

# **Three-Dimensional Ultrasound Mosaicing**

Diploma Thesis in Computer Science

Christian Wachinger

Department of Computer Science  
Technical University of Munich

in collaboration with

Siemens Corporate Research Inc.  
Princeton, NJ, USA



# Three-Dimensional Ultrasound Mosaicing

Diploma Thesis in Computer Science

Christian Wachinger



**Technische Universität München**  
**Fakultät für Informatik**

This thesis was conducted in collaboration with  
Siemens Corporate Research Inc. (SCR), Princeton, NJ, USA.

Autor: Christian Wachinger  
Aufgabensteller: Prof. Dr. Nassir Navab  
Frank Sauer, Ph.D.  
Betreuer: Wolfgang Wein  
Ali Khamene, Ph.D.  
Abgabedatum: 30. March 2007



Ich versichere, dass ich diese Diplomarbeit selbständig verfaßt und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Garching, den 30. März 2007

Christian Wachinger

---



## Abstract

The creation of 2D ultrasound mosaics is becoming a common clinical practice with a high clinical value. The next step coming along with the increasing availability of 2D array transducers is the creation of 3D mosaics. In the literature of ultrasound registration, the alignment of multiple images has not yet been addressed. Therefore, we propose registration strategies, which are able to cope with problems arising by multiple image alignment. We use pair-wise registration with a consecutive *Lie normalization* and *simultaneous registration*, which urges the usage of *multivariate similarity measures*. In this thesis, we propose alternative multivariate extensions based on a maximum likelihood framework. Due to the higher computational cost of simultaneous registration, we describe possibilities for speeding them up, among others, the usage of a stochastic pyramid.

We also present methods to reduce speckle noise and to detect shadow in ultrasonographic volumes, to improve the overall registration performance. For the compounding of the final volume, a variety of approaches are listed, ranging from general to advanced ones. Experimental results, on multiple ultrasound data sets, show the good performance of the proposed registration strategies and similarity measures.

---



## Acknowledgments

I want to thank Frank Sauer for setting the general conditions of my internship at SCR. Thanks to Nassir Navab for the enduring support over the last years. To my supervisor, Wolfgang Wein, for the close collaboration, fruitful discussions and good company, not just limited to office hours. Further from the interventional imaging group I want to thank Ali Khamene for the good hints. Romain for his commitment to the project. And my friends and office mates Stefan, Mike, and Markus. Stefan, for letting me win in tennis and score in soccer. Markus, for being a permanent source of incredible stories. Mike, for inviting us to the nice Japanese restaurant in NY and dedicated discussions about American football. All three for the nice time on lake Carnegie in our canoes. Also Claudia, for that incredible Thanksgiving meal. Patric, for picking the youth hostel with bed bugs. Moritz, for organizing tickets for the Metropolitan opera. The Austrians, the French guys and all the other interns at SCR, thanks a lot, for a sheer infinite number of ice-cream breaks, BBQs, and parties. Back in Munich, I want to thank the guys from the CAMP group. Tobias, for not leaving me so lonesome in the Lab. Lilla, for the feedback on my work.

Seeing this thesis as the last brick of my studies, I want to thank my roommates from the Haus 7 in the StuSta for all the nice moments over the last years. My parents and the rest of my family for their permanent support. Et Méline, dont l'amour a réussi à chauffer mon coeur, en dépit de l'océan qui nous séparait.

---



# Contents

1	Introduction	1
2	Registration Strategies	5
2.1	Problem statement	5
2.2	Bibliography	5
2.3	Imaging Scenarios	7
2.4	Pairwise Strategies	8
2.4.1	Sequential pairwise	8
2.4.2	Complete pairwise	10
2.5	Simultaneous Strategies	12
2.5.1	Full-Simultaneous	13
2.5.2	Semi-Simultaneous	14
2.6	Conclusion	15
3	Similarity Measures	17
3.1	Basic statistical concepts	17
3.2	A maximum likelihood approach for registration	18
3.3	Extension to Multivariate Similarity Measures	22
3.3.1	Summed-up bivariate extension	23
3.3.2	Joint density function	23
3.3.3	Voxel-wise measures	24
3.4	Detail analysis of multivariate measures	25
3.4.1	Sum of Squared Differences	26
3.4.2	Sum of Absolute Differences	28
3.4.3	Normalized Cross Correlation	28
3.4.4	Correlation Ratio	29
3.4.5	Mutual Information	31
3.4.6	Conclusion	33
3.5	Speeding up the similarity calculation	33
3.5.1	Region of overlap	34
3.5.2	Partial Compounding	34
3.5.3	Stochastic similarity measures	35
4	Reconstruction	37
4.1	General Methods	37
4.2	Advanced Methods	38
5	Experiments on ultrasound images	41
5.1	Research prototype	41

---

---

5.2	Non-linear Optimization . . . . .	41
5.3	Ultrasound Images . . . . .	43
5.3.1	Image Formation . . . . .	43
5.3.2	Image pre-processing . . . . .	43
5.4	Experiments and results . . . . .	46
5.4.1	Clay models . . . . .	46
5.4.2	Abdominal phantom . . . . .	48
5.4.3	Baby phantom . . . . .	51
6	Conclusion and Further Development . . . . .	53
6.1	Conclusion . . . . .	53
6.2	Further Development . . . . .	54
6.2.1	Phase information . . . . .	54
6.2.2	Model-based registration . . . . .	55
6.2.3	Optimization . . . . .	55
6.2.4	Non-rigid registration . . . . .	56
6.2.5	Parallelization . . . . .	56
A	Estimation problems on Lie groups . . . . .	59
A.1	Differentiable manifolds and Lie groups . . . . .	59
A.2	Left invariant metric and distance . . . . .	61
A.3	Exponential and logarithmic map . . . . .	61
A.4	Mean and covariance matrix . . . . .	62
B	Detailed experimental results . . . . .	65
	Bibliography . . . . .	77

---

# Chapter 1

## Introduction

The definition of the word *Mosaic* from the Encyclopedia Britannica<sup>®</sup> is:

Surface decoration of small coloured components - such as stone, glass, tile, or shell - closely set into an adhesive ground. Mosaic pieces, or tesserae, are usually small squares, triangles, or other regular shapes.

Briefly, many small pieces of a regular shape are set together to create the large mosaic. In ultrasound (US) mosaicing, which refers to the process of the creation of a US mosaic, something very similar is done, with the 3D ultrasound volumes as the basic building blocks. The creation of ancient mosaics, dating back until the 8th century BC, was very time consuming because of its manual construction. We are interested in an automatic alignment of the ultrasound images. To make this automation possible we have to guarantee a certain amount of overlap between neighboring components. The necessity of overlap between neighboring volumes is the main point, in which mosaicing for ultrasound differs from the given, original definition.

In the literature of ultrasound imaging, the terminology for mosaicing is precisely defined. Other terms, like extended field-of-view (FOV) imaging, volume stitching, and compounding are either used to refer to mosaicing in general or to focus on a specific aspect of mosaicing. For extended FOV imaging, for instance, the interest lies on combining images with a minimal amount of overlap in between to gain the largest possible field-of-view with the least images possible. When referring to compounding, in contrast, the combination of the content of multiple images, taken from the same object from different viewing angles, to increase the quality, is desired. For this purpose, a high amount of overlap between the images has to be guaranteed.

The afore mentioned aspects of ultrasound mosaicing, larger field-of-view and improved image quality, outline its major advantages. While the FOV extension may be evident, we will describe the associated quality improvement with regard to ultrasound imaging in more detail. The crucial question is, why imaging the same situation from different perspectives should increase the image quality? Thinking about other modalities like computed tomography (CT) or magnetic resonance (MR) imaging, the gain of images from different viewing angles - as far as it is possible - may mainly lie in the increase of resolution. For ultrasound imaging, the situation is different. US images are so-called gradient images. Meaning, that interfaces between structures are emphasized. The depiction of these interfaces depend on the incident angle of the ultrasound beam, leading to images that look different from varying perspectives. Mosaicing has the ability to fuse these informations into one consistent reconstruction volume, having the complete contour of the object assembled from the local ones. Another aspect is the occlusion caused by structures

---

with high acoustical impedance like bones. Combining the content of multiple images is therefore very beneficial for ultrasound images because imaging artifacts may be reduced and occluded structures may be augmented in the final reconstructed image.

But the high potential of quality improvement for mosaicing, comes along with a challenging registration scenario. The search for the right global alignment is complicated by the viewing dependent nature of US images, because structures that are necessary to guide the registration may either not be present or look different in various images. We will discuss these issues in more detail throughout this thesis and develop specific registration strategies for this purpose. The search for the correct alignment is performed with a *rigid intensity-based registration*, urging the usage of similarity measures to assign each alignment a certain score. For mosaicing the registration scenario changes since the perfect alignment does not correspond to a maximal overlap, like it is in most cases, putting a special interest on the overlap invariance of these similarity measures. Like the results of our experiments proved, the overlap invariance of the similarity measures is a major issue, being a source of misregistrations between the images. With the introduction of *simultaneous registration* strategies for mosaicing we address this problem.

## Motivation for 3D ultrasound mosaicing

2D ultrasound mosaicing is becoming common clinical practice with a high clinical value, which we further describe in the following section. The next step coming along with the increasing availability of 2D array transducers is the creation of 3D ultrasound mosaics. Since the introduction of 2D array ultrasound transducers into the market is very recent, 3D ultrasound mosaicing is still an active field of research.

### 3D ultrasound imaging

Among several methods available for 3D ultrasonography, the majority rely on the acquisition of a series of 2D frames that are then reassembled to a 3D volume dataset also called ultrasound fan. One can differentiate between a free-hand acquisition using a conventional 2D ultrasound transducer without position sensing, a free-hand acquisition using a conventional 2D ultrasound transducer with position sensing, and an automated acquisition using dedicated mechanical volume probes, so-called "wobbler" probes. These techniques do not provide *real* three-dimensional volumetric imaging since the same technology as for the 2D acquisition is used (linear array probes) and no instantaneous volume acquisition is possible.

During the last years, numerous ultrasound machine vendors introduced 2D array transducers on the market. They enable the instantaneous acquisition of volumetric ultrasound images, leading to a paradigm shift in ultrasonographic imaging, moving from 2D to 3D image acquisition. Due to the high acquisition rate, 4D ultrasonographic imaging (i.e., three spatial dimensions plus motion) can be done with less motion artifacts and higher image quality. Currently available 2D array transducers are based on piezo-electronics arranged in different layers, leading to a massive packaging problem within the probe. The next generation of 2D array US transducers with CMUT<sup>1</sup> technology could accelerate the

---

<sup>1</sup>Capacitive Micromachined Ultrasound Transducer

paradigm shift from 2D to 3D by offering superior and efficient volumetric imaging at a lower cost. The bright prospects for this technology may soon lead to a high availability of 3D ultrasound transducers in the clinical practice, forming the prerequisite for 3D ultrasound mosaicing.

### Clinical value of mosaicing

The usage of ultrasound mosaicing provides the sonographers not just with a compounded volume of higher quality; recent studies also state a couple of other clinical advantages that come along with the extended FOV. First, the spatial relationship among structures that are too large for a single volume is easier to understand [25]. Second, sonographers have the flexibility to visualize anatomical structures from a variety of different angles [40, 31]. Third, size and distance measurements of large organs are possible [72, 25]. Fourth, individual structures within a broader context can be identified by having an image of the whole examination area [9]. And last, because of the increased features in the compounded view, specialists that are used to other modalities than ultrasound can better understand the spatial relationships of anatomical structures [18]; helping to bridge the gap between the modalities and making it easier to convey sonographic findings to other experts.

But it is not just the improvement of already existing workflows, the creation of high quality 3D mosaics may also create new medical applications for ultrasound that do not yet exist at all or are reserved for other modalities. One step into this direction is the creation of special devices that support the image acquisition for US mosaicing. For instance, a device that consists of several US transducers allowing the acquisition of multiple images simultaneously. The advantage is that, due to the acquisition of all the images at the same moment, the deformation of soft tissue is similar in all images, facilitating the rigid registration. One specific application may be the creation of a ultrasound breast scanner, posing an alternative to the X-ray based mammograms. This is especially interesting for countries in which no mammograms are taken because of the caused radiation exposure.

### Outline of this thesis

The crucial point for the construction of high quality mosaics is the good alignment of the images. In chapter 2, we present registration strategies that are adapted for the alignment of multiple images. These strategies include the usage of simultaneous registration, which is based on multivariate similarity measures that we will discuss in chapter 3. After the correct alignment of the images, the reconstruction or compounding of the final image has to be done, for which we present methods in chapter 4. To evaluate the proposed strategies we perform experiments on several data sets that are described together with the obtained results in chapter 5. In the last chapter 6, we give a conclusion of the thesis and present an outlook on further developments that could further increase the performance of mosaicing.

---



# Chapter 2

## Registration Strategies

*In this section, we make an analysis of different imaging scenarios and propose strategies for multiple image alignment adapted to the creation of ultrasound mosaics. The first group of strategies considers pairwise transformations and the problem of how to combine them to a consistent global alignment. The second group consists of tow approaches for simultaneous registration. Each of the strategies has its advantages and drawbacks so that no general purposes best strategy exists, instead specific imaging scenarios require specific strategies.*

### 2.1 Problem statement

In the literature of ultrasound mosaicing, the global alignment of multiple images is deduced from a sequence of pairwise ones. Gee et al. [13] reduce the 3D-3D registration problem to a 2D-2D one by registering the dividing planes to each other. Poon et al. [49] use a block-based rigid and block-based warping approach for the registration. The disadvantages that come along with the usage of pairwise registrations for ultrasound mosaicing are twofold. First, by stitching together pairwise aligned images, registration errors can be accumulated leading to a non-consistent global alignment, see figure 2.1 for a schematic illustration. Second, during the pairwise registrations only a fraction of the available information is taken into account making it prone to misregistrations. The registration is further complicated by the viewing angle dependent US images and the high demands on the overlap invariance by mosaicing.

### 2.2 Bibliography

In this section, we will present an overview of approaches for mosaicing, extended field-of-view imaging, volume stitching and ultrasound registration that were proposed in the literature. One of the first groups registering volumetric ultrasound breast data was Meyer et al. in 1999 [35]. Since no real 3D US probes were available at that time, the volumes were created by linear sweeping. They obtained good registration results for aligning two volumes with rigid registration using mutual information. In the same year, they published another article about 3D extended field-of-view imaging with three volumes, Kruecker et al. [26]. Pairwise affine and warping transformations were used to calculate the alignment

---

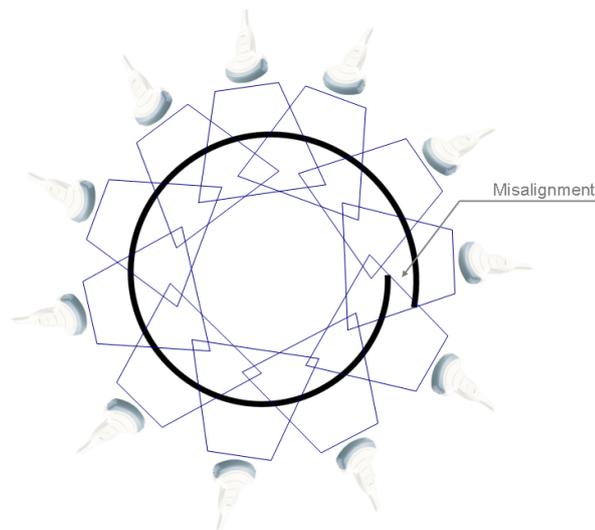


Figure 2.1: Global misalignment due to accumulated pairwise registration errors.

of the data sets. In the following years, they concentrated on non-rigid registration of ultrasound volumes [22, 23]. An elastic registration algorithm was developed that divides the volumes in subvolumes, which are separately, rigidly registered. These blocks serve as control points for thin-plate splines, which are used to calculate the transformation of the whole image.

An even earlier article about extended field-of-view but for 2D ultrasound images comes from Weng et al. [28]. The alignment is done by splitting the images up in blocks for which the local motion vector is calculated. To get a global motion out of the local ones, fuzzy techniques are applied. This method is used in the SieScape<sup>®</sup> technology (Siemens Medical Systems) for 2D mosaicing.

More recent articles about ultrasound stitching come from Gee et al. [13] and Poon et al. [49]. Gee et al. transform the 3D-3D registration problem into a 2D-2D one by picking a plane in the overlapping area on which the registration is performed, the so called dividing plane. This is coupled with a multi-resolution framework to lower the computation time. The volumetric data sets are created with freehand sweeps. Poon et al. create mosaics with an ultrasound “wobbler” probe in combination with a position tracking system. The information of the tracking system serves as initial estimate for the position of the ultrasound volumes which is further refined later on with an intensity-based registration. Therefore two registration approaches, a block-based rigid one and a block-based warping one, were presented. The amount of overlap between the images was about 20-40%.

Other works from Sanches et al. [54], who did a joint registration and volume reconstruction, and Aiger et al. [1], who did mosaicing for visual simulation, were less related to our work. Another approach, that is already in use in the clinical practice is the volume stitching for 4D echocardiography, done by Brekke et al. [5]. At a fixed position, the US probe acquires volumes between different deflection angles. Later on, a simple merging of these volumes by considering the ECG signal is sufficient to get the stitched volume. This

technique is called ECG-gated stitching.

In all the approaches that we have presented for 3D ultrasound mosaicing, the focus was always on registering two volumes and no attention was paid on the changing imaging scenario when dealing with multiple images. Simultaneous or groupwise registration methods address this issue and were already proposed in the literature; intensity-based by Marsland, Twinning et al. [64, 65, 63, 33] and point-based by Fookes et al. [12]. Up to now, simultaneous registration has not yet been used for ultrasound registration.

## 2.3 Imaging Scenarios

In this section, we are analyzing three different imaging scenarios that frequently occur during the image acquisition. An illustration of the scenarios is shown in figure 2.2. Each setup is modeled by a graph, called the *acquisition graph*. It is an undirected, weighted graph with nodes representing the images and edges representing an overlap between the images. The weights of the graph are the amounts of overlap between the image pairs. This graph helps us to assign to each scenario the pairwise registration strategy that fits best. The pairwise strategies are described in the section 2.4.

The *first scenario* shows a linear acquisition of volumes where only neighboring images are overlapping. We have only one possibility how to choose the pairs so that the *sequential pairwise* registration strategy is used.

In the *second scenario*, the situation is a bit more complex because in this case more than only neighboring volumes are overlapping. But there is a very dominant overlap of the neighboring volumes and only a small amount of overlap from the others. Since, in general, the quality of the registration is direct proportional to the amount of overlap we can neglect these small overlaps. Neglecting pairs that only have small overlap is even desired as they can easily produce misregistrations that, in consequence, negatively affect the whole registration. Finding the pairs with the maximum overlap is similar to the calculation of the maximum spanning tree on the acquisition graph. This is done with the Kruskal minimum spanning tree algorithm by negating the graph weights [27]. The *sequential pairwise* registration strategy is used once again, this time on the spanning tree.

The *third scenario*, which simulates the scanning around an extremity or the trunk, is the most complex one because there exists a high overlap between all the images. Therefore, neglecting pairs like in the in the previous case does not work anymore. In this scenario, we are interested in calculating registrations between all the available image pairs. The acquisition graph in this scenario is a complete graph which has up to  $\frac{n \cdot (n-1)}{2}$  edges, with  $n$  the number of nodes. Hence, the number of pairwise registrations that have to be performed increases quadratically with the number of images. In this scenario we use the *complete pairwise* registration.

Like we mentioned in section 2.1, there are several problems that can occur during the registration. In the previous paragraphs, we considered pairwise strategies for the imaging scenarios in figure 2.2. We assigned best strategies to the scenarios based on accumulated errors and misregistration due to small overlap. For all the scenarios, *simultaneous* registration strategies may make sense, which are described in section 2.5. The reason for its usage lies in the, in general, better conditioned optimization problem on cost functions created by *multivariate* similarity measures. They consider the information of the whole imaging

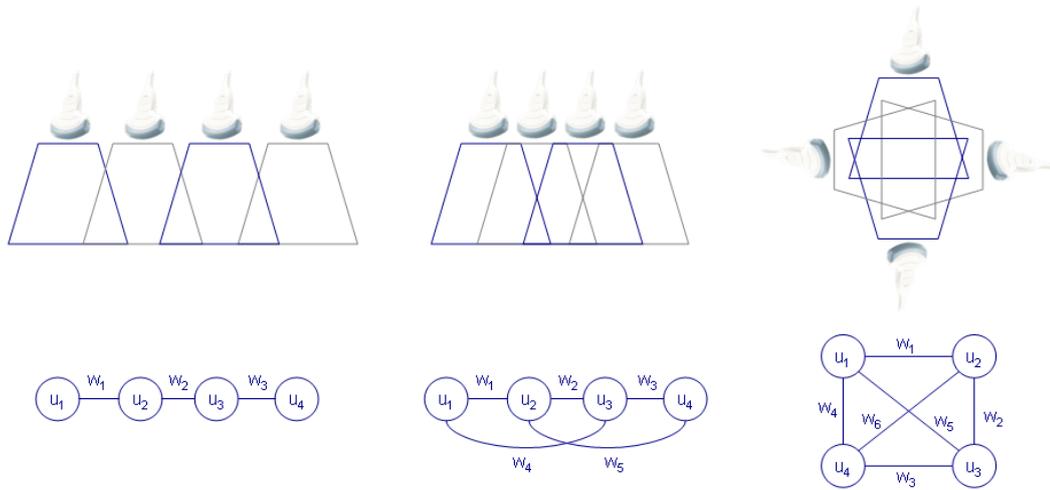


Figure 2.2: Three imaging situations with their associated acquisition graphs.

setup, trying to get the maximum out of the depicted structures. That aspect is especially important for viewing angle dependent US images. Moreover, the accumulated error does not occur during simultaneous registration, because errors are dealt with intrinsically during the registration. These advantages are more relevant for complex scenarios. We have to point out that higher computational costs have to be accepted when using simultaneous strategies, so that pairwise ones are not superfluous.

## 2.4 Pairwise Strategies

All the overlapping image pairs are aligned separately with a pairwise rigid registration. Afterwards, the calculated transformations are used to deduce the correct global alignment, which corresponds to assigning world coordinates to the images. The acquisition graph like it is shown in figure 2.2 is very useful to have an overview of the imaging setup. In the following paragraphs, we will explain the sequential and complete pairwise strategies in more detail.

The notation that we use throughout this report is to denote the  $n$  images by  $u_1, \dots, u_n$ . The pairwise transformation between  $u_i$  and  $u_j$  is  $T_{i,j}$  with  $i, j \in \{1, \dots, n\}$ . The global position of the image  $u_i$ , meaning the location with reference to the world coordinate system, is  $T_i$ . The set of all global transformations is  $\mathcal{T} = \{T_1, \dots, T_n\}$ .

### 2.4.1 Sequential pairwise

In the case of an acquisition graph that has the tree property, the sequential strategy is used. This property is important as it assures that there is exactly one path between two nodes, which guarantees the ambiguity free calculation of the world coordinates. Any of

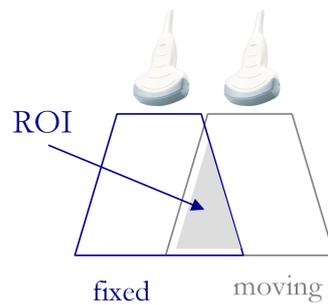


Figure 2.3: ROI defined in the overlapping region

the nodes in the graph can be picked as root, which coordinate system is, without loss of generality, seen as world coordinate system. Calculating then the world coordinates of the other images is done by traversing the graph where the edges are augmented with the associated transformations see figure 2.4(a). The transformation  $T_{i,j}$  is the result of the pairwise registration between the images  $u_i$  and  $u_j$ . By augmenting the graph with the pairwise transformations it becomes directed because transformations show the relationship between two images in a specific order. But this does not pose any restrictions on the traversal of the graph because rigid transformations are invertible.

These 3D rigid registrations have 6 degrees-of-freedom (DOF), 3 DOF account for the rotation and 3 DOF for the translation of the images. Conform with the standard, we first apply the rotation and then the translation. One has to pay attention to this detail, since the two parts are not commutative. During the registration we use a standard multi-resolution framework having a. o. the advantage of making the registration less sensitive to local minima, see section 3.5.3 for more details. The smoothed images of the Gaussian pyramid are especially beneficial for ultrasound images because the filtering reduces speckle noise like described in section 5.3.2.2.

### Volumes overlap

The major difference between the usual registration scenario and the mosaicing one is the amount of overlap between the volumes. Usually, the perfect alignment corresponds to the total overlap of the images. Since we are interested in extending the field-of-view, a high overlap is counterproductive. A compromise has to be found between having enough overlap to enable a robust and accurate registration, and enough non-overlapping parts to guarantee a fast extension of the field-of-view. It is hard to quantify the amount of overlap that is sufficient because it mainly depends on the number of structures being present in the overlapping region to guide the registration. In [49] the overlapping region made up about 20% to 40% of the size of each volume. Often, the scanning surface is convex; therefore deeper parts are more overlapping.

The limited amount of overlap makes high demands on the similarity measure because we are interested in a smooth cost function even by changing amounts of overlap. There are particular designed measures, like normalized mutual information, for being invari-

ant to changing overlap. But since the percentage of changing overlap compared to the total overlap per optimization step is very high in the mosaicing scenario, these special measures are not able to cope with it, see section 5. Therefore, we defined a region of interested (ROI) within the overlap over which the similarity measure is calculated. The ROI stays constant for the whole registration so that the values are more consistent and therefore better comparable.

The ROI is calculated by eroding the initial overlap. Since it is only the moving image that changes the position during the registration it is sufficient to erode it along the border of the moving image. In detail, we apply an erosion kernel of the intended size to the moving image and then joining it with the fixed volume by a logical conjunction. In figure 2.3 a schematic representation of the ROI is shown. The erosion should not be too large because otherwise the already small overlap may get too small for a robust registration.

## 2.4.2 Complete pairwise

For the complete pairwise strategy we do not pose any constraints on the acquisition graph. Therefore, several paths between two nodes in the graph can exist. This can lead to ambiguities. Imagine an acquisition graph like in figure 2.4 with the subgraph  $(u_1, u_2, u_3)$ . Theoretically the equation  $T_{2,3} \cdot T_{1,2} = T_{1,3}$  holds, but due to registration errors we have to add an error term  $\varepsilon$

$$T_{2,3} \cdot T_{1,2} = T_{1,3} \cdot \varepsilon.$$

The problem is now to find global transformations  $T_k$  so that they best fit to the registration transformations  $T_{i,j}$ , with  $k \in \{1, \dots, n\}$  and  $(i, j)$  an edge in the acquisition graph. To illustrate the problem, we augment the acquisition graph with a world node  $w$  which has edges to all the images representing the world (global) transformations  $T_k$ , see figure 2.5. By using the path  $(w, u_i, u_j)$ , we can address a certain error  $\varepsilon_{i,j}$  to each registration transformation:

$$\varepsilon_{i,j} = T_i^{-1} \cdot T_j \cdot T_{i,j}.$$

This finally leads to an acquisition graph where the edges are augmented with the registration errors, which depend on the estimated global transformations shown in figure 2.5. Our goal is it to estimate a set of consistent global transformations by using all the available pairwise rigid registration results. A good choice is to consider a least-squares approach

$$[\hat{T}_1, \dots, \hat{T}_n] = \arg \min_{[T_1, \dots, T_n]} \sum_{(i,j)} \mu^2(\varepsilon_{i,j})$$

with a so far not specified error function  $\mu$ .

### Lie group based error function

The error function  $\mu$  assigns each transformation  $\varepsilon_{i,j}$  a quality or distance value. The classical notions for distance are not trivial to generalize to rigid body transformations since they do not form a vector space but rather a Lie group that can be considered as a Riemannian manifold. Paradoxes such as Bertrand's paradox [38, 45] appear when one considers a Lie group as a vector space within an estimation problem.

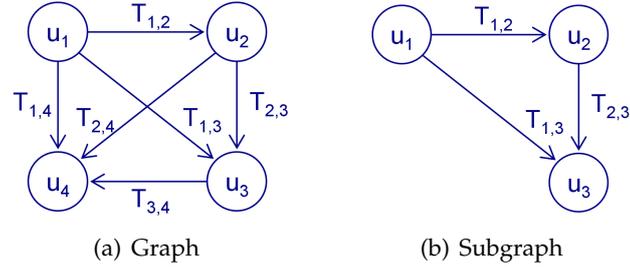


Figure 2.4: (a) An acquisition graph with 4 nodes (images) and 6 edges (overlaps) which is augmented by the registration transformations. As the transformation represents an ordered relationship the graph is directed. (b) Subgraph ( $u_1, u_2, u_3$ ) of the acquisition graph.

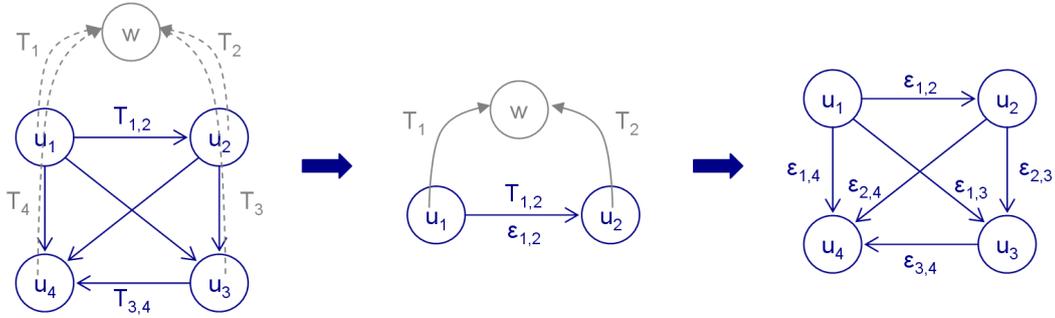


Figure 2.5: The augmentation of the graph with the world node  $w$  leads to an extended acquisition graph. The subgraph ( $w, u_i, u_j$ ) is used to assign each edge an error term  $\epsilon_{i,j}$ , leading to an acquisition graph weighted with pairwise errors.

There exist various approaches for averaging transformations. Davis et al. [7] solve a linear system of equations in a least square sense. Unfortunately, this approach works only for projective matrices and constraints would have to be placed on the optimization framework to work with rigid matrices. Gramkow et al. [14] propose a quaternion based approach for averaging. Pennec [44] showed that the only valid approach is based on Lie groups. In Appendix A we present basic techniques for estimation problems on Lie groups, especially focusing on the derivation of the statistical concepts of mean, covariance matrix, and Mahalanobis distance.

We adapt the Mahalanobis distance, see A.4, to our estimation problem like described in [67]. Let  $\epsilon$  be the random error with Fréchet matrix mean identity and covariance matrix  $\Sigma_{\epsilon\epsilon}$ . The samples of the random error are denoted by  $\epsilon_{i,j}$ . The Lie group based error function is then

$$\mu_{\epsilon}^2(\epsilon_{i,j}) = \log_{\text{Id}}(\epsilon_{i,j})^T \cdot \Sigma_{\epsilon\epsilon}^{-1} \cdot \log_{\text{Id}}(\epsilon_{i,j})$$

and the optimization criterion

$$[\hat{T}_1, \dots, \hat{T}_n] = \arg \min_{[T_1, \dots, T_n]} \frac{1}{2} \sum_{(i,j)} \mu_{\varepsilon}^2(\varepsilon_{i,j}).$$

The covariance matrix can be estimated with the data as explained in appendix A.4. We start with a diagonal covariance matrix that weights the angles with respect to the translation. The proposed criterion is invariant to a composition of all transformations with a fixed transformation. To avoid a drifting of the images, an arbitrary transformation  $T_k$  is set to the identity.

During our tests we found out there are significant differences in registration quality depending on the amount of overlap between images. Images with a high amount of overlap, having many structures to guide the registration, produce better results than images with a small overlap. Therefore we are introducing quality weights in the criterion. We use the amount of overlap  $\omega_{i,j}$  between image  $i$  and  $j$  to estimate the quality. The modified criterion is

$$[\hat{T}_1, \dots, \hat{T}_n] = \arg \min_{[T_1, \dots, T_n]} \frac{1}{2} \sum_{(i,j)} \omega_{i,j} \cdot \mu_{\varepsilon}^2(\varepsilon_{i,j}),$$

focusing on registration pairs with high overlap. One can even think of using  $\exp(\omega_{i,j})$  to further strengthen their influence.

### Algorithm

Like already mentioned in the problem statement in section 2.1 we have to differentiate between two different kind of registration errors. First, accumulated errors leading to a non-consistent global alignment shown in figure 2.1, and second, registration errors during the pairwise registration. For errors of the first kind we can assume that the Lie averaging results in a good global alignment. This does not hold anymore for errors of the second kind, where we can only try to detect misregistrations and restart the registration from the averaged transformations. The detection of misregistrations is done by comparing the total error

$$\varepsilon_t = \sum_{(i,j)} \omega_{i,j} \cdot \mu_{\varepsilon}^2(\varepsilon_{i,j})$$

to a certain error threshold  $\delta$ . The registration is restarted until the total error is below the threshold, then the averaged transformations are returned. The final algorithm for the Lie registration is shown in table 2.1.

## 2.5 Simultaneous Strategies

In this section, we describe a *simultaneous* approach for solving the registration problem for multiple images. In the last section, we described the issues that arise when trying to calculate the global transformations from the pairwise ones. To avoid the normalization problems, we now directly optimize the global transformations. This optimization makes only sense if the registration score depends on all the images and not just on two. Therefore, we have to use multivariate similarity measures, which have the ability to produce

- 
1. Start with initial global transformations  $\mathcal{T} = \{T_1, \dots, T_n\}$
  2. **Do**
    - 2.1 Deduce initial pairwise transformations  $T_{i,j}$  from  $\mathcal{T}$
    - 2.2 Compute all pairwise registrations  $T_{i,j}$
    - 2.3 Estimate new  $\mathcal{T}$  from calculated  $T_{i,j}$  with Lie normalization
  3. **While** ( $\varepsilon_t > \delta$ )
  4. Return  $\mathcal{T}$
- 

Table 2.1: Algorithm for pairwise registration with Lie normalization.

better conditioned cost functions for the optimizer because they consider the whole imaging scenario. The extension of the similarity measures is not trivial and described in detail in section 3.

Simultaneous registration, also referred to as groupwise registration, is a recent technique used for population studies [75, 3] but has not yet been used for ultrasound mosaicing. We will present a *full-simultaneous* and a *semi-simultaneous* registration approach that vary in their computational complexity.

### 2.5.1 Full-Simultaneous

In the full-simultaneous approach we optimize over  $6 \cdot n$  parameters. Having a best neighbor optimizer this leads to  $12 \cdot n$  evaluations of the cost function per optimization step. Moreover, the evaluation of the multivariate cost function is computationally more complex than in the pairwise case. This indicates the, in general, higher computational complexity of the simultaneous registration. But it is hard to prove the increase of complexity because one can argue that the simultaneous registration steps are of higher quality and therefore more efficient so that less steps are necessary. We present possible techniques to speed the calculation up in section 3.5.

A question that comes up when using the simultaneous registration is whether to optimize over  $n$  or  $n - 1$  images. Theoretically,  $n - 1$  images are sufficient and having one image fixed would avoid a common drift of the images. But imagining the situation that the fixed image is pretty bad aligned to the other  $n - 1$  moving ones. The easiest way is to transform one image, the fixed one, and not all the other  $n - 1$  moving ones. However, this easy solution is blocked making it necessary to move all  $n - 1$  images, leading to a higher computational cost and a higher probability of registration errors. In order to prevent this situation we always optimize over all  $n$  images, making a subsequent transformation normalization necessary.

#### Transformation Normalization

The transformation normalization eliminates a general drift outside of the world coordinate center. It is less crucial than the normalization for the pairwise registrations. The similarity measures only depend on the relative position of the images to each other, so there

---

- 
1. **While** global alignment not satisfying
    - 1.1 **For each**  $i$  **in**  $\{1, \dots, n\}$ 
      - 1.1.1 Simultaneously register image  $u_i$  to  $\{u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n\}$  for  $k$  steps, optimizing the matrix  $T_i$
    - 1.2 **END For**
    - 1.3 Increment  $k$
  2. **END While**
  3. Return  $\mathcal{T}$
- 

Table 2.2: Algorithm for semi-simultaneous registration

is no constraint that keeps the images near the coordinate center. One way to limit this perturbation would be to ensure an identity transformation for e.g. the first image after each registration cycle. Another way would be to ensure that the average transformation is the identity like it was done in [74]. Each transformation would have to be multiplied by an unknown matrix  $T_0$

$$T_I = T_0 \left[ \frac{1}{n} \sum_{i=1}^n T_i \right] \Leftrightarrow T_0 = \left[ \frac{1}{n} \sum_{i=1}^n T_i \right]^{-1}.$$

For the estimation of the mean transformation we can use the Lie group estimation framework described in Appendix A.

### 2.5.2 Semi-Simultaneous

The semi-simultaneous approach is a modification of the full one by just optimizing over 6 pose parameters of one image at a time. This has implications on the design of the registration, which is now divided into several cycles, and the multivariate similarity measure. The idea of using cycles is further explained in the following paragraph. Specific multivariate similarity measures can be developed by concentrating on one image at a time, and measuring how well the other ones describe it. Moreover, we limit the evaluation of the measure to the area of that image therewith making it faster. For more information, the reader is referred to section 3.

An example algorithm for semi-simultaneous registration is stated in table 2.2. In general, one could first register the first image until convergence, then the second one and so forth. We think that this is not a good strategy to approach a good global alignment. Instead, we use a registration technique based on cycles. In each cycle, all images are registered one after another for a certain number  $k$  of registration steps. The question comes up in which order we process the images per cycle. Possibilities are either to use simple linear order or a random one. When the cycle is completed it is checked whether a satisfying registration accuracy is reached. If not, a new cycle is started with an incremented number of registration steps. Otherwise the algorithm terminates and returns the global transformations  $\mathcal{T}$ .

---

## 2.6 Conclusion

In table 2.3, the described approaches are characterized by the number of parameters per optimization step, the kind of similarity measure, and the kind of overlap considered. The demands on multivariate measures are slightly different for the semi- and full-simultaneous registration, leading to two different variants of summed-up measures in section 3. Sequence overlap means that only sequentially overlapping images are considered.

	Sequential PW	Complete PW	Semi-Simultaneous	Full-Simultaneous
Optimization	6	6	6	$n \cdot 6$
Similarity	Bivariate	Bivariate	Multivariate	Multivariate
Overlap	Sequence	All	All	All

Table 2.3: Summary of registration strategies.



# Chapter 3

## Similarity Measures

*In this chapter, we describe multivariate similarity measures and how they can be deduced from the bivariate ones. First, we present a maximum likelihood approach for pairwise registration which we consecutively extend to multiple images. This allows us to derive a new class of multivariate measures that consist of summing up the bivariate equivalents. Later on, we consider four selected similarity measures, sum of squared differences, normalized cross correlation, correlation ratio, and mutual information in detail. For sum of squared differences, we deduce a voxel-wise criterion. At the end, we present more practical considerations, e.g., how to speed the calculation up.*

### 3.1 Basic statistical concepts

A famous quotation of Pierre-Simon Laplace about probability theory is:

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge.

Since we are also interested in benefiting of this most important object of human knowledge, we artificially consider images as random variables. But this is not just due to the advice of Monsieur Laplace, it also offers us a convenient framework for the similarity estimation. Although it may seem strange to consider data as random variables, the probabilistic concepts provide powerful means to design similarity measures. This artificial randomness should not be confused with the stochastic nature of the noise that affects the images.

When randomly selecting a pixel in the image  $u$  the probability that its intensity is  $i$ , is proportional to the number of pixels,  $N_i$ , in  $u$  having the intensity  $i$

$$P(i) = \frac{N_i}{N}$$

with  $N$  the number of pixels in  $u$ . The image histogram has therefore the role of the probability density function (pdf). Sometimes one is interested in having a smooth pdf, especially when calculating the gradient. For this purpose, estimators like smooth histogramming or parzen windowing [68] can be used. To extend the pdf to a joint-pdf we consider two images  $(u, v)$  and a spatial transformation  $T$  that maps the grid of  $v$ ,  $\Omega_v$ , to

---

the grid of  $u$ ,  $\Omega_u$ . The images  $u$  and  $v$  have a common set  $\mathcal{A}$  of intensity values, normally  $\mathcal{A} = \{0, 1, \dots, 255\}$ :

$$u : \Omega_u \rightarrow \mathcal{A}, \quad v : \Omega_v \rightarrow \mathcal{A}.$$

In order to find corresponding pixels in the both images we have to transform the image  $v$  with the transformation  $T$  leading to the transformed image  $v^\downarrow$ :

$$\begin{aligned} v^\downarrow = T(\Omega_v) &\rightarrow \mathcal{A}, \\ x &\mapsto v(T^{-1}(x)). \end{aligned}$$

Since, in general, points in  $\Omega_v$  do not transform to points in  $\Omega_u$  we have to interpolate. We were using trilinear interpolation that considers the eight neighboring voxels. The joint-pdf between  $u$  and  $v^\downarrow$  is then

$$P_T(i, j) = \frac{\text{Card} \{x | (u(x), v^\downarrow(x)) = (i, j)\}}{\text{Card} \Omega_u}.$$

*Statistics* of random variables can be calculated that sum up its long term behavior. An example of a statistic is the *mean* value that can be approximated by summing over the grid points

$$\mathbb{E}[u] = \frac{1}{N_u} \sum_{x \in \Omega_u} u(x).$$

A second order statistic is the *variance* defined by

$$\text{Var}(u) = \mathbb{E}[(u - \mathbb{E}[u])^2] = \mathbb{E}[u^2] - \mathbb{E}[u]^2.$$

Another kind of statistic is the *entropy* that summarizes the randomness of a random variable. A variable has higher entropy the more random it is. In can be written in form of an expectation

$$H(u) = -\mathbb{E}[\log P(i)] = -\sum_{i \in \mathcal{A}} P(i) \cdot \log P(i).$$

As a last statistical concept, we describe the Gaussian distribution which is one of the most popular ones. Having the mean  $\mu$  and the variance  $\sigma^2$  it is defined by

$$g_\sigma(x - \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}.$$

## 3.2 A maximum likelihood approach for registration

After introducing some basic statistical concepts in the last section, we now focus on the maximum likelihood estimation (MLE) and how it can be used for image registration [68, 52]. The maximum likelihood estimation is used to get the most probable model  $m$  when having a set of observations  $a$ . If the observations are statistically independent we can write the pdf as

$$P(a = \{a_1, \dots, a_n\} | m) = \prod_{i=1}^n P(a_i | m).$$

This expresses the probability of the observations having the model. Since we are interested in the inverse problem, we define the *likelihood*  $\mathcal{L}$  by reversing the roles of observations and model:

$$\mathcal{L}(m|a) = P(a|m).$$

The estimation can be simplified by maximizing the logarithm of the likelihood

$$\log \mathcal{L}(m|a) = \sum_{i=1}^n \log P(a_i|m).$$

Often we simply write  $\mathcal{L}(m)$  if the related observations are obvious.

In order to apply the MLE to the registration problem we first have to model the imaging process. Once again we have the images  $u$  and  $v$ , the transformation  $T$ , and an intensity mapping  $f$ . The intensity mapping is necessary because we cannot assume that the commonly imaged object has exactly the same intensities in both images. Moreover, a stationary<sup>1</sup> white Gaussian noise  $\varepsilon$  corrupting the images is added, so that we finally end up with

$$u(x) = f(v(T(x))) + \varepsilon.$$

To get to the MLE, we have to express the registration in the form of a probability function. The way it is done, is to measure how well the image  $u$  is described by  $v$ :

$$P(u|v, T, \varepsilon, f) = P(\varepsilon = u(x) - f(v(T(x_k))))$$

leading to the negative log likelihood function

$$\begin{aligned} -\log \mathcal{L}(T, \varepsilon, f) &= -\log P(u|v, T, \varepsilon, f) \\ &= -\log P(\varepsilon = u(x) - f(v(T(x_k)))) \end{aligned} \quad (3.1)$$

Before further concentrating on the MLE we have to introduce some more notations. We shortly write the intensities  $u(x_k) = i_k$  and  $v(T(x_k)) = j_k^\downarrow$ . The intensity function maps each tissue class  $j$  in the image  $u$  with  $f(j) = f_j$ . Using the formula for the Gaussian we write

$$P(u|v, T, f, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(i_k - f(j_k^\downarrow))^2}{\sigma^2}}.$$

By setting it into the MLE framework and assuming that they are independently distributed we get

$$\begin{aligned} -\log \mathcal{L}(T, f, \sigma) &= -\sum_{x_k \in \Omega_u} \log P(u(x_k)|v^\downarrow(x_k), T, f, \sigma) \\ &= N \log \sqrt{2\pi}\sigma + \frac{1}{2} \sum_{x_k \in \Omega_u} \frac{(i_k - f(j_k^\downarrow))^2}{\sigma^2} \end{aligned} \quad (3.2)$$

When further setting  $\Omega_u^j = \{x_k \in \Omega_u, j_k^\downarrow = j\}$ :

$$-\log \mathcal{L}(T, f, \sigma) = N \log \sqrt{2\pi}\sigma + \frac{1}{2} \sum_j \sum_{x_k \in \Omega_u^j} \frac{(i_k - f_j)^2}{\sigma^2}. \quad (3.3)$$

The transformation is implicitly modeled in  $f_j$ . Starting from this equation will deduce various similarity measures in the following paragraphs.

<sup>1</sup>Stationary means independent on the spatial position.

### Sum of Squared Differences

To derive the sum of squared differences we set the intensity function to the identity. Equation 3.2 has then the form

$$-\log \mathcal{L}(T, \sigma) = N \log \sqrt{2\pi}\sigma + \frac{1}{2} \sum_{x_k \in \Omega_u} \frac{(i_k - j_k^\downarrow)^2}{\sigma^2}.$$

The important term for the optimization is the sum over the intensities so that a cost function of the following form is written:

$$\begin{aligned} C(T) &= \mathbb{E} [(i_k - j_k^\downarrow)^2] \\ &\propto -\log \mathcal{L}(T). \end{aligned} \quad (3.4)$$

This corresponds to the well known sum of squared differences.

### Normalized Cross Correlation

$C(T)$  cannot be used to align images that have an intensity mapping  $f$  other than the identity. However,  $C(T)$  can be generalized to work with images that have been transformed by an affine transformation, meaning:  $u \approx \alpha v + \beta$ . This generalization is done by normalizing the images. A normalized image has a zero mean and an one standard deviation and is calculated by

$$\tilde{u}(x) = \frac{u(x) - \mathbb{E}[u(X)]}{\text{Var}(u(X))}. \quad (3.5)$$

Setting this in equation 3.4 leads to the *normalized sum of squared differences* (NSSD)

$$\begin{aligned} \text{NSSD}(T) &= \mathbb{E}[(\tilde{v}(T(X)) - \tilde{u}(X))^2] \\ &= \mathbb{E}[\tilde{v}^2(T(X)) - 2\tilde{v}(T(X))\tilde{u}(X) + \tilde{u}^2(X)] \\ &= \mathbb{E}[\tilde{v}^2(T(X))] - 2\mathbb{E}[\tilde{v}(T(X))\tilde{u}(X)] + \mathbb{E}[\tilde{u}^2(X)] \\ &= 1 - 2 \frac{\mathbb{E}[v(T(X))u(X)] - \mathbb{E}[v(T(X))]\mathbb{E}[u(X)]}{\text{Var}(v(T(X))) \text{Var}(u(X))} + 1 \\ &= 2 - 2 \text{NCC}(v(T(X)), u(X)) \end{aligned}$$

with the standard formula for the *normalized cross correlation* NCC

$$\text{NCC}(T) = \mathbb{E} \left[ \frac{(u(x_k) - \mathbb{E}[u(X)]) (v(x_k^\downarrow) - \mathbb{E}[v(X)])}{\text{Var}(u(X)) \cdot \text{Var}(v(X))} \right].$$

This shows that the two measures are proportional to each other,  $\text{NSSD}(T) \propto \text{NCC}(T)$ , and therefore, that the NCC can be deduced from the MLE approach.

### Correlation Ratio

To derive the Correlation Ratio (CR) we continue with our maximum likelihood approach from equation 3.3. We consider a general intensity mapping  $f$  so that the parameter vector

that has to be estimated is  $\vartheta = (f_0, f_1, \dots, \sigma)$ . We therefore assume a functional relationship between the intensity values of the images. For finding the right parameters the log likelihood is differentiated by  $\vartheta$

$$\begin{aligned} \frac{\partial \log P}{\partial f_j} &= \frac{1}{\sigma^2} \sum_{\Omega_u^j} (i_k - f_j) & \implies \hat{f}_j &= \frac{1}{N_j} \sum_{\Omega_u^j} i_k \\ \frac{\partial \log P}{\partial \sigma} &= -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_j \sum_{\Omega_u^j} (i_k - f_j)^2 & \implies \hat{\sigma}^2 &= \sum_j \frac{N_j}{N} \sigma_j^2 \end{aligned}$$

with  $N_j = \text{Card } \Omega_u^j$  and the image variances  $\sigma_j^2 = \frac{1}{N_j} \sum_{x_k \in \Omega_u^j} (i_k - \hat{f}_j)^2$ . Substituting the optimal parameter vector  $\vartheta$  leads to the registration energy  $U(T)$

$$\begin{aligned} U(T) &= \min_{\vartheta} \left[ - \sum_{x_k \in \Omega_u} \log P(u(x_k) | v(x_k), T, \vartheta) \right] \\ &= \frac{N}{2} \log \left( 2 \pi e \sum_j \frac{N_j}{N} \hat{\sigma}_j^2 \right) \\ &= \frac{N}{2} \log (2 \pi e \text{Var}(u - \hat{f}(v^\downarrow))) . \end{aligned}$$

$U(T)$  decreases with the variance of the difference between  $u$  and the intensity corrected  $\hat{f}(v)$ . When considering the formula for the correlation ratio

$$\eta^2(u|v) = 1 - \frac{\text{Var}(u - \hat{f}(v^\downarrow))}{\text{Var}(u)}$$

then registration energy  $U(T)$  is related to it by

$$\eta^2(u|v) = 1 - \frac{1}{2\pi e \text{Var}(u)} e^{\frac{2 \cdot U(T)}{N}}$$

with  $\hat{f}(v)$  the least squares optimal non-linear approximation of  $u$  in the terms of  $v$ . Minimizing  $U(T)$  is then strictly equivalent to maximizing the correlation ratio.

### Mutual Information

A functional relationship between the images is sometimes too restrictive. For this purpose, a more general, *information theoretic* measure is used. The most popular one is the mutual information that is derived in the following paragraph. Therefore, the formal constraints on our model are further relaxed so that we only assume that the densities  $P(i_k | j_k^\downarrow)$  are stationary. The log likelihood that has to be maximized is

$$\log P(u|v, T, \vartheta) = \sum_{x_k \in \Omega_u} \log f(i_k | j_k^\downarrow)$$

with  $\vartheta = (f(0|0), f(1|0), \dots, f(0|1), \dots, f(0|2), \dots)$  and the constraints:

$$\forall j, C_j = \sum_i f(i|j) = 1. \quad (3.6)$$

The intensity pairs  $(i_k, j_k^\downarrow)$  having the same values are regrouped so:

$$\log P(u|v, T, \vartheta) = \sum_{i,j} N_{i,j} \cdot \log f(i|j),$$

with

$$\Omega_{i,j} = \{x_k \in \Omega_u, u(x_k) = i, v(T(x_k)) = j\}, \quad N_{i,j} = \text{Card } \Omega_{i,j}.$$

Using Lagrange multipliers, there exist constants  $\lambda_0, \lambda_1, \dots$  such that for any  $j$ :

$$0 = \frac{\partial \log P}{\partial f(i|j)} - \sum_{j'} \frac{\partial C_{j'}}{\partial f(i|j)} = \frac{N_{i,j}}{f(i|j)} - \lambda_j.$$

Together with the constraints in equation 3.6, the optimal parameters verify:

$$\hat{f}(i|j) = \frac{N_{i,j}}{N_j} = \frac{P(i,j)}{P(j)},$$

where  $P(i,j) \equiv N_{i,j}/N$  is the image normalized 2D histogram and  $P(j) \equiv \sum_i P(i,j)$  the corresponding marginal distribution for  $v^\downarrow$ . The energy function  $U(T)$  is a decreasing function of mutual information:

$$U(T) = N \sum_{i,j} P(i,j) \log \frac{P(i,j)}{P(j)} = N [\text{H}(u) - \text{NMI}(u, v^\downarrow)],$$

with  $\text{H}(u)$  the entropy of image  $u$  and  $\text{NMI}(u,v)$  the normalized mutual information

$$\text{NMI}(u, v) = \sum_i \sum_j P(i,j) \log \frac{P(i,j)}{P(i) \cdot P(j)}.$$

## Conclusion

In the last paragraphs, we used a MLE framework to derive four popular similarity measures by following the works of [68] and [52]. Each of these measures assumes different relationships of the intensities between the images, modeled in the intensity mapping  $f$ . The following list summarizes these relationships:

- Identity relationship: Sum of Squared Differences, Cross-Correlation
- Affine relationship: Normalized Cross Correlation
- Functional relationship: Correlation Ratio
- Statistical relationship: Mutual Information

## 3.3 Extension to Multivariate Similarity Measures

In this section, we try to extend the above used MLE to multiple images to derive multivariate similarity measures. The set of images is denoted by  $\mathcal{U} = \{u_1, \dots, u_n\}$  and the set of transformations by  $\mathcal{T} = \{T_1, \dots, T_n\}$ . Like argued in section 2.5.1, we deal with  $n$  and not  $n - 1$  transformations. To simplify the notation like in the bivariate case, we frequently write  $u_i^\downarrow(x_k)$  instead of  $u_i(T_i(x_k))$ . The grid in the common reference frame, where all the volumes are mapped to, is  $\Omega$ .

### 3.3.1 Summed-up bivariate extension

One way to extend the MLE framework is to try to generalize the idea from the bivariate scenario to which extent one image describes the other one. In terms of probabilities this was expressed by the conditional probability  $P(u|v^\downarrow)$ . For the multivariate case we make the following deduction by assuming conditional independent images:

$$\begin{aligned}
-\mathcal{L}(T) &= -P(u_1^\downarrow | u_2^\downarrow, \dots, u_n^\downarrow, \vec{f}, \vec{\varepsilon}) \\
&= -P(\varepsilon_2 = u_1^\downarrow - f_2(u_2^\downarrow), \varepsilon_3 = u_1^\downarrow - f_3(u_3^\downarrow), \dots, \varepsilon_n = u_1^\downarrow - f_n(u_n^\downarrow)) \\
&= -\prod_{i=2}^n P(\varepsilon_i = u_1^\downarrow - f_i(u_i^\downarrow)) \\
&= -\prod_{i=2}^n \prod_{x_k \in \Omega} P(\varepsilon_i = u_1^\downarrow(x_k) - f_i(u_i^\downarrow(x_k)))
\end{aligned}$$

with intensity mappings  $\vec{f} = (f_2, \dots, f_n)$  and Gaussian noises  $\vec{\varepsilon} = (\varepsilon_2, \dots, \varepsilon_n)$ . The negative log likelihood is

$$\begin{aligned}
-\log(\mathcal{L}(T)) &= -\log \prod_{i=2}^n \prod_{x_k \in \Omega} P(\varepsilon_i = u_1^\downarrow(x_k) - f_i(u_i^\downarrow(x_k))) \\
&= -\sum_{i=2}^n \sum_{x_k \in \Omega} \log P(\varepsilon_i = u_1^\downarrow(x_k) - f_i(u_i^\downarrow(x_k))).
\end{aligned}$$

In this equation, each summand corresponds to the bivariate formula in equation 3.1 and the deduction of the four similarity measures can therefore be done analogously to the bivariate deduction in section 3.2. Finally, we end up with a multivariate extension of similarity measures that sums up the bivariate ones. This approach is very interesting for the semi-simultaneous registration because, although considering all available images, we put one in the focus (above  $u_1$ ). That is the same way like it is done in the semi-simultaneous registration by optimizing the pose parameters of one image at a time. Partial compounding can hereby be used for speeding up the registration, see section 3.5.2.

During our experiments we found out, that the extension of the measures can be improved by weighting the summed-up pairs by the amount of overlap between the involved images. Like in section 2.4.2 about the complete pairwise registration, the amount overlap between the images  $u_i$  and  $u_j$  is modeled by  $\omega_{i,j}$ , emphasizing the influence of pairs with high overlap. This leads to the slightly changed formula

$$-\log(\mathcal{L}(T)) = -\sum_{i=2}^n \omega_{1,i} \sum_{x_k \in \Omega} \log P(\varepsilon_i = u_1^\downarrow(x_k) - f_i(u_i^\downarrow(x_k))).$$

### 3.3.2 Joint density function

The approach above has the disadvantage that the resulting similarity measures is not full-simultaneous because we concentrate on one image at a time. Another approach that gets

rid of this problem is to consider joint probabilities [74]

$$\begin{aligned} -\log(\mathcal{L}(T)) &= -\log P(u_1^\downarrow, u_2^\downarrow, \dots, u_n^\downarrow) \\ &= -\sum_{x_k \in \Omega} \log P(u_1^\downarrow(x_k), u_2^\downarrow(x_k), \dots, u_n^\downarrow(x_k)). \end{aligned}$$

In this scenario no reference image is chosen, instead, all images are treated equally. This measure was used by Studholme et al. [58] under the name *average self-information*, as it builds up on

$$\log P(u_1(T_1(x)), u_2(T_2(x)), \dots, u_n(T_n(x)))$$

what is referred to as *self-information* by Reza [51]. A drawback of this approach is that it requires the construction of a high dimensional joint density function. This leads to two severe disadvantages. First, the size of the density function grows exponentially with the number of input images. In contrast to that, the amount of data available only grows linearly, but the number of samples required for a good density estimation grows exponentially [75]. Second, the joint density function is not flexible to changing numbers of overlapping images. This is especially problematic for mosaicing, where we do not have large regions with constant numbers of overlapping images. Instead, we have regions with changing numbers of overlapping images; some images do not even overlap at all. Taking an imaging situation with 5 images and constructing, e.g., a 3-variate pdf, only regions where three images are overlapping are considered. This is not satisfying for mosaicing and will finally lead to only two groups of extensions for similarity measures: *summed-up bivariate* ones and *voxel-wise* ones.

### 3.3.3 Voxel-wise measures

A third extension are *voxel-wise* measures. They focus on a voxel position of the grid at a time and make an estimation based on the values at this location. The similarity between the images is obtained by estimating the similarity at each location  $x_k \in \Omega$ . For 2D images, one can think of putting all the images on a stack and then piercing through that stack at a certain location, shown in figure 3.1. The idea of using the pixel stack was first described by Miller [36]. Since not all  $n$  images are overlapping at each point, we use the index function  $\iota : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$  to identify the  $m_k$  overlapping ones. The principle of the voxel-wise measures is to get a similarity measure for the whole imaging scenario based on the voxel-wise ones. We can integrate this into our MLE by assuming independent but not identical distributed coordinate samples. Therefore, for each coordinate location  $x_k$ , a different distribution  $P^k$  has to be estimated:

$$\begin{aligned} -\log(\mathcal{L}(T)) &= -\log P(u_1^\downarrow, u_2^\downarrow, \dots, u_n^\downarrow) \\ &= -\sum_{x_k \in \Omega} \log P^k(u_{\iota(1)}^\downarrow(x_k), u_{\iota(2)}^\downarrow(x_k), \dots, u_{\iota(m_k)}^\downarrow(x_k)). \end{aligned}$$

During our experiments, we found out that an additional weighting of the voxel-locations, emphasizing such locations with a high number of overlapping images, further increases the performance of voxel-wise measures. Let  $\omega_k$  be a coefficient that characterizes the

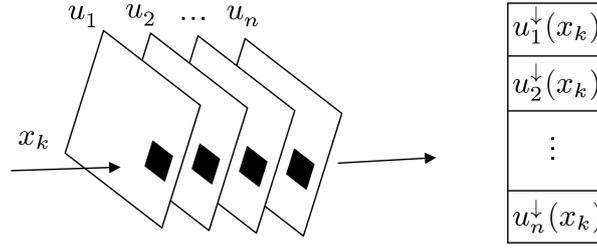


Figure 3.1: The pixel stack considers in all transformed images the values at one certain location  $x_k$ .

number of overlapping images, the slightly different negative log likelihood function is

$$-\log(\mathcal{L}(T)) = - \sum_{x_k \in \Omega} \omega_k \cdot \log P^k(u_{i(1)}^\downarrow(x_k), u_{i(2)}^\downarrow(x_k), \dots, u_{i(m_k)}^\downarrow(x_k)).$$

A possible assignment for the coefficient is  $\omega_k = \frac{m_k}{n}$ . It exists the possibility to extend the considered voxels to the neighborhood of a certain location. This can be especially beneficial for the density estimation of locations, where only a few images are overlapping, to make a superior density estimation. Voxel-wise measures have the advantages that they offer enough flexibility to deal with varying numbers of overlapping images and no joint pdf has to be estimated.

### 3.4 Detail analysis of multivariate measures

In this section, we consider the similarity measures one after another in detail. The bivariate extension is for all the measures pretty similar, therefore we will describe it here in general and just mention it shortly in the specific sections.

Of great value for the multivariate measures is the *similarity matrix*  $M$  which has a similar structure like the covariance matrix:

$$M = \begin{pmatrix} SM(u_1, u_1) & SM(u_1, u_2) & \cdots & SM(u_1, u_n) \\ SM(u_2, u_1) & SM(u_2, u_2) & \cdots & SM(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ SM(u_n, u_1) & SM(u_n, u_2) & \cdots & SM(u_n, u_n) \end{pmatrix}.$$

Since most of the measures are symmetric and reflexive it is sufficient to consider the right

upper triangular part

$$M^\nabla = \begin{pmatrix} SM(u_1, u_2) & SM(u_1, u_3) & \cdots & SM(u_1, u_n) \\ & SM(u_2, u_3) & \cdots & SM(u_2, u_n) \\ & & \ddots & \vdots \\ 0 & & & SM(u_{n-1}, u_n) \end{pmatrix}.$$

The multivariate extension, by summing up the pairwise similarity measures that we described in section 3.3.1, can also be seen as summing over the first line in the similarity matrix

$$SM(u_1|u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \cdot SM(u_1|u_i). \quad (3.7)$$

Like already mentioned, this works good for the semi-simultaneous case, but not for the full-simultaneous one. To deal with this case we sum up all the entries of the upper triangular matrix  $M^\nabla$

$$\begin{aligned} SM(u_1, u_2, \dots, u_n) &= \sum_{i=2}^n \sum_{j=i}^n \omega_{i,j} \cdot SM(u_i|u_j) \\ &= \sum_{i<j} \omega_{i,j} \cdot SM(u_i|u_j). \end{aligned}$$

with the weights  $\omega_{i,j}$ . Having these preparations, we will list in the subsequent paragraphs for each similarity measure how the multivariate extensions for the semi- and full-simultaneous case looks like and also mention the associated voxel-wise measure.

### 3.4.1 Sum of Squared Differences

The formula for the bivariate SSD is

$$SSD(u, v) = \mathbb{E} [(u - v^\downarrow)^2].$$

The extension that we use for the semi-simultaneous registration is

$$SSD(u_1|u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \cdot \mathbb{E} [(u_1 - u_i^\downarrow)^2].$$

The bivariate extension for the full-simultaneous is

$$SSD(u_1, \dots, u_n) = \sum_{i<j} \omega_{i,j} \cdot \mathbb{E} [(u_i^\downarrow - u_j^\downarrow)^2].$$

Summing just over the upper triangular part of the similarity matrix is fine since the metric is symmetric and reflexive.

For the voxel-wise measure we assume, like for the congealing later on, independent but not identical distributed coordinate samples. Therefore, for each coordinate location  $x_k$ , a different distribution  $P^k$  has to be estimated. The only difference, with regard to the concealing, lies in the assumption of having a Gaussian distribution, like we already did it for the deduction of the bivariate similarity measures. Here, the mean  $\mu_k$  and variance  $\sigma_k^2$  of the distribution depend on the voxel location. The intensity mapping is again  $\vec{f} = \{f_1, \dots, f_n\}$ . The MLE is as follows

$$\begin{aligned}
-\log(\mathcal{L}(T)) &= -\frac{1}{N} \log P(u_1^\downarrow, \dots, u_n^\downarrow) \\
&= -\frac{1}{N} \log \prod_{x_k \in \Omega} P^k(u_1^\downarrow(x_k), \dots, u_n^\downarrow(x_k)) \\
&= -\frac{1}{N} \sum_{x_k \in \Omega} \log P^k(u_1^\downarrow(x_k), \dots, u_n^\downarrow(x_k)) \\
&\approx -\frac{1}{N} \sum_{x_k \in \Omega} \log \prod_{i=1}^n P^k(u_i^\downarrow(x_k)) \\
&= -\frac{1}{N} \sum_{x_k \in \Omega} \sum_{i=1}^n \log \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(f_i(u_i^\downarrow(x_k)) - \mu_k)^2}{\sigma_k^2}} \right) \\
&= -\frac{1}{N} \sum_{x_k \in \Omega} \sum_{i=1}^n \left( \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2} \frac{(f_i(u_i^\downarrow(x_k)) - \mu_k)^2}{\sigma_k^2} \right) \\
&\approx \frac{1}{N} \sum_{x_k \in \Omega} \frac{1}{\sigma_k^2} \sum_{i=1}^n (f_i(u_i^\downarrow(x_k)) - \mu_k)^2
\end{aligned}$$

with

$$\mu_k = \mathbb{E}_i [f_i(u_i(x_k))].$$

Analogous to the deduction of the pairwise SSD we assume the identity as intensity mapping, leading to

$$-\log(\mathcal{L}(T)) \approx \frac{1}{N} \sum_{x_k \in \Omega} \frac{1}{\sigma_k^2} \sum_{i=1}^n (u_i^\downarrow(x_k) - \mu_k)^2.$$

This corresponds to summing up the variances at each voxel location leading to the name *sum of voxel-wise variances* (SVV). In practice, we do not have  $n$  images overlapping at each location but  $m_k$ , resulting in

$$\begin{aligned}
\text{SVV}(u_1, \dots, u_n) &= \frac{1}{N} \sum_{x_k \in \Omega} \frac{1}{m_k} \sum_{i=1}^{m_k} (u_{\iota(i)}^\downarrow(x_k) - \mu_k)^2 \\
&\approx \sum_{x_k \in \Omega} \omega_k \cdot \mathbb{E}_i \left[ (u_{\iota(i)}^\downarrow(x_k) - \mu_k)^2 \right]
\end{aligned}$$

with the weighting coefficient  $\omega_k$ ,  $\iota$  the index function, and the voxel wise mean

$$\mu_k = \frac{1}{m_k} \sum_{i=1}^{m_k} u_{\iota(i)}^\downarrow(x_k).$$

### 3.4.2 Sum of Absolute Differences

The extension of the SAD is analogous to the SSD so that we just state the formulas

$$\text{SAD}(u, v) = \mathbb{E} [|u - v^\downarrow|].$$

Semi-simultaneous:

$$\text{SAD}(u_1 | u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \cdot \mathbb{E} [|u_1 - u_i^\downarrow|].$$

Full-simultaneous:

$$\text{SAD}(u_1, \dots, u_n) = \sum_{i < j} \omega_{i,j} \cdot \mathbb{E} [|u_i^\downarrow - u_j^\downarrow|].$$

Voxel-wise:

$$\text{SAD}(u_1, \dots, u_n) = \sum_{x_k \in \Omega} \omega_k \cdot \mathbb{E}_i [|u_{i(i)}^\downarrow(x_k) - \mu_k|]$$

### 3.4.3 Normalized Cross Correlation

The bivariate normalized cross correlation (NCC) is calculated by

$$\text{NCC}(u, v) = \frac{1}{N} \sum_{x_k \in \Omega_u} \frac{(u(x_k) - \mathbb{E}[u(X)])(v(x_k^\downarrow) - \mathbb{E}[v(X)])}{\text{Var}(u(X)) \cdot \text{Var}(v(X))}.$$

Taking normalized images

$$\tilde{u}(x) = \frac{u(x) - \mathbb{E}[u(X)]}{\text{Var}(u(X))}$$

we ease the notation to

$$\text{NCC}(u, v) = \mathbb{E}[\tilde{u} \cdot \tilde{v}^\downarrow].$$

Semi-simultaneous:

$$\text{NCC}(u_1 | u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \cdot \mathbb{E}[\tilde{u}_1 \cdot \tilde{u}_i^\downarrow].$$

Full-simultaneous:

$$\text{NCC}(u_1, \dots, u_n) = \sum_{i < j} \omega_{i,j} \cdot \mathbb{E}[\tilde{u}_i^\downarrow \cdot \tilde{u}_j^\downarrow].$$

We also use a voxel-wise extension of NCC, which generalizes the basic idea of NCC by multiplying the values at a specific location. Unfortunately, we cannot state a deduction from our MLE framework for this extension.

$$\text{NCC}(u_1, \dots, u_n) = \sum_{x_k \in \Omega} (\omega_k \cdot \tilde{u}_{i(1)}^\downarrow \cdot \tilde{u}_{i(2)}^\downarrow \cdots \tilde{u}_{i(m)}^\downarrow).$$

### 3.4.4 Correlation Ratio

The correlation ratio is defined by

$$\text{CR}(u|v) = 1 - \frac{\text{Var}(u - \hat{f}(v^\downarrow))}{\text{Var}(u)}$$

with  $\hat{f}$  the least-squares approximation of  $u$  in terms of  $v$ . This is uniquely solved by the conditional expectation [38] of  $u$  with respect to  $v$

$$\begin{aligned} \text{CR}(u|v) &= 1 - \frac{\text{Var}[u - \mathbb{E}[u|v^\downarrow]]}{\text{Var}(u)} \\ &= \frac{\text{Var}[\mathbb{E}(u|v^\downarrow)]}{\text{Var}(u)}. \end{aligned}$$

When using the correlation ratio one has to choose between four different realizations of the measure [53]. These four cases arise from the asymmetry of the measure ( $\text{CR}(u, v) \neq \text{CR}(v, u)$ ). And the question, important for each similarity measure, of which image to choose as reference and floating image ( $\text{CR}(u, v^\downarrow) \neq \text{CR}(u^\downarrow, v)$ ). The choice of the order of the images should depend on which image serves as a good model for the other one. The floating image should be the one with the higher resolution to guarantee a high quality of the interpolated image. For our experiments, all the images had the same resolution and also their modeling capabilities were similar so that there was no significant difference between the four choices.

The semi-simultaneous extension is as usual

$$\text{CR}(u_1|u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \frac{\text{Var}[\mathbb{E}(u_1|u_i^\downarrow)]}{\text{Var}(u_1)}.$$

For the full-simultaneous extension it would normally be necessary to sum up the upper and lower triangular part of the similarity matrix because the measure is asymmetric. But our experiments show much better for summing up just the upper triangular part:

$$\text{CR}(u_1, \dots, u_n) = \sum_{i < j} \omega_{i,j} \frac{\text{Var}[\mathbb{E}(u_i^\downarrow|u_j^\downarrow)]}{\text{Var}(u_i^\downarrow)}.$$

In our search for another extension of the measure to multiple images, we first give some more details about the structure for the bivariate case:

$$\begin{aligned} \text{Var}[\mathbb{E}[Y|X]] &= \text{Var}[Y] - \mathbb{E}[\text{Var}[Y|X]] \\ &= \text{Var}[Y] - \mathbb{E}[\mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2] \\ &= \text{Var}[Y] - \mathbb{E}[Y^2] + \mathbb{E}[\mathbb{E}[Y|X]^2] \\ &= \text{Var}[Y] - \mathbb{E}[Y^2] + \sum_x \left( \sum_y y \cdot P(Y = y|X = x) \right)^2 P(X = x) \end{aligned}$$

by using the property that

$$\mathbb{E}[\mathbb{E}[Y^2|X]] = \mathbb{E}[Y^2].$$

This property holds because of

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y^2|X]] &= \sum_x \mathbb{E}[Y^2|X=x] \cdot P(X=x) \\
&= \sum_x \left( \sum_y y^2 \cdot P(Y=y|X=x) \right) P(X=x) \\
&= \sum_x \sum_y y^2 \cdot P(X=x|Y=y) \cdot P(Y=y) \\
&= \sum_y y^2 \cdot P(Y=y) \cdot \left( \sum_x P(X=x|Y=y) \right) \\
&= \sum_y y^2 \cdot P(Y=y) \\
&= \mathbb{E}[Y^2].
\end{aligned}$$

A voxel-wise approach for getting a multivariate measure, as we did for the others, is not possible for the correlation ratio. The design concept of the CR is to analyze how well one image is predicted by another one. When we have a couple of voxels on a certain location we would have to choose one as reference, what is not desired. In the following, we list two ideas, how extensions for the correlation ratio may look like. A first idea is to start with

$$\begin{aligned}
\text{CR}(u_1|u_2, \dots, u_n) &= \frac{\text{Var}[\mathbb{E}[u_1|u_2, \dots, u_n]]}{\text{Var}(u_1)} \\
&= \frac{\text{Var}[u_1] - \mathbb{E}[u_1^2] + \mathbb{E}[\mathbb{E}[u_1|u_2, \dots, u_n]^2]}{\text{Var}(u_1)} \\
&= 1 - \frac{\mathbb{E}[u_1^2]}{\text{Var}(u_1)} + \frac{\mathbb{E}[\mathbb{E}[u_1|u_2, \dots, u_n]^2]}{\text{Var}(u_1)}
\end{aligned}$$

further

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X_1, \dots, X_n]^2] &= \sum_{x_1} \dots \sum_{x_n} \left( \sum_y y \cdot P(Y=y|X_1=x_1, \dots, X_n=x_n) \right)^2 \\
&\quad P(X_1=x_1, \dots, X_n=x_n) \\
&= \sum_{x_1} \dots \sum_{x_n} \left( \sum_y y \cdot P(Y=y|X_1=x_1) \dots P(Y=y|X_n=x_n) \right)^2 \\
&\quad P(X_1=x_1) \dots P(X_n=x_n) \\
&\approx \sum_{x_1} \left( \sum_y y \cdot P(Y=y|X_1=x_1) \right)^2 \cdot P(X_1=x_1) + \dots + \\
&\quad \sum_{x_n} \left( \sum_y y \cdot P(Y=y|X_n=x_n) \right)^2 \cdot P(X_n=x_n) \\
&= \mathbb{E}[\mathbb{E}[Y|X_1]^2] + \dots + \mathbb{E}[\mathbb{E}[Y|X_n]^2]
\end{aligned}$$


---

leading to

$$\begin{aligned}
\text{CR}(u_1|u_2, \dots, u_n)_{idea1} &= \frac{\text{Var}[\mathbb{E}[u_1|u_2, \dots, u_n]]}{\text{Var}(u_1)} \\
&\approx 1 - \frac{\mathbb{E}[u_1^2]}{\text{Var}(u_1)} + \frac{\mathbb{E}[\mathbb{E}[u_1|u_2]^2] + \dots + \mathbb{E}[\mathbb{E}[u_1|u_n]^2]}{\text{Var}(u_1)} \\
&= 1 - \frac{\mathbb{E}[u_1^2]}{\text{Var}(u_1)} + \frac{\sum_{i=2}^n \mathbb{E}[\mathbb{E}[u_1|u_i]^2]}{\text{Var}(u_1)}
\end{aligned}$$

A second idea for an extension

$$\begin{aligned}
\text{CR}(u_1|u_2, \dots, u_n)_{idea2} &= \frac{\text{Var}[\mathbb{E}(u_1|u_2, \dots, u_n)]}{\text{Var}(u_1)} \\
&= \frac{\text{Var}\left[\frac{1}{N} \sum_{x_k} i_k \cdot P(i_k|u_2, \dots, u_n)\right]}{\text{Var}(u_1)} \\
&= \frac{\text{Var}\left[\frac{1}{N} \sum_{x_k} \prod_{j=2}^n i_k \cdot P(i_k|u_j)\right]}{\text{Var}(u_1)} \\
&\approx \frac{\text{Var}\left[\prod_{j=2}^n \frac{1}{N} \sum_{x_k} i_k \cdot P(i_k|u_j)\right]}{\text{Var}(u_1)} \\
&= \frac{\text{Var}\left[\prod_{j=2}^n \mathbb{E}[u_1|u_j]\right]}{\text{Var}(u_1)} \\
&= \frac{\prod_{j=2}^n \text{Var}[\mathbb{E}[u_1|u_j]]}{\text{Var}(u_1)}
\end{aligned}$$

### 3.4.5 Mutual Information

Mutual information became, since its introduction in 1995, one of the most popular similarity measures for medical image registration [48].

#### 3.4.5.1 Entropy

$$\begin{aligned}
H(X) &= - \sum_{x \in X} P(x) \cdot \log P(x) \\
H(X_1, \dots, X_n) &= - \sum_{x_1 \in X_1, \dots, x_n \in X_n} P(x_1, \dots, x_n) \cdot \log P(x_1, \dots, x_n)
\end{aligned}$$

#### 3.4.5.2 Bivariate

Mutual Information for two images:

$$\text{MI}(u, v) = H(u) + H(v) - H(u, v).$$

An overlap invariant similarity measure, the normalized mutual information, was introduced by Studholme et al. [60]

$$\text{NMI}(u, v) = \frac{H(u) + H(v)}{H(u, v)}.$$

### 3.4.5.3 Multivariate

In contrast to the other measures, a couple of articles can be found in the literature about the theoretical extension of mutual information [62, 57, 34]. Practically, it was already used for the simultaneous registration of 3 images [4, 59, 32, 73]. The formula for the generalized mutual information is

$$\text{MI}_n(u_1, u_2, \dots, u_n) = \sum_{k=1}^N (-1)^{k-1} \sum_{i_1 < \dots < i_k} H(u_{i_1}, \dots, u_{i_k}).$$

Since this formula is a bit hard to understand we give an example for  $n = 3$  to have a better idea of its structure

$$\text{MI}_3(u_1, u_2, u_3) = H(u_1) + H(u_2) + H(u_3) - H(u_1, u_2) - H(u_1, u_3) - H(u_2, u_3) + H(u_1, u_2, u_3)$$

Using this extension leads to problems that were already addressed in section 3.3.2. First, the missing flexibility to deal with varying numbers of overlapping images. Second, the problems for estimating multivariate probability density functions for larger  $n$ . According to [20] it is still tractable for small  $n$ , without further quantifying  $n$ .

Frequently, an extension is used that was introduced by Watanabe [69] under the name *total correlation*. It expresses the amount of redundancy or dependency (and thus structure) existing among a set of variables

$$C(u_1, u_2, \dots, u_n) = \sum_{i=1}^n H(u_i) - H(u_1, u_2, \dots, u_n).$$

But despite the easier structure of this formula in comparison to the mathematical correct extension, it poses the same problems.

An approximation for mutual information was proposed by Kern et al. [24] that is pretty much the same as what we called the full-simultaneous summed-up bivariate extension, by summing up the upper triangular part of the similarity matrix. Instead, Kern et al. sum up the upper and lower triangular part, but because of the symmetry of mutual information this is redundant:

$$\text{MI}_{app}(u_1, u_2, \dots, u_n) = \sum_{i,j \in \{1, \dots, n\}, i \neq j} \text{MI}(u_i, u_j). \quad (3.8)$$

For the sake of completeness we also mention the semi-simultaneous bivariate extension

$$\text{MI}(u_1 | u_2, \dots, u_n) = \sum_{i=2}^n \omega_{1,i} \cdot \text{MI}(u_1, u_i).$$

### 3.4.5.4 Stochastic Congealing

For the voxel-wise extension of mutual information we use stochastic congealing. The idea of congealing was first introduced in the machine learning and vision literature [36]. Zoellei modified the original framework to let it work with medical data and added an efficient stochastic optimization, leading to the name stochastic congealing [74].

The assumption by congealing is to assume independent but not identical distributed coordinate samples. Therefore, for each coordinate location  $x_k$ , a different distribution  $P^k$  has to be estimated. If we further assume that the images are independently and identically distributed the log likelihood can be approximated by

$$\begin{aligned}
 \log(\mathcal{L}(T)) &= \frac{1}{N} \log P(u_1^\downarrow, \dots, u_n^\downarrow) \\
 &= \frac{1}{N} \log \prod_{x_k \in \Omega} P^k(u_1^\downarrow(x_k), \dots, u_n^\downarrow(x_k)) \\
 &= \frac{1}{N} \sum_{x_k \in \Omega} \log P^k(u_1^\downarrow(x_k), \dots, u_n^\downarrow(x_k)) \\
 &\approx \frac{1}{N} \sum_{x_k \in \Omega} \log \prod_{i=1}^n P^k(u_i^\downarrow(x_k)) \\
 &= \frac{1}{N} \sum_{x_k \in \Omega} \sum_{i=1}^n \log P^k(u_i^\downarrow(x_k)).
 \end{aligned}$$

The inner sum is closely related to the sample entropy leading to

$$\log(\mathcal{L}(T)) \approx \sum_{x_k \in \Omega} H(P^k).$$

The MLE can therefore be approximated by the *sum of voxel-wise entropies*. In practice, this measure is used within a stochastic optimization framework and not all locations are summed up, but only certain number of randomly selected ones. Depending on how many percent of the locations we evaluate, this leads to an enormous speed up. For more information see section 3.5.3.

The congealing has tremendous advantages for the mosaicing. First, it offers enough flexibility to deal with varying numbers of overlapping images. And second, no joint pdf has to be estimated, instead, one-dimensional distributions are estimated that are fast to calculate, even for a growing number of input images. Problems can arise when having not enough voxels at a certain location to make reliable estimations. Here, also the voxels in the neighborhood can be taken into consideration, the so called cylinder.

### 3.4.6 Conclusion

In table 3.1, we summarize the previous paragraphs and state the extensions for the similarity measures that we are working with. In order to have the table clearly arranged, we relax the mathematical strictness.

## 3.5 Speeding up the similarity calculation

The gain of quality, when using multivariate similarity measures, comes along with a rise in computational complexity. In order to keep the registration time still at a reasonable level, we introduce methods that speed up the registration.

---

	Pairwise	Semi-Simultaneous	Full-Simultaneous	Voxel-Wise
SSD	$\mathbb{E}[(u - v^\downarrow)^2]$	$\sum_{i=2}^n \omega_{1,i} \mathbb{E}[(u_1 - u_i^\downarrow)^2]$	$\sum_{i < j} \omega_{i,j} \mathbb{E}[(u_i^\downarrow - u_j^\downarrow)^2]$	$\sum_{x_k \in \Omega} \omega_k \mathbb{E}_i[(\mu_k - u_i^\downarrow(x_k))^2]$
NCC	$\mathbb{E}[\tilde{u} \cdot \tilde{v}^\downarrow]$	$\sum_{i=2}^n \omega_{1,i} \mathbb{E}[\tilde{u}_1 \cdot \tilde{u}_i^\downarrow]$	$\sum_{i < j} \omega_{i,j} \mathbb{E}[\tilde{u}_i^\downarrow \cdot \tilde{u}_j^\downarrow]$	$\sum_{x_k \in \Omega} (\omega_k \cdot \tilde{u}_1^\downarrow \cdot \tilde{u}_2^\downarrow \cdots \tilde{u}_n^\downarrow)$
CR	$\frac{\text{Var}[\mathbb{E}(u v^\downarrow)]}{\text{Var}(u)}$	$\sum_{i=2}^n \omega_{1,i} \frac{\text{Var}[\mathbb{E}(u_1 u_i^\downarrow)]}{\text{Var}(u_1)}$	$\sum_{i < j} \omega_{i,j} \frac{\text{Var}[\mathbb{E}(u_i^\downarrow u_j^\downarrow)]}{\text{Var}(u_i^\downarrow)}$	-
MI	$\text{MI}(u, v^\downarrow)$	$\sum_{i=2}^n \omega_{1,i} \text{MI}(u_1, u_i^\downarrow)$	$\sum_{i < j} \omega_{i,j} \text{MI}(u_i^\downarrow, u_j^\downarrow)$	$\sum_{x_k \in \Omega} \omega_k \text{H}(P^k)$

Table 3.1: Summary of bi- and multivariate similarity measures in shortened notation.

### 3.5.1 Region of overlap

For mosaicing, the target volume is normally much larger than the input volumes. For the calculation of the similarity, we have to iterate over each voxel in the target volume. Since the grid of the target volume is quite large, this iteration has a high computational cost. Increasing the speed of this iteration, has a strong influence on the overall performance of the registration. Depending on the imaging scenario, large regions of the target volume may just be covered by one image, where no similarity metric can be evaluated, see figure 3.2. The definition of a region of overlap (ROO), to which the iteration is limited, can therefore speed up the calculation. Since we are interested in keeping the estimation of the ROO simple, we approximate it by a cuboid.

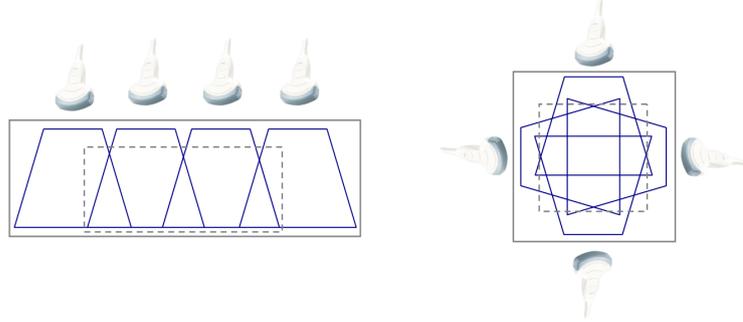


Figure 3.2: The target volume (solid line) and the region of overlap (dashed line) are drawn for two imaging scenarios. Iterating over the region of overlap is sufficient for the similarity estimation.

### 3.5.2 Partial Compounding

For the semi-simultaneous registration, a technique, which we refer to as *partial compounding*, can be used to gain on registration speed. In the semi case, only one image is moving

at a time while the other  $n - 1$  images stay fixed. These  $n - 1$  images are temporary compounded so that we end up with two images that are used for the registration. For the construction of the compounded volume two different strategies exist. First, to really reduce the possible  $n - 1$  voxel values at a certain location to one value, leading to the usage of bivariate similarity measures. Second, to create a volume where up to  $n - 1$  values can be saved at each voxel location; allowing for the usage of multivariate similarity measures.

The creation of the compounded volume takes some time, so that it only makes sense for a high number of registration steps  $k$ , see table 2.2. Since for the semi-simultaneous approach it suffices to limit the similarity measure evaluation to its grid, also the compounding can be limited to that part. Leading to a smaller partial compounded volume.

### 3.5.3 Stochastic similarity measures

*Stochastic similarity measures* are evaluated at a randomly picked subset of the grid  $\Omega_{rand} \subset \Omega$ . They break with usual approach of getting a similarity estimation by evaluating the similarity on all locations. Stochastic similarity measures have to be used with a stochastic optimization framework, which is able to deal with its variations. Zoellei [74] used this technique in combination with the congealing and had good results by only selecting 0.05% - 2.5% of the total voxels. The gained speed up is enormous. For mosaicing we have to choose slightly higher values because the amount of overlap is generally low. Additionally, its implementation is easy and it can be used for most of the similarity measures.

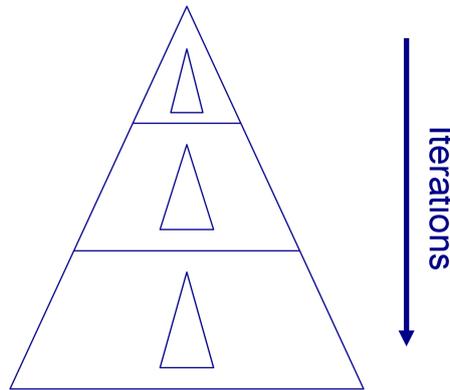


Figure 3.3: The Gaussian pyramid (large) combined with the stochastic pyramids (small) can lead to an enormous speed-up of the registration.

We propose the combination of a progressive stochastic approach with a multi-resolution framework. The multi-resolution framework uses the so called Gaussian pyramid that is described for instance in [21]. At each level of the pyramid a low-pass filtered and down-sampled version of the original image is used. The registration is started on the highest level of the pyramid corresponding to the images with the lowest resolution. When stepping down the levels of the pyramid, the registration result of the previous level serves as initial value. Using this method has several advantages. First, it speeds up the registration because good estimations are already achieved at the coarse levels, so less optimization

steps are necessary on the high resolution images. Second, it avoids the registration to become too easily trapped in local minima [6]. And third, ultrasound images are corrupted by speckle noise which can be reduced by Gaussian filtering, see section 5.3.2.2. The number of used levels varies depending on the resolution of the images, but three levels is a good reference value.

We add to the idea of the Gaussian pyramid the usage of stochastic similarity measures. We use a progressive version of the stochastic measures meaning that the percentage of considered voxels increases during the registration. This can be seen as a *stochastic pyramid* in contrast to the Gaussian one. Replacing the Gaussian pyramid with the stochastic one does not work because we would lose the advantages coming from the filtering. Therefore, we combine the two approaches. At each level of the Gaussian pyramid we start with a low percentage of considered voxels and increase it progressively, see figure 3.3. In contrast to the Gaussian pyramid, where it takes some computational effort to create the images for each level, the stochastic pyramid comes without computational cost. The relatively fine-grained approach of the stochastic pyramid, forms therefore a good complement to the coarse-grained Gaussian pyramid.

---

# Chapter 4

## Reconstruction

*In the previous chapters, various strategies were described, focusing on how the global alignment of the input images can be calculated. In this chapter, we will concentrate on the last part that is still missing for the creation of mosaics, the reconstruction. There are two different demands on the reconstruction that are slightly different. First, the creation of an extended field-of-view. Here, the focus is on stitching several images together to enlarge the captured area with a minimal amount of overlap in-between the images. Second, the improvement of quality in the final image. Here, acquisitions of the same part of the body are taken from different perspectives, what is also known as spatial compounding. In the following sections, we describe general and advanced methods for the reconstruction. More than for the registration, the right strategy for the reconstruction depends to a large extent on the imaging scenario so that specific solutions for specific scenarios have to be developed.*

### 4.1 General Methods

In the following, general methods for the compounding of intensity images are described. They are sufficient for the creation of extended field-of-view images and can lead to an increase of quality in the overlapping regions. But for really focusing on quality improvement, specific methods like in section 4.2 should be chosen. The following methods do not consider any neighborhood relationships. At each location  $x_k$  in the grid  $\Omega$  we suppose to have  $m$  values  $X_k = \{x_k^1, \dots, x_k^m\}$ . These values are equally weighted.

1. **Selection:** The intensity of one image  $j$  is selected at a certain location  $x_k$ .

$$\text{Sel}_k^j = x_k^j \quad (4.1)$$

2. **Mean:** The mean value over all pixels at  $x_k$  is calculated.

$$\text{Mean}_k = \mathbb{E}[X_k] \quad (4.2)$$

3. **Median:** The median value at  $x_k$  is chosen.

$$\text{Median}_k = \text{median}\{x_k^1, \dots, x_k^m\} \quad (4.3)$$

---

4. **Max:** The maximal intensity value is chosen at  $x_k$ .

$$\text{Max}_k = \max\{x_k^1, \dots, x_k^m\} \quad (4.4)$$

5. **Min:** The minimal intensity value is chosen at  $x_k$ .

$$\text{Min}_k = \min\{x_k^1, \dots, x_k^m\} \quad (4.5)$$

6. **Geometric Mean:** The geometric mean calculates the mean on log-compressed images [66] at  $x_k$ .

$$\text{Geometric Mean}_k = \sqrt[n]{x_k^1 \cdot x_k^2 \cdots x_k^m} \quad (4.6)$$

When using mean and geometric mean one has to pay attention that new intensities are created, which are not existent in the input images. Thus, the compounded volume may be augmented with artificial structures that do not correspond to real ones, and therefore seriously affect the clinical acceptance. For multiple overlapping images, we achieved good results with the maximum intensity method. It assures that any depiction of anatomy gets transferred into the final image. This is especially interesting for imaging scenarios with occlusion. But also image artifacts and noise get accumulated in the final image, leading to many confusing structures. In contrast, methods like mean and geometric mean are less sensitive to noise and artifacts but produce a blurred output. The median gets rid of outliers in individual volumes, but does not average the intensities like the mean, resulting in a less smooth volume. The selection method makes most sense in combination with prior knowledge, indicating which value to choose for a certain location.

At the reconstruction process, we are profiting from our multiple image framework in terms of interpolation. Methods, which consider mosaicing as a sequence of pairwise compounding, interpolate several times in the overlapping regions, leading to a decrease of quality. Since we are having all images at the same time available, we can use the method proposed by Wein et al. [70] for efficient backward-warping compounding. The voxel coordinates of the reconstruction volume are back-projected to the original images so that an interpolation at exactly this position, in the original image, is done. This assures a high quality compounded volume.

## 4.2 Advanced Methods

Not so much work has been done yet to explore more sophisticated compounding strategies. In this section, we will present methods that have been proposed in the literature. Leotta et al. [8] improve the compounding by incident angle weighting. They use the dependence of echo intensity on the angle of incidence of the ultrasound beam on a reflecting surface, see figure 4.1. Therefore, local weights have to be assigned depending on the local incidence angle of the ultrasound beam. This can lead to a significant enhancement of the result when comparing to the mean calculation. However, the algorithm is based on the segmentation of the surface in the data, which is challenging for complex data.

Grau et al. [16] also use an approach that assigns local weights. They use the image phase to obtain multi scale information about local structure definition and orientation to weight the contributions of the images. The image phase has the advantage that it

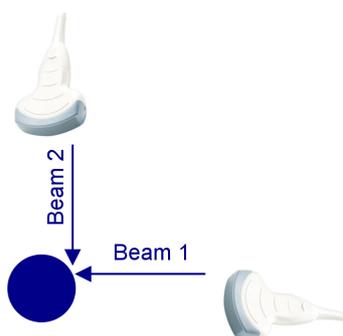


Figure 4.1: Weighting by incidence angle: for the target region the beam 1 would be given a higher weight because its incidence angle is closer to the normal.

is invariant to image contrast and brightness, what makes it particularly interesting for ultrasound. For the calculation of the local phase they used an approach proposed by Felsberg et al. [11] to calculate the phase in n-dimensions. For further details, we refer to section 6.2.1.

An algorithm that uses lateral beam steering for producing spatial compounded images was presented by Treece et al. [61]. Angular compounding, where the beam is steered laterally to various angles and the images averaged, is a frequently used technique. It increases the signal to speckle ratio and reduces the dependency of reflection from planar interfaces. However, angular compounding has the effect that artifacts are blurred in the final image. To avoid these problems Treece et al. use scans with known angles, which are used to deduce the attenuation for which they set up a specific model.

Another approach that addresses more the problem of general image quality improvement in ultrasound images than shadow reduction and surface reconstruction was proposed by Ng et al. [37]. Ultrasound images are corrupted by inherent blurring. The blurring can be characterized by a point-spread-function (psf) which models the spatial extent of the blurring of a single point. Ng et al. first use a line target to determine the psf and utilize it then to filter the images for improving their quality. Another approach is it to acquire images from known angles to make a blind deconvolution, where the first step, the estimation of the psf, is not needed anymore.



# Chapter 5

## Experiments on ultrasound images

*In this section, we will describe the experiments that we conducted and the results that we obtained. To be able to conduct the experiments and to test the developed strategies and similarity measures, we implemented a mosaicing prototype. This prototype also has a bundle of methods to pre-process the ultrasound images, which is inevitable because ultrasound images are especially prone to noise and shadow. For the experiments we use images of clay models and ultrasound phantoms. The acquisitions from an abdominal phantom are done in combination with optical tracking to have ground-truth data. To analyze the quality we show similarity measure plots and the results of random studies. The best results are obtained from the baby phantom.*

### 5.1 Research prototype

To conduct the experiments a research prototype for mosaicing was implemented with a special focus on visualization during the registration, to better understand the procedure. The platform independent prototype was written in C++ by using OpenGL<sup>®</sup> for the 3D visualization, FLTK<sup>®</sup> for building the graphical user interface, and the CAMP-Lib<sup>1</sup> for image processing. The design is based on the model-view-controller architectural pattern, to decouple the user interface from the data. Volume renderings of the images help to find the initial alignment. During the registration, the online compounded volume or a multiplanar reconstruction (MPR) slice through the volume can be visualized. Moreover, a triple view with three orthogonal slices through the fused volume helps to make a visual assessment of the registration result. Screenshots of the prototype and the triple view are shown in figures B.10 and B.11.

### 5.2 Non-linear Optimization

The two core parts that make up a registration are the similarity measure and the optimization algorithm. For an illustration of the registration framework see 5.1. During our work we focused on similarity measures, which were described in detail in section 3. The

---

<sup>1</sup>Object-oriented library developed at the Chair for Computer Aided Medical Procedures - <http://campar.in.tum.de/Chair/ResearchCamp>

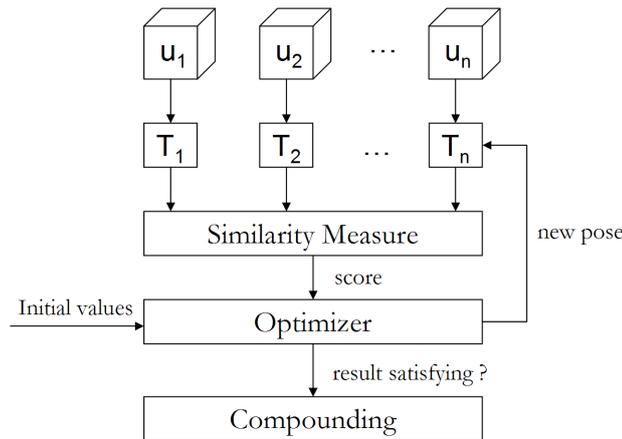


Figure 5.1: Schematic registration framework for multiple image alignment. The current alignment of the transformed images is evaluated and depending on the score the compounded volume is created or a new registration cycle with a new pose is started.

optimization, more precise the non-linear optimization, is a complex subject by itself so that we just describe the basic idea of the used optimizers, which are frequently utilized in the field of medical image processing.

### Best Neighbor Search

The best neighbor search is a simple optimization method also referred to as hill-climbing [50]. In every iteration, the pose in each of the  $n$  degrees-of-freedom is changed by a specific step size in both directions. The cost function is evaluated at these positions. After having the score of all  $2 \cdot n$  neighbors calculated, the best one is chosen and the same procedure is repeated based on this value. If the optimum corresponds to the current position, the step size is either downscaled or the search terminates, assuming that the optimum has been found.

### Downhill Simplex

The downhill simplex method [50] uses the concept of a simplex, which is a polytope of  $n + 1$  vertices in  $n$  dimensions; a line segment on a line, a triangle on a plane, a tetrahedron in three-dimensional space and so forth. A starting simplex is defined, the cost function is evaluated at the corners, and depending on the results the shape of the simplex is changed. The worst corner is replaced by a new point that is hopefully better.

### Powell's Direction Set Method

Powell's Direction Set Method starts at a given position in the parameter space, and minimizes the cost function successively along certain directions. Therefore the problem is split

up in two parts: Finding the best directions in  $n$ -dimensional space, and doing efficient line minimization on a new cost function with only one parameter. The latter problem can be solved very well with an algorithm called Brent Line Minimization. It uses both parabolic interpolation and golden section search, choosing dynamically in each step which method is more appropriate. For a detailed description we refer to [50].

## 5.3 Ultrasound Images

US images are commonly said to be "gradient images" since they enhance the interfaces between anatomical structures. The physical reason is that the amplitudes of the US echoes are proportional to the difference between acoustical impedances caused by successive tissue layers. Ideally, the US signal should be high at the interfaces, and zero within homogeneous tissues. In reality, US reflections also occur within tissues due to small inhomogeneities (compared with the US wavelength) which are almost invisible in MR. At this scale, there are significant interference patterns between the ingoing and outgoing pulsed waves, resulting in speckle. As a consequence, homogeneous tissue regions generally appear in the US with nonzero mean intensity and strong texture. During this section, we will describe how the image formation of ultrasound images works and present techniques for pre-processing them to reduce speckle and detect occlusion. Leading to an improvement of the registration and reconstruction.

### 5.3.1 Image Formation

A sound wave is typically produced by a piezoelectric transducer encased in a probe. Strong, short electrical pulses from the ultrasound machine make the transducer ring at the desired frequency. The sound is focused either by the shape of the transducer, a lens in front of the transducer, or a complex set of control pulses from the ultrasound scanner machine. The return of the sound wave to the transducer results in the same process that it took to send the sound wave, except in reverse. The return sound wave vibrates the transducer, the transducer turns the vibrations into electrical pulses that travel to the ultrasound scanner where they are processed and transformed into a digital image. The ultrasound scanner must determine three things from each received echo: First, the direction of the echo. Second, how strong the echo was. And third, how long it took the echo to be received from when the sound was transmitted. Once the ultrasound scanner determines these three things, it can locate which pixel in the image to light up and to what intensity.

### 5.3.2 Image pre-processing

Ultrasound images are prone to noise, especially speckle noise and shadow. It is necessary to use methods to reduce them because they influence the quality of the registration and reconstruction.

---

- 
1. Start with beam  $B = (b_0, \dots, b_{end})$  in image
  2.  $Iter := 0$
  3. Do
    - 2.1 Compute variance  $\sigma$  of  $(b_0, \dots, b_{Iter})$
    - 2.2  $Iter = Iter + 1$
  4. While  $((\sigma < \text{threshold}) \text{ AND } (Iter \leq \text{end}))$
  5. Use  $(b_0, \dots, b_{Iter})$
- 

Table 5.1: Algorithm for top-down variance based shadow suppression.

### 5.3.2.1 Shadow suppression

Structures with a high acoustical impedances e.g. bone lead to a total reflection of the ultrasound beam, causing occlusion of the underlying parts. In these shadow parts, we normally have values with zero intensity but also artifacts from double reflection. It is obvious that we cannot use these regions for the registration so that we have to think about methods how to single them out. One approach that is very easily integrated into the registration workflow is to set all voxels in the shadow region to zero. Since we don't want to use them for the registration, we consider only values larger than 0 for the registration. This has the disadvantage that the registration focuses on the alignment of structures and background information is not considered anymore. Another approach is to mark the shadow regions explicitly not to be considered. This can either be done by using a separate boolean volume in which for each voxel is noted if it is shadow or not, or by reserving one of the intensity values in  $\mathcal{A} = \{0, \dots, 255\}$  for that purpose. In both cases, the registration process has to be slightly adapted to this issue.

We try three different approaches for the shadow suppression. First, the intensity values along the direction of the beam heading towards the transducer face are compared to an intensity threshold. All values are neglected until the first time one value exceeds the threshold [47, 30]. Second, the variance is calculated along the beam until it exceeds a variance threshold [71]. Depending on calculating the variance top-down (TD) or bottom-up (BU), we consider the values until the threshold is reached or the values that are remaining. A sample algorithm for the top-down approach is shown in table 5.1. And third, a robust maximum search along the beam. The robustness is achieved by filtering the beam with a low-pass filter in advance.

Before we apply the shadow suppression methods the first 2cm of the top of the images were cut away. On patient data, this corresponds to the skin and is prone to intensity variations that are undesired [47]. Moreover, it leads to better rigid registration results to remove the due to probe pressure non-rigidly deformed parts.

## Results

In figure 5.2 the results of the four described approaches are illustrated. The test image is a sonographic acquisition of the human thigh. Like already mentioned before, we cut the

---

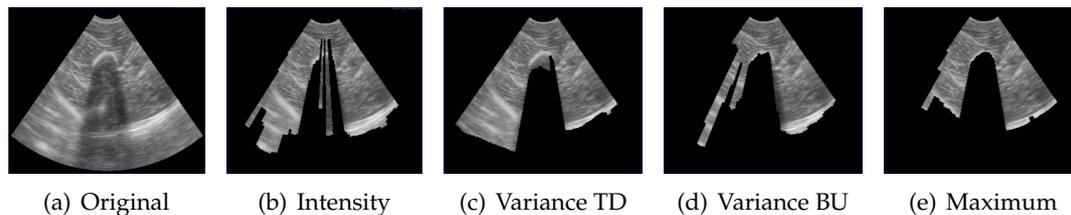


Figure 5.2: Comparison of four shadow suppression methods. For larger images see figure [B.1](#).

top part of the images. For the intensity approach we used an intensity threshold of 170. For the top-down approach we used a variance threshold of 775 and for the bottom-up approach we set it to 550. The maximum method runs without any further parameterization. For all the four approaches we used a consecutive median filtering to get a smoother borderline.

The experiments show that the intensity based suppression is not very robust and parts of the bone were also marked as shadow. The variance based approaches perform better, although they either do not cover the whole shadow region or detect too much. The best results were obtained with the robust maximum approach. The boundary corresponds exactly with the bone. But this works only if there is a significant boundary corresponding to high intensity values. Otherwise the maximum may lie anywhere in the image.

The high influence of the shadow suppression for the registration becomes apparent from the experiments on the abdominal phantom, see [5.4.2](#).

### 5.3.2.2 Speckle reduction

Sonography suffers from artifacts caused by coherent wave interference known as speckle. Speckle limits low resolution image contrast and may even obscure true structures in high contrast regions. Nevertheless, a recent work on freehand 3D ultrasound uses speckle for the registration of the slices [\[19\]](#). This requires a permanent stream of high-resolution scans from a similar viewing angle with a small spacing between slices. Hence, the same speckle patterns will be visible on successive 2D images. For the alignment of 3D volumes for mosaicing, these requirements do not hold because we do not have a permanent stream of 3D volumes but, instead, volumes from varying viewpoints. On these images, speckle patterns at the same physical location look different. Therefore, we have to pre-process the input data sets to reduce the influence of speckle noise.

A well known technique to reduce it is by low-pass filtering the image e.g. with a Gaussian filter [\[13, 47\]](#). For the reconstruction either the original or the filtered images can be compounded. When using the filtered images, the final image will be less corrupted by speckle noise but, in contrast, more blurred. On the other hand, when using the original images the influence of speckle will be reduced by averaging the images. This comes with the advantage of no additional blurring.



Figure 5.3: Imaged clay models.

## 5.4 Experiments and results

In this section, we present the experiments that we conducted and the results that we obtained. Our imaging objects are clay models and ultrasound phantoms. The best results are obtained for the baby phantom. The acquisitions are done with a Siemens Sonoline Antares<sup>®</sup> 2.2-4.7Mhz with a 3D/4D curved 1D array wobbler probe and a prototype of a 2D array probe with CMUT<sup>2</sup> technology.

### 5.4.1 Clay models

A first set of experiments is done with clay models because they produce nice sonographic images when put into a water bath. The temperature of the water was about the body temperature. This was necessary because the speed of sound depends on the temperature of the medium and ultrasound machines are calibrated for the acquisitions of body tissue. An experimental set up is shown in figure B.4.

#### Spine model

A clay model of the human spine was created<sup>3</sup> see figure 5.3(a), which is basically a long tube with some attached features. The model is too long to be visible on one image. We made 12 successive acquisitions with about 3cm displacement and stitched them together. 3D renderings of the compounded volume are shown in figures B.2 and B.3.

According to our analysis in section 2.3, for this imaging setup good registration results should be achieved with the pairwise sequential or simultaneous registration. To get a better understanding of the scene from a registration point of view, we plot the in table 3.1 proposed similarity measures, see table 5.2. For the plot, image number three was translated in the interval [-50; 50] mm along the main axis of the model. One clearly sees the high overlap dependence of the bivariate measures, being a source for misregistrations (total overlap at about -30mm displacement). In contrast to that, the multivariate

<sup>2</sup>Capacitive Micromachined Ultrasound Transducer.

<sup>3</sup>Despite lacking pottery experience of the author.

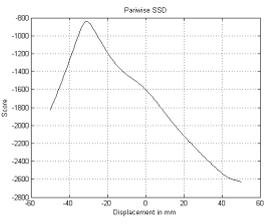
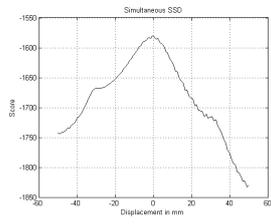
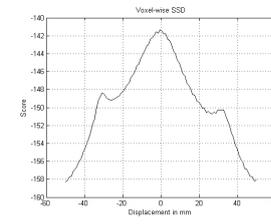
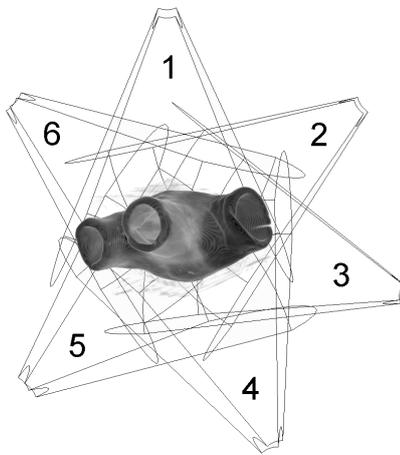
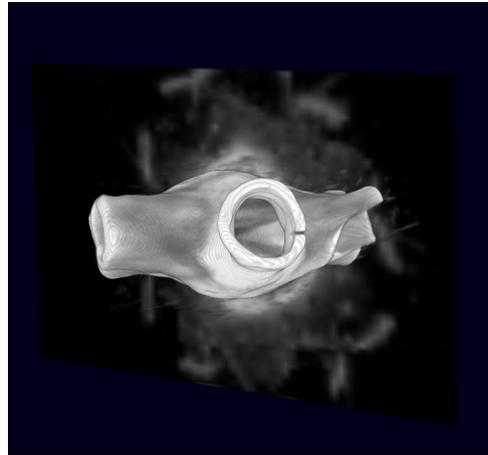
	Pairwise	Full-Simultaneous	Voxel-Wise
SSD			

Table 5.2: Excerpt of similarity plots from table B.1 for the *spine*.

(a) Imaging setup



(b) 3D Rendering with cutting plane

Figure 5.4: Rendering of CT scan from heart model augmented with US fan and cutting plane.

extensions of SSD provide a smooth cost function with a clear maximum at the correct position 0, and only a very small peak at the locations of total overlap ( $\pm 30\text{mm}$  displacement). The extensions for NCC, CR, and MI do not perform very well in this scenario. An exception is voxel-wise MI which also produces a smooth cost function. Reasons for the bad performance of the measures may lie in the only surface enhancing nature of the clay acquisitions.

#### Heart model

3D images of a heart clay model (figure 5.3(b)), in the water bath were acquired from six different angles, see figure B.4. The imaging setup is shown in figure 5.4(a). We use a cutting plane through the reconstruction volume to visualize the registration error, see figure 5.4(b). When using pairwise registration the summed up error leads to a large displacement between the first and sixth volume, figure 5.5(a). The pairwise registration with a

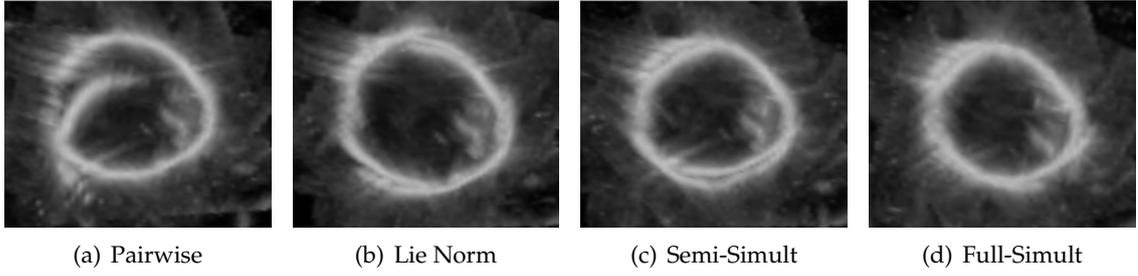


Figure 5.5: MPR slice through the compounded volume after the registration with the 4 discussed strategies. When using pairwise registration, error terms sum up, leading to large displacement that becomes visible between the first and sixth volume. Applying a subsequent Lie normalization reduces the error. Simultaneous registration intrinsically deals with these errors leading to better registration results.

successive Lie normalization corrects this error, but the alignment is not perfect, figure 5.5(b). The semi-simultaneous registration provides good results, figure 5.5(c), but superior results are obtained with the full-simultaneous registration, figure 5.5(d).

#### 5.4.2 Abdominal phantom

We acquired eight volumetric images from a multimodal abdominal phantom, see figure 5.7(a) for the 3D rendering of a CT scan. During the scanning, the pose of the transducer was measured by an accurate 6-DOF optical tracking system, see figure B.5 for the experimental setup. This enabled us to make a random registration study based on ground truth data. The accuracy of the tracking with the manual probe placement is about 2mm for translation and  $2^\circ$  for rotation.

To be able to use the tracked pose information, the calibration matrix  $T_c$ , mapping from the probe to the image coordinate system, has to be estimated. Therefore, an imaging scenario with coordinate transformations like it is shown in figure 5.6 was set up. Traversing the coordinate systems yields the following equations:

$$\begin{aligned} T_{1,2} T_c T_1 &= T_c T_2 \\ \Leftrightarrow T_{1,2} T_c &= T_c T_2 T_1^{-1} \end{aligned}$$

This is an  $AX = XB$  type equation, which is also encountered in the field of wrist-mounted robotic sensors [39, 55]. The calibration matrix has the form of a homogeneous rigid transformation with 6 DOF. Each of the equations puts three constraints on the matrix, so that at least two equations are needed. A closed-form solution is presented in [55]. However, we prefer a least-squares minimization because it is more robust to noise:

$$\arg \min_{T_c} \sum_{i \neq j} d(T_{i,j} T_c, T_c T_j T_i^{-1}) \quad (5.1)$$

with an appropriate distance function  $d(\cdot)$ . For this purpose, the Lie group estimation framework like in section 2.4.2 could be used with the error  $\varepsilon = T_c T_j T_i^{-1} T_c^{-1} T_{i,j}^{-1}$ . How-

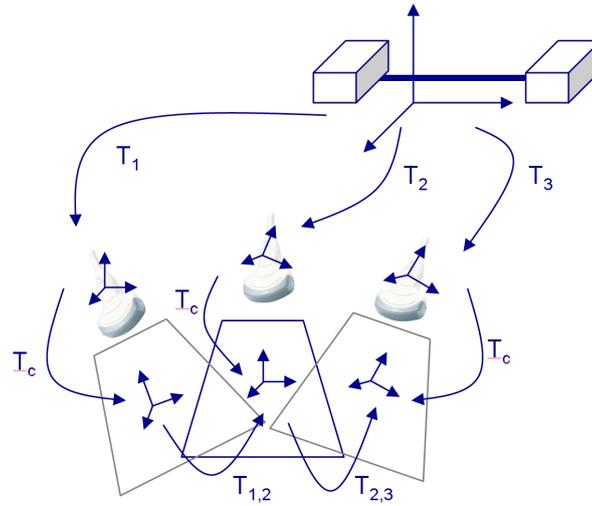
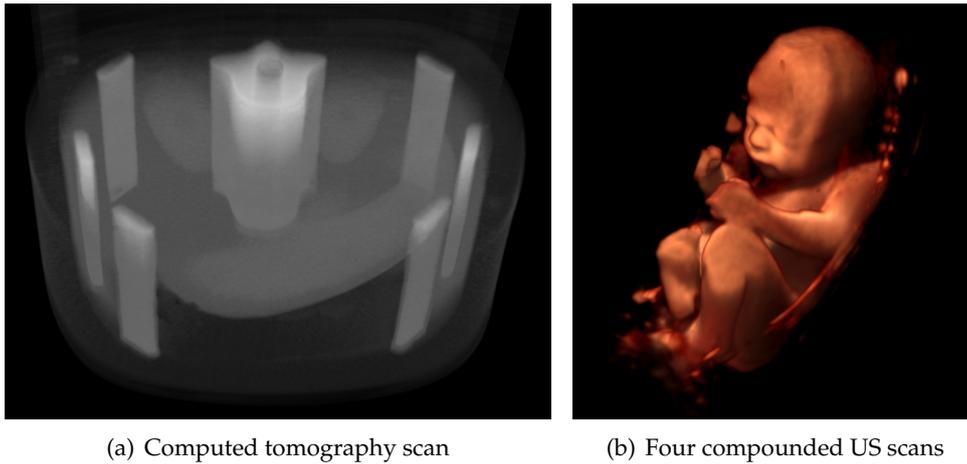


Figure 5.6: Coordinate transformations for ultrasound acquisition with optical tracking.



(a) Computed tomography scan

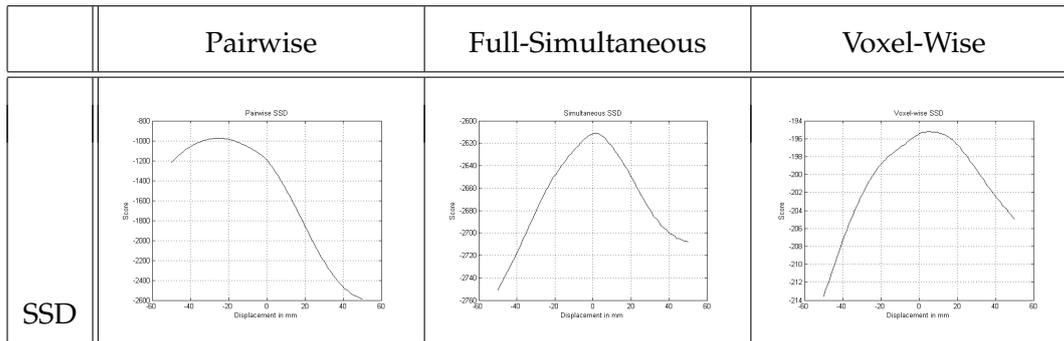
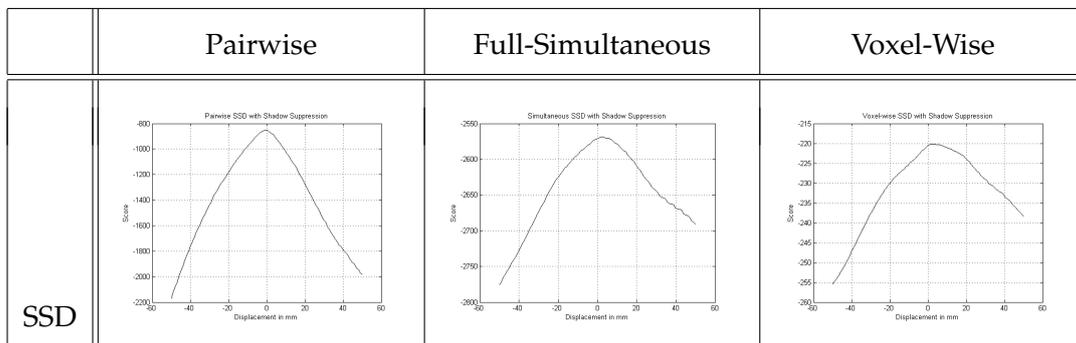
(b) Four compounded US scans

Figure 5.7: Rendering of abdominal and baby phantom.

ever, we are using a method where we do not consider the transformation matrices themselves but, instead, their effect on points. To guarantee a high precision in the workspace, we are placing a cuboid around it with the corners  $\{p_1, \dots, p_n\}$ . The fiducial registration error is then

$$d(T_x, T_y) = \frac{1}{n} \sum_{i=1}^n |T_x p_i - T_y p_i|^2.$$

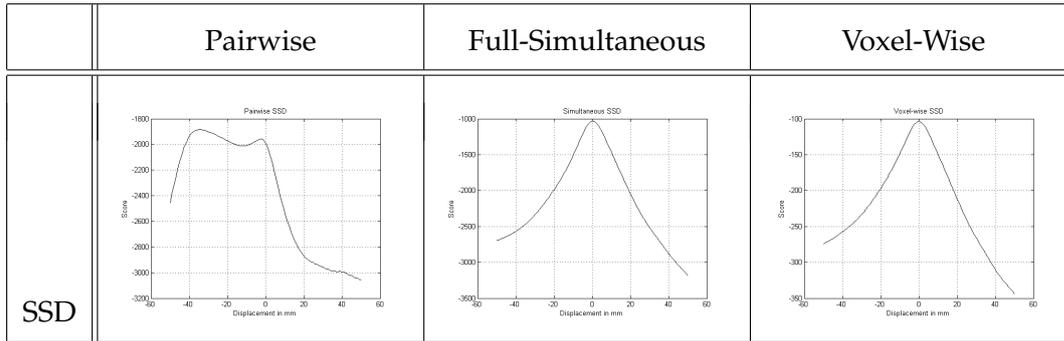
Like for the spine model, we plot the proposed similarity measures by translating one image in the interval  $[-50; 50]$  mm, see table 5.3 for the results. Once again, the bivariate measures are misled from the right alignment at the position 0 by their overlap dependence. The multivariate extensions of SSD and CR produce a smooth cost function with a

Table 5.3: Excerpt of similarity plots from table B.2 for the *abdominal phantom*.Table 5.4: Excerpt of similarity plots from table B.3 for the *abdominal phantom* with *shadow suppression*.

clear maximum at the correct position. The extensions of MI only have a local maximum at that position. Simultaneous NCC has a very sharp maximum close to 0, but the cost function is not very smooth making it hard to find. For voxel-wise NCC, the maximum is at about 18mm displacement.

A second similarity measure plot is done after suppressing the shadow, caused by the ribs and spine, see table 5.4 for the plots. This has a very positive effect on the bivariate measures, having now a clear maximum at the correct position. The influence to the multivariate measures is mixed. Measures like SSD, having already good results without shadow suppression remain constant. Some like voxel-wise MI become better, some like CR become worse.

We also ran a registration study, with an initial random deviation of maximal  $\pm 20$  mm in translation and  $\pm 20^\circ$  in rotation from the correct pose. The mean and standard deviation of each pose parameter of the seven moving images after 100 registrations are shown in table B.9. The phantom presents a difficult registration scenario for the pairwise as well as for the simultaneous registration. The result of the study for the pairwise registration shows in general a large displacement from the correct position. Since registration errors that were done at the first images are transferred to the following ones, the displacement from the correct position is worse for the last images. The simultaneous registration leads to much better results. The mean values are closely distributed around zero. But there are also some large displacements in the rotational components of the images.

Table 5.5: Excerpt of similarity plots from table B.4 for the *baby*.

### 5.4.3 Baby phantom

The last experiments were done on four sequentially taken acquisitions from a baby phantom, see figure 5.7(b) for the compounded result. Also for this data set we plot the proposed similarity measures, see table 5.5. One clearly sees the high overlap dependence of the bivariate measures by having one peak at 0 and the other one at -37mm displacement, corresponding to the total overlap. This is obviously a source for misregistrations. The multivariate measures provide a smooth cost function with a clear maximum at the correct position 0. This also holds for other measures than SSD, see table B.4 in the appendix.

We also ran a registration study, with an initial random deviation of maximal  $\pm 20$  mm in translation and  $\pm 20^\circ$  in rotation from the correct pose. The mean and standard deviation of each pose parameter of the three moving images after 100 registrations are shown in figure 5.8. The pairwise registration leads to a misalignment because of the total overlap of the images 2 and 3, indicated in figure 5.8(a) by a mean of -34.9 mm of parameter 7. The close distribution of the mean values around 0 after the simultaneous registration, together with low variances, indicates good registration results, see figures 5.8(b) and 5.8(c). In the appendix in figure B.8, we also state the results for other similarity measures than SSD. The pairwise registration results are always very similar, with an overlap of the images 2 and 3, although the normalized mutual information performs a bit better. With the simultaneous extension for NCC and CR we achieve also very promising results. The results for the extensions of MI show a large displacement from the correct alignment.

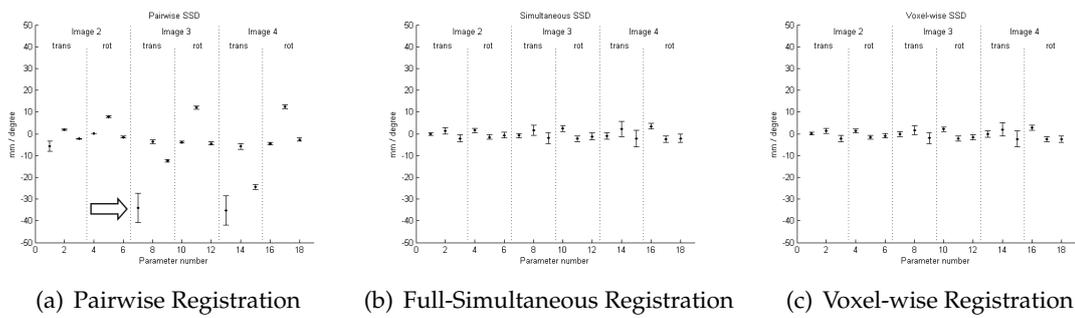


Figure 5.8: Excerpt of random study results from figure B.8 for the *baby*.

# Chapter 6

## Conclusion and Further Development

*In this section, we present a conclusion of the thesis and list methods that could further increase the performance of the registration. This means the increase of quality of the registration as well as a lower computation time. Unfortunately, we were not able to prove their advantages in practice, so that we only sketch the general idea and their possible positive effects.*

### 6.1 Conclusion

In this thesis, we treated the problem of three-dimensional ultrasound mosaicing by focusing on methods to calculate the correct global alignment. In the literature for ultrasound registration, we haven't found approaches that deal with multiple image alignment. Therefore, we extended our search to general image registration where we came across the simultaneous registration idea and the pairwise registration with Lie normalization. For the simultaneous registration we used two variants, the semi-simultaneous and full-simultaneous registration. During our work with simultaneous registration, we found out that there is only a limited number multivariate similarity measures, being introduced in the literature. Thus, we propose our own extensions of multivariate measures that we deduce from a maximum likelihood framework. Basically, there are two classes of extensions, which we are considering; first, summed-up bivariate measures; and second, voxel-wise measures for which we deduced an extension of SSD. Due to the higher computational cost of simultaneous registration we presented methods to make the calculation more efficient, e.g., the stochastic pyramid.

Up to this point, our work was not specialized on ultrasound and is therefore applicable for any modality. In order to achieve superior results for ultrasound registration we applied methods to reduce speckle noise and detect shadow in the images. For the shadow suppression we described four approaches and discussed their performance on images of the human leg.

In extensive experiments, we demonstrated the advantages and drawbacks of the proposed registration strategies. Pairwise registration has the problem of accumulated errors and the tendency of the similarity measures to favor a total overlap of the images. Especially for accumulated errors, a consecutive Lie normalization like it was described helps. But also the normalization is limited and builds upon good pairwise registration

---

results. For most cases the simultaneous approaches perform better, but they come with a higher computational cost. In general, multivariate cost functions are superior to the pairwise ones because they consider the whole imaging scenario. A further advantage of the simultaneous strategies is that accumulated errors are dealt with intrinsically during the registration. The full-simultaneous variant produces better alignments than the semi-simultaneous one, but in return the semi-simultaneous one is faster.

The evaluation of the proposed similarity measures was insofar astonishing that the best results were obtained with SSD. Since we are dealing with unimodal registration SSD was expected to perform well, but other similarity measures that are able to deal with complex mappings of the image intensities should produce similar results. This was not the case. While the similarity plots of the measures for the baby phantom showed good results of all similarity measures, the performance on spine and abdominal phantom were more diverse. This indicates that the performance of the similarity measures largely depends on the imaging scenario and tests on the particular data sets have to be performed. For this purpose, the visualization of the registration process in our prototype was of great value.

The proposed registration strategies have shown their good performance in the mosaicing scenario in comparison to the standard sequential pairwise registration. Especially the simultaneous registration is of great value for ultrasound mosaicing. Multivariate similarity measures have the highest potential to cope with the difficult imaging scenario of viewing angle dependent ultrasound images, which are only partly overlapping.

## 6.2 Further Development

### 6.2.1 Phase information

In the literature, the phase information is used to improve the registration [15, 7] and the compounding [16]. Phase-based analysis has been proposed as an alternative to intensity-based one for many image processing tasks. Phase provides invariance to changes in brightness and contrast within the image. This property makes it particularly interesting for ultrasound images, in which beam attenuation is present and echo intensity depends on the angle of incidence of the ultrasound beam. Local phase is usually calculated by combining the output of a set of filters with different angles. In [11], the authors introduce a new approach to calculate the phase in n-dimensional signals: the monogenic signal, an isotropic extension of the analytic signal which preserves its basic properties. Analogous to the analytic signal for 1-D, to build the monogenic signal the Riesz transform, a generalization of the Hilbert transform for higher dimensional signals, is used.

The phase is calculated by first filtering the images with a log-Gabor filter to achieve frequency as well as spatial localization. The log-Gabor filter can hereby be replaced by any other band-pass filter. Afterwards the monogenic signal is computed. This is done by the applying the Riesz transform, extension of the Hilbert transform, to the images. The Hilbert transform is necessary to calculate the imaginary part of the signal which is essential for the phase calculation. The Riesz transform is obtained by multiplying the image in the frequency domain by

$$H_i(u) = \frac{u_i}{|u|}$$

where  $u = [u_1, \dots, u_n]$ , with  $u_i$  representing the  $i$ -th coordinate unit vector. There are as many filters as image dimensions. In the case of 3D images, the monogenic signal consists of 4 values, one real value (original signal)  $re$  and an imaginary 3-vector (result of Riesz transform)  $\mathbf{im}$ . The angles defined by this vector represent the local phase and structure orientation, respectively. So what we refer to phase is defined by

$$\text{phase} = \text{atan} \left( \frac{\sqrt{\mathbf{im}(1)^2 + \mathbf{im}(2)^2 + \mathbf{im}(3)^2}}{re} \right).$$

with  $\text{atan}$  (arcus tangent) the inverse tangent.

For the registration, the phase information is used in form of a phase based similarity measure. It may be possible to find a multivariate extension of this measure. For the compounding, the values are weighted based on the phase information, see section 4.2.

### 6.2.2 Model-based registration

The term model-based registration has several meanings. It can for example refer to the introduction of prior knowledge into the registration process in form of joint statistics of correctly registered images or the introduction of a model of the underlying imaged object. In this section, we discuss the introduction of a model of the imaged object. It is clear, that it is not possible to have a model of the underlying object for all body parts, but for the extremities it may be possible. The dominant structure in the extremities is the bone which can be approximated by a cylinder, see figure 6.1.

A possible registration algorithm is to first "dock" each image to the cylinder model. Then, the rotation axis of the cylinder is set to be the registration center and the images are aligned to each other. First experiments show that the docking works well, because of the strong and clear bone structure in the images. To enlarge the capture range of the model, the intensity values decrease linearly from the surface of the cylinder. For the later on registration of the volumes, the optimization over 4 parameters, 3 for translation and 1 for rotation along the cylinder axis, is sufficient.

Through the breakdown of the process into two parts and the usage of the prior-knowledge, the overall registration becomes more robust. Unfortunately, from a current point of view, its usage may remain limited to the extremities.

### 6.2.3 Optimization

Like already mentioned, the simultaneous registration leads to a high computational cost. Methods for a faster evaluation of the cost function were shown in section 3.5. Another possibility to speed up the registration, is to use more efficient non-linear optimizers. For our experiments we used the simplex, the best neighbor, and the Powell-Brent optimizer. Faster convergence could be gained by using gradient based ones or stochastic ones. For the gradient based optimizers like gradient descent, newton's method, or Levenberg-Marquardt, the gradient of the cost function has to be approximated. The optimizers follows the direction of the gradient to faster approach the optimum. But the estimation of the gradient leads to an additional computational cost.

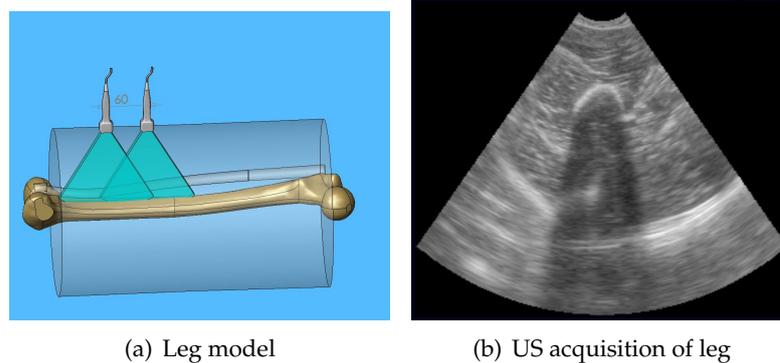


Figure 6.1: Model of the human leg and ultrasound scan of leg.

Stochastic optimization adds randomness in the optimization process. The randomness may be present as either noise in measurements or Monte Carlo randomness in the search procedure, or both. Stochastic algorithms have already been used for the image registration e.g. the simultaneous perturbation stochastic approximation (SPSA) algorithm proposed by Spall [56]. Moreover, Zöllei et al. used an stochastic optimizer within the stochastic congealing framework [75]. Stochastic optimizers are necessary for working with stochastic similarity measures.

#### 6.2.4 Non-rigid registration

There are roughly two sources of non-rigid deformation in the images. Either the deformation of soft tissue due to organ motion, or the compression of tissue caused by the ultrasound probe. For the second kind of deformations, one could avoid the usage of a non-rigid registration by uncompressing the image in a pre-processing step. Therefore, tests would be necessary to determine the compression profile of the transducer for the specific body parts. For all other kind of deformations, non-rigid registrations have to be used, which come with a much higher computational cost than rigid ones. Also the chance of misregistrations increases because of the increased degree-of-freedom.

#### 6.2.5 Parallelization

At the moment, there is a paradigm shift in development of microprocessors from single-core to multi-core processors. Further processors will possess multiple cores and the parallelization of the programs will be necessary to profit from their increasing performance. This could be very beneficial for the registration, especially for the simultaneous registration, which comes along with a much higher computational cost than the pairwise one. Parallelization can be done at different granularities. Either, fine-grain parallelism that has relatively small amounts of computational work being done between communication events and therefore facilitating load balancing. Or, coarse-grain parallelism with large amounts of computational work between the communication events. For the registration, the load balancing could be integrated into the optimizer, that sends to different

---

worker units the task to evaluate a cost function at a certain position. The evaluation of the cost function is very well suited for the optimization because it comes with a reasonable amount of work and can be done completely independent, without communication to other processes. Moreover, the amount of information that has to be communicated is very small (only the pose). The optimizer would therefore have the additional task to manage the parallelization.

Another possibility for a speed-up through parallelization would be the usage of the graphics processing unit (GPU). The GPU is originally designed for the graphics rendering, where their highly parallel structure makes them more effective than typical CPUs. Because of its increasing performance, it became more and more interesting to "misuse" it to perform ordinary calculation. The translation of complex programs from CPU to GPU is not trivial but the new generations of graphic boards support an ever increasing instruction set, making the port of parts or the whole registration to the GPU possible.

---



# Appendix A

## Estimation problems on Lie groups

Bertrand's Paradox [38, 45] is a well-known example for the problems that arise when treating a Lie group like a vector space within an estimation problem. Similar problems also come up when calculating estimates for 2D and 3D rigid transformations because they form real Lie groups. The following generalization of concepts that are designed for usual vector spaces uses the profound techniques of differential geometry. For an introduction to differential geometry and Lie groups the reader is referred to the standard textbooks [10, 29, 17]. In this section we mainly refer to the basic statistical tools for geometric features were developed by Xavier Pennec ([41, 42]).

In order to illustrate the problematic, let us take the example of the mean 3D rotation which is also of central interest in this report. We could calculate the mean rotation matrix  $\bar{R} = \frac{1}{n} \sum_i R_i$ , the mean quaternion  $\bar{q} = \frac{1}{n} \sum_i q_i$  by using unit quaternions, or the mean rotation vector  $\bar{r} = \frac{1}{n} \sum_i r_i$ . All these methods lead to different results, whereas the first two are not even rotations. The main problem is that the addition and the multiplication by a scalar are operators, which are only defined in a vector space (for example each chart). For the development of a computational framework on manifolds the minimization of intrinsic values shows to be a powerful concept [44]. For the mean the minimization of an intrinsic distance is necessary, which is provided by the Riemannian metric of the manifold. Pennec developed an estimation and optimization framework for rotations and rigid body motions where composition and inversion replaced addition and subtraction. The interesting parts of this framework together with some basics of differential geometry are presented in the following paragraphs.

### A.1 Differentiable manifolds and Lie groups

A central role in differential geometry play *differentiable n-manifolds* and *Lie groups* which are introduced in the following.

**Definition:** An *n-chart* on a set  $X$  is a pair  $(U, \Phi)$ , where  $U \subset X$  and  $\Phi$  is a bijection of  $U$  onto an open subset  $\Phi(U) \subset \mathbb{R}^n$ . Two *n-charts*  $(U, \Phi)$  and  $(V, \Psi)$  are said to be *compatible* if two conditions are satisfied:

1.  $\Phi(U \cap V)$  and  $\Psi(U \cap V)$  are open in  $\mathbb{R}^n$ .
  2.  $\Psi \circ \Phi^{-1} : \Phi(U \cap V) \rightarrow \Psi(U \cap V)$  and  $\Phi \circ \Psi^{-1} : \Psi(U \cap V) \rightarrow \Phi(U \cap V)$  are both infinitely differentiable.
-

**Definition:** An  $n$ -atlas on a set  $X$  is a family  $\{(U_i, \Phi_i)\}_{i \in I}$  of  $n$ -charts such that:

1. Any two charts in the family are compatible.
2.  $X = \bigcup\{U_i : i \in I\}$ .

Two  $n$ -atlases are said to be compatible if their union is also an  $n$ -atlas.

**Definition:** An  $n$ -smoothness structure  $\mathcal{S}$  on  $X$  is a set of  $n$ -atlases on  $X$  such that

1. Any two members of  $\mathcal{S}$  are compatible.
2. Any  $n$ -atlas on  $X$  that is compatible with some member of  $\mathcal{S}$  is also a member of  $\mathcal{S}$ .

**Definition:** A differentiable  $n$ -manifold  $M$  is a pair  $(X, \mathcal{S})$ , where  $X$  is a set and  $\mathcal{S}$  is an  $n$ -smoothness structure on  $X$  such that the countability axiom and the Hausdorff axiom are satisfied.

**Definition:** A Lie group  $\mathcal{G}$  is a triple  $(G, \mathcal{S}, \circ)$ , where  $G$  is a set,  $\mathcal{S}$  is an  $n$ -smoothness structure on  $G$  and  $\circ : G \times G \rightarrow G$  such that

1.  $(G, \mathcal{S})$  is a manifold.
2.  $(G, \circ)$  is a group.
3.  $\circ : G \times G \rightarrow G$  is smooth.
4.  $g \mapsto g^{-1} : G \rightarrow G$  is smooth.

**Definition:** The left- and right-compositions by an element  $g \in \mathcal{G}$  are

$$\begin{aligned} L_g : \mathcal{G} &\rightarrow \mathcal{G} \\ x &\mapsto L_g(x) = g \circ x \\ R_g : \mathcal{G} &\rightarrow \mathcal{G} \\ x &\mapsto R_g(x) = x \circ g \end{aligned}$$

with the differential maps

$$\begin{aligned} DL_g(x) : T_x \mathcal{G} &\rightarrow T_{g \circ x} \mathcal{G} \\ u &\mapsto DL_g(x) \cdot u = \left. \frac{\partial g \circ y}{\partial y} \right|_{y=x} \cdot u \\ DR_g(x) : T_x \mathcal{G} &\rightarrow T_{x \circ g} \mathcal{G} \\ u &\mapsto DR_g(x) \cdot u = \left. \frac{\partial y \circ g}{\partial y} \right|_{y=x} \cdot u \end{aligned}$$

which define maps from  $T_x \mathcal{G}$  the tangent space at the point  $x$  to its counterpart  $T_{g \circ x} \mathcal{G}$  or  $T_{x \circ g} \mathcal{G}$ .

---

## A.2 Left invariant metric and distance

We are interested in measuring the distance between the various transformations. A Riemannian metric can be defined for Lie groups which provides a distance metric.

**Definition:** A *Riemannian metric* for a Lie group  $\mathcal{G}$  is a continuous collection of dot products on the tangent space  $T_x\mathcal{G}$  at  $x$

$$\langle v|w \rangle_x = v^T \cdot G(x) \cdot w. \quad (\text{A.1})$$

Each dot products can be transformed to the Id-tangent space  $T_{\text{Id}}\mathcal{G}$  with the above defined differential right- and left-compositions

$$G(x) = DL_x(\text{Id})^{-T} \cdot G(\text{Id}) \cdot DL_x(\text{Id})^{-1}. \quad (\text{A.2})$$

The left-composition differential map  $DL_x$  is a left invariant metric. The distance between two points of Lie group can be measured by measuring the minimum length of the curves connecting these points. The Riemannian metric provides therefore an intrinsic measure. The minimal curves between points are referred to as *geodesics*  $\gamma$ . It can be shown with the calculus of variation that there exists one and only one geodesic  $\gamma_{x,u}(\cdot)$  starting at  $x$  and having  $u$  as tangent vector.

## A.3 Exponential and logarithmic map

A central role in differential geometry plays the exponential map together with its inverse the logarithmic map. For a point  $x \in \mathcal{G}$ , a vector  $u \in T_x\mathcal{G}$ , and the uniquely associated geodesic  $\gamma_{x,u}(\cdot)$  the exponential map is

$$\begin{aligned} \exp_x : T_x\mathcal{G} &\rightarrow \mathcal{G} \\ u &\mapsto \exp_x(u) = \gamma_{x,u}(1) \end{aligned}$$

with  $\gamma_{x,u}(1)$  the point reached after a unit time by the geodesic. Within the neighborhood of  $x$  the inverse logarithmic map  $\log_x(\cdot)$  is defined and

$$\exp_x(u) = y \iff u = \log_x(y). \quad (\text{A.3})$$

For an illustration on the unit sphere, see figure [A.1](#).

The notion of an exponential map can be ambiguous because either the Lie group or the Riemannian one can be used, which do for rigid transformations not agree [2]. Because of our major interest in distance measurements we continue by using the Riemannian exponential map.

With the usage of left-invariant metrics the exponential and logarithmic maps at any point of a Lie group can be related to their counterpart at the identity with:

$$\begin{aligned} \log_x(y) &= DL_x(\text{Id}) \circ \log_{\text{Id}}(x^{-1} \circ y) \\ \exp_x(u) &= x \circ \exp_{\text{Id}}(DL_x(\text{Id})^{-1}u). \end{aligned}$$

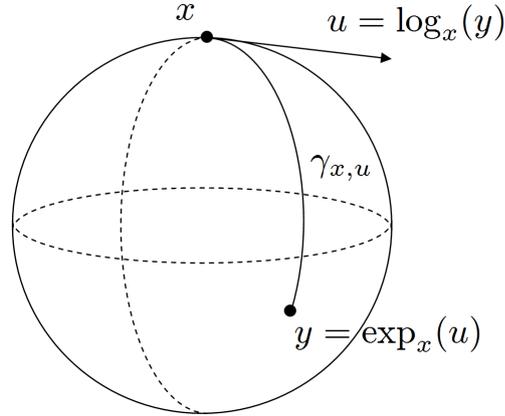


Figure A.1: Riemannian exponential and logarithmic maps on unit sphere together with the geodesic.

The (geodesic) distance between two points is then defined by:

$$\begin{aligned} \text{dist}(x, y) &= \text{dist}(\text{Id}, x^{-1} \circ y) = \left\| \log_{\text{Id}}(x^{-1} \circ y) \right\| \\ &= \sqrt{\langle \log_{\text{Id}}(x^{-1} \circ y) \mid \log_{\text{Id}}(x^{-1} \circ y) \rangle} \end{aligned}$$

with  $\langle \cdot \mid \cdot \rangle$  the inner product. We can limit our interest to the identity tangent space as the above equations allow us to relate each other one to it.

#### A.4 Mean and covariance matrix

For the normalization of rigid transformations, in order to make them as consistent as possible, the statistical ideas of mean and covariance are essential. As already mentioned, Lie groups are not vector spaces so these statistical terms cannot be easily applied. However, for any metric space, it is possible to define probabilistic spaces, random elements and probability density functions [38]. Expectations and other usual tools are then defined for random variables, which are real-valued functions of the probabilistic events, but not directly for random elements of the group [46]. But, by changing the definition of the expectation and using the Riemannian geometry tools from the last sections, it turns out that a consistent statistical framework (including the mean, the covariance matrix, and the Mahalanobis distance) can be defined [43].

Thinking of the mean as the element that minimizes the expected distance to a random vector it is possible to extend this idea to Lie groups. Let  $x$  be a random element and let

$$\sigma_x^2(y) = \mathbb{E} \left[ \text{dist}(y, x)^2 \right]$$

be its variance at the (fixed) element  $y$ . This is well defined as  $\text{dist}(y, \cdot)$  is a real-valued function.

Now let  $x$  be a random element of a Lie group  $\mathcal{G}$ . If the variance  $\sigma_x^2(y)$  is finite for all elements  $y \in \mathcal{G}$ , every element minimizing the variance is called a Fréchet mean element. The set of Fréchet mean elements is thus given by

$$\mathcal{F}_x = \arg \min_{y \in \mathcal{G}} (\mathbb{E}[\text{dist}(y, x)^2])$$

It can be shown that under suitable conditions, that are fulfilled here, there exists one and only one Fréchet mean that is denoted as  $\mathbb{E}[x]$ .

To define higher order moments of a distribution, on a Lie group, the exponential map at the mean point is used. The random feature is thus represented as a random vector with zero mean in a star-shaped domain. With this representation, the covariance matrix can be defined by:

$$\begin{aligned} \Sigma_{xx} &= \mathbb{E} [\log_{\mathbb{E}[x]}(x) \cdot \log_{\mathbb{E}[x]}(x)^T] \\ &= DL_g(\mathbb{E}[x]) \cdot \mathbb{E} [\log_{\text{Id}}(\mathbb{E}[x]^{-1} \circ x) \cdot \log_{\text{Id}}(\mathbb{E}[x]^{-1} \circ x)^T] \cdot DL_g(\mathbb{E}[x])^T \end{aligned}$$

Finally, the Mahalanobis distance, which plays a key role to get a robust estimation of the global positioning, is defined. The extension of the Mahalanobis distance to Lie groups can be easily done with the above described covariance matrix. The Mahalanobis distance of a point  $y$  to a random feature with Fréchet mean  $\mathbb{E}[x]$  and covariance matrix  $\Sigma_{xx}$  is given by

$$\mu_x^2(y) = \log_{\mathbb{E}[x]}(y)^T \cdot \Sigma_{xx}^{-1} \cdot \log_{\mathbb{E}[x]}(y). \quad (\text{A.4})$$



# Appendix B

## Detailed experimental results

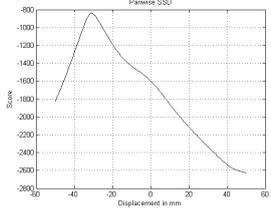
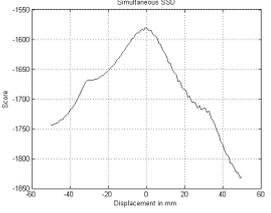
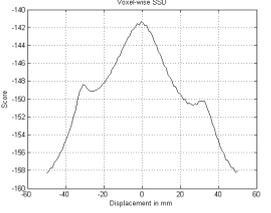
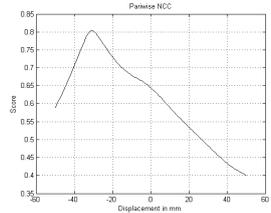
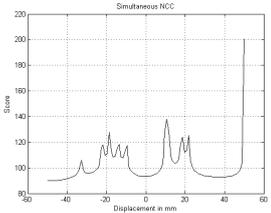
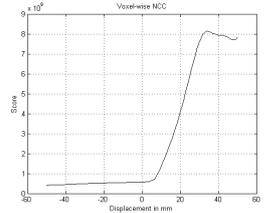
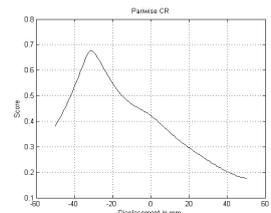
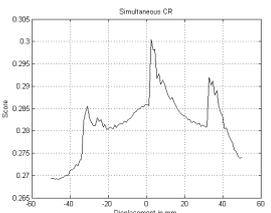
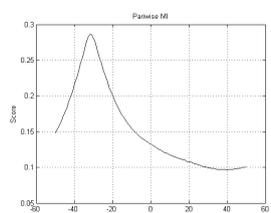
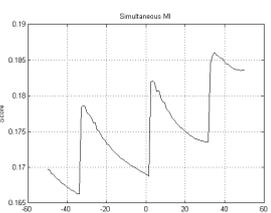
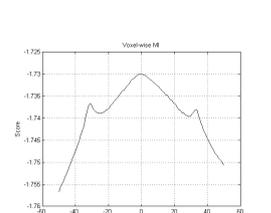
	Pairwise	Full-Simultaneous	Voxel-Wise
SSD			
NCC			
CR			-
MI			

Table B.1: Similarity plots of the in table 3.1 proposed measures for the *spine*.

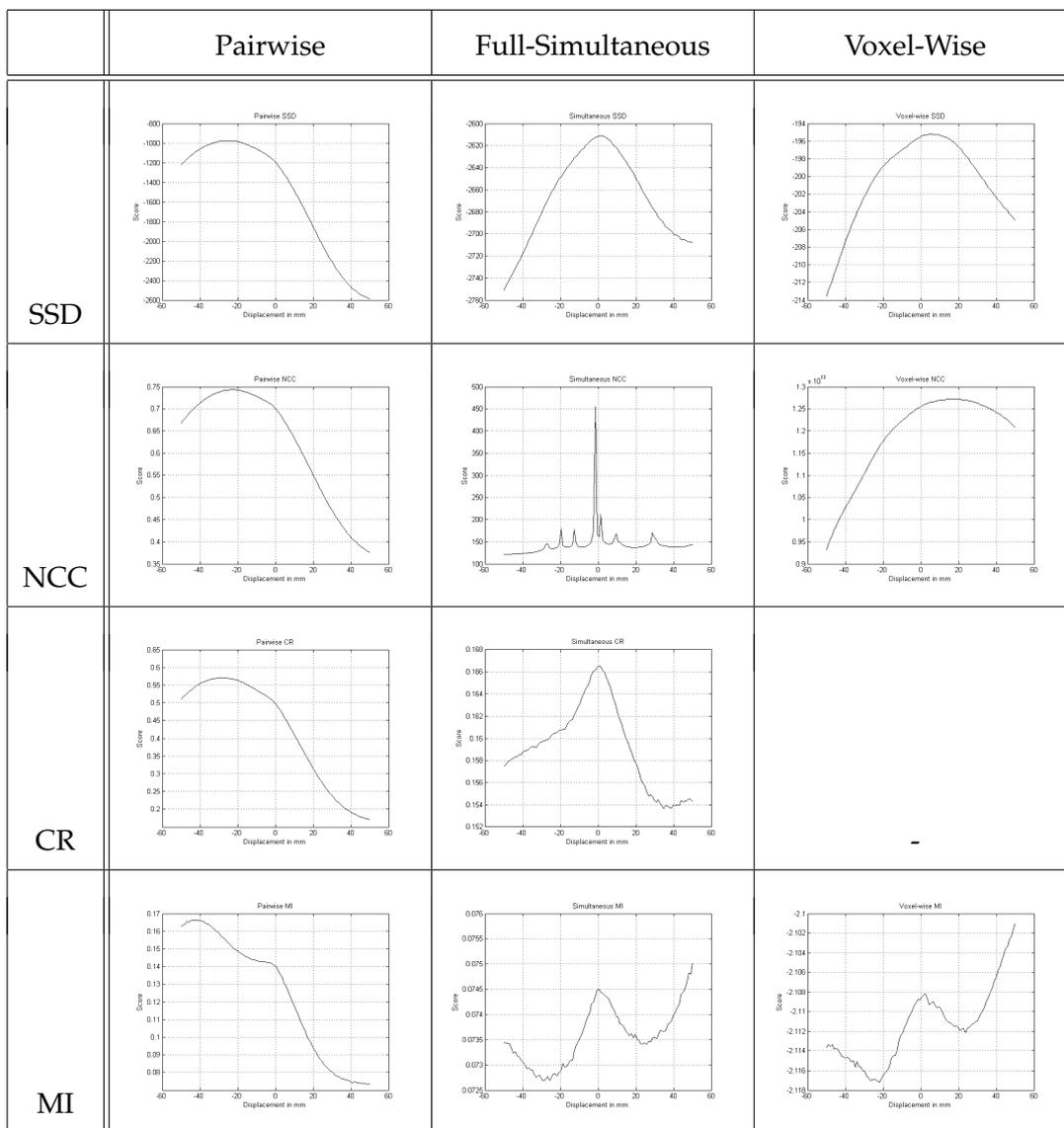


Table B.2: Similarity plots of the in table 3.1 proposed measures for the *abdominal phantom*.

	Pairwise	Full-Simultaneous	Voxel-Wise
SSD			
NCC			
CR			
MI			

Table B.3: Similarity plots of the in table 3.1 proposed measures for the *abdominal phantom with shadow suppression*.

	Pairwise	Full-Simultaneous	Voxel-Wise
SSD			
SAD			
NCC			
CR			-
MI			

Table B.4: Similarity plots of the in table 3.1 proposed measures for the *baby phantom*.

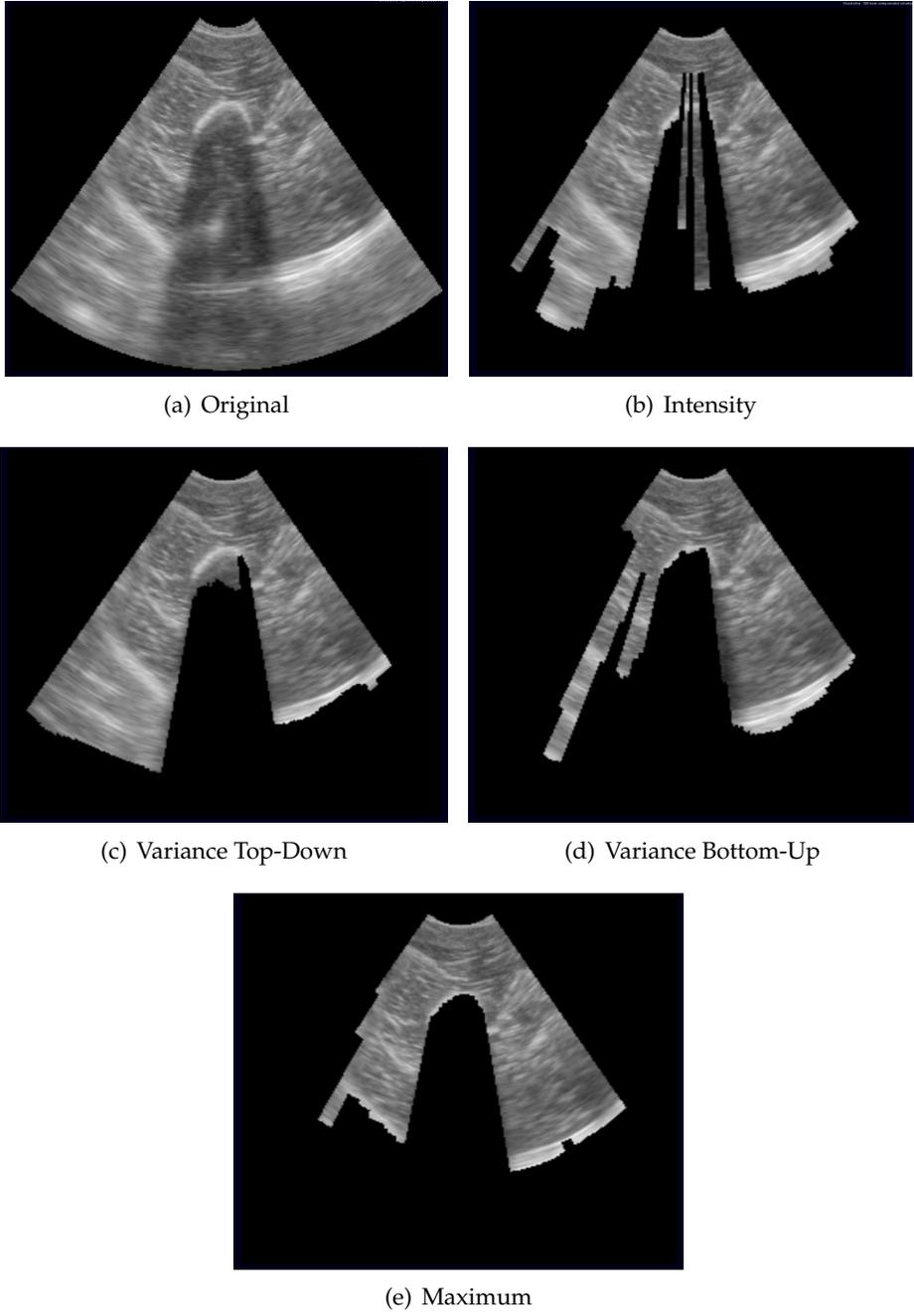


Figure B.1: Comparison of shadow suppression methods.

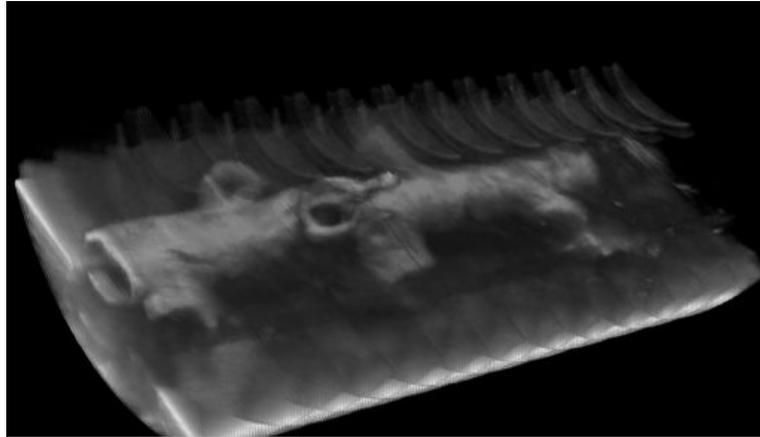
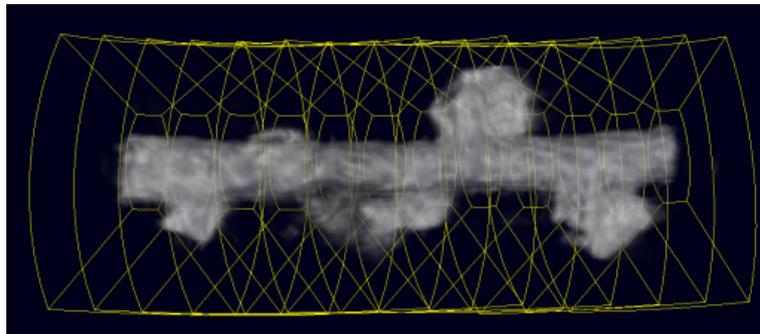
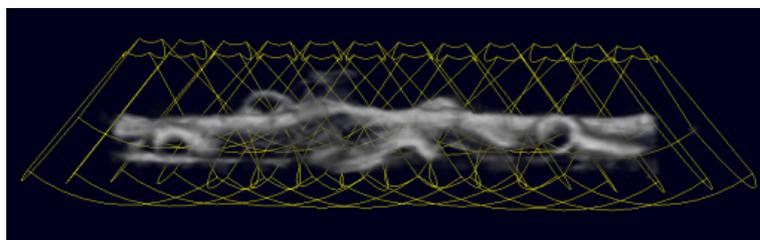


Figure B.2: 3D rendering of the compounded *spine* acquisitions.



(a) Top view



(b) Side view

Figure B.3: 3D rendering of *spine* model with US fan.

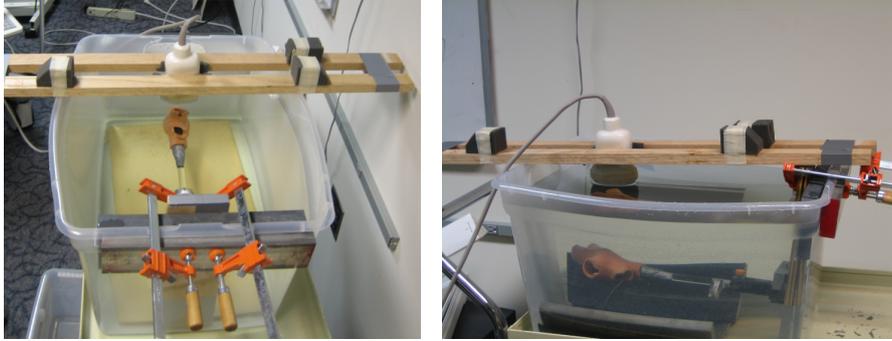


Figure B.4: Image acquisition of *heart* clay model in water bath.

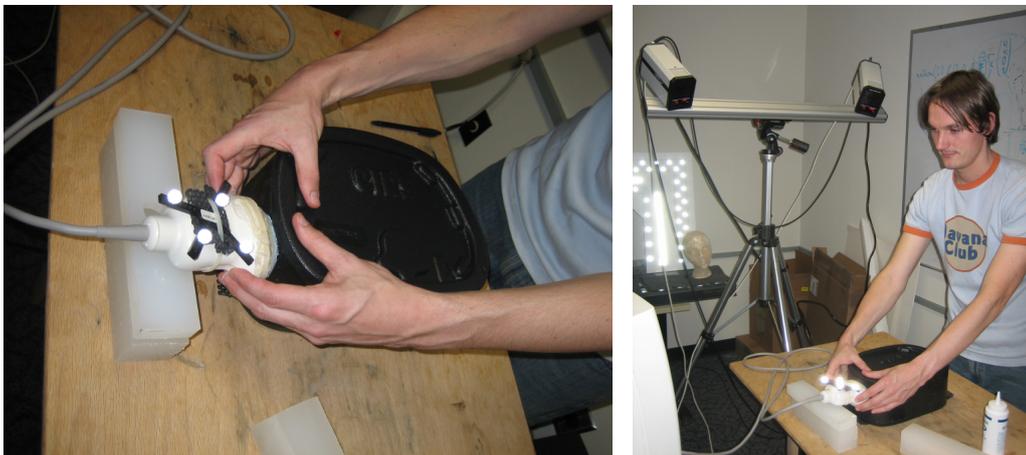
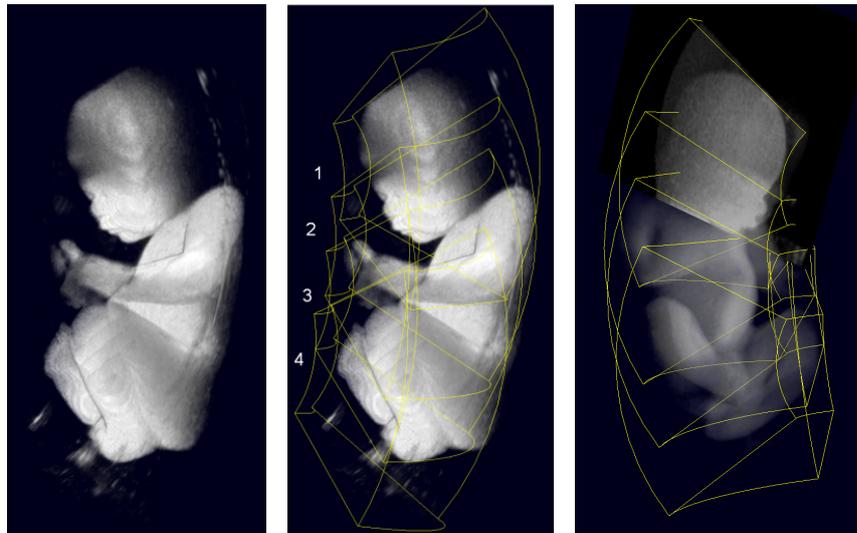


Figure B.5: Image acquisition with optical tracking of *abdominal phantom*. Optical markers are attached to the ultrasound transducer. Optical infrared cameras, pointing at the work space, track the pose.



(a) Compounded volume (b) Augmentation of fan (c) MPR slice and Rendering

Figure B.6: Renderings of baby phantom.



Figure B.7: Advanced 3D Renderings of baby phantom. Done by the Imaging and Visualization department at SCR.

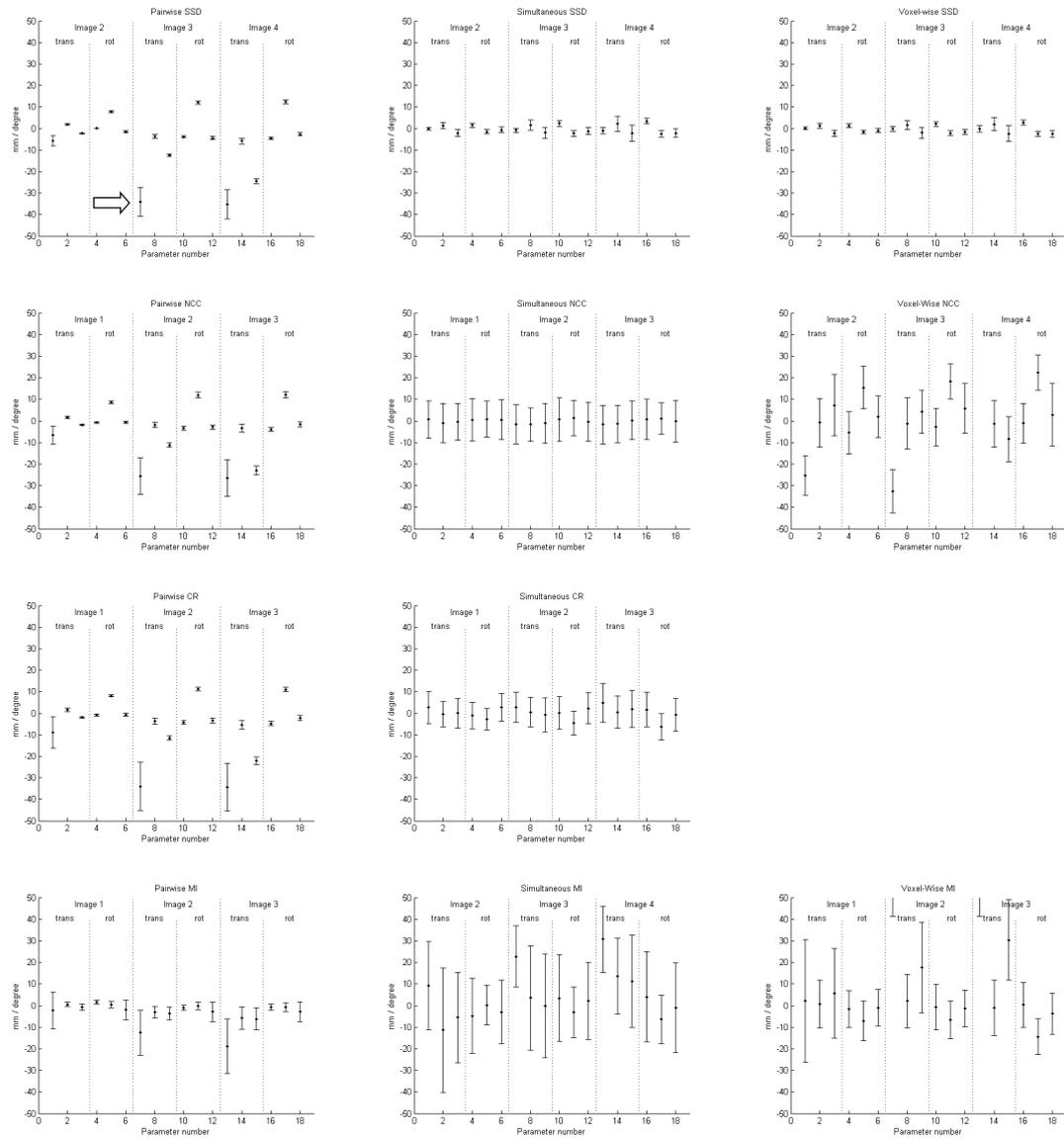


Figure B.8: Random study results for the in table 3.1 proposed measures for the *baby phantom*. Calculated over 100 registrations with maximal  $\pm 20$  mm/ $^\circ$  initial displacement.

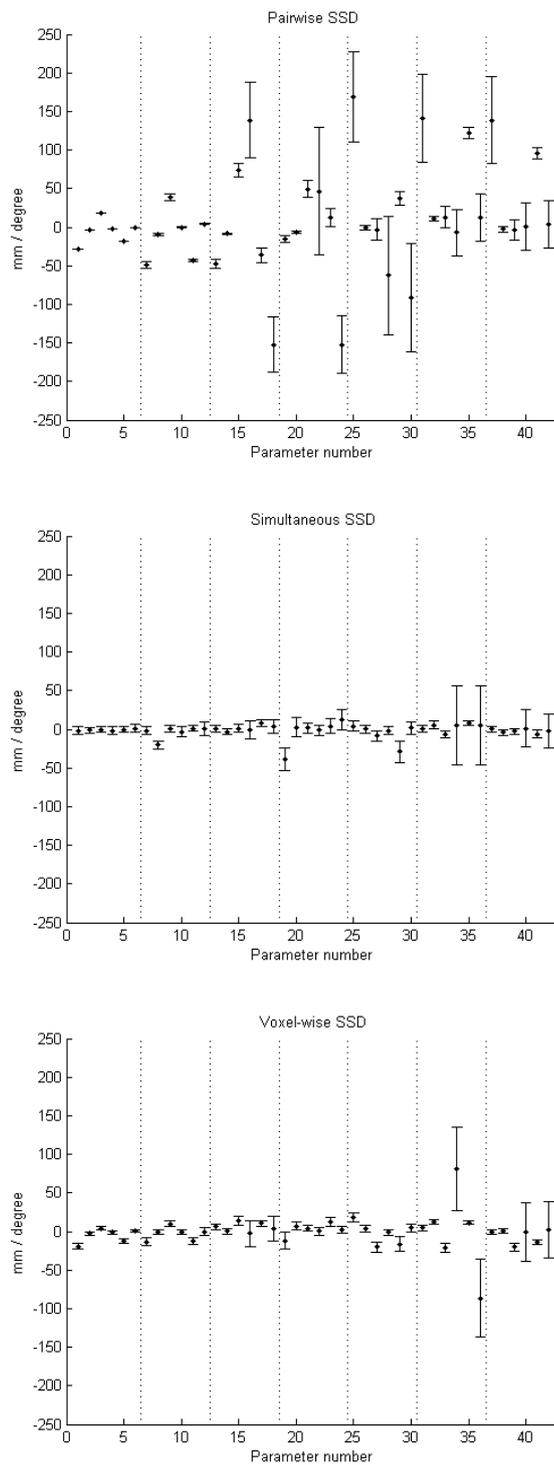


Figure B.9: Random registration study using SSD on *abdominal phantom* with 100 overall registrations and maximal  $\pm 20$  mm/ $^\circ$  initial displacement.

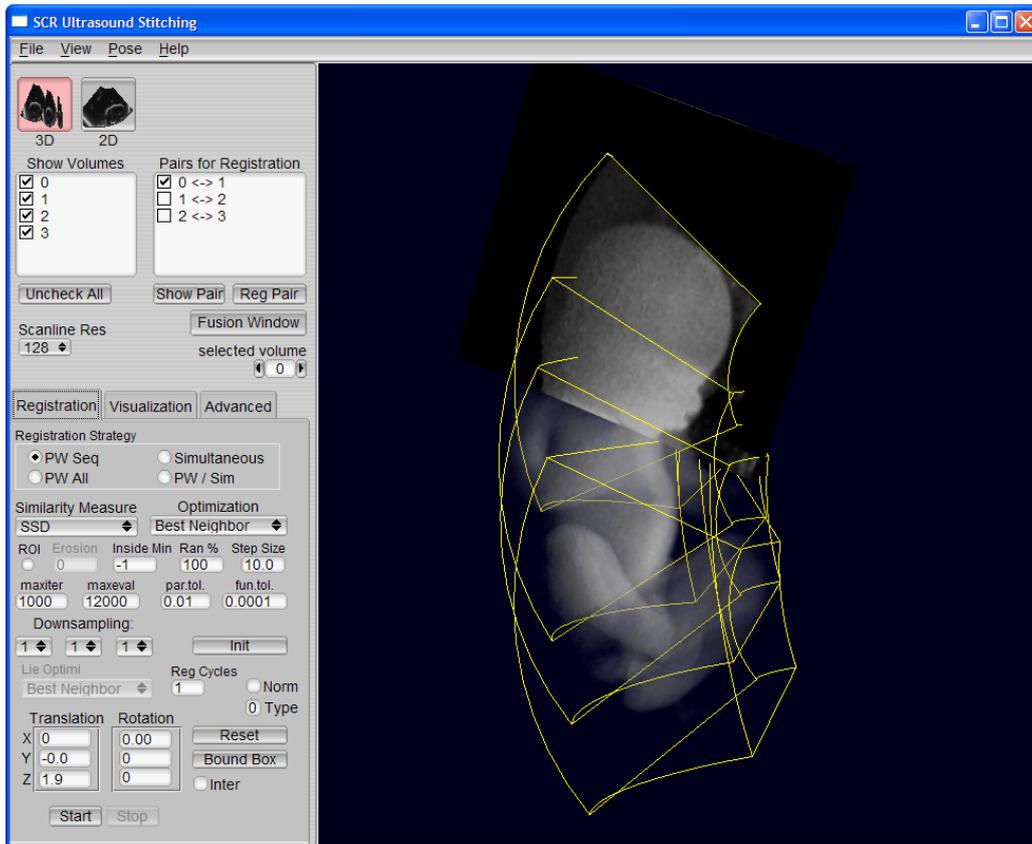


Figure B.10: Screenshot of mosaicing prototype. Visualization of baby phantom with volume rendering and MPR slice.

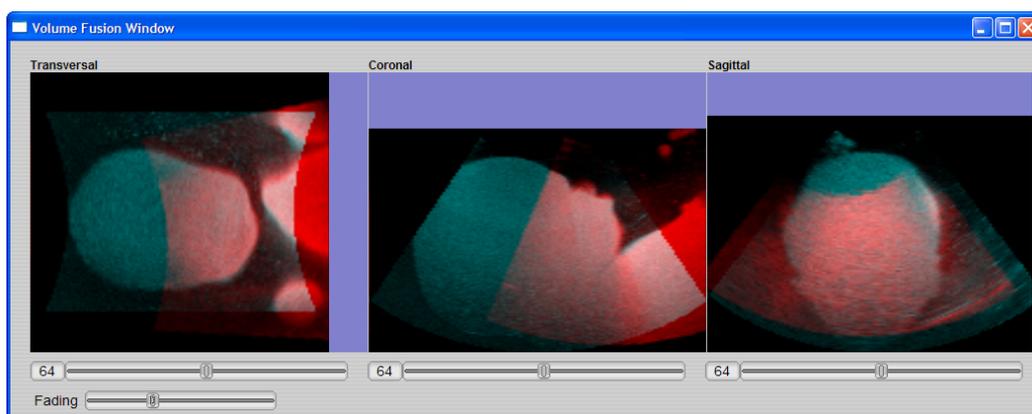


Figure B.11: Screenshot of triple view from mosaicing prototype.



# Bibliography

- [1] Dror Aiger and Daniel Cohen-Or. Mosaicing ultrasonic volumes for visual simulation. *IEEE Comput. Graph. Appl.*, 20(2):53–61, 2000.
  - [2] Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Bi-invariant means in Lie groups. application to left-invariant polyaffine transformations. Research report RR-5885, INRIA Sophia-Antipolis, April 2006.
  - [3] Kanwal K. Bhatia, Joseph V. Hajnal, Basant K. Puri, A. David Edwards, and Daniel Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *ISBI*, pages 908–911, 2004.
  - [4] J. L. Boes and C. R. Meyer. Multi-variate mutual information for registration. *Medical Image Computing and Computer-Assisted Intervention*, 1679:606–612, 1999.
  - [5] S. Brekke, S. I. Rabben, A. Haugen, G. U. Haugen, E. N. Steen, and H. G. Torp. Real-time volume stitching in 4d echocardiography. In *Ultrasonics Symposium, 2005 IEEE*, volume 2, pages 1228–1231, 2005.
  - [6] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
  - [7] James Davis. Mosaics of scenes with moving objects. In *CVPR*, pages 354–360, 1998.
  - [8] Leotta D.F. and Martin R.W. Three-dimensional spatial compounding of ultrasound scans with incidence angle weighting. *Ultrasonics Symposium*, 2:1605–1608, 1999.
  - [9] C.F. Dietrich, A. Ignee, M. Gebel, B. Braden, and G. Schuessler. Imaging of the abdomen. *Z Gastroenterol*, 40:965–970, 2002.
  - [10] Manfredo P. do Carmo. *Riemannian Geometry*. Birkhäuser, Boston, MA, 1992.
  - [11] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, December 2001.
  - [12] C. Fookes, J. Williams, and M. Bennamoun. Global 3d rigid registration of medical images. In *International Conference on Image Processing*, volume 2, pages 447–450, 2000.
  - [13] Andrew H. Gee, Graham M. Treece, Richard W. Prager, C. J. C. Cash, and Laurence H. Berman. Rapid registration for wide field-of-view freehand 3d ultrasound. *IEEE Trans. Med. Imaging*, 22(11):1344–1357, 2003.
  - [14] Claus Gramkow. On averaging rotations. *J. Math. Imaging Vis.*, 15(1-2):7–16, 2001.
-

- 
- [15] Vicente Grau, Harald Becher, and J. Alison Noble. Phase-based registration of multi-view real-time three-dimensional echocardiographic sequences. In *MICCAI (1)*, pages 612–619, 2006.
- [16] Vicente Grau and J. Alison Noble. Adaptive multiscale ultrasound compounding using phase information. In *MICCAI*, pages 589–596, 2005.
- [17] S. Helgason. *Differential Geometry, Lie Groups and Symmetric Spaces*. American Mathematical Society, 2001.
- [18] Wolfgang Henrich, Annette Schmider, Siri Kjos, Boris Tutschek, and Joachim W. Dudenhausen. Advantages of and applications for extended field-of-view ultrasound in obstetrics. *Archives of Gynecology and Obstetrics*, V268:121–127, Jun 2003.
- [19] R. J. Housden, A. H. Gee, G. M. Treece, and R. W. Prager. Subsample interpolation strategies for sensorless freehand 3d ultrasound. *Ultrasound Med Biol*, 32:1897–904, 2006.
- [20] Alan Julian Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, 1991.
- [21] Bernd Jähne. *Digital Image Processing*. Springer, 5 edition, 2002.
- [22] Krücker JF, Meyer CR, LeCarpentier GL, Fowlkes JB, and Carson PL. 3d spatial compounding of ultrasound images using image-based nonrigid registration. *Ultrasound in Medicine and Biology*, 26(9):1475–1488, 2000.
- [23] Krücker JF, LeCarpentier GL, Fowlkes JB, and Carson PL. Rapid elastic image registration for 3-d ultrasound. *IEEE Transactions on Medical Imaging*, 21(11):1384–1394, 2002.
- [24] Kern J.P., Pattichis M., and Stearns S.D. Registration of image cubes using multivariate mutual information. In *Signals, Systems and Computers*, volume 2, pages 1645–1649, 2003.
- [25] Se Hyung Kim, Byung Ihn Choi, Kyoung Won Kim, Kyoung Ho Lee, and Joon Koo Han. Extended Field-of-View Sonography: Advantages in Abdominal Applications. *J Ultrasound Med*, 22(4):385–394, 2003.
- [26] J. F. Krücker, C. R. Meyer, T. A. Tuthill, G. L. Lecarpentier, J. B. Fowlkes, and P. L. Carson. 3D compounding of B-scan ultrasound images. *Acoustical Society of America Journal*, 105:1209–+, February 1999.
- [27] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.
- [28] Weng L., Tirumalai A. P., Lowery C. M., Nock L. E., Gustafson D. E., Von Behren P. L., and Kim J. H. Us extended-field-of-view imaging technology. *Radiology*, 203(3):877–880, 1997.
- [29] J.M. Lee. *Introduction to Smooth Manifolds*. Springer-Verlag New York, 2002.
-

- 
- [30] Antoine Leroy, Pierre Mozer, Yohan Payan, and Jocelyne Troccaz. Rigid registration of freehand 3d ultrasound and ct-scan kidney images. In *MICCAI (1)*, pages 837–844, 2004.
- [31] Y.L. Leung, A.L. Roshier, S. Johnson, R. Kerslake, and D.S. McNally. Demonstration of the appearance of the paraspinal musculoligamentous structures of the cervical spine using ultrasound. *Clin Anat*, 18(2):96–103, 2005.
- [32] Jonathan Lok-Chuen Lo, Michael Brady, and Niall Moore. Simultaneous multiple image registration method for t1 estimation in breast mri images. In Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 4191 of *Lecture Notes in Computer Science*, pages 865–872. Springer, October 2006.
- [33] Stephen Marsland, Carole Twining, and Chris Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *Medical Image Computing and Computer-Aided Intervention (MICCAI)*, number 2879 in *Lecture Notes in Computer Science*, pages 771–779. Springer, 2003.
- [34] Hiroyuki Matsuda. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Phys. Rev. E*, 62(3):3096–3102, Sep 2000.
- [35] C.R. Meyer, J.L. Boes, B. Kim, P.H. Bland, G.L. LeCarpentier, J.B. Fowlkes, M.A. Roubidoux, and P.L. Carson. Semiautomatic registration of volumetric ultrasound scans. *Ultrasound med. biol.*, 25(3):339–347, 1999.
- [36] E. Miller. *Learning from One Example in Machine Vision by Sharing Probability Densities*. Ph.d. thesis, Massachusetts Institute of Technology, 2002.
- [37] J. K. H. Ng, R. W. Prager, N. G. Kingsbury, G. M. Treece, and A. H. Gee. Tomographic estimation of the point-spread-function of an ultrasound imaging system for deconvolution. In *Cambridge University Engineering Department Technical Report*, 2005.
- [38] Athanasios Papoulis and Unnikrishna S. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, December 2001.
- [39] F.C. Park and B.J. Martin. Robot sensor calibration: Solving  $ax = xb$  on the euclidean group. *IEEE Transactions on Robotics and Automation*, 10:717–721, October 1994.
- [40] P. Peetrons. Ultrasound of muscles. *European Radiology*, 12(1):35–43, 2002.
- [41] Xavier Pennec. Computing the mean of geometric features - application to the mean rotation. Technical Report RR-3371, INRIA, March 1998.
- [42] Xavier Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In *Proc. of Nonlinear Signal and Image Processing (NSIP'99)*, volume 1, pages 194–198, 1999.
- [43] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, July 2006. A preliminary appeared as INRIA RR-5093, January 2004.
-

- 
- [44] Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia-Antipolis, December 2006.
- [45] Xavier Pennec and Nicholas Ayache. Randomness and geometric features in computer vision. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'96)*, pages 484–491, San Francisco, Cal, USA, June 1996. Published in *J. of Math. Imag. and Vision* 9(1), July 1998, p. 49-67.
- [46] Xavier Pennec and Nicholas Ayache. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67, July 1998. A preliminary version appeared as INRIA Research Report 2820, March 1996.
- [47] G. P. Penney, J. M. Blackall, M. S. Hamady, T. Sabharwal, A. Adam, and D. J. Hawkes. Registration of freehand 3d ultrasound and magnetic resonance liver images. *Medical Image Analysis*, 8(1):81–91, March 2004.
- [48] Josien P. W. Pluim, J. B. Antoine Maintz, and Max A. Viergever. Mutual information based registration of medical images: A survey. *IEEE Trans. Med. Imaging*, 22(8):986–1004, 2003.
- [49] T.C. Poon and R.N. Rohling. Three-dimensional extended field-of-view ultrasound. *Ultrasound in Medicine and Biology*, 32(3):357–369, 2005.
- [50] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, October 1992.
- [51] F.M. Reza. *An Introduction to Information Theory*. McGraw-Hill, New York, 1961.
- [52] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80, 2000.
- [53] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Multimodal image registration by maximization of the correlation ratio. Research Report RR-3378, INRIA, August 1998. Published in MICCAI'98, Cambridge, USA, LNCS 1496, p. 1115-1124.
- [54] Joao M. Sanches and Jorge S. Marques. Joint image registration and volume reconstruction for 3d ultrasound. *Pattern Recogn. Lett.*, 24(4-5):791–800, 2003.
- [55] Y.C. Shiu and S. Ahmad. Calibration of wrist-mounted robotic sensors by solving homogenous transform equations of the form  $ax = xb$ . *IEEE Transactions on Robotics and Automation*, 5:16–29, February 1989.
- [56] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37:332–341, 1992.
- [57] Sunil Srinivasa. Multivariate mutual information. *University of Notre Dame*, 2006.
-

- 
- [58] C. Studholme and V. Cardenas. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recogn. Lett.*, 25(10):1191–1202, 2004.
- [59] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Incorporating connected region labelling into automatic image registration using mutual information. In *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 23, Washington, DC, USA, 1996. IEEE Computer Society.
- [60] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1):71–86, January 1999.
- [61] G.M. Treece, R.W. Prager, and A.H. Gee. Ultrasound compounding with automatic attenuation compensation using paired angle scans. In *Cambridge University Engineering Department Technical Report*, 2006.
- [62] Toru Tsujishita. On triple mutual information. *Adv. Appl. Math.*, 16(3):269–274, 1995.
- [63] Carole Twining, Tim Cootes, Stephen Marsland, Roy Schestowitz, Vladimir Petrovic, and Chris Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In Gary Christensen and Milan Sonka, editors, *19th International Conference on Information Processing in Medical Images (IPMI'05)*, volume 3565 of *Lecture Notes in Computer Science*, pages 1–14. Springer, 2005.
- [64] Carole Twining and Stephen Marsland. A unified information-theoretic approach to the correspondence problem in image registration. In *International Conference on Pattern Recognition*, volume 3, pages 704–709, 2004.
- [65] Carole Twining, Stephen Marsland, and Chris Taylor. Groupwise non-rigid registration: The minimum description length approach. In *British Machine Vision Conference*, 2004.
- [66] Behar V., D. Adam, and Z. Friedman. A new method of spatial compounding imaging. *Ultrasonics*, 41:377–384, 2003.
- [67] Tom Vercauteren, Aymeric Perchant, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Robust mosaicing with correction of motion distortions and tissue deformation for in vivo fibered microscopy. *Medical Image Analysis*, 10(5):673–692, October 2006.
- [68] Paul A. Viola. *Alignment by Maximization of Mutual Information*. Ph.d. thesis, Massachusetts Institute of Technology, 1995.
- [69] S. Watanabe. Information-theoretical analysis of multivariate correlation. *IBM J. Res. Develop.*, 4:66–82, 1960.
- [70] W. Wein, F. Pache, B. Roeper, and N. Navab. Backward-warping ultrasound reconstruction for improving diagnostic value and registration. In *MICCAI 2006 Proceedings*, Lecture Notes in Computer Science. Springer, October 2006.
-

- [71] W. Wein, B. Roper, and N. Navab. Automatic registration and fusion of ultrasound with ct for radiotherapy. In *MICCAI 2005 Proceedings, Lecture Notes in Computer Science*. Springer, October 2005.
  - [72] Michael Ying and Man-Hong Sin. Comparison of extended field of view and dual image ultrasound techniques: Accuracy and reliability of distance measurements in phantom study. *Ultrasound in Medicine & Biology*, 31:79–83, Jan 2005.
  - [73] Jie Zhang and Anand Rangarajan. Multimodality image registration using an extensible information metric and high dimensional histogramming. In *IPMI*, pages 725–737, 2005.
  - [74] Lilla Zöllei. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images*. Ph.d. thesis, MIT; MIT-CSAIL, 2006.
  - [75] Lilla Zöllei, E Learned-Miller, E Grimson, and WM Wells III. Efficient population registration of 3d data. In *ICCV*, 2005.
-