimental feasibility study using a CT-US fusion imaging system. Invest Radiol 2008;43(1):33–39.

[26] Cuijpers, C.F., Klink, C., Stappers, P.J., Freudenthal, A.; Comparing image guidance systems to improve complex navigation in medicine. In: 12th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, IFAC HMS 2013, Las Vegas, Nevada, USA, August 11-15, 2013. 2013. p. 226–231.

[27] Donoghue, C.R., Rao, A., Bull, A.M.J., Rueckert, D.; Robust global registration through geodesic paths on an empirical manifold with knee MRI from the osteoarthritis initiative (oai). In: WBIR. 2012. p. 1–10.

[28] Edwards, A.L.; An introduction to linear regression and correlation. W.H.Freeman &

Co Ltd, 1976.

[29] Fehr, J.; Local rotation invariant patch descriptors for 3D vector fields. In: ICPR. 2010. p. 1381–1384.

[30] Fehr, J., Burkhardt, H.; 3D rotation invariant local binary patterns. In: ICPR. 2008. p. 1–4.

[31] Fischler, M.A., Bolles, R.C.; Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 1981;24(6):381–395.

[32] Foroughi, P., Abolmaesumi, P., Hashtrudi-Zaad, K.; Intra-subject elastic registration of 3D ultrasound images. Medical Image Analysis 2006;10(5):713–725.

[33] Forsyth, D.A., Ponce, J.; Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference, 2002.

[34] Francois, R., Fablet, R., Barillot, C.; Robust statistical registration of 3D ultrasound images using texture information. In: ICIP (1). 2003. p. 581–584.

[35] Friedman, L.S., Keeffe, E.B., editors; The Handbook of Liver Disease. 3rd ed. Elsevier Health Sciences, 2011.

[36] Fuerst, B., Wein, W., Müller, M., Navab, N.; Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D $LC^2$ metric. Medical Image Analysis 2014;18(8):1312–1319.

[37] Fuller, R.B.. U.s. patent no. 2,682,235 (building construction). 1954.

[38] G, G., KR, G., A, S., et al, ; Effect of minimally invasive surgery on the risk for surgical site infections: Results from the national surgical quality improvement program (nsqip) database. JAMA Surgery 2014;149(10):1039–1044.

[39] Gluckman, J.; Kurtosis and the phase structure of images. In: in 3rd International Workshop on Statistical and Computational Theories of Vision (in conjunction with ICCV '03). 2003. .

[40] Gonzalez, R.C., Woods, R.E.; Digital Image Processing. 2nd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.

[41] Grau, V., Becher, H., Noble, J.A.; Registration of multiview real-time 3-D echocardiographic sequences. IEEE Trans Med Imaging 2007;26(9):1154–1165.

[42] Green, R.; Spherical harmonic lighting: The gritty details. Archives of the Game Developers Conference 2003;.

[43] Guo, Z., Zhang, L., Zhang, D.; A completed modeling of local binary pattern operator for texture classification. IEEE Transactions on Image Processing 2010;19(6):1657–1663.

[44] Guo, Z., Zhang, L., Zhang, D.; Rotation invariant texture classification using lbp variance (lbpv) with global matching. Pattern Recognition 2010;43(3):706–719.

[45] Haaga, J.R.; Interventional CT: 30 years' experience. European Radiology Supplements ;15(4):d116–d120.

[46] Haber, E., Modersitzki, J.; Intensity gradient based registration and fusion of multi-modal images. In: MICCAI (2). 2006. p. 726–733.

[47] Harris, E.J., Miller, N.R., Bamber, J.C., Symonds-Tayler, J.R.N., Evans, P.M.; Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound. Physics in Medicine and Biology 2010;55(12):3363.

[48] Heimann, T., van Ginneken, B., Styner, M., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G.K., Bischof, H., Bornik, A., Cashman, P., Chi, Y., Cordova, A., Dawant, B.M., Fidrich, M., Furst, J.D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmüller, D., Kitney, R., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H., Németh, G., Raicu, D.S., Rau, A., van Rikxoort, E.M., Rousson, M., Ruskó, L., Saddi, K.A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J.M., Wimmer, A., Wolf, I.; Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging 2009;28(8):1251–1265.

[49] Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T.N., Gleeson, F., Brady, M., Schnabel, J.A.; MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. Medical Image Analysis 2012;16(7):1423–1435.

[50] Heinrich, M.P., Jenkinson, M., Papiez, B.W., Brady, M., Schnabel, J.A.; Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013 - 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I. 2013. p. 187–194.

[51] Hjertaas, J.J., FossÃě, H., Dybdahl, G.L., GrÃijner, R., Lunde, P., Matre, K.; Accuracy of real-time single- and multi-beat 3-d speckle tracking echocardiography in vitro. Ultrasound in Medicine & Biology 2013;39(6):1006 – 1014.

[52] III, W.M.W., Viola, P.A., Atsumi, H., Nakajima, S., Kikinis, R.; Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1996;1(1):35–51.

[53] Jun, B., Kim, T., Kim, D.; A compact local binary pattern using maximization of mutual information for face analysis. Pattern Recognition 2011;44(3):532–543.

[54] Kaar, M., Figl, M., Hoffmann, R., Birkfellner, W., Hummel, J., Stock, M., Georg, D., Goldner, G.; Automatic patient alignment system using 3D ultrasound. Medical Physics 2013;40(4).

[55] Kang, T.W., Rhim, H.; Recent Advances in Tumor Ablation for Hepatocellular Carcinoma. Liver Cancer 2015;4(3):176–187.

[56] Kang, U., Hebert, M., Park, S.; Fast and scalable approximate spectral graph matching for correspondence problems. Inf Sci 2013;220:306–318.

[57] Karnik, V., Fenster, A., Bax, J., Cool, D., Gardi, L., Gyacskov, I., Romagnoli, C., Ward, A.; Assessment of image registration accuracy in three-dimensional transrectal ultrasound-guided prostate biopsy. Med Phys 2010;37(2):802–13.

[58] Kazhdan, M.M., Funkhouser, T.A., Rusinkiewicz, S.; Rotation invariant spherical

harmonic representation of 3D shape descriptors. In: Symposium on Geometry Processing. 2003. p. 156–165.

[59] Kim, K.R., Thomas, S.; Complications of image-guided thermal ablation of liver and kidney neoplasms. Semin Intervent Radiol 2014;31(2):138–148.

[60] Kingham, T.P., Scherer, M.A., Neese, B.W., Clements, L.W., Stefansic, J.D., Jarnagin, W.R.; Image-guided liver surgery: intraoperative projection of computed tomography images utilizing tracked ultrasound. HPB (Oxford) 2012;14(9):594–603.

[61] Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.; elastix: A toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 2010;29(1):196–205.

[62] Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.. elastix: parameter file database. http://elastix.isi.uu.nl/wiki.php; 2010. Accessed: 2014-06-22.

[63] Krajina, A., Hulek, P., Fejfar, T., Valek, V.; Quality improvement guidelines for Transjugular Intrahepatic Portosystemic Shunt (TIPS). Cardiovasc Intervent Radiol 2012;35(6):1295–1300.

[64] Kuklisova-Murgasova, M., Cifor, A., Napolitano, R., Papageorghiou, A., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Noble, J.A., Schnabel, J.A.; Registration of 3d fetal neurosonography and MRI. Medical Image Analysis 2013;17(8):1137–1150.

[65] Kutter, O., Wein, W., Navab, N.; Multi-modal registration based ultrasound mosaicing. In: MICCAI (1). 2009. p. 763–770.

[66] LaBerge, J.M., Ring, E.J., Lake, J.R., Ferrell, L.D., Doherty, M.M., Gordon, R.L., Roberts, J.P., Peltzer, M.Y., Ascher, N.L.; Transjugular intrahepatic portosystemic shunts: preliminary results in 25 patients. J Vasc Surg 1992;16(2):258–267.

[67] Lange, T., Papenberg, N., Heldmann, S., Modersitzki, J., Fischer, B., Lamecker, H., Schlag, P.; 3D ultrasound-ct registration of the liver using combined landmark-intensity information. International Journal of Computer Assisted Radiology and Surgery 2009;4:79–88. 10.1007/s11548-008-0270-1.

[68] Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., Fichtinger, G.; PLUS: open-source toolkit for ultrasound-guided intervention systems. IEEE Trans Biomed Engineering 2014;61(10):2527–2537.

[69] Learned-Miller, E.G.; Data driven image models through continuous joint alignment. IEEE Trans Pattern Anal Mach Intell 2006;28(2):236–250.

[70] Lediju, M., Byram, B., Harris, E., Evans, P., Bamber, J.; 3D liver tracking using a matrix array: Implications for ultrasonic guidance of IMRT. In: Ultrasonics Symposium (IUS), 2010 IEEE. 2010. p. 1628–1631.

[71] Lediju Bell, M.A., Sen, H.T., Iordachita, I., Kazanzides, P., Wong, J.; In vivo reproducibility of robotic probe placement for a novel ultrasound-guided radiation therapy system. Journal of Medical Imaging 2014;1(2):025001.

[72] Lee, M.W.; Fusion imaging of real-time ultrasonography with CT or MRI for hepatic intervention. Ultrasonography 2014;33(4):227–239.

[73] Lee, S.L., Lerotic, M., Vitiello, V., Giannarou, S., Kwok, K.W., Visentini-Scarzanella, M., Yang, G.Z.; From medical images to minimally invasive intervention: Computer assistance for robotic surgery. Computerized Medical Imaging and Graphics 2010;34(1):33 – 45. Image-Guided Surgical Planning and Therapy.

[74] Leordeanu, M., Hebert, M.; A spectral technique for correspondence problems using pairwise constraints. In: ICCV. 2005. p. 1482–1489.

[75] Lewin, J.S.; Interventional mr imaging: Concepts, systems, and applications in neuroradiology. American Journal of Neuroradiology 1999;20(5):735–748.

[76] Liang, T., Yung, L.S., Yu, W.; On feature motion decorrelation in ultrasound speckle tracking. IEEE Trans Med Imaging 2013;32(2):435–448.

[77] Liao, S., Law, M.W.K., Chung, A.C.S.; Dominant local binary patterns for texture classification. IEEE Transactions on Image Processing 2009;18(5):1107–1118.

[78] Liu, L., Zhao, L., Long, Y., Kuang, G., Fieguth, P.W.; Extended local binary patterns for texture classification. Image Vision Comput 2012;30(2):86–99.

[79] Liu, X., Annangi, P., Gupta, M., Yu, B., Padfield, D., Banerjee, J., Krishnan, K.; Learning-based scan plane identification from fetal head ultrasound images. In: Proc. SPIE. volume 8320; 2012. p. 83200A–83200A–6.

[80] Lowe, D.G.; Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 2004;60(2):91–110.

[81] Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S., Alvarado, M., Anderson, H., Anderson, L., Andrews, K., Atkinson, C., Baddour, L., Barker-Collo, S., Bartels, D., Bell, M., Benjamin, E., Bennett, D., Bhalla, K., Bikbov, B., Bin, A.A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S., Coffeng, L., Colan, S., Colquhoun, S., Colson, K., Condon, J., Connor, M., Cooper, L., Corriere, M., Cortinovis, M., de, V.K., Couser, W., Cowie, B., Criqui, M., Cross, M., Dabhadkar, K., Dahodwala, N., De, L.D., Degenhardt, L., Delossantos, A., Denenberg, J., Des, J.D., Dharmaratne, S., Dorsey, E., Driscoll, T., Duber, H., Ebel, B., Erwin, P., Espindola, P., Ezzati, M., Feigin, V., Fl, ; Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. LANCET 2012;380:2095–2128.

[82] De Luca, V., Tschannen, M., Székely, G., Tanner, C.; A learning-based approach for fast and robust vessel tracking in long ultrasound sequences. In: MICCAI (1). 2013. p. 518–525.

[83] Luca, V.D., Benz, T., Kondo, S., König, L., Lübke, D., Rothlübbers, S., Somphone, O., Allaire, S., Bell, M.A.L., Chung, D.Y.F., Cifor, A., Grozea, C., Günther, M., Jenne, J., Kipshagen, T., Kowarschik, M., Navab, N., Rühaak, J., Schwaab, J., Tanner, C.; The 2014 liver ultrasound tracking benchmark. Physics in Medicine and Biology 2015;60(14):5571.

[84] Markelj, P., Tomazevic, D., Likar, B., Pernus, F.; A review of 3D/2D registration methods for image-guided interventions. Medical Image Analysis 2012;16(3):642–661.

[85] Mattes, D., Haynor, D.R., Vesselle, H., Lewellyn, T.K., Eubank, W.; Nonrigid multimodality image registration 2001;:1609–1620.

[86] McDermott, S., Gervais, D.A.; Radiofrequency ablation of liver tumors. Semin Intervent Radiol 2013;30(1):49–55. 30049[PII].

[87] Metz, C., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J.; Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach. Medical Image Analysis 2011;15(2):238–249.

[88] Mukherjee, R., Sprouse, C., Pinheiro, A., Abraham, T., Burlina, P.; Computing myocardial motion in 4-dimensional echocardiography. Ultrasound in Medicine & Biology 2012;38(7):1284 – 1297.

[89] Nagpal, S., Abolmaesumi, P., Rasoulian, A., Hacihaliloglu, I., Ungi, T., Osborn, J., Lessoway, V.A., Rudan, J., Jaeger, M., Rohling, R.N., Borschneck, D.P., Mousavi, P.; A multi-vertebrae CT to US registration of the lumbar spine in clinical data. Int J Computer Assisted Radiology and Surgery 2015;10(9):1371–1381.

[90] Nam, W.H., Kang, D.G., Lee, D., Lee, J.Y., Ra, J.B.; Automatic registration between 3D intra-operative ultrasound and pre-operative ct images of the liver based on robust edge matching. Physics in Medicine and Biology 2012;57(1):69.

[91] Nanni, L., Lumini, A., Brahnam, S.; Local binary patterns variants as texture descriptors for medical image analysis. Artif Intell Med 2010;49:117–125.

[92] Neemuchwala, H., Hero, A.O., Carson, P.L.; Image matching using alpha-entropy measures and entropic graphs. Signal Processing 2005;85(2):277–296.

[93] Ni, D., Chui, Y.P., Qu, Y., Yang, X.S., Qin, J., Wong, T.T., Ho, S.S.H., Heng, P.A.; Reconstruction of volumetric ultrasound panorama based on improved 3D sift. Comp Med Imag and Graph 2009;33(7):559–566.

[94] Ni, D., Yang, X., Chen, X., Chin, C.T., Chen, S., Heng, P.A., Li, S., Qin, J., Wang, T.; Standard plane localization in ultrasound by radial component model and selective search. Ultrasound in Medicine & Biology 2014;40(11):2728 – 2742.

[95] Ojala, T., Pietikäinen, M., Harwood, D.; A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 1996;29(1):51–59.

[96] Ojala, T., Pietikäinen, M., Mäenpää, T.; Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24(7):971–987.

[97] Ourselin, S., Roche, A., Prima, S., Ayache, N.; Block matching: A general framework to improve robustness of rigid registration of medical images. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2000, Third International Conference, Pittsburgh, Pennsylvania, USA, October 11-14, 2000, Proceedings. 2000. p. 557–566.

[98] Øye, O.K., Wein, W., Ulvang, D.M., Matre, K., Viola, I.; Real time image-based tracking of 4D ultrasound data. In: MICCAI (1). 2012. p. 447–454.

[99] Patel, I., Pirasteh, A., Prologo, J.; Liver intervention: Angiography. In: Hamm, B., Ros, P., editors. Abdominal Imaging. Springer Berlin Heidelberg; 2013. p. 969–982.

[100] Pavan, M., Pelillo, M.; Dominant sets and pairwise clustering. IEEE Trans Pattern Anal Mach Intell 2007;29(1):167–172.

[101] Penney, G.P., Blackall, J.M., Hamady, M.S., Sabharwal, T., Adam, A., Hawkes, D.J.; Registration of freehand 3d ultrasound and magnetic resonance liver images. Medical Image Analysis 2004;8(1):81–91.

[102] Peters, T.M.; Image-guidance for surgical procedures. Physics in Medicine and Biology 2006;51(14):R505.

[103] Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.; Mutual information based registration of medical images: A survey. IEEE Trans Med Imaging 2003;22(8):986–1004.

[104] Pons, F., Varela, M., Llovet, J.M.; Staging systems in hepatocellular carcinoma. HPB (Oxford) 2005;7(1):35–41.

[105] Preiswerk, F., Luca, V.D., Arnold, P., Celicanin, Z., Petrusca, L., Tanner, C., Bieri, O., Salomir, R., Cattin, P.C.; Model-guided respiratory organ motion prediction of the liver from 2D ultrasound. Medical Image Analysis 2014;18(5):740–751.

[106] Rajpoot, K., Noble, J.A., Grau, V., Szmigielski, C., Becher, H.; Multiview RT3D echocardiography image fusion. In: FIMH. 2009. p. 134–143.

[107] Ripamonti, R., Ferral, H., Alonzo, M., Patel, N.H.; Transjugular intrahepatic portosystemic shunt-related complications and practical solutions. Semin Intervent Radiol 2006;23(2):165–176.

[108] Rivaz, H., Chen, S.J., Collins, D.L.; Automatic deformable mr-ultrasound registration for image-guided neurosurgery. IEEE Trans Med Imaging 2015;34(2):366–380.

[109] Roche, A., Malandain, G., Pennec, X., Ayache, N.; The correlation ratio as a new similarity measure for multimodal image registration. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI'98, First International Conference, Cambridge, MA, USA, October 11-13, 1998, Proceedings. 1998. p. 1115–1124.

[110] Roche, A., Pennec, X., Malandain, G., Ayache, N.; Rigid registration of 3d ultrasound with MR images: A new approach combining intensity and gradient information. IEEE Trans Med Imaging 2001;20(10):1038–1049.

[111] Rodolà, E., Bronstein, A.M., Albarelli, A., Bergamasco, F., Torsello, A.; A game-theoretic approach to deformable shape matching. In: CVPR. 2012. p. 182–189.

[112] Rosch, J., Hanafee, W.N., Snow, H.; Transjugular portal venography and radiologic portacaval shunt: an experimental study. Radiology 1969;92(5):1112–1114.

[113] Rössle, M., Richter, G.M., Nöldge, G., Palmaz, J.C., Wenz, W., Gerok, W.; New non-operative treatment for variceal haemorrhage. The Lancet 1989;334(8655):153.

[114] Rubin, J.M., Feng, M., Hadley, S.W., Fowlkes, J.B., Hamilton, J.D.; Potential use of ultrasound speckle tracking for motion management during radiotherapy: preliminary report. J Ultrasound Med 2012;31(3):469–81.

[115] Rucker, D.C., Wu, Y., Clements, L.W., Ondrake, J.E., Pheiffer, T.S., Simpson, A.L., Jarnagin, W.R., Miga, M.I.; A mechanics-based nonrigid registration method for liver surgery using sparse intraoperative data. IEEE Trans Med Imaging 2014;33(1):147–158.

[116] Saad, W.E.; The history and future of transjugular intrahepatic portosystemic shunt: Food for thought. Semin Intervent Radiol 2014;31(3):258–261. 25177087[pmid].

[117] Schneider, R.J., Perrin, D.P., Vasilyev, N.V., Marx, G.R., del Nido, P.J., Howe, R.D.; Real-time image-based rigid registration of three-dimensional ultrasound. Medical Image Analysis 2012;16(2):402–414.

[118] Shekhar, R., Zagrodsky, V., Garcia, M.J., Thomas, J.D.; Registration of real-time 3-D ultrasound images of the heart for novel 3-D stress echocardiography. IEEE Trans Med Imaging 2004;23(9):1141–1149.

[119] Shi, W., Jantsch, M., Aljabar, P., Pizarro, L., Bai, W., Wang, H., O'Regan, D.P., Zhuang, X., Rueckert, D.; Temporal sparse free-form deformations. Medical Image Analysis 2013;17(7):779–789.

[120] Skibbe, H., Reisert, M., Schmidt, T., Brox, T., Ronneberger, O., Burkhardt, H.; Fast rotation invariant 3D feature computation utilizing efficient local neighborhood operators. IEEE Trans Pattern Anal Mach Intell 2012;34(8):1563–1575.

[121] Sørensen, L., Shaker, S.B., de Bruijne, M.; Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Trans Med Imaging 2010;29(2):559–569.

[122] Späth, H.; Fitting affine and orthogonal transformations between two sets of points. Mathematical Communications 2004;9(1):27–34.

[123] Sun, S., Gilbertson, M.W., Anthony, B.W.; Probe localization for freehand 3d ultrasound by tracking skin features. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014 - 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II. 2014. p. 365–372.

[124] Torresani, L., Kolmogorov, V., Rother, C.; A dual decomposition approach to feature correspondence. IEEE Trans Pattern Anal Mach Intell 2013;35(2):259–271.

[125] Torsello, A., Bulò, S.R., Pelillo, M.; Grouping with asymmetric affinities: A game-theoretic perspective. In: CVPR (1). 2006. p. 292–299.

[126] Tuceryan, M., Jain, A.K.; Handbook of pattern recognition & computer vision. River Edge, NJ, USA: World Scientific Publishing Co., Inc.; 1993. p. 235–276.

[127] Varga, E., Pattynama, P.M.T., Freudenthal, A.; Manipulation of mental models of anatomy in interventional radiology and its consequences for design of human-computer interaction. Cognition, Technology & Work 2013;15(4):457–473.

[128] Varma, M., Zisserman, A.; A statistical approach to texture classification from single images. International Journal of Computer Vision 2005;62(1-2):61–81.

[129] Vijayan, S., Klein, S., Hofstad, E.F., Lindseth, F., Ystgaard, B., Langø, T.; Validation of a non-rigid registration method for motion compensation in 4d ultrasound of the liver. In: ISBI. 2013. p. 792–795.

[130] Vijayan, S., Klein, S., Hofstad, E.F., Lindseth, F., Ystgaard, B., Langø, T.; Motion tracking in the liver: Validation of a method based on 4D ultrasound using a nonrigid registration technique. Medical Physics 2014;41(8).

[131] Wachinger, C., Klein, T., Navab, N.; Locally adaptive nakagami-based ultrasound similarity measures. Ultrasonics 2012;52(4):547–554.

[132] Wachinger, C., Navab, N.; Simultaneous registration of multiple images: Similarity metrics and efficient optimization. IEEE Trans Pattern Anal Mach Intell 2013;35(5):1221–1233.

[133] Wachinger, C., Wein, W., Navab, N.; Three-dimensional ultrasound mosaicing. In: MICCAI (2). 2007. p. 327–335.

[134] Wang, Z.W., Slabaugh, G.G., Unal, G.B., Fang, T.; Registration of ultrasound images using an information-theoretic feature detector. In: ISBI. 2007. p. 736–739.

[135] Weibull, J.W.; Evolutionary Game Theory. MIT Press, 1995.

[136] Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.; Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 2008;12(5):577–585.

[137] Xu, Z.F., Xie, X.Y., Kuang, M., Liu, G.J., Chen, L.D., Zheng, Y.L., Lu, M.D.; Percutaneous radiofrequency ablation of malignant liver tumors with ultrasound and ct fusion imaging guidance. Journal of Clinical Ultrasound 2014;42(6):321–330.

[138] Zhao, G., Ahonen, T., Matas, J., Pietikäinen, M.; Rotation-invariant image and video description with local binary pattern features. IEEE Transactions on Image Processing 2012;21(4):1465–1477.

[139] Zöllei, L., Learned-Miller, E.G., Grimson, W.E.L., Wells III, W.M.; Efficient population registration of 3d data. In: CVBIA. 2005. p. 291–301.

[140] Zweers, D., Geleijns, J., Aarts, N.J., Hardam, L.J., Laméris, J.S., Schultz, F.W., Kool, L.J.S.; Patient and staff radiation dose in fluoroscopy-guided tips procedures and dose reduction, using dedicated fluoroscopy exposure settings. The British Journal of Radiology 1998;71(846):672–676. PMID: 9849393.

# Samenvatting

In dit proefschrift hebben we algoritmen voor beeldgeleiding in leverinterventies gepresenteerd en geëvalueerd. We hebben de berekening van kenmerken die invariant zijn onder rotatie en beeld-gebaseerde bewegingscompenstatie van real-time 3-dimensionale (3D) ultrageluid beelden geïntroduceerd. Daarnaast hebben we een methode voor de fusie van computer tomografie (CT) en echografie (ultrasound, US) gepresenteerd en geëvalueerd. In dit hoofdstuk vatten we de belangrijkste vindingen van dit werk samen en bespreken richtingen voor verder onderzoek.

Real-time 3D echografie zou gebruikt kunnen worden voor beeldgeleiding in minimaal invasieve ingrepen van de lever. Ademhaling van de patiënt zorgt ervoor dat de lever verschuift, en maakt het moeilijk om continu een specifiek doelgebied af te beelden. De beweging veroorzaakt door ademhaling maakt ook een nauwkeurige combinatie met pre-operatieve data moeilijk. In hoofdstuk 2 ontwikkelden we een snelle affine registratiemethode die kan compenseren voor leverbewegingen veroorzaakt door de ademhaling. De registratie van twee opeenvolgende echografie beelden gebeurde door een gecombineerde snelle blok-vergelijking en verwerping van uitschieters. Voor een verzameling van gelijkmatig verdeelde punten in het ene beeld werden de corresponderende posities in het andere beeld bepaald door blok-vergelijking. Vervolgens werd een robuuste methode om uitbijters te detecteren toegepast, en werden de uitbijters verworpen. De overgebleven punten werden gebruikt om de transformatie te bepalen. Deze aanpak is geëvalueerd op 13 4D echografie series die verkregen zijn van 8 personen. Voor 91 paren van 3D echografie beelden kregen we een gemiddelde registratiefout van 1.8 mm. De methode is geïmplementeerd in een GPU, en deze implementatie verwerkt 3D beelden met een snelheid van 8 Hz.

In hoofdstuk 3 presenteerden we een uitbreiding van het werk van hoofdstuk 2 naar 4D echografie beeldseries. Voor 4D registratie registeerden we een serie opeenvolgende 3D echografie volumes. De transformatie tussen twee volumes werd geschat door de combinatie van alle tussenliggende transformaties. Om opeenstapeling van fouten over een serie transformaties te voorkomen werd een langeafstand terugkoppel mechanisme gepresenteerd. Een totale gemiddelde fout van 1 mm werd verkregen over zes 4D echografie beeldseries van de menselijke lever. De methode verwerkte beelden met een snelheid van 10 Hz. De nadelen van de methode zijn dat het een grote zoekruimte nodig heeft, en niet kan omgaan met grote rotaties.

In hoofdstuk 4 presenteerrden we een verbetering van de methode uit hoofdstuk 3, die slechts getest was op een beperkte dataset, op korte beeldseries. In hoofdstuk 4 beschreven we een snelle en robuuste 4D registratie methode voor het real-time stabiliseren van echografie beeld om beeldgeleiding in transjugulaire intrahepatische portosystemische shunt (TIPS) ingrepen te verbeteren. Hiermee wordt een continue afbeelding van de relevante anatomische vlakken, die vooraf in een planning bepaald zijn, mogelijk gemaakt. De methode vereist een transformatie van de planningsinformatie naar de beelden die tijdens de ingreep gemaakt zijn. Dit werd gerealiseerd in twee stappen: in de eerste stap werd de beweging tussen opeenvolgende beelden in de 4D echografie beeldserie bepaald. Hiermee werd een schatting gemaakt van de transformatie tussen het referentiebeeld (met de planning) en het real-time echografie beeld. Met behulp van deze transformatie werd vervolgens een registratie uitgevoerd van het real-time echografie beeld met het referentiebeeld, om accumulatie van fouten te voorkomen. Deze twee-stappen aanpak hielp in het klein maken van de zoekruimte en is robuust onder rotatie van de beelden. De methode is geëvalueerd op 13 echografie beeldseries van 8 personen. Middels een GPU implementatie verwerkte de methode beelden met een snelheid van 9 Hz met een gemiddelde registratiefout van 1.7 mm. Experimenten met lange beeldseries lieten zien dat er geen accumulatie van de registratiefout plaatsvindt. We presenteerden ook een methode om de registratie te initialiseren en te verifiëren. De verificatie methode hielp in het detecteren en valideren van de registratieresultaten, en werd gebruikt om de registratie zoals hierboven beschreven te initialiseren. De methode bepaalde of het referentiebeeld (met de planningsinformatie) goed geregistreerd is met de beelden uit de real-time 4D echografie serie.

In hoofdstuk 5 presenteerden we een methode voor het volgen van anatomische herkenningspunten in de lever. Deze methode was gebaseerd op een globale en lokale rigide registratie. De methode is gebaseerd op het werk dat gepresenteerd is in de hoofdstukken 2 en 4. De methode is geëvalueerd op de beelden die gepresenteerd zijn in de "MICCAI 2015 Challenge on Liver Ultrasound Tracking (CLUST 2015). Met de beelden die beschikbaar waren om de methode te testen verkregen we een fout van 1.80 ś 1.64 mm.

Beeldkenmerken die niet gevoelig zijn voor rotatie kunnen helpen in het registreren van beelden die verkregen zijn onder verschillende oriëntaties, wat bijvoorbeeld van belang is bij het registreren van beelden die verkregen zijn met verschillende echografie apparaten. Lokaal binaire patronen (LBP) is een populaire beschrijving van textuur die met name in 2D beeldanalyse gebruikt wordt. In hoofdstuk 6 werd een 3D rotatie-invariante beschrijving gepresenteerd die gebaseerd is op 3D LBP. Deze methode genereert een aantal textuurkenmerken per beeldpunt, die onafhankelijk zijn van de oriëntatie van het beeld. Het aantal kenmerken is equivalent met het aantal sferische harmonischen plus de kurtosis. De verdeling van kwantitatief beschrijvende waarden van elk van de kenmerken kan gebruikt worden om een gebied te karakteriseren. De vraag of twee gebieden in twee beel-

den op elkaar lijken (of niet) kan dan beantwoord worden door afstandsmaten die toe te passen zijn op de waardeverdelingen te gebruiken, zoals bijvoorbeeld de Bhattacharyya maat of the Chi-square maat. We hebben de methode toegepast om vaten in beelden van een fantoom te detecteren, en ook in een beeld van een patiënt. De resultaten lieten zien dat het mogelijk was om met deze maat visuele classificatietaken uit te voeren.

In hoofdstuk 8 presenteerden we een methode voor registratie van CT-beelden en echografie beelden voor beeldgeleide ingrepen in de lever. De methode was gebaseerd op de eerder beschreven blok-vergelijkingsmethode, en bevatte ook een nieuwe maat die voor beelden van verschillende oorsprong kan bepalen of ze op elkaar lijken. De blok-vergelijking gebruikte de intensiteit van het CT beeld en de grootte van de afgeleide van het CT-beeld en een echografie beeld. De blok-vergelijking gebruikte een meervoudige correlatie coëfficiënt om de correlatie tussen het CT-beeld en het echografie beeld te bepalen. De geometrische consistentie voor het detecteren van uitschieters was uitgebreid met maten voor de gladheid. De blok-vergelijking resultaten nadat de uitschieters verwijderd zijn, werden weer gebruikt om de transformatie te bepalen. Acht paren van een CT-beeld en echografie beeld van vijf patiënten zijn gebruikt om de instellingen van de methode te optimaliseren, wat resulteerde in een fout van 3.5 mm. Evaluatie van de methode op 21 andere paren van beelden van 12 patiënten lieten een fout zien van 4.3 mm.

## Toekomstige ontwikkelingen

In ons werk hebben we een een snelle registratiemethode voor leveringrepen ontwikkeld, en we hebben hierbij gekeken naar het stabiliseren van de beelden onder invloed van bewegingen door de ademhaling. Het leverweefsel is niet rigide, en de lever kan ook vervormen tijdens de ademhalingsbeweging. We hebben ons in ons werk echter gericht op rigide en affine transformaties, vanwege de volgende overwegingen:

- We richtten ons op een methode die snel kan werken op een GPU, wat realtime toepassing mogelijk maakt. Bij een affine registratie model zijn interpolatie operaties veel minder rekenintensief dan bij niet-rigide registratie modellen.
- De optimalisatie-schema's die nodig zijn voor niet-rigide vervormingen kosten vaak veel meer rekenkracht.
- Het beeldvolume in echografie is niet heel groot, en vaak is een groot orgaan als de lever niet geheel in beeld. Daarom zou een affine transformatie een voldoende goede benadering kunnen zijn.
- Voor de taak van het stabilizeren van de planningsvlakken voor beeldgeleiding is een affine registratie voldoende.

Met bovenstaande in gedachten, zijn we toch van mening dat het onderzoeken en uitbreiden van de methode naar een niet-rigide methode zin kan hebben voor het verder verbeteren van de nauwkeurigheid en verbreding van het toepassingsgebied. Hiervoor is meer onderzoek en ontwikkeling nodig. Ook zou onze methode uit hoofdstuk 4, die in eerste instantie ontwikkeld was voor TIPS ingrepen, relevant kunnen zin voor andere toepassingen, zoals bijvoorbeeld radiotherapie.

In hoofstuk 7 hebben we een methode gepresenteerd voor de registratie van CT- en echografie beelden. Het kan interessant zijn om de relatie tussen de kwaliteit van de afbeelding van de vaten in deze beelden te relateren aan de kwaliteit van de registratie. Een andere interessante studie zou het integreren van de registratie tussen beelden van verschillende scanners in de registratie met beeldseries zijn. Zo'n methode zou relevant kunnen zijn voor percutane behandelingen, zoals van tumoren in de lever en andere organen. Het gebruiken van een bewegingsmodel om nog beter de plaats van de lever te kunnen schatten zou ook nuttig zijn. En ook dit werk zou uitgebreid kunnen worden naar niet-rigide registratie. De CT - echografie registratie zou ook uitgebreid kunnen worden naar andere soort beelden, zoals magnetische resonantie tomografie (MRI) en echografie beelden. Vanuit een klinisch perspectief zou deze technologisch vooruitgang gepaard moeten gaan met ergonomisch goed ontwikkeld gebruiksgemak. De vlakken die loodrecht staan op de vlakken zoals die in de echografie beeldseries verkregen worden hebben echter nog vaak te lijden onder een lage resolutie. Daarom is ook verdere ontwikkeling van de beeldkwaliteit van 4D echografie nodig, omdat deze vlakken relevant zijn voor adequate beeldgeleiding.

Concluderend, in dit proefschrift hebben we methoden gepresenteerd en hun nauwkeurigheid geëvalueerd om real-time 3D echografie beelden te registreren en te fuseren met CT-beelden. Deze methoden kunnen mogelijk gebruikt worden om echografie geleide ingrepen aan de lever te verbeteren. De snelle echografie registratie methode is geïmplementeerd in een systeem voor 4D echografie beeldgeleiding voor TIPS, zoals zichtbaar in figuur 8.1. Verdere uitbreiding naar niet-rigide registratie en registratie met ander type beelden zijn relevante richtingen voor verder onderzoek. Vanuit klinisch gezichtspunt is een grotere validatie studie nodig voordat deze systemen in de kliniek toegepast kunnen worden.

# Acknowledgement

I am blessed with wonderful parents. They encouraged me to follow my heart. My mother has always been my best guide. When I wanted to pursue further studies abroad, they let me to do so with a heavy heart. Thank you Baba and Ma.

I would like to express my heart felt gratitude to Prof Wiro J. Niessen and Dr Theo van Walsum for accepting me as a PhD candidate. The project was a great learning experience. Wiro, I have always admired and looked up to you.

I would like to thank Dr Theo van Walsum for his immense support and guidance. His simplicity and humble nature always stood out. He always had time for us and was prompt in his feedback. His advice rose above narrow thoughts and was of high standards. Theo, I have high regards for you.

I would like to thank Dr Adriaan Moelker for his support. Adriaan, your polite nature and in depth knowledge on the clinician aspects were very helpful. I would like to thank Dr Yu Song for bringing in a lot of energy and exuberance in the project. I would like to remember Dr Adinda Freudenthal for her contributions in the project, may her soul rest in peace.

I would like to thank Dutch Technology Foundation STW for funding the project. I would like to thank all the members of the user committee for their constructive feedback and support. I would like to thank the reviewers for their constructive feedback.

Special mention to Camiel, who took a lot of effort and time in acquiring the data and in performing the joint experiments required for the project. Thank you Renske for helping me with the annotations, without you it would have been very difficult. I would like to thank Peter for helping us with the GPU programming. I would also like to thank Cecile, Edit, Pierre, Valerio, Erwin, Luu and all my BIGR colleagues. Work would not have been so much fun without you all. I also enjoyed all the get-together we had with our families. You all were our friends in a foreign land. Thanks Petra and Desiree for your administrative support. All the past and present members of Medical Informatics and Radiology, I really enjoyed the PhD dinner, lunches, drinks and cakes.

During my growing up years I always nurtured the dream of being a researcher and now I am happy being one.

I would like to thank my brother for his invaluable advice in my career and his family for the love and affection. Special mention to the childrens of my family - Dhruv, Tanmoy and Niloy. Their happy faces make us smile.

Finally I would like to thank Deepleena for being always there. Niloy has brought lots of joy and happiness in our lives.

# Publications

**Journal Papers**

- **J. Banerjee**, C. Klink, R. Gahrmann, W.J. Niessen, A. Moelker and T. van Walsum, Multiple-correlation similarity for block-matching based fast CT and ultrasound fusion in liver interventions, *submitted*
- **J. Banerjee**, C. Klink, W.J. Niessen, A. Moelker and T. van Walsum, 4D Ultrasound Tracking of Liver and its Verification for TIPS guidance, IEEE Transactions on Medical Imaging, 2016
- **J. Banerjee**, C. Klink, E.D. Peters, W.J. Niessen, A. Moelker and T. van Walsum, Fast and Robust 3D Ultrasound Registration - Block and Game Theoretic Matching, Medical Image Analysis, 2015

**Conference Papers**

- H. Ma, G. Dibildox, **J. Banerjee**, W.J. Niessen, C. Schultz, E. Regar and T. van Walsum, Layer Separation for Vessel Enhancement in Interventional X-ray Angiograms Using Morphological Filtering and Robust PCA, Augmented Environments for Computer-Assisted Interventions, LNCS 9365, Springer, 2015
- **J. Banerjee**, C. Klink, E.D. Peters, W.J. Niessen, A. Moelker and T. van Walsum, 4D Liver Ultrasound Registration, Biomedical Image Registration, LNCS 8545, Springer, 2014
- **J. Banerjee**, A. Moelker, W.J. Niessen and T. van Walsum, 3D LBP-based Rotationally Invariant Region Description, ACCV 2012 Workshops, Part I, LNCS 7728, Springer, 2013

**Conference Abstracts**

- S.J.C. Klink, C. Cuijpers, E. Varga, **J. Banerjee**, Y Song, A. Moelker and T. van Walsum, A real-time three-dimensional ultrasound user interface for TIPS: lessons learned, Cardiovascular and Interventional Radiological Society of Europe (CIRSE), 2015
- C. Cuijpers, E. Varga, C. Klink, **J. Banerjee**, P.J. Stappers and A. Freudenthal, A real-time three-dimensional ultrasound user interface for TIPS: Preliminary results, Dutch Biomedical engineering conference (BME), 2015

- E. Varga, **J. Banerjee**, C. Cuijpers, C. Klink, Y Song, A. Moelker and T. van Walsum, Real-time image guidance in TIPS using 4D ultrasound, Cardiovascular and Interventional Radiological Society of Europe (CIRSE), 2015
- C. Klink, C. Cuijpers, E. Varga, **J. Banerjee**, P.J. Stappers, A. Freudenthal, A. Moelker and T. van Walsum, A real-time three-dimensional ultrasound user interface for TIPS: Preliminary results, European congress of radiology (ECR), 2015
- **J. Banerjee**, C. Klink, E.D. Peters, W.J. Niessen, A. Moelker and T. van Walsum, Real-time respiratory motion compensation of the liver using 4D ultrasound and GPU, European congress of radiology (ECR), 2015
- **J. Banerjee**, C. Klink, E.D. Peters, W.J. Niessen, A. Moelker and T. van Walsum, Real-time 3D Ultrasound Registration of Liver, 5th Dutch Bio-Medical Engineering Conference (BME), 2015
- C. Cuijpers, E. Varga, C. Klink, **J. Banerjee**, P.J. Stappers and A. Freudenthal, A real-time three-dimensional ultrasound user interface for TIPS: Preliminary results, Computer Assisted Radiology Surgery (CARS), 2014
- **J. Banerjee**, A. Moelker, W.J. Niessen and T. van Walsum, 3D Rotationally Invariant Region/Landmark Description in Medical Images, 4th Dutch Bio-Medical Engineering Conference (BME), 2013

**Challenges**

- **J. Banerjee**, S.J.C. Klink, E. Vast, W.J. Niessen, A. Moelker and T. van Walsum, A combined tracking and registration approach for tracking anatomical landmarks in 4D ultrasound of the liver, MICCAI - Challenge on Liver Ultrasound Tracking (CLUST), 2015

# PhD portfolio

| | |
|---|---|
| PhD Period | : 2011 - 2015 |
| Department | : Radiology and Medical Informatics |
| Research School | : ASCI, Erasmus MC |

| **Courses** | **Year** |
|---|---|
| Presentation course (Erasmus MC) | 2012 |
| Knowledge driven image segmentation (ASCI) | 2012 |
| Frontend Vision and Multiscale Image Analysis (ASCI) | 2013 |
| Computer Vision by Learning (ASCI) | 2014 |
| A Programmer's Guide for Modern High- Performance Computing (ASCI) | 2014 |
| Research Integrity (Erasmus MC) | 2015 |

| **Conferences and workshops (attendance/presentation)** | |
|---|---|
| Local Binary Patterns (LBP) 2012 Workshop, Daejeon, Korea | 2012 |
| 4th Dutch Bio-Medical Engineering Conference, Egmond aan Zee, Netherlands | 2013 |
| Cardiovascular and Interventional Radiological Society of Europe (CIRSE), Barcelona, Spain | 2013 |
| International Workshop on Biomedical Image Registration, London, UK | 2014 |
| 5th Dutch Bio-Medical Engineering Conference, Egmond aan Zee, Netherlands | 2015 |
| European congress of radiology (ECR), Vienna, Austria | 2015 |
| Cardiovascular and Interventional Radiological Society of Europe (CIRSE), Lisbon, Portugal | 2015 |

**Summer schools**

| | |
|---|---|
| 3R's of Computer Vision: Recognition, Registration, Reconstruction, International Computer Vision Summer School (ICVSS), Sicily Italy | 2012 |
| Modalities, Methodologies and Clinical Research, Biomedical Imaging Summer School, Paris, France | 2013 |

**Others**

| | |
|---|---|
| Anatomy Session (Erasmus MC) | 2013 |
| Integrity in research - Roadshow (Erasmus MC) | 2013 |
| MISP 2014, Leiden, Netherlands | 2014 |
| Erasmus MC Seminars | 2011- 2015 |
| BIGR Seminars (Erasmus MC) | 2011- 2015 |
| MI Research colloquia (Erasmus MC) | 2011- 2015 |

# About the author

Jyotirmoy Banerjee received his Master's Degrees from International Institute of Information Technology at Hyderabad, India in 2009. He joined GE Global Research as a Scientist in 2009. His work at GE John F. Welch Technology Centre in Bangalore, primarily involved developing image analysis algorithms with applications to obstetrics. In 2011, he began his PhD at Erasmus MC, Netherlands. His PhD work aimed at replacing classical surgical interventions by minimally invasive alternatives for liver interventions. In 2016, he joined as a Post-doc at University College London, UK. He is currently involved in the GIFT-Surg project, which is aimed at improving image guided fetal surgery.