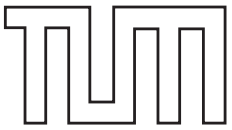
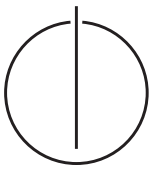


FAKULTÄT FÜR INFORMATIK
TECHNISCHE UNIVERSITÄT MÜNCHEN



Multimodal Integration of Medical Ultrasound for Treatment Planning and Interventions

Wolfgang Wein

Dissertation

Lehrstuhl Univ.-Prof. N. Navab, Ph.D.
Computer Aided Medical Procedures

Multimodal Integration of Medical Ultrasound for Treatment Planning and Interventions

Wolfgang Wein

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. D. Burschka
Prüfer der Dissertation:	
	1. Univ.-Prof. N. Navab, Ph.D.
	2. Prof. D. J. Hawkes, Ph.D. University College London / UK

Die Dissertation wurde am 22. Juni 2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 28. September 2007 angenommen.

Abstract

Ultrasound is a popular, cost-effective and non-invasive medical imaging modality. If an ultrasound probe is equipped with a 3D position sensor, the acquired images can be obtained in their spatial context, a technique commonly denoted "3D freehand ultrasound".
Combining ultrasound and a pre-operative Computed Tomography (CT) scan of the same patient can be beneficial for a number of clinical applications. The core of this thesis is the development of novel methods for fully automatic alignment (i.e. registration) of 3D freehand ultrasound and CT data, based on the image content and the physical properties of both modalities.
In particular, a new method is proposed which simultaneously optimizes the linear combination of different ultrasonic effects simulated from CT, and the parameters for spatial alignment. This original concept allows local as well as global optimization of the simulation parameters, resulting in optimal registration of any modalities, where usual registration solutions do not succeed and explicit simulation of complex effects are necessary.
Furthermore, we introduce new techniques for 3D freehand ultrasound calibration and reconstruction, as well as visualization of fused CT and ultrasound data.
Two clinical applications are investigated in detail. We use a designated version of an automatic registration algorithm to integrate diagnostic ultrasound into radiation treatment planning for head and neck cancer. Our simultaneous optimization of simulation and registration is applied, in the context of treating liver and kidney metastases, for fusing CT with both diagnostic and interventional ultrasound of the abdomen. While diagnostic fusion helps doctors to assess indeterminate lesions in those organs, interventional fusion using our techniques allows for advanced image-guided navigation, in particular, for needle biopsies and radio-frequency ablations.

Zusammenfassung

Ultraschall ist ein weit verbreitetes, kostengünstiges und nicht-invasives bildgebendes Verfahren in der Medizin. Wird der Schallkopf mit einem 3D-Positionssensor versehen, so können die Aufnahmen in ihrem räumlichen Zusammenhang gewonnen werden, eine Technik die 3D-Freihand-Ultraschall bezeichnet wird.
Für bestimmte klinische Anwendungen ist es vorteilhaft, Ultraschallaufnahmen mit präoperativer Computertomographie (CT) desselben Patienten zu verknüpfen. Der Schwerpunkt dieser Arbeit liegt in der Entwicklung neuartiger Methoden, die einen vollautomatischen räumlichen Abgleich (auch: Registrierung) von 3D-Freihand-Ultraschall und CT-Daten ermöglichen, basierend auf dem Bildinhalt und den physikalischen Eigenschaften beider Bildverfahren.
Wir stellen insbesondere ein neues Verfahren vor, welches gleichzeitig eine Linearkombination von aus dem CT simulierten Ultraschalleffekten und den räumlichen Registrierparametern optimiert. Dieses neuartige Konzept ermöglicht sowohl lokale als auch globale Optimierung der Simulationsparameter für bestmögliche Registrierung beliebiger Modalitäten, wo Standardmethoden scheitern und eine explizite Simulation komplexer Vorgänge notwendig ist.
Darüber hinaus wurden neue Methoden für Kalibrierung und Rekonstruktion von 3D-Freihand-Ultraschall, sowie Visualisierung fusionierter CT-Ultraschalldaten entwickelt.
Zwei klinische Anwendungsfelder werden im Detail untersucht. Ein designiertes automatisches Registrierverfahren wird verwendet, um diagnostischen Ultraschall in die Strahlentherapieplanung von Kopf-Hals-Tumoren zu integrieren. Die Methoden zur gleichzeitigen Optimierung von Simulation und Registrierung erlauben die Fusion von sowohl diagnostischem als auch interventionellem Ultraschall des Abdomens mit CT-Aufnahmen, im Zusammenhang der Behandlung von Leber- und Nierenmetastasen. Während die Fusion in einem diagnostischen Kontext den Ärzten hilft, die Dignität suspekter Läsionen in diesen Organen genauer einzuschätzen, ermöglicht die Fusion bei der Intervention, basierend auf den vorgestellen Methoden, eine fortgeschrittene bildbasierte Navigation, insbesondere für Biopsien und Radiofrequenzablationen.

Acknowledgements

First of all, I would like to thank my advisor Nassir Navab, for his restless efforts to shape a proper scientist out of me, and so much more. Starting at his newly created CAMP chair at TU Munich was an incredibly exciting and great experience. In this context I would also like to thank all my friends and colleagues at CAMP+AR, most importantly the first generation Jörg, Tobias, Marco, and all Martin's (yes, all of you), who helped anytime and anywhere with anything, and who, together with Nassir, made it the certainly most collaborative and fun academic environment on the planet.
I was privileged to work with a number of great Diplomarbeit students in the past years, including Diego Vivancos, Peter Lücke, Razvan Chisu, Fabian Pache, Darko Zikic, Oliver Kutter, Moritz Blume and Christian Wachinger.
Many thanks to Barbara Röper, who is responsible for most of the medical knowledge I acquired during the course of this work. Her incredible patience and devotion to our research, despite the overworked clinical environment, have both helped and impressed me very much.
I am very indebted to Frank Sauer, head of the Imaging & Visualization department at Siemens Corporate Research. He arranged that my work in Munich was supported by SCR, and later encouraged me to join his team in Princeton, making sure that I could seamlessly continue my thesis work. Besides, I am really grateful for his understanding and support in balancing work with private matters.
Ali Khamene, one of SCR's brightest researchers, has accompanied me ever since the first time I came to SCR for an internship more than four years ago. I would like to express my gratitude for everything that I learned from him, as well as the time, insight and fun we have shared.
Many other colleagues from SCR have provided valuable input for this work, including Chenyang Xu, Christoph Gütter, Yiyong Sun and Hesam Najafi.
Last but not least, big thanks go to all my wonderful friends both in Munich and Princeton, as well as my family.

- Princeton, June 2007

Contents

List of Figures

1. Introduction

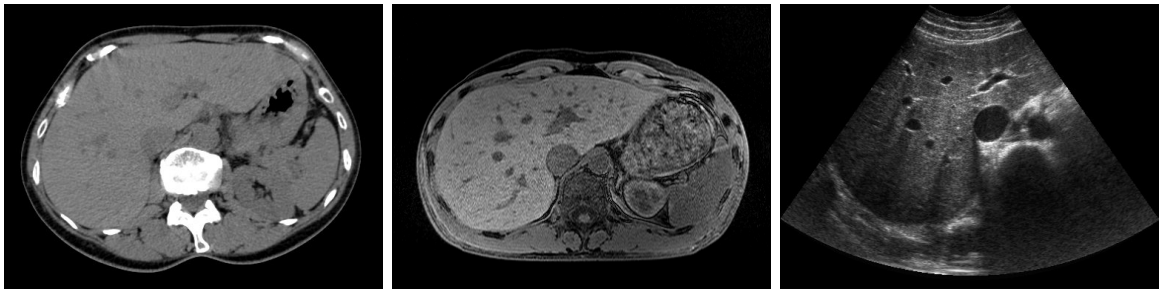
Healthcare is in the middle of an important and exciting transition. In the past, most medical imaging devices have been regarded as separate sources of information, integrated only by written documents about patient history, pathology and treatment, and inherently in the mind of the clinicians. While the oldest imaging modality, X-Ray, used to be a completely analogue device, most other imaging techniques have always relied to some extend on electronics and computers. Today, virtually all imaging is available in a digital representation, and the possibilities arising there have by no means been fully exploited yet. Hospitals are aiming at converting to integrated digital patient records combining all written documents with all available imaging. The now available networking bandwidth allows the seamless transfer of large three-dimensional patient images not only within a hospital, but on a global scale. This has dramatic consequences, and has for instance lead to the whole new market of Tele-Radiology.
Every single medical imaging technology has enjoyed tremendous improvements over the last decades, based on research in physics, electrical, biological and mechanical engineering, and computer science. Not only the quality of the depicted information has increased, but often also its applicability, cost effectiveness and, minimal-invasiveness has improved. Many modalities are now able to provide three-dimensional and four-dimensional information (i.e. 3D imaging over time). This is also changing the way images are presented and interpreted. For an increasing number of applications, 3D- and 4D-visualization techniques are used instead of traditional representations, like cross-sectional 2D slice images that are still predominantly used in radiology. High-quality digital displays are emerging from radiology reading rooms into interventional settings and even portable devices. Instead of looking at X-Ray films and side-by-side CT slice images on a back-lit panel in the operating room, surgeons can now use any appropriate three- and four-dimensional representation of image data, combined with planning data and live interventional imaging.
This combination of multiple imaging data of the same patient, acquired at different times and by different modalities, is termed *Multi-Modal Fusion*. Putting together pre-operative 3D anatomical imaging and 3D ultrasound, is the focus of this dissertation.

1.1. Medical Imaging

1.1.1. Anatomical Imaging

X-Ray

The patient is exposed to a small amount of ionizing electromagnetic radiation, generated by an X-Ray tube. A detector panel on the other side of the patient creates a 2D image of the incoming radiation. Therefore an X-Ray image, or radiograph, is a projection image of the X-Ray attenuation along the lines from the source through the patient to the detector. X-Ray Fluoroscopy is a variation, where the X-Ray system is imaging for a longer period



(a) CT slice (b) MRI slice (c) Ultrasound image

Figure 1.1.: Comparison of three modalities as 2D slice images of a liver

of time with less dose than for acquiring static X-Rays, such that live imaging is obtained, e.g. for guiding catheter interventions. Many of the technological improvements in the recent years have been centered around replacing the image intensifiers, which are detecting and amplifying the remaining X-Ray radiation behind the patient, by superior (however also way more expensive) flat detectors. Besides, the reconstruction of 3D images is possible if an adequate system, usually a C-arm, is slowly rotated around the patient, acquiring 200-400 individual X-Ray images.

Computed Tomography (CT)

Here X-Ray physics are used to create three-dimensional images. A CT scanner has an X-Ray tube mounted on a rotating gantry, one or several line detectors are on the opposite side. Sophisticated mathematical reconstruction techniques allow to create cross-sectional images of X-Ray attenuation inside the human body. During a CT scan, the patient couch is slowly moved through the gantry, and the scanner acquires a stack of axial cross-section images, comprising a three-dimensional dataset. Current scanners have up to 64 detector lines, and the gantry rotates so fast that four-dimensional data sets of the human heart can be acquired. Both for X-Ray and CT, the patient's vasculature can be emphasized by injecting a fluid with high X-Ray attenuation, a so-called *Contrast Agent*, prior to imaging. The overall imaging is then denoted *Angiography* or Computed Tomography Angiography (CTA).

Magnetic Resonance Imaging (MRI)

In contrast to the X-Ray based imaging techniques above, MRI does not use any ionizing radiation, and is therefore considered a non-invasive modality. The magnetic nuclei (mostly protons) of the anatomy are aligned in a strong uniform magnetic field generated by the scanner. They absorb energy from tuned radiofrequency pulses, and emit radiofrequency signals as their excitation decays. These signals, which vary in intensity according to nuclear abundance and molecular chemical environment, are converted into sets of tomographic (selected planes) images by using field gradients in the magnetic field, which permits 3-dimensional localization of the point sources of the signals.

Ultrasound

Ultrasound is substantially different from the tomographic modalities. Unlike CT or MRI scanners, an ultrasound system is a portable device, and it allows for interactive real-time imaging. In its basic operation, it acquires 2D slice images from within the patient, depicting the amount of ultrasonic reflections (figure 1.1(c)). Since this work is centered around the use of ultrasound, we provide some more detailed information about this modality in section 1.2.

1.1.2. Functional Imaging

The modalities described so far fall within the category of anatomical imaging. On the contrary, functional imaging includes techniques which measure biological and physiological processes. In the field of *Nuclear Medicine*, radioactive tracers are injected into the patient to be examined. They accumulate in certain tissue due to biochemical reactions, often they are used to label malignant tumors within the body. A nuclear imaging device like *Positron Emission Tomography* (PET), *Single Photon Emission Computed Tomography* (SPECT) or hand-held devices like Beta- and Gamma-probes measure the activity originating from within the patient's body after a given uptake time.
Some of the anatomical imaging techniques can be modified in order to perform functional imaging as well. In *functional Magnetic Resonance Imaging* (fMRI), a designated MR pulse sequence detects different levels of blood oxygenation, and can therefore be used to image brain activity. MR studies after injection of certain contrast agents, and MRI *Diffusion Tensor Imaging* (DTI) are related techniques. Regarding ultrasound, there is active research in developing functional contrast agents for ultrasound imaging, which accumulate at tumor sites and hence increase their echogeneity.

1.2. Ultrasound

1.2.1. Basic Operation

A hand-held transducer is put in direct contact with the patient's skin, using some coupling gel to avoid any air in-between. The transducer emits short ultrasound pulses with a frequency of 1-15 MHz. As the sound pulses propagates deeper into the body, part of them are reflected by the anatomic structures. The transducer records and amplifies the received echoes. In B-Mode imaging, those echoes are converted into brightness values and displayed on the screen as a 2D image, representing a slice of anatomic information pointed from the transducer downwards into the patient.
Both for creating the necessary mechanical vibrations and converting received echoes to electrical signals, an array of transducer elements is used. These elements are denoted the *aperture* of an ultrasound system. They were exclusively piezoelectric elements for many decades, but recently new silicon-based technology, so-called Capacitive Micro-machined Ultrasonic Transducers (CMUT) is evolving, potentially enabling to manufacture ultrasound systems with smaller transducer size (some of the signal processing can be done on the transducer itself), higher bandwidth, and flexible 2D-arrays for imaging [96].
In order to steer and focus the ultrasonic pulses for optimal imaging, neighboring transducer elements are controlled together with a certain delay pattern, a technique denoted as *beamforming* (figure 1.2). Derived from linear systems theory, the impulse response of imaging an

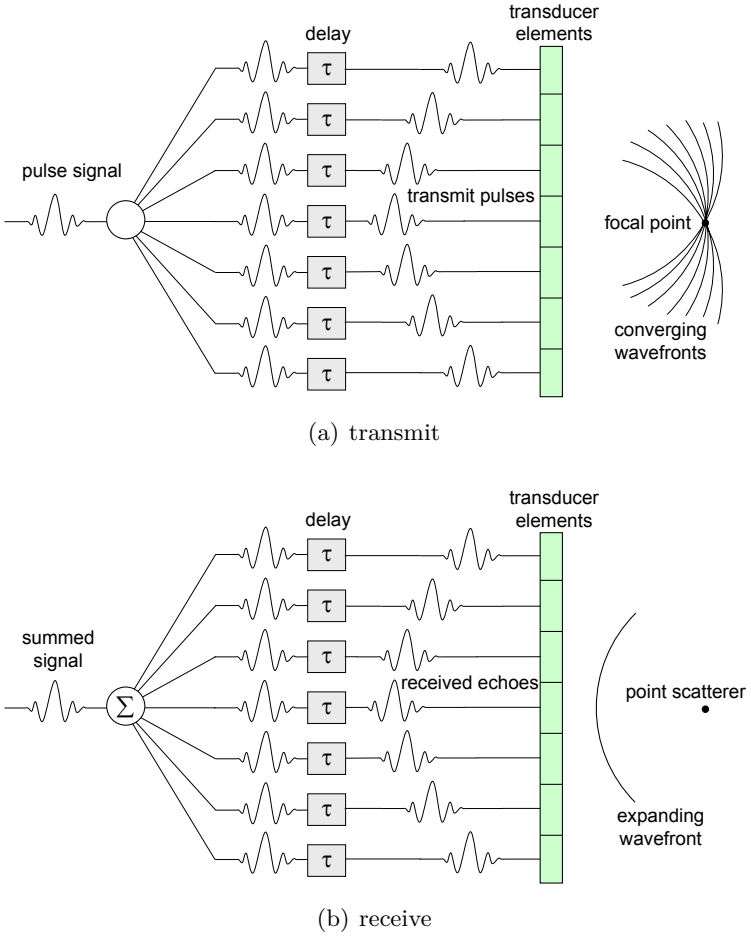
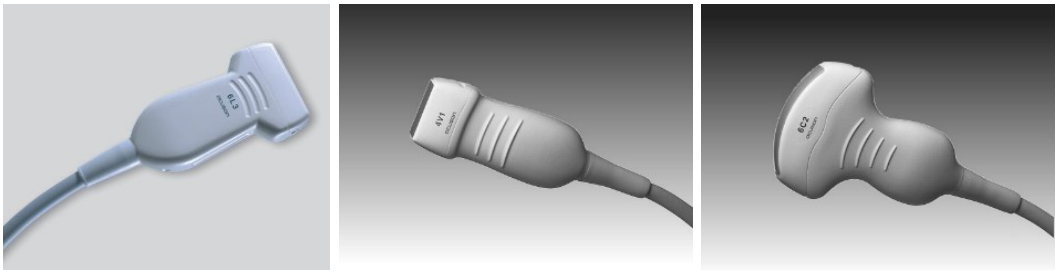


Figure 1.2.: Transmit and receive beamforming on a phased array ultrasound transducer.

ideal single target point is denoted as the Point Spread Function (PSF). Its size and shape ultimately determines the fidelity of an ultrasound image. The base frequency (and therefore in inverse relation the wavelength) and bandwidth of the used pulse sequence affect the PSF in axial direction, i.e. along the beam. Hence the resolution is higher for higher frequencies, but unfortunately the ultrasound pulse is attenuated proportional to the square of it. Therefore high-frequency transducers are used mainly for applications requiring imaging in shallow areas of the body (e.g. neck, muscles), while lower frequencies are better for acquiring larger images e.g. of the abdomen. The PSF's shape perpendicular to the beam, i.e. in the lateral and elevational directions, depends upon the beamforming.
For reconstructing the images, the ultrasound system assumes that the sound propagates through the human tissue with a constant velocity, and the received echoes are directly backscattered from point reflectors. Unfortunately, those assumptions are in reality often violated, causing a number of image artifacts whose severity depends on the particular patient and anatomy. Some of these artifacts provide useful information to the clinicians, while others just degrade the images. Most of the advancements made in recent years on ultrasound systems are targeted to reduce those unwanted image degradations.



(a) linear (b) phased (c) curvilinear

Figure 1.3.: Different 2D ultrasound transducer types.

Besides B-mode imaging, a number of *Doppler* imaging modes are usually available. Spectral Doppler retrieves the blood flow at a particular location as 1D-signal over time. Color Doppler displays the flow information within a selected region, overlaid on the B-Mode image.

1.2.2. Types of Systems

The most common types of 2D transducers are depicted in figure 1.3. A linear array transducer has a flat aperture and parallel beam-steering, therefore the images acquired are rectangular strips from the aperture down into the body (see figure 1.4 left). They are used for clinical applications that do not require a large field of view within the body, and often feature relatively high frequencies. A phased array transducer has a flat surface as well, but performs angulated beam-steering, therefore the imaged area within the body is larger. They are used primarily for anatomical sites where a larger imaging area is desirable while the access on the patient's skin is limited, e.g. for imaging the heart or kidney through the ribs. Curvilinear array transducers have a curved surface, and acquire images similar to phased arrays in terms of the geometry. However, since many more transducer elements can be used, higher quality images are obtained, given that its larger footprint is not an issue (e.g. for scans of the liver & abdomen).
Nowadays, most of the ultrasound vendors offer 3D transducers, which allow volumetric 3D and 4D (i.e. 3D volumes with real-time update) imaging. Two main techniques are used. *Wobblers* basically consist of a regular curvilinear array transducer that can be rotated (wobbled) around the lateral axis within the transducer housing. Depending on the quality settings, the acquisition of a single 3D volume can take $0.2 - 20$ seconds (4D imaging vs. acquisition of high-quality volumes, optionally with Doppler information). On the contrary, 2D array transducers do not require any mechanical system, but rather acquire 3D volumes by electronic beam steering in both the lateral and elevational direction. They are potentially better for 4D imaging, but due to the very large number of array elements required within the 2D grid, the imaging quality is somewhat limited. It is expected that the upcoming CMUT-based products will significantly shuffle around this market segment, since they will allow volumetric acquisitions with both higher frame rate and imaging quality.

1.2.3. Image Characteristics

The appearance of B-mode ultrasound images is composed by two main effects. Different acoustical impedance of human tissue causes strong reflection of the ultrasonic pulses. Therefore the interfaces between different anatomical structures and tissue types are highlighted as strong reflections. This can be seen in figure 1.1(c) for instance on the lower surface of the liver, at the bottom left of the image. Some of the hepatic vasculature in this image is tagged with a border as well, due to a fat layer surrounding the actual vessel. The other effect is caused by tissue inhomogeneities that are smaller than the ultrasound pulse wavelength. They cause seemingly random reflection and scattering, resulting in the typical ultrasound speckle patterns. While often just considered as useless noise by beginners, they actual carry significant information about the imaged tissue. Based on both the direct appearance of speckle and its dynamic change when slightly moving the probe, sonographers can draw conclusions about important tissue properties. For image processing, ultrasonic speckle is statistically modeled as Raleigh noise [159]. Examples for further ultrasound effects that convey information are refraction at tubular structures (causing the streaks downward from some of the vessels in figure 1.1(c)), and full occlusion behind bone surfaces (the spine in the lower right of the image). Aberration is an unwanted imaging artifact, caused by inhomogeneities mostly in the compressed fat layers on top of the image. Essentially, they randomly delay the individual ultrasound pulses, impairing the resolution of the lower part of the image.

1.2.4. 3D Freehand Ultrasound

If a 2D ultrasound transducer is equipped with a 3D position tracking system, every single ultrasound image can be regarded in its spatial context. This can be used to either construct three-dimensional information (potentially on a more global scale than with designated 3D-ultrasound devices, see figure 1.5 left for an example), or for establishing a 3D visualization of the live ultrasound image, e.g. for navigating certain clinical interventions. We use 3D freehand ultrasound throughout this work, since we require both. There is a number of choices to be made when setting up such a system, the most important one being the type of tracking system used to record the motion of the ultrasound probe. Besides, a calibration has to be performed to precisely know the position and orientation of the ultrasound image plane (which is basically invisible to the user) with respect to the tracking sensor or target. Further questions are how to present or process a set of arbitrarily placed ultrasound images.

1.3. Multi-Modal Fusion

The term multi-modal fusion describes the combination of two or more medical images of the same patient, acquired by different types of imaging systems. Different imaging often reveals different information about the same anatomy, hence providing complementary value. Combining the images in an appropriate manner can yield more information to the clinicians than just the “sum” of individual findings. In Nuclear Medicine, for instance, the combination of anatomical and functional imaging has become so essential that integrated PET-CT scanners are now generally used.
In the following we describe scenarios where 3D ultrasound is to be combined with a Computed Tomography of the same patient, for improving clinical diagnostics, treatment planning and interventions.

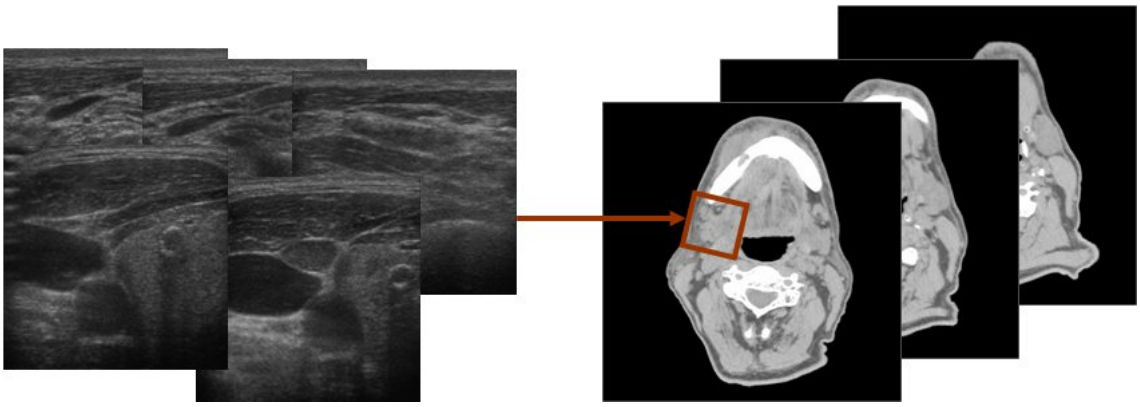


Figure 1.4.: Ultrasound images of neck lymph nodes (left) and their relation to the CT slices (right) used for radiotherapeutic treatment planning.

1.3.1. Clinical Need

Radiation Therapy

For inoperable tumors, *External Beam Radiation Therapy* is a technique where the malignant target volume is irradiated with a high-energy beam (typically $\approx$ 6 MV) generated by a linear accelerator (LINAC). To spare surrounding healthy tissue as much as possible, the target volume within the patient is aligned with the iso-center of the radiation device, and the irradiation is repeated with beams from a number of angles, always passing through this iso-center. The treatment planning involves the definition of target volumes and safety margins to assure that the malignant tissue can effectively be destroyed in spite of positioning uncertainty and breathing motion, as well as the definition of critical areas to be spared. It is done based on a CT-scan of the patient.
For some anatomical sites, in particular the head and neck, diagnostic ultrasound can reveal important information not obtained by CT, e.g. about the internal nodal architecture. To improve the accuracy of the treatment planning, it is therefore desirable to combine the high-resolution information of the small ultrasound images with the rather global anatomy of CT on which the planning is based upon, as illustrated by figure 1.4. In this context, three-dimensional ultrasound would add even more value, since small and difficult to localize lesions can be assigned more easily, given that the CT and ultrasound data is properly aligned.
Ultrasound could then also be used for patient positioning, i.e. the proper setup of the patient in the treatment room, such that the target volume is aligned with the iso-center of the Linear Accelerator. Extending upon this idea, the anatomy could be monitored during irradiation for breathing-induced motion and patient shift. The potential applications are limited though, as ultrasound can not be used in some of the anatomic sites primarily targeted by Radiation Therapy. Ultrasound imaging of the chest and head is not possible. Echography of the prostate is possible and actually used, but the resulting images are of limited quality and more difficult to interpret than for other organs. Radiation Therapy is currently not a treatment option for the liver due to its enhanced sensitivity to ionizing radiation, while ultrasound would provide excellent imaging there. It is however being introduced on an experimental basis using advanced respiration-gating techniques [31].

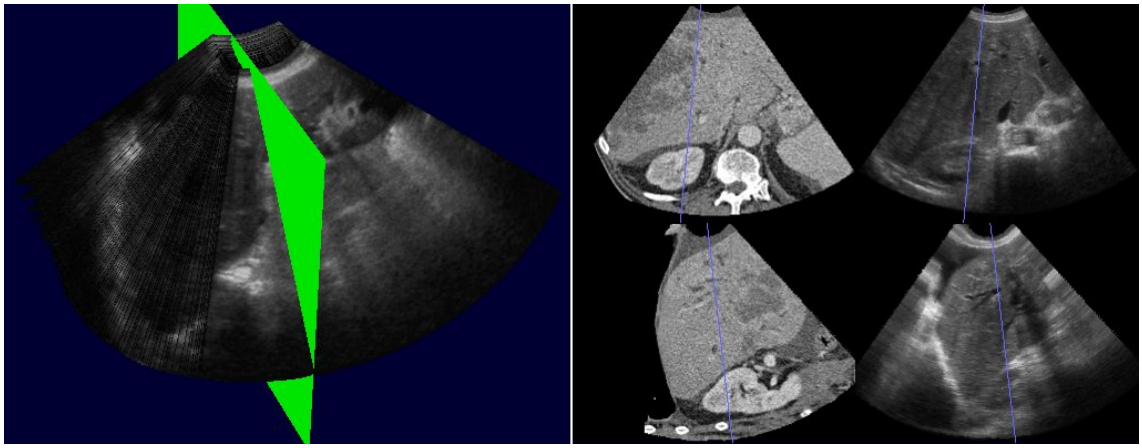


Figure 1.5.: Ultrasound and CT of a liver with widespread metastatic disease. The left side depicts the whole 3D freehand ultrasound sweep. On the right, two CT-Ultrasound planes are shown side-by-side. The top image represents an original recorded ultrasound plane from the sweep, at the bottom an oblique reconstruction from the ultrasound is shown, at the location illustrated by the green plane in the left image. The intersection of the respective other plane is drawn as blue line.

Diagnostic Fusion

Often suspicious masses are identified in patients using CT or MRI imaging. If they are indeterminate, additional imaging like Ultrasound or PET can reveal the missing clues to conclude if it is a benign or malignant tumor mass. As last resort, a tissue sample from the volume in question is obtained using a biopsy procedure. Here, a special needle is inserted into the patient, and a mechanism at the needle tip carves out a small tissue sample. Then the needle is retracted, and the sample is transferred for pathological examination.
Examining the CT combined with ultrasound can reveal more information compared to separate reading of the images, potentially sparing a biopsy. Figure 1.5 shows an example for metastases in the liver.

Interventional Navigation

Radio-Frequency Ablation is one of the treatment options for liver and kidney cancer. One or several needle electrodes are inserted into the patient, depending on the number and size of the lesions. Electrical current is induced between the needles and a grounding pad attached to the patient's skin, such that the malignant tissue is essentially burned (coagulated). The insertion of the needles has to be guided by interventional imaging, to ensure their accurate placement within the lesions. Mostly CT and ultrasound are used. With CT, the intervention is performed in a CT room with the patient outside of the bore of the scanner. Occasionally, the patient couch is moved into the scanner, and a small number of slices are acquired to verify the current needle location. The patient is moved out of the scanner, then the needle(s) are further advanced. Those steps are repeated until the tip of the needles reach the planned target. If the insertion is guided by ultrasound, the physician places the ultrasound transducer arbitrarily to obtain images, whenever necessary, of both the lesion and the needles (figure

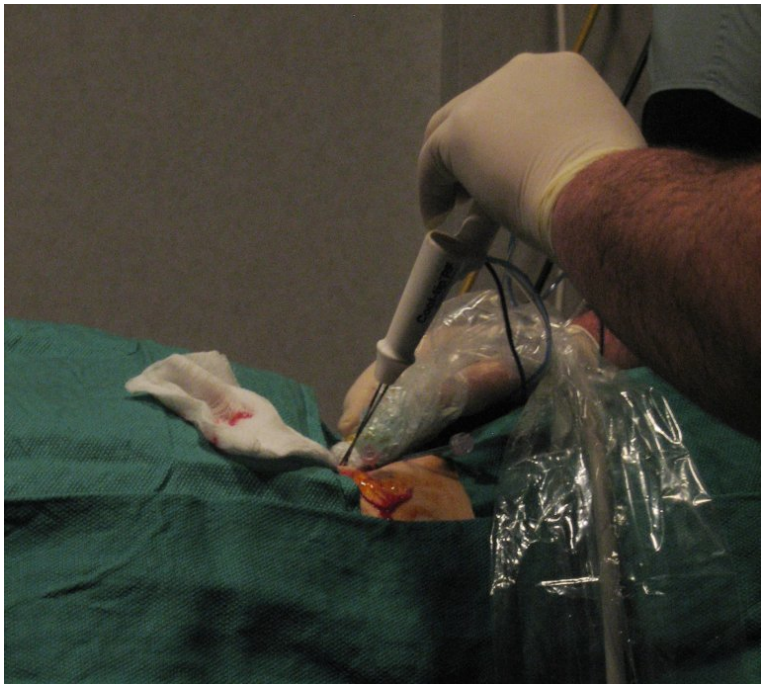
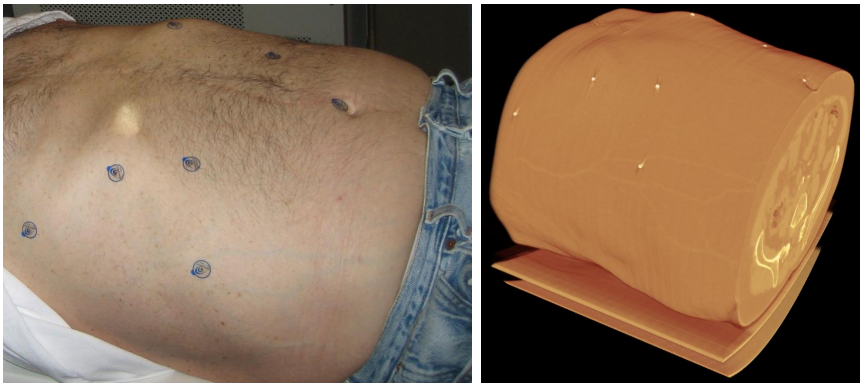


Figure 1.6.: Ultrasound-guided Radio-Frequency Ablation of a kidney. Two electrodes are inserted in the patient, the ultrasound transducer is visible behind, enclosed in a sterile plastic wrap.

1.6). If a single RF electrode is sufficient, a needle guide mounted on the ultrasound transducer can be used instead. Since the needle path is rigidly connected to the image location, it can be outlined right on the ultrasound image. Guidance is then achieved by lining up the lesion with the shown needle path in the ultrasound image and advancing the electrode until it reaches the lesion. A related treatment technique is Cryo-Ablation, here the tumor tissue is basically "frozen to death". Some research-oriented clinics use interventional MRI to guide those procedures. While a promising approach, it is not in widespread use yet due to the extremely high cost of such a system, and limited image quality & patient access. It is controversial which of the imaging modalities is best for guiding RF- and Cryo-Ablations. Many interventional radiologists prefer ultrasound, since the imaging can be done whenever needed, and is very cost-effective. However it requires profound expertise in sonography, and hepatic lesions might be not well visualized unless contrast-enhanced ultrasound is used. CT-based procedures take longer, since the needle insertion has to be interrupted for imaging; this not only makes the treatment more expensive, but also requires the patient to be sedated longer, if general anesthesia is used.
The pre-operative contrasted CT scan is used to plan those procedures, in terms of optimal needle access, hepatic vasculature that must not be ruptured, and overall needle placement to fully ablate all lesions. If tracked ultrasound is used during procedure, and - again - CT and ultrasound are correctly aligned, both oblique slices from CT and planning information can be shown in real-time. This can provide a tremendous advantage during execution of those procedures, potentially revolutionizing the way they are performed. In fact, a future workflow



(a) Photo (b) CT volume rendering

Figure 1.7.: Radio-opaque markers attached to a patient's skin

for RF-Ablation might resemble more the way Radiation Therapy is done nowadays; it would include very sophisticated treatment planning. Using intra-operative ultrasound and proper alignment with the CT, the plan can then be implemented precisely.

This application belongs to the discipline of interventional radiology, where navigation is often achieved using registered pre- and intra-operative images. A broad review about this field of Image-Guided Surgery or Image-Guided Intervention can be found in [106].

1.3.2. Challenges

Registration

For the scenarios described above, CT and ultrasound have to be precisely aligned within a common coordinate system, such that the anatomy can be correctly superimposed from both modalities. This problem is commonly referred to as *Registration*, which has become a fairly large field of research.
A straightforward method to achieve registration in the described scenarios uses small fiducial markers attached to the patient's skin (figure 1.7). Those radio-opaque markers are visible in the CT scan, and can be located with a calibrated, tracked pointer. By computing the 3D rigid motion of the corresponding points from CT to the tracked points, the CT data is registered to the tracking coordinate system, which is also used by the tracked ultrasound probe. A drawback is that the CT scan has to be acquired right before the freehand ultrasound exam or the intervention - which is often not possible depending on the clinical setting. Besides, the quality of the alignment will be limited within the patient, since the registration is based on skin markers.
Alternatively, manual registration can be performed by the physician. One or several ultrasound images with tracking are recorded, then the CT reconstruction planes are moved and re-oriented such that the anatomy lines up. Here the tracking of the ultrasound transducer can serve as a convenient user input device with the necessary 6 Degrees-Of-Freedom (DOF) for translation and rotation. However, it is very cumbersome and challenging to achieve a good registration based on visual alignment of 2D planes. On one hand, it might be feasible

to align a single plane reasonably quick. In practice, an image that contains as much uniquely identifiable structures as possible is used, for abdominal applications e.g. an umbilical plane (producing a large shadow in the ultrasound image). That does by no means guarantee that different planes are correctly registered as well, since there can be multiple ambiguous alignments of a 2D plane within a 3D volume. On the other hand, it is difficult and time-consuming to manually align multiple oblique ultrasound images within the CT volume. Last but not least, yet another option for manual registration is to define anatomical landmarks in both modalities, which will serve as internal fiducial markers for computing the spatial mapping. Those landmarks have to be precisely locatable in 3D. This both limits the possible structures to be used (e.g. vessel bifurcations or small lesions), and poses additional requirements for the user interface - optimally the CT/Ultrasound data should be presented with two or three oblique planes to ascertain the true three-dimensional location of the landmarks.
It would be an enormous advantage to have an automatic method that computes the registration between CT and ultrasound images. A lot of the additional effort required for registration could be spared, improving the clinical workflow and therefore increasing the acceptance of CT-Ultrasound fusion imaging amongst clinicians. Given the unique properties of ultrasound images and the very different representation of anatomy compared to CT, one can see that this is not going to be easy. Automatic image-based registration of CT and ultrasound is the core topic of this dissertation. We will introduce methods that bring the two modalities closer together, and devise new mathematical means to compare them. Some of the specific artifacts of ultrasound imaging can be turned from being a hassle to providing important clues for registration.

Fusion

Once the spatial relationship between CT and tracked ultrasound is established, both modalities have to be presented in a way such that their fusion provides meaningful information for a given clinical application. This is only possible with a profound understanding of the clinical practice without multi-modal fusion. We will discuss the actual clinical workflow for a number of applications, and make suggestions on how to adapt it to integrate CT-ultrasound fusion.
Visualizing CT and freehand ultrasound is not a trivial issue either, since both are inherently three-dimensional, and each of them shows a very different representation of the same anatomy. Therefore a simple overlay will not be sufficient in general. An arsenal of existing and new techniques can be used, including showing the live or recorded ultrasound in its 3D context with adapted volumetric visualization of CT, using a number of different superimposition methods, and computing and displaying oblique cross-sections from both modalities. Besides, additional planning information can be shown as well. We will propose specific visualization techniques for every investigated clinical application.

1.4. Contributions

Following is a summary of the technical contributions in this thesis, along with the corresponding publications:

- We have developed new methods for image-based registration of CT and ultrasound data, and applied it to integrate diagnostic ultrasound of the neck with the planning CT

for Radiotherapy [169, 120]. Particular focus has been the derivation of an appropriate similarity measure, taking the physical properties of ultrasound, and its artifacts into account. An overall workflow that includes the freehand ultrasound acquisition, global pre-registration based on the head surface, and validation, is described in detail in the journal publication [170].

- Continued research resulted in a technique to simulate ultrasound from CT in real-time. This is combined with a novel similarity measure that assesses the correlation of a combination of multiple signals extracted from CT (using the simulation) with ultrasound. Together, it comprises a framework for simultaneous optimization of simulation and registration parameters, serving as the foundation of a fully automatic registration, which aligns a freehand ultrasound sweep with the corresponding 3D modality using a rigid or an affine transformation model [166]. Those methods have been validated on abdominal CTA and ultrasound imaging of 10 patients, within the context of a diagnostic fusion study.

- In order to reconstruct the ultrasound B-Mode slices scattered in space into rectilinear 3D-volumes, we have devised an efficient algorithm for spatial compounding, based on a backward-warping strategy [167]. It also enables one to compute arbitrary MPR slices directly from the freehand ultrasound slice set, without the need of an extra volumetric reconstruction step.

- Building upon the spatial compounding technique, we also developed a method for freehand ultrasound calibration based on the acquisition of one or two sweeps with interleaved information [165]. The calibration parameters are estimated such that the similarity of a number of ultrasound slices with reconstructions computed from perpendicular slices becomes maximal. This calibration can be applied in-vivo on the patient, no designated calibration phantom is required.

Besides, I was involved in research efforts regarding the fusion of intracardiac ultrasound with C-Arm CT for electrophysiology [65]. We also investigated methods for mono-modal registration of 3D ultrasound data sets for various application scenarios. This includes the use of variational methods for deformable registration of abdominal 3D ultrasound volumes [183], and the stitching (or "mosaicking") of overlapping 3D ultrasound volumes using multivariate registration techniques [156]. In further work we developed and applied a blind shift-variant deconvolution method to create ultrasound reconstructions of superior quality from a number of acquisitions taken at different angles [17]. A novel visualization paradigm for multi-modal imaging is introduced in [24], where contextual cutaways within a CT volume rendering allow a clear yet informative view on a live tracked ultrasound image.
At the beginning of my Ph.D. studies, I had followed up research on 2D-3D registration (based on the subject of my diploma thesis [163], see also [70]). This resulted in a novel method for 2D/3D registration based on volume gradients [168], it also describes the parallel implementation used. Inspired by the advantages of 2D/3D registration, we came up with a projection-based 3D-3D registration technique that uses GPU acceleration, and explores the possibility of computing image similarity metrics on a graphics processor [71]. Last but not least, we developed a technique to incorporate voxel-wise quality information into a 3D-3D registration algorithm for optical tomography volumes [164].

Some of the technology described in this document has been protected by Siemens Corporate Research Inc., please contact me for further details.

1.5. Outline

Chapter 2 provides a comprehensive technical introduction to image-based registration. The underlying mathematical concepts in particular of image similarity measures are described in detail. Practical issues for anyone who is to develop an automatic registration method, are addressed as well.
In **chapter 3**, 3D freehand ultrasound systems are introduced in terms of the required hardware, system setup, calibration and reconstruction issues. The newly developed methods for image-based calibration and efficient backward-warping compounding can be found here as well.
Then, **chapter 4** describes the novel methods for automatic registration of freehand ultrasound to CT, constituting the core contribution of this thesis.
The clinical applications are described in **chapter 5**, including the evaluation of the new registration methods within their context.
Chapter 6 concludes the work by summarizing the outcome, discussing benefits and drawbacks, and pointing at future work.
The appendices each present a self-contained publication or internal report, which is not within the core research subject, but still is closely related in terms of the methodology.

2. Image-Based Registration

Medical Image Registration has become a huge scientific field, with research efforts by many interdisciplinary groups all over the world. It is about establishing a common coordinate system for a number of different medical images. In fact, it seems to be a rather special meaning of the english word “register”, as can be seen in figure 2.1. The dictionary entry for “registration” (left) does not contain anything related to the topic we would like to address. Only after digging further into "1) the act of registering" (right box), we find that item 2. matches to what we call registration here.
There are many reasons why one would like to register medical images, e.g. to create anatomical atlases, for follow-up studies by matching images of the same patient taken at different times, to enable interventional guidance by integrating pre-operative with intra-operative imaging, and many more. Some clinical scenarios that are the focus of this thesis have been pointed out in the introduction.
The book [53], as well as chapter 8 of [143] provide a detailed introduction to Medical Image Registration for a novice reader. An overview article proposing a classification of registration techniques into nine criteria is [86]. A later literature survey can be found in [58]. We were unable to find any surveys after 2001, which might be due to the explosion of literature in that field (which has also been pointed out in [108], 2003).

From a technical standpoint, an important distinction is the underlying information used in the registration process.
Feature-based methods use indirect information extracted from the images beforehand, like point sets or surfaces. Aligning those structures then allows one to align the images as well. If a set of corresponding points is determined, using e.g. skin surface markers, markers implanted into the patient, or manually selected anatomical landmarks (see also section 1.3.2), the rigid motion between the two point sets can be computed with a closed-form solution [152]. If surfaces are extracted, they can be matched with the famous Iterative Closest Point (ICP) algorithm [179]. In both cases, often one of the information entities arises not from an image, but from the actual physical patient (e.g. when establishing points or surfaces using pointers or laser-scanners before an intervention).
Image-based methods, on the contrary, use the images themselves rather than an indirect representation to compute the registration. The synonymous terms *Intensity-based* and *Voxel-based* registration are often used as well. All terms imply that the image content is directly used to compute the correct alignment, however a certain amount of pre-processing can be performed on the images. This category of registration techniques if often associated with automatic algorithms. This is not always true, since defining an initial alignment, a region-of-interest (ROI), or other parameters, might require manual interaction. In this context, registration methods are often also denoted *semi-automatic*. On the other hand, feature-based methods can be automatic as well, given that the feature extraction step is automatic. This thesis centers around the use of image-based registration to integrate ultrasound with other modalities, with a certain degree of automation. Therefore we will introduce the math-

Main Entry:
reg · is · tra · tion
Pronunciation:
`"re-j&-'strA-sh&n`
Function: noun
1. the act of registering
2. an entry in a register
3. the number of individuals registered : ENROLLMENT
4. a) the art or act of selecting and adjusting pipe organ stops b) the combination of stops selected for performing a particular organ work
5. a document certifying an act of registering

Main Entry: [2]**register**
Function: verb
Inflected Form(s): **reg · is · tered**; **reg · is · ter · ing**
`/-st(&-)ri[ng]/`
transitive verb
1. a) to make or secure official entry of in a register b) to enroll formally especially as a voter or student c) to record automatically : INDICATE d) to make a record of : NOTE e) PERCEIVE; also : COMPREHEND
2. to make or adjust so as to correspond exactly
3. to secure special protection for (a piece of mail) by prepayment of a fee
4. to convey an impression of : EXPRESS
5. ACHIEVE <registered an impressive victory>

intransitive verb
1. a) to enroll one's name in a register <registered at the hotel> b) to enroll one's name officially as a prerequisite for voting c) to enroll formally as a student
2. a) to correspond exactly b) to be in correct alignment or register
3. to make or convey an impression

Figure 2.1.: Entries about *registration* and *register* from the Merriam-Webster Online Dictionary (www.m-w.com)

ematical and technical foundation of image-based registration, as well as practical issues, in the remainder of this chapter. The main focus is on describing multi-modal registration techniques that use a global transformation model (typically rigid), rather than local deformable models (which is a large sub-field of research termed *Non-linear registration*).

2.1. General Formulation

In the following we model Images as scalar functions defined on vectors:

$$I : \Omega \to \mathbb{R} \tag{2.1}$$

with $\Omega \subseteq \mathbb{R}$ being the region where the image I is defined (i.e. the *domain*). At the position $\vec{x} \in \Omega$, the Image has the intensity $i = I(\vec{x})$. Typically, $\vec{x}$ denotes a location in a two- or three-dimensional cartesian coordinate system, in physical units like mm or voxel indices. This will be the case for CT and MRI images, mostly for 2D and 3D ultrasound as well. However, depending on the dimensionality and type of images used, every component of $\vec{x}$ can carry a different physical meaning. For example, four-dimensional images are often defined on three spatial coordinates and time. The two coordinate components of an X-Ray projection denotes the 2D position on a detector plane, however the image intensity can not be associated with one physical location, since it is an integral over the X-Ray attenuation along the line from the X-Ray source to the detector. If ultrasound images are available as so-called scanline data, $\vec{x}$ might refer to a location in a curvilinear coordinate system, which has to be converted to cartesian coordinates before presentation on a screen.
Medical imaging devices in fact only provide measurements at a discrete number of locations

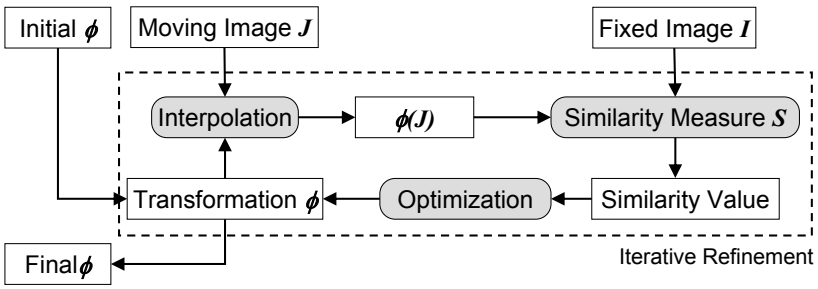


Figure 2.2.: Iterative scheme of an image-based registration algorithm.

$\{\vec{x_k}\} \subset \Omega, k = 1 \dots N$. For cartesian cordinate systems, those sample locations often represent an equal-spaced grid in two or three dimensions, N is the total number of measurements (aka. pixels for 2D images, voxels for 3D). However, a large number of image-processing algorithms (including image registration) needs to evaluate the image function at non-grid locations. Therefore, whenever $I(\vec{x})$ is queried with $\vec{x} \notin \{\vec{x_k}\}$, some sort of interpolation is necessary (section 2.4). The assumption that interpolation is implicitly performed when required, allows us to stick to the definition of images on a continuous region Ω. Likewise, the measured image intensities are only available as a discrete number of values (e.g. as 12 Bit = 4096 values for CT and MRI, 8 Bit = 256 values for most ultrasound systems), rather than the set of real numbers $\mathbb{R}$. For a theoretical discussion of image registration, incorporating the discreteness of I does not provide any benefit, though. Besides, the results of interpolation are usually floating-point numbers with 32 or 64 Bit machine precision that can be approximately treated as $\mathbb{R}$.

If two images I and J are to be coregistered, we are searching for the parameters of a transformation ϕ, that correctly maps every point $\vec{x} \in \Omega_I$ of the Image I onto a point $\phi(\vec{x})$ in J, such that the alignment of the imaged anatomy is correct. Here we have assumed without loss of generality, that I is used as *fixed* image, since it's original imaging area is used. J is then the *moving* image, because it's content is evaluated at the transformed locations. Another synonymous terminology declares I as *reference* and J as *template* image.
An image-based registration algorithm seeks the optimal parameters of ϕ, such that a *Similarity Measure* S between the two images is maximized:

$$\phi' = \arg\max_{\phi} S\left(I, \phi(J)\right) \tag{2.2}$$

Here we use the symbol ϕ in an additional meaning, namely as an operator which transforms the whole image J. The Similarity Measure assesses the quality of anatomical alignment between the two modalities, and is a function of the image gray values. This prohibits elegant closed-form solutions that can often be defined for feature-based methods. Therefore, the registration is usually an iterative process, as outlined in Figure 2.2. The following sections describe the components of an image-based registration algorithm in further detail. In 2.2, the commonly used Similarity Measures S are introduced, 2.3 describes possible spatial transformations ϕ and their parameterization, interpolation is treated in 2.3 and 2.5 lists numerical optimization algorithms that can optimize those parameters.

2.2. Similarity Measures

The best choice of a similarity measure S depends heavily on the underlying images to be registered, however some general requirements can be stated. A similarity measure should have its global maximum when the two images are correctly aligned. It should also increase smoothly towards this optimum for slightly wrong transformation parameters, as the optimizer iteratively refines ϕ based on the Similarity Measure. Besides, S should be robust with respect to noise in the images, as well as outliers. The latter is typically some image content that is only present in one of the images, and hence will not line up even for a correct alignment of the images.
In sections 2.2.2 and 2.2.3 we will first restrict ourselves to Similarity Measures that only evaluate tuples of image intensities, regardless of their location in the images. The general registration formulation in equation 2.2 can then be rewritten as

$$\phi' = \arg\max_{\phi} S\left(\{i_k, j_k\}\right) \text{ with } i_k := I(\vec{x_k}); j_k := J(\phi(\vec{x_k})) \tag{2.3}$$

For simplifying the notation, we define j_k as the intensity of image J at the *transformed* location $\phi(\vec{x_k})$. Within the context of similarity measures for registration, there is usually no need to evaluate J at its original voxel locations. Whenever the original voxel grid positions of J are used, we will explicitely state it.
The maximization of a similarity measure is equivalent to the mimimization of an error metric, which expresses the dissimilarity of two images. Similarity can be transformed to dissimilarity in many ways, for instance by using its negative value $-S$ or computing its inverse $1/S$.

2.2.1. Images as Random Variables

A convenient representation is to treat the actual image intensities as random variables. This allows one to apply many concepts from statistics and probability theory. The basic notation assigns a probability to an image having a particular intensity:

$$P(I{=}i) = P_I(i) \tag{2.4}$$

Without assigning a relationship to the images, those probabilites can be estimated from a given image just by using the actual occurences of each intensity. Typically, the image intensities are treated as discrete values. If necessary, some scaling and rounding can be applied, resulting in so-called *binning* of the values. The probability of an image intensity is then the number of its occurences divided by the total number of samples:

$$P_I(i) = \frac{N_i}{N} = \frac{|\Omega_I^i|}{|\Omega_I|} \text{ with } \Omega_I^i = \{\vec{x_k} | I(\vec{x_k}) = i\} \tag{2.5}$$

Ω_I^i is the iso-intensity set of image I for the intensity value i. That describes the probability density function (PDF) of the random variable I, estimated from the image histogram. For image registration, the joint probability of I and J is often of importance:

$$P_{IJ}(i,j) = P(I{=}i, J{=}j) = \frac{|\{\vec{x_k} | I(\vec{x_k}) = i \wedge J(\phi(\vec{x_k})) = j\}|}{|\Omega_{IJ}|} \tag{2.6}$$

2.2.2. Maximum-Likelihood Formulation

Various researchers have proposed to regard image-based registration as a Maximum Likelihood Estimation [116, 154]. It allows to derive a number of popular similarity measures using certain assumptions about the image intensities. Therefore we will briefly introduce the statistical concepts behind the Maximum-Likelihood. We model the dependency of the image intensities I and J as follows:

$$I(\vec{x}) = f\left(J(\phi(\vec{x}))\right) + \epsilon \tag{2.7}$$

That means the image intensity of I can be described by a functional mapping f on the intensities of J at locations transformed by ϕ, plus an extra noise term ϵ, which represents the measurement error of the imaging system. The probability of an image pixel $I(\vec{x})$ is then described by

$$P(I|J,\phi,\epsilon,f) = P\left(I(\vec{x}) - f(J(\phi(\vec{x}))) = \epsilon\right) \tag{2.8}$$

Assuming that the image pixels of individual images are independent of their spatial location, the probability of the image I, given image J, the transformation ϕ, error term ϵ and intensity mapping f, then writes as

$$\begin{aligned} P(I|J,\phi,\epsilon,f) &= \prod_k P\left(I(\vec{x_k}) - f(J(\phi(\vec{x_k}))) = \epsilon\right) \\ &= \prod_k P\left(i_k - f(j_k) = \epsilon\right) \\ \text{with} \quad k &\in \{k|\vec{x_k} \in \Omega_I \wedge \phi(\vec{x_k}) \in \Omega_J\} \end{aligned} \tag{2.9}$$

The entities J, ϕ, ϵ and f constitute the model that I is predicted from. Here they act as parameters of a probability density function (PDF). We now want to maximize the likelihood of that PDF to the known images, which is the same equation 2.9 for given I and J. In particular we are interested in obtaining optimal parameters for ϕ. The logarithm of eq. 2.9 leads to the so-called log-likelihood, which is often preferred, as it transforms products to sums that are easier to deal with. Since the logarithm is a continuous strictly increasing function over the range of the likelihood, maximizing the likelihood is equivalent to maximizing its logarithm. We therefore seek a maximum of the following term:

$$\mathcal{L}(\phi,\epsilon,f) = \sum_k \log P\left(i_k - f(j_k) = \epsilon\right) \tag{2.10}$$

Now ϵ is assumed to be stationary zero-centered Gaussian noise with variance σ. The probability of a particular pixel intensity is then

$$P\left(i_k - f(j_k) = \epsilon\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(i_k - f(j_k))^2}{2\sigma^2}\right) \tag{2.11}$$

and the corresponding log-likelihood writes as

$$\mathcal{L}(\phi,\sigma,f) = -N \log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\sum_k \frac{(i_k - f(j_k))^2}{\sigma^2} \tag{2.12}$$

The similarity measures presented below each make a specific assumption about the intensity mapping f that supposedly underlies the two images I and J to be registered.

2.2.3. Commonly Used Measures

Sum of Squared Differences (SSD)

$$SSD = \frac{1}{N}\sum_{k=1}^{N}(i_k - j_k)^2 \tag{2.13}$$

In its original form stated here, SSD sums over the squared differences of the image intensities, and is therefore an error measure. To use it in a similarity maximization framework, its negated value $-SSD$ can be used instead. The error will be minimal if the image intensities at the correct alignment are identical. Therefore this measure is mostly useful for registering images of the same modality, e.g. CT or MRI acquisitions acquired at different times [52]. The local error contribution of every voxel pair makes it especially useful for deformable registration techniques, where local image dissimilarity or force fields based on intensity errors are often required. Besides, it can be optimized by least-squares optimization techniques (see section 2.5). The main disadvantage of SSD is the restrictive assumption that the image intensities are identical. Therefore non-matching structures entering one of the images easily cause larger increases of the error value than the shift to a correct alignment. In that sense, its robustness with respect to outliers and noise is generally bad. In fact, SSD is to be considered an optimal error measure, if the image intensities only differ by stationary Gaussian noise, as defined above. That can easily be shown by inserting $f(j) = j$ in equation 2.12

$$\begin{aligned}\mathcal{L}_{SSD}(\phi,\sigma) &= -N\log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\sum_{k}\frac{(i_k - j_k)^2}{\sigma^2} \\ &= -N\left(\log\left(\sqrt{2\pi}\sigma\right) + \frac{1}{2\sigma^2}SSD\right)\end{aligned} \tag{2.14}$$

which is increasing with $-SSD$. One common modification to reduce the sensitivity with respect to outliers is to use the Sum of Absolute Differences (SAD) instead:

$$SAD = \frac{1}{N}\sum_{k=1}^{N}|i_k - j_k| \tag{2.15}$$

Mathematically, this would be the correct measure for a Lagrangian noise model. Unfortunately, the property of the absolute value operator being non-differentiable at zero, makes it difficult to integrate it into analytical expressions. In general, an arbitrary hull function can be designed that is applied to the intensity differences, details will be explained in section 2.2.4.

Normalized Cross-Correlation (NCC)

$$NCC = \frac{1}{\sigma_i\sigma_i N}\sum_{k}\left(i_k - \bar{i}\right)\left(j_k - \bar{j}\right) \tag{2.16}$$

$$\text{with } \bar{i} = \frac{1}{N}\sum_{k} i_k; \quad \sigma_i = \sqrt{\mathrm{Var}(i)} = \sqrt{\frac{1}{N}\sum_{k}\left(i_k - \bar{i}\right)^2} \tag{2.17}$$

Normalized Cross-Correlation, often also denoted as Correlation Coefficient (CC) is a very common technique throughout statistics, signal processing and many other fields. It assesses

the amount of linear correlation of two signals by computing the average product of its demeaned values, divided by their standard deviation. The arithmetic mean $\bar{i}$ is used, equivalent to the expectation value $\mathbb{E}(i)$, if i is interpreted as random variable. The standard deviation σ_i, is the square root of the statistical variance. The mean and standard deviation of j is then defined equivalently. Note that there is a bit of confusion among mathematicians, if N or $N-1$ should be used in the denominator for computing the variance, see [115]. The bottom line is, that if that difference matters, the number of samples (i.e. image content) is by far not sufficient anyway. Putting it together, the equation for Normalized Cross Correlation writes

$$NCC = \frac{\sum_k (i_k - \bar{i})(j_k - \bar{j})}{\sqrt{\sum_k (i_k - \bar{i})^2 \sum_k (j_k - \bar{j})^2}} \tag{2.18}$$

$$= \frac{\sum_k i_k j_k - N\bar{i}\bar{j}}{\sqrt{\left(\sum_k i_k^2 - N\bar{i}^2\right)\left(\sum_k j_k^2 - N\bar{j}^2\right)}} \tag{2.19}$$

The mean intensities $\bar{i}$ and $\bar{j}$ have to be established before equation 2.18 can be computed. In contrast, equation 2.19 has the advantage that it can be computed in one traversal of the images, by basically accumulating five sums (the intensity product $i_k j_k$, the original intensities i_k, j_k, as well as their squares i_k^2, j_k^2, respectively). It is however less numerically stable (compare [115], section 14.1), since potentially large sums are substracted from small ones.

In the case of our two images, the assumed intensity mapping is a linear one, $f(j) = \alpha j + \beta$. Therefore the NCC measure is independent of both a change in the brightness (β) and contrast (α) between the two images. If the intensities i and j are completely independent, the value of NCC will be around zero. If the linear mapping is correct for all intensities, its value will be 1 or -1 (depending if the linear mapping is positive or negative). This is easily confirmed by inserting $f(j) = \alpha j + \beta$ in eq. 2.17 and then eq. 2.16:

$$\bar{i} = \alpha\bar{j} + \beta; \quad \sigma_i = \sqrt{\alpha^2 \frac{1}{N}\sum_k (j_k - \bar{j})^2} = |\alpha|\sigma_j \tag{2.20}$$

$$NCC = \frac{1}{\sigma_i \sigma_j N}\sum_k \alpha (j_k - \bar{j})(j_k - \bar{j}) = \frac{\alpha\sigma_j^2}{|\alpha|\sigma_j^2} = \mathrm{sgn}(\alpha) = \pm 1$$

Equation 2.20 can in a reverse manner be used to estimate the parameters α and β. Assuming a positive linear relationship, i.e. $\alpha > 0$, we obtain:

$$\alpha = \frac{\sigma_i}{\sigma_j}; \quad \beta = \bar{i} - \frac{\sigma_i}{\sigma_j}\bar{j} \tag{2.21}$$

Note that this is not necessarily the best method to estimate the parameters, since a linear relationship between i_k and j_k was implicitly assumed in deriving it. Therefore, some other technique like linear regression might produce better results if the correlation of the images is poor. We will do just that later in this thesis (section 4.3.2) to yield an improved similarity measure for CT-Ultrasound registration. By using eq. 2.21, we can show that Normalized Cross Correlation is consistent with the maximum log-likelihood. To help us in the derivation, we use the normalized image intensities

$$\widetilde{i_k} = \frac{i_k - \bar{i}}{\sigma_i}; \quad \widetilde{j_k} = \frac{j_k - \bar{j}}{\sigma_j} \tag{2.22}$$

that have some convenient properties:

$$\mathbb{E}\left(\widetilde{i_k}\right) = \mathbb{E}\left(\widetilde{j_k}\right) = 0; \quad \mathrm{Var}\left(\widetilde{i_k}\right) = \mathrm{Var}\left(\widetilde{j_k}\right) = 1$$
$$\mathbb{E}\left(\widetilde{i_k}^2\right) = \mathbb{E}\left(\widetilde{j_k}^2\right) = 1$$
$$NCC(I,J) = \tfrac{1}{N}\sum_k \widetilde{i_k}\widetilde{j_k} = \mathbb{E}\left(\widetilde{i_k}\widetilde{j_k}\right)$$

Expanding only the sum in equation 2.12 yields

$$\begin{aligned}
\sum_k (i_k - f(j_k))^2 &= \sum_k (i_k - \alpha j_k - \beta)^2 \\
&= \sum_k \left(i_k - \frac{\sigma_i}{\sigma_j} j_k - \bar{i} + \frac{\sigma_i}{\sigma_j}\bar{j} \right)^2 \\
&= \sigma_i^2 \sum_k \left(\frac{i_k - \bar{i}}{\sigma_i} - \frac{j_k - \bar{j}}{\sigma_j} \right)^2 \\
&= \sigma_i^2 \sum_k \left(\widetilde{i_k} - \widetilde{j_k} \right)^2 \\
&= \sigma_i^2 \sum_k \left(\widetilde{i_k}^2 - 2\widetilde{i_k}\widetilde{j_k} + \widetilde{j_k}^2 \right) \\
&= \sigma_i^2 N \left(\mathbb{E}\left(\widetilde{i_k}^2\right) - 2\mathbb{E}\left(\widetilde{i_k}\widetilde{j_k}\right) + \mathbb{E}\left(\widetilde{j_k}^2\right) \right) \\
&= 2\mathrm{Var}(I) N \left(1 - NCC(I,J)\right)
\end{aligned}$$

The log-likelihood for Normalized Cross Correlation is then

$$\mathcal{L}_{NCC}(\phi, \sigma) = -N \log\left(\sqrt{2\pi}\sigma\right) + \frac{\mathrm{Var}(I) N}{\sigma^2} \left(NCC(I,J) - 1\right) \tag{2.23}$$

Normalized Cross Correlation has been used extensively for mono-modal registration problems, in particular 2D-3D registration algorithms [105, 60, 129]. An original idea was to use NCC to non-iteratively register images in the spatial frequency domain , see [68] and [53], page 54-55. This approach however suffers from the common problems of applying fourier transformations to discrete images on a limited domain.

Correlation Ratio

$$\eta(I|J) = \frac{\mathrm{Var}(\mathbb{E}(I|J))}{\mathrm{Var}(I)} \tag{2.24}$$

Again, the images I and J are treated as random variables. Introduced by Roche et al. [117], Correlation Ratio then measures how well I is explained by J. Here's how it works intuitively: Assuming that I is totally independent of J, the expectation $\mathbb{E}(I|J) = \mathbb{E}(I)$ is constant and its variance is zero. On the other hand, full functional dependency implies that every value of I can be predicted from J, therefore $\mathbb{E}(I|J) = I$, resulting in a measure value of 1. Note that this similarity measure is not symetric. One therefore has to decide in advance which of the images can be used better to predict the other one (and therefore serves as a model). Without loss of generality, we assume for the derivations here that the moving image J is the model, however note that those roles can be interchanged.

Without assuming any parameterization of the functional relationship, a popular method to compute η uses the variances of the prediction for all discrete intensities in J. The required one- and two-dimensional probability density functions are commonly derived from the respective image histograms.

$$\begin{aligned} \eta(I|J) &= 1 - \frac{\mathbb{E}_j(\mathrm{Var}(X|J=j))}{\mathrm{Var}(I)} = 1 - \frac{1}{\mathrm{Var}(I)} \sum_j \sigma_j^2 P_J(j) \\ \sigma_j^2 &= \left(\frac{1}{P_J(j)} \sum_i i^2 P_{IJ}(i,j) \right) - \left(\frac{1}{P_J(j)} \sum_i i P_{IJ}(i,j) \right)^2 \\ \mathrm{Var}(I) &= \sigma_I^2 = \left(\sum_i i^2 P_I(i) \right) - \left(\sum_i i P_I(i) \right)^2 \end{aligned} \tag{2.25}$$

The actual prediction of the I intensities based on J, i.e. $f(j_k)$, is now a function that is piece-wise defined on the PDFs:

$$f_j = \frac{1}{P_J(j)} \sum_i i P_{IJ}(i,j) \tag{2.26}$$

Roche et al. had further generalized this measure [118] to include arbitrary functional mappings f (not necessarily estimated from the image PDFs):

$$\eta(I|J) = 1 - \frac{\sum_k (i_k - f(j_k))^2}{N \mathrm{Var}(I)} \tag{2.27}$$

That equation is now closely related to the formulation of the log-likelihood in equation 2.12. This actually causes a lot of fame for the authors, since almost everything can now be called Correlation Ratio. They probably deserve it though, since they nicely describe a consistent theory, which helps to deeply understand this class of similarity measures.

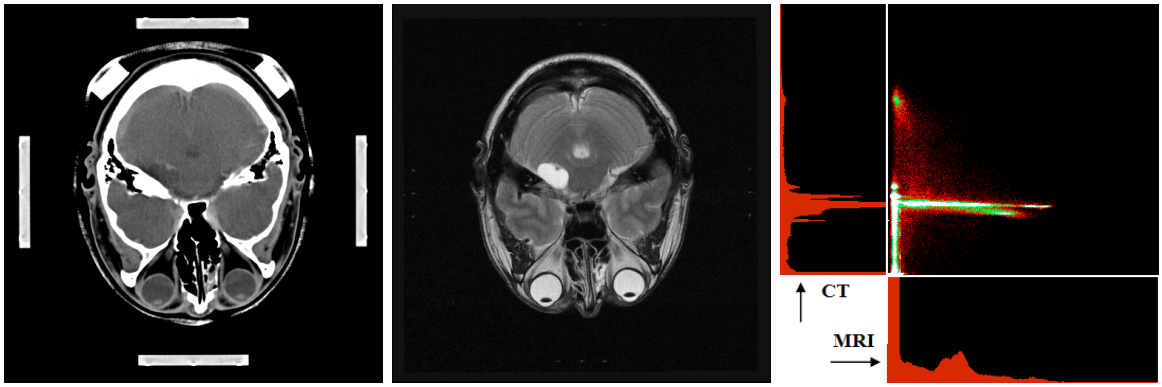
The original version of Correlation Ratio can be put in relation to the log-likelihood as well. We rewrite equation 2.12 to incorporate the piece-wise defined function mapping:

$$\mathcal{L}(\phi, \sigma, f) = -N \log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2} \sum_j \sum_{\vec{x_k} \in \Omega^j} \frac{(i_k - f_j)^2}{\sigma^2} \tag{2.28}$$

with $\Omega^j = \{\vec{x_k} \in \Omega | f(\vec{x_k}) = j\}$ being the iso-intensity set in the transformed image for every j. Both the intensity mapping f and σ should be estimated such that the likelihood is maximized, therefore we derive $\mathcal{L}$ and obtain

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial f_j} = \frac{1}{\sigma^2} \sum_{\Omega^j} (i_k - f_j) \quad &\Longrightarrow \quad \widehat{f_j} = \frac{1}{|\Omega^j|} \sum_{\Omega^j} i_k \\ \frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_j \sum_{\Omega^j} (i_k - f_j)^2 \quad &\Longrightarrow \quad \widehat{\sigma}^2 = \sum_j \frac{|\Omega^j|}{N} \widehat{\sigma_j}^2 \end{aligned}$$

where $\widehat{\sigma_j}^2 = \frac{1}{|\Omega^j|} \sum_{\Omega^j} (i_k - \widehat{f_j})^2$ are the variances for every iso-intensity set. Note that both $\widehat{f_j}$ and $\widehat{\sigma_j}^2$ correspond exactly to the histogram-based estimations in equations 2.26 and 2.25,



(a) CT slice (b) MRI slice (c) Joint Histogram

Figure 2.3.: Registered head CT and MRI slices and their joint and individual histograms.

respectively. Using equation 2.25 again, the final maximum likelihood relates to Correlation Ratio:

$$\mathcal{L}_\eta = N \log \left[\sqrt{2\pi} \mathrm{Var}(I)(1 - \eta(I|J)) \right]$$

Figure 2.3 illustrates the intensity relationship for a multi-modal CT-MRI registration problem. Note that the CT slice (a) is displayed within a very narrow intensity window, due to the poor soft tissue contrast. Looking at the joint histogram in (c), it becomes obvious that MRI should be used as the model image (in our notation J), since it distinguishes the soft tissue with more detail. On the other hand, high intensities in CT can not be predicted, as MRI does not image bony structures. Thus, while Correlation Ratio might still work reasonably well in this scenario, one can see that the assumption of an intensity mapping f is violated in several parts of the histogram.

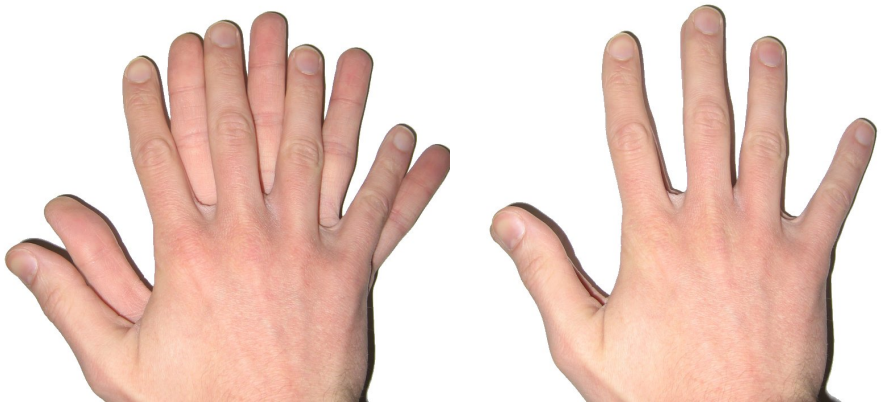
Mutual Information

$$MI = \sum_i \sum_j P_{IJ}(i,j) \log \frac{P_{IJ}(i,j)}{P_I(i) P_J(j)} \tag{2.29}$$

The idea behind Mutual Information is that a correct alignment of two images minimizes the amount of information in a shared representation of them. Put your hands together while spreading your fingers, and look at them from the front (figure 2.4). If they're misaligned, you can distinguish all ten fingers. If they're fully aligned, you only see five fingers, which can be considered the least information possible in the shared representation (the projection of your hands that you're looking at).
Transferring this idea to images, we again make use of their incarnation as random variables. A measure of how much information is present in an image is needed. The most common measure of information is the Shannon entropy:

$$H(I) = -\sum_i P_I(i) \log P_I(i) \tag{2.30}$$

One can easily see that the value of $H(I)$ will be zero if there is just one constant intensity in the images, since $p \log p$ equals 0 for both $p = 0$ and $p = 1$. On the other hand, the maximum



(a) misaligned (b) aligned

Figure 2.4.: Example illustrating the information content in a shared representation of two images.

entropy is obtained if every intensity is equally likely to occur.
As a joint representation (the entropy of which expresses the misalignment) we use the intensity tuples $\{i_k, j_k\}$ and their joint probability. The joint entropy of I and J is then

$$H(I,J) = -\sum_i \sum_j P_{IJ}(i,j) \log P_{IJ}(i,j) \tag{2.31}$$

In the majority of applications, the individual and joint probability distributions are estimated using histogramming, as defined in section 2.2.1. Mutual Information now simple writes as

$$MI = H(I) + H(J) - H(I,J) \tag{2.32}$$

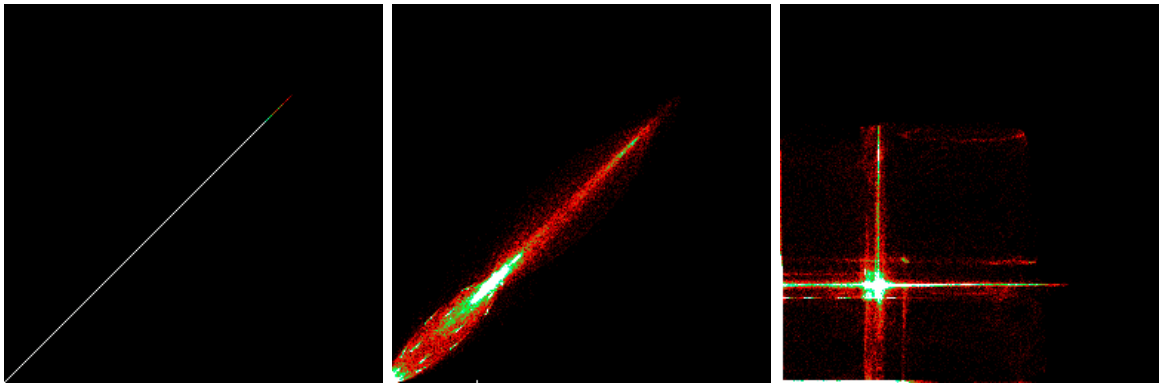
which corresponds to equation 2.29. Historically, the minimization of the Joint Entropy $H(I,J)$ alone was used for registration too [59], just before Mutual Information was developed. The value of Mutual Information is in the range $[0 \dots H_{max}]$, with H_{max} being the maximum entropy of either of the images. Normalizing it to $[0 \dots 1]$ is easily possible by [53]:

$$MI' = \frac{2MI}{H(I) + H(J)} = 2 - \frac{2H(I,J)}{H(I) + H(J)} \tag{2.33}$$

An alternative normalization scheme is to use the ratio between the individual entropies and the joint entropy. This has been shown to be more invariant to changes in the overlapping area of the images [147], but nevertheless is only a rewriting of the first normalized equation:

$$MI'' = \frac{H(I) + H(J)}{H(I,J)} = \frac{1}{MI' - 2} \tag{2.34}$$

Equation 2.33 becomes zero if one of the images is constant. It assumes a value of one if the images are identical; it remains one even if any perturbation of the histogram intensity bins is applied to I and/or J. The number of histogram bins has a significant effect on MI cost function. The more bins are used, the sharper the peak will be at the best alignment. However, using less bins might ensure a more smooth monotonic increase towards the optimum.



(a) identical (b) 0.3mm displaced (c) 2cm displaced

Figure 2.5.: Joint histograms of a CT slice with itself, for various displacements.

Figure 2.5 shows a joint histogram with 256 bins for various displacements of an identical CT slice. For a misalignment as small as $0.3mm$ (b), the intensity values are spread away from the identity axis significantly.

Mutual Information is perfectly suited for multi-modal registration. In particular for CT-MRI registration (as shown in figure 2.3), it is well established and used in commercial products. Many other registration problems are addressed with Mutual Information, either used exclusively or in combination with application-specific additions. See Pluim et al. [109] for an overview on Mutual Information based registration.

Last but not least, following is the maximum-likelihood derivation, similar to [116]. The assumption of a functional mapping with a gaussian noise channel is not used anymore, we rather only require that the conditional densities $P(i_k|j_k)$ are stationary, i.e. independent of $\vec{x_k}$. Therefore the following likelihood term is to be maximized:

$$\mathcal{L}(\phi, F) = \log P(I|J, \phi, F) = \sum_{\vec{x_k} \in \Omega} \log f(i_k|j_k) \tag{2.35}$$

with parameter vector F now representing the unknown conditional densities

$$F = (f(0|0), f(1|0), \ldots, f(1|1), \ldots, f(2|0), \ldots)$$

with the contraints

$$\forall j: \quad C_j = \sum_i f(i|j) - 1 = 0$$

The groups of intensity pairs are written as

$$\Omega_{i,j} = \{\vec{x_k} \in \Omega | I(\vec{x_k}) = i \wedge J(\phi(\vec{x_k})) = j\}; \quad N_{i,j} = |\Omega_{i,j}|$$

The log-likelihood becomes:

$$\mathcal{L}(\phi, F) = \sum_{i,j} N_{i,j} \log f(i|j)$$

In order to maximize it, we force its partial derivatives with respect to all $f(i|j)$ to be zero. To incorporate the constraints C_j, additional Lagrange multipliers λ_j are used:

$$\frac{\partial \mathcal{L}}{\partial f(i|j)} - \sum_{j'} \lambda_{j'} \frac{\partial C_{j'}}{\partial f(i|j)} = \frac{N_{i,j}}{f(i|j)} - \lambda_j = 0$$

Subject to the constraints $\sum_i f(i|j) = 1$, the optimal parameters are:

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

The joint probabilities $P_IJ(i,j)$ and the marginal probabilities $P_J(j)$ are again computed from the joint and individual histograms, respectively. The final log-likelihood writes as

$$\mathcal{L}_{MI}(\phi) = N \sum_{i,j} P_{IJ}(i,j) \log \frac{P_{IJ}(i,j)}{P_J(j)} = N\left(H(I) - MI(I, \phi(J))\right)$$

Under the simplification of a constant marginal entropy $H(I)$ of the fixed image, it is directly proportional to Mutual Information.

2.2.4. Robust Functions

All similarity measures except Mutual Information can be written such that they incorporate a sum of a function ρ of intensity differences:

$$\sum_k \rho(i_k - f(j_k))$$

usually with $\rho(d) = d^2$ (for Sum of Absolute Differences $\rho(d) = |d|$). This formulation is not robust with respect to outliers, therefore large intensity differences can affect this error term significantly. Since the similarity measures have been defined under somewhat idealized assumptions (namely that the image intensities correspond up to a stationary gaussian noise term), such outliers will occur in most practical situations. It is therefore wise to restrict the influence of large intensity differences with a *Robust Function* ρ, a popular choice is

$$\rho(d) = \frac{d^2}{\sigma^2 + d^2}; \quad \rho(0) = 0; \quad \lim_{d\to\infty}\rho(d) = 1$$

The sensitivity of the similarity measure on the differences d can be adjusted by σ. $\rho'(d)$ is maximal for

$$\rho''(d) = \frac{-2\sigma^2(3d^2 - \sigma^2)}{(d^2 + \sigma^2)^3} = 0 \quad \Longrightarrow \quad d = \pm\frac{\sigma}{\sqrt{3}}$$

Compare [163], page 14. The same function is used with a slightly different notation in [118] for MR-Ultrasound registration. General information on robust statistics can be found e.g. in [62].

2.2.5. Remarks about Normalization

All similarity measures use some sort of normalization, to make the resulting value independent of the number of voxels in the overlapping region, some of them also normalize basic image properties like the intensity variance or entropy. It is not a trivial step though, one has to be aware about the problems that can arise due to normalization. Choosing the right equation is always a trade-off favoring one or another scenario when the image similarity will become maximized, therefore the right choice depends highly on the application. Here are some examples to illustrate problems of the introduced similarity measures with normalization:

- The Sum of Squared Differences (SSD) contains a division by the number of voxels in the overlapping region. If the SSD value does only have a small minimum at the correct alignment (which can happen if there is significant amount of noise in the images, or a lot of modality-specific differences), a global minimum might be reached just by fully overlapping the images, since the denominator (the number of overlapping voxels) reaches a maximum there. On the contrary, if that division is omitted, a minimum error can arise, if the transformation just overlaps a couple of similar voxels (in the extreme case one voxel with identical intensity, the error becomes zero then).
- In addition to the number of voxels, Normalized Cross-Correlation (NCC) divides by the standard deviations of both images. Therefore, the measure can reach a maximum if either of those is small, which usually corresponds with little structure in the images. If that division is omitted for one or both of the standard deviations, there is danger of reaching a maximum when some structures with a lot of contrast are in the overlapping region (but not necessarily well aligned).
- Correlation Ratio divides by the variance of image I, the potential normalization problems are similar to NCC.
- Mutual Information somehow has to weight between the individual image entropies and the joint entropy. It is obvious that this is a difficult issue just by the fact that at least four different equations have been commonly used in the literature - the joint entropy alone, as well as equations 2.32, 2.33 and 2.34.

2.2.6. Incorporating Spatial Information

All measures presented so far were just functions of image intensity tuples, not including the location within the images. Likewise, in the maximum likelihood formulation, the image intensities were defined to be independent. In reality that is not true though, and therefore we may actually be discarding a lot of valuable information. In the following we would like to give an overview about ways of treating images within their spatial context.

Neighborhood Operations

Any of the presented basic similarity measures can be applied on smoothed versions of the original images. This includes e.g. smoothing with a Gaussian kernel filter, as well as downsampling the images to a different resolution. This may have the advantage that the similarity criterion converges more monotonically towards its optimum, since the image structures gradually line up. Imagine a thin line structure in 2D images to be registered. Only for very close

alignment, the overlap of the lines contributes to the image similarity, while larger displacements will produce the same value, since the lines do not touch. For strongly smoothed images, the similarity measure will increase when the lines are moved towards each other, since their blurred seams start to overlap long before the actual lines.
Working on smaller copies of the original images by down-sampling can also drastically improve the computation speed. It is in fact a popular technique to run a registration algorithm subsequently on resampled images with increasing resolution. It is commonly referred to as *Image Resolution Pyramid*. In computer vision, local image operators and feature detectors work in *Scale Space*.
Similarity Measures can be computed on gradient images, computed from the original images e.g. using a Sobel filter kernel. This has been used especially for 2D-3D registration [163]. *Gradient Correlation* computes the average of the NCC values of both horizontal and vertical gradient images. Similarly, *Gradient Difference* applies a robust function on the differences of gradient images. *Pattern Intensity* assesses the amount of structure in a substraction of the original image intensities, by computing local differences within that difference image. Those three measures are described and compared in [104].

Local Averaging

Another option is to compute a Similarity Measure repeatedly on small patches of the images and average the result. A prominent example is Local Normalized Cross Correlation (LNCC) as described in [78]. Since the mean and standard deviation values are now computed locally, the new measure is robust against local brightness and contrast changes in the images. Important parameters that affect the performance are the size of the patches, as well as their overlap. An extension of LNCC is to weight each of the local values with the variance of one of the images [78]. It makes the measure more immune regarding newly introduced structures present only in the other image. Local Normalized Cross-Correlation has been extensively used for 2D-3D registration [163], but has also applications for monomodal 3D-3D registration (e.g. for registering different contrast phases of a CT study). In general, similarity measures that can be defined locally are suited for deformable registration methods, see section 2.3.4.

Using Image Location

Similarity Measures can be extended by incorporating the location vectors, then the underlying information is $\{i_k, j_k, \vec{x_k}\}$. First of all, the location $\vec{x_k}$ can be used to derive a weighting for every intensity pair. The weight might be defined as proximity to the position of a clinical target, where the maximum accuracy is desired. It could also make use of some knowledge of the image acquisition process, and therefore define higher weights at locations where the intensity measurement is supposedly more accurate (we will apply such a technique later).
As a continuation of that idea, the image location can be directly integrated into Information Theoretic Measures. A significant amount of research has been done regarding *Higher-Dimensional Mutual Information*. It allows one to treat not only a scalar image intensity i_k, but a whole vector, e.g. $\{i_k, \vec{x_k}\}$ as a single random variable. It helps to overcome one of the drawbacks of Mutual Information: the fact that it considers no proximity in both image and intensity space. It adds a lot of additional problems, though. Entropy and joint entropy have to be estimated in high dimensions. A histogram-based technique is usually

not feasible then, since every occurence of a multi-component vector is likely to have a very sparse distribution. Therefore, different techniques have been developed to directly estimate entropy (for an older overview see [13]), without actually computing the underlying PDFs. One solution is to use the distances of the vectors representing the random variables, and compute their minimum spanning tree (MST) for entropy estimation [57]. High-dimensional entropy estimation based on Voronoi & Delaunay regions is introduced in [92]. If a stable entropy estimation is in place, any kind of extra information can be used, besides intensity and location. In [75], higher-dimensional Mutual Information using both 3-dimensional color image feature vectors and 25-dimensional neighborhood feature vectors is investigated. The Maximum Distance-Gradient-Magnitude (MDGM) is developed in [46] as a feature with more global properties then local image gradients, and combined with the image intensity as 2D feature vectors. Mutual Information is estimated based on histogramming (resulting in a four-dimensional joint PDF), and applied to register CT and MRI volumes.

2.2.7. Incorporating Prior Information

Especially in multi-modal registration scenarios, where the two images do not have much in common, feeding in a little more a priori knowledge can boost the robustness, or allow a working algorithm in the first place. Often knowledge about the joint probability distribution function of registered images is used as an intensity prior. The Kullback-Leibler divergence (KLD) expresses the error between the observed joint PDF P_{IJ} and a reference PDF $\widehat{P}$:

$$D(P_{IJ}|\widehat{P}) = \sum_{i,j} P_{IJ}(i,j) \log \frac{P_{IJ}(i,j)}{\widehat{P}(i,j)}$$

In the registration community it is often referred to as Kullback-Leibler Distance as well, strictly speaking it is not a distance though, since it is not symetric (symetric versions have been proposed as well). This error term can be used to exclusively drive the registration, as done in [29] for rigid 2D-3D registration of DSA to MRA images, as well as 3D-3D registration of T1 and T2 phase MRI. It can also be used as a weighted component that also includes a regular similarity metric, see [51] for deformable registration of functional PET and SPECT imaging with CT. For a stable registration based on such intensity priors, it is crucial that the intensity distributions are normalized (or equal in the first place) with respect to contrast and brightness, which is often not trivial to achieve. If the intensity prior information is obtained by computing the distributions based on a number of training data sets, the algorithm falls within the category of *Learning-based methods*.
There are additional means to include prior, learned information. An image pre-processing operator can be trained to extract significant features for registration, as done in [102] for MRI-Ultrasound registration. Boosting of a similarity measure can be achieved by learning which image features contain information about the alignment, such as in [181] for motion estimation on cardiac echography.
Learning-based methods can help to solve tough image registration problems, however they often require large additional effort to establish the training database.

2.2.8. Multi-signal Similarity

Similarity Measures can be further generalized beyond using intensity information and the spatial context. The original formulation in equation 2.2 was $S(I, \phi(J))$. We extend it to

$$S\left(\Theta_1(I), \Theta_2(\phi(J))\right) \tag{2.36}$$

The Θ_i operators perform some processing on the whole image and return a vector-valued image with the same input dimensions and an arbitrary number of output dimensions. In the simplest case, both $\Theta_i = \mathrm{Id}$ are the identity operator, therefore the similarity measure is computed directly on the input images. The point is, that an arbitrary number of signals can be extracted from each image. This is in particular important in multi-modal registration, where the two images to be compared have very different properties. An example from the literature is the MR-Ultrasound registration from Roche et al. [118]. Here the MR volume I is pre-processed such that the intensity and gradient information are compared to the ultrasound image J using a Correlation Ratio framework. Using the notation just introduced, we have the operators $\Theta_1(I) = (I, \nabla I)$ and $\Theta_2(J) = J$. Another example is the MRI-Ultrasound registration method presented by Penney et al. [102]. There, the operators convert both modalities to vessel probability maps, learned from a number of training data sets, which are then registered using a simple cross-correlation similarity measure. In a complex scenario, each Θ_i can create a whole structural tensor from its input image. In fact, those operators could even describe a full simulation of one modality from the other. We make use of this idea when developing our new CT-ultrasound registration algorithms in chapter 4.

2.2.9. Multi-variate Similarity

Scenarios exist where it is appropriate to register more than two images at once. It basically requires that the pair-wise similarity measure $S(I, \phi(J))$, as defined in equation 2.2, be extended to an arbitrary number of n images $S(\phi_1(I_1), \phi_2(I_2), \ldots, \phi_n(I_n))$. In the limited body of literature on this, mainly Mututal Information has been extended to register multiple MRI phases. In [20], MI is extended with three-dimensional histogramming, [84] use just the sum of all combinations of regular MI on image pairs, and a new entropy estimation on sparse high-dimensional histograms combined with a multi-variate similarity measure related to MI is used in [178].
In [156] we systematically extended all similarity measures from section 2.2.3, with the goal of "mosaicking" multiple 3D-ultrasound volumes. More details on multi-variate registration, including the discussion on normalizing multiple pair-wise registration results versus full-simultaneous optimization of all images, can be found in [155].

2.3. Spatial Transformations

Now we will introduce different transformation models that can be used to transform the moving image J to eventually achieve a correct alignment to I. Figure 2.6 depicts an overview of the transformation models we will cover in this section. Not only the type of transformation is important, but also its parameterization, i.e. the way a transformation is described as parameter vector ϕ for optimization. In the following, transformations are only described for three dimensions, since most registration problems are three-dimensional.
We start using homogenous coordinates at this point, and expect the reader is familiar with

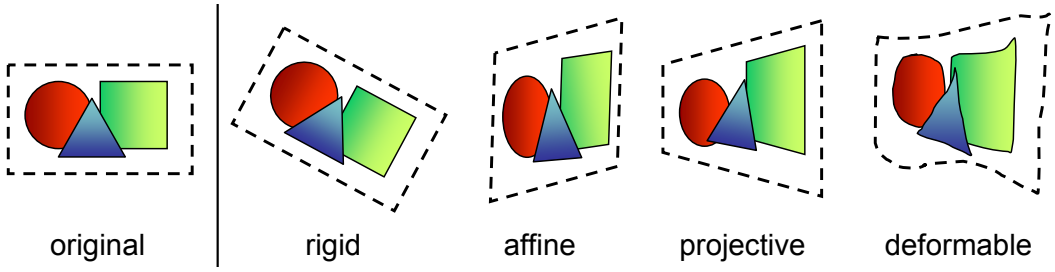


Figure 2.6.: Illustration of different transformation models applied to a simple 2D scene.

this notation. Basically, a 3D-vector is extended with a fourth coordinate that is initially set to 1. In particular, it allows for the compact representation of all global 3D transformations as a $4x4$ matrix. Its full strength is deployed when it comes to projective transformations, see [38] or [54] for details. Homogenous coordinates are also the standard notation in the area of computer graphics [41]. Nowadays, all 3D graphics hardware natively supports computations on 4-vectors and $4x4$ matrices.

2.3.1. Rigid

A rigid transformation applies a translation and rotation to every location in J. In fact, it only changes its coordinate system, all geometric properties (length, volume, parallelity etc.) are preserved. It has six degrees of freedom (DOF), three for translation and three for rotation.

$$\phi(\vec{x}) = R\vec{x} + \vec{t} \tag{2.37}$$

where $\vec{t} = (t_x, t_y, t_z)^T$ is a 3-vector describing the translation, and R is a special orthogonal $3x3$-matrix. In homogenous coordinates, a $4x4$ rigid transformation matrix is written as

$$T_{rigid} = \begin{pmatrix} R & \vec{t} \\ \vec{0} & 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{11} & r_{31} & t_x \\ r_{12} & r_{22} & r_{32} & t_y \\ r_{13} & r_{23} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

As R is a special orthogonal matrix, its determinant is one, its transposed equals its inverse $R^T = R^{-1}$, and its three column vectors are the basis vectors of the rotated coordinate systems, i.e. they have unit length and are all perpendicular. Because of those constraints, the values of R can not be directly modified, but an appropriate parameterization is needed.

Euler Angles

This is a very common parameterization of 3D-rotations, it describes R as consecutive rotations around the three coordinate axes with the angles θ_x, θ_y and θ_z:

$$\begin{aligned} R &= R_z(\theta_z) \cdot R_y(\theta_y) \cdot R_x(\theta_x) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_z & -s_z & 0 \\ s_z & c_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_y & 0 & s_y \\ 0 & 1 & 0 \\ -s_y & 0 & c_y \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_x & -s_x \\ 0 & s_x & c_x \end{pmatrix} \\ &= \begin{pmatrix} c_y c_z & (s_x s_y c_z - c_x s_z) & (c_x s_y c_z + s_x s_z) \\ c_y s_z & (s_x s_y s_z + c_x c_z) & (c_x s_y s_z - s_x c_z) \\ -s_y & s_x c_y & c_x c_y \end{pmatrix} \end{aligned} \tag{2.38}$$

$$s_x = \sin(\theta_x), c_x = \cos(\theta_x), s_y = \sin(\theta_y), c_y = \cos(\theta_y), s_z = \sin(\theta_z), c_z = \cos(\theta_z)$$

Note that this is only one of many ways of composing the rotations. Any permutation of both the rotation direction (clockwise vs. counter-clockwise with respect to the positive axis vectors) and the order might be used.
Vice versa, the angles can be computed from the matrix R in equation 2.38:

$$\begin{aligned} \theta_x &= \tan^{-1}\left(\frac{R_{23}}{R_{33}}\right) \\ \theta_y &= \sin^{-1}\left(R_{31}\right) \\ \theta_z &= \tan^{-1}\left(\frac{R_{12}}{R_{11}}\right) \end{aligned}$$

where R_{ij} denots the element in row i and column j from R. An alternative which we suspect to be more stable, computes θ_z as above, and the other two angles based on it:

$$\begin{aligned} \theta_z &= \tan^{-1}\left(\frac{R_{12}}{R_{11}}\right) \\ \theta_x &= \tan^{-1}\left(\frac{R_{31}\sin(\theta_z) - R_{32}\cos(\theta_z)}{R_{22}\cos(\theta_z) - R_{21}\sin(\theta_z)}\right) \\ \theta_y &= \tan^{-1}\left(\frac{-R_{13}}{R_{11}\cos(\theta_z) + R_{12}\sin(\theta_z)}\right) \end{aligned}$$

This exploits more of the information inherent in R, and uses the arcus tangens $\tan^{-1}$ as only inverse trigonometric function. For the actual computation, one should always use the function atan2(a, b), available in virtually all programming environments, instead of $\tan^{-1}(a/b)$, since it distincts between the signs of both nominator and denominator and therefore properly computes angles in all quadrants.
The Euler Angles parameterization is minimal, since it only comprises three values. Its biggest disadvantage is that it shows singularities at some specific angles. This can be shown

by calculating the derivatives of a rotated point $R\vec{x}$ with respect to the rotation angles:

$$\begin{aligned}
\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= R \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
\frac{\partial}{\partial\theta_x} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= \begin{pmatrix} (c_x s_y c_z + s_x s_z)y + (-s_x s_y c_z + c_x s_z)z \\ (c_x s_y s_z - s_x c_z)y + (-s_x s_y s_z - c_x c_z)z \\ c_x c_y y - s_x c_y z \end{pmatrix} \\
\frac{\partial}{\partial\theta_y} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= \begin{pmatrix} -s_y c_z x + s_x c_y c_z y + c_x c_y c_z z \\ -s_y s_z x + s_x c_y s_z y + c_x c_y s_z z \\ -c_y x - s_x s_y y - c_x s_y z \end{pmatrix} \\
\frac{\partial}{\partial\theta_z} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= \begin{pmatrix} -c_y s_z x - (s_x s_y s_z + c_x c_z)y - (c_x s_y s_z - s_x c_z)z \\ c_y c_z x + (s_x s_y c_z - c_x s_z)y + (c_x s_y c_z + s_x s_z)z \\ 0 \end{pmatrix}
\end{aligned}$$

If we now specify the angles θ_x and θ_y accordingly, some of the derivatives become identical:

$$\begin{gathered}
\theta_x = \tfrac{\pi}{2}, \theta_y = -\tfrac{\pi}{2} \Rightarrow \\
\tfrac{\partial}{\partial\theta_x} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \tfrac{\partial}{\partial\theta_z} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} s_z y + c_z z \\ -c_z y + s_z z \\ 0 \end{pmatrix}
\end{gathered}$$

Thus the three rotations are not independent of each other anymore - the angles θ_x and θ_z rotate around the same axis. This problem is also referred to as *Gimbal Lock*. Nevertheless, Euler Angles can well be used for registration if only relative rotations are applied such that configurations with large angles never occur.
Last but not least, another slight disadvantage is that Euler Angles can be ambiguous, mostly at large angles as well. That means that two different sets of angles can describe the same rotation in 3D. However, as a particular Euler Angles configuration describes just one rotation R, it can be used without restriction to describe rotations, as long as the exact rotation order and direction is documented (compare eq. 2.38). One should only exercise caution when comparing rotations solely based on their Euler Angles.

Unit Quaternion

Representing 3D transformations with quaternions is very common in computer graphics and robotics. A rigid transformation can be described as a vector $[t_x, t_y, t_z, q_x, q_y, q_z, q_w]^T$, which combines three translational parameters and the four elements of a quaternion, denoting a rotation in space.
Quaternions are defined in the form $Q = iq_x + jq_y + kq_z + q_w = [(q_x, q_y, q_z), q_w]$, where i, j, and k have the following properties:

$$\begin{gathered}
i^2 = -1, j^2 = -1, k^2 = -1, \\
ij = k, ji = -k, jk = i, kj = -i, ki = j, ik = -j
\end{gathered} \tag{2.39}$$

All basic arithmetic operations can be derived from those properties, the behavior is consistent with the one of complex numbers, as quaternions can be seen as an extension of them. Refer to

[6] for a more detailed introduction to quaternions and their use in 3D computer applications. Here we will just present the very convenient way of describing a rotation in 3D with them. For this to work, the quaternion has to be normalized, i.e. $q_x'^2 + q_y'^2 + q_z'^2 + q_w'^2 = 1$. This can be achieved easily by dividing by the size of the original quaternion

$$q_x' = \frac{q_x}{|Q|}, q_y' = \frac{q_y}{|Q|}, q_z' = \frac{q_z}{|Q|}, q_w' = \frac{q_w}{|Q|} \qquad (2.40)$$
$$|Q| = \sqrt{q_x^2 + q_y^2 + q_z^2 + q_w^2}$$

If this is valid, the vector (q_x', q_y', q_z') represents the axis of rotation, and the angle θ is denoted indirectly by $\theta = 2\cos^{-1}(q_w')$. The respective rotation matrix is

$$R(q_x, q_y, q_z, q_w) = \begin{pmatrix} 1 - 2q_y'^2 - 2q_z'^2 & 2q_x'q_y' - 2q_w'q_z' & 2q_x'q_z' + 2q_w'q_y' & 0 \\ 2q_x'q_y' + 2q_w'q_z' & 1 - 2q_x'^2 - 2q_z'^2 & 2q_y'q_z' - 2q_w'q_x' & 0 \\ 2q_x'q_z' - 2q_w'q_y' & 2q_y'q_z' + 2q_w'q_x' & 1 - 2q_x'^2 - 2q_y'^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad (2.41)$$

As this only valid for unit quaternions, the normalization in equation 2.40 should be carried out for every set of parameters implicitly. We can therefore use the original scaled quaternion $Q = [(q_x, q_y, q_z), q_w]$ as rotational parameters. However, this results in having four parameters for the rotation - one more as degrees of freedom.

2.3.2. Affine

An affine transformation includes a rigid one, and provides in addition shearing and non-uniform scaling. It might be well suited for registration scenarios where geometric distortion of the imaging system has to be corrected. Besides, large-scale anatomic deformations caused e.g. by respiratory motion can be approximately recovered with this transformation model (how well depends on the exact anatomical site, region/volume of interest etc.).

$$\phi(\vec{x}) = HSR\vec{x} + \vec{t}$$

with

$$H = \begin{pmatrix} 1 & h_{xy} & h_{xz} \\ 0 & 1 & h_{yz} \\ 0 & 0 & 1 \end{pmatrix}; \quad S = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{pmatrix}$$

S is a scaling matrix, the components s_x, s_y, s_z define a non-uniform scaling in the three axis directions. H is a shearing matrix, the components h_{xy}, h_{yz} and h_{xz} define a shearing in the x-y, x-z and y-z-planes, respectively. In fact, every plane can be sheared in each of their spanning directions, therefore additional components h_{yx}, h_{zy} and h_{zx} would be needed, arranged in a matrix like H^T. It can be shown however, that those additional shearing parameters can be expressed with the first three shearing parameters and the rotation and scaling values [133]. Therefore a total of 12 parameters (each three for translation, rotation, scaling and shearing) define an affine transformation in 3D. This confirms with the homogenous representation

$$A = HSR; \quad T_{affine} = \begin{pmatrix} A & \vec{t} \\ \vec{0} & 1 \end{pmatrix}$$

where 3 degrees of freedom are in the translation and the remaining 9 in the now unconstrained matrix A. Usually it is not a good idea to directly optimize the entries of A though, since the effects on the image transformation vary greatly depending on the particular element of A and the actual value.

2.3.3. Projective

The most prominent use of a projective transformation is to describe the geometric mapping from locations in 3D onto the 2D view of e.g. the human eye or a camera. In terms of a homogenous $4x4$-matrix, all entries are unconstrained for this type of transformation, resulting in 15 degrees of freedom, as the matrix is defined up to scale. The vast majority of applications are in Computer Vision [54]. In the area of medical image registration, projective transformations arise mostly for 2D-3D registration e.g. of X-Ray projections to CT volumes. However, typically not the full projective transformation is sought for registration, since most of its parameters would be included in the geometric calibration of the X-Ray unit.

2.3.4. Deformable

If a global rigid, affine, or projective transformation is not appropriate, but local tissue motion is to be compensated as well, *deformable* registration is required. Also termed *Non-linear registration*, it has become a large field of research on its own [93].
The methodology introduced in this chapter can be extended in a straightforward manner for deformable registration, if a control point scheme is used. The sought transformation parameters are defined as the displacements of a set of control points, arranged in a uniform grid or adaptively selected. The deformable mapping of the moving onto the fixed image is interpolated using certain basis functions (popular ones being Thin-Plate Splines, Radial Basis Functions, B-Splines). An optimization algorithm still tries to maximize a global similarity measure, however with respect to a much larger number of parameters (up to many thousand). Such an approach is then only feasible if a gradient estimation of the similarity measure is available.

Dense-field deformable registration is a very different approach. Here a deformation field $\vec{u}$ is to be established that denotes the displacement of every voxel in the moving image:

$$u : \Omega \rightarrow \Omega$$
$$\phi(J(\vec{x}) = J(\vec{x} + \vec{u}(\vec{x}))$$

It is integrated in a global energy minimization formulation:

$$S(I, \phi(J)) + \alpha R(\vec{u}) = \min \tag{2.42}$$

Here S represents a dissimilarity measure (e.g. SSD for mono-modal registration), R is a regularization term that penalizes physically unlikely deformations $\vec{u}$, and $\alpha > 0$ a weighting factor. Usually, methods of variational calculus are used to find the solution for $\vec{u}$. Equation 2.42 is derived with respect to $\vec{u}$, which yields Euler-Lagrange equations for both S and R. A large system of non-linear Partial Differential Equations (PDE) is obtained, which can be solved by iterative methods, see [182] and [93] for details.
Deriving the (dis-)similarity measure S with respect to $\vec{u}$ implies that it is used as a force term, that "drags" the image at every voxel location toward the correct alignment. As a consequence, only similarity measures are eligible which can be locally defined, or at least feature meaningful local derivatives.

2.3.5. Parameter Scaling and Centering

For any optimizer to work best, changing each of the parameters in ϕ should have a comparable affect on the cost function value. This implies that altering any parameter value should change the similarity of the two images in the same order of magnitude.

Rigid Translation versus Rotation

In the case of a rigid transformation, we need to introduce an appropriate scaling between the translational and rotational parameters. To derive it, we use the corners of the bounding box of a volume of interest used in the registration. Translating the box a specific amount should then displace its corners equally than if rotating it around the center about the same amount. Using a cube with edge width d around the origin and considering one corner point p, the scaling can be derived as follows:

$$\begin{aligned} t_x &= \Delta p \\ \theta_x s &= \Delta p = \theta_x d \frac{\sqrt{3}}{2} \\ s &= d \frac{\sqrt{3}}{2} \end{aligned}$$

t_x is a translation in millimeters, θ_x a rotation in radiant, Δp is the displacement of the point p and s is the sought scaling between translational and rotational values. It is convenient to use translation values in millimeters and rotation in degrees, corresponding to a scaling $s = \frac{180}{\pi}$. Then the width of the respective bounding box satisfying the former constraints is $d = \frac{360}{\pi\sqrt{3}} \approx 66mm$. If the size of the images involved in the registration problem is roughly in that order of magnitude, one might stick with it, otherwise a different scaling has to be applied.

Affine Scaling

Concerning the additional parameters for affine registration, the non-uniform scaling parameters s_x, s_y and s_z deserve a little attention. They should not be directly used in an optimization, since $\lim_{s_x \to 0}$ makes an image dimension infinitely small, while for $s \gg 1$ the size barely changes. This is a consequence of the fact that s_x expresses the ratio between the size of images I and J. Since ratios are well expressed using logarithmic units (e.g. decibel), we should do likewise and therefore compute the image scaling as

$$\phi_{s_x} = \log\left(\frac{s_x}{1}\right) \Rightarrow s_x = \exp\left(\phi_{s_x}\right)$$

Centering

For rigid and affine transformations, the choice of the used coordinate system can significantly affect the registration performance. The origin of the physical coordinate system used to specify image locations, will be the effective center of rotation. Defining it at the corner of the images seems straightforward, but is a rather bad choice, since it causes a strong coupling of translation and rotation parameters. Depending on the used optimizer, such a dependency might be difficult to handle, and result in non-convergence even for relatively easy registration

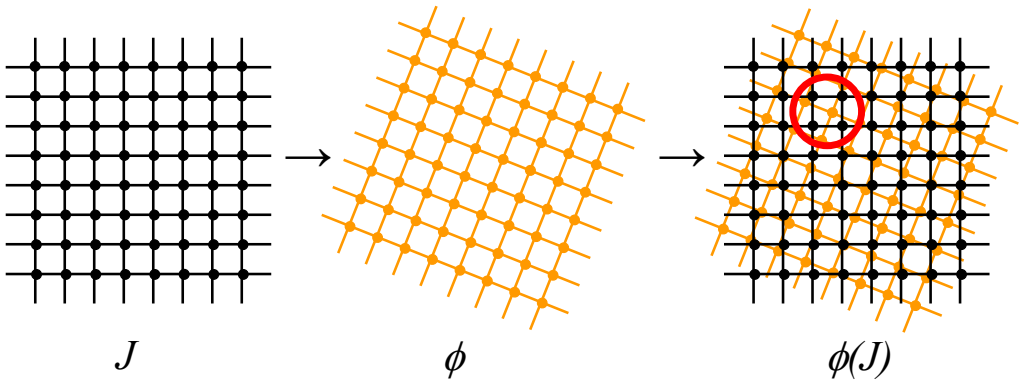


Figure 2.7.: Rotating an image while keeping the rasterization grid requires interpolation.

tasks. Using the center of the images as coordinate system offset should be a good general choice, however there might be application-specific indications that suggest a different center. As an example, for patient positioning in Radiation Therapy, one would use the iso-center of the linear accelerator, as the target anatomy is located there.

2.4. Interpolation

So far we have assumed that the image I is evaluated at its original grid locations, while J is mapped with the transformation ϕ. We did not point out, however, how the transformed image intensities $j_k := J(\phi(\vec{x_k}))$ are computed. It requires *Interpolation*, as illustrated for a 2D rotation in figure 2.7. There are two contrary approaches to it. An empty target image in the desired grid of I can be created and every pixel of J be "splatted" onto it with a certain footprint (using the inverse transformation ϕ^{-1}), typically a gaussian kernel. This is known as *Forward-Warping*; it is mostly used for interpolating scattered data into grids (see also section 3.3), in general it is problematic due to potential holes in the target image (depending on the transformation & kernel size), and limited efficiency. Therefore, in the majority of cases *Backward-Warping* is used, as described in the following.

2.4.1. Kernels

For every pixel $\vec{x_k}$ of the target image, an interpolation kernel centered at $\vec{x_k}$ considers neighboring grid values around $\phi(\vec{x_k})$. Mathematically, this corresponds to a convolution of the image J with a filter kernel h. In the following we make the simplification that physical image locations are equivalent to pixel/voxel coordinates, i.e. the grid spacing is 1. In reality, one has to divide by the actual spacing in every dimension before applying the subsequent equations. The convolution expands to:

$$J(\vec{x}) = (J * h)(\vec{x}) \tag{2.43}$$

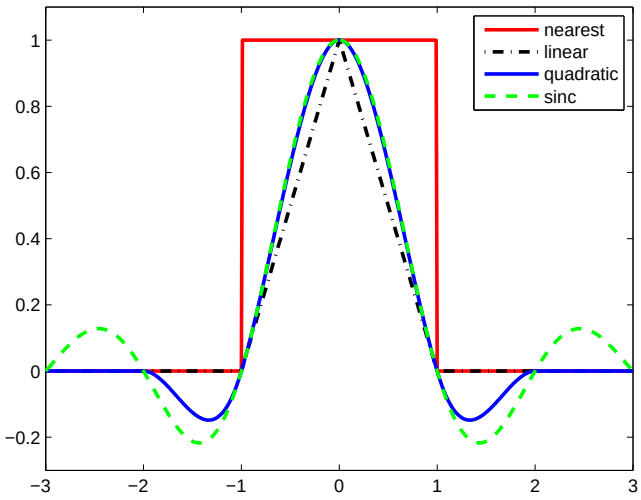


Figure 2.8.: Plots of 1D-Interpolation kernel functions.

Discretized for one, two or three dimensions, this expands to:

$$\begin{aligned} J(x) &= \sum_k J(k)h(x-k) \\ J\left((x,y)^T\right) &= \sum_k \sum_l J\left((k,l)^T\right) h_{2D}\left((x-k,y-l)^T\right) \\ J\left((x,y,z)^T\right) &= \sum_k \sum_l \sum_m J\left((k,l,m)^T\right) h_{3D}\left((x-k,y-l,z-m)^T\right) \end{aligned}$$

where k, l and m are integer pixel grid locations of J, and $\vec{x}$ (and thus its components x, y and z as well) denotes an arbitrary non-grid location.

Usually, interpolation kernels are symmetric and separable, hence two- and three-dimensional interpolation can be achieved by successively using h in each dimension:

$$\begin{aligned} h_{2D}\left((x,y)^T\right) &= h(x)h(y) \\ h_{3D}\left((x,y,z)^T\right) &= h(x)h(y)h(z) \end{aligned}$$

Following are the most relevant choices for a polynomial interpolation kernel:

$$\text{nearest neighbor: } h(x) = \begin{cases} 1 & |x| < 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{2.44}$$

$$\text{linear: } h(x) = \begin{cases} 1-|x| & |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.45}$$

$$\text{cubic: } h(x) = \begin{cases} 1-2|x|^2+|x|^3 & |x| < 1 \\ 4-8|x|+5|x|^2-|t|^3 & 1 \le |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.46}$$

A plot of those kernels is shown in figure 2.8. In the listed order, they are fast → slow, and low-quality → high-quality. The nearest neighbor interpolator just picks the closest grid point, which can easily be seen by the fact that for any x there is exactly one integer satisfying $|x - k| < 0.5$. Linear interpolation uses two grid points in every dimensions, cubic interpolation four.
There is many more choices for interpolation kernels, see Lehmann et al. [80] for an excellent survey. Further options are quadratic, cubic B-Spline, Lagrange, Gaussian and sinc-Kernels. An optimal interpolation kernel reflects the multiplication with a rectangular function in the fourier domain, which is in theory achieved by the sinc function (also shown in figure 2.8):

$$h(x) = \frac{\sin(\pi x)}{\pi x} = \mathrm{sinc}(x) \tag{2.47}$$

However, this kernel requires infinite support and is therefore computationally prohibitive in most applications. A variety of windowed and truncated versions exist; look-up tables are used since sine computations are expensive.

2.4.2. Implementation

A note about the nomenclature beforehand: If interpolation is applied in two or three dimensions, the prefixes "bi-" and "tri-" are often used, therefore one would write e.g. "bi-cubic" or "tri-linear" interpolation. In the following, the operators $\lfloor\rfloor$ and $\lceil\rceil$ determine the next smaller and larger integer, respectively. Computing the intensity of J at an arbitrarily (transformed) location $\vec{x} = (x, y, z)^T$ with nearest neighbor interpolation in 3D is simple:

$$J(\vec{x}) = J\left(\lfloor x + 0.5\rfloor, \lfloor y + 0.5\rfloor, \lfloor z + 0.5\rfloor\right) \tag{2.48}$$

Linear interpolation in 3D is computed as:

$$u = x - \lfloor x\rfloor; \quad v = y - \lfloor y\rfloor; \quad w = z - \lfloor z\rfloor$$

$$\begin{aligned}
J(\vec{x}) = (1-u)(1-v)(1-w) &\cdot J\left(\lfloor x\rfloor, \lfloor y\rfloor, \lfloor z\rfloor\right) + \\
u(1-v)(1-w) &\cdot J\left(\lceil x\rceil, \lfloor y\rfloor, \lfloor z\rfloor\right) + \\
(1-u)v(1-w) &\cdot J\left(\lfloor x\rfloor, \lceil y\rceil, \lfloor z\rfloor\right) + \\
(1-u)(1-v)w &\cdot J\left(\lfloor x\rfloor, \lfloor y\rfloor, \lceil z\rceil\right) + \\
u(1-v)w &\cdot J\left(\lceil x\rceil, \lfloor y\rfloor, \lceil z\rceil\right) + \\
(1-u)vw &\cdot J\left(\lfloor x\rfloor, \lceil y\rceil, \lceil z\rceil\right) + \\
uv(1-w) &\cdot J\left(\lceil x\rceil, \lceil y\rceil, \lfloor z\rfloor\right) + \\
uvw &\cdot J\left(\lceil x\rceil, \lceil y\rceil, \lceil z\rceil\right)
\end{aligned} \tag{2.49}$$

The image intensities are evaluated at the 8 corners of the grid cube where $\vec{x}$ falls within, and summed with the corresponding linear weights. A special adaption of this tri-linear interpolation is *Partial Volume Interpolation* (PVI), for improving the smoothness of histogram-based similarity measures. Here the Joint Histogram is updated with the intensity of all 8 voxels, for each of them the corresponding weight at the left side of equation 2.49 is added.

2.4.3. Choosing the Right Method

The requirements of registration algorithms with respect to interpolation are different than for other imaging applications. In general, applying interpolation can introduce artifacts and act as a low-pass filter, i.e. results in smoothing. For high-fidelity visual displays, none of these are acceptable, therefore often more sophisticated techniques are used. For registration, however, a certain amount of smoothing might even have a positive effect. On the other hand, an efficient technique is necessary here, since voxels need to be transformed millions of times per second. Following are some aspects that might clarify the right choice.

- Linear interpolation is often the best tradeoff. It introduces a limited amount of smoothing, however also results in a smooth transition of intensity values. In contrast, Nearest Neighbor interpolation just picks the closest voxel, therefore an image similarity criterion might end up with discontinuities when transformed voxels "jump" from one neighbor to the next. Cubic interpolation in turn can be too expensive to compute, especially in 3D, where $4 \times 4 \times 4 = 64$ voxels need to be queried.

- If one rather smooth image is used as the moving image in registration, nearest neighbor interpolation can be sufficient, while providing a significant speed-up (one voxel access in 3D instead of 8). An example is multi-modal registration of MRI with functional PET/SPECT data, the latter containing rather fuzzy clouds of nuclear activity.

- Different interpolation can be used along different axes in two- or three-dimensional images, especially when their scaling is non-isotropic. Nearest neighbor interpolation might make sense for a rather over-sampled dimension in an image, while the other axes should be linearly interpolated. On the other hand, e.g. a CT scan often has high resolution within the axial cross-sections, but a very large spacing within the slices ($10mm$ or more for a chest CT, for instance). Here cubic interpolation might be used for the head-feet axis, and linear one for the other two. This idea taken to an extreme, one can spend even more effort to reconstruct within sparse CT slices by registration-based interpolation [103].

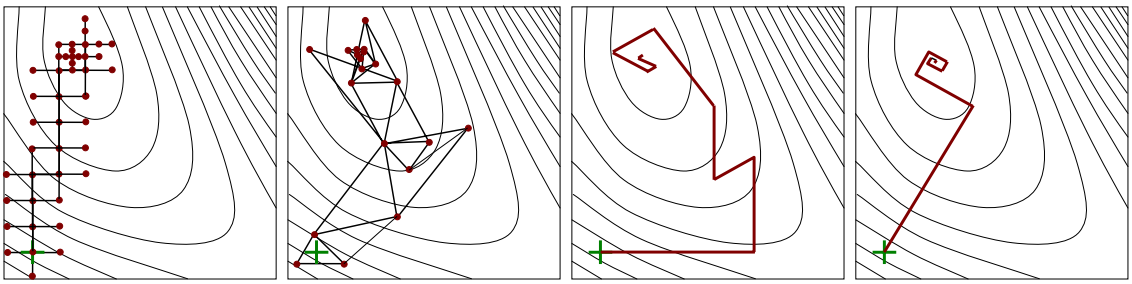
- Partial Volume Interpolation is a very good idea for Correlation Ratio, Mutual Information, or other similarity measures that require a joint histogram. The only additional computational effort with respect to tri-linear interpolation are the extra insertions (and therefore memory accesses) into the joint histogram, 8 instead of a single one. The smaller the size of the used histograms is, the less improvement PVI will yield over ordinary interpolation.

2.5. Optimization Algorithms

The goal of an optimization algorithm is to find the parameter vector $\hat{x}$ that minimizes (or maximizes) the value of a cost function F.

$$\hat{x} = \arg\min_{\vec{x}} F(\vec{x}) \tag{2.50}$$

The algorithm searches the parameter space with a specific scheme iteratively. It terminates once some abortion criteria has been satisfied, for instance if the change in the cost function value is below a limit (functional tolerance) or the search distances in the parameter space



(a) Hill Climbing (b) Simplex (c) Powell-Brent (d) Steepest Descent

Figure 2.9.: Search strategies of different optimizers on a two-dimensional function. Red dots in (a) and (b) represent single cost function evaluations, red lines in (c) and (d) whole line minimizations.

drop below a given value (parameter tolerance). Optimization schemes have always been developed with the aim of getting along with as few cost function evaluations as possible, therefore taking rare samples of the cost function value and using a nifty strategy to decide where to set the next try in the parameter space. This is especially crucial for image-based registration, as the cost function F equals the similarity measure S, whose computation requires the traversal of possibly quite large images. Besides, since it depends on the actual image content, this similarity is usually highly non-linear. Therefore the class of optimization problems that we have to solve is *unconstrained non-linear optimization*.

2.5.1. Direct Search Methods

Hill Climbing

Probably the simplest optimization scheme, Hill Climbing (also: Best Neighbor Search) evaluates a number of neighbors of the current parameter estimate in every iteration. The neighbor which yields the best cost function value is adopted as the central estimate for the successive iteration. A popular strategy to select the neighbors is to just add and subtract a certain step size to every parameter separately. This results in $2N$ cost function evaluations per iteration, N being the dimension of the cost function (e.g. $N = 6$ for a similarity measure wrt. a rigid transformation). If no better estimate is obtained, either the step size is reduced, or the algorithm terminates. The popularity of this Optimizer probably arises from its simplicity, however it has several limitations that one should be aware of. Since the step size is altered for all parameters together, it is important that a proper parameter scaling has been applied before, such that the effect of individual parameters on the cost function are in the same order of magnitude (see section 2.3.5). Besides, if the path toward the optimum is along a narrow rim, the step size will be drastically reduced until the optimizer starts heading in the right direction, converging then requires a large number of evaluations. The straightforward implementation does not include a strategy to ever increase the step size again, after it had been reduced. Figure 2.9(a) depicts an example where the step size is actually too small from the beginning. Because this search strategy is so simple, different researchers often implement their own variations. We have used a method that evaluates all combinations of the tree parameter changes [forward, keep, backward], and termed it *Exhaustive Hill Climbing*

[169]. For a rigid transformation, it requires $3^6 = 729$ cost function evaluations per iteration, however it achieves better convergence for rough cost functions.

Nelder-Mead Simplex

A more advanced search strategy is the Simplex algorithm [115]. A simplex is a minimal geometric shape, consisting of $N + 1$ corners in N-dimensional space (a triangle in 2D, a tetrahedron in 3D, and so on). Around the initial parameter estimate, a starting simplex is defined, and the cost function is evaluated at its corners. Depending on the results, the shape of the simplex is changed according to rules for reflection, extension and contraction, see figure 2.9(b). Advantages of this method are that it relies on no assumptions on the cost function whatsoever, and it uses the minimal number of evaluations that span the parameter space. Besides, since the Simplex shape is flexible in all dimensions, this optimizer is much more efficient traveling through narrow regions, as opposed to Hill Climbing.

Pattern Search

The concept of the described two methods can be generalized as Pattern Search algorithms. They all have in common that they use an adaptive search pattern, independent of the type of cost function, in order to explore the cost function. According to the evaluation results, the position, orientation and shape of that pattern is altered successively in each iteration. A general discussion of Pattern Search algorithms, including a theory of the convergence behavior, can be found in [151]. Besides the Hill Climbing and Simplex methods, this category contains for instance the Hooke-Jeeves pattern search method. Surrogate optimization is a related approach where the cost function values are approximated based on some underlying knowledge on F and prior evaluations. Due to the non-predictability of image similarity metrics, it is usually not well applicable for image registration.

Powell's Direction Set Method

This algorithm 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 efficiently with the *Brent Line Minimization* method. It uses both parabolic interpolation and golden section search, choosing dynamically in each step which is is more appropriate. It is described very detailed in [115], other methods for bracketing a minimum and root finding can be found in [87].
The second problem is now to find directions in the parameter space, such that a line minimization along one of them does not spoil the minimizations done along the former directions. According to Powell's scheme, the first set of directions are the individual degrees of freedom themselves. After N line minimizations, a new first direction is established as the vector that leads from the last to the current estimate. This makes sure that the first line minimization of each new iteration is done along the way that the parameters improve most. From the old directions, the one where the largest move took place among the previous line minimizations gets "fired". It is most likely to be contained in the newly selected direction anyway, therefore the directions are kept as independent as possible. Figure 2.9(c) shows an example optimization run.

The only settings that need to be defined affect the line minimizer, such as termination constraints. The implementation proposed in [115] spends a large number of evaluations in each line minimization. Depending on the actual problem, one might consider adapting the line minimization termination parameters, linking them to the overall number of iterations (i.e. perform faster line searches in the beginning).

2.5.2. Least-Squares Methods

Another important set of methods can be used if the cost function is a sum of squares of non-linear functions:

$$F(\vec{x}) = \frac{1}{2}\sum_{i=1}^{m} f_i(\vec{x})^2 = \frac{1}{2}||\vec{f}(\vec{x})||_2^2 \tag{2.51}$$

This type of function often occurs if a model is to be fitted to data, where $\vec{x}$ is a parameter vector for the model function and the individual $f_i(\vec{x})$ express the distance from a specific part of the model to a data element. Both the gradient vector and the Hessian matrix of eq. 2.51 have a special structure, allowing adapted optimization algorithms like *Gauss-Newton* and *Levenberg-Marquardt* to be used [48].
In particular the SSD similarity measure can efficiently be optimized by such methods.

2.5.3. Derivative-based Methods

If the gradient vector of F is available, it can be used as base for determining the next steps, as done in the *Gradient Descent* or *Conjugate Gradient* methods. There are applications where the cost function gradient can be calculated much faster than the function value itself, boosting the optimization time accordingly. Some registration approaches use derivative approximations, especially when Mutual Information is used as similarity measure [154, 172].

Gradient Descent Method

A very basic way to find a maximum is to step successively in the direction of the function gradient.

$$\vec{x_{k+1}} = \lambda \frac{df(\vec{x_k})}{d\vec{x_k}}$$

The factor λ is a constant known as the learning rate. It is positive if we are seeking for a maximum, and negative otherwise. Unfortunately the success of the optimization depends very much on the right choice of λ. If it is too big, steps are taken too far and may skip the optimal position repeatedly. However, if it is set too small, the algorithm may get stuck before the optimum, as the step size depends both on λ and the absolute size of the function gradient. For that reason, various approaches have been developed to adapt the learning rate, often based on particular knowledge about the cost function.
An extension of this algorithm conducts line maximizations along the direction of the gradient, resulting in the *Steepest Descent* method. This does not rely on the empirical definition of a learning rate anymore. Every new iteration then starts from on optimum on the direction of the previous gradient, thus the component of the new gradient in that old direction will always be zero. This means that all iterative steps are perpendicular to each other, which is not very efficient (figure 2.9).

This drawback has lead to the development of more sophisticated algorithms involving multiple old directions, the *Conjugate Gradient* methods. Yet another category are *Quasi-Newton* methods, which construct an estimate of the Hessian matrix of F to achieve faster convergence. See [115] for details.

2.5.4. Quasi-global Methods

For many optimization problems a danger exists of getting trapped in local optima of the cost function. One possible solution is to repeatedly start the optimization from different starting points in the parameter space, eventually choosing the best result as optimum. Another workaround comprises adding noise to the cost function value, either purposedly, or as byproduct of an randomized cost function approximation [154]. This results in a stochastic optimization approach. Another widely used technique to overcome local optimas are *Simulated Annealing* methods, which sometimes take steps in bad directions with decreasing probability. That "bouncing around" in the parameter space reflects the movement of molecules when matter is in the transition from fluid to solid state. *Particle Swarm* methods perform a number of concurrent searches with a certain collaboration strategy [158]. Yet another class are *Genetic Algorithms* [107].
Note that one has to decide beforehand if the goal is indeed to find a global optimum, since many similarity measures produce only a local optimum at the correct alignment.

2.5.5. The Right Optimizer for Registration

The selection of an optimizer for an image-based registration affects mostly its computation speed, robustness and capture range, in the sense of a trade-off between those factors.
Direct search techniques can mainly be distinguished by how broad the path is they explore within the parameter space. Hill Climbing uses more evaluations than Simplex, and therefore might work better in certain cases, in particular if the individual parameters have little dependencies. On the other hand, Simplex is generally more efficient and flexible at the same time. For highly non-linear and "bumpy" cost functions, a more exhaustive local search might be required. The Powell-Brent method is based on robust successive line minimizations, which might come in handy, on the other hand it is usually considered one of the less efficient techniques.
Least-squares methods can provide a significant boost in speed and robustness, given that it is possible to express the similarity measure as a sum of squares. This is natively given for the simple SSD (dis-)similarity measure. A word of caution here: Many of the other measures presented in section 2.2.3 appear to have a least-squares form, in particular if they are rewritten based on Generalized Correlation Ratio (equation 2.27). Since other variables in the equations, like the functional intensity mapping f or the image variances, change in every computation, they are in fact not valid least-squares expressions, and should therefore be maximized by general non-linear optimizers.
Derivative-based methods should definitely be used when there is an efficient means to compute the similarity measure gradient, and its values are numerically stable. It is a bad idea to estimate the gradient by using finite difference evaluations around the current estimate, since this will negatively affect both speed and robustness.
Quasi-global methods are only eligible if one's intention is indeed to achieve the global optimum in the parameter space, independently from the starting location. Designing the

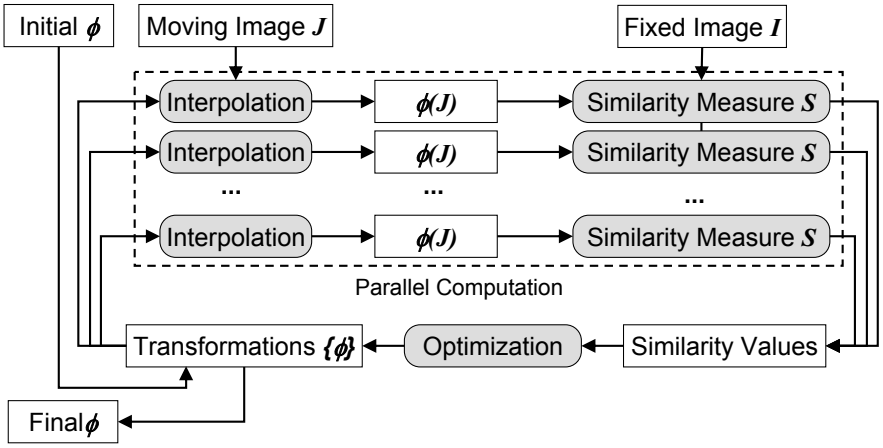


Figure 2.10.: Parallelized scheme of an image-based registration algorithm.

similarity measure in a way that it actually produces the highest similarity value at the correct alignment is not trivial, to some extend because of the normalization issues pointed out in section 2.2.5. Often however, a registration algorithm requires an initial estimate within a certain proximity to the correct solution. A local search method then iterates towards the closest local optimum, which supposedly represents the correct alignment.

2.6. Acceleration Techniques

The requirements in terms of the computation time that a registration algorithm may need, vary greatly depending on the clinical application. They can be quite demanding in situations where the patient is present when the registration is performed (often an interventional setup).

2.6.1. Parallel Computing

In general, the future of high-performance computing seems to be highly parallel and distributed. Common buzz-words in this context are *Grid Computing* and *Ubiquitous Computing*. Computing clusters, where a number of workstations are connected with a very fast network, have been around for a long time. Consumer PCs used to have a single CPU, though, which has changed rapidly. It started with tricks like Intel's Hyperthreading, now Core Duo chips (two CPUs), IBM cell (8+1 CPUs) etc. Today, virtually no new PCs are sold anymore that feature only a single CPU, so software application developers better get used to the new opportunities & challenges of multi-threaded programming.
Most of the non-linear optimization strategies suited for image-based registration provide a good anchor point for distributed computing. All cost function evaluations within one iteration, i.e. before the algorithm decides on a new parameter estimate from which to launch all further evaluations, can be processed in parallel. An illustration of this is provided in figure 2.10. The parallel implementation of a number of different optimizers for image registration is described in [157]. In own work, [168] and Appendix A, we had achieved a speedup of one

order of magnitude for a 2D-3D registration algorithm, by distributing the 12 evaluations per iteration of a Hill-Climbing optimizer onto an Infiniband-based cluster.

2.6.2. GPU Processing

The demand of consumer industry for more and more complex graphics has driven a fast-paced development of affordable yet powerful Graphics Processing Units (GPU) in the last decade. Since it is straightforward to parallelize the visualization of 3D geometry, graphics hardware today features a large number of computing units. A recent consumer graphics card at the time of writing (summer 2007, NVidia GeForce 8800 Ultra) features 128 parallel processing units, 768 MB of memory and 103.7 GB/sec memory bandwidth, for less than 800 dollars. This by far exceeds the specifications of current PC hardware. Together with full programmability, supported by high-level GPU programming languages and APIs (Cg, GLSL, Cuda, to name a few), GPUs have become an ideal platform for scientific computing. The research field that has emerged is called General Purpose GPU (GPGPU), see [97] for an overview.
Graphics hardware has been used in particular for 2D-3D registration, since the involved computation of DRRs is a special case of Volume Rendering which can ideally be performed on a GPU [70, 163]. We had evaluated projection-based 3D-3D registration method on graphics hardware in [71]. Additional focus was on the computation of image similarity measures on the GPU [27].

2.6.3. Random Sampling

In particular when registering 3D volumes, most of the computation time is spent traversing the volumes while evaluating the similarity measure. An alternative approach is to just pick a number of intensity tuples (i, j), randomly selected from Ω. Such a random sampling approach can potentially yield an immense speed up, however it requires significant adaption of the overall algorithm. A similarity measure computed based on random samples will have a certain amount of jitter, depending on the number of samples used. Special optimization algorithms that take randomness into account, can be used to deal with it. A well-known technique based on a Stochastic Maximization of Mutual Information has been proposed by Viola et al. [154, 153].

2.7. Evaluation

A crucial part of the development of an image-based, potentially automatic, registration algorithm is its evaluation. Most importantly, the quality of alignment has to be assessed.
This can be done purely qualitatively by visual assessment. A number of experts in the involved image modalities should visually determine on a statistically significant number of image pairs and executed registrations, if the alignment is satisfactory. Unfortunately, the outcome will depend to a certain degree on "human factors", as well as the actual visualization used to present the registration result (e.g. color overlay, checkerboard overlay, superimposition of outlines, linked pointer etc.).
Quantitative assessment is generally preferred, but often requires more effort. In general, a so-called *Gold Standard* technique is used to compute an image alignment that is known to be better than the method under investigation. If that is not guaranteed, then at least

its accuracy has to be known; in this context manual techniques are often denoted "Bronze Standard" [64]. Applying a Gold Standard method results in a so-called *Ground Truth* registration, which is then compared with the image-based algorithm.

2.7.1. Establishing Ground Truth

Since image-based registration registers anatomy within the patient, the method for establishing the Ground Truth has to do likewise. Often a registration based on corresponding points is used, which can be computed in a closed-form manner with [152] or [160]. The points can be anatomical landmarks defined by an expert, or actual fiducial markers visible in both modalities. The first is again subject to human variability, the locatability of the landmarks, and the used visualization. With reasonable effort though, a solid Ground Truth registration can be established using such a manual landmark definition. Fiducial landmarks provide ultimate accuracy, however they are invasive. Implanting fiducials into patients therefore might be difficult to justify within research about a proposed registration technique, unless it is part of the current clinical procedure to achieve registration (as the case e.g. in stereotactic radiosurgery). Often an iterative evaluation is done starting with synthetic data and ending with real patient data. As for synthetic data, one of the images involved in the registration process is usually generated, therefore the Ground Truth is known. Then a number of more or less realistic phantoms can be used, which enclose fiducial markers or other means to establish the correct registration. If appropriate, cadaver studies might be conducted. While marker implantation is not an issue here, imaging on a dead body might not yield useful results, depending on the modalities and target anatomy. Eventually, the final stage of evaluation has to be done on real clinical data, with whatever methods to establish Ground Truth are feasible.

2.7.2. Alignment Errors

Target Registration Error

The most direct measure of misalignment stems from the use of point correspondences. Let $\vec{p}$ be a point in image I, and $\vec{q}$ represent the same point, i.e. the same anatomical location, in image J. The Target Registration Error (TRE) then simply writes as

$$\mathrm{TRE} = |\phi(\vec{p}) - \vec{q}| \tag{2.52}$$

The significance of this error lies within the notion that $\vec{p}, \vec{q}$ should reflect a clinical target point. For registration in the context of interventional oncology, it would typically be the center of a malignant structure that is to be ablated, irradiated or resected. Therefore it denotes the location within the patient's body, where the highest accuracy of alignment is required.
It is possible, and often meaningful, to establish multiple clinical targets. Those points can reflect several lesions spread over the organ, or border points of a single large tumor. The error is then usually expressed as Root-Mean-Square (RMS) value:

$$\mathrm{TRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\phi(\vec{p}) - \vec{q}|^2} \tag{2.53}$$

Generally it is preferred to use multiple target points for evaluation, as the size of the target is reflected in the error value then. Besides, rotational errors of a registration algorithm can be detected better, while in the case of a single target a wrong rotation might not increase the TRE significantly (given that the rotation center is close to the target). If this is not possible, e.g. for a rather homogeneous tumor mass, the target might be defined as a sphere. A special variant of a TRE (with reciprocal meaning) would then be the amount of sphere overlap.

Fiducial Registration Error

This error uses point correspondences as well, however they are specifically selected for best possible locatability. They might also be the location of implanted fiducial markers for establishing Ground Truth, as discussed above. If $\{(\vec{p_i}, \vec{q_i})\}$ is a set of corresponding fiducial points in both images, the Fiducial Registration Error (FRE) is in analogy to the TRE expressed as an RMS value:

$$\mathrm{FRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\phi(\vec{p}) - \vec{q}|^2} \tag{2.54}$$

An FRE error value does not have clinical significance, but it might be more precise, since the fiducial points are selected for their uniquely identifiable location in 3D. For a rigid transformation, at least three non-collinear targets also allow to establish the Ground Truth transformation by directly computing the motion between the point sets (see above). For this, and the computation of FRE values, the accuracy of the actual point localization plays an important role. It is denoted Fiducial Localization Error (FLE) and should be taken into account when establishing error statistics. See also [53] chapter 6, and Fitzpatrick et al. [40] for a detailed study of the mathematical dependencies between TRE, FRE and FLE.

Deviation of Transformation Parameters

A measure of misalignment that is often used, but has purely technical meaning, is to compute some distance of the resulting registration transformation to the Ground Truth one. Rigid transformations are often parameterized as two 3-vectors denoting translation and rotation in Euler angles (see section 2.3.1). Stating the deviation in all components can reveal more information about the spatial behavior of an registration algorithm, however it results in six values which makes it difficult to summarize the algorithm performance. For more compactness, often the RMS or mean values of each the translation and rotation components are computed, resulting in two error values. Note that it is not meaningful to directly average translation and rotation error, since they describe two different geometric properties. To work around this, people often translate the errors by computing the displacement of the corners of a target bounding box, which is affected by both translation and rotation. This brings us back to the Target Registration Error though, with the box as an approximation for the actual clinical target.

2.7.3. Robustness

The robustness of a registration algorithm determines how well the correct alignment is reached for repeated execution of the algorithm, and on different image data. The first case is also defined as repeatability. A very common tool in this context are randomized

studies. The initial transformation parameters are randomly displaced a certain amount from the Ground Truth, then the registration algorithm is executed. This process is repeated, typically $100 - 1000$ times. The resulting distribution allows one to draw conclusions on the overall robustness, the number of outliers, and the performance in the individual transformation parameters. Often the mean and standard deviation of the individual transformation parameters are computed for studying the spatial behavior of the algorithm. If such studies are repeated with successively increased variance of the random displacements, one can also systematically evaluate the capture range of the algorithm, i.e. determine how far off the initial transformation may be to still reach the correct alignment.
Eventually, such an evaluation has to be done on a large number of realistic clinical data sets, to be statistically meaningful.

2.7.4. Similarity Measure

Often the Similarity Measure is the most crucial component in a new registration algorithm, it therefore deserves special attention with respect to evaluation as well. Investigating the measure usually happens at an earlier stage in the development. When it is known to behave as expected, one can turn toward optimizing the remaining algorithmic components and conduct the overall evaluation (by means described above).
When designing and tweaking a similarity measure, one usually computes a large number of plots that show the measure value with respect to a transformation parameter changed within some range from the correct alignment. Two parameters can be changed as well, resulting in a surface plot (see e.g. figure 5.6 further down in this document). This allows in addition to assess the amount of dependency between the chosen parameters, for instance if there is a rim of high similarity diagonally through the plot. One has to be aware though, that it only reflects a small portion of the parameter space to be evaluated. Any third parameter altered somewhat from the Ground Truth might totally change a plot with respect to the first two transformation parameters. Unfortunately there are no practical means to visualize plots in more than three dimensions in an intuitive way. Therefore one has to get back to a pure mathematical framework to cover, say, a full rigid parameter space. Skerl et al. [139] have proposed an evaluation protocol that includes scheme to properly sample the parameter space.

2.8. Summary

Designing a successful image-based registration algorithm requires careful consideration of their essential components.
In the case of multi-modal registration emphasized in this thesis, the similarity measure is arguably the most crucial ingredient. It should be designed based on all available knowledge about the involved imaging modalities. We have introduced the most common similarity measures with respect to assumed functional relationships of the image intensities and a noise term. A maximum-likelihood derivation has proven their consistency. Additional strategies exist to incorporate more information about the registration problem, and to make the similarity measure more robust with respect to deviations from the underlying model.
Deciding on the appropriate transformation model, as well as an interpolation which applies this transformation to the moving image, is usually much easier, yet very important for the overall outcome. An optimization algorithm has to be picked as well. This is often done

in a rather experimental way, however the best choice here also depends on some similarity measure properties, such as its smoothness with respect to the transformation parameters, the size of local and/or global optima, the availability of gradient information and the applicability of a least-squares formulation.
Last but not least, the evaluation and validation of a registration technique is not to be underestimated. For precise and realistic evaluation, Ground Truth information eventually has to be established right within the human body.

3. 3D Freehand Ultrasound

While 3D ultrasound transducers are now available by most of the manufacturers, tracked freehand ultrasound is nevertheless a very important technology. On one hand, it allows for having larger field-of-view required for acquisitions of whole organs. On the other hand, it is a preferred modality for interventional navigation applications, as both the ultrasound images and their location in space are acquired in real-time. A broad overview about different approaches to acquire three-dimensional ultrasound is provided in [39].

3.1. General Setup

A 3D freehand ultrasound system consists of an ultrasound machine, a tracking system, and a computer workstation (figure 3.1). The tracking system provides the transformation T_T, which is the position and orientation (commonly denoted as *Pose*) of a tracking sensor or target attached to the ultrasound transducer. The computer workstation records both the tracking information and the ultrasound images. The latter is typically achieved by using frame grabber hardware, which retrieves the analogue video signal from a corresponding output of the ultrasound machine. Depending on the ultrasound machine vendor, digital interfaces can be used instead to transfer the image information without loss of quality. The workstation runs a software which records the streams of tracking and image information in a proprietary file format on the hard drive. Often, it also features algorithms to load, view and reconstruct 3D ultrasound from the recorded sweeps.

Tracking systems that can be used for 3D freehand ultrasound, are briefly described in section 3.2.
To exploit the three-dimensional information contained in 3D freehand ultrasound recordings, often a reconstruction into a cartesian volume, or oblique plane, is required. We review possible strategies for this Spatial Compounding, and present a new efficient technique for it, in section 3.3.
Given a 3D pose provided by a tracking system, the transformation T_C to the actual ultrasound image plane has to be established. This problem of spatial calibration is addressed in section 3.4. Eventually, when combining tracking and image information, they should be synchronized as well. A constant delay between tracking and images can be recovered using temporal calibration methods. Depending on the hardware setup, this delay might vary over time - in particular when the tracking and ultrasound imaging occur with different update frequencies. This problem is further amplified if video grabbing is used, since the video frame rate is usually not synchronized with the update frequency of the actual ultrasound image. A method to recover the constant temporal lag is presented in section 3.4 as well.

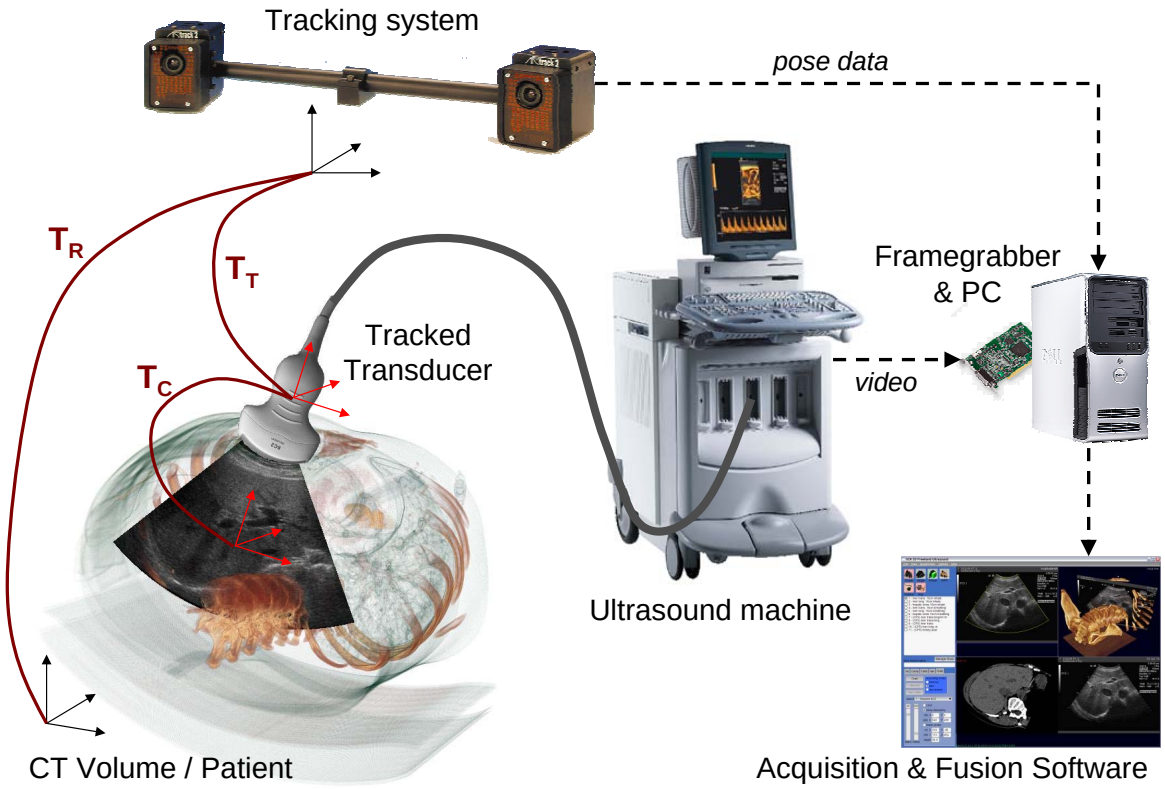


Figure 3.1.: Overview of a 3D freehand ultrasound system.

3.2. Tracking

A detailed description of the physical principles underlying most tracking systems can be found in [171]. In this work we use the optical and magnetic position sensing systems shown in figures 3.2 and 3.3.

3.2.1. Optical Tracking

An optical tracking system usually consists of two or more video cameras mounted at a fixed location, and optical markers that are attached in a certain pattern to the target to be tracked. Markers can be active (i.e. light-emitting) or passive (reflective to visible or infrared light). Exemplary systems are the NDI Polaris and Optotrack products, or ARTTrack2 (figure 3.2) by Advanced Realtime Tracking (ART) GmbH, Germany. The 3D location of every marker is precisely computed by triangulation from detected 2D locations in the video images. A special case is tracking with just one camera, possible if the marker arrangement is known. An example is the RAMP Augmented Reality System developed by SCR [73], which has watched over me during the last year of my dissertation (since my office is in the SCR Imaging Lab). Optical tracking can be extremely precise, depending on the amount of money one is willing to spend. Some systems are scalable in the sense that any number of cameras can be mounted somewhere in the workspace (16 for ART systems, for instance), a room calibration proce-

(a) camera pair (b) targets

Figure 3.2.: The ARTtrack2 optical tracking system.

dure then recognizes their spatial relationship. The greatest disadvantage of optical tracking systems is that they always require a direct line of sight between the tracking markers and the cameras (actually, at least two cameras if more are installed). Besides, marker balls can be too large and bulky for not being in the way, depending on the clinical use.

3.2.2. Magnetic Tracking

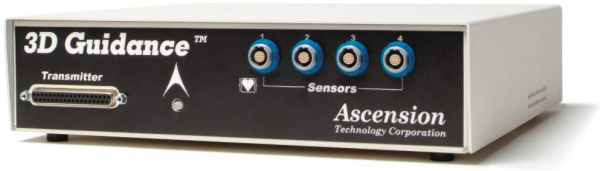
Magnetic position sensing uses a field generator, which generates a weak yet precisely defined electromagnetic field over a certain working volume. Small sensors, containing two or three miniature coils arranged in a perpendicular fashion, pick up the magnetic field, allowing to compute their location and orientation in space. Popular systems are the NDI Aurora and Ascension 3D Guidance (figure 3.3, their previous system was MicroBIRD) products.
Magnetic tracking sensors can be extremely small, and can therefore be integrated in catheters that are inserted into the human body for particular interventions. No line of sight is required between the field generator and the sensors. However, magnetic tracking is very susceptible to magnetic field distortion, arising whenever other ferro-magnetic material is brought close to the system. Note that it can be very challenging to remove all metal around the patient in a clinical environment. In particular, most patient beds and bed-rails contain a large amount of metal. Addressing this issue, Ascencion has developed a new *Flat Transmitter* for the 3D Guidance system (figure 3.3(a)), which can easily be placed below the mattress of the patient bed, shielding all metal below it. Even in an ideal metal-free environment, accuracy of magnetic tracking is rather low compared to their optical counterparts.

3.2.3. Image-based Tracking

If no external position sensing system is to be used, the content of the 2D ultrasound images can be used to estimate the frame-to-frame displacement. This is easy if the transducer is moved such that the ultrasound images stay in the same plane, and is used in panoramic imaging techniques which are nowadays available on products by all major ultrasound vendors [42]. On the contrary, estimating the out-of-plane motion is a more challenging task. Researchers have mainly used the fact that ultrasonic speckle patterns enter and exit the image plane when the transducer is slowly moved in elevational direction. Linear regression [112] or speckle decorrelation [47] are proposed techniques to compute the inter-frame dis-

(a) flat and mid-range transmitter

(b) sensors



(c) electronics unit

Figure 3.3.: The 3D Guidance magnetic tracking system.

placement. This technology is still in its infancy, the problem mainly being that the speckle statistics needs to be precisely known. However, depending on the anatomy, every portion of the image may contain anything from no speckle (e.g. a clear, bright tissue interface) to fully developed speckle. Besides, either the raw RF-signals have to be used to derive speckle statistics, or the parameters used for preparing the data for visual display, in particular the log-compression, have to be known. Since the motion between successive frames is estimated, this approach also suffers from error accumulation, resulting in a drift of later parts of the acquisition, and thus potentially a wrong scaling and/or shearing of the 3D sweep.

3.3. Compounding

Using the three-dimensional characteristics of 3D freehand ultrasound sweeps, in particular any-plane views, has many medical advantages, including

- independency from examiner and, to some extent, used probe positions
- freedom to display pathological process in any angle, e.g. along and perpendicular to its main axis for visualizing its full extend, or planes that focus on relations to relevant neighbouring normal tissue structures
- possibility to visualize planes parallel to the skin, that cannot directly be derived from diagnostic sweeps in B-mode
- upvalue ultrasonography into a comparable line with other sectional imaging modalities that allow for free choice of plane at acquisition (e.g. MRT) or reconstruction (e.g. multi-slice CT)

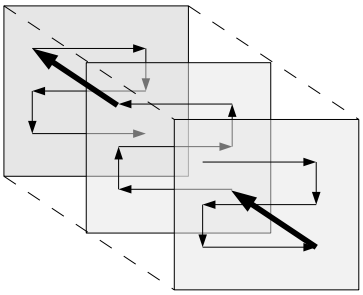


Figure 3.4.: Volume traversal scheme for advancing only one voxel at a time.

To exploit them, a reformatting either into a cartesian volume or a plane arbitrarily located in space is necessary. We denote this process *Spatial Compounding*, however there is some confusion about this term. Spatial Compounding is sometimes used to describe the integration of ultrasound pulses focused in different directions, to decrease the amount of speckle in the images (e.g. the SonoCT feature available on Philipps systems). This is a feature directly implemented on the respective ultrasound systems, while the compounding we are addressing is the reconstruction of 3D ultrasonic information from scattered 2D slices. For that reason, a number of researchers use the general term *Ultrasound Reconstruction* instead [122]. There are two distinct approaches to it, as already pointed out by [12].
A footprint of each of the B-mode images scattered in space can be created in the initially empty 3D volume. If an additional volume channel is used, an averaging can be achieved where several ultrasound planes intersect the same voxels, otherwise often the maximum is used as final intensity [123]. This so-called forward-warping is computationally efficient, while it has the potential to cause gaps in the reconstruction. It can also be used to directly create Multi-Planar Reconstructions (MPRs) by assuming a constant elevational thickness of each ultrasound image [111]. For this purpose, a polygon depicting the intersection of each ultrasound image with the desired plane is drawn onto the screen with hardware-accelerated texture-mapping.
A backward-warping strategy would traverse the target plane or volume, for each grid point identify the relevant original ultrasound information and accumulate the voxel intensity using e.g. distance-weighted interpolation or other merging schemes. In the following, we present an algorithm implementing this approach efficiently, despite the obviously high computational effort. A simplified approach is taken in [32], where a continuous probe motion without any ultrasound plane intersections is assumed, henceforth each voxel intensity is weighted from the two neighboring ultrasound slices, using the probe trajectory rather than the perpendicular projections.

3.3.1. New Methods

For every discrete position $\vec{x_i} \in V$ in the reconstruction volume V, we need to compute a set of tuples $A_i = \{(d, y)\}$ where d is the distance of an original ultrasound data point to $\vec{x_i}$ and y its intensity value.

$$\forall \vec{x_i} \in V : A_i = \{(d, y) | d < D; d = \|\vec{p} - \vec{x_i}\|\} \tag{3.1}$$

$$\vec{p} = H_j(u, v, 0, 1)^T; \quad y = Y_j(u, v) \tag{3.2}$$

Here, D is the maximum distance at which points should be taken into account for accumulation of the voxel intensity. The homogenous 4x4 transformation matrix H_j for a particular ultrasound slice j maps a 2D point $(u, v)^T$ of the slice plane into a point $\vec{p}$ in 3D. The corresponding ultrasound image intensity is denoted $Y_j(u, v)$. The main effort is now to determine the set A_i for each voxel from the whole ultrasound information available. However, we can restrict ourselves to points $\vec{p}$ originating from slices whose perpendicular distance to $\vec{x_i}$ is smaller than D:

$$S_i = \{j | d_{i,j} < D; d_{i,j} = (0, 0, 1, 0) H_j^{-1} \vec{x_i}\} \tag{3.3}$$

Fast slice selection

In the following we devise an efficient means to successively compute S_i for all voxels, using the following facts:

- We can traverse the volume in a way such that the distance of two successive voxels is always $\|\vec{x_{i+1}} - \vec{x_i}\| \leq k$, where k is the maximum voxel spacing. Figure 3.4 illustrates a simple possible scheme.
- If the distance of ultrasound slice j to $\vec{x_i}$ is $d_{i,j}$, then the distance of the same slice to voxel $\vec{x_{i+1}}$ can not be smaller than $d_{i,j} - k$.
- Hence this mentioned slice j can only be required for an voxel index $i + \lfloor d_{i,j}/k \rfloor$ or later.

We use a rotation queue with $\lceil k/D_V \rceil$ elements (D_V being the volume diagonal), which is equivalent to the maximum distance of a slice contributing to the reconstruction, from a particular voxel. Each element contains a set of slice indices, at the beginning all slices are in the head element of the queue. For every voxel $\vec{x_i}$, all slices in the queue head are removed, their distance d is computed and they are reinserted into the rotation queue corresponding to that distance. If $d < D$, then the slice is added to S_i and considered for accumulation of the voxel intensities. For the next voxel $\vec{x_{i+1}}$, the rotation queue head is advanced.
This will allow us to implement efficient backward-warping compounding methods to reconstruct three-dimensional volumes of ultrasound information. In order to create online Multi-Planar Reconstructions (MPRs) directly from the original data, we will just consider a reconstruction volume with a single slice, arbitrarily located in space.

Intensity accumulation

For a voxel $\vec{x_i}$, all pixels on each slice $\in S_i$ that satisfy $d < D$ are added to the set of distance-intensity tuples A_i defined in equation 3.1. For a given set $A = \{(d_i, y_i)\}$, the reconstructed voxel intensity can be any weighting or selection function $f(A)$. We considered the following functions in our work:

Inverse Distance Weighting. Originally defined in [135], it assures that the reconstructed intensity approximates the original data values for $d \to 0$ ($\mu > 1$ being a smoothness parameter):

$$f(A) = \sum_{i=1}^{n} y_i \frac{d_i^{-\mu}}{\sum_{j=1}^{n} d_j^{-\mu}} \tag{3.4}$$

Gaussian Weighting. A 3D Gaussian kernel with size σ can be used to weight the data values as well:

$$f(A) = \frac{\sum_{i=1}^{n} y_i e^{-d_i^2/\sigma^2}}{\sum_{i=1}^{n} e^{-d_i^2/\sigma^2}} \tag{3.5}$$

Nearest Sample. Here, the data value closest to the considered voxel is directly accepted as reconstructed intensity:

$$f(A) = y_i | d_i = min\{d_i\} \tag{3.6}$$

Weighted Median. Using the median has the advantage that one of the original intensities is chosen, which would be the centermost one from the sorted intensities $[y_i]$. In order to incorporate the distances, we "stretch" the sorted list with their respective inverse linear weightings [1]

$$\begin{aligned} \forall k \in [1 \dots n] : y_{k+1} \geq y_k,\ w_k = 1 - \frac{d_k}{D} \\ f(A) = y_i \left| \sum_{k=1}^{i-1} w_k \leq \frac{\sum_k^n w_k}{2} \leq \sum_{k=1}^{i} w_k \right. \end{aligned} \tag{3.7}$$

3.3.2. Results

Computation time

We implemented a forward-warping compounding algorithm for comparison purposes, which averages all ultrasound pixel hits onto reconstruction volume voxels, as described in [123]. The following table compares the impact of volume and slice resolution on both forward and backward compounding.

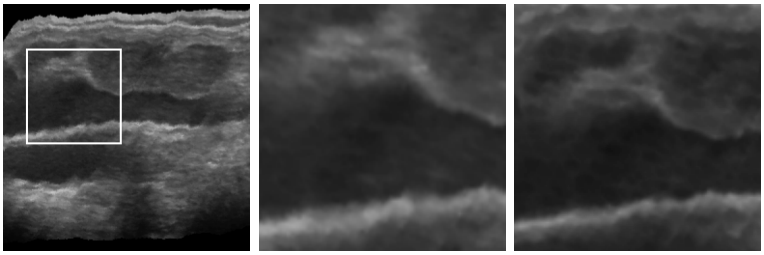
	forward		backward multiple		backward single	
D_V	$D_S = 256$	$D_S = 454$	$D_S = 256$	$D_S = 454$	$D_S = 256$	$D_S = 454$
16	517s	1687s	226s	401s	119s	295s
134	538s	1915s	712s	942s	442s	510s

D_V denotes the number of million voxels of the reconstruction volume, D_S is the width and height of the input ultrasound slices. The times were taken using a set of 1024 slices, on an AMD64 3200+ with 1GB RAM. While forward compounding is largely unaffected by the number of voxels, the amount of pixels to be processed linearly affect the computation time. Backward compounding reacts to changes in both input and target data, however the increase depending on the size of the slices can be largely eliminated if only a single pixel of each slice is required for each voxel as in a nearest neighbor accumulation.

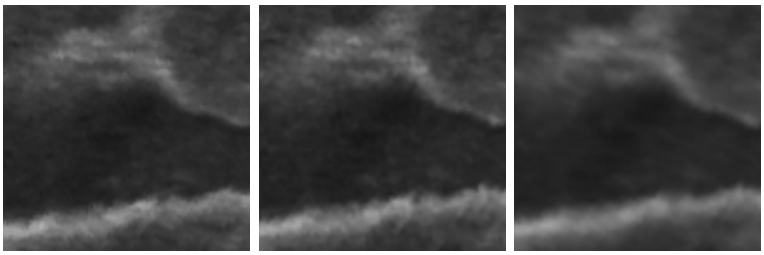
Qualitative and Quantitative Comparison

Figure 3.5(a) shows one reslice approximately orthogonal to the original slices. 200 slices were processed using different accumulation schemes. Mean-Squared-Error (MSE) values were calculated according to the leave-one-out strategy for the whole reconstruction volumes,

[1] Linear mapping $[0, D] \rightarrow [1, 0]$, pixel beyond distance D are disregarded and therefore have weight 0.



(a) MPR slice (b) backward weighted median (MSE 54) (c) forward with smoothing (MSE 59)



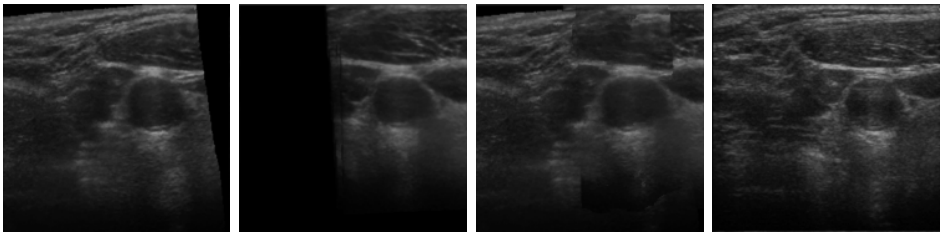
(d) online MPR, backward w. median (e) backward nearest sample (MSE 55) (f) backward gauss (MSE 57)

Figure 3.5.: Longitudinal MPR (a) from an axial freehand sweep of a right neck created using backward compounding with weighted median. Anatomical details, from superficial to deep structures: skin, subcutaneous tissue, platysma,sternocleidomastoid muscle, lymph nodes (left: normal, center+right: malignant), internal jugular vein with physiological pulsation, marginal section of carotid artery with arteriosclerotic plaque. Enlarged details of images interpolated from compounded volume (b,c,e,f) and online MPR (d).

as described in [32]. Each method has its own merits for specific applications. The more homogenous appearance of the gaussian and smoothing kernels are favourable for volume rendering and segmentation. On the other hand, methods that retain original intensities, median and nearest sample produce more diagnostically relevant images than methods recomputing the values. Nearest Neighbor is the fastest but also the most unforgiving on noisy datasources and jittery tracking information. The online MPR reconstruction (fig. 3.5(a)(d)) provides the sharpest and detailed images, as one data resampling step is skipped.

Interleaved 3D-Ultrasound Data

Using the online backward-warping MPR reconstruction it is possible to fuse multiple freehand ultrasound sweeps regardless of their relative spatial positions. The samples presented here were created using three sequences. A selected slice of the center sequence was recreated using data from sequences to the right and left, see figure 3.6. Reconstruction using backward compounding with weighted median accumulation of one 256×256 slice from sweeps with a total of over 1000 slices takes about 1 second on an AMD64 3200+ processor using high quality settings.



(a) left sequence (SMD 667) (b) right sequence (SMD 1007) (c) combination (SMD 476) (d) reference

Figure 3.6.: The MPR Viewport was frozen in place at the spatial location of the original slice (d). The individual contributions of the right and left sequence are (a) and (b). The SMD (Squared Mean Differences) of the reconstruction drops significantly when both sequences are combined in (c).

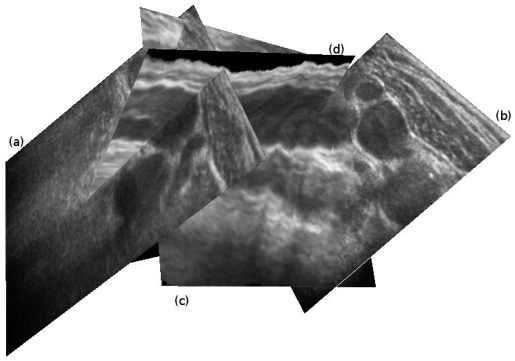


Figure 3.7.: Examples for spatially related ultrasound images from the neck: Original axial slices (a+b) and longitudinal MPRs (c+d) using our reconstruction technique.

Improving Registration

For multi-modal registration of freehand ultrasound images with a CT scan, we present automatic image-based registration techniques later in section 4.2. There, a set of axial images from a continuous caudo-cranial sweep along the neck is selected and used for registration, see section 5.1 for details on the data acquisition. Using our compounding methods, arbitrary planes of ultrasound information can be considered in addition. Figure 3.7 shows two original transversal images from a freehand ultrasound sequence along the jugular vessels, as well as two additional MPR-reconstructed slice images.
One could argue that the examiner should rotate the ultrasound probe after the continuous caudo-cranial motion to acquire longitudinal images supporting the registration. This however will introduce significant deformation errors due to tissue compression distributed differently on the patient's skin, as well as motion induced by the blood pulsating through the vessels (see section 5.1.8 for a detailed discussion of the error sources). For a rigid registration, it is preferred to use additional planes derived directly from the data of the original continuous probe motion. This is done using our MPR reconstruction method, which allows

in addition to create planes of information parallel to the skin surface within the body, which cannot be acquired by ultrasound itself.
The robustness of automatic rigid CT-Ultrasound registration on such a sequence improved significantly when oblique reconstruction planes were considered for registration. The standard deviation of the Target Registration Error on a lymph node decreased from $3.2mm$ to $1.5mm$ for a random displacement study for the sequence depicted in figure 3.7, when two MPRs where considered in addition to 5 slices from the original sweep (only two of which are shown in the figure for better spatial impression).

3.3.3. Discussion

We have developed new methods for spatial compounding of freehand ultrasound data, using a backward-warping approach which collects the scattered image information for each voxel in an efficient manner. They allow to perform reconstructions with superior quality and shorter computation time compared to forward-projection techniques known from literature, while a choice of smoothness and continuity versus retaining original image characteristics can be made using different accumulation functions. These algorithms can also be used to compute Multi-Planar Reconstructions (MPRs) in real-time from the original data. This further increases the image quality, as an extra interpolation step is saved that would be necessary when rendering MPRs from previously compounded 3D-volumes. As a result, more detailed diagnostic information can be gathered and visualized not only for the person performing the ultrasound examination, but also for demonstration to colleagues at a later point of time without the presence of the patient necessary. Pathological changes in tissue texture and their relation to anatomical landmarks can be demonstrated in an optimized plane without haste, which gives the possibility to combine the advantages of ultrasonography (high spatial resolution and tissue contrast depending on the frequency of the ultrasound device) with the advantages of sectional imaging in any plane. Furthermore, multi-modal image-registration can be improved by adding oblique ultrasound information in addition to original image data. Finally, the online MPR algorithm can yield high-quality real-time visualization of oblique slices for freehand ultrasound systems.

3.4. Calibration

For 3D freehand ultrasound systems it is crucial that the spatial relation between the position sensor and the image plane is precisely determined. This is a rather complex problem, which tends to be underrated. A large body of literature deals with spatial ultrasound calibration, a concise survey is [91]. Most of the methods are based on imaging a designated phantom object with some known geometric properties. Acquiring a number of images from different orientations then allows to establish the relation of the image content to the tracking sensor's coordinate system. Optimal phantoms contain material mimicking human tissue, and are precisely manufactured according to a geometric model used for the mathematical computation of the calibration parameters. Such phantoms are mostly used for commercial freehand ultrasound systems, and potentially allow for a very convenient calibration workflow, that can be performed by sonographers & physicians. However, purchasing or manufacturing them is often not an option for research facilities. On the contrary, the Single-Wall calibration method [113] only relies on a rough-textured plane submerged in water. A number of subsequent research effort has been conducted to further automate this method [124]. However,

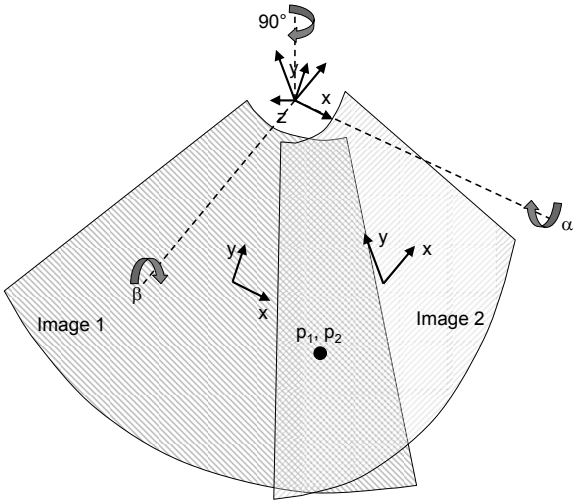


Figure 3.8.: Used coordinate systems and image planes for computing reconstruction errors.

its precision is still limited by the narrow range of imaging orientations and depths. The varying elevational beam thickness causes the Single Wall phantom to appear as line with different thickness and fuzziness, depending on the depth and focus settings used. Another problem arises as the speed of sound in water is different than the reference speed assumed by ultrasound machines (1540m/s in average human soft tissue). An additional source of errors is introduced when trying to overcome this problem, either by using a particular fluid with the desired speed (mixture of water with NaCl), or estimating the actual speed of sound and henceforth compensating for the image distortion. Phantom-less calibration methods [72, 94] use the intersections of a tracked tool with the ultrasound image plane to derive the calibration. They suffer from the speed-of-sound problem as well, as the calibration has to be performed underwater. Besides, two more error sources are added: The tracking accuracy of the tool itself, as well as the calibration of the tool. For magnetic tracking, especially the rotational accuracy is rather poor, making it difficult to achieve a good tool calibration. For optical tracking, the target markers have to be considerably far away from the tool tip, as they always need to be above the water surface.
In [15], an image-based approach is used, which maximizes Mutual Information of tracked ultrasound images and reconstructions from an MRI scan of a custom-built Agar gel phantom.

Building upon the efficient compounding method presented in section 3.3, we developed a new calibration approach, that can avoid a number of the problems pointed out above. It uses the spatial consistency of interleaved freehand ultrasound acquisitions as the underlying information for calibration. The proposed method can be performed using in vivo tissue, making it possible to calibrate ultrasound prior to the procedure, or even validate the calibration during an exam. While in [19] tracking of in-plane motion on successive frames is used to monitor some calibration parameters, we are using large-scale consistency of the anatomy to recover all parameters, including the temporal lag.

3.4.1. New Sweep-based Calibration Method

If two slow angulated freehand ultrasound sweeps are acquired from the same anatomy at approximately perpendicular orientation, they should measure the same image intensities at every point in 3D-space, assuming the right calibration parameters T_c are known. Hence a maximization of the similarity measure S of images I_j from the first sweep with reconstructions $\widetilde{I_j}$ from the second sweep (and vice versa) should yield the correct optimization parameters:

$$\widetilde{T_c} = \arg\max_{T_c} \sum_j S\left(I_j(\vec{p}), \widetilde{I_j}(\vec{p})\right) \tag{3.8}$$

Geometric Formulation

In homogenous coordinates, we denote the sought calibration T_c as rigid transformation from image to sensor coordinates, and $T_{ij} = T_j^{-1} T_i$ the relative transformation between two sensor measurements. Then the transformation between two images is described as

$$\begin{aligned} T_c^{-1} T_{ij} T_c &= \begin{bmatrix} R_c^{-1} & -R_c^{-1} t_c \\ 0\,0\,0 & 1 \end{bmatrix} \begin{bmatrix} R_{ij} & t_{ij} \\ 0\,0\,0 & 1 \end{bmatrix} \begin{bmatrix} R_c & t_c \\ 0\,0\,0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} R_c^{-1} & -R_c^{-1} t_c \\ 0\,0\,0 & 1 \end{bmatrix} \begin{bmatrix} R_{ij} R_c & R_{ij} t_c + t_{ij} \\ 0\,0\,0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} R_c^{-1} R_{ij} R_c & R_c^{-1}(R_{ij} t_c + t_{ij} - t_c) \\ 0\,0\,0 & 1 \end{bmatrix} \end{aligned} \tag{3.9}$$

If the rotation R_{ij} is close to identity, the influence of both R_c and t_c in equation 3.9 diminishes, which means that frames used for calibration should be angulated as far as possible from each other. For further examination, we use a more specific setup with $T_{12} = T_2^{-1} T_1$ and two locations $\vec{p_1} = (u_1, v_1, 0, 1)^T$ and $\vec{p_2} = (u_2, v_2, 0, 1)^T$ in image coordinate systems that represent the same point $\vec{p_w}$ in 3D-space:

$$T_c^{-1} T_2^{-1} T_1 T_c \vec{p_1} = \vec{p_2} \tag{3.10}$$

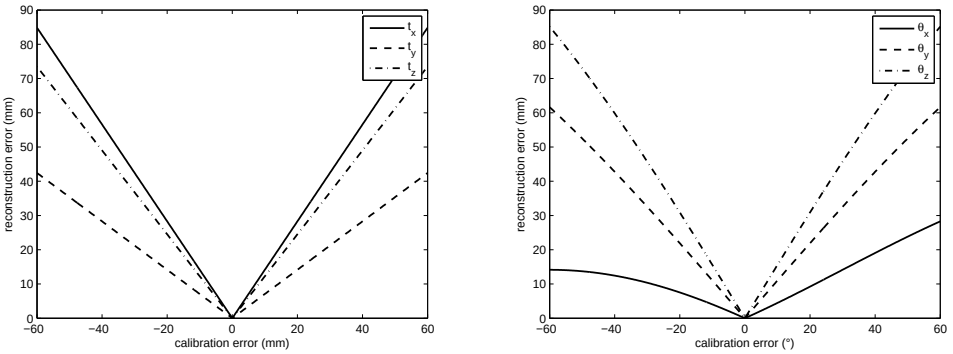
The image locations $\vec{p_1}$ and $\vec{p_2}$ are in physical units (mm), i.e. we assume the correct pixel spacing to be known. We are now interested in the reconstruction error e of the 3D positions, if an error-prone calibration matrix $\widetilde{T_c}$ is used.

$$e = |T_2 \widetilde{T_c} \vec{p_2} - T_1 \widetilde{T_c} \vec{p_1}| \tag{3.11}$$

$$\text{Using eq. 3.10: } e = |T_2 \widetilde{T_c} T_c^{-1} T_2^{-1} T_1 T_c \vec{p_1} - T_1 \widetilde{T_c} \vec{p_1}| \tag{3.12}$$

In the setup used to analytically derive the error, T_c consists solely of a translation $d = 100mm$ along the negative y-axis, T_1 a rotation of α around the x-axis, T_2 a rotation of β around the x-axis first and 90° around the y-axis (see figure 3.8). The 3D point $\vec{p_w}$ is

$$\vec{p_w} = T_1 T_c \vec{p_1} = \begin{pmatrix} u_1 \\ (v-d)\cos\alpha \\ (v-d)\sin\alpha \\ 1 \end{pmatrix} \tag{3.13}$$



(a) translation error

(b) rotation error

Figure 3.9.: Plots of reconstruction error. The configuration is $\vec{p_1}=(40,20,0,1)^T$ and $\alpha=\beta=-30°$.

The angle β is defined such that the second imaging plane intersects $\vec{p_w}$ as well:

$$\beta = -\arctan\frac{-u_1}{(v-d)\cos\alpha} \tag{3.14}$$

An error-prone calibration $\widetilde{T_c}$ is defined as a translation about $(t_x, t_y - d, t_z)^T$. Inserting in equation 3.12 yields

$$e = \left\| \begin{pmatrix} \sin(\beta)t_y + \cos(\beta)t_z - t_x \\ \cos(\beta)t_y - \sin(\beta)t_z - \cos(\alpha)t_y + \sin(\alpha)t_z \\ -t_x - \sin(\alpha)t_y - \cos(\alpha)t_z \\ 0 \end{pmatrix} \right\| \tag{3.15}$$

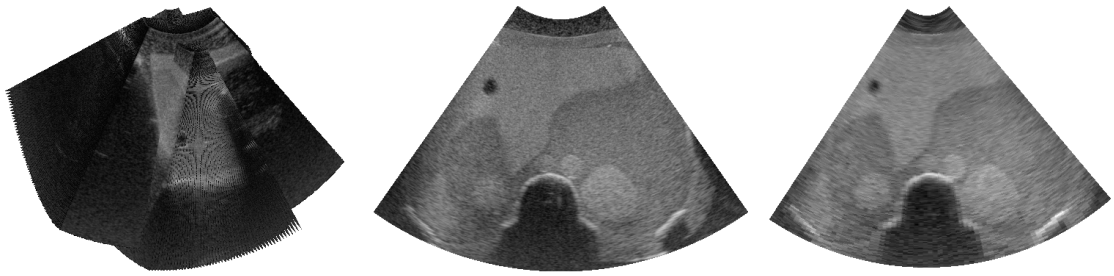
For applying error in one translation component at a time, the error results to:

$$e_x = |t_x|\sqrt{2} \tag{3.16}$$

$$e_y = |t_y|\sqrt{2(1-\cos\alpha\cos\beta)} \tag{3.17}$$

$$e_z = |t_z|\sqrt{2(1-\sin\alpha\sin\beta)} \tag{3.18}$$

The reconstruction error linearly increases with respect to the calibration error, and is independent of the position in the image plane (apart from the dependency of the angle β in equation 3.14). Only if the angles α and β are small, the error in the y direction remains unchanged (eq. 3.17). This is obvious as in that case the change of t_y moves the two image planes along parallel lines in 3D space. Hence the calibration sweeps have to be sufficiently angulated, two linear motions would not recover the y translation of the calibration transformation. Similarly, eq. 3.18 shows that for angles approaching 90° a change in t_z would not affect the reconstruction error. However, such a configuration is not possible with transcutaneous ultrasound anyway (the image planes would be parallel to the patient's skin).
Expanding the equations with rotations in $\widetilde{T_c}$ yields very large trigonometric equations. Essentially, the only situations where a wrong rotation would not increase the reconstruction



(a) 3D view (b) ultrasound image (c) reconstruction

Figure 3.10.: Phantom images, acquired with the Antares wobbler. The NCC measure for those images is 0.86.

error, are: a) that $\vec{p_1}$ and $\vec{p_2}$ both lie on the x- or y-axis of the respective image coordinate system (which would only be the case for a minority of the pixels in an ultrasound image); b) both α and β are very small, then the θ_y rotation cannot be recovered. Figure 3.9 displays the reconstruction error for all 6 calibration parameters.
We have shown that the described setup allows to recover all 6 calibration parameters. If an appropriate image similarity measure between original ultrasound intensities $I_{im}(u, v)$ and a reconstruction at its 3D location $I_{rec}(\vec{p_w})$ from the respective other sweep can be described as a monotonous function of the reciprocal reconstruction error e, its maximization will result in the correct calibration.

Algorithm

We use the compounding algorithm described in section 3.3 to efficiently reconstruct arbitrary planes directly from a freehand ultrasound sweep. 5-10 images from every sweep are compared against reconstructions from the other sweep, respectively. Normalized Cross-Correlation (NCC) is computed on every pair, its average is used as a cost function for non-linear optimization (i.e. $S = NCC$ in equation 3.8). This assures that brightness and contrast differences of an ultrasound image and its reconstruction do not impose the registration accuracy. Such differences can occur, as the respective intensities originate from different scan orientations. The Amoeba Simplex algorithm (see section 2.5.1) then finds the optimal calibration transformation maximizing the image similarity. Instead of directly modifying the calibration parameters, a relative calibration matrix composed from zero-initialized translation and Euler angles is right-multiplied onto the initial estimate. This avoids the inherent Gimbal lock problem of the Euler-angles parameterization.

3.4.2. Results

Ground Truth Study

To derive the absolute precision of our new calibration method, we used a Siemens Antares ultrasound machine with the C5F1 3D transducer. It acquires 30-70 2D frames per volume, mechanically wobbling over an angulation of $35 - 75°$. Two perpendicular volumes have been acquired from a multi-modal abdominal phantom (CIRS Inc, Norfolk VA USA, figure 3.10(a)). Using the imaging geometry from the saved pre-scan-converted DICOM volumes,

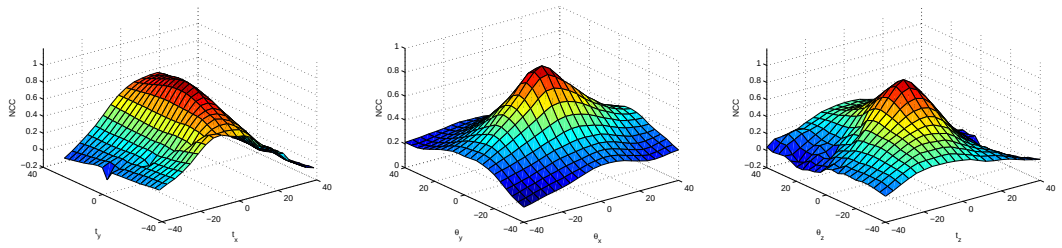


Figure 3.11.: Similarity measure plots for the Antares phantom data, altering two calibration parameters at once.

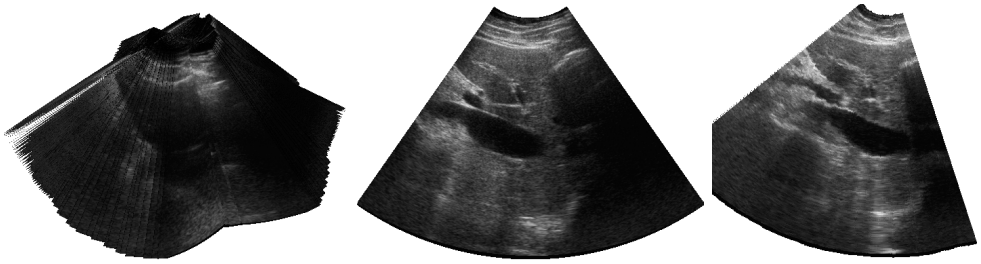
	t_x(mm)	t_y(mm)	t_z(mm)	θ_x(°)	θ_y(°)	θ_z(°)	NCC	time(s)	error(mm)
Antares Ground Truth data									
mean	-0.15	2.01	-0.63	-0.06	0.00	0.01	0.91	168.6	2.39
σ	0.20	0.75	0.30	0.38	0.17	0.15	0.00	45.9	0.74
Sequoia freehand ultrasound, phantom vs. in-vivo									
phantom	115.5	-3.5	27.3	-172.7	7.3	-90.3	0.85	115.9	
liver	114.4	-2.0	25.8	-172.8	6.8	-86.6	0.68	89.2	

Table 3.1.: Results of precision and robustness study.

we perform a 2D scan-conversion into a set of cartesian 2D images, and tag the images with calibration and tracking matrices to resemble a freehand ultrasound acquisition. The tracking effectively consists of an x-rotation angle α, the calibration of a negative y offset d, similarly to the setup in figure 3.8. A standard image-based registration technique is used to register 3D-scan-converted representations of the volumes, using NCC as similarity metric. The registration result is then applied to the tracking matrices of the second sweep. Table 3.1 (top) depicts the mean and standard deviation values for execution of 183 optimizations, each randomly displaced from the ground truth calibration (every parameter in $\pm 20mm$, $\pm 20°$ mean distribution). The method reliably finds the same calibration. Only in the y axis is a systematic bias of $2mm$. This might be due to interpolation in the 2D- and 3D-scan-conversion algorithms, or a difference between the 3D imaging geometry provided by the manufacturer, and the physical transducer properties. The error value in the right column depicts the relocation error of a point in the center of the ultrasound image (around $7cm$ depth). It is mostly composed from that deviation in t_y. Figure 3.11 depicts 2D plots of the sum-of-NCCs similarity function used, for changing two parameters from the ground truth calibration each.

Calibration on Human Liver

Here we used a freehand ultrasound system based on a Siemens Sequoia ultrasound machine in conjunction with an Ascension 3D Guidance magnetic tracking system and a PC with a video grabber. Three transducers (4C1, 4V1 and 6C2) were calibrated for a permanent clinical setup, using the same CIRS phantom. For every transducer, two perpendicular sweeps were recorded at 12cm, 18cm and 26cm depth setting, and two focal zones (the minimum on that



(a) 3D view (b) ultrasound image (c) reconstruction

Figure 3.12.: Freehand ultrasound sweep of the human liver. NCC for the shown images is 0.60.

machine) evenly distributed over the image.
We repeated a particular calibration (4C1 transducer at 18cm depth) in-vivo on a human liver. The two required sweeps were combined to achieve a consistent data set within one breath-hold, and without removing the probe from the skin (see figure 3.12(a), angulated, twisted 90°, angulated back). The result deviates in the calibration parameters from the phantom-based results less than $2mm$ and 4°, see table 3.1 (bottom). If the patient or volunteer steadily maintains the breath-hold and does not shift his position during the acquisition (10-20 seconds), the only additional error source with respect to phantom data arises from lower structures pulsating with the heart-beat (e.g. the portal vein).

Temporal Calibration

We achieved temporal calibration by performing a linear forward-backward motion on the phantom. Each 4 images in the forward motion were reconstructed from the backward motion, and vice versa. A one-dimensional Brent-line-optimization [115] was run on the temporal delay parameter. It alters the spatial information of every recorded frame by interpolating between the neighboring tracking matrices, using a quaternion-based method [137]. We consistently obtained a lag between the video-frames and tracking information of $100ms$.
The linear motion used here makes the temporal calibration mostly independent of the spatial calibration parameters. For bootstrapping the calibration, it is sufficient to optimize the spatial calibration parameters with zero temporal lag, successively run the temporal calibration, and refine the spatial calibration with the temporal result. In our experience, the temporal delay is very similar for all different transducers and depth settings, and in fact only needs to be established once for a certain hardware setup.

3.4.3. Discussion

We presented novel methods for spatial and temporal ultrasound calibration, that can be performed on any ultrasound phantom or in-vivo. We have expanded on the theoretical foundation of the method, and proved that the chosen acquisition geometry allows one to recover all spatial calibration parameters. Our result on a ground-truth 3D ultrasound machine, as well as freehand ultrasound calibration on both phantom and human liver data show that the method is reliable and precise. We believe it is a very convenient alternative to other methods published in the literature, that might especially be beneficial if the availability of specialized

tools, phantoms and 3D imaging of those phantoms, is limited. As an advanced application, freehand ultrasound systems continuously recording during an exam or procedure could detect perpendicular image frames from analyzing the tracking trajectories, and automatically verify and correct some or all calibration parameters in the background. Besides, our algorithm might be used in conjunction with methods based on motion of successive ultrasound frames[19], to increase overall stability and accuracy by combining small- and large-scale anatomic clues.

4. CT-Ultrasound Registration

4.1. Related Work

4.1.1. Modeling and Simulation of Ultrasound

A first step toward registration methods involving ultrasound images, is to model and simulate the physical effects that take place in ultrasound imaging. The book [85], chapters 9 and 10, explain the physics of ultrasound imaging in terms of linear system theory. A detailed description of ultrasound physics, as well as transducer instrumentation is contained in the textbook [56].
In [35] the authors model the ultrasound wave propagation for a multi-layered medium, with constant properties within each layer. Experimental validation on a phantom with a single-element ultrasound transducer is also conducted.
A detailed mathematical description of ultrasound imaging systems is [67]. The concept of spatial impulse response is used in order to approximate the whole ultrasound imaging process in a numerical manner. It leads to the full simulation of 2D images based on a ultrasonic scattering map as input. This is implemented in the software FIELD II [66], a MATLAB based simulation program. It has become a standard tool in the community to study, teach and research ultrasound systems, however the used approach is hardly applicable for multi-modal registration. On one hand, the extremely realistic simulation of aperture, scattering and speckle noise yields no benefit for comparison with other modalities. On the other hand, the computation times are quite large, ranging from several minutes to many hours for a single image. The nature of ultrasound speckle is examined in [34], leading to a model which allows to use less point scatterers, while producing the same simulated speckle data.
Another, more general (i.e. not restricted to medical ultrasound) simulation software is *Wave2500*, capable of simulating the wave propagation for two-dimensional models, with arbitrarily placed senders and receivers [69]. It solves the acoustic wave equations using a finite differences method. In the medical domain it is used e.g. for assessment of ultrasound interaction with bone [10].
Most of the real-time medical ultrasound simulation systems available today, use previously established 3D ultrasound data sets. An overview of systems in use and their clinical benefit in particular in prenatal diagnosis is given in [88]. A specific implementation is presented in [4]. This work contains many implementation details about efficient voxel traversal for slice interpolation, which probably can be solved differently today, e.g. by GPU processing. A similar system for education and training is described in [37].

4.1.2. Image-based Mono-modal Registration

Before addressing multi-modal registration, one should have a glance at how mono-modal registration of ultrasound images is adressed in the community. This can provide clues about dealing with the common ultrasonic properties and artifacts within an image-based registra-

tion method.

Intra-modal registration of multiple compounded ultrasound volumes is done in [123]. In particular, the correlation of the 3D gradient magnitude, together with a hierarchical resolution approach leads to a robust registration technique, which is used to compound several free-hand 3D ultrasound sweeps.
The popular Mutual Information similarity measure is also used in this context, for rigid and non-rigid registration of ultrasound volumes [133]. An application of both temporal and spatial registration using MI is 3D stress echocardiography [134]. [175] have applied a non-rigid registration of free-hand US volumes of the breast, using the correlation coefficient and local statistics for the measure. [95] perform 2D rigid registration of breast ultrasound using an adapted alpha-entropy similarity criteria. Registration using a deformation model which consists of a global affine transformation and local free-form deformations, applied to 3D ultrasound data acquired during neurosurgery (before and after opening the dura, respectively), is presented in [82]. Here an existing software package for non-rigid registration of MR data [126] is used.
In [30], a new similarity measure for motion estimation from ultrasound images is proposed. It is derived by modeling the ultrasound intensities with multiplicative Raleigh noise (instead of an additive Gaussian noise term, as used for deriving the common similarity measures in section 2.2.3). This measure is also used in a block-matching based velocity estimation in [22], and a learning-based algorithm in [181]. In the latter, weak learners based on local rectangle features are used to boost a discriminative similarity function, for motion estimation of stress echocardiography.
In [15, 16] an ultrasound calibration method is derived, which uses intensity-based registration of acquired ultrasound slices of a gelatin phantom with a MR scan of it. Normalized Mutual Information and a Best Neighbor optimizer are used in order to find the most appropriate calibration transformation.
Texture information from ultrasound images can also serve as significant features for registration. In [43], spatial Gabor filters are applied to create localized feature vectors, describing the texture components in various directions and frequencies. These features are in turn used for deriving a statistical similarity measure in order to register two 3D ultrasound volumes.
In [5] the authors determine a global rigid transformation for two ultrasonic volumes, which is later refined using an elastic registration on 2D seam planes.
In own studies we registered multiple 3D-ultrasound volumes of the author's liver, acquired with a Siemens Antares C5F1 probe at different breathing states, using a non-linear variational approach [183]. A scenario which has been adressed rarely in literature so far, is the simultaneous registration of more than two 3D ultrasound datasets. We conducted a study of registering multiple 3D-US volumes by extending popular similarity measures to the multi-variate case [156].

4.1.3. Feature-based Multi-modal Registration

Due to the difficult properties of ultrasound, a lot of research has been carried out using features extracted from the ultrasound images, in order to align them with corresponding structures in other modalities. In the context of thermal ablation for liver cancer, the most significant structures in the ultrasound images are the liver surface and major vessels inside the liver. In [104], those structures are segmented in an MRI scan, and points on the surface

of the same structures are manually picked in the ultrasound images. A method based on the Iterative Closest Point (ICP) algorithm is then used to compute the transformation between the two modalities. The organ surface of the prostate is used in [174].
Bone structures can be well identified using ultrasonic imaging, as they produce a strong (though also specular and hence position-dependant) reflection and full occlusion behind the reflection. Thus it is feasible to use such structures in a more automated manner for registration [8, 23]. In [8], a modified ICP algorithm is used in order to register the bone surface extracted from CT with points in the ultrasound images likely to reflect the bone surface. Therefore they minimize the point distances weighted with a probability of each point being on the bone surface. This probability is computed from three components: The so-called spatial prior, which is the proximity of the point to the surface from the CT, the US intensity, and the result of a directional edge filter applied to the ultrasound images. Thus this method pursuits a combined registration and segmentation approach. In [11], a similar point-based registration based on point features is executed simultaneously while refining the freehand ultrasound calibration parameters, which improves the accuracy by compensating speed-of-sound induced scaling changes in soft tissue.
Manual segmentation of MRI scans of the brain is used in [9] to create “Pseudo US” images, which are non-linearly registered to intra-operative 3D ultrasound (compounded from a freehand acquisition), using a correlation-based similarity metric.
The use of color Doppler ultrasound aids the automatic feature extraction of vessel structures, which can in turn be used for feature-based registration [77, 110, 140]. In [77], liver vessel features are extracted from both preoperative MRI/CT data and intraoperative 3D Power-Doppler ultrasound data. The registration is initiated with a few manually selected landmarks, and a rigid transformation is then estimated using a modified ICP method, which takes the vessel topology into account. Eventually, the registration is further improved using a transformation grid modeled with deformable B-Splines. Rigid registration of Power-Doppler 3D ultrasound with MRI scans of the carotid artery was evaluated in [140].
A special yet interesting case is when no pre-operative data is to be used. An alignment with respect to a *Statistical Shape Model (SSM)* can be desirable then. In [26], bone surface points originating from intraoperative 3D Ultrasound are registered with SSMs of the pelvis and femur. SSMs serve a different purpose in [74], where they are created as liver surface motion models from pre-operative MR sequences. They are then registered to intra-operative tracked ultrasound acquisitions in a Bayesian framework, to compensate for breathing motion.

4.1.4. Intensity-based Multi-modal Registration

Pure intensity-based registration with other modalities has been performed mainly for 3D ultrasonic data. Roche et al. [118] use an adapted correlation ratio similarity measure in order to register the ultrasonic data simultaneously to both the intensity and the gradient information of an MRI scan. This method is used in [100] together with successive non-rigid tracking of intra-operative deformations for neurosurgery. In [146], the Kullback-Leibler divergence (KLD) is minimized for the registration of 3D-US and MRI. A registration involving a trained mapping of MRI and ultrasound data to “vessel probability values" and successive registration of this information driven by the NCC similarity measure is proposed in [102]. The liver registration method is also extended in [14] in order to incorporate a deformable model generated from MRI acquisitions of the liver at different stages in the breathing cycle. Single tracked ultrasound slices can hence be registered to the pre-operative information in a

non-rigid manner. In [90] the image phase is proposed as underlying data for a Mutual Information based registration. It includes experiments on 2D data with brain MRI and simulated ultrasound images, where early results on deformable registration are presented.
To our knowledge the published data on true image-based registration of CT with ultrasound is sparse. A Mutual Information based registration of CT with tracked 3D ultrasound for prostate localization is presented in [25]. It is aided by the manual segmentation of prostate and bladder and the application of pre-processing steps on both modalities. In [81], in the CT data of a kidney, the intensity values are enhanced with strong edges from the gradient. Successively an automatic registration with freehand 3D ultrasound is performed. In [101], an approach related to their previous work [102] uses bone probability values generated from the CT intensity and gradient in order to perform a rigid registration of CT with tracked ultrasound, for orthopedic applications.

In this thesis, we focus on developing advanced intensity-based techniques. In section 4.2, we present methods based on a combination of different information and physical properties of both modalities to introduce a more stable measure for automatic registration. They are targeted on small linear ultrasound data of the head and neck, and have also been published in [169, 170]. In section 4.3, those methods are expanded upon by introducing a simulation approach that reproduces the major ultrasonic imaging effects, for making the modalities more comparable. A special new similarity measure based on the correlation of multiple simulated effects with ultrasound is then presented. Here, larger curvilinear abdominal ultrasound images are used. These methods have been introduced to the community in [166].

4.2. New Methods 1: Semi-automatic Image-based Registration

4.2.1. General Considerations

A tomographic data set will usually enclose the respective ultrasound data, be it two- or three-dimensional. Hence, data from the CT volume, according to the spatial extent of the ultrasound data (a slice or cone with specific thickness etc.), can be extracted at a location where the anatomy contained in the ultrasound images is presumed. This should establish a representation that can be compared with the respective ultrasound data. In the ideal case, it would be a realistic reconstruction of ultrasound image intensities, i.e. a *simulation* of ultrasound from CT. Using the terminology we had introduced in chapter 2, the CT volume is therefore the moving image, ultrasound the fixed one. This might be somewhat counter-intuitive, since the fact, that we are looking for the location of ultrasound images within the CT volume, suggests opposite roles.
Any computation step that brings the modalities closer together in terms of comparability, is useful. If an intermediate representation is established for comparison, it might be a simulation of ultrasound from CT, but just as well a simulation of CT from ultrasound, as well as anything in between. An ultrasound image depicts the strength of echoes whose magnitude increases at boundaries between different types of tissue. This can be related to CT data as the gradient magnitude of the X-Ray attenuation (from a simplified point of view, see e.g. [119] section 2.2). Hence the derivative of the CT voxel intensity yields information that can be compared to ultrasound. On the other hand, numerical integration of the edge information depicted in ultrasound unfortunately does not result in values that can be put in relation to the CT data, as we are missing directional information.

As a consequence, the resemblance of ultrasound images to CT can only be increased by reducing imaging artifacts, however a simulation of CT intensities from ultrasound is not possible solely based on the image data (but might be feasible with learning-based approaches and methods for incorporating prior information).

4.2.2. Information Extraction from CT

Instead of a realistic, possibly very slow, simulation of ultrasound, we need an intelligent and efficient intermediate representation of the CT data at arbitrary cut-planes, such that an iterative registration can be performed in an acceptable time. These slices have multiple components containing intensity, gradient and edge information, which are used to derive various parts of a similarity metric, so that the correspondence of anatomy contained therein with structures in 2D B-mode ultrasound images can be determined.
The use of gradient and edge information is justified, as medical ultrasound mainly depicts tissue interfaces, caused by ultrasonic reflections at a boundary of different acoustic impedances. Those will also produce different intensities in CT and hence a visible edge. The strength of such a boundary gradient can however not be put into a mathematical relation to the strength of the respective ultrasound reflection. On the other hand, the original Computed Tomography attenuation values are of importance as well, as they reveal different types of human soft tissue, which in turn cause different effects in ultrasound, such as attenuation, speckle characteristics, reflection etc.
In our approach, first the three-dimensional gradient vector values are computed from the CT data set by convolution with a Sobel filter cube. They are stored in a 4-channel volume together with the original voxel intensity (i.e. channels 1-3 contain the gradient vector, channel 4 the CT Hounsfield intensity). The interpolated slices contain four channels as well. For each pixel, the 4-vector is computed from the volume using trilinear interpolation. In the first channel of the slice, the original CT intensity is stored. The 3D gradient vector is scalar multiplied with each of the vectors indicating the horizontal and vertical slice plane directions, respectively. The resulting values, corresponding to the 2D gradient of the CT intensity within the slice, are stored in the second and third channel.
The 2D slice gradient values are then used to perform Canny edge-detection on the slice data, and the result is stored in the fourth channel. The most time-consuming steps within the Canny algorithm are the computation of the 2D gradients, as well as filtering them with a sufficiently large Gaussian kernel for smoothing. As we compute the 2D gradients directly from the precomputed 3D gradient values, we do not need to run a 2D filtering for gradient computation. In addition, those gradients are fairly smooth, as they originate from a three-dimensional Sobel filter using a 27-neighborhood. This makes further Gaussian filtering unnecessary. The two remaining steps for the Canny algorithm, non-maxima suppression and hysteresis thresholding, can each be performed in one traversal of the 2D slice. The horizontal gradient is weighted with a user-defined factor between 0 and 1, as the ultrasound data tends to show mainly edges along the lateral direction, parallel to the transducer array.
Thus we are able to construct intermediate slices from the CT data at estimated transformations of the US scan plane in very short time ($1.1ms$ for a 128^2 pixel slice, interpolated from a $512^2 \cdot 100$ CT/gradient volume, on an AMD Opteron 2.2GHz machine). The individual components of the slice pixels are then used to compute a similarity metric with the ultrasound data.

4.2.3. Occlusion Handling

If an ultrasonic pulse hits bony structures, all image intensities in the ultrasound image further along the specific ray are occluded, and mainly determined by noise. Therefore, all ultrasound intensity values on a ray below such an occlusion should be disregarded in the registration method. In our implementation, we scan the US image from bottom to top, updating the variances for all ultrasonic pulse rays. Where they exceed a threshold σ_y^2 (which is easily determined in the user interface), the first pixel to be considered is defined. Thus, our Region of Interest (ROI) Ω is expressed by the following equations:

$$\Omega = \{(x,y) \mid (y < y_{top}) \wedge (y \geq b(x))\} \tag{4.1}$$

$$b(x) = \min\left\{ y : \frac{1}{y}\sum_{i=0}^{y-1} U(x,i)^2 - \left(\frac{1}{y}\sum_{i=0}^{y-1} U(x,i)\right)^2 < \sigma_y^2 \right\} \tag{4.2}$$

where x is the lateral (increasing to the right) and y the axial (increasing upward) pixel index of an ultrasound image U. By applying a median filter on the bottom function $b(x)$, discontinuities are removed before defining the ROI. In addition, we discard all pixels which are located above $y_{top} = \frac{9}{10} size_y$, as we observed that the anatomy is highly compressed there due to the probe pressure on the patient's skin. This compressed region is very distinct from the remaining anatomical structures; its size ($3.6mm$) being consistent on all data we obtained from patients (Figure 5.4(a)). This ROI definition is similar to the ones used in [102] and [81]. It can be adapted to curved-array transducers as well, by accumulating Equation 4.2 along the actual ultrasound pulse rays (which do not coincide with the image columns as in our case; hence some interpolation would be required). However, using a sector probe will markedly enhance all remaining problems relating to tissue compression and distortion, as there is no simple function available for correcting the tissue shift.

4.2.4. Similarity Measure

Deriving a similarity measure for image-based registration of ultrasound with CT, based on tissue attenuation values and their edges reconstructed from CT, is a demanding issue. Based on both the physical properties of the imaging modalities, as well as the visible appearance of their images, we introduce several components for a similarity measure, which can in turn be weighted to define a global cost function with respect to the transformation parameters.

Skin Surface Clamping

In the topmost part of the ultrasound image, which contains subcutaneous compressed tissue, we run the interpolation from CT with 6 times the vertical scaling (see Figure 5.4(a) on top, the 6× factor has been manually optimized). As result, the border between skin and air in the CT volume always has to be within that region, producing a large vertical gradient in the interpolated slice. When all vertical gradient pixels in this uncompressed stripe are summed to t, high and low thresholds t_h, t_l can be defined in order to decide if the skin surface lies

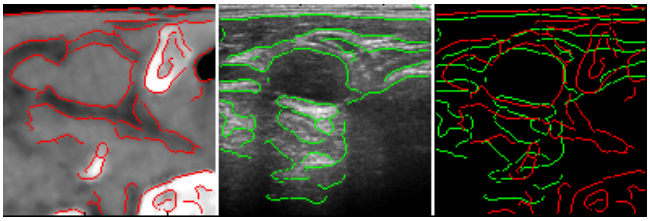


Figure 4.1.: Registered image pair and edge detection result (red=CT, green=US)

inside, outside or close to it:

$$\begin{aligned} f(t) &= \left\{ \begin{array}{ccc} 1 & if & t > t_h \\ 0 & if & t < t_l \\ (t-t_l)/(t_h-t_l) & otherwise & \end{array} \right\}; \\ S_0 &= 3f(t)^2 - 2f(t)^3 \end{aligned} \tag{4.3}$$

A cubic polynomial is used instead of the linear one in order to avoid discontinuities. As a component in a cost function, S_0 penalizes transformations that are physically impossible, as the patient's skin is always located at the top of the ultrasound images. t_h and t_l have been manually defined by displaying the value of t while gradually moving the transformation towards the patient, and visually assessing the skin surface alignment, on sweeps from all patients.
Note that another option for evaluating the distance to the patient's skin would be to perform lookups in the distance volume that was used for the global pre-registration in section 5.1.4. However, we prefer the described method as we do not have to load another volume (or another intensity channel) into the computer's main memory, which is already largely occupied by the CT/Gradient volume. Furthermore, the above computations can be performed with little additional computational cost.

Edge Alignment

As we have detected the edges in the simulated images, we would like to derive a similarity estimate based on the distance to edge structures in the ultrasound images. The straightforward approach would be to 1) compute an edge-detection for the ultrasound images, 2) compute a 2D distance map for those edges and 3) sum over the distance map values at the locations indicated by the edges of the simulated data. Steps 1) and 2) need to be performed once for each ultrasound slice, while 3) establishes a similarity metric and thus has to be computed for each simulated slice during pose estimation.
However, due to the different nature of CT and ultrasound data, the detected edges do not correspond in general, as shown by figure 4.1. We therefore propose to skip the ultrasound edge detection, using the original ultrasound intensity just as an indicator for edges instead. Given a binary edge image, the distance of an image point $\vec{x}$ to the edge structures $Y = \{\vec{y_i}\}$ is defined as $d(\vec{x}) = \min_i |\vec{x} - \vec{y_i}|$. Instead of the Euclidean distance, we can also express the proximity to edges by using a Gaussian expression, which allows us to adjust the sensitivity of the cost function value with respect to the distances, using σ^2:

$$p(\vec{x}) = \max_i \exp - \frac{(\vec{x} - \vec{y_i})^2}{\sigma^2} \tag{4.4}$$

Taking into account that we do not have precise edge information, a proximity value can be defined as

$$p(\vec{x}) = \sum_i u_i \exp -\frac{(\vec{x} - \vec{y_i})^2}{\sigma^2} \tag{4.5}$$

where $u_i \in [0 \dots 1]$ is the probability for the image pixel $\vec{y_i}$ being an edge. Assuming that the ultrasound image intensity directly scales with the edge probability (i.e. $u_i \propto U(\vec{y_y})$, a two-dimensional proximity function $p(\vec{x})$ can be computed according to eq. 4.5 by just convoluting the ultrasound image with a large Gaussian kernel. From this, we define a similarity measure component $S_1 = (\overline{p_e} - \overline{p})/\sigma_p$, where $\overline{p}$ is the mean of all values in the proximity image, $\overline{p_e}$ is the mean of just the pixels at locations where an edge is present in the simulated image, and σ_p is the standard deviation of the proximity image values.

Statistical Correspondence

In addition to tissue interfaces, different tissues by themselves cause different ultrasonic scattering characteristics, in particular characteristic speckle patterns and echogeneity. These are in turn reflected in the average ultrasound image intensities for a given tissue type. It is therefore applicable to assess the statistical dependence of the CT intensities, which classify the tissue according to the X-Ray attenuation property, with the intensity in the ultrasound image. We use Mutual Information (see section 2.2.3) on the CT and ultrasound intensities. The Normalized Mutual Information term uses the entropies of the combined and individual images, which are computed with the Shannon entropy from probability distributions of the image intensities:

$$\begin{aligned} NMI(U,C) &= 2 - 2H(U,C)/(H(U) + H(C)) \\ H(U) &= -\sum_j p_u(j) \log p_u(j) \\ H(C) &= -\sum_i p_c(i) \log p_c(i) \\ H(U,C) &= -\sum_i \sum_j p(i,j) \log p(i,j) \end{aligned}$$

Here U denotes an ultrasound image, and C the corresponding simulated image, i.e. the slice interpolation of CT attenuation values. The probability distributions can be estimated using histogram information from the images:

$$p_u(i) = \frac{1}{n_\Omega} |\{(x,y) \in \Omega | U(x,y) = i\}| \tag{4.6}$$

$$p_c(j) = \frac{1}{n_\Omega} |\{(x,y) \in \Omega | C(x,y) = j\}| \tag{4.7}$$

$$p(i,j) = \frac{1}{n_\Omega} |\{(x,y) \in \Omega | U(x,y) = i \wedge C(x,y) = j\}| \tag{4.8}$$

Here we assume that each intensity value is mapped into one histogram bin, and n_Ω is the number of pixels in the region of interest,

$$n_\Omega = |\Omega| = \sum_{x=0}^{n_x - 1} (y_{top} - b(x)) \tag{4.9}$$

An equivalent formulation for constructing the probability distribution from a histogram can be written using a binary count function c_u

$$p_u(i) = \frac{1}{n_\Omega} \sum_{(x,y)\in\Omega} c_u(x,y,i) \tag{4.10}$$

$$c_u(x,y,i) = \begin{cases} 1 & if \quad U(x,y) = i \\ 0 & otherwise \end{cases} \tag{4.11}$$

Due to the various physical effects in ultrasound imaging, both the chance that an image intensity reflects the anatomy (due to refraction, reverberation and other artifacts), and the Signal to Noise Ratio (SNR), decrease with the distance from the ultrasound transducer. Thus we would like to give more weight to image pixels that are closer to the probe, i.e. with higher y values. In our approach, we introduce an integer weighting for estimating the probability distribution:

$$p'_u(i) = \frac{1}{n'_\Omega} \sum_{(x,y)\in\Omega} (y + c_0) c_u(x,y,i) \tag{4.12}$$

$$n'_\Omega = \sum_{x=0}^{n_x-1} \sum_{y=b(x)}^{y_{top}-1} (y + c_0) \tag{4.13}$$

Every intensity value is inserted $y + c_0$ times into the histograms and the joint histogram. c_0 is a shifting constant that affects the amount of weighting. We set $c_0 = n_y$ (n_y being the number of image rows); hence the top of the image is weighted twice as much as the bottom. For $c_0 \to \infty$ the original Mutual Information notation is obtained. Our weighted Mutual Information component NMI' of the similarity measure is assembled by inserting all used ultrasound slice images and the corresponding simulations into one histogram, as it increases the statistical significance of the derived entropy terms. See section 4.2.6 for details on the weighting of Mutual Information.

Cost Function

The final similarity measure from a set of n ultrasound slices $\{U_i\}$ and their CT simulations $\{C_i\}$ is

$$\begin{aligned} cf = w_0 \frac{1}{n} \sum_{i=1}^{n} S_0(C_i) + w_1 \frac{1}{n} \sum_{i=1}^{n} S_1(U_i, C_i) \\ + w_2 NMI'(\{U_i\}, \{C_i\}) \end{aligned} \tag{4.14}$$

where w_0, w_1, w_2 are fractional weights of the individual measure components adding to one ($w_0 + w_1 + w_2 = 1$), and S_i are the measure components as defined in the previous sections.

4.2.5. Registration

For automatic registration, a non-linear optimization method maximizes the cost function cf iteratively with respect to the parameters of a rigid transformation (6 DOF, translation and Euler angles), which is initialized with zero and affects the current location of all slices. We

used three optimization schemes: simple Hill-Climbing (aka Best-Neighbor search), Powell-Brent and an Exhaustive Hill Climbing (see also section 2.5). The latter one evaluates all combinations of [forward, keep, backward] for all parameters, using the best result of all $3^6 = 729$ evaluations as estimate for the next iteration. When the optimization terminates, the resulting relative transformation is multiplied onto the overall registration transformation.

4.2.6. Weighted Mutual Information

In this section we would like to study the effect of weighted image histograms in more detail, by using an artificial set of test images.

Basics

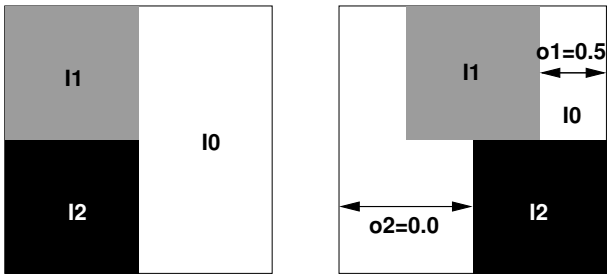


Figure 4.2.: Test images containing 3 intensities, the boxes I_1 and I_2 are displaced in the moving image, resulting in relative overlaps o_1 and o_2. An overlap of 1 means the boxes are at identical positions, 0 corresponds to no overlap.

We consider the images shown in figure 4.2. Half of the fixed (reference) image A has intensity I_0, and each one fourth is covered with a box of intensity I_1 and I_2, respectively. In the moving image B those boxes are shifted to the right, resulting in an overlap $o_1, o_2 \in [0 \dots 1]$ with respect to the boxes in the fixed image. The entropy in the images is independent of the overlaps:

$$p(I_0) = \frac{1}{2};\ p(I_1) = \frac{1}{4};\ p(I_2) = \frac{1}{4} \tag{4.15}$$

$$H_a = H_b = -\sum_i p(i) \log p(i) =$$

$$= -\left(\frac{1}{2}\log\frac{1}{2} + 2\frac{1}{4}\log\frac{1}{4}\right) = \frac{3}{2}\log 2 \tag{4.16}$$

The joint probability distribution is a table listing the probability of every pair of intensities in the left and right images, respectively, and depends on the overlap values:

$\mathbf{p_{ab}}$	$\mathbf{I_0}$	$\mathbf{I_1}$	$\mathbf{I_2}$
$\mathbf{I_0}$	$\frac{1}{4}(o_1 + o_2)$	$\frac{1}{4}(1 - o_1)$	$\frac{1}{4}(1 - o_2)$
$\mathbf{I_1}$	$\frac{1}{4}(1 - o_1)$	$\frac{1}{4}o_1$	0
$\mathbf{I_2}$	$\frac{1}{4}(1 - o_2)$	0	$\frac{1}{4}o_2$

The joint entropy computed from it results in:

$$H_{ab} = -\frac{1}{4}\left(o_1 \log\frac{o_1}{4} + 2(1-o_1)\log\frac{1-o_1}{4} + o_2\log\frac{o_2}{4} + 2(1-o_2)\log\frac{1-o_2}{4} + (o_1+o_2)\log\frac{o_1+o_2}{4}\right) \tag{4.17}$$

Histogram Weighting

If we are more confident that the overlap of the upper box reflects the image alignment than the lower one, we would like to weight the image intensities depending on the vertical position. We therefore multiply the histogram values with two integer count variables c_1 and c_2. They correspond to weights w_1, w_2 adding to one:

$$w_1 = \frac{c_1}{c_1+c_2};\ w_2 = \frac{c_2}{c_1+c_2} \tag{4.18}$$

The marginal image entropies are still independent of the overlap, however they depend on the weights:

$$p(I_0) = \frac{1}{2};\ p(I_1) = \frac{w_1}{2};\ p(I_2) = \frac{w_2}{2} \tag{4.19}$$

$$H_a = H_b = -\sum_i p(i)\log p(i) = \tag{4.20}$$

$$= -\frac{1}{2}\left(w_1\log\frac{w_1}{2} + w_2\log\frac{w_2}{2} + \log\frac{1}{2}\right) \tag{4.21}$$

The joint entropy is affected by both the overlap and the weight values now.

$\mathbf{p_{ab}}$	$\mathbf{I_0}$	$\mathbf{I_1}$	$\mathbf{I_2}$
$\mathbf{I_0}$	$\frac{1}{2}(o_1w_1 + o_2w_2)$	$\frac{1}{2}(1-o_1)w_1$	$\frac{1}{2}(1-o_2)w_2$
$\mathbf{I_1}$	$\frac{1}{2}(1-o_1)w_1$	$\frac{1}{2}o_1w_1$	0
$\mathbf{I_2}$	$\frac{1}{2}(1-o_2)w_2$	0	$\frac{1}{2}o_2w_2$

$$l(x) = -x\log x \tag{4.22}$$

$$H_{ab} = l\left(\frac{1}{2}o_1w_1\right) + 2l\left(\frac{1}{2}(1-o_1)w_1\right) + l\left(\frac{1}{2}o_2w_2\right) + 2l\left(\frac{1}{2}(1-o_2)w_2\right) + l\left(\frac{1}{2}(o_1w_1+o_2w_2)\right) \tag{4.23}$$

Figure 4.3 plots the Weighted Mutual information (WMI) term over a translation to the left of the moving image from figure 4.2. The left diagram illustrates that WMI contains a higher peak at the alignment of the upper box (translation=0.25) for an overlap function rising linearly with respect to the translation (derived from the test images). For a smooth overlap function that causes a single similarity measure peak in the center (right diagram), that peak is shifted towards the alignment of the upper box with our weighting. Here not the linear overlap o_1, o_2 wrt. translation is used, but a Gaussian function. The Mutual Information term used is $2-(2H_{ab})/(H_a+H_b)$ (equal to eq. 2.33).

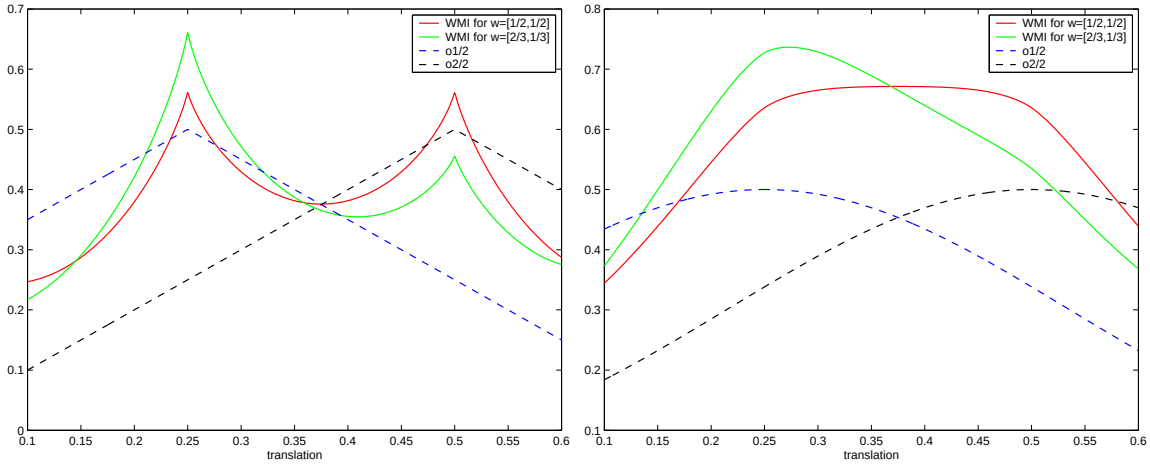


Figure 4.3.: Weighted Mutual Information (WMI) wrt. translation of the moving image, for different overlap functions and weightings. The dashed lines show the relative overlaps o_1, o_2 for the test images (upper diagram) and a Gaussian overlap function (lower diagram).

4.3. New Methods 2: Improved Registration Based on Ultrasound Simulation

The methods presented above are only capable of achieving a local alignment, and also require manual selection of adequate ultrasound frames, as well as the adjustment of a number of threshold values and other algorithm parameters. We therefore investigated further ideas, heading in a somewhat different direction. It resulted in methods that allow better simulation of ultrasonic effects from CT without compromising the computation speed. Besides, a robust new similarity measure is developed that assesses the correlation of a combination of multiple signals extracted from CT with ultrasound, without knowing the influence of each signal. It overcomes the main weaknesses of the previous methods, especially the instability of the Mutual Information component.

4.3.1. Simulation of Ultrasound from CT

An ultrasound wave is partly reflected whenever a change in acoustic impedance is encountered in the imaged tissue. The acoustic impedance $Z = \rho c$ depends on the tissue density ρ and the speed of sound c. Ultrasound machines assume a constant $c = 1540 m/s$ in human soft tissue, while a significantly different speed of sound occurs e.g. in air and bone. Table 4.1 lists the relevant values for various tissue types, from [131]. The ratio of an ultrasound wave intensity reflected at a tissue interface with different acoustic impedances Z_1 and Z_2 is $(Z_2 - Z_1)^2/(Z_2 + Z_1)^2$, given a specular interface with angle of incidence equal to the angle of reflection. The diffuse reflection, reflected straight back to the ultrasound transducer depends

Material	μ	ρ	c	Z
Bone	1000	1912	4080	7.8
Muscle	10..40	1080	1580	1.7
Liver	40..60	1060	1550	1.64
Blood	40	1057	1575	1.62
Kidney	30	1038	1560	1.62
Brain	43..46	994	1560	1.55
Water	0	1000	1480	1.48
Fat	-100..-50	952	1459	1.38
Air	-1000	1.2	330	0.0004

Table 4.1.: CT Hounsfield Units μ and physical properties of various tissues (density ρ, speed of sound c and acoustic impedance Z).

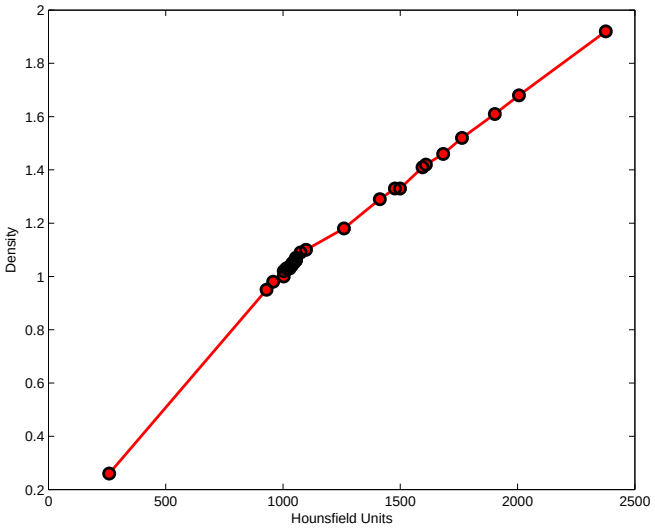


Figure 4.4.: Plot of CT Hounsfield Units against Tissue Density.

on the angle:

$$\Delta r(Z_1, Z_2, \theta) = (\cos\theta)^n \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 \tag{4.24}$$

$$t(Z_1, Z_2) = 1 - \left(\frac{Z_2 - Z_1}{Z_2 + Z_1}\right)^2 = \frac{4Z_2Z_1}{(Z_2 + Z_1)^2} \tag{4.25}$$

The exponent n describes the heterogenity on the interface, causing the amount of reflection to be more or less narrow around the perpendicular of the tissue interface. We lack detailed physical knowledge from CT, hence we use $n = 1$, it produces good results and simplifies the equations. Higher values would restrict the reflections of non-perpendicular interfaces, possibly missing to extract some features from the CT intensities. On the other hand, the similarity measure that will be introduced below is to some extent capable of ignoring additional information not present in ultrasound. The transmitted intensity $t(Z_1, Z_2)$ does not

depend on the angle of incidence, if refraction is neglected.
The X-Ray attenuation μ measured by a CT scanner is approximately proportional to the tissue density, see figure 4.4. As tissue density is in turn proportional to acoustic impedance, we can directly derive the an incremental acoustic intensity reflection from it:

$$\Delta r(\vec{x}, \vec{d}) = \left(\vec{d}^T \frac{\nabla\mu(\vec{x})}{|\nabla\mu(\vec{x})|}\right)^n \left(\frac{|\nabla\mu(\vec{x})|}{2\mu(\vec{x})}\right)^2 \tag{4.26}$$

$$\text{for } n = 1: \quad \Delta r(\vec{x}, \vec{d}) = \left(\vec{d}^T \nabla\mu(\vec{x})\right) \frac{|\nabla\mu(\vec{x})|}{(2\mu(\vec{x}))^2} \tag{4.27}$$

$$t(\vec{x}) = 1 - \left(\frac{|\nabla\mu(\vec{x})|}{2\mu(\vec{x})}\right)^2 \tag{4.28}$$

$\mu(\vec{x})$ is the CT attenuation value at position $\vec{x}$, $\nabla\mu(\vec{x})$ its spatial derivative. $\vec{d}$ is a unit vector denoting the direction of the ultrasound wave propagation, the scalar multiplication with the normed CT gradient vector yields the angular dependency equivalent to $cos(\theta)$. The ultrasound wave intensity is reduced according to $t(\vec{x})$ at each tissue interface, while $\Delta r(\vec{x}, \vec{d})$ contributes to the wave intensity detected by the probe. Integrating over this reflection and transmission behavior yields for any depth along a scanline:

$$I(\vec{x}) = I_0 \exp\left(-\int_0^{\lambda_x} \left(\frac{|\nabla\mu(\vec{x_0} + \lambda\vec{d})|}{2\mu(\vec{x_0} + \lambda\vec{d})}\right)^2 d\lambda\right) \left(\vec{d}^T \nabla\mu(\vec{x})\right) \frac{|\nabla\mu(\vec{x})|}{(2\mu(\vec{x}))^2} \tag{4.29}$$

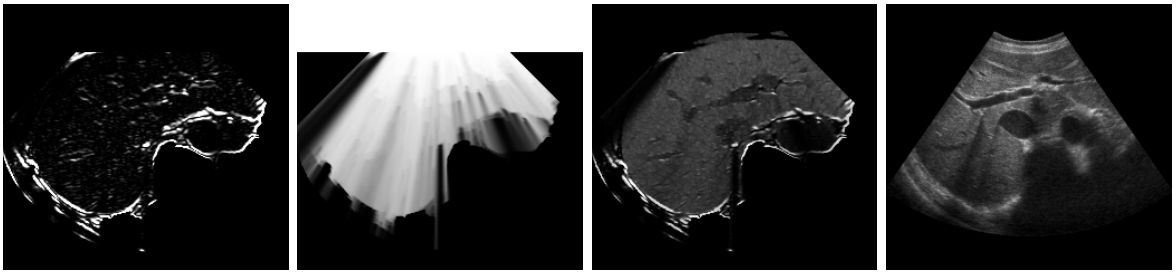
where I_0 is the original intensity of the ultrasound pulse, we define it as $I_0 = 1$. In addition, we apply a log-compression with one parameter a, which amplifies smaller reflections (resembling the Dynamic Range knob on the ultrasound machine), yielding the resulting value of the simulation:

$$r(\vec{x}) = (log(1 + aI(\vec{x})))(log(1 + a)) \tag{4.30}$$

For a linear array probe, the integral in equation 4.29 can be computed efficiently by traversing the columns in the simulated ultrasound image from top to bottom while updating the transmitted intensity based on the interpolated CT intensity and gradient values. For curvilinear arrays, we compute the image row-wise from top to bottom, while using an auxiliary channel storing the remaining transmitted ultrasound wave intensity (starting with 1 in the first row). For every pixel, this transmission value is retrieved by linear interpolation from two pixels in the above row, according to the ultrasound ray angle derived from the curvilinear geometry.

This provides a means to simulate large-scale ultrasonic reflection at tissue boundaries, and the related shadowing effects at strong interfaces like bone. However, individual tissue types have specific echogeneity and speckle patterns by themselves, based on the microscopic tissue inhomogenities. There is no simple relationship between tissue echogeneity and CT hounsfield units, therefore we add an intensity mapping $p(\vec{x})$ (further described in section 4.3.2) on a narrow soft-tissue range to the simulated large-scale reflection $r(\vec{x})$. Figure 4.5 depicts the simulation result for a transversal liver image, computed from a native CT scan. Figure 4.6 shows similar results for a portal-venous phase CTA scan.

4.3.2. Registration Algorithm



(a) reflection r (b) transmission t (c) simulation $r + p$ (d) ultrasound

Figure 4.5.: Simulation of ultrasonic effects from CT.

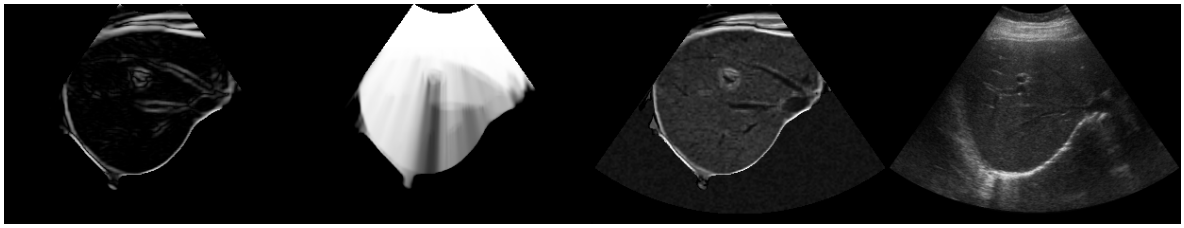


Figure 4.6.: Ultrasound simulation from portal-venous phase CTA, from left to right: Reflection r, transmission t, simulation $r + p$, original ultrasound. Here, 3-dimensional Perlin noise has been added to the occluded part of the simulation.

Automatic frame selection

Since we simulate ultrasound imaging effects with respect to the probe geometry, the original B-mode scan planes of the sweep have to be used rather than a 3D reconstruction. Neighboring frames of the freehand sweep contain similar information, hence we use always the one out of n frames that has the highest image entropy. This assures that frames which contain unique fine vascularity, that can be located in CT as well, are picked for registration. If two neighboring frames have the highest entropy out of their group of n, only one of them (again with the highest entropy) is used. Furthermore, a threshold is used to discard frames at the beginning and end of the sweep with little structures. In our experiments, n=3 was used, resulting in 15-22 frames per sweep for registration.

Idealized Intensity Prior

It seems appropriate to use statistical similarity metrics like Mutual Information (MI) and Correlation Ratio (CR) for assessing the correspondence of original CT and ultrasound intensities. In their general formulation, however, they do not work well for our registration problem, since there are too many possible configurations where the Joint Entropy is minimal (for MI), or the intensities from one image can be predicted well from the other one (for CR). At correct alignment of CT and US, they typically produced only a small local optimum. Known approaches for restricting the possible intensity distributions are distance metrics to Joint Histograms learnt from correct registrations (e.g. Kullback-Leibler-Divergence), as well as bootstrapping parameters for a polynomial intensity mapping in the actual registration process itself [118]. In both cases, very important information is disregarded, as e.g. small

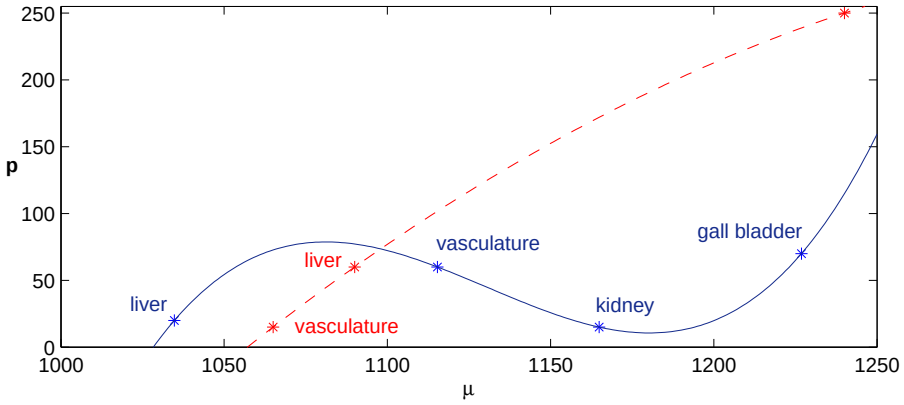


Figure 4.7.: Intensity mapping p for CT (red, dashed) and portal-venous CTA (blue) soft tissue. Note that the liver-vasculature relation is inverted in the two modalities.

vascularity is essential for a correct registration within the liver, but due its appearance on a relatively small fraction of the image content, it would neither affect a Joint Histogram or a least-squares estimate of a polynomial intensity mapping. Since CT attenuation measurements are mostly reproducible, we define a mapping function $p(\mu)$ based on a number of correspondences (liver tissue, liver vasculature, kidney, gall bladder) between CT/CTA intensities and tissue echogeneity in ultrasound, see figure 4.7.

Similarity Measure

In a Correlation Ratio framework, the registration transformation parameters are modified in order to maximize

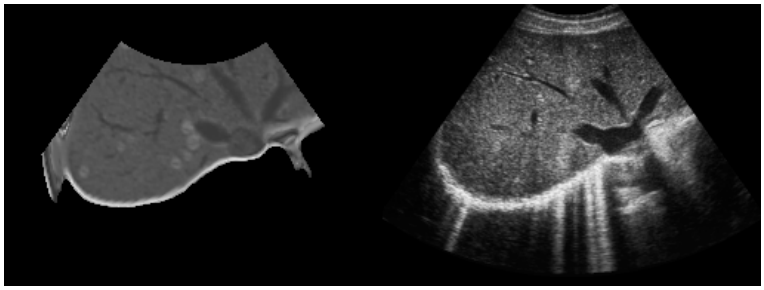
$$CR = 1 - \frac{\sum_{x\in\Omega}(U(x) - f(\mu(T(x))))^2}{|\Omega|\mathrm{Var}(I)} \tag{4.31}$$

with f denoting the mapping function which estimates the intensities of the image U from the transformed image μ. If a linear mapping $f(\mu) = \alpha\mu + \beta$ is assumed, equation 4.31 can be directly related to the common Normalized Cross-Correlation (NCC) similarity metric (see section 2.2.3).
For a pixel intensity in the ultrasound image, it is unknown how much the contribution of large-scale reflections and general tissue echogeneity is. Hence both the mapped CT intensity $p(\mu)$ and the simulated reflection r have to be integrated in a correlation framework with the ultrasound intensity. Using the notation $p_i = p(\mu(T(\vec{x_i}))), r_i = r(T(\vec{x_i})), u_i = U(\vec{x_i})$ for the intensity triple at a certain voxel, we define the intensity function as

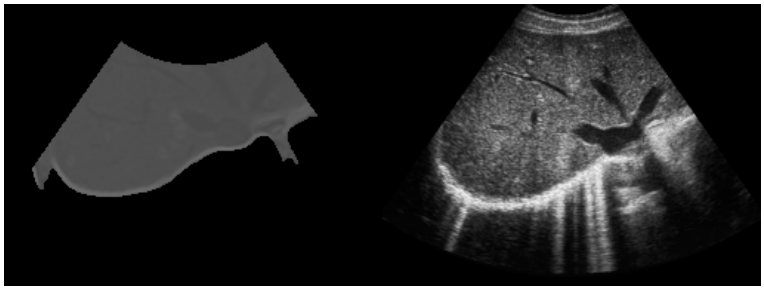
$$f(\vec{x_i}) = \alpha p_i + \beta r_i + \gamma \tag{4.32}$$

The unknown parameters α, β and γ then have to minimize

$$\left\| M \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} - \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \right\|^2 \text{ with } M = \begin{pmatrix} p_1 & r_1 & 1 \\ \vdots & \vdots & \vdots \\ p_n & r_n & 1 \end{pmatrix} \tag{4.33}$$



(a) $\alpha = 0.58$, $\beta = 0.64$, $\gamma = 0.20$



(b) $\alpha = 0.11$, $\beta = 0.14$, $\gamma = 0.29$

Figure 4.8.: The effect of simultaneous simulation and registration. (a) is well registered, (b) is $1cm$ displaced.

Therefore the solution is

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = (M^T M)^{-1} M \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} \sum p_i^2 & \sum p_i r_i & \sum p_i \\ \sum p_i r_i & \sum r_i^2 & \sum r_i \\ \sum p_i & \sum r_i & n \end{pmatrix}^{-1} \begin{pmatrix} \sum p_i u_i \\ \sum r_i u_i \\ \sum u_i \end{pmatrix} \tag{4.34}$$

Direct inversion of the symetric matrix $M^T M$ results in a closed-form solution for the parameters. They are then inserted in equation 4.31 to yield a novel registration similarity metric, which we denote *Linear Correlation of Linear Combination (LC^2)*. It assesses the correlation of ultrasound intensities u_i and a linear combination with unknown weights of signals p_i, r_i extracted from CT. The value of LC^2 is constant with respect to brightness and contrast changes of the ultrasound image (as NCC), but also independent to how much of the two described physical effects contributes to the image intensities. The latter is important, since e.g. hepatic vasculature or the gall bladder is represented mostly by p (different intensities due to echogeneity in ultrasound, no borders), while large-scale tissue interfaces correspond to r (strong edge in ultrasound, comparable intensities on both sides).
We compute equations 4.34 and 4.31 for every ultrasound frame in the set, and use the mean of the results as cost function. Note that an extension similarly to Local Normalized Cross-Correlation (LNCC) is possible by averaging over smaller overlapping patches in every ultrasound frame. More specifically, arranging the patches over 8 rings at different depths can fully compensate the time gain compensation (TGC) sliders on the ultrasound machine.

This implicit computation of the parameters α, β and γ in every pose evaluation equals a simultaneous optimization of simulation and registration parameters. Figure 4.8 illustrates it by showing the simulated intensity according to $\alpha p_i + \beta r_i + \gamma$ for an aligned and displaced

image of the liver. In the aligned image, the parameters α and β are higher, denoting a good least-squares matching of the structures. In the displaced image in turn, they are very small, causing both the vasculature and liver border to almost disappear. The value of γ is expressed with respect to a normalized intensity range $[0 \dots 1]$.
Both for disregarding fine speckle information in the registration, as well as speed-up of the computation, the ultrasound images are down-scaled to ∼156x116 pixels. The top 3cm along the ultrasonic rays are ignored for the measure computation, since they contain only compressed subcutaneous tissue.

Note: The fact that the weightings for p and r are unknown, suggests that a higher-dimensional Mutual Information approach could be used as well. Here, each p, r and u would represent one axis in a three-dimensional joint probability distribution. We had investigated this approach, computing modified versions of both Correlation Ratio and Mutual Information on a 3D joint histogram; it however resulted in a quite unstable similarity metric. One reason is that the reflection term r has an intensity distribution containing mostly small values (no reflections) and few yet important large values (representing reflections at tissue interfaces). A non-linear histogram equalization approach would be a pre-requesite to use r in such a framework. Another problem, as pointed out before in section 4.3.2, is the unconstrained huge number of intensity configurations that lead to a local optimum.

Optimization Strategy

A rough initial estimate of the orientation is obtained either from the tracking setup or by mapping the center slice of a sweep onto a transformation preset of typical transversal/longitudinal liver, and kidney sweeps, respectively. The large-scale translation is determined by performing a brute-force scan of the translation space. On the configuration which yields the highest similarity measure value (for a number of similar high results, the one closest to the preset transform is used), a local optimization of the translation is executed using a Simplex-based non-linear optimizer (see section 2.5.1). Successively, all six parameters of the rigid transformation are refined. As an optional last step, an optimization is executed on all rigid and three selected affine transformation parameters. These are the two scaling parameters and the one shearing of the sagittal plane, since respiratory motion mainly causes deformation in that plane [121].

4.4. Summary

After reviewing related work, we have devised two new methods for image-based registration of CT and ultrasound. They both incorporate knowledge about the underlying physics and image formation process in the two modalities, and are therefore able to compare the images without the need for learning-based techniques. The algorithm presented in section 4.2 is centered around a multi-component similarity measure. A combination of a weighted Mutual Information term, edge correlation, clamping to the patient's skin surface and occlusion detection is able to assess the alignment of structures in ultrasound images and information extracted from CT.
In section 4.3 we have developed a means to simulate the two main ultrasonic effects from CT, namely large-scale reflections and tissue echogeneity. Instead of weighting them in a multi-component measure, we presented the novel LC^2 similarity measure, which assesses

the correlation of a linear combination of multiple signals, independently of the influence of each of the signals. It implicitly optimizes the weight of each of the signals in the simulation for every similarity measure evaluation. While being an ideal measure for comparing our simulation result with ultrasound, this multi-signal correlation is a general concept than can serve in other difficult multi-modal registration problems as well.

5. Clinical Applications

5.1. Radiation Therapy of Head and Neck Cancer

5.1.1. Clinical Context

In radiation treatment planning for inoperable head and neck cancer, the identification of metastatic neck lymph nodes is mandatory for the correct target volume delineation. This can be achieved with a reported accuracy of 80-95% using high-frequency ultrasound [3, 161, 176]. In direct comparison with ultrasonography, diagnostic CT was equally predictive in revealing lymph node size, but achieved lower performance in depicting internal nodal architecture, leading to a lower sensitivity and specificity than ultrasonography [149]. In planning CTs for radiotherapy, contrast medium is usually omitted, and consequently their diagnostic properties are particularly poor.
Omitting the CT and performing the treatment planning and execution solely based on ultrasound has its place only in brachytherapy (e.g. in prostate cancer), where a small treatment volume can be visualized together with the radiation source(s) by ultrasound without artifacts from bone or air, the dose distribution being dependent on distance rather than on tissue properties. However, in the context of external-beam radiotherapy (EBRT) delivered from different angles, axial sections of the whole body are mandatory and cannot be provided by ultrasound.
Thus, for external-beam radiotherapy, CT remains the base imaging modality of choice for treatment planning and simulation, as it provides a coordinate systems with stable geometric fidelity and the necessary electron density information for the computation of the accurate dose distribution within the CT anatomy. Consequently, the wish to integrate the diagnostic properties of other imaging modalities into the planning process has usually been met by registration with CT, either based on external markers or anatomy information (e.g. registration of MRI and CT) or with dedicated combined systems like PET-CT. Similar solutions of registration between freehand ultrasonography and planning CT seem to be a valuable goal, as the additional US information may enable the radiation oncologist to refine the target volume definition and individualize treatment planning. For example, in studies on integrating PET into the treatment planning process of brain or lung tumors significant changes in gross tumor volume (GTV) delineation could be demonstrated (review by [50]). It is quite possible to have a similar impact of ultrasound information for GTV definition in head and neck cancer.

5.1.2. Related Systems

In the context of radiotherapy, tracked ultrasound has been used mainly to quantify and reduce daily set-up errors so far. The commercial NOMOS B-mode acquisition and targeting system (BAT)[1] allows to integrate the CT planning contours with tracked ultrasound acqui-

[1] Website: `http://www.nomos.com`

sitions prior to each treatment fraction. A description of the system, as well as a comparative study of prostate localization based on BAT and CT can be found in [89]. Though the largest experience has been gathered with prostate cancer [44], some data on other sites such as the upper abdomen is available as well [45]. A similar commercial solution is SonArray [44, 130] by Varian Medical Systems[2]. The RESTITU system (Resonant Medical)[3] uses two calibrated 3D ultrasound devices, in the CT and treatment room, respectively [33]. Hence the problem of aligning planning information within the coordinate system of the treatment room is reduced to a monomodal registration of the fractional 3DUS acquisition with the reference 3DUS scan taken in the CT room.

5.1.3. Proposed Medical Workflow

For the described application of radiotherapy planning, a workflow (steps 1-8) is needed to integrate the diagnostic ultrasound with the current treatment planning processes. The planning CT is usually performed with the patient fixed onto a head pad with a thermoplastic mask individually molded. This minimizes the spatial deviation between the anatomy depicted in the CT scan and the daily radiation treatment delivery.

1. Just before ultrasonography the patient is immobilized likewise to ensure the same reclination of the head. Then, the mask is carefully removed while the patient is told to keep his position for the following examination.

2. The diagnostic ultrasound is performed by slow freehand sweeps of the probe in lateral and craniocaudal directions along the patient's neck and chin with special regards to the lymph node regions (i.e. submental, submandibular, digastric, jugular chain, spinal accessory and occipital nodes). In our scenario, a position sensor is attached to the ultrasound probe. All diagnostic sweeps are recorded as a combination of videos containing the actual ultrasound images, alongside the tracker readings. An off-line calibration step allows us to place all recorded ultrasound images in the spatial context of the tracking world coordinate system (Figure 5.1). Hence, three-dimensional ultrasound information is obtained with full coverage of the patient's neck and chin.

3. Our system now applies an automatic pre-registration step by aligning the lines denoting the top of the numerous ultrasound images in 3D-space with the skin surface extracted from the CT data set, see section 5.1.4 for details.

4. From all sweeps with relevant lymph nodes, the physician picks some particular frames containing structures that are identifiable in both modalities, like the carotid artery (maybe with individual calcifications), internal jugular vein (preferably with little or no compression), certain muscles or thyroidal/salivary gland tissue.

5. These frames serve for reviewing the first results from global skin surface registration.

6. In case of excessive deviation, manual drag and drop of the stack of ultrasound pictures into a closer range of corresponding CT anatomy is performed. Any frame can be selected and its relative location and orientation modified, while the spatial relation to all other frames (given by the position sensor) is maintained. This allows a convenient

[2]Website: `http://www.varian.com`
[3]Website: `http://www.resonantmedical.com`

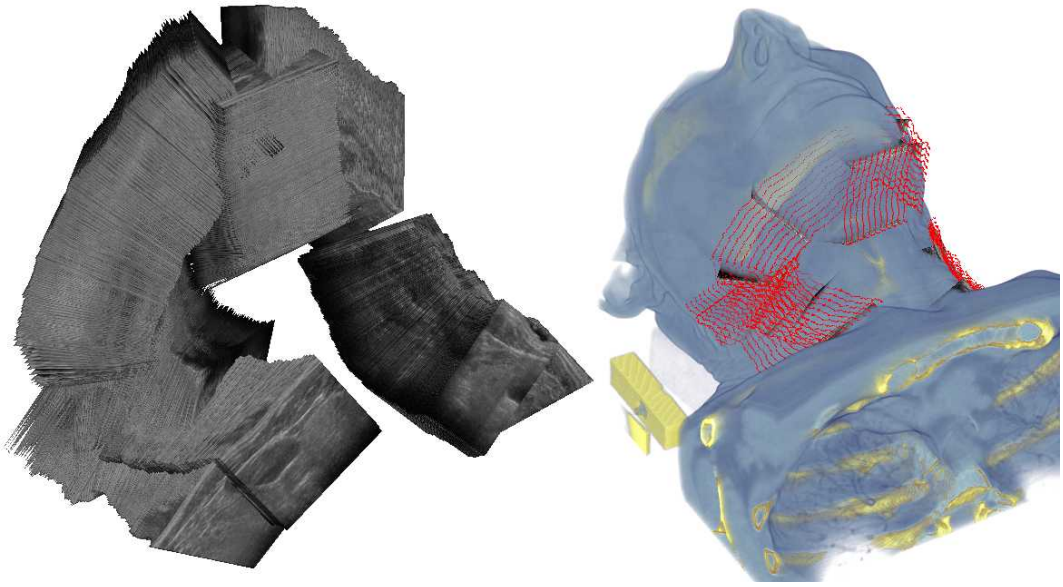


Figure 5.1.: Left: Spatial overview of all ultrasound images of Patient 3. Right: Extracted surface points (red) with volume rendering of CT and some individual ultrasound slices.

navigation, as in-plane and out-of-plane transformation parameters can be manipulated separately. Either a 3D display of the data or side-by-side visualization is used, as shown in figures 5.4(b), 5.5.

7. This allows an automatic image-based registration to be performed on these images successively, as has been described in Section 4.2.

8. The results from image-based registration are evaluated and may serve for target volume delineation. In particular, our method overlays pathological lymph nodes areas, which need to be identified and irradiated with a certain dose, but are barely visible in and extractable from the CT data alone.

An adapted treatment planning software should allow the physician to scroll through the registered ultrasound frames, displayed alongside corresponding Multi-Planar Reconstructions (MPRs) interpolated from the CT data, and vice versa through the original axial CT slices overlaid with the registered three-dimensional US data in order to account for all additional information during target volume delineation.
In the following sections, we present details of the proposed registration steps, involving the image-based algorithm introduced in section 4.2, and describe the setup and results of experiments used for validation.

5.1.4. Global skin surface registration

The set of all ultrasound images recorded from a patient including the submental and submandibular regions as well as both sides of the neck (Robbins level I-VI) can be used to derive an approximate registration to the CT scan. In each ultrasound image, the horizontal

line denoting the top of the image roughly lies on the skin surface. As the images are continuously recorded and tagged with position information, all frames acquired from a patient (up to several thousand images) yield a complete coverage of the neck surface in the tracking coordinate system. Thus a surface-based registration with the three-dimensional skin surface segmented from CT will provide a fairly good global initial registration.
The CT scan is automatically segmented with a region-growing method [1], using a seed point in the surrounding air and inverting the result afterwards. Successively, a border detection algorithm is run, and a Chamfer distance map transformation [21] is applied to the surface. It results in a distance volume where the voxel values reflect the closest approximate Euclidean distance to the surface. 10 equal-spaced points on the top of each ultrasound image form the 3D point set. The global rigid transformation is initialized with a 90-degree rotation that brings the tracking world coordinate system in the same orientation than the CT volume, and a translation which consists of the subtracted mean of all points (hence the point set is centered in the CT volume). For each estimate of the global rigid transformation, the distance of the transformed points to the CT surface can be efficiently looked up in the distance volume. The transformation parameters are iteratively refined until the error converges. Figure 5.1 shows the recorded ultrasound images in a three-dimensional visualization, as well as the 3D point set derived from them, overlaid onto a volume rendering of the CT scan.

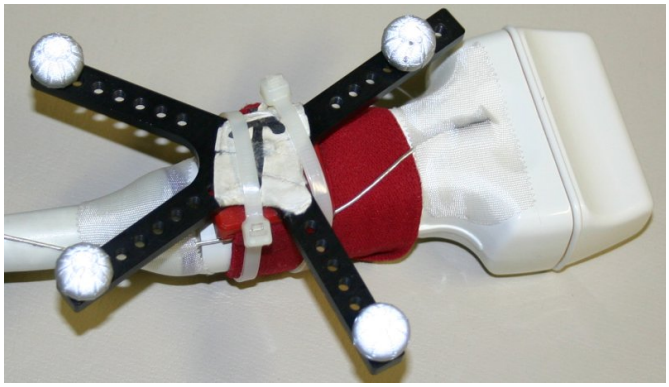
For refining the obtained initial registration, the image-based registration algorithm described in section 4.2 is then executed.

5.1.5. Data Acquisition

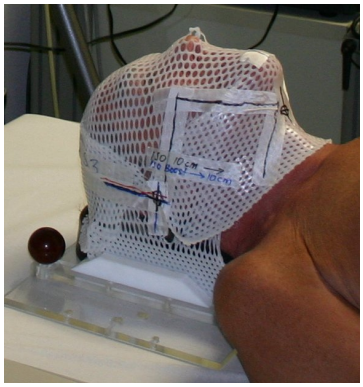
We conducted a study on five patients. For the freehand ultrasound acquisitions, we decided to install both optical and magnetic tracking. However for the evaluation of our algorithms, we used solely the optical tracking data. In the future, information from both tracking technologies will allow interesting comparative studies. All ultrasound examinations were performed with a GE Logiq 500-scanner and a 8.5-11MHz transducer with a 40mm linear probe (LA39, GE Healthcare Technologies, Waukesha, Wisconsin, USA). An optical tracking target, consisting of four infrared marker balls, was mounted on the handle of the ultrasound probe using cable ties and a layer of Varihesive bandage-aid between probe and target (see (Figure 5.2(a)). We tested the tracking setup extensively to assure that the target does not hinder the physicians flexibility when scanning a patient, in particular that the optical target does not touch the patient's skin. On the other hand, the target should be recognized by the tracking cameras at all times.

The final setup contained four optical ARTtrack2 cameras (A.R.T. GmbH, Weilheim, Germany) arranged behind the head of the patient (Figure 5.2(c)). A two-camera system would have been sufficient if mounted somewhat higher. As we had four cameras at our disposal, we used them to obtain a symmetric setup, so that the expected tracking precision is comparable on both sides of the patient's neck. The ultrasound data was recorded using an IDS FALCON frame grabber card (IDS Imaging Development Systems GmbH, Obersulm, Germany)[4]. The ART tracking data was received over UDP network transmission from the separate PC running the ART DTrack software. We used the CAMPAR framework for Medical Augmented Reality Applications [138], which was developed by colleagues in our group, for the acqui-

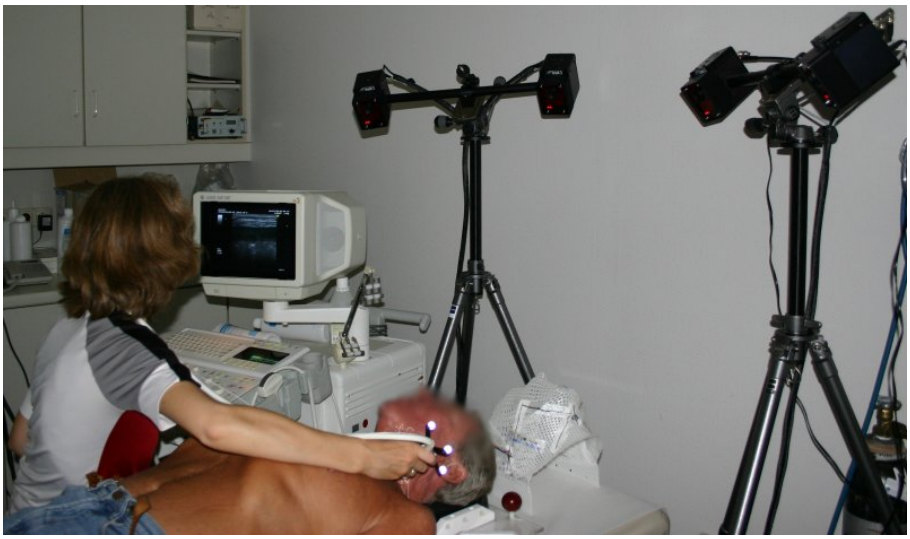
[4]Website: `www.ids-imaging.com`



(a) Ultrasound transducer with optical and magnetic tracking targets



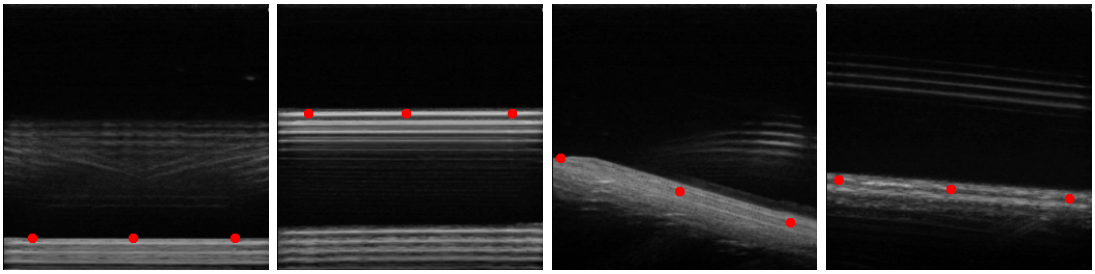
(b) Positioning with mask



(c) Actual patient scanning

Figure 5.2.: Setup for 3D freehand ultrasound acquisitions.

sition. We implemented the recording of sequences by storing the gray-valued video region containing the actual ultrasound image (size 454x454) on disk, alongside with its timestamp and the corresponding 4x4 tracking matrices. The resulting performance was the frame grabber's full speed of 25 frames per second; 5 MB/s are written to disk. On mouse clicks, the corresponding 2D position in the ultrasound video was logged in an additional text file, along with the current tracker poses. We included an additional pose of a tracked tool, whose tip intersected the ultrasound image, for calibration purposes (see section 5.1.6).
For the actual freehand ultrasound examinations, patients were positioned supine on the examination couch with the head pad supporting the neck and the individual thermoplastic immobilization mask covering the face, as shown in Figure 5.2(b). Then the mask was removed, taking care not to move the head that was still stabilized by the preformed pad beneath, and the patient was told to keep this posture. From each patient, 11-16 ultrasound sequences, up to 45 seconds each, were recorded. It resulted in 5000-11000 ultrasound images



(a) Straight, deep (b) Straight, shallow (c) Rotated Probe (d) Tilted Probe

Figure 5.3.: Ultrasound images from the water bath calibration scan, red dots show points used for the floor reconstruction.

per patient.
Right after the examination, we took a CT scan of each patient in the radiology department. The device was a Siemens Sensation Cardiac 64, used with the standard neck scanning protocol and reconstruction with 0.6mm slice thickness.

5.1.6. Calibration

In order to obtain the position and orientation of the ultrasound image plane with respect to the tracker's world coordinate system, a calibration is necessary to derive the transformation that relates the image plane to the tracking sensor coordinates, see section 3.4. Here, we adapted the Single-Wall calibration technique [113] to our experimental setting, in order to yield an easy yet sufficiently accurate calibration, as described in the following.
For any point $\vec{p} = (u, v, 0, 1)^T$ in the ultrasound imaging plane, the corresponding point $\vec{p_w}$ in the tracker's world coordinate system is obtained by:

$$\vec{p_w} = T_T T_C \vec{p} \tag{5.1}$$

The calibration transformation T_C maps a point from the ultrasound coordinate system to the tracking target, while T_T in turn transforms it to the tracker's world coordinate system. The ultrasound probe is dispensed in a shallow water bath, and moved in a variety of locations and orientations. The floor then appears as a line in the ultrasound image, and the user clicks on positions of that line in the video image. 100-200 of those 2D positions in the ultrasound image (2-3 positions per still frame) are recorded alongside the tracking matrices in a file.
Using a water-level device, we assured that both the floor of the water bath and the rim of the container are exactly parallel to the water surface. By means of the ART DTrack room calibration [2], the tracker world coordinate system could be defined such that the water bath floor is exactly in the x-y plane.
Thus the calibration transformation T_C can be recovered by minimizing the variance of the z component of the points $\{p_w\}$ reconstructed with Equation 5.1. The Matlab Simplex optimizer was used to estimate the Euler angles parameterization of T_C. The resulting standard deviation of the reconstructed points in the z direction is about $1.1mm$ for multiple calibration runs.
We manually selected points on the line representing the floor in the ultrasound image. Depending on the angle of the probe, the appearance of the floor becomes manifold and blurred,

P.-S./side	anatomy	target LN / diameter	surface TRE	auto. TRE
1-6/left	LNL IV-II, CA, IJV with some compression	normal / 5 mm	28.9 mm	4.1 mm
2-1/left	LNL IV-II, CA, IJV, chain of LN metastases	metast. / 12 mm	54.5 mm	3.3 mm
2-10/right	LNL IV-II, CA, IJV, sternocleidoid muscle	normal / 7 mm	53.4 mm	2.8 mm
3-6/right	LNL III-II, larynx, bifurcation, submand. gland	suspect / 12 mm	4.9 mm	3.5 mm
3-7/left	LNL IV-II medial, trachea, thyroid, larynx, CA	normal / 5 mm	4.9 mm	3.8 mm
4-2/right	LNL IV-III, thyroid, CA, IJV	metast. / 12 mm	4.1 mm	3.0 mm
4-6/right	IB-IIA-IIB-V, bulky lymph node disease	metast. > US probe	7.7 mm	6.1 mm
5-5/right	IIA-craniocaudal, parotis, submandibular gland	suspect / 19 mm	6.1 mm	4.6 mm
5-6/right	IA-IIA, chin, jaw, digastric muscle, sm.gland	normal / 10 mm	10.7 mm	4.0 mm

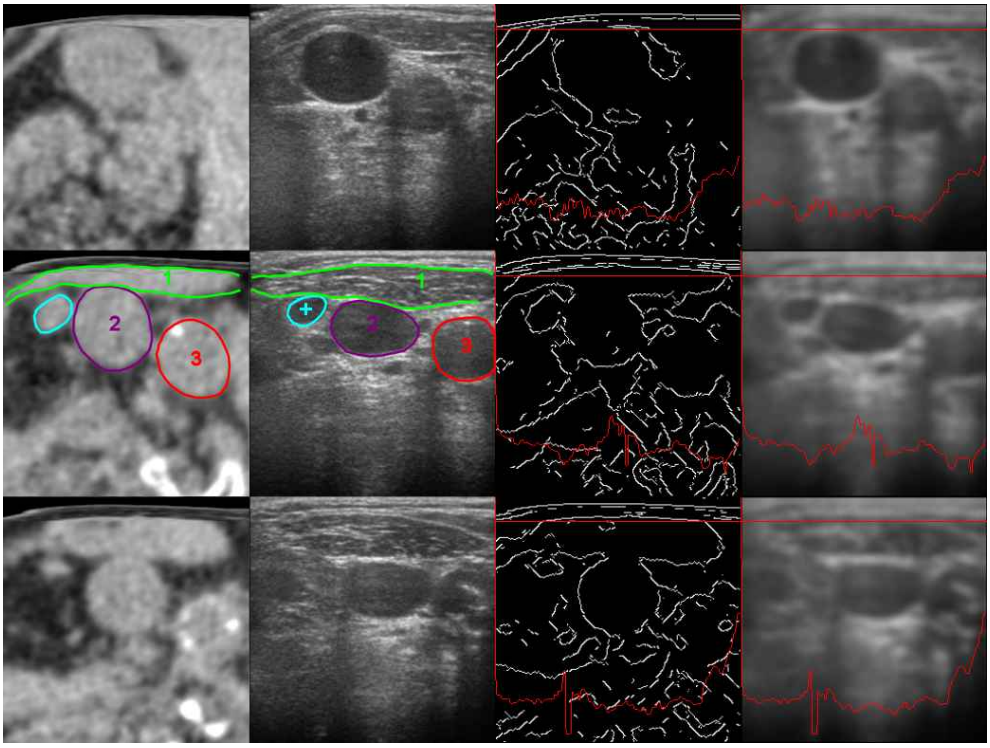
Abbreviations: P.-S.=patient and sweep number, LN=lymph node, diameter=maximal nodal diameter according to US sections, surface TRE=Target registration error after skin surface registration, auto. TRE=Target registration error left after additional image-based registration, LNL=lymph node level (Robbins), CA=carotid artery, IJV=internal jugular vein

Table 5.1.: Description of datasets and target registration errors.

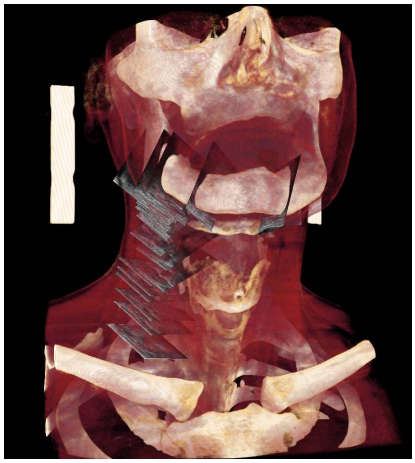
rather than showing a sharp line (Figure 5.3). Using careful adjustment of a single focal zone, the floor plane can be determined even for difficult angles. In particular, when slowly moving the probe from a perpendicular to a steep position, one can see how the first characteristic sharp line of the floor merges into a more complex structure, while the actual position of the floor is still identifyable. This approach allowed us to avoid manufacturing calibration hardware such as the Cambridge phantom [113], which overcomes this problem by providing an obstacle always perpendicular to the image plane. Furthermore, as the calibration did not need to be done on a regular basis by physicians, we did not consider more automated approaches involving complex algorithms for estimating the correct floor points, as in [124]. For Single-Wall calibration with automatic floor line extraction, a reconstruction precision of $3.4mm$ is reported in [113] with magnetic tracking, and $2.7mm$ in [124] using optical tracking. In a more recent study [125], these two methods are quantitatively evaluated both with optical tracking, the authors report mean 3D point localization errors of $2.7mm$ and $1.67mm$, respectively.

Medical ultrasound devices usually assume a constant speed of sound of $1540m/s$, which is the average in human soft tissue. In our water bath at room temperature, however, the speed is approximately $1485m/s$. Objects visible in the images are thus closer to the probe than indicated by the rulers on the image borders. Hence for our calibration, the vertical direction of the ultrasound image was scaled accordingly with a factor of $1485/1540 = 0.964$ (compare [91], page 458, Equation 7).

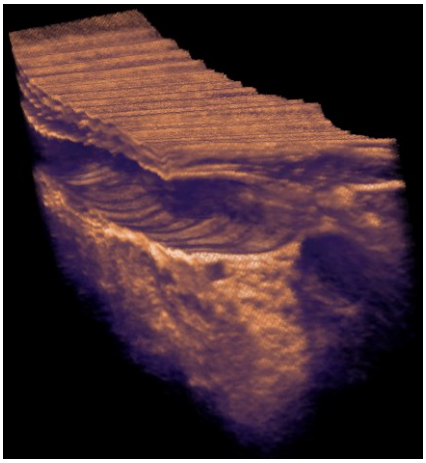
We used a second tracked tool to intersect the ultrasound image plane in the water bath. By means of a 3D visualization of the tool and the ultrasound image we visually confirmed the calibration was correct (i.e. the optimization had not minimized the floor plane to a wrong local optimum). It also allowed us to assess the latency of the ultrasound image with respect to the tracker readings, by means of periodic pointer movement. We concluded that no systematic temporal calibration is necessary. One would expect the ultrasound image to be up to one frame older than the tracker readings (due to the image processing and reconstruction on the ultrasound machine, as well as frame grabber delay). This equals to a maximum of $40ms$ delay, or $0.2mm$ error if a continuous sweep with the typical speed of $5mm/s$ is considered.



(a) Three corresponding CT (1st column) and ultrasound slices (2nd column) from Patient 2, right neck, with edges from CT (3rd column) and blurred ultrasound for edge alignment (4th column), positioned above, in the middle and below a normal target lymph node (blue outline, cross denotes center for TRE computation). Examples for corresponding anatomy: 1-sternocleidomastoid muscle, 2-internal jugular vein, 3-carotid artery. ROI is delineated in red on the 3rd and 4th column. The physical image size is $4 \times 4cm$.



(b) Fused visualization of several registered sequences of Patient 5



(c) Volume Rendering of compounded ultrasound sweep from Patient 2, clipped to show a profile of the internal jugular vein (note the ripple artifacts due to blood pulsation)

Figure 5.4.: Image data from different patients used for the evaluation

5.1.7. Results

For each of the five patients, all of the recorded tracked ultrasound sequences were registered onto the skin surface extracted from the CT scan, as described in Section 5.1.4. This registration always converged, with computation times being 1-5 seconds. The mean residual point distances (based on the Chamfer distance lookups) for the five patients were $2.36, 3.67, 2.41, 2.31, 2.70mm$, respectively.
For studying the accuracy and robustness of the intensity-based registration methods (section 4.2), one or two ultrasound sweeps were selected from each patient, 9 in total. Together, the selections should cover the whole range of the expected clinical applications, including the typical lymph node sites on both sides of the neck and displaying pathological as well as normal lymph nodes of different sizes. Of each sweep in turn, $4 - 7$ frames were picked for automatic image-based registration. For validation purposes, the physician carefully established a Ground-Truth transformation for each set of images using anatomical landmarks such as calcifications in the carotid artery, well-defined lymph nodes or gland tissue etc. For each set, one particular lymph node of interest was selected as the target, its deviation after the automatic registration with respect to the manually defined Ground Truth represented the Target Registration Error (TRE). The Hill-Climbing optimizer was used with an initial step size of $5mm/5°$, and the similarity measure weightings were $w_0 = w_1 = w_2 = 0.33$. Before registration, the ultrasound images were downsampled to 128 x 128 resolution, as their original high spatial sampling rate is not given in the CT data set and hence cannot be exploited for registration. In addition, downsampling the ultrasound images positively affects the computation speed.
Table 5.1 depicts the registration accuracy with respect to the selected target for all 9 sequences, after surface registration (column 5, surface TRE) and image-based registration (column 6, auto. TRE). In the first two patients, large translational misalignments between pure surface registration and Ground Truth were noted, mainly in the lateral or superior/inferior dimension. In the next three patients the deviations for surface registration ranged between 4.1 and 10.7mm. After subsequent image-based automatic registration (in case of the first three sequences after some manual adjustment in order to lie within the capture range), the remaining mean Target Registration Error was 3.9mm. Apart from sweep 4-6, where the size of the lymph node metastasis exceeded the 40mm linear ultrasound probe and little anatomic distinction was given in the CT data, the Target Registration Error after image-based fusion remained below 5mm in all cases.
Based on the findings above the physician would have to expect the following clinical workload and estimated additional time consumption per patient to enable registration of ultrasonography with planning CT (with the setup of the camera system/calibration of optical tracking provided and planning-CT performed routinely):

1. Patient positioning with mask and head pad on the ultrasound examination table, removing the mask (preferable following the diagnostic ultrasound examination): **1 min**

2. Acquisition of 1-6 freehand ultrasound sweeps (each 20-30 sec) containing all pathological lymph nodes on both sides of the neck: **4 min**

3. Computation of US-CT global skin surface registration: **5 sec**

4. Selection of 4-7 US frames per sweep for image-based registration: **4 min**

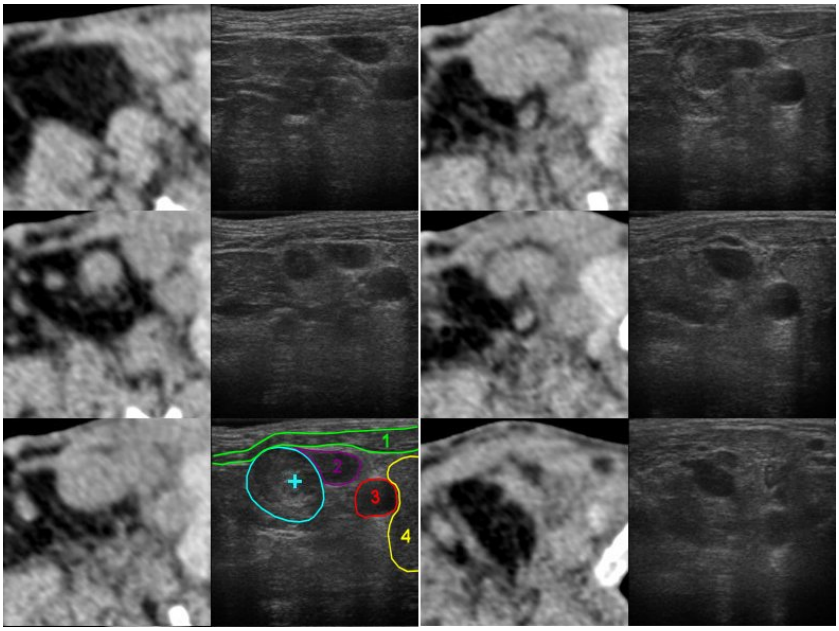


Figure 5.5.: Images used for registration from Patient 4, 6 consecutive ultrasound and CT slice pairs; a malignant target lymph node (blue outline and TRE cross; 1-sternocleidomastoid muscle, green, 2-internal jugular vein, purple, 3-carotid artery, red, 4-thyroid, yellow)

5. Visual control, if global skin surface registration is good enough to allow for image-based registration: **3 min**
6. if not, manual drag and drop of the $> 1cm$ deviating sweep(s) into the range of successful image-based registration: **5 min**
7. Computation of image-based registration: **30 sec**
8. Re-evaluation: **2 min**

Thus, in case the global surface registration deviates substantially, performing all steps would amount to approximately 20 minutes additional working time for the physician (patients 1 and 2), which could be cut down to 15 minutes by leaving out step 6 in case of sufficient global surface registration to allow for a meaningful image-based registration (patients 3-5).

In order to evaluate the robustness of the different optimization strategies, 130 registrations were launched from initial transformations randomly displaced up to $4mm/4°$ in each parameter around the ground truth pose. It was executed for all three optimizers on a set of 7 ultrasound images from sequence 10 of Patient 2, three of which are shown in Figure 5.4(a). The same set of random numbers was used in each case, to assure sufficient statistical significance of the randomized study. Table 5.2 denotes the variance of the error in the translational and rotational components of the resulting transformation in Euler angle parameterization, as well as the mean target registration error (TRE) for a lymph node (Figure 5.4(a)) picked as target, and the computation time (on a 2.2GHz AMD Opteron machine with 2GB RAM running Linux). The Hill Climbing Optimizer is stable while requiring the least number of

Optimizer	σ_t(mm)	σ_r(mm)	TRE(mm)	time(s)
Hill Climbing	1.74	2.92	2.78	4.70
Powell-Brent	1.96	2.90	3.04	23.38
Exhaustive H.C.	1.11	3.14	2.26	336.30

Abbrev: σ_t=translational deviation, σ_r=rotational deviation, TRE=mean Target Registration Error, time=computation time

Table 5.2.: Robustness and speed of different optimizers, evaluated by repeatedly registering from displaced starting estimates.

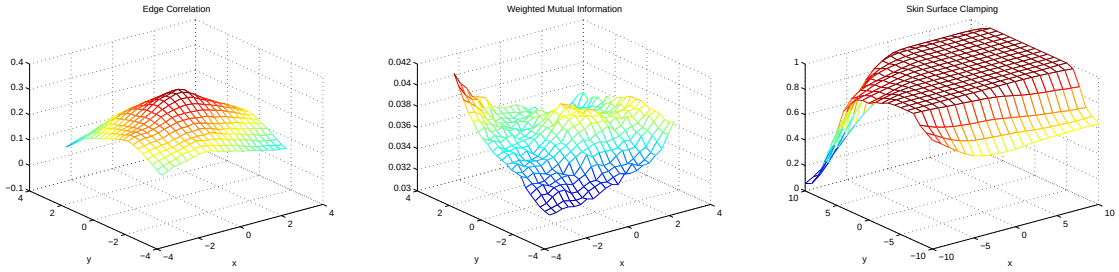


Figure 5.6.: Plots of the individual similarity measure components against two translational transformation parameters.

cost function evaluations, resulting in a very small computation time. Due to the expensive line minimizations, the Powell-Brent optimizer takes significantly more time but does not outperform the Hill-Climbing method. Exhaustive Hill-Climbing achieves the highest robustness, as it evaluates all combinations of possible search directions in each iteration, and hence always reaches the closest local optimum in the parameter space. As its computation time is very high though, we prefer to use the Hill-Climbing strategy as best tradeoff between speed and robustness. To show the contribution of the various similarity measure components to the overall registration process, Figure 5.6 plots their value against changes in two translational parameters, with respect to the ground-truth transformation. The Edge correlation term provides a smooth component, globally converging towards the optimum, hence driving the stability of the algorithm. Weighted Mutual Information contributes with a local peak denoting the highest statistical correspondence, however for larger translations, the individual image entropies used for normalization might change significantly, as the image contents become different. This can affect the Mutual Information value in a way that it produces wrong local optima (left side on the center plot). In order not to compromise the global convergence range, we have to limit the influence of the WMI term with respect to Edge Correlation. As its values are much smaller $(0.01 - 0.1)$ compared to Edge Correlation $(0.02 - 0.35)$, an equal weighting $w_1 = w_2 = 0.33$ represents a good balance between the two components. The skin surface clamping plot depicts a rim of value 1, dropping smoothly to zero when some of the ultrasound images drift away from the patient's skin. In fact, for our weight $w_0 = 0.33$, the optimization never allows any values other than one for this component.

5.1.8. Discussion

The described setup allows acquisition of freehand ultrasound data from all sides and orientations required for thorough examination of the patient's neck, while knowing the location of each image in 3D-space. However, several sources for error might affect a registration of this information to the corresponding CT data:

Tracking inaccuracy. In our particular case, this error is very small due to the use of the high-end ART tracking system with four cameras[5].

Calibration error. Our calibration method yielded a reconstruction plane standard deviation of 1.1mm (see Section 5.1.6). Considering that we perform sweeps in a continuous motion (without many angulations) and register them individually, we expect the calibration accuracy to have limited impact on the registration. However we did not perform a complete calibration accuracy study in the scope of this work.

Patient immobilization. Even though the patient was placed on the immobilization head pad also used for treatment, slight movements of the patient's head might have happened during the examination, as the thermoplastic mask was removed after the initial positioning. In addition, involuntary movements are possible in the context of counterbalancing the mild ultrasound probe pressure. The head pad was not fixed on the examination couch, thus only gravity and friction prevented its displacement. Similarly, the examination couch itself was a mobile design with parking brakes in use that did not completely prevent translational movement if leaning against it (e.g. examination of the patient´s left neck from his right side). These latter two aspects could be responsible for the large deviations in patient 1 and 2 and should be addressed in subsequent studies by using a stationary examination couch and firm fixation of the head pad. The former aspects are more difficult to counteract: leaving the thermoplastic mask in place would impede access to two thirds of the lymph nodes under question, cutting it according to the requirements of ultrasonography would lead to an unstable mask no longer sufficient for subsequent radiotherapy. Special face masks that gain stability from an individually moulded mouth piece are poorly tolerated by head and neck cancer patients due to hypersalivation and tumoros obstruction of airways. Because of these problems, the global skin surface registration of the whole exam may serve as initial estimation, but a more precise registration should be performed for individual sweeps, which is the case in our approach.

Internal tissue movement. The heart-beat causes the vessels to pulsate with a characteristic rhythm over time, with a single high pressure pulse per heart beat for arteries and a low pressure double pulse for veins resulting in tissue movements of 1-3mm; see Figure 5.4(c). In addition, some of the patients had to swallow during the examination, resulting in a strong shift of internal structures in the whole neck region. If deep breath-taking happens, other deformations occur. Hence we asked the patients for quiet, regular breathing throughout the examination, where only minor changes in the anatomy are expected.

Ultrasound image distortion. Speed-of-sound variations can cause geometric distortion in the ultrasound images. The speed of sound varies slightly in soft tissue (fat $1476 m/s$, water at body temp. $1529 m/s$, muscle $1568 m/s$, blood $1570 m/s$), while the ultrasound machine assumes $1540 m/s$ for the image reconstruction. The maximum deviation here is about 4%, hence e.g. a $1 cm$ fat layer would cause the tissue below to appear $0.4 mm$ displaced. Further distortion can appear due to refraction. In the given data a lateral shadowing effect on the carotid artery was a frequent finding, however we do not expect implications for the

[5] 0.4mm RMS position error over the whole measurement volume, as stated on `http://www.ar-tracking.de`

registration. The shadowing itself is excluded by our algorithm (see Section 4.2.3), and the slightly displaced echos below the carotid artery contribute with a low weight to the similarity measure, as they are on the bottom of the image.

Ultrasound probe pressure. In any ultrasound exam, some minimum pressure has to be applied in order to keep direct contact with the patient's skin. Where the skin surface is convex, the linear ultrasound probe will flatten the surface underneath, and the compression should be strongest in the middle of the probe. This causes deformation of the skin, subcutaneous fat, muscular layers and vessels. Placing and removing the ultrasound probe repeatedly at neighboring locations will thus lead to different distortion of the same structures. In particular, vessel structures, as well as lymph nodes close to the skin are affected: Due to different inherent blood pressures between the arterial (60-140 mmHg) and venous (2-10 mmHg) blood system, the external compression effect imposed by the ultrasound probe is much more pronounced in veins; see anatomy contours in figure 5.4(a). In some of the sweeps, we could observe that the internal jugular vein would be totally collapsed due to compression in some parts of the sweep, while it was visible with little or no deformation in others. Non-fixed lymph nodes may be shifted slightly sideways by the ultrasound probe, depending on the consistency of surrounding tissue. In the study presented, we handled these issues by manual selection of ultrasound image frames well-suited for registration, as opposed to a 3D-3D registration of all information available from a sweep with the CT scan. Thus, when two sweeps of freehand ultrasound with overlapping structures of interest are recorded, one can not expect to have an exact match of the corresponding anatomy in both sequences. Again, this supports our supposition that a precise image-based co-registration of anatomic structures to CT is restricted to individual freehand ultrasound sweeps.

Some of the issues mentioned above suggest the integration of deformable models into our registration methods, which we consider an important topic for future research. However, this will require a high degree of distinction between different tissue types with different deformability (arteries, veins, subcutaneous layers, muscle, fat etc.) and thus in turn will depend on segmentation algorithms and feature-based methods. Furthermore, validating the correctness of all registered CT-Ultrasound data in terms of a deformable mapping would be a very time-consuming effort, left with many uncertainties concerning the relation of physiological flexible soft tissues and rigid tumor tissue.

Considering a feature-based approach as an option for automatic registration between US and CT, the carotid artery would be a structure of choice, as it appears well in the ultrasound images, is less prone to artifacts from probe pressure as opposed to veins, and sited in close proximity to several lymph node regions. But for CT, it is angiographic acquisitions with contrast agent that allow one to precisely extract the tubular shape of the carotids [128]. However, standard planning CT scans for radiotherapy are performed without contrast agent; thus we made use of native CT scans in order to minimize the additional workload in the treatment planning process and to enhance the acceptance for the clinical routine.

According to our time estimates for the proposed procedure, the physician faces a maximum of 20 min additional workload in case of the global surface registration not meeting the requirements for subsequent image-based registration (patients 1-2). Avoiding pitfalls like an unstable head pad fixation and examination couch along with careful patient positioning will supposedly lead to a sufficient global surface registration like in our patients 3-5, in which case step 6 is dispensable, resulting in 15 min workload per patient. In view of the complexity of planning process in Intensity-Modulated Radiotherapy (IMRT), usually involving the ra-

diation oncologist and medical physicist for more than an hour, this may deem acceptable, if substantial sparing of tissue irradiation in non-effected lymph node sites and dose-escalation in clearly identified nodal metastases due to on-site ultrasound visualization is to be the gain. Still, from the clinician´s point of view further reduction of personal involvement is desirable. As a next step, the manual selection of US-frames should be replaced by automation (e.g. every 20th frame of a sweep). In fact, in this study we selected frames not only for being less distorted than others, but also for being helpful to establish the ground truth validation. If image-guided registration is possible on frames selected automatically as well, preceded by a stable skin surface registration, steps 4 and 5 will become dispensable as well, cutting the manual involvement time down to below 8 minutes, which should be the aim of coming research.
Altogether, we found a Target Registration Error of $3-5mm$ for all lymph nodes other than bulky disease, which is in accordance with the literature for the BAT-system compared to gold marker verification of the prostate [89]. In view of the internal movements and tissue deformation mentioned above this seems to be acceptable. Studies on positioning head and neck patients with a thermoplastic mask reported on a daily setup variation of around $3mm$, but the reported measurements usually relied on osseous landmarks in portal films. Translating the traditional three-point laser setup error of mean $3.33mm$ in any single direction into a mean composite vector, deriving from a high-precision, optically guided patient localization system and considering all six degrees of freedom (like in our TRE), the offset was $6.97mm$ with a standard deviation of $3.63mm$ [61]. In fact, the planning CT alone, imaging the neck tissue at a certain time point, pretends to resemble an accurate picture of a region that is in fact in motion all the time due to pulsating vessels, breathing, swallowing etc. These effects are mirrored in the ultrasound examination, which gives a more realistic imaging along the time scale, e.g. when watching the pulsation without moving the probe. Thus, part of the registration error cannot be attributed to pressure effects from the ultrasound probe but rather relies on inherent tissue movement. This will not be overcome by optimizing positioning and should be allowed for when defining the margins (additional zones around GTV) for planning target volume definition. There is a growing interest on this issue of soft tissue movement, as the widths of margins determine to what extent dose escalation and normal tissue sparing is possible in high-precision radiotherapy. The magnitude of TRE values derived from CT-US registration may help to define the minimal margin to ensure full dose coverage in macroscopic tumor surrounded by soft tissue.
From a clinician's point of view, $3-5mm$ uncertainty about the exact site of each image voxel seems acceptable for radiotherapy planning purposes in soft tissue. Given the tracking system is set up and calibrated, the expected additional workload for the physician amounts to 15 min per patient with adequate fixation. This might be further halved by automation of frame selection for image-based registration, which should be addressed in studies to come. In principle, ultrasonography findings have been made available for treatment planning, where benefits for target volume definition might be expected.

5.2. Abdominal Cancer

5.2.1. Clinical Context

Liver cancer is one of the leading causes of death world-wide, Hepatocellular Carcinoma (HCC) being one of the most common primary cancers. Regarding metastatic disease, col-

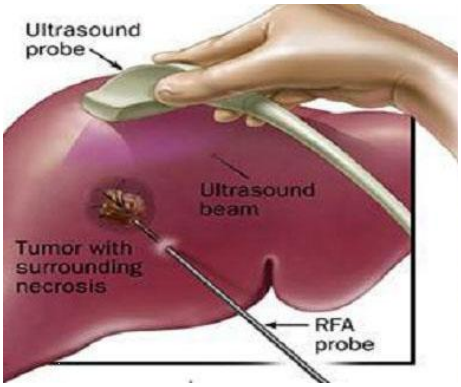


Figure 5.7.: Radiofrequency ablation of the liver guided by ultrasound.

orectal cancer is the most common malignancy in the liver. Despite all the recent advances in cancer therapy, treatment of primary and metastatic tumors of the liver remains a challenge, with often very small survival rates.
Potentially curative therapy options are surgical resection and liver transplantation, however the majority of patients is not eligible for these treatments. There is nowadays a large focus on ablative procedures for the treatment of unresectable liver tumors [28]. They use image-guided placement of an electrode within the target area in the liver parenchyma (figure 5.7). In Radiofrequency Ablation (RFA), heat created around the electrode affects the surrounding tumor tissue, causing coagulative necrosis between 50° and 100°. Often local tissue ablation is performed with lower morbidity than surgical resection, besides it can be performed as a minimally invasive procedure, including percutaneously and laparascopically. RFA is the most common ablative procedure, however other techniques are in use as well, such as cryoablation and ethanol injection.
Successful RF-ablation requires the precise placement of the end-effector tip, typically within the volumetric center of the tumor in order to achieve adequate destruction. The tumor itself and a $\sim 1cm$ margin of surrounding normal parenchyma can then be ablated. The tumor(s) are identified in pre-operative imaging, primarily contrasted CT and MRI. For percutaneous (i.e. through the skin) ablation, often ultrasonography is used as the guidance modality. While it provides excellent visualization of tumors, in particular in combination with ultrasonic contrast agents, its inherently two-dimensional nature and strong dependency on the sonographer's skills still limit its effectiveness [173]. Liver deformation due to needle placement, ultrasound probe pressure, respiratory and cardiac motion poses additional problems.

One can clearly see that the use of tracked ultrasound, alongside with multi-modal registration of ultrasound to the pre-operative CT data, bears immense potential of improving RF-ablation treatments. Many of the uncertainties regarding intra-operative tumor localization and delivery can be removed by precisely mapping the pre-operative and planning data into the coordinate system of tracked ultrasound during the actual procedure.

Even before an actual procedure, multimodal fusion can benefit the diagnostic process. For assessment of indeterminate lesions, usually a contrasted CT scan is used in the first place.

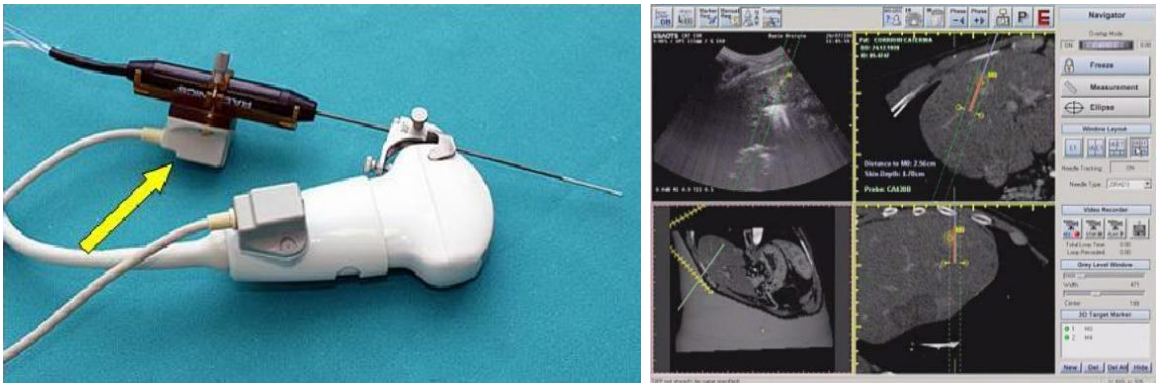


Figure 5.8.: Tracked transducer & electrode (yellow arrow), and software user interface of the Virtual Navigator system, from [141].

Optionally, an additional ultrasound exam is performed. If the malignancy can still not be determined with sufficient confidence, a biopsy has to be performed, where a small needle is inserted into the patient to obtain a tissue sample for further pathological examination. In this context, the fusion of CT and ultrasound can improve the diagnostic value to an extend beyond the "sum" of the individual modalities, potentially sparing an invasive biopsy.

5.2.2. Related Systems

A company named *Ultraguide* (est. 1996, Haifa, Israel) had offered a system that used magnetically tracked ultrasound along with an additional tracking sensor on the electrode to guide ablation procedures [145]. It seems that this system was not well perceived among interventional radiologists. A number of studies could not provide strong evidence of the clinical benefit of it, in particular in comparison to procedures with a needle guide affixed to the ultrasound transducer [132]. The company went into bancruptcy in 2003.

The ultrasound vendor Biosound Esaote (Genoa, Italy)[6] features a solution denoted *Virtual Navigator*, integrated on some of its products, the software of which is developed as *NaviSuite* by MedCom (Darmstadt, Germany)[7]. It uses magnetic tracking as well, of both the ultrasound transducer and, optionally, one ablation electrode (figure 5.8). The pre-operative CT data is integrated before the procedure using manual registration, two approaches are available:
For skin marker registration, 6-9 CT-visible markers are attached to the patient before CT acquisition. The CT scan is preferably done in full inhale, if there is a choice. If a previously recorded CT (e.g. from another hospital) is used, the patient is asked to redo the same state of inhalation for registration. The skin markers are then pointed at with a tracked and calibrated tool just before the procedure.
The second manual registration option is to select the umbilical plane with the ultrasound probe (it shows a shadow in the center of the image), and then manually navigate to the

[6] Website: `http://www.biosound.com`
[7] Website: `http://www.medcom-online.de`

Figure 5.9.: Magnetic tracking sensor hot-glued onto the ultrasound probe.

corresponding plane in CT. This seems to be sufficient for registration in some cases [141]. If not, 3-5 internal landmarks (entrance of portal vein into liver, liver margin etc.) are located in both modalities, to allow a more precise point-based registration. At any point before or even during the intervention, the registration can be manually refined - the physician freezes an ultrasound image, and can then move the ultrasound probe which alters the registration pose, until the CT MPR is aligned. In particular if contrast-enhanced ultrasound is used, opportunities arise to confirm the registration directly on the target lesion.
According to [141], the clinical indications of using this fusion system are lesions difficult to detect solely with ultrasound, large lesions only partially visible in ultrasound, lesions requiring multiple insertions, lesions adjacent to anatomical structures at risk, as well as re-treatment of previously partially ablated tumors.

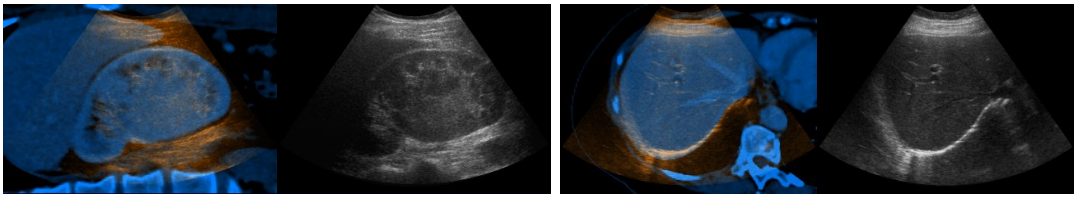
Traxtal Inc. (Toronto, Canada)[8] recently released the PercuNav system for Image-Guided Needle Interventions. It uses miniature electromagnetic tracking sensors integrated in the tip of ablation and biopsy needles, providing accurate navigation even for flexible instruments. Integration of pre-operative imaging and intra-operative tracked ultrasound is featured as well.

The proposed automatic registration techniques could directly add value to such multi-modal fusion systems, and improve the clinical workflow and reliability.

5.2.3. Experiments

In order to evaluate the performance of the novel registration algorithm proposed in section 4.3, we performed a study on abdominal data of 10 patients with various pathology. Since that project is ultimately targeted towards the interventional application, the decision was made to use a magnetic tracking system. The small tracking sensor can be attached right to the ultrasound transducer (optionally inside a sterile plastic wrap), and does not require a line of sight to the transmitter. This is advantageous in the crammed RFA setup (see also figure 1.6), furthermore tracking of one or more of the ablation electrodes is possible as well. Our freehand ultrasound system uses an Ascension MicroBird tracking system with a Siemens Sequoia ultrasound machine and progressive RGBS video fed into a PC with frame grabber.

[8] Website: `http://www.traxtal.com`



(a) kidney of patient 5, rigid (b) liver of patient 6, affine

Figure 5.10.: Result of the automatic registration on kidney and liver images.

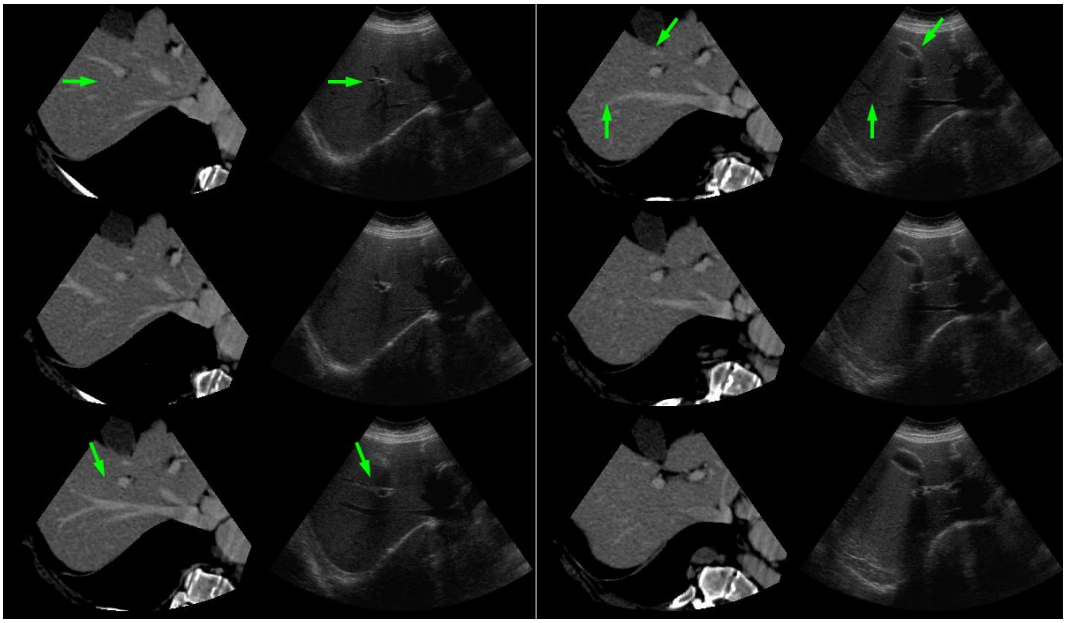
Patient	no. points	manual	pt-based	rigid	affine	remarks
1	8	13.8	9.0	17.0	11.4	strong compresson at top
2	7	16.8	10.0	14.4	8.5	
3	11	10.6	8.9	12.0	11.2	10cm renal tumor
5	5	10.0	8.4	15.5	8.7	kidney
6	7	8.0	6.2	10.7	9.9	
7	11	9.1	6.5	10.8	9.3	pt-based reg. visually bad
9	15	4.2	3.5	7.6	6.8	rigid and affine reg. excellent
11	8	11.1	5.6	8.2	8.2	
13	5	11.6	10.7	13.4	12.3	
14	13	6.6	5.4	7.8	8.0	

Table 5.3.: Registration results on 10 patient data sets in terms of the Fiducial Registration Error (FRE) as root mean square (RMS) values in mm.

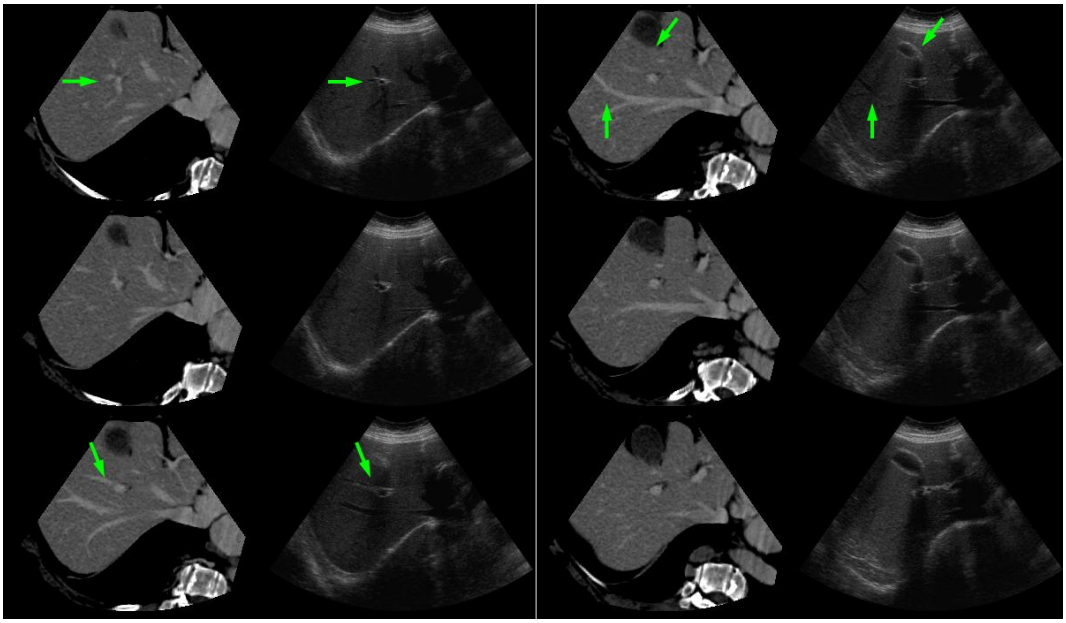
The position sensor was affixed to the transducer using hot-melt adhesive (figure 5.9), a method based on [124] was used to determine the calibration. Transversal liver sweeps on 9 patients, and one kidney sweep from the 10th patient were used, acquired during breath-hold. They were co-registered with portal-venous phase CTA scans, acquired on a dual-source Siemens Somatom Definition scanner before, however not necessarily within the same position in the breathing cycle.

5.2.4. Results

After manually aligning each of the data sets, a physician selected 5-15 point correspondences on anatomical landmarks, including portal & hepativ vein, biliary duct, aorta vena cava and heart atrium. Table 5.3 lists the RMS distances after manual alignment, point-based rigid registration according to [160], and rigid & semi-affine registration using our methods. The automatic registration converges correctly for all patients with an execution time of ~ 20 seconds. At the initial estimate (before the translation search), the FRE was between $11-62mm$. The errors after automatic alignment are in the same range of the manual ones, but larger than the residual errors after point-based registration. Since all of the registrations seem visually correct (exemplary results for liver & kidney are depicted in figure 5.10), we assume to have a fairly large uncertainty in the definition of point correspondences, especially in cranio-caudal direction. This confirms that manual CT-ultrasound registration is error-prone, as it usually reduces the problem to definition of points on 2D-planes, or manually aligning a single 2D



(a) rigid



(b) affine

Figure 5.11.: Comparison of rigid and affine registration results.

plane (as in use in existing products for interventional CT-US navigation) - not guaranteeing a correct matching in 3D. This confirms again, that an automatic method which intrinsically takes full 3D image information into account, is beneficial.
Affine registration with the three parameters of the sagittal plane in general yielded better registration results, accounting for the majority of errors caused by probe pressure, breathing

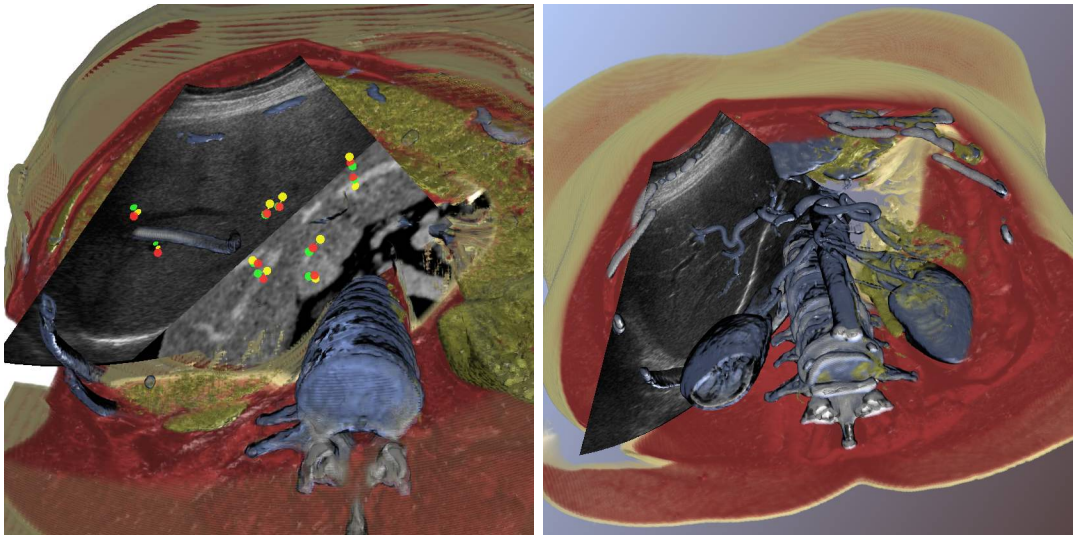


Figure 5.12.: Left: Registered liver of patient 9 with fiducial points (yellow=CT, green=initial US, red=registered US), an oblique CT plane and contextual cutaway volume rendering of CT. Right: Another example of the importance-driven visualization, transversal ultrasound image of the liver with early-arterial phase CTA.

and different patient setup. As can be seen in figure 5.11, displacements on top of the images were further reduced, in particular large shifts of the gall bladder were decreased. In fact, the rigid-body assumption might be valid for a sub-volume of hepatic vasculature, but certainly not for the overall shape of the liver. Therefore the rigid registration usually finds a local optimum where the liver surface lines up best, not achieving a precise alignment of the fine vasculature. The semi-affine model was more often able to match up both the organ surface and the vasculature (figure 5.11).
All fiducial registration error (FRE) values are below $2cm$. We expect that they represent an upper bound for a target registration error (TRE) on liver lesions, which we did not define in the scope of this study due to difficult locatability of relevant clinical targets in most of the data sets.

Regarding the diagnostic value of the study, reading of the registered CT/US data could exclude a number of suspicions on a total of five patients, including partial portal vein thrombosis, acute inflammation of the gall bladder and infiltration of renal cancer into liver tissue.

5.2.5. Visualization

It is desirable to visualize the pre-operative CT together with real-time ultrasound imaging for guiding ablation procedures. However, it is difficult to provide a fused visualization that allows sufficient spatial perception of the anatomy of interest, as derived from the rich pre-operative scan, while not occluding the real-time image displayed embedded within the volume. In [24], we have proposed an importance-driven approach that presents the embedded data such that it is clearly visible along with its spatial relation to the surrounding volumetric material. To

allow this, novel techniques for importance specification, feature emphasis, and contextual cutaway generation have been developed. Essentially, an importance function is defined in addition to the standard one- or two-dimensional transfer function for volume rendering (the latter maps CT intensity and gradient magnitude on red/green/blue color components and opacity). Contextual cutaways then specify the regions of the volume that should only show and emphasize vasculature in order not to occlude the live ultrasound image. A more opaque representation of the anatomy is exposed in the surrounding area, in order maintain the spatial context of the ultrasound image. Results of this visualization approach are depicted in figure 5.12.

5.2.6. Discussion

We have initially evaluated a system for fully automatic alignment of a single freehand ultrasound sweep with CT and CTA data, based on the algorithm developed in section 4.3. We expect this to greatly increase the acceptance of multimodal fusion for diagnosis and treatment, since it provides a simple workflow and enables more precise registration.

Currently, a further study is conducted with a 3D freehand ultrasound system permanently installed at a clinical site. For this system, we used the spatial calibration method presented in section 3.4.1 to prepare a number of transducers for freehand acquisitions. Based on the problems pointed out above, we put particular focus on deriving appropriate ground truth data. For liver sweeps, we visualize both original ultrasound frames and an arbitrary number of cross-sections, compounded using the direct MPR technique from section 3.3.1, each with the respective CT plane (similar to figure 1.5). Using a linked pointer and superimposition options, physicians precisely locate vessel bifurcations, defining fiducial landmarks truly in 3D.
The developed similarity metric LC^2 can easily be extended to handle a larger number of signals from both modalities. We are particularly interested in registering CT with a combination of ultrasound B-mode and contrast (CPS) imaging, since a dual imaging mode is commonly used for ultrasound-guided procedures. Further issues to be adressed are real-time compensation of respiratory motion, as well as deformable mapping techniques.

5.3. Intracardiac Ultrasound for Electrophysiology

The following studies have been mainly executed by John, Sun et al. [65]. I was involved in the development of the image-based registration technique, however it is not related to the novel methods introduced in chapter 4. Since it describes a further very interesting application, where CT/Ultrasound fusion can provide navigation for a very difficult clinical intervention, we present it here as well.

5.3.1. Clinical Context

Atrial fibrillation is the most common heart arrhythmia and the major cause of stroke. Over 2 million people are affected in the U.S. alone. Atrial fibrillation can be treated using a pulmonary vein isolation procedure. Here the pulmonary veins are electrically isolated from the left atrium by a radio frequency ablation catheter, with image guidance provided today mainly by x-ray fluoroscopy. Only highly trained electrophysiologists are able to perform the

procedure, because of the complex and patient specific anatomy of the left atrium. Cardiac CT and MR can deliver high resolution 3D images of the individual heart anatomy. In the future, this pre-interventional imaging could be replaced by a rotational C-arm technique, that is able to produce such images immediately before or during the intervention [79]. An imaging modality that is already used today in many EP labs is Intracardiac Echo (ICE) - a steerable catheter that contains an ultrasound transducer in its tip. By placing the catheter tip into the right atrium the physician is able to image the whole left atrium and some neighboring structures in real time. Therefore it has become an excellent tool for visualization of anatomical structures and instruments, and to monitor critical events.
The combination of 3D cardiac rotational C-arm imaging with ICE and a future integration in the EP lab could help the electrophysiologists to guide ICE and the ablation catheter. It could improve the learning curve for the use of ICE and therefore the whole Pulmonary vein isolation procedure.
Also for electrophysiological applications ultrasound was fused with CT data. In [180] a registration between conventional cardiac CT data and Intracardiac Echo is described. A point-to-surface registration, by first extracting surface point sets of the left atrium from the ICE images, is used. To the best of our knowledge, our results are the first on fusing the two image modalities Intracardiac Echo and cardiac C-arm CT.

5.3.2. Experiments

System Setup

In our system 3D images are acquired on a Angiographic C-arm system (AXIOM Artis, Siemens Medical Solutions). To image the left atrium of a patient we acquire images during 4 consecutive rotational 190° C-arm runs. Therefore we get enough images to reconstruct a 3D image of one cardiac phase. The images are reconstructed and processed on a PC workstation. The left atrium and other heart structures can be segmented using dedicated software.
The images acquired by the ICE catheter (AcuNav, Siemens Medical Solutions) are transferred via a frame grabber card into the PC. To track the position of the ICE catheter tip we used a magnetic tracking system (Microbird, Ascension). Its position sensor has been integrated in the same tubing with the ultrasound transducer. The transmitter is installed under the table of the C-arm system, such that an ICE catheter above the table can be tracked during an intervention.
During ICE imaging we record the ECG signal of the patient, and track the position of the ICE position sensor and a position sensor at the patient's chest (for correcting respiratory motion) synchronously.

Acquired Animal Data

In an animal experiment a 3D cardiac C-arm CT data set from a pig was taken. We additionally took various Intracardiac Echo image sequences from the pig's heart. The catheter was inserted by the physician into the right atrium of the pig. All image sequences were taken from this position by rotating and slightly moving the catheter tip. As mentioned above we also recorded the ECG signal and the coordinates of the two position sensors at catheter tip and chest. Based on these images we performed the following preprocessing and registration procedure offline.

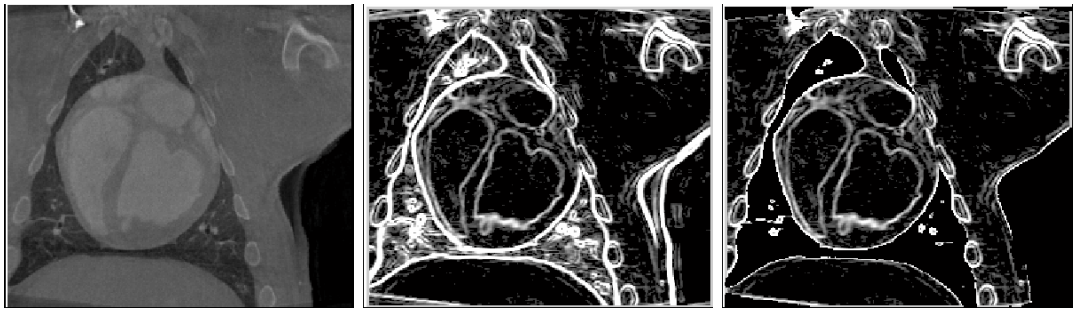


Figure 5.13.: Preprocessing of a pig heart cardiac C-arm CT image: original image (left), gradient magnitude image (middle), and after applying a mask with a threshold of 67 (right). From [65]

Motion Gating for ICE Images

The main difficulty of a good registration is the cardiac motion from the beating heart and the motion due to respiration in the ICE images. Therefore we need a good preselection of useful ICE images.
If we ignore patient movement we can observe the cyclic respiration motion in the graph of the vertical dimension of the 3D position sensor at the chest. This graph has regular peaks separated by long plateaus. The plateaus are of nearly constant height. We compute for every image the variation of this image and its previous images for a fixed time frame. For images with a low variation we can assume that these images were taken in a respiratory phase corresponding to a plateau. So we select those 'low variation' images from the whole sequence.
To compensate for cardiac motion we further select those images with a fixed time distance to the previous R-wave in the ECG signal.

Registration

For an initial registration the user has to select a point in the 3D data set that is close to the position of the catheter tip of the ultrasound image sequence. This gives an initial translation. For an initial rotation we assume that the tracking device is installed under the C-arm table in a fixed and given direction. This initial registration is followed by an automatic local optimization step to find a good rigid body transformation. First we preprocess the C-arm CT data. We extract the magnitudes of the local gradients by applying a 3D Sobel filter [49]. Because there are many regions outside the heart with strong gradients that can worsen the registration quality, we apply a mask based on the grey values of the original volume to focus on the edges representing the heart walls (see fig. 5.13). The ultrasound images are downsized by a factor of 4 in each dimension to improve the runtime. We optimize the transformation according to the following similarity measure: We re-slice the C-arm CT gradient magnitude data in the ultrasound image planes using tri-linear interpolation. Now the similarity measure is computed using Normalized Cross-Correlation (see section 2.2.3) on the sequence of gated US images and resliced planes from the C-arm CT gradient magnitude volume.
The optimization is done using a best neighbor method. For an initial step size all trans-

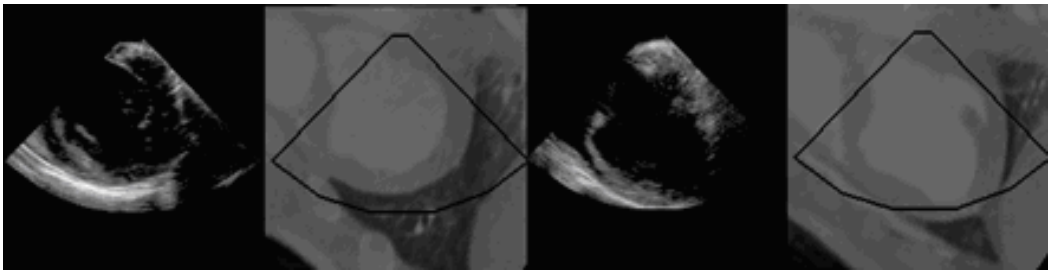


Figure 5.14.: Registration using a sequence of 12 gated ICE images. From [65]

formation parameters are changed in turn with the step size and the resulting value of the similarity measure is computed. The change with the best improvement is taken and the step is repeated until there is no further improvement. Then the step size is decreased and the whole procedure starts again. We use an initial step size of 5.5 mm and 5.5 degrees. Both are reduced by a factor of 2 repeatedly until we reach a step size of 0.1 mm or 0.1 degree.

5.3.3. Results

The registration was done offline with the cardiac C-arm CT and ICE data obtained from the animal experiment. A visual comparison of the registration can be obtained by aligning the ICE images side by side with their corresponding cardiac C-arm CT cut planes (see fig. 5.14).
For a quantitative validation of the registration results we compared segmentations of cardiac chambers. For the 3D segmentation of the C-arm CT data set we used a semi-automatic tool developed for cardiac CT data. The segmentation of the ICE data was done manually by an expert.
We generated registrations and segmentations of 29 pairs of ICE images and their corresponding C-arm CT cut planes. For visual assessment we compared the contour of the ultrasound segmentation and the registered C-arm CT segmentation contour (see fig. 5.15). For quantitative assessment we computed the shortest distance from each contour pixel of the C-arm CT segmentation to the registered ultrasound contour. The mean error was 3.14 ± 3.13 mm. The whole registration procedure implemented in C++ took less than a minute on a system with an Intel P4 processor with 2.8 GHz and 2 GB DDR memory.

5.3.4. Discussion

The results show a good alignment of ICE images and their registered cardiac C-arm CT planes. Nevertheless, the quantitative analysis of registration accuracy shows some variation. The question is, whether this is related to the registration algorithm. A non-negligible additional source of registration and fusion errors that we currently ignore might be patient movement. Furthermore the registration algorithm is based on grey values, whereas the validation of these results is based on segmentations. Some segmented anatomical details and their contour lines might be different in both image modalities. In particular, a precise manual segmentation of the ultrasound images is difficult to achieve.
In the future our proposed methods can be used to build a system that makes it easy to integrate and fuse 3D cardiac rotational C-arm imaging and Intracardiac Echo in the EP

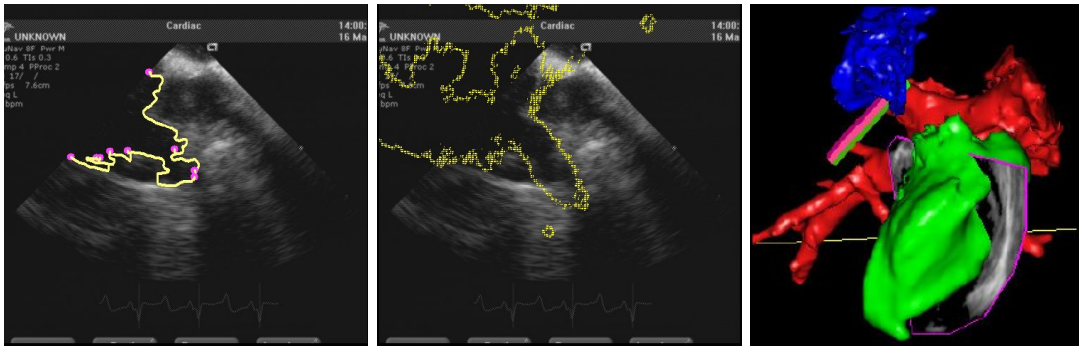


Figure 5.15.: Contour of a cardiac chamber manually segmented from ultrasound (left). The contour of the same chamber segmented from the cardiac C-arm CT image and projected to the registered ultrasound image (middle). A fusion of an ICE image with segmented cardiac chambers from the 3D cardiac C-arm CT image (right). From [65]

suite. Both image modalities will be available during the EP procedure, in contrast with MR and CT which are acquired pre-intervention. A further step could be the fusion with Electro-anatomical mapping data. We believe that these image integrations make it much easier for physicians to learn and perform complex EP ablation procedures.

6. Conclusion

6.1. Summary

The main problem adressed in this thesis is automatic image-based registration of CT and 3D freehand ultrasound imaging of the same patient. The dynamic nature of ultrasonography and the very different representation of anatomy in the two modalities make this an extremely difficult problem. Hence no complete solution, but rather only initial ideas can be found in the literature to date. In order to tackle it, profound knowledge of the state of the art is a prerequesite. Therefore, we provided a comprehensive introduction to current image-based registration and its mathematical and technical foundation in chapter 2, with particular emphasis on image similarity measures, the key component to multi-modal registration algorithms.

The system basics and technical issues of 3D freehand ultrasound have been described in chapter 3. When setting up different hardware and developing software for 3D freehand ultrasound acquisition, we came along some novel ideas regarding calibration and compounding, they are described in this chapter as well.

Our new CT-Ultrasound registration methods are proposed in chapter 4, which constitutes the core contribution of this dissertation. In section 4.2, we first developed a method based on an adapted similarity measure, which considers the most important ultrasound imaging effects that can be recovered to some extent from CT. A combination of a weighted Mutual Information term, edge correlation, clamping to the skin surface and occlusion detection is able to assess the alignment of structures in ultrasound images and information reconstructed from the CT data.
While those methods achieve a good local alignment, they also require manual selection of adequate ultrasound frames, and the adjustment of a number of threshold values and other algorithm parameters. Based on these limitations, we tried to come up with an improved, more integrated registration approach. In section 4.3, we discovered that the transmission and reflection of ultrasonic pulses can be simulated from CT, based on the known geometry of the transducer. This allows to reproduce the imaging of tissue interfaces and shadowing & occlusion effects. Combined with an approximated intensity mapping of the CT hounsfield units to ultrasonic echogeneity, we have achieved a simulation of ultrasound from CT with higher accuracy than we expected. Instead of weighting the two effects reproduced from CT, we developed a novel similarity metric that assesses the linear correlation of their combination with ultrasound, regardless of their individual influence. In fact this equals a simultaneous optimization of simulation and registration parameters, which is applicable for many other multi-modal registration problems. Put together with an automatic frame selection strategy and a global initial transformation parameter search, it resulted in a fully automatic algorithm for CT-ultrasound registration.

The clinical applications have been presented in chapter 5. In section 5.1, we developed a clinical workflow for integrating diagnostic ultrasound of the head and neck with CT for improved radiation treatment planning (section 5.1). A global initial registration is achieved by matching the skin surface derived from both CT and all freehand ultrasound sweeps. Successively, the first semi-automatic registration algorithm from section 4.2 is applied. In a study on five patients with head and neck tumors and cervical lymph node metastases, we found a mean target registration error of $3.9mm$. In a detailed discussion, we described the individual error sources, and pointed out that there is significant overall clinical benefit, despite the additional time required for the 3D ultrasound exam and registration.
The second registration algorithm was evaluated in a diagnostic fusion study on 10 patients with hepatic and renal cancer (section 5.2). Fully automatic alignment was achieved on all data sets, we reported a registration error of always $< 2cm$ with respect to a number of fiducial points throughout liver or kidney, for both rigid and affine transformation models. Furthermore, a novel importance-driven visualization, based on contextual cutaway rendering, has been applied to the registration results.

6.2. Discussion and Future Work

6.2.1. Clinical Evaluation

Regarding the Radiation Therapy application, the next steps would be to integrate our proposed fusion workflow into a prototypical planning software, by collaborating closely with a vendor of existing software solutions. A clinical study can then be executed on a large number of patients, to precisely determine the amount of improved planning confidence. Besides, the registration procedure could be further advanced by adapting the second, automatic algorithm from section 4.3 to the head and neck data.
On abdominal imaging of liver and kidney, the developed automatic registration proved to be a valuable tool to achieve alignment of CTA and ultrasound, as opposed to cumbersome manual registration. With respect to diagnostic CTA-Ultrasound fusion, a larger clinical follow-up study is necessary, as it can reveal how much better indeterminate lesions can be assessed, and how the overall diagnostic workflow is changed & improved.
The most challenging clinical application is certainly the fusion for interventional procedures. The performance of our automatic registration techniques has to be evaluated in an online fusion system, to see if it can keep up with its promises in a realistic environment with unpredictable patients and hard time constraints.
A field of further investigation that might benefit all aforementioned applications, is how to cope with anatomic deformations between the CT and ultrasound acquisitions. Providing real-time update and validation on the quality of alignment, will be particularly important for the interventional application. In the following, we would like to highlight some of the technical research topics that should be adressed in the future, based on our presented automatic registration method.

6.2.2. Deformable Registration

In all clinical scenarios described in this thesis, the rigid registration of CT and ultrasound is only an approximation, since the anatomy will always be deformed to some extent. This does not only impair the resulting accuracy of alignment, but can also prevent a robust registration

in the first place, if the data just does not line up within the assumption of rigidity. Before enthusiastically assuming all those problems can be solved by deformable registration, one has to carefully examine the requirements of the clinical scenario. Deforming three-dimensional measurements of patient anatomy is a significant modification, which might not be acceptable in certain settings. Besides, validation of deformable registration results is a very difficult issue which has not been solved satisfactorily to date. This can also pose additional difficulties for approving respective medical systems with regulatory authorities, such as the U.S. Food and Drug Administration (FDA).
Nevertheless, deformable registration can and should be used to more precisely register CT and ultrasound imaging. In most cases, there is a certain target area (usually lesions) within the images where the highest accuracy of registration is desired. Given that the target site contains enough anatomical structures, a simple improvement is possible by weighting the similarity measure contributions with their proximity to the target. Going a step further, a good deformable registration of CT and ultrasound can be used to obtain a precise rigid alignment of the target subvolume.
We have started heading in the direction of deformable registration by executing our registration techniques with an affine transformation model, that can compensate large-scale liver motion. Our new LC^2 similarity measure (section 4.3.2) can be computed locally, an efficient implementation would use recursive filters. This will allow to use it as force in a dense-field deformable registration algorithm. Together with an adapted regularization term, which should adress ultrasound-specific properties like larger deformation close to the skin due to probe pressure, an automatic image-based deformable registration of CT and ultrasound will be feasible.

6.2.3. Real-time Update

In the interventional scenario, a single tracked ultrasound sweep is used to register CT into the tracking coordinate system. Henceforth, the alignment is only correct if there is no patient shift, anatomic deformation, accidental movement of the tracking reference, etc. It would therefore be desirable to have real-time feedback about the correctness of the registration. Due to the challenges of the multi-modal registration scenario, it is difficult do register a single live ultrasound image with CT. We propose to take a detour instead, computing a real-time update of the alignment by registering the live ultrasound image with the 3D ultrasound sweep recorded beforehand. The latter has been registered to CT, and therefore allows to connect the actual live frame with CT. While computing those incremental registration updates, one should decide based on the image content, if the current frame provides a sufficient amount of structures to yield a relevant update of the alignment. Besides, it has to be studied which transformation models are suited here. While a single live image alone would not provide sufficient information to compute a deformable update of the alignment, rigid or local affine models might make sense.
Such a real-time update will be computationally demanding. It can be regarded as repeated slice-to-volume registration of the live ultrasound frame with a 3D-ultrasound sweep that has been compounded into a cartesian grid. This is perfectly suited for GPU acceleration, as slices can be extracted from 3D-textures with just insane speed. Combined with predictive filtering of the incremental transformation changes, it would certainly be possible to assess, and optionally correct, the multi-modal alignment in real-time during an intervention.

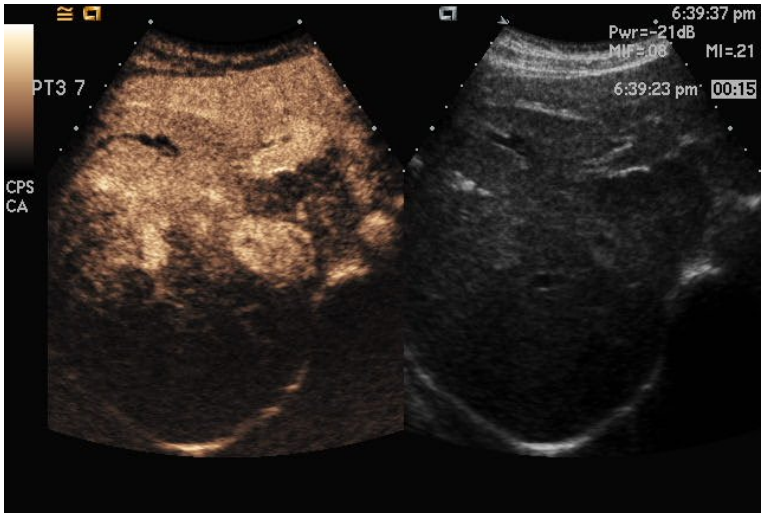


Figure 6.1.: Simultaneous CPS contrast (left) and B-mode imaging on the Sequoia system.

6.2.4. Motion Management

In particular for abdominal and cardiac interventions, the single most significant source of mis-alignment is periodic respiratory and cardiac motion. If four-dimensional pre-operative imaging is available, a patient-specific motion model can be established. The actual intra-operative imaging can then be used to measure the motion and correlate it with the model [55]. That way, even deformable updates can be achieved based on sparse live ultrasound imaging [14], because the image-to-model alignment dramatically reduces the number of required registration parameters. While four-dimensional imaging might not be available on a regular basis (due to high radiation exposure for 4D-CT, and high costs & long duration for 4D-MRI), alternatives that provide approximate motion models can be used as well. These include organ-specific statistical shape models (SSM) and atlas-based models. We have achieved promising registration results in section 5.2.4 using a sub-set of the affine transformation parameters, as it has been shown before that they sufficiently represent large-scale abdominal respiratory motion [121]. This can actually be considered as an initial, very coarse motion model.
We have, and continue to acquire a large number of 3D freehand ultrasound exams with an additional tracking sensor attached to the patient's stomach. Together with aforementioned techniques for real-time update, both intra- and pre-operative imaging can be used to gradually refine a model of the patient-specific motion, starting from a patient-independent prior motion model of the respective organ.

6.2.5. Other Imaging Channels

We have used the novel registration framework based on the LC^2 similarity measure, to register two information channels (large-scale reflections and echogeneity) from CT with the ultrasound image intensity. However, it can be extended to assess the alignment of any number of imaging signals. In ultrasound-guided RF-ablation procedures, contrast-enhanced ultrasound (CEUS) is nowadays the preferred imaging technique, since it can visualize tu-

mors with superior distinction and highlight fine vasculature [142]. The Siemens Sequoia ultrasound platform features a split-screen mode, simultaneously showing Contrast Pulse Sequencing (CPS) and regular B-mode images (figure 6.1). We intend to use both of those ultrasound information channels to achieve an improved registration, especially within the liver. A further extension would be to integrate multiple CTA contrast phases, since fusion of multi-phase CT is a common trend in liver perfusion imaging [98].
New ultrasonic imaging modes, the most prominent one being elasticity imaging, bear great potential for improving both diagnostics and ultrasound-guided interventions. Caused by the large academic interest in this field, ultrasound vendors have started to provide research interfaces that allow to customize the beam-forming and pulse sequencing, as well as access raw radiofrequency data of the returned echoes. Along with the shift away from proprietary digital signal processing (DSP) boards towards using powerful PC hardware for the signal processing, it is now easier than ever to modify the internals of ultrasound systems. In this context, not only elasticity imaging could provide an extra channel of information for CT-ultrasound registration. One could furthermore design specific ultrasound imaging modes that carry CT-like characteristics, e.g. using statistical classification on the raw echoes. This would also circumvent the main problem of ultrasound texture classification approaches operating on the final images, namely the dependency on many of the parameters that are adjustable by the sonographer.

Its unique flexibility, versatility and real-time capabilities have always distinguished ultrasound from the other modalities. Enhancing it with 3D position sensing turns it into a full-fledged three-dimensional imaging modality in addition. Based on the achievements in this thesis regarding multi-modal registration and fusion, as well as the proposed future research, seamless integration with other pre-operative imaging, as well as interventional navigation, will be easier than ever before. We believe that this will further strengthen the position of medical ultrasound as a premier, integrated imaging modality for diagnosis and treatment, especially in the areas of radiation therapy and interventional oncology.

A. 2D/3D Registration Based on Volume Gradients

We present a set of new methods for efficient and precise registration of any X-Ray modality (fluoroscopy, portal imaging or regular X-Ray imaging) to a CT data set. These methods require neither feature extraction nor 2D or 3D segmentation. Our main contribution is to directly perform the computations on the gradient vector volume of the CT data, which has several advantages. It can increase the precision of the registration as it assesses mainly the alignment of intensity edges in both CT and X-Ray images. By using only significant areas of the gradient vector volume, the amount of information needed in each registration step can be reduced up to a factor of 10. This both speeds up the registration process and allows for using the CT data with full precision, e.g. 512^3 voxels. We introduce a *Volume Gradient Rendering (VGR)* as well as a *Volume Gradient Correlation (VGC)* method, where the latter one can be used directly for computing the image similarity without DRR generation. This work was published in [168].

A.1. Introduction

2D-3D Registration has numerous applications in computer-aided diagnosis and therapy, including intraoperative navigation with fluoroscopy, patient positioning for radiation therapy and multimodal data fusion for diagnosis and therapy planning. These applications usually involve a two-dimensional X-Ray projection image and a preoperative CT data set. The task is to define a common coordinate frame, in a way that corresponding structures of the patient's anatomy are properly aligned in both data sets. The general workflow of a 2D-3D registration algorithm is to simulate X-Ray images by computing two-dimensional projections of the CT volume at an estimate of the pose of the real X-Ray image. This digitally reconstructed radiograph (DRR) is iteratively compared to the real X-Ray image and the DRR pose is altered. Finding a rigid pose which maximizes the similarity of two images is a non-linear optimization problem, where adequate algorithms can be used in order to find the maximum as fast as possible.

A.2. Related Work

For performing an intensity-based 2D/3D registration, numerous DRRs have to be created, each of them requiring processing of the full volumetric data set. Plenty of research has been done on speeding up DRR computation. Contrary to the classical back-projection approach, *Ray Casting*, algorithms like *Shear-Warping* [162, 76], precomputation of DRR rays [78], light field rendering [127] and hardware-accelerated 2D Texture rendering [78] are more efficient in some orders of magnitude, however they all use simplifications which reduce the DRR quality and produce artifacts which may affect the registration accuracy. Using the 3D texturing feature of modern graphics hardware, combined with rendering to floating-point

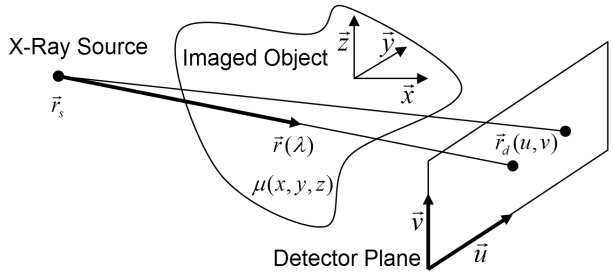


Figure A.1.: Scheme of X-Ray projection imaging

color buffers, it is possible to accumulate slices trilinearly interpolated from a 3D volume. Based on this feature, the 3D texture volume rendering algorithms are today the methods of choice, because they are efficient and produce high-quality DRRs without artifacts. However, their implementation imposes a lot of technical difficulties, and designated software usually works only for a specific graphics board manufacturer.

The DRR and X-Ray images to be compared can be considered as two different modalities, mainly because of different energy spectrums used for the acquisition of CT slices and X-Ray images. This difference can be enormous if Portal Images with MegaVolt energies are acquired [70]. By applying an adequate transfer function on the volume this can be taken care of partly. Nevertheless the images will have different structures, and it is tough to assess the quality of alignment automatically. In this context many *Similarity Measures* have been developed. Some are using just sum-of-squares or correlation of the intensity differences, some are working on two-dimensional gradient images. Information theory approaches assess the alignment based on the properties of histograms and joint histograms, as the very popular *Mutual Information* measure [154], or *Correlation Ratio* [117]. Correlation of small image regions weighted with local variances is another very stable measure [78]. Various researchers compared the performance of 2D/3D registration using different similarity measures [105, 60] and [163].

Our approach yields efficient use of volume gradients, in order to obtain either the two-dimensional DRR image gradients or directly the gradient based similarity measures. The underlying idea of using volume gradient vectors has been described in two other recent papers. Tomazevic et al. [150] backproject the gradients from the X-Ray image into the volume and compute a similarity based on their correspondence with 3D gradient vectors, which are defined on the surface of (previously segmented) bony structures. Livyatan et al. [83] use, as final stage in their registration algorithm, the correspondence of the 2D gradients on the X-Ray image with ray-casted sums of 3D gradient vectors in the volume. These correspondences are only evaluated for border areas in the 2D image, which they define by applying a Canny edge detector algorithm on the X-Ray image. In contrast to those algorithms, we are using the CT gradient information in order to perform a purely intensity-based registration, without prior feature extraction.

A.3. Methods

A.3.1. Foundation for DRR Computation

The pixel values in a X-Ray image originate from the number of particles reaching the detector plane after passing the imaged object. The basic equation is

$$I(u,v) = \int_0^{E_{max}} I_0(E) exp\left(-\int_{r(u,v)} \mu(x,y,z,E)dr\right) dE \tag{A.1}$$

where u, v defines a pixel of the X-Ray detector, $r(u,v)$ is the ray from the source to this pixel, $\mu(x,y,z,E)$ is the attenuation coefficient at a specific position in space and a X-Ray energy E (figure A.1). As X-Ray sources are always polychromatic, the attenuation has to be integrated not only along the ray, but also over the incident energy spectrum $I_0(E)$. Those spectrums unfortunately vary greatly depending on the imaging devices. Therefore attenuation coefficients are being treated in a simplified manner, denoting the attenuation of a monochromatic X-Ray source at an effective energy E_0:

$$I(u,v) = I_0 exp\left(-\int_{r(u,v)} \mu(x,y,z,E_{eff})dr\right) \tag{A.2}$$

Most X-Ray imaging devices measure the logarithm of equation A.2, thus the pixel intensities are reduced to a simple integral term, which we will use for our computations.
A heuristic means to compensate for the different effective energies in CT and X-Ray images is to apply a scaling and truncation of the CT intensities. The CT values are cropped with a user-defined window/level setting, the line integrals are evaluated (e.g. with a ray-casting algorithm) and the resulting pixels are scaled in order to fully cover the gray-scale range of the X-Ray image. This process can be referred to as *radiometric calibration* [70], it has to be done only once for a specific combination of CT and X-Ray imagers. This even allows to compute highly realistic DRRs of MegaVolt Portal Images from CT.

A.3.2. Determining Quality of Alignment

In order to search for the best DRR pose, the DRR and X-Ray images have to be compared iteratively and the quality of alignment has to be assessed by some measure of similarity. *Gradient Correlation* [105] is a very powerful one, its robustness originating from the use of gradient images. It uses *Normalized Cross Correlation*, which expresses the linear dependency between the intensities in the images I_1 and I_2:

$$\begin{aligned} NCC(I_1,I_2) = & \quad \tfrac{1}{\sigma_1\sigma_2}\tfrac{1}{n}\textstyle\sum_{u,v}\left(I_1(u,v)-\overline{I_1}\right)\cdot\left(I_2(u,v)-\overline{I_2}\right) \quad = \\ = & \quad \frac{\sum\left(I_1(u,v)-\overline{I_1}\right)\cdot\left(I_2(u,v)-\overline{I_2}\right)}{\sqrt{\sum(I_1(u,v)-\overline{I_1})^2}\cdot\sqrt{\sum(I_2(u,v)-\overline{I_2})^2}} \end{aligned} \tag{A.3}$$

Gradient Correlation is then the mean of the NCC values for both pairs of the horizontal and vertical gradient images, respectively:

$$GC = \frac{1}{2}\left(NCC\left(\frac{\partial I_1}{\partial u},\frac{\partial I_2}{\partial u}\right) + NCC\left(\frac{\partial I_1}{\partial v},\frac{\partial I_2}{\partial v}\right)\right) \tag{A.4}$$

It reaches its maximum value 1 if both the horizontal and vertical components of the DRR and X-Ray gradients are fully linearly dependent. This assumes a looser dependency on the image content itself [163], which is especially beneficial if the data to be registered has been taken at very different energies, as the is case with Portal Images / CT.

A.3.3. Direct Computation of DRR Gradient

A DRR pixel is the attenuation along the ray originating in the X-Ray source $\vec{r_s}$ and passing through the location of the pixel in the image plane $\vec{r_d}(u,v)$. Each position on this ray can be parameterized:

$$\vec{r}(u,v,\lambda) = \vec{r_s} + \lambda\left(\vec{r_d}(u,v) - \vec{r_s}\right), \quad \lambda \in [0\ldots1] \tag{A.5}$$

The refined equation for the resulting intensity value at the pixel u,v (corresponding to the logarithm of equation A.2) is

$$I(u,v) = \int \mu(\vec{r}(u,v,\lambda)^T)d\lambda \tag{A.6}$$

For computing the similarity measure, equation A.4, we need the partial derivatives of $I(u,v)$ with respect to the pixel locations u and v, i.e. $\frac{\partial I}{\partial u}(u,v)$ and $\frac{\partial I}{\partial v}(u,v)$. Given the three-dimensional gradient $\nabla\mu(x,y,z)$ of the attenuation in space, we can directly compute this information:

$$\frac{\partial I}{\partial u}(u,v) = \left(\int \lambda\nabla\mu(\vec{r}(u,v,\lambda)^T)d\lambda\right)^T \cdot \vec{u} \tag{A.7}$$

$$\frac{\partial I}{\partial v}(u,v) = \left(\int \lambda\nabla\mu(\vec{r}(u,v,\lambda)^T)d\lambda\right)^T \cdot \vec{v} \tag{A.8}$$

The volumetric gradient $\nabla\mu(x,y,z)$ has to be incorporated into the line integral by scaling it with λ. This is due to the fact that the rays belonging to neighbored pixels are further apart from each other proportional to the distance from the X-Ray source (see figure A.1).
The length of the gradient vectors is to a big extent close to zero. It is therefore applicable to consider only gradient vectors of significant size, e.g. 1% of the maximum size. Neglecting all smaller vectors has the side effect of improving the registration quality further, as the important edges from rigid, bony anatomy are emphasized, while too small ones, likely to belong to deformable tissue, are discarded.

Gradient Ray Casting

Equations A.7 and A.8 can be directly evaluated using a ray-casting technique. For each pixel u,v in the DRR gradient image, the line equation A.5 is sampled at constant intervals, while $\nabla\mu(x,y,z)$ is computed at the respective positions by trilinear interpolation from the gradient volume. For comparison purposes, we also implemented equation A.6 in the same way, which produces regular DRRs. It is important to mention that both approaches take the same computation time, if we make use of Single Instruction Multiple Data (SIMD) machine commands, which are available on today's ix86 CPUs. Hence summation and interpolation of a 3-vector takes the same number of operations than using just scalar data. We used the GNU C++ Compiler under Linux, where Intel's SSE2 instructions are supported.
For skipping the huge number of unimportant small gradient vectors, we implemented a binary Octree data structure. For every sampling step along a ray, the octree is queried. The largest bounding cube that is empty (side width $32, 16, \ldots 1$ voxels) is skipped, and the first sampling position on the ray is set, which is outside of this cube. In order to efficiently implement this skipping, we use an algorithm similar to the one originally proposed by Amanatides and Woo [7] for ray tracing purposes. We refer to our rendering method as *Volume Gradient Rendering (VGR)*.

Gradient Splatting

Splatting is a forward-projection method, where the volume is traversed voxel-wise, and a footprint of each projected voxel is created in the 2D image. Thus a great reduction in computation time can be achieved by just considering the gradient voxels larger than a specific magnitude. The footprint is usually a gaussian kernel, its size varying with the distance of the voxels from the image plane. For each voxel, the pixel intensities of all pixels affected by the kernel have to be updated. Unless hardware-accelerated methods are used (e.g. drawing the kernels as OpenGL textures), this method is too slow in order to be superior to ray-casting, even if only a small percentage of the gradient voxels is considered.
Therefore, if the objective is DRR generation, splatting would not be the preferred approach. However, a smart algorithm could take advantage of the splatting formulation to directly use the 3D gradient data for the similarity computation, without DRR generation.
We can leave the use of kernels out, given that the size of a projected voxel is negligible with respect to the pixel size in the 2D image. Each gradient voxel then adds to the intensity of just one pixel:

$$\frac{\partial I}{\partial u}(u,v) = \left(\sum_i \lambda_i \nabla \mu_i\right)^T \cdot \vec{u} \tag{A.9}$$

$$\frac{\partial I}{\partial v}(u,v) = \left(\sum_i \lambda_i \nabla \mu_i\right)^T \cdot \vec{v} \tag{A.10}$$

Let $I_1(u,v) = \frac{\partial I}{\partial u}(u,v)$ be the resulting vertical gradient values, and $I_2(u,v)$ the respective values in the X-Ray image. The Normalized Cross Correlation equation (A.4) can then be rewritten:

$$\frac{1}{\sigma_1\sigma_2}\frac{1}{n}\left(\sum I_1(u,v)I_2(u,v) - \overline{I_1 I_2}\right) \qquad = \tag{A.11}$$

$$\frac{1}{\sigma_1\sigma_2}\frac{1}{n}\left(\sum\sum\left[\left((\lambda_i\nabla\mu_i)^T \cdot \vec{u}\right) I_2(u,v)\right] - \overline{I_1 I_2}\right) \tag{A.12}$$

The expression in squared brackets can be directly evaluated for each voxel, independently of the sum of projected voxels resulting in pixel values. We therefore compute this correlation by multiplying each projected gradient voxel with the intensity $I_2(u,v)$ of the X-Ray image in full resolution, i.e. 1024^2 pixel. This approach directly creates the correlation values, without producing a high-quality DRR. As we do not assemble ray integrals, we do not have access to the mean and standard deviation of the simulated image, though.
This results in a very efficient computation of gradient correlation values without normalization, which uses both all significant gradient voxels directly, without interpolation and kernel functions, and the full content of the X-Ray image. We denote this technique *Volume Gradient Correlation (VGC)*.

A.3.4. Implementation

A prototypical application was developed, which demonstrates the advantages of the described methods. A CT volume is loaded with full precision (12 bit) into memory. The volume intensities are then cropped with a defined window/level setting. Successively the gradient volume $\nabla\mu(x,y,z)$ is computed using a three-dimensional Sobel filter cube, which takes 27 values into account for each voxel, resulting in very smooth gradient information. For the

VGR method, a binary octree is constructed from the gradient volume. For the VGC method, the gradient voxels above the threshold are saved in a run-length encoded structure for fast traversal.
The actual registration is performed by iteratively computing the DRR/X-Ray similarity with one of the described methods. The pose is refined according to a simple best neighbor search strategy, which has proven to be stable and sufficiently fast [163]. At the same time it allows to distribute the computations on multiple CPUs, as (for a 6 DOF optimization problem) 12 cost function evaluations can be done in parallel, before the pose for the next iteration is defined. We implemented a distributed version of the described registration methods using the *Message Passing Interface* (MPI), allowing for transparent distribution on any number of nodes and/or CPUs. Several client processes receive job messages with a pose description, and return a single scalar similarity measure value to the coordinating process. Therefore the communication overhead is very small and the computation speed scales directly with the number of CPUs used.

A.4. Experimental Results

The subsequent registrations were computed on a cluster comprising 3 nodes with each 4 Intel Itanium-2 CPUs running at 1.3 GHz[1]. Accordingly, 12 processors were used simultaneously, resulting in a speedup of 9.7 compared to the original sequential implementation, executed on a Pentium M 1.6 GHz Notebook.

A.4.1. Phantom Data

For assessing the registration robustness and accuracy, we used CT and portal image data acquired from a Rando Body Phantom, including Ground Truth information [163, 70]. The registration was launched iteratively from a pose randomly displaced (up to 20 mm / 20° in each degree of freedom) from the ground truth pose. The following table denotes the mean displacement with the respective standard deviation of the translational and rotational components (in mm and degrees, respectively) of the resulting pose. As we primarily assessed the registration robustness with respect to various pose parameters, we preferred this error description over a target registration error (TRE) value, especially as we were dealing with phantom data. We compared our novel methods to a standard registration using the 2D gradients of a DRR rendered with a regular volume rendering technique (VRT), i.e. software based Ray-Casting, which yields very high accuracy whilst performing relatively slow.

method	Δ_{trans}	Δ_{rot}	σ_{trans}	σ_{rot}	time
VRT	1.31	0.41	0.08	0.09	32 s
VGR	0.79	0.64	0.12	0.28	3.2 s
VGC	1.24	0.48	0.10	0.14	2.5 s

A.4.2. Patient Data

A second set of experiments was performed on real pelvis data of two prostate cancer patients, comprising a pretherapeutic CT scan and scanned simulator X-Ray images for radiotherapeutic purposes. Before the gradient computation the CT data was thresholded in order to

[1] InfiniBand Cluster at TU Munich, http://infiniband.in.tum.de

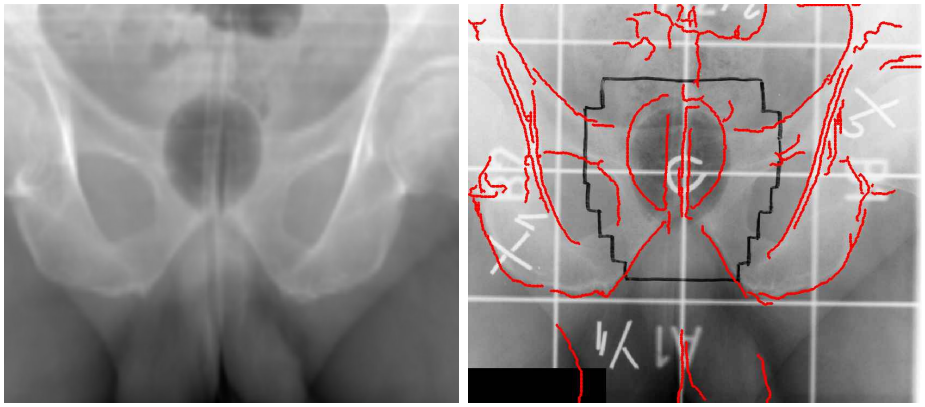


Figure A.2.: DRR image registered with VGR, X-Ray image with edges from DRR

restrain the registration to the pelvis bone, successively a gradient threshold was applied. This resulted in just 1% of the voxels being used for the similarity computation, using our new methods. Ground Truth information was not available, thus we depict the robustness in terms of the standard deviation of the resulting pose parameters, registered from randomly displaced starting poses, as above. The alignment was visually assessed by a physician and declared as optimal for all three methods (figure A.2).

method	σ_x	σ_y	σ_z	σ_{alpha}	σ_{beta}	σ_{gamma}	time
VRT	0.07	0.44	0.05	0.06	0.05	0.08	22.2 s
VGR	0.18	2.73	0.18	0.23	0.16	0.15	1.8 s
VGC	0.15	4.12	0.41	0.47	0.10	0.21	3.1 s

While the standard deviations (in mm and °, resp.) are exceedingly small with regular volume rendering (VRT), the values for our new methods are still well within the range to be considered robust, apart from σ_y. The y values reflect the out-of-plane translation, and therefore are higher due to the small field of view (9°) of the images. This can be compensated by concurrently registering a second, perpendicular X-Ray image. The inferior results of the VGC method probably originates from the missing normalization of the correlation, in conjunction with the relatively small region of interest of these data sets (due to the positioning grid in the X-Ray images). Even though the pelvis was registered correctly, the highly deformable balloon inserted manually into the patients rectum was displaced as expected (figure A.2). Therefore both the thresholding of intensity / gradient voxels and the use of a gradient-based similarity measure was essential for the successful registration, as the balloon and tissue had to be excluded for registration.

A.5. Conclusion

We presented a set of new techniques for performing 2D/3D registration of X-Ray modalities, which are based on the immediate use of gradient information. The volumetric data can be used with both full precision (16 bit intensities) and full size for highly precise registration. By

thresholding the gradient magnitudes, the amount of information used for the registration can be reduced dramatically, delineating only significant (bony or contrasted) structures. These are in fact well adapted to the use of a rigid registration algorithm, as other parts of the patient's anatomy often undergo deformable transformations between the image acquisitions (which is the case in experiment A.4.2). Note moreover that no segmentation is necessary to achieve this. The VGC method computes the gradient correlation values directly from the thresholded gradient voxels and the full-resolution X-Ray image, therefore using all of the available, pertinent image information.
We tested the proposed methods on both phantom (CT / Portal) and patient (CT / X-Ray) data. In terms of robustness, they are slightly inferior to an algorithm comprising DRR generation with highest possible quality (which can be considered the "Gold Standard" of intensity-based registration lacking Ground Truth information), while they are an order of magnitude faster in execution. Using a purely CPU-based registration solution, we were also able to exploit the speed advantage of distributed computations, which resulted in overall registration times of a few seconds on a system with 12 processors. This leads toward new intra-operative and real-time applications of automatic image registration. In addition, it motivates further research on parallel software-based registration solutions, as opposed to GPU-based techniques, which are difficult to distribute.

B. A Volume Fusion Approach to 3D Medical Ultrasound Imaging

The following is a initial technical report which I wrote at Siemens Corporate Research in summer 2005, while investigating the feasibility of combining several 3D ultrasound volumes to improve the imaging quality within a slice or volume of interest. Subsequent work, carried out together with students, resulted in novel approaches for ultrasound mosaicking by multivariate registration [156], and new techniques for ultrasound volume reconstruction [17].

B.1. Introduction

When using medical ultrasound, physicians often want to depict a particular slice of interest, which might be hard to acquire directly, as a structure entailing occlusion effects is in the way. This is for instance the case for transcutaneous ultrasound imaging of the heart, where chest ribs are covering the theoretically optimal scanning position, and either small designated probes have to be used, or other orientations approaching from the abdomen/liver, pointing upwards toward the heart.
Using modern 3D ultrasound devices, a whole volume of ultrasonic reflectivity information can be acquired at once. Successively, the physician can navigate within a 3-slice view, an oblique MPR, or a volume-rendered representation of the data in order to see the desired anatomy.
We propose a mosaicking of an arbitrary number of 3D ultrasound volumes, i.e. one reference volume should be expanded with every new acquisition, while assuring insertion of the new information at the correct spatial position. This overcomes position-dependant artifacts, as volumes can be acquired from any orientation. In particular it can resolve occlusion problems by merging ultrasound information from various angles aside the structure causing the occlusion.
The workflow is as follows: The user first performs a reference acquisition, telling the system what the volume or slice of interest is. Then an arbitrary number of additional volumes are acquired, either in real-time 4D mode, or high resolution 3D. The information is merged with the reference scan, using both intensity-based registration and some tracking system (optical, magnetic or gyroscopic). In terms of the user interface, the required interaction could be very simple and convenient for physicians used to only 2D imaging: The probe is maneuvered to a particular position of interest, which is partly obscured. A button is pressed, the slice is frozen, and any further probe movement creates additional information that enhances the slice, until a satisfying view of the anatomy has been achieved.

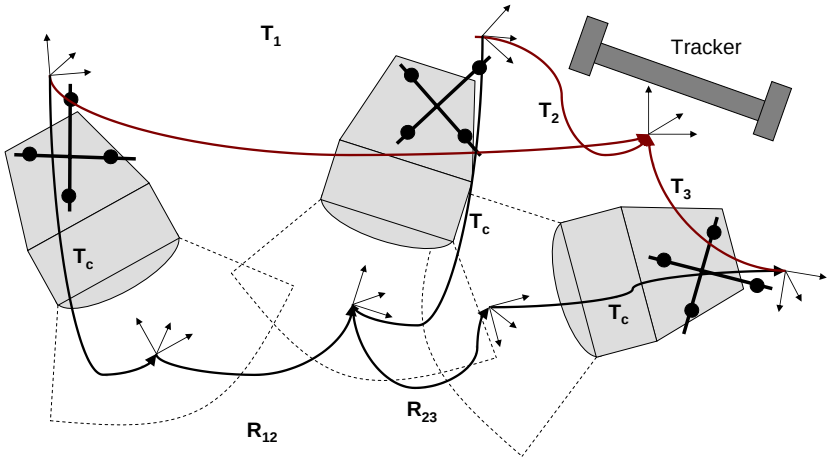


Figure B.1.: Transformations involved in calibration using three volumes.

B.1.1. Potential Applications

- Obstetrics: Part of the baby's shape can be occluded from its own hands or an odd position.
- Monitoring and Treatment of Abdominal Aortic Aneurysms (AAA). The Aorta can be occluded partly due to refraction artifacts from the portal vein, or total reflection at other tissue interfaces. Fusion can be also beneficial to reconstruct 3D-Powerdoppler Information, while the registration can be achieved using the regular B-Scan data.
- Imaging of bony anatomy, for instance spine vertebrae. Scanning from slightly different insonification angles can be used to reconstruct a completed bone surface, as a single snapshot yields incomplete information due to the more specular reflection properties of bone.

B.2. Methods

B.2.1. Calibration

For an efficient workflow, the position and orientation of the ultrasound volumes has to be established in real-time. For now, we focus on the case of using an optical tracking system, which delivers this information accurately with its full 6 Degrees of Freedom (DOF). Therefore, the relative transformation between the tracking target coordinate system, which is reported by the tracking system, and the coordinate system of the ultrasound volume has to be computed.
A related approach for "quality control" of the calibration transformation during a procedure is proposed in [18], where only in-plane transformations from specific sliding motions are used to update and verify the calibration solving an $AX = XB$ type of equation systems as well.
We approach this problem by scanning a phantom object submerged in a water bath, which yields reasonably sharp outlines in the images. Two volumes taken at varying angles, can be

coregistered using intensity-based image registration, which results in their relative transformation R_{12}. Given the tracking transformation matrices T_1 and T_2, respectively, the following equation holds:

$$R_{12}T_c = T_cT_2^{-1}T_1 \tag{B.1}$$

We are looking for the calibration transformation T_c. This is a $AX = XB$ type of problem on homogenous rigid transformation matrices, which is also encountered in the context of calibrating wrist-mounted robotic sensors [136, 99]. A solution to equation B.1 has one rotational and one translational degree of freedom. Hence, at least two such equations are needed in order to yield a unique solution. In other words, two relative motions of the ultrasound transducer have to be recorded, requiring the acquisition of a minimum of three volumes. A closed-form solution to two $AX = XB$ equations has been described in [136], however we prefer a least-squares minimization, as its behavior is better in the presence of noise. Furthermore, the acquisition of three volumes produces three motions, which should all be taken into account to yield an optimal result. Hence we would like to minimize the following term:

$$argmin_{T_c} \sum_{i \neq j} d\left(R_{ij}T_c, T_cT_j^{-1}T_i\right) \tag{B.2}$$

A non-linear least-squares optimization computes the parameters of T_c, expressed in a 6-vector containing the translation and rotation in Euler-angles. For the function d, an appropriate distance metric on rigid transformation matrices has to be chosen, several of which have been suggested in the literature. The general problem is that one error value is to be computed from two physically incompatible units, i.e. angular measurements (in degree or radiant units) and distances (in millimeters or inches), respectively. A wrong weighting can over-emphasize the translational accuracy at the expense of the rotational one. A solution to this problem is not to consider the transformation matrices themselves, but their effect on a bunch of points located in the workspace where high precision is required. This results in the fiducial registration error

$$d(T_a, T_b) = \frac{1}{n}\sum_{i=1}^{n} |T_a\vec{p_i} - T_b\vec{p_i}|^2 \tag{B.3}$$

For our problem, we chose the fiducials $\{\vec{p_i}\}$ as the corners of a cuboid with approximate extensions as the ultrasound volumes.

All acquired ultrasound volumes are reconstructed into a volume with rectangular spacing, whose voxels are interpolated from the ultrasound scanline data using the known geometry of the pyramidal measurement volume. Each two volumes are then registered using a manual pre-alignment, and a successive automatic optimization of the intensity correlation of the reference volume voxels with the interpolated ones from the floating volume:

$$\frac{1}{n}\sum_{i=1}^{n}(I_i - \overline{I})(J_i - \overline{J}) \tag{B.4}$$

where I_i are all non-zero voxel intensities in the reference volume and J_i the intensities at the transformed positions in the floating volume, computed using trilinear interpolation. Note that we do not normalize the correlation with the standard deviation of the intensity, as this would yield a dependency on the overlap volume.
For the three volumes required for calibration, the relative motions of the volume centers R_{12},

R_{13} and R_{23} are established using this registration. As the registration result directly affects the precision of the calibration, we assess the quality of the registration by comparing parts of the transformation loop (see figure B.1):

$$d(R_{13}, R_{23}R_{12}) \tag{B.5}$$

B.2.2. Fusion

Once the reference volume V_1 is taken, every further acquisiton V_i can be merged into it instantly, using the relative motion recovered with the tracking data T_i and the calibration transformation T_c:

$$R_{1i} = T_c T_i^{-1} T_1 T_c \tag{B.6}$$

For every new acquisition V_i, an intermediate volume I_i is updated, using the algorithm for traversing a reference and floating volume, that performs the image registration as well. For averaging all intensities, the following weighting is used:

$$I_i = \frac{i-1}{i} I_{i-1} + \frac{1}{i} V_i \tag{B.7}$$

where the variables V_i, I_i and I_{i-1} reflect the indidividual voxel values here. Other possible merging methods are to take the minimum or maximum value of the intermediate and new volume, respectively.

If it is bearable to keep all acquired volumes in memory, one can also use more sophisticated fusion methods. The ultimate goal is to emphasize tissue interfaces depicting anatomy of interest, and disregarding all position-dependant artifacts. Even if physicians would disagree on this, I consider speckle noise as an undesired effect of ultrasound imaging in this context. The acquisition of data from different viewpoints will produce different speckle patterns on the same physical location, hence altering the characteristics for different tissue types. Thus speckle noise in 3D fused imaging can not be understood in the same way than in regular 2D imaging.
Towards this goal, the first method to consider is median merging, i.e. the median of many acquisitions is taken as correct intensity for the fused volume.

B.3. Experiments

B.3.1. Clay Phantom Reconstruction with Calibration and Tracking

In order to study the proposed calibration method, a clay phantom depicting several tubular structures was scanned in a water bath. The probe with an attached optical tracking target was mounted on a holder. In addition to the phantom, some rubber sheets were submerged in the water to minimize multiple reflections of ultrasonic pulses on the glass walls of the used container (figure B.2). The used ultrasound system was a Siemens Sonoline Antares with a 2.2-4.7Mhz 3D/4D curved 1D array wobbler probe.
For the first set of experiments, 10 volumes were acquired, 6 of them with an additional plastic bar placed on top of the object to produce some occlusion. For each volume, the optical tracking data was recorded as well. The non-occluded acquisitions were used for

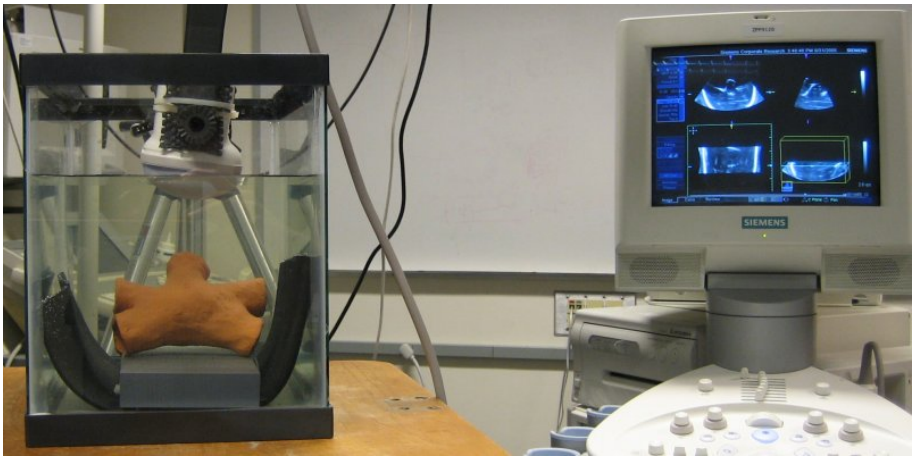


Figure B.2.: Clay phantom submerged in water

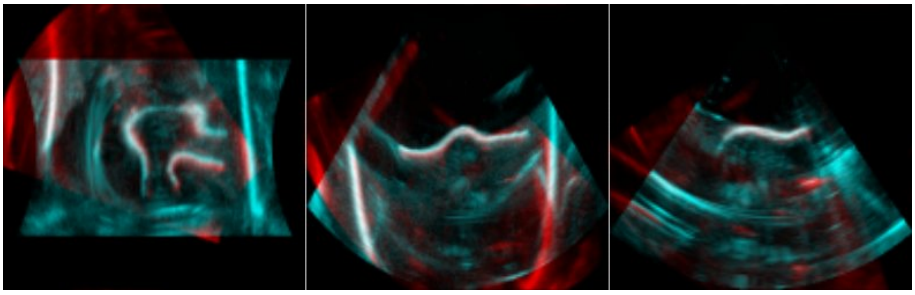


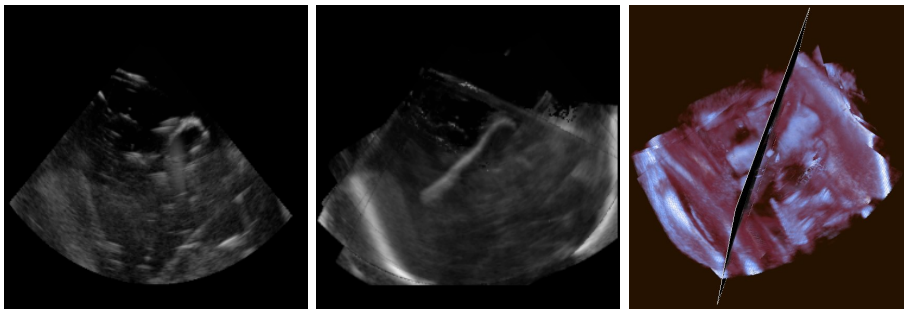
Figure B.3.: 3-Slice overlay of registered clay phantom volumes.

calibration, hence an intensity-based registration was performed on all volume pairs. The three volumes that yielded the best accuracy in terms of the transformation loop equation B.5 were used to derive the calibration transformation T_c. The resulting fiducal error of a cube with $100mm$ edge length was $1.8mm$. The mean calibration error (equation B.3 without the square, in order to produce euclidian distances) was $2.1mm$. Hence the resulting accuracy depended mainly on the intensity-based registration. Possible sources of error are position-dependant artifacts involved in the registration process, as well as a slightly wrong scaling of the reconstructed ultrasound volumes due to the speed of sound in water. Figure B.3 shows a 3-slice overlay visualization of one of the registration results.

Using the tracking data and the calibration transformation, the 6 volumes of the object with the occlusion bar were averaged using the successive fusion, equation B.7. The result is depicted in figure B.4. A particular slice of interest that is obscured in the straight acquisition, is enhanced using the successive 5 volumes from slightly altered angles.

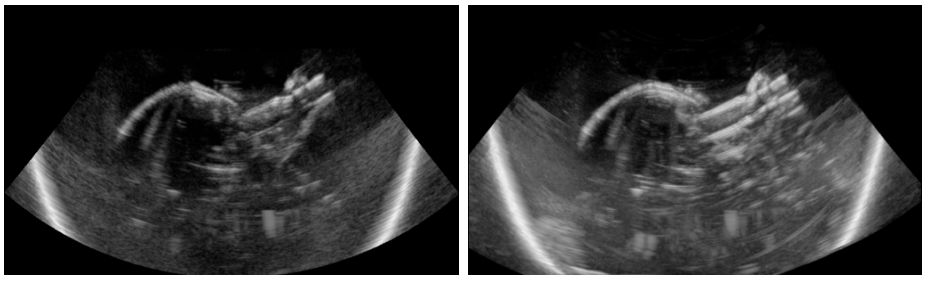
B.3.2. Heart Phantom Reconstruction solely based on intensity information

For this experiment, a plastic phantom of a heart was used, whose surface is more smooth compared to the clay model used before. Therefore the ultrasonic reflections are mainly

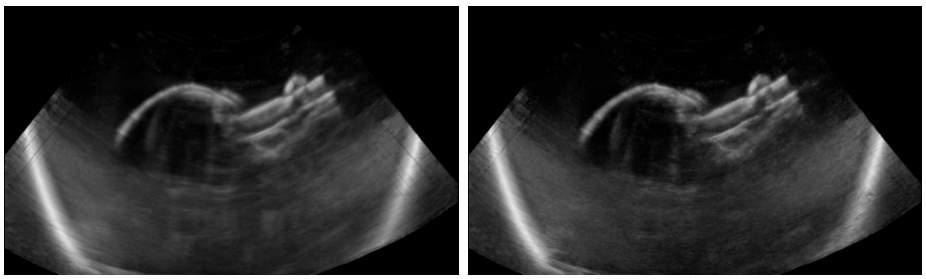


(a) from reference volume (b) from fused volume (c) VRT with slice

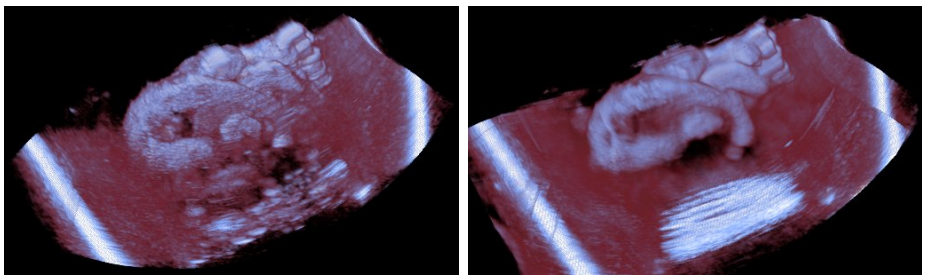
Figure B.4.: A slice of interest is enhanced using fusion of 6 volumes.



(a) Reference Volume (b) Maximum Fusion

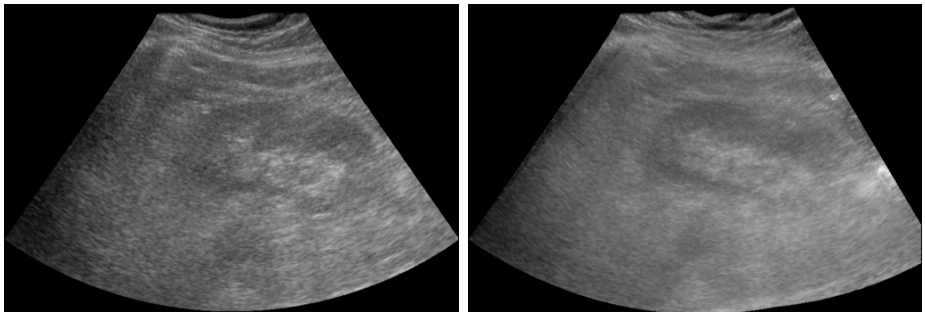


(c) Mean Fusion (d) Median Fusion



(e) Reference VRT (f) Mean VRT

Figure B.5.: Intensity-based fusion of 9 volumes from a heart phantom.



(a) Single Slice (b) Mean Fusion of 3 Volumes

Figure B.6.: Fusion of 3 Volumes from a kidney.

specular, resulting in an incomplete surface reconstruction that can be improved using volume fusion. The Sonoline Antares machine supports storing a sequence of 15 volumes acquired in 4D imaging mode, hence any slow movement over the heart could be used for reconstruction of a single fused volume. However, as the scans are taken with a 1D "wobbling" probe which is continuously moving, the scans would be distorted if taken while moving the probe. Therefore I decided to perform the scans in 3D mode, manually adjusting the holder arm for each new acquisition.
Nine volumes of the heart were established and all starting with the second were registered to the first one using the described method. The fusion was then performed using a variety of methods, the results can be seen in figure B.5. The reconstruction using the reference volume only is incomplete and grainy (e), while the one using the mean of all 9 volumes clearly depicts the shape of the heart (f). On the MPR slice one can see very well the effect of the different fusion methods. Using the maximum intensity of all volumes assures that any depiction of anatomy in a volume will be transferred into the final image, hence it would be very effective in overcoming occlusion. On the other hand, all position-dependant artifacts, in our case in particular multiple reflections in the water bath, will be accumulated from all volumes, resulting in a lot of confusing structures in the final image (b). The mean fusion (c) produces the most smooth image and is therefore well-suited for volume rendering or surface reconstructions. Using the median of all volumes (d) removes outliers in individual volumes, while it still keeps more of the noise characteristics from a single acquisition, resulting in a less smooth volume. Hence it might be the method of choice for improving on a single slice of interest.

B.3.3. B-Mode Reconstruction on Human Kidney and Liver

Acquisitions on the kidney of a collegue were done with 3.3 Mhz and intermediate quality settings (a scan took approximately 4 seconds). The organ itself is fairly rigid, hence several acquisitions from similar positions could be well registered with the automatic algorithm. Fusion of several volumes however barely improved the visual appearance of the organ. The edges are barely visible in the image, so the impression of the organ shape arises to a big extent from the different speckle characteristics (figure B.6). Fusion of several volumes averages out most of the speckle, which is hence an undesired result in this case.

Several acquisitions with the highest quality settings were performed from the author's liver. The portal vein, as well as some vessel structures inside the liver and the liver border are very well outlined. Fusion of several volumes can even improve the appearance, while it might blur regions were small deformations occur. These are due to the pulsating aorta, and slightly different breathing stages for the different volumes. A simple deformable registration should overcome these problems, while requiring more processing time though.
Note: Several 3D ultrasound volumes of the author's spine were acquired as well, however at the time of writing no evaluation had been taken place yet.

B.3.4. Powerdoppler reconstruction of the aorta

For a reliable diagnosis of abdominal aortic aneurisms (AAA), the aorta, located close to the spine beneath the liver, has to be imaged and measured throughout its whole length. Using doppler ultrasound, the shape of the aorta vessel can be extracted more precisely and conveniently, as it images the blood flowing through it. As only the flow component parallel to the ultrasound scan pulses can be retrieved, the physician has to apply the probe at certain angles with respect to the aorta. The Siemens Antares ultrasound machine supports the simultaneous recording of B-Mode and Doppler 3D volumes. Fusing several volumes acquired at different positions and angles along the patient'a abdomen can yield a complete doppler reconstruction of the aorta. As the doppler information might not correspond due to different insonification angles, one would use the B-mode information for registration of the individual volumes, while a fused dataset can be created from both the B-Mode and Doppler data.
A preliminary experiment was conducted by recording several dual-channel volumes of the author's abdomen, however it was very difficult to get nice doppler images of the aorta without the assistance of a physician. Hence the resulting reconstructions show mainly vessel structures inside the liver where a sufficient flow towards the probe occurs.

B.4. Conclusion

Extended-field-of-view Imaging and similar technologies have been around for quite some time [42], providing means to enhance 2D ultrasonic data. Now that 3D transducers are used more and more often, it is time to create solutions to improve the experience of physicians dealing with ultrasound, using three-dimensional data of the anatomy. We demonstrated methods for automatic fusion of an arbitrary number of 3D ultrasound volumes, that possess the capability of recovering occluded structures and enhancing both 2- and three-dimensional regions of interest. While the registration and fusion was performed offline in our preliminary experiments, we believe that an interactive solution, that can register and merge volumes as they are acquired, is well feasible. The numerical computing capabilities of modern graphics hardware are already used for real-time visualization of the 4D ultrasound volumes in the pyramidal grid that they were acquired in [148]. Hence a GPU-based volume registration on these volumes would be a straightforward extension (using GPU pixel shader programs), yielding registration times possibly below a second.
On the other hand, we demonstrated how to utilize additional tracking information in order to fuse the data without any registration, and introduced an adapted calibration method.
Even a hybrid solution using a very compact tracking system integrated in the ultrasound probe (e.g. magnetic or gyroscopic), that delivers either not all degrees of freedom of the transformation, or has very limited accuracy, could be thought of. The incomplete position

information can then be supplemented by intensity-based registration of the volumes.
Some potential applications demand the use of a 2D-array ultrasound transducer, which allows to acquire volumetric information in an instant of time, so that keeping the probe steady is not necessary.

C. Quality-based Registration of Optical Tomography Volumes

A novel optical tomographic imaging modality, related to fluorescence microscopy, allows to acquire cross-sectional slices of small specially prepared biological samples with astounding quality and resolution. However, scattering of the fluorescence light causes the quality to decrease proportional to the depth of the currently imaged plane. Scattering and beam thickness of the excitation laser light cause additional image degradation. We perform a physical simulation of the light scattering in order to define a quantitative function of image quality with respect to depth. This allows us to establish 3D-volumes of quality information in addition to the image data. Volumes are acquired at different orientations of the sample, hence providing complementary regions of high quality. We propose an algorithm for 3D-3D registration of these volumes incorporating voxel quality information, based on maximizing an adapted linear correlation term. The quality ratio of the images is then used, along with the registration result, to create improved volumes of the imaged object. The methods are applied on acquisitions of a mouse brain and mouse embryo to create outstanding three-dimensional reconstructions. This work was published in [164].

C.1. Introduction

Fluorescence microscopy is a technique to depict in-situ biological tissue at cell resolution level [177]. Fluorescent molecules absorb light used for excitation, and emit light at a longer wavelength. Using a microscope with an appropriate optical filter, the emitted light can be measured solely. Usually cells are stained with a particular fluorescent dye, which allows to tag e.g. proteins and to observe their interaction within the cell. A novel optical tomography system was developed and combined with another technique [36]. Anatomical preparations become transparent if both the tissue and surrounding liquid exhibit the same optical refraction index, which can be achieved by basically exchanging water with a mixture of benzoin and benzyl alcohols [144, 63]. The microscope's focal plane, arbitrarily placed within the sample, is sideways illuminated with a laser and the fluorescent light is measured through the microscope with a digital camera (resolution 1392 x 1024, 12 Bit grayscale). A micropositioning device advances the tray with the sample in steps of $12\mu m$, hence a stack of slices is recorded. The resulting data is a comprehensive three-dimensional reconstruction with approximate isotropic voxel size of $10\mu m$, whose size can comprise several Gigabytes. A great number of biological questions can be addressed using this new imaging modality.
Due to tissue inhomogenity, the fluorescent light is still scattered to some extent while passing through the substance. Hence lower slices suffer a blurring effect, in relation to the distance that light travels through the object to the microscope. Our approach to overcome this problem is to acquire volumes with different orientations of the sample, while establishing corresponding volumes with quality information at the same time. This quality information will be used for both spatial registration of the different recordings, as well as the reconstruc-

tion of improved volumes disposed of blurring.

C.2. Quality Function

We want to establish a function

$$Q : \Omega \rightarrow [0..1]; \quad \Omega \subset \mathbb{R}^3 \tag{C.1}$$

which returns the relative quality at any position in the image space Ω. It is determined by the amount of scattering of the measured light, which in turn depends on the depth that the light is traveling through the object. Assuming that light is only being scattered in the sample and not the surrounding liquid, we can reduce our problem to computing the amount of scattering with respect to tissue depth. This is done using a Monte Carlo simulation of light propagation similar to [114], which we briefly describe in the following.
Instead of tracing single photons, we consider photon packets with a certain initial weight for efficiency. We assume that "centers" where both scattering and absorption occur, are distributed uniformly throughout the tissue. Photon packets are initialized with the emitting position $\vec{x_0} = (0, 0, -z)^T$ and weight $w = 1$. Then they repeatedly travel from their actual position $\vec{x_i}$ a certain distance s_i in direction $\vec{d_i}$, until a scattering and absorption event occurs. The photon absorption obeys the classical attenuation relationship

$$N(s) = N_0 e^{-\mu_t s} \tag{C.2}$$

where μ_t is the transmission coefficient, N(s) is the number of photons remaining at distance s from an original number N_0. An adequate generating function $g(x)$ for the probability variable s from a uniformly distributed variable X is

$$g(x) = \frac{1}{\mu_t} \log(1 - x) \tag{C.3}$$

The mean free pathlength is $< s >= 1/\mu_t$. The scattering in tissue can be characterized by the Henyey-Greenstein phase function, which is a probability density function of the scattering angle, given an anisotropy factor g:

$$f_{HG}(\phi) = \frac{1 - g^2}{4\pi(1 + g^2 - 2g\cos\phi)^{\frac{3}{2}}} \tag{C.4}$$

For our simulation it needs to be transformed to a generating function from a uniformly distributed random variable X as well:

$$\cos\phi = \frac{1}{2g}\left(1 + g^2 - \left(\frac{1 - g^2}{1 - g + 2gX}\right)^2\right) \tag{C.5}$$

In each iteration, the photon position, orientation and weight is then updated:

$$\begin{gathered} \vec{x_{i+1}} = \vec{x_i} + s_i\vec{d_i}; \quad w_{i+1} = w - \frac{\mu_a}{\mu_t}; \quad \vec{d_i} = (d_x, d_y, d_z)^T \\ \vec{d_{i+1}} = \begin{pmatrix} \frac{\sin\theta}{\sqrt{1-d_z^2}}(d_x d_z \cos\phi - d_y \sin\phi) + d_x \cos\theta \\ \frac{\sin\theta}{\sqrt{1-d_z^2}}(d_y d_z \cos\phi + d_x \sin\phi) + d_y \cos\theta \\ -\sin\theta\cos\phi\sqrt{1 - d_z^2} + d_z \cos\theta \end{pmatrix} \end{gathered} \tag{C.6}$$

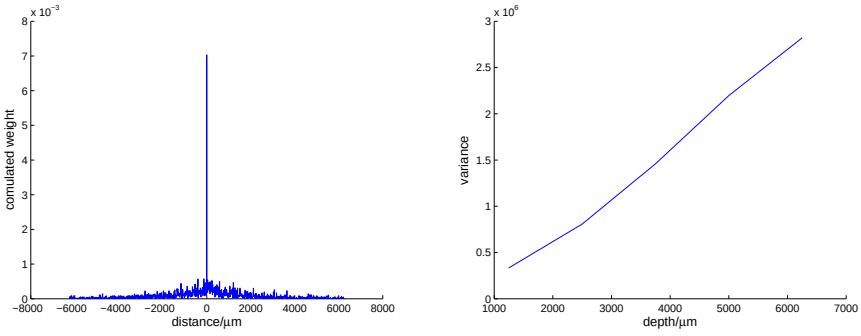


Figure C.1.: Results of simulated scattering and function of standard deviation per depth.

The simulation is terminated if the photon packet reaches the top of the object, $x_3 \geq 0$, where $\vec{x_i} = (x_1, x_2, x_3)^T$. If the weight falls below a threshold $w_i < w_T$, a roulette approach decides if the photon packet is terminated. With a defined probability p_t the photon packet is discarded, otherwise it is reinserted in the simulation with a new weight $w_0 = w_i/p_t$. This makes sure that the energy conservation is not violated.
We are interested in the distance of the virtual point, where the light seems to come from assuming a straight line through the image plane, to the point where the simulation was started. The variance of this distance for many photon packets directly relates to the amount of blurring, i.e. our sought-after quality. Figure C.1 depicts the deviation results for a simulation at particular depth, as well as the function of quality versus depth. The latter is approximately a linear relationship, which we accordingly use for assembling volumes of quality information $Q(x)$.

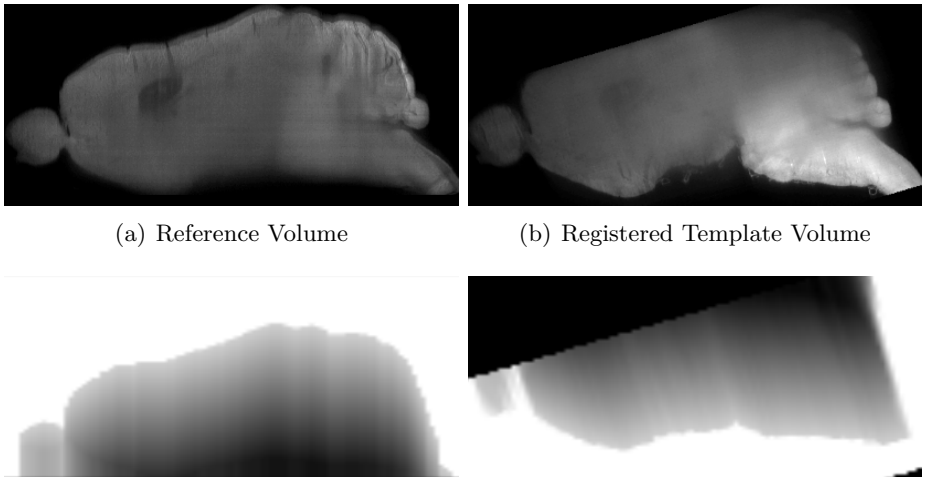
C.3. Quality-Based Registration and Merging

For multiple acquisitions, the preparation is carefully re-oriented within the test tube. No significant deformations occur in this context, however the coordinate system of the second acquisition has to be mapped onto the first one with very high precision, in order to use the combined information for reconstruction. This alignment is hence performed using an automatic rigid intensity-based registration method (see [86] for an overview of image registration techniques). Such methods conduct a non-linear optimization of the transformation parameters, in order to maximize a similarity criterion defined on the voxel intensities of the reference and template volumes R and T, respectively:

$$\phi_{reg} = \arg\max_{\phi} S\left(\{(R(\vec{x_i}), T(\phi(\vec{x_i}))) \,|\vec{x_i} \in \Omega_\phi\}\right) \tag{C.7}$$

where $\{\vec{x_i}\}$ are all discrete voxel positions of the reference volume, ϕ is a 6-DOF rigid transformation, and Ω_ϕ is the volume overlap region for a given ϕ. We use Normalized Cross-Correlation (NCC) as similarity criterion:

$$\begin{aligned} r'_i = r_i - \overline{r}; \quad t'_i = t_i - \overline{t} \\ S = \frac{\sum_i r'_i t'_i}{\sqrt{\sum_i r'^2_i \sum_i t'^2_i}} \end{aligned} \tag{C.8}$$



(a) Reference Volume (b) Registered Template Volume

(c) Reference Quality Volume (d) Registered Template Quality

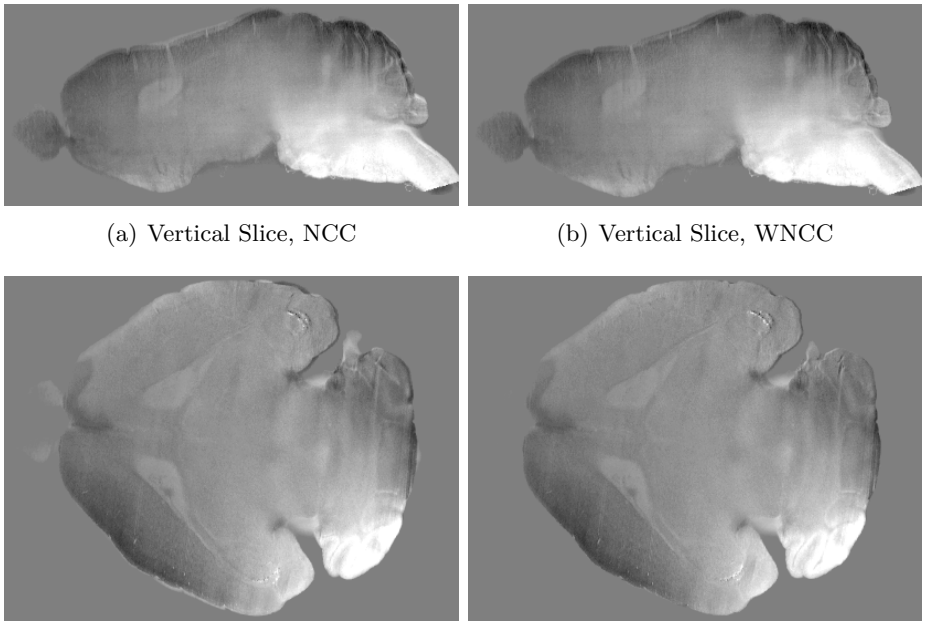
Figure C.2.: Vertical slice of intensity and quality information from registered brain data.

For all voxels $r_i = R(\vec{x_i})$ of the reference volume, the corresponding voxel $t_i = T(\phi(\vec{x_i}))$ is trilinearly interpolated from the template volume.
In order to incorporate the voxel quality information Q_R and Q_T, we do not need to alter the registration algorithm itself. Only an adapted insertion of the voxel values with a weight $w_i \in [0..1]$ into equation C.8 is needed:

$$\begin{aligned} w_i &= Q_R(\vec{x_i})Q_T(\phi(\vec{x_i})) \\ r_i^* = w_i(r_i - \overline{r}); \quad t_i^* &= w_i(t_i - \overline{t}) \\ S^* &= \frac{\sum_i r_i^* t_i^*}{\sqrt{\sum_i r_i^{*2} \sum_i t_i^{*2}}} \end{aligned} \tag{C.9}$$

Using this weighting, voxels with high quality in both volumes affect the individual sums of the NCC equation more. We denote this similarity measure Weighted Normalized Cross-Correlation (WNCC).
Note that a simple, approximative alternative is to use a limited joint volume of interest Ω, where the quality is sufficiently high in both volumes. This is in our case a manually defined slab from the center slices. However, we would like to provide a general framework for incorporating quality information into registration rather than a quick specialized solution. In addition, the precision and especially robustness (as large portions of the images have to be omitted) of this center-slab approach is not convincing, as we experienced in an early registration study. Eventually, when the registered datasets are to be combined, the quality information is a prerequesite in order to allow a smooth transition. For merging two registered volumes, we consider the quality information in the following way:

$$M(\vec{x_i}) = \frac{R(\vec{x_i})Q_R(\vec{x_i}) + T(\phi(\vec{x_i}))Q_T(\phi(\vec{x_i}))}{Q_R(\vec{x_i}) + Q_T(\phi(\vec{x_i}))} \tag{C.10}$$



(a) Vertical Slice, NCC (b) Vertical Slice, WNCC

(c) Horizontal Slice, NCC (d) Horizontal Slice, WNCC

Figure C.3.: Difference of reference and template volumes of the brain preparation after registration.

C.4. Results

C.4.1. Registration Accuracy

An in-vitro preparation of a mouse brain was imaged from the top and bottom side (figure C.2). Its total length is $9mm$, the volume was downsampled to size $256x256x189$ for registration. Figure C.3 shows a vertical and horizontal difference slice for the two registration methods. The standard method results in slightly larger borders, and especially a wrong displacement in vertical direction, as the blurred regions, located in opposite directions in both images, are fully considered for the similarity measure. The robustness of the registration was assessed with a randomized study: 236 registration computations were executed with initial transformations randomly displaced up to $1mm$ and $6°$ from the manually defined starting estimate. Both methods perform equally stable, the standard deviation of the resulting translational parameters is $6.4\mu m$, which corresponds to the parameter abortion criteria of the used Hill-Climbing optimizer. The translations of the two methods are $0.1mm$ displaced (figure C.4).

C.4.2. Merging

Figure C.5 shows the result of merging two volumes of a mouse embryo, the preparation was flipped sideways (approx. $180°$) between the two acquisitions. Precise image registration is crucial, as the resulting voxels are taken from both volumes. Each single data set is heavily blurred on one side, while the final reconstruction depicts very sharp and detailed features throughout the whole volume without any visible reconstruction artifacts.

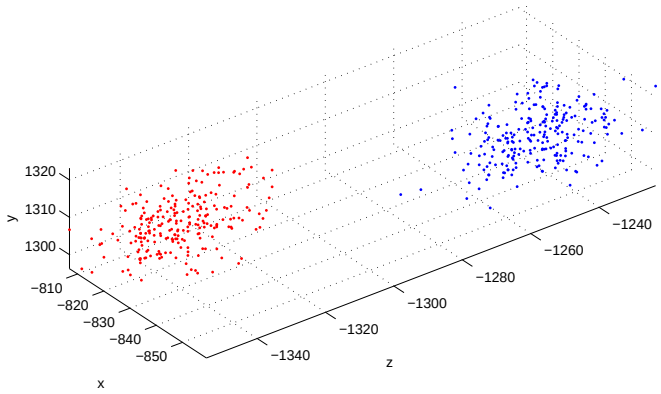
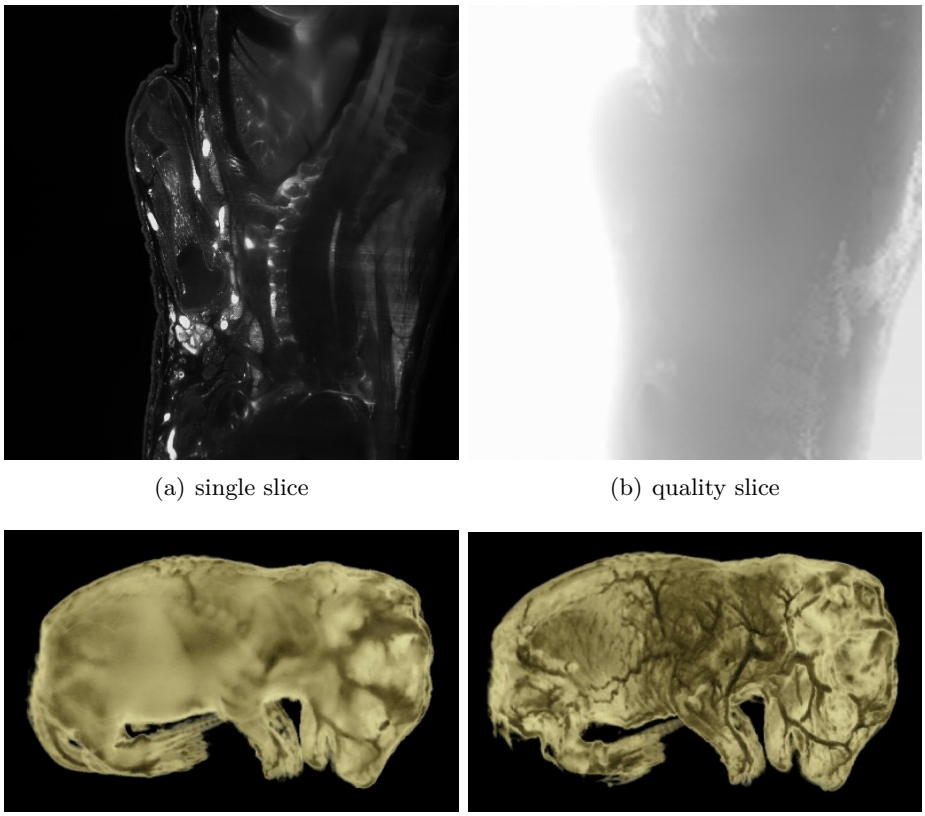


Figure C.4.: Translation vectors from different registrations, Blue=NCC, Red=WNCC

C.5. Conclusion

We presented an algorithm to reduce artifacts arising from a novel optical tomographic imaging modality. Depth-wise degradation of image quality can be overcome by registering multiple volumetric acquisitions. A physical simulation of the light scattering in the object allows us to derive additional volumes of relative voxel quality information. These are both used in an adapted registration algorithm, and for weighting multiple intensities during merging of the volumes. We believe that this straight-forward extension can be easily applied to other modalities where quality-related information is available. We demonstrated the increased precision of our quality-based registration on an optical tomography volume. The subsequent merging of registered data produces continuously high quality throughout the whole image space. The result are three-dimensional reconstructions of in-vitro biological tissue samples, with a resolution and quality which, to our knowledge, has never been achieved before.



(a) single slice (b) quality slice

(c) VRT of single acquisition (d) VRT of reconstruction result

Figure C.5.: Slices and volume rendering (VRT) & reconstruction result from two flipped acquisitions of a whole mouse embryo.

D. Abbreviations

CT. Computed Tomography

CAT. Computer Aided Tomography = CT

CEUS. Contrast-Enhanced Ultrasound

CMUT. Capacitive Micro-machined Ultrasonic Transducer

CR. Correlation Ratio

CPS. Contrast Pulse Sequencing

CPU. Central Processing Unit

DOF. Degree of Freedom

DSA. Digital Substraction Angiography

EBRT. External-Beam Radiotherapy

FLE. Fiducial Localization Error

FRE. Fiducial Registration Error

GPGPU. General Purpose GPU

GPU. Graphics Processing Unit

HCC. Hepato-Cellular Carcinoma

ICE. Intracardiac Echography

ICP. Iterative Closest Point

IMRT. Intensity-Modulated Radiotherapy

KLD. Kullback-Leibler Divergence

MI. Mutual Information

MRA. Magnetic Resonance Angiography

MRI. Magnetic Resoncance Imaging (also: MR)

MST. Minimum Spanning Tree

NCC. Normalized Cross Correlation

PDF. Probability Density Function

PSF. Point Spread Function

PVI. Partial Volume Interpolation

RFA. Radiofrequency Ablation

RMS. Root Mean Square

ROI. Region of Interest

SAD. Sum of Absolute Differences

SNR. Signal to Noise Ratio

SSD. Sum of Squared Differences

SSM. Statistical Shape Model

TRE. Target Registration Error

US. Ultrasound

E. Own Publications

Publications as First Author

- W. Wein, B. Röper, and N. Navab. Integrating diagnostic B-mode ultrasonography into CT-based radiation treatment planning. *IEEE Trans. Med. Imag.*, 26:866–879, June 2007.

- W. Wein and A. Khamene. Image-based method for in-vivo freehand ultrasound calibration. In *SPIE Medical Imaging 2008, San Diego*, Feb. 2008.

- W. Wein, A. Khamene, D. Clevert, O. Kutter, and N. Navab. Simulation and fully automatic multimodal registration of medical ultrasound. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

- W. Wein, M. Blume, U. Leischner, H. Dodt, and N. Navab. Quality-based registration and reconstruction of optical tomography volumes. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

- W. Wein, F. Pache, B. Röper, and N. Navab. Backward-warping ultrasound reconstruction for improving diagnostic value and registration. In *MICCAI 2006 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2006.

- W. Wein, B. Röper, and N. Navab. Automatic registration and fusion of ultrasound with CT for radiotherapy. In *MICCAI 2005 Proceedings*, volume 3750 of *Lecture Notes in Computer Science*, pages 303–311. Springer, Oct. 2005.

- W. Wein, B. Röper, and N. Navab. 2D/3D registration based on volume gradients. In *SPIE Medical Imaging 2005, San Diego*, Feb. 2005.

Publications as Co-Author

- M. Blume, D. Zikic, W. Wein, and N. Navab. A new and general method for blind shift-variant deconvolution of biomedical images. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

- C. Wachinger, W. Wein, and N. Navab. Three-dimensional ultrasound mosaicing. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

- M. Burns, M. Haidacher, W. Wein, I. Viola, and E. Groeller. Feature emphasis and contextual cutaways for multimodal medical visualization. In *EuroVis 2007 Proceedings*, May 2007.

- M. John, Y. Sun, S. Kadoury, W. Wein, Y. Li, J. Resnick, G. Plambeck, A. Dempsey, A. Al-Ahmad, R. Fahrig, and F. Sauer. Fusion of intracardiac ultrasound with 3D cardiac C-arm CT from animal data for electrophysiology. In *BVM 2007 Proceedings*, Mar. 2007.

- A. Khamene, R. Chisu, W. Wein, N. Navab, and F. Sauer. A novel projection based approach for medical image registration. In *Biomedical Image Registration*, volume 4057 of *Lecture Notes in Computer Science*, pages 247–256, 2006.

- A. Khamene, P. Bloch, W. Wein, F. Sauer, and M. Svatos. Automatic portal image based patient positioning for radiation therapy. *Medical Image Analysis*, 10:96–112, Feb. 2006.

- B. Roeper, W. Wein, P. Kneschaurek, M. Molls, M. Dobritz, and N. Navab. Image-based fusion of diagnostic ultrasound and planning CT for radiotherapy of head and neck cancer – a way towards improvement of regional target volume delineation. In *DEGRO 2006 Proceedings*, May 2006.

- D. Zikic, W. Wein, A. Khamene, D. Clevert, and N. Navab. Fast deformable registration of 3D-ultrasound data using a variational approach. In *MICCAI 2006 Proceedings*, volume 4190 of *Lecture Notes in Computer Science*, pages 915–923, 2006.

Bibliography

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. Pattern Anal. Machine Intell.*, 16:641–647, 1994.

[2] Advanced Realtime Tracking GmbH (ART). ARTtrack & DTrack manual, 2005.

[3] A. Ahuja and M. Ying. Sonography of neck lymph nodes. part II: abnormal lymph nodes. *Clin Radiol*, 58:359–366, May 2003.

[4] D. Aiger and D. Cohen-Or. Real-time ultrasound imaging simulation. *Real-Time Imaging*, 4:263–274, 1998.

[5] D. Aiger and D. Cohen-Or. Mosaicing ultrasonic volumes for visual simulation. *IEEE Computer Graphics and Applications*, 20:53–61, Mar. 2000.

[6] B. D. Allen, G. Bishop, and G. Welch. Tracking: Beyond 15 minutes of thought. SIGGRAPH Course 11, 2001.

[7] J. Amanatides and A. Woo. A fast voxel traversal algorithm for ray tracing. In *Eurographics '87*, pages 3–10. Elsevier Science Publishers, Amsterdam, North-Holland, 1987.

[8] D. Amin, T. Kanade, A. M. D. Gioia, and B. Jaramaz. Ultrasound registration of the bone surface for surgical navigation. *Computer Aided Surgery*, 8(1):1 – 16, Jan. 2003.

[9] T. Arbel, X. Morandi, R. Comeau, and D. Collins. Automatic non-linear MRI-Ultrasound registration for the correction of intra-operative brain deformations. In *MICCAI 2001 Proceedings*, volume 2208 of *Lecture Notes in Computer Science*, pages 913–922. Springer, 2001.

[10] R. Barkmann, W. Timm, S. Sakata, M. Heller, and C. Glüer. Simulation of ultrasound interaction with bone structure based on 3D high-resolution-magnetic resonance- and micro-ct-images. In *22nd Annual EMBS International Conference*, pages 2666–2669, July 2000.

[11] D. Barratt, G. Penney, C. Chan, M. Slomczykowski, T. Carter, P. Edwards, and D. Hawkes. Self-calibrating 3D-ultrasound-based bone registration for minimally invasive orthopedic surgery. *IEEE Trans. Med. Imag.*, 25:312–323, Mar. 2006.

[12] C. Barry, C. Allott, N. John, P. Mellor, P. Arundel, D. Thomson, and J. Waterton. Three-dimensional freehand ultrasound: Image reconstruction and volume analysis. *Ultrasound in Med. & Biol.*, 23:1209–1224, 1997.

[13] J. Beirlant, E. Dudewicz, L. Gyorfi, and E. van der Meulen. Nonparametric entropy estimation: An overview. *International Journal of the Mathematical Statistics Sciences*, 6:17–39, 1997.

[14] J. Blackall, G. Penney, A. King, and D. Hawkes. Alignment of sparse freehand 3-D ultrasound with preoperative images of the liver using models of respiratory motion and deformation. *IEEE Trans. Med. Imag.*, 24:1405–1416, Nov. 2005.

[15] J. Blackall, D. Rueckert, C. M. Jr., G. Penney, D. Hill, and D. Hawkes. An image registration approach to automated calibration for freehand 3D ultrasound. In *MICCAI Proceedings*, pages 462–471, 2000.

[16] J. M. Blackall. *Respiratory Motion in Image-Guided Interventions of the Liver*. PhD thesis, University of London, Computational Imaging Science Group, 2002.

[17] M. Blume, D. Zikic, W. Wein, and N. Navab. A new and general method for blind shift-variant deconvolution of biomedical images. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

[18] E. Boctor, I. Iordachita, G. Fichtinger, and G. Hager. Real-time quality control of tracked ultrasound. In *MICCAI Proceedings*, Oct. 2005.

[19] E. Boctor, I. Iordachita, G. Fichtinger, and G. Hager. Ultrasound self-calibration. In *Proceedings of SPIE Medical Imaging*, volume 6141, 2006.

[20] J. Boes and C. Meyer. Multi-variate mutual information for registration. In *MICCAI 1999 Proceedings*, volume 1679 of *Lecture Notes in Computer Science*, pages 606–612. Springer, 1999.

[21] G. Borgefors. Distance transforms in arbitrary dimensions. *Computer Vision, Graphics and Image Processing*, 27:321–345, 1984.

[22] D. Boukerroui, J. Noble, and M. Brady. Velocity estimation in ultrasound images: A block matching approach. In *IPMI 2003 Proceedings*, volume 2732 of *Lecture Notes in Computer Science*, pages 586–598. Springer, 2003.

[23] B. Brendel, S. Winter, A. Rick, M. Stockheim, and H. Ermert. Registration of 3D CT and ultrasound datasets of the spine using bone structures. *Computer Aided Surgery*, 7:146–155, 2002.

[24] M. Burns, M. Haidacher, W. Wein, I. Viola, and E. Groeller. Feature emphasis and contextual cutaways for multimodal medical visualization. In *EuroVis 2007 Proceedings*, May 2007.

[25] C. Castro-Pareja, V. Zagrodsky, L. Bouchet, and R. Shekhar. Automated prostate localization in external-beam radiotherapy using mutual information-based registration of treatment planning CT and daily 3D ultrasound images. In *International Congress Series*, volume 1281, pages 435–440, 2005.

[26] C. Chan, D. Barratt, P. Edwards, G. Penney, M. Slomczykowski, T. Carter, and D. Hawkes. Cadaver validation of the use of ultrasound for 3D model instantiation of bony anatomy in image guided orthopaedic surgery. In *MICCAI 2004 Proceedings, Part II*, page 397ff, 2004.

[27] R. Chisu. Techniques for accelerating intensity-based rigid image registration. Master's thesis, Technische Universität München, Feb. 2005.

[28] M. A. Choti. Surgical management of hepatocellular carcinoma: Resection and ablation. *J. Vasc. Interv. Radiol.*, 13:S197–S203, 2002.

[29] A. Chung, W. W. III, A. Norbash, and W. Grimson. Multi-modal image registration by minimising kullback-leibler distance. In *MICCAI proceedings, Lecture Notes in Computer Science*, volume 2489, pages 525–532, 2002.

[30] B. Cohen and I. Dinstein. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35:455–463, 2002.

[31] C. Coolens, M. White, D. Hawkes, D. Atkinson, L. Charles-Edwards, D. Tait, and M. Hawkins. Free-breathing liver gated radiotherapy with external markers using MRI derived models of hepatic motion. In *ICCR 2007 Proceedings*, June 2007.

[32] P. Coupe, P. Hellier, N. Azzabou, and C. Barillot. 3D freehand ultrasound reconstruction based on probe trajectory. In *MICCAI Proceedings*, pages 597–604, 2005.

[33] F. Cury, G. Shenouda, L. Souhami, M. Duclos, S. Faria, M. David, F. Verhaegen, R. Corns, and T. Falco. Ultrasound-based image guided radiotherapy for prostate cancer - comparison of cross-modality and intramodality methods for daily localization during external beam radiotherapy. *Int J Radiat Oncol Biol Phys*, 66:1562–1567, 2006.

[34] R. Dantas, E. Costa, and S. Leeman. Ultrasound speckle and equivalent scatterers. *Ultrasonics*, 43:405–420, 2005.

[35] J. Dey and T. Kanade. Forward modeling of ultrasound propagation in layered medium. Technical report, Robotics Institute, Carnegie Mellon University, June 1999.

[36] H.-U. Dodt, U. Leischner, A. Schierloh, N. Jährling, C. Mauch, K. Deininger, J. Deussing, M. Eder, W. Zieglgänsberger, and K. Becker. Ultramicroscoy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nature Methods*, 2007.

[37] H. Ehricke. SONOSim3D: a multimedia system for sonography simulation and education with an extensible case database. *Eur J Ultrasound*, 7:225–300, Aug. 1998.

[38] O. Faugeras. *Three-Dimensional Computer Vision, A Geometric Viewpoint*. The MIT Press, 1993.

[39] A. Fenster, D. B. Downey, and H. N. Cardinal. Three-dimensional ultrasound imaging. *Physics in Medicine and Biology*, 46(5):R67–R99, 2001.

[40] J. Fitzpatrick, J. West, and C. Maurer. Predicting error in rigid-body point-based registration. *IEEE Trans. Med. Imag.*, 17:694–702, Oct. 1998.

[41] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics, Principles and Practice*. Addison-Wesley, 1993.

[42] F. Forsberg. Ultrasound biomedical technology: marketing versus clinical reality. *Ultrasonics*, 42:17–27, 2004.

[43] R. Francois, R. Fablet, and C. Barillot. Robust statistical registration of 3D ultrasound images using texture information. In *Proceedings. International Conference on Image Processing*. IEEE, Sep 2003.

[44] A. Fung, K. Ayyangar, D. Djajaputra, R. Nehru, and C. Enke. Ultrasound-based guidance of intensity-modulated radiation therapy. *Medical Dosimetry*, 31:20–29, 2006.

[45] M. Fuss, B. Salter, S. Cavanaugh, C. Fuss, A. Sadeghi, C. Fuller, A. Ameduri, J. Hevezi, T. Herman, and C. T. Jr. Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies. *Int J Radiat Oncol Biol Phys*, 59:1245–1256, July 2004.

[46] R. Gan and A. C. Chung. Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude. In *IPMI 2005 Proceedings*, volume 3565 of *Lecture Notes in Computer Science*, pages 210–221, 2005.

[47] A. Gee, R. Housden, P. Hassenpflug, G. Treece, and R. Prager. Sensorless freehand 3D ultrasound in real tissue: Speckle decorrelation without fully developed speckle. *Medical Image Analysis*, pages 137–149, 2006.

[48] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, Inc., 1981.

[49] R. Gonzalez and R. Woods. *Digital Image Processing.* Prentice Hall, 2002.

[50] A. Grosu, M. Piert, W. Weber, B. Jeremic, M. Picchio, U. Schratzenstaller, F. Zimmermann, M. Schwaiger, and M. Molls. Positron emission tomography for radiation treatment planning. *Strahlenther Onkol*, 181:483–499, Aug. 2005.

[51] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. Learning based non-rigid multi-modal image registration using kullback-leibler divergence. In *MICCAI 2005 proceedings*, volume 3750 of *Lecture Notes in Computer Science*, pages 255–262, 2005.

[52] J. Hajnal, N. Saeed, E. Soar, A. Oatridge, I. Young, and G. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *J. Comput. Assist. Tomogr.*, pages 289–296, 1995.

[53] J. V. Hajnal, D. L. Hill, and D. J. Hawkes, editors. *Medical Image Registration.* CRC Press, 2001.

[54] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision, 2nd edition.* Cambridge University Press, 2004.

[55] D. Hawkes, D. Barrat, J. Blackall, C. Chan, P. Edwards, K. Rhode, G. Penney, J. McClelland, and D. Hill. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Medical Image Analysis*, 9:163–175, Apr. 2005.

[56] W. Hedrick, D. Hykes, and D. Starchman. *Ultrasound Physics and Instrumentation.* Mosby - Year Book, Inc., 1995.

[57] A. Hero and O. Michel. Robust entropy estimation strategies based on edge weighted random graphs, July 1998.

[58] D. Hill, P. Batchelor, M. Holden, and D. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46:R1–R45, Mar. 2001.

[59] D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff. A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics. In *IPMI Proceedings*, pages 182–196, 1993.

[60] J. H. Hipwell, G. P. Penney, T. C. Cox, J. V. Byrne, and D. J. Hawkes. 2D-3D intensity based registration of DSA and MRA - a comparison of similarity measures. *Lecture Notes in Computer Science*, 2489:501–508, 2002.

[61] T. Hong, W. Tome, R. Chappell, P. Chinnaiyan, M. Mehta, and P. Harari. The impact of daily setup variations on head-and-neck intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys*, 61:779–788, Mar. 2005.

[62] P. Huber. *Robust Statistics*. Wiley, 1981.

[63] W. D. J. The history of Werner Spalteholz's Handatlas der Anatomie des Menschen. *Journal of Audiovisual Media in Medicine*, 22(4):164–170, 1999.

[64] P. Jannin, J. Fitzpatrick, D. Hawkes, X. Pennec, R. Shahidl, and M. Vannier. Validation of medical image processing in image-guided therapy. *IEEE Trans. Med. Imag.*, 21:1445–1449, Dec. 2002.

[65] M. John, Y. Sun, S. Kadoury, W. Wein, Y. Li, J. Resnick, G. Plambeck, A. Dempsey, A. Al-Ahmad, R. Fahrig, and F. Sauer. Fusion of intracardiac ultrasound with 3D cardiac C-arm CT from animal data for electrophysiology. In *BVM 2007 Proceedings*, Mar. 2007.

[66] Jørgen Arendt Jensen. Field II ultrasonic simulation program.

[67] Jørgen Arendt Jensen. Linear description of ultrasound imaging systems. Technical report, Notes for the International Summer School on Advanced Ultrasound Imaging at the Technical University of Denmark, June 2001.

[68] A. Kassam and M. Wood. Fourier registration of three-dimensional brain mr images: exploiting the axis of rotation. *J Magn Reson Imaging*, 6:894–902, Nov. 1996.

[69] J. Kaufman, G. Luo, and R. Siffert. Ultrasound simulation for 3D-axisymetric models. In *IEEE Ultrasonics Symposium*, pages 2065–2068, 2003.

[70] A. Khamene, P. Bloch, W. Wein, F. Sauer, and M. Svatos. Automatic portal image based patient positioning for radiation therapy. *Medical Image Analysis*, 10:96–112, Feb. 2006.

[71] A. Khamene, R. Chisu, W. Wein, N. Navab, and F. Sauer. A novel projection based approach for medical image registration. In *Biomedical Image Registration*, volume 4057 of *Lecture Notes in Computer Science*, pages 247–256, 2006.

[72] A. Khamene and F. Sauer. A novel phantom-less spatial and temporal ultrasound calibration method. In *MICCAI Proceedings*, Oct. 2005.

[73] A. Khamene, S. Vogt, F. Azar, T. Sielhorst, F. Sauer, and H. Niemann. Local 3D reconstruction and augmented reality visualization of free-hand ultrasound for needle biopsy procedures. In *MICCAI 2003 Proceedings*, volume 2879 of *Lecture Notes in Computer Science*, pages 344–355. Springer, 2003.

[74] A. King, J. Blackall, G. Penney, and D. Hawkes. Tracking liver motion using 3-D ultrasound and a surface-based statistical shape model. In *IEEE MMBIA 2001 Workshop*, pages 145–152, 2001.

[75] J. Kybic. High-dimensional mutual information estimation for image registration. In *International Conference on Image Processing*, pages 1779–1782, 2004.

[76] P. Lacroute and M. Levoy. Fast volume rendering using a shear-warp factorization of the viewing transformation. *Computer Graphics*, 28(Annual Conference Series):451–458, 1994.

[77] T. Lange, S. Eulenstein, M. Hunerbein, H. Lamecker, and P.-M. Schlag. Augmenting intraoperative 3D ultrasound with preoperative models for navigation in liver surgery. In *MICCAI 2004 Proceedings, Part II*, page 534ff, 2004.

[78] D. A. LaRose. *Iterative X-Ray/CT registration using accelerated volume rendering*. PhD thesis, Carnegie Mellon University, May 2001.

[79] G. Lauritsch, J. Boese, L. Wigstrom, H. Kemeth, and R. Fahrig. Towards cardiac c-arm computed tomography. *IEEE Trans. Med. Imag.*, 25:922–934, 2006.

[80] T. Lehmann, C. Gönner, and K. Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Trans. Med. Imag.*, 18:1049–1075, Nov. 1999.

[81] A. Leroy, P. Mozer, Y. Payan, and J. Troccaz. Rigid registration of freehand 3D ultrasound and CT-Scan kidney images. In *MICCAI Proceedings*, page 837ff, Sept. 2004.

[82] M. M. Letteboer, P. W. Willems, M. A. Viergever, and W. J. Niessen. Non-rigid registration of 3D ultrasound images of brain tumours acquired during neurosurgery. *Lecture Notes in Computer Science*, 2879:408–415, Nov. 2003.

[83] H. Livyatan, Z. Yaniv, and L. Joskowicz. Gradient-based 2-D/3-D rigid registration of fluoroscopic x-ray to ct. *IEEE Trans. Med. Imag.*, Nov 2003.

[84] J. L.-C. Lo, M. Brady, and N. Moore. Simultaneous multiple image registration method for T_1 estimation in breast MRI images. In *MICCAI 2006 Proceedings*, volume 4190 of *Lecture Notes in Computer Science*, pages 865–872. Springer, 2006.

[85] A. Macovski. *Medical Imaging*. Prentice-Hall, Inc., 1983.

[86] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2:1–36, 1998.

[87] J. H. Mathews. *Numerical Methods for Computer Science, Engineering and Mathematics*. Prentice Hall, Inc., 1987.

[88] H. Maul, A. Scharf, P. Baier, M. Wüstemann, H. Günter, G. Gebauer, and C. Sohn. Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound Obstet Gynecol*, 24:581–585, Aug. 2004.

[89] H. McNair, S. Mangar, J. Coffey, B. Shoulders, V. Hansen, A. Norman, J. Staffurth, S. Sohaib, A. Warrington, and D. Dearnaley. A comparison of CT- and ultrasound-based imaging to localize the prostate for external beam radiotherapy. *Int J Radiat Oncol Biol Phys*, 65:678–687, July 2006.

[90] M. Mellor and M. Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9:330 – 343, Sept. 2005.

[91] L. Mercier, T. Langø, F. Lindseth, and D. Collins. A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Med. & Biol.*, 31:449–471, 2005.

[92] E. Miller. A new class of entropy estimators for multi-dimensional densities. In *International Conference on Acoustics, Speech, and Signal Processing*, 2003.

[93] J. Modersitzki. *Numerical methods for image registration.* Oxford University Press, 2004.

[94] D. Muratore and J. R.L. Galloway. Beam calibration without a phantom for creating a 3-D freehand ultrasound system. *Ultrasound in Med. & Biol.*, 27:1557–1566, 2001.

[95] H. Neemuchwala, A. Hero, and P. Carson. Image matching using alpha-entropy measures and entropic graphs. *Signal Processing*, 85:277–296, 2005.

[96] O. Oralkan, A. Ergun, C.-H. Cheng, J. Johnson, M. Karaman, T. Lee, and B. Khuri-Yakub. Volumetric ultrasound imaging using 2-D CMUT arrays. *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, 50:1581–1594, Nov. 2003.

[97] J. D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krüger, A. E. Lefohn, and T. J. Purcell. A survey of general-purpose computation on graphics hardware. In *Eurographics 2005, State of the Art Reports*, pages 21–51, Aug. 2005.

[98] P. V. Pandharipande, G. A. Krinsky, H. Rusinek, and V. S. Lee. Perfusion imaging of the liver: Current challenges and future goals. *Radiology*, 234:661–673, 2005.

[99] F. Park and B. Martin. Robot sensor calibration: Solving AX = XB on the euclidean group. *IEEE Transactions on Robotics and Automation*, 10:717–721, Oct. 1994.

[100] X. Pennec, N. Ayache, A. Roche, and P. Cachier. Non-rigid MR-US registration for tracking brain deformations. In *MIAR 2001 Proceedings*, pages 79–86, 2001.

[101] G. Penney, D. Barratt, C. Chan, M. Slomczykowski, T. Carter, P. Edwards, and D. Hawkes. Cadaver validation of intensity-based ultrasound to CT registration. *Medical Image Analysis*, pages 385–395, 2006.

[102] G. Penney, J. Blackall, M. Hamady, T. Sabharwal, A. Adam, and D. Hawkes. Registration of freehand 3D ultrasound and magnetic resonance liver images. *Medical Image Analysis*, 8:81 – 91, Mar. 2004.

[103] G. Penney, J. Schnabel, D. Rueckert, M. Viergever, and W. Niessen. Registration-based interpolation. *IEEE Trans. Med. Imag.*, 23:922– 926, July 2004.

[104] G. P. Penney, J. M. Blackall, D. Hayashi, T. Sabharwal, A. Adam, and D. J. Hawkes. Overview of an ultrasound to CT or MR registration system for use in thermal ablation of liver metastases. In *Proc. Medical Image Understanding and Analysis*, pages 65–68, 2001.

[105] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2D-3D medical image registration. *Lecture Notes in Computer Science*, 1496, 2001.

[106] T. Peters. Image-guidance for surgical procedures. *Phys. Med. Biol.*, 51:R505–R540, 2006.

[107] D. Pham and D. Karaboga. *Intelligent Optimisation Techniques*. Springer, 2000.

[108] J. P. W. Pluim and J. M. Fitzpatrick. Image registration. *IEEE Trans. Med. Imag. (editorial)*, 22:1341–1343, Nov. 2003.

[109] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual information based registration of medical images: a survey. *IEEE Trans. Med. Imag.*, 22, Aug. 2003.

[110] B. Porter, D. Rubens, J. Strang, S. Totterman, and K. Parker. Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. *IEEE Trans. Med. Imag.*, 20:354–359, Apr. 2001.

[111] R. Prager, A. Gee, and L. Berman. Stradx: Real-time acquisition and visualization of freehand three-dimensional ultrasound. *Medical Image Analysis*, 3:129–140, 1999.

[112] R. Prager, A. Gee, G. Treece, C. Cash, and L. Berman. Using image-based regression to acquire freehand 3D ultrasound. Technical report, Cambridge University Engineering Department, June 2002.

[113] R. Prager, R. Rohling, A. Gee, and L. Berman. Rapid calibration for 3-D freehand ultrasound. *Ultrasound in Med. & Biol.*, 24:855–869, 1998.

[114] S. A. Prahl, M. Keijzer, S. L. Jacques, and A. J. Welch. A Monte Carlo Model of Light Propagation in Tissue. *SPIE Institute Series*, 5, 1989.

[115] W. Press, S. T. abd W.T. Vetterling, and B. Flannery. *Numerical Recipes in C, Second Edition*. CRC Press, Inc., 1992.

[116] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. Technical report, INRIA, July 1999.

[117] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. Technical report, INRIA, 1998.

[118] A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE Trans. Med. Imag.*, 20:1038 – 1049, Oct. 2001.

[119] A. Roche, X. Pennec, G. Malandain, N. Ayache, and S. Ourselin. Generalized correlation ratio for rigid registration of 3D ultrasound with MR images. Technical report, INRIA, July 2000.

[120] B. Roeper, W. Wein, P. Kneschaurek, M. Molls, M. Dobritz, and N. Navab. Image-based fusion of diagnostic ultrasound and planning CT for radiotherapy of head and neck cancer – a way towards improvement of regional target volume delineation. In *DEGRO 2006 Proceedings*, May 2006.

[121] T. Rohlfing and J. C.R. Maurer. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Medical Physics*, 31:427–432, 2004.

[122] R. Rohling. *3D Freehand Ultrasound: Reconstruction and Spatial Compounding.* PhD thesis, University of Cambridge, Department of Engineering, Sept. 1998.

[123] R. Rohling, A. Gee, and L. Berman. Automatic registration of 3-D ultrasound images. Technical report, Cambridge University Engineering Department, May 1997.

[124] F. Rousseau, P. Hellier, and C. Barillot. Confhusius: A robust and fully automatic calibration method for 3D freehand ultrasound. *Medical Image Analysis*, 9, Sept. 2005.

[125] F. Rousseau, P. Hellier, M. Letteboer, W. Niessen, and C. Barillot. Quantitative evaluation of three calibration methods for 3-D freehand ultrasound. *IEEE Trans. Med. Imag.*, 25:1492–1501, Nov. 2006.

[126] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Trans. Med. Imag.*, 18:712–721, 1999.

[127] D. B. Russakoff, T. Rohlfing, A. Ho, D. H. Kim, R. Shahidi, J. R. Adler, Jr., and C. R. Maurer, Jr. Evaluation of intensity-based 2D-3D spine image registration using clinical gold-standard data. In *Biomedical Image Registration – Second International Workshop, WBIR 2003, Philadelphia, PA, USA, June 23-24, 2003*, volume 2717 of *Lecture Notes in Computer Science*, pages 151–160, June 2003.

[128] M. Sanderse, H. Marquering, E. Hendriks, A. van der Lugt, and J. Reiber. Automatic initialization algorithm for carotid artery segmentation in cta images. In *MICCAI 2005 Proceedings*, volume 3750 of *Lecture Notes in Computer Science*, pages 846–853. Springer, Oct. 2005.

[129] D. Sarrut and S. Clippe. Geometrical transformation approximation for 2D/3D intensity-based registration of portal images and ct scan. In *MICCAI 2001 Proceedings*, pages 532–540, 2001.

[130] T. Scarbrough, N. Golden, J. Ting, C. Fuller, A. Wong, P. Kupelian, and C. T. Jr. Comparison of ultrasound and implanted seed marker prostate localization methods: Implications for image-guided radiotherapy. *Int J Radiat Oncol Biol Phys*, 65:378–387, June 2006.

[131] U. Schneider, E. Pedroni, and A. Lomax. The calibration of CT hounsfield units for radiotherapy treatment planning. *Phys. Med. Biol.*, 41:111–124, 1996.

[132] D. H. Sheafor, E. K. Paulson, M. A. Kliewer, D. M. DeLong, and R. C. Nelson. Comparison of sonographic and ct guidance techniques: does ct fluoroscopy decrease procedure time? *Am J Roentgenol.*, 174:939–942, Apr. 2000.

[133] R. Shekhar and V. Zagrodsky. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Trans. Med. Imag.*, 21:9 – 22, Jan. 2002.

[134] R. Shekhar, V. Zagrodsky, M. Garcia, and J. Thomas. Registration of real-time 3-D ultrasound images of the heart for novel 3-D stress echocardiography. *IEEE Trans. Med. Imag.*, 23:1141 – 1149, Sept. 2004.

[135] D. Shepard. A two-dimensional interpolation function for irregularly spaced data. In *Proc. 23 Nat. Conf. ACM*, pages 517–524, 1968.

[136] Y. 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, Feb. 1989.

[137] K. Shoemake. Animating rotation with quaternion curves. In *SIGGRAPH proceedings*, 1985.

[138] T. Sielhorst, M. Feuerstein, J. Traub, O. Kutter, and N. Navab. CAMPAR: A software framework guaranteeing quality for medical augmented reality. *International Journal of Computer Assisted Radiology and Surgery*, 1(Supplement 1):29–30, June 2006.

[139] D. Skerl, B. Likar, and F. Pernus. A protocol for evaluation of similarity measures for rigid registration. *IEEE Trans. Med. Imag.*, 25:779–791, 2006.

[140] P. Slomka, J. Mandel, D. Downey, and A. Fenster. Evaluation of voxel-based registration of 3-D power doppler ultrasound and 3-D magnetic resonance angiographic images of carotid arteries. *Ultrasound Med. Biol.*, 27:945–955, July 2001.

[141] L. Solbiati, L. Cova, and A. Bellobuono. New improvement of real-time US: CT fusion—the anatomical markers and the "Virtual Needle". In *RSNA 2005 Proceedings (Online Presentation)*, 2005.

[142] L. Solbiati, T. Ierace, M. Tonolini, and L. Cova. Guidance and monitoring of radiofrequency liver tumor ablation with contrast-enhanced ultrasound. *European Journal of Radiology*, 51S:S19–S23, 2004.

[143] M. Sonka and J. Fitzpatrick, editors. *Handbook of Medical Imaging. Volume 2: Medical Image Processing and Analysis*. SPIE Press, 2000.

[144] W. Spalteholz. *Über das Durchsichtigmachen von menschlichen und tierischen Präparaten und seine theoretischen Bedingungen*. S. Hirzel Leipzig Verlag, 2nd extended edition, 1914.

[145] D. Stippel, S. Böhm, K. Beckurts, H. Brochhagen, and A. Hölscher. Experimental evaluation of accuracy of radiofrequency ablation using conventional ultrasound or a third-dimension navigation tool. *Langenbecks Arch Surg.*, 387:303–308, Nov. 2002.

[146] G. Stippel, J. Ellsmere, S. Warfield, W. Wells, and W. Philips. A new technique for multi-modal 3D image registration. In *WBIR 2003 proceedings*, volume 2717 of *Lecture Notes in Computer Science*, pages 244–253. Springer, 2003.

[147] C. Studholme, D. Hill, and D. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32, 1998.

[148] T. Sumanaweera. Applying real-time shading to 3D ultrasound visualization. *GPU Gems, R. Fernando (Editor)*, pages 693–707, 2004.

[149] M. Sumi, M. Ohki, and T. Nakamura. Comparison of sonography and CT for differentiating benign from malignant cervical lymph nodes in patients with squamous cell carcinoma of the head and neck. *AJR Am J Roentgenol*, 176:1019–1024, Apr. 2001.

[150] D. Tomazevic, B. Likar, T. Slivnik, and F. Pernus. 3-D/2-D registration of ct and mr to x-ray images. *IEEE Trans. Med. Imag.*, Nov 2003.

[151] V. J. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 1997.

[152] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Patt. Anal. Machine Intell.*, 13:376–380, Apr. 1991.

[153] P. Viola and W. Wells. Alignment by maximization of mutual information. *Int. J. of Computer Vision*, 24:137–154, 1997.

[154] P. A. Viola. *Alignment by Maximization of Mutual Information.* PhD thesis, Massachusetts Institute of Technology, June 1995.

[155] C. Wachinger. Three-dimensional ultrasound mosaicing. Master's thesis, Technische Universität München, Mar. 2007.

[156] C. Wachinger, W. Wein, and N. Navab. Three-dimensional ultrasound mosaicing. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

[157] M. Wachowiak and T. Peters. High-performance medical image registration using new optimization techniques. *IEEE Trans. on Information Technology in Biomedicine*, 10:344–353, Apr. 2006.

[158] M. Wachowiak, R. Smolikova, Y. Zheng, J. Zurada, and A. Elmaghraby. An approach to multimodal biomedical image registration utilizing particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, 8:289–301, June 2004.

[159] R. Wagner, S. Smith, J. Sandrik, and H. Lopez. Statistics of speckle in ultrasound B-scans. *IEEE Trans. Sonics. Ultrason.*, 30:156–163, 1983.

[160] M. Walker, L. Shao, and R. Volz. Estimating 3-D location parameters using dual number quaternions. *CVGIP: Image Understanding*, pages 358–367, Nov. 1991.

[161] Q. Wang, S. Takashima, F. Takayama, J. Wang, S. Kawakami, A. Saito, T. Matsushita, and S. Sone. Detection of occult metastatic lymph nodes in the neck with gray-scale and power doppler us. *Acta Radiol*, 42:312–319, May 2001.

[162] J. Weese, R. Göcke, G. P. Penney, P. Desmedt, T. M. Buzug, and H. Schumann. Fast voxel-based 2D/3D registration algorithm using a volume rendering method based on the shear-warp factorization. In *SPIE International Symposium on Medical Imaging*, 1999.

[163] W. Wein. Intensity based rigid 2D-3D registration algorithms for radiation therapy. Master's thesis, Technische Universität München, Dec. 2003.

[164] W. Wein, M. Blume, U. Leischner, H. Dodt, and N. Navab. Quality-based registration and reconstruction of optical tomography volumes. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

[165] W. Wein and A. Khamene. Image-based method for in-vivo freehand ultrasound calibration. In *SPIE Medical Imaging 2008, San Diego*, Feb. 2008.

[166] W. Wein, A. Khamene, D. Clevert, O. Kutter, and N. Navab. Simulation and fully automatic multimodal registration of medical ultrasound. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

[167] W. Wein, F. Pache, B. Röper, and N. Navab. Backward-warping ultrasound reconstruction for improving diagnostic value and registration. In *MICCAI 2006 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2006.

[168] W. Wein, B. Röper, and N. Navab. 2D/3D registration based on volume gradients. In *SPIE Medical Imaging 2005, San Diego*, Feb. 2005.

[169] W. Wein, B. Röper, and N. Navab. Automatic registration and fusion of ultrasound with CT for radiotherapy. In *MICCAI 2005 Proceedings*, volume 3750 of *Lecture Notes in Computer Science*, pages 303–311. Springer, Oct. 2005.

[170] W. Wein, B. Röper, and N. Navab. Integrating diagnostic B-mode ultrasonography into CT-based radiation treatment planning. *IEEE Trans. Med. Imag.*, 26:866–879, June 2007.

[171] G. Welch and E. Foxlin. Motion tracking: no silver bullet, but a respectable arsenal. *IEEE Computer Graphics and Applications*, 22(6):24–38, Nov. 2002.

[172] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1:35–51, 1996.

[173] T. Wood, D. Rose, M. Chung, D. Allegra, L. Foshag, and A. Bilchik. Radiofrequency ablation of 231 unresectable hepatic tumors: indications, limitations, and complications. *Ann Surg Oncol.*, 7:593–600, Sept. 2000.

[174] R. Wu, K. V. Ling, W. Shao, and W. S. Ng. Registration of organ surface with intraoperative 3D ultrasound image using genetic algorithm. In *MICCAI 2003 Proceedings*, pages 383–390, 2003.

[175] G. Xiao, J. Brady, J. Noble, M. Burcher, and R. English. Nonrigid registration of 3-D free-hand ultrasound images of the breast. *IEEE Trans. Med. Imag.*, 21:405 – 412, Apr. 2002.

[176] K. Yonetsu, M. Sumi, M. Izumi, M. Ohki, S. Eida, and T. Nakamura. Contribution of doppler sonography blood flow information to the diagnosis of metastatic cervical nodes in patients with head and neck cancer: assessment in relation to anatomic levels of the neck. *AJNR Am J Neuroradiol*, 22:163–169, Jan. 2001.

[177] R. Yuste and A. Konnerth, editors. *Imaging in Neuroscience and Development: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 2004.

[178] J. Zhang and A. Rangarajan. Multimodality image registration using an extensible information metric and high dimensional histogramming. In *IPMI 2005 Proceedings*, page 725–737, 2005.

[179] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *Int. J. of Computer Vision*, 13:119–152, Oct. 1994.

[180] H. Zhong, T. Kanade, and D. Schwartzman. Virtual touch: An efficient registration method for catheter navigation in left atrium. In *MICCAI 2006 proceedings*, pages 437–444, 2006.

[181] S. K. Zhou, J. Shao, B. Georgescu, and D. Comaniciu. Boostmotion: Boosting a discriminative similarity function for motion estimation. In *CVPR 2006 Proceedings*, volume 2, pages 1761–1768, 2006.

[182] D. Zikic. Variational deformable registration for medical applications. Master's thesis, Technische Universität München, Apr. 2006.

[183] D. Zikic, W. Wein, A. Khamene, D. Clevert, and N. Navab. Fast deformable registration of 3D-ultrasound data using a variational approach. In *MICCAI 2006 Proceedings*, volume 4190 of *Lecture Notes in Computer Science*, pages 915–923, 2006.