Statistical Intensity Prior Models with Applications in Multimodal Image Registration

Der Technischen Fakultät der
Universität Erlangen–Nürnberg

zur Erlangung des Grades

DOKTOR–INGENIEUR

vorgelegt von

Christoph Gütter

Erlangen — 2011

Deutscher Titel:
Statistische Modelle für Intensitäts-basiertes Vorwissen mit Anwendungen in der multimodalen Bildregistrierung

Als Dissertation genehmigt von der
Technischen Fakultät der
Universität Erlangen-Nürnberg

Tag der Einreichung:	19.07.2010
Tag der Promotion:	28.01.2011
Dekan:	Prof. Dr.-Ing. habil. R. German
Berichterstatter:	Prof. Dr.-Ing. J. Hornegger
	Prof. Dr. D. Cremers
	Prof. Dr. med. T. Kuwert

Acknowledgment

I would like to thank my research advisor, Prof. Joachim Hornegger, for his friendship, encouragement, and direction during the course of my research. From the very first moment, he has been an excellent source of ideas and stimulating discussions. I am very thankful to him for the opportunity to earn my degree in a truly mixed, industrial and academic, environment. I am very indebted to my research advisor at Siemens Corporate Research (SCR), Dr. Chenyang Xu, for his invaluable guidance in both academic and industrial topics that provided me with a manifold research education. His trust, constant motivation, and insights carried me through the ups and downs of research and decision making. I truly appreciate his caring supervision and protection as a supervisor and friend and all the inspiring brainstorming sessions that we had over the years. I am very grateful to Prof. Daniel Cremers for not only being a committee member but also for being an instrumental part in this thesis. His research expertise and enthusiasm greatly inspired me, and I truly appreciate working together. Our very visual and stimulating discussions that made the most complex problem seem so approachable are unforgettable. I am also very thankful to Prof. Torsten Kuwert for providing his medical expertise to this thesis and for being a committee member.

I am very grateful to Dr. Frank Sauer, leader of the Global Technology Field Medical Imaging, and Dr. Jens Guehring, Program Manager Interventional Imaging, for their continued support at SCR in pursuing this research. Among my close industrial and clinical collaborators, I would like to mention Dr. Hans Vija, Dr. Timor Kadir, Dr. Jerome Declerck, and Dr. Timm Dickfeld. Our joint projects further enhanced my knowledge about numerous product-related and clinical topics and this thesis would be incomplete without the tremendous amount of clinical image data provided by them. I was privileged to work with a number of great colleagues on exciting projects supporting my research, including Dr. Christophe Chefd'Hotel, Prof. Kazunori Okada, Dr. Dieter A. Hahn, Gabriele Wolz, Zhe Fan, Christoph Vetter, Kalpit Gajera, and Ponraj Chinnadurai.

I owe a tremendous amount of gratitude to my friends and unofficial thesis reviewers: Ponraj, Klaus J. Kirchberg, and Dr. Timo Kohlberger. Their input and constructive comments helped shape this thesis to its final form. All of my friends and colleagues whose support and encouragement have made my experience at the SCR research lab a memorable one far away from home and family.

Finally I would like to thank my parents Angelika and Stefan Gütter and my sister Juliane Gütter back home for the eternal support that I have received from them. Without the sacrifices they have made and the hardships that they went through, I would have never reached where I am at the moment. I would especially like to thank my wife Marcela Gütter who with all her patience, support, and love has given me the strength to achieve this goal. This dissertation is dedicated to them.

Christoph Gütter

Abstract

Deriving algorithms that automatically align images being acquired from different sources (multimodal image registration) is a fundamental problem that is of importance to several active research areas in image analysis, computer vision, and medical imaging. In particular, the accurate estimation of deformations in multimodal image data perpetually engages researchers while playing an essential role in several clinical applications that are designed to improve available healthcare. Since the field of medical image analysis has been rapidly growing for the past two decades, the abundance of clinical information that is available to medical experts inspires more automatic processing of medical images.

Registering multimodal image data is a difficult task due to the tremendous variability of possible image content and diverse object deformations. Motion patterns in medical imaging mostly originate from cardiac, breathing, or patient motion (i.e. highly complex motion patterns), and the involved image data may be noisy, furnished with image reconstruction artifacts, or rendered with occluded image information resulting from imaged pathologies. A key problem with methods reported in the literature is that they purely rely on the quality of the available images and have, therefore, difficulties in reliably finding an accurate alignment when the underlying multimodal image information is noisy or corrupted.

In this research, we leverage prior knowledge about the intensity distributions of accurate image alignments for robust and accurate registration of medical image data. The following contributions to the field of multimodal image registration are made. First, we developed a prior model called integrated statistical intensity prior model that incorporates both current image information and prior knowledge. It shows an increased capture range and robustness on degenerate clinical image data compared to traditional methods. Second, we developed a generalization of the first model that allows for modeling all available prior information and greater accuracy in aligning clinical multimodal image data. The models are formulated in a unifying Bayesian framework that is embedded in the statistical foundations of information theoretic similarity measures. Third, we applied the proposed models to two clinical applications and validated their performance on a database of approximately 100 patient data sets. The validation is performed using a systematic framework and we further developed a criteria for assessing the quality of non-rigid or deformable registrations.

The experiments on synthetic and real, clinical images demonstrate the superior performance, i.e. in terms of robustness and accuracy, of statistical intensity prior models to traditional registration methods. This suggests that fully automatic multimodal registration (i.e. rigid and non-rigid) is achievable for clinical applications. Statistical intensity prior models deliver great accuracy from a "relatively small" amount of prior knowledge when compared to traditional machine learning approaches that is appealing in both theory and in practice.

Übersicht

Die Herleitung von Algorithmen, die automatisch Bilder aus unterschiedlichen Aufnahmequellen registrieren können, man spricht auch von Multimodaler Bildregistrierung, ist ein fundamentales Problem von grosser Bedeutung für Forschungsgebiete in der Bildanalyse, der Computervision und der medizinischen Bildgebung. Insbesondere, die genaue Berechung von Deformationen in multimodalen Bilddaten beschäftigt fortwährend Wissenschaftler und spielt zur gleichen Zeit eine immens wichtige Rolle in verschiedenen klinischen Anwendungen, die zu einer höheren Qualität des Gesundheitswesens beitragen sollen. Da das Gebiet der medizinischen Bildanalyse in den letzen zwei Jahrzehnten rapide gewachsen ist, verlangt die Fülle der klinischen Informationen, die den Experten in der Medizin zur Verfügung stehen, nach mehr Automation der Algorithmen in der medizinischen Bildverarbeitung.

Die Registrierung von multimodalen Bilddaten in der Medizin ist eine komplizierte Aufgabenstellung aufgrund der hohen Variabilität der möglichen Bildinhalte und der Mannigfaltigkeit der vorkommenden Objektverformungen. Die Bewegungsmuster in der medizinischen Bildgebung begründen sich meist in Herz-, Atem-, oder Patientenbewegungen (d.h. sehr komplexe Bewegungsmuster), und die zugrundeliegenden Bilddaten können verrauscht sein, mit Bildrekonstruktions-Artefakten versehen sein, oder verdeckte Bildinformation, das aus einigen Krankheitsbildern resultiert, in der Modalität aufweisen. Ein Kernproblem der Methoden, die in der Fachliteratur vorhanden und bekannt sind, ist, dass sie auf die Qualität der vorhandenen Bilder angewiesen sind. Dadurch kann es schwierig werden, eine zuverlässige und akkurate Registrierung von verrauschten oder fehlerbehafteten, multimodalen Bilddaten zu erreichen.

In dieser Forschungsarbeit nutzen wir Vorwissen über die Intensitätsverteilungen von exakten vorhergehenden Registrierungen aus, um eine robuste und genaue Registrierung von multimodalen medizinischen Bilddaten zu erreichen. Die folgenden Beiträge zum Gebiet der multimodalen Bildregistrierungen werden von der vorliegenden Arbeit gemacht. Erstens, wir haben ein integriertes statistisches Model für intensitäts-basiertes Vorwissen entwickelt, das Bildinformationen der zu registrierenden Bilder und das Vorwissen miteinander verbindet. Das Model ermöglicht eine robustere Registrierung mit einem grösseren Erfassungsgebiet als gewöhnliche Methoden. Zweitens, wir haben eine Generalisierung dieses Modells entwickelt, das die Modellierung von verschiedenartigem Vorwissen erlaubt und damit eine höhere Genauigkeit in der Registrierung von klinischen, multimodalen Bilddaten aufweisen kann. Die Modelle sind in einem vereinigenden Bayes'schen System formuliert, das in die statistischen Grundlagen von Informationstheoretischen Ähnlichkeitsmaßen eingebettet ist. Drittens, wir haben die vorgestellten Modelle angewandt auf zwei klinische Anwendungen und deren Registrierungsqualitäten auf einer Datenbank von ungefähr 100 Patienten ausgewertet. Diese Auswertung wurde anhand eines systematischen Validierungssystem vorgenommen und im Zuge dessen haben wir weiterhin ein Kriterium zur Evaluierung von nicht-starrer oder deformierbarer Registrierung entwickelt.

Die Experimente auf synthetischen und realen, klinischen Daten zeigen ein besseres Ergebnis, d.h. in Bezug auf Robustheit und Genauigkeit, der statistischen Modelle für intensitäts-basiertes Vorwissen gegenüber gewöhnlichen Registrierverfahren. Unsere Ergebnisse zeigen auch, dass voll-automatische multimodale, starre sowie nicht-starre,

Bildregistrierung erreicht werden kann in klinischen Anwendungen. Statistische intensitäts-basierte Vorwissenmodelle liefern eine hohe Genauigkeit erzeugt von einer "relativ kleinen" Menge an Vorwissen im Vergleich zu traditionellen Verfahren auf dem Gebiet von machine learning: eine interessante Eigenschaft in der Theorie als auch in der Praxis.

Contents

List of Figures

List of Tables

Chapter 1

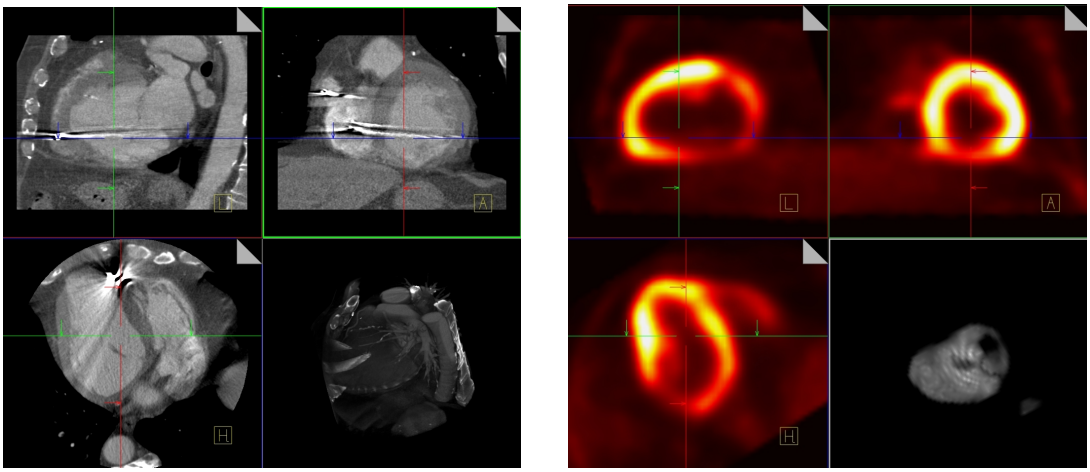
Introduction

Over the past two decades, there has been an increasing research activity in deriving algorithms that automatically align images being acquired from different sources. This task, also known as *multimodal image registration*, is a fundamental component for many active research areas in image analysis, computer vision, and medical imaging. Multimodal image registration is aimed at determining a spatial transformation that will bring homologous points in images from different domains into correspondence such that complementary information can be linked and compared correctly, e.g. for a fused visualization. The applications of multimodal image registration include sensor data fusion [Fish 01, Denz 07], medical image fusion in diagnostic imaging and medical image analysis [Bari 93, Loat 93, Bosc 03, Lee 05], image segmentation [Chrs 04, Dros 07], and image guided surgery [Grim 96, Rizz 97, Vogt 06, Tian 08, Wein 08].

Registering multimodal image data is a difficult task due to the tremendous variability of possible image content and diverse object deformations. For example, one important task in medical imaging is the registration of potential cardiac, breathing, and patient motion in 3-d myocardial positron emission tomography (PET) and computed tomography (CT) images (Fig. 1.1), where we are facing complex motion patterns, imaging noise, reconstruction artifacts, and large-scale image data sets. Imaging noise and artifacts especially may lead to failures of registration methods that are generally known to be robust. How to solve the multimodal registration problem on such degenerate clinical data is a challenging problem that a robust registration algorithm has to address.

1.1 Medical Image Analysis And Processing

One of the important tools in the field of medical imaging is the automated analysis of medical images. It allows for non-invasive inspection of the human bodies on the look for diseases or damages to internal structures. Medical image analysis provides effective diagnostic tools in medicine derived from data produced by imaging instruments that leverage physical phenomena such as X-rays, ultrasound, radioactivity, and magnetic resonance, and the medical imaging community is capable to probe into the structure, function, and pathology of the human body with a diversity of imaging systems. Data sets in two, three, and more dimensions convey increasingly vast and detailed information for clinical and research applications. In order to benefit health care,



(a) Cardiac CT (b) Cardiac PET

Figure 1.1: Three orthogonal views plus a 3-d volume rendering of a patient cardiac scan in two modalities: (a) CT, and (b) PET.

this information has to be analyzed and interpreted in a timely and accurate manner. Moreover, the imaging systems are also used for planning treatment and surgery. Fundamental classes of algorithms in medical imaging analysis are concerned with image enhancement, segmentation, quantification, registration, visualization, and the area of compression, storage, and communication. For a comprehensive introduction to the field of medical image analysis including the description of concepts and techniques for processing and analyzing medical images after they have been generated, we refer the interested reader to [Bank 09].

Image Segmentation, the separation of structures of interest from the background and from each other, is an essential class of algorithms in medical image analysis. The automated delineation of different components is used for analyzing spatial distribution of activity, anatomical structure and tissue types, and pathological regions. It can also be used as an initial step for visualization. Among the numerous applications of segmentation are lesion quantification, surgery simulations, surgical planning, measurement of tumor volume, and its response to therapy, study of brain development, image registration, atlas-matching, heart image extraction from cardiac cine data, and so on.

Image Registration, the alignment of images and structures to establish proper correspondence between the multiple medical images, needs to be identified before any type of medical image comparison can be performed. The automated registration of images and structures is used in image classification, motion estimation and analysis, template matching, and image reconstruction. Once proper correspondence is established, interesting applications can be built upon: e.g. (i) the fused visualization of complementary information, or (ii) the correction of motion that occurred between the two or more acquired images. These applications help address a wide range of medical questions including analyzing relationships between structure and function in complex organs such as the brain and the heart, full-body screening for lesions,

surgical planning and navigational guidance during minimally-invasive interventions, monitoring of treatment response in radiation treatment, image quality enhancement through reduction of artifacts induced by patient or organ motion, and many more.

Despite the variation of context in which medical image registration approaches have appeared in the literature, there are several commonalities between these approaches. The objective of pair-wise image registration is to find a transformation that transforms, in general, one of the images called *alignment image* so that it best matches the *reference image*. This transformation can be generated through manual, semi-automatic, or fully-automatic means, and belongs to a certain class of transformations dependent on the application. Section 2.1.3 portrays the most popular types of transformations that are sometimes categorized in rigid and non-rigid transformations based on the amount of flexibility that is offered by the transformation. How well a particular transformation matches the two images is described by the image similarity, and we point to the most popular measures in section 2.1.4. Registration approaches can further be categorized based on the type of sensors that were used to acquire the image data, and are called *monomodal*, when the images to be registered are generated by the same sensor type or scanner, or are called *multimodal*, when they originate from different scanners. Depending on the application, image registration can either be done using data sets from the same patient, *intrasubject*, or from patient populations or atlases, *intersubject*. Multimodal registration problems arise in both settings. For example, multimodal intersubject registration to a shape atlas can be employed to obtain segmentations of brain anatomy in functional images that are used for surgery planning [Rizz 97], and multimodal intrasubject registration can be used to correct for organ and patient motion in hybrid scanner image acquisitions [Guet 07] or to align multimodal data sets from different time points. One can imagine that multimodal image registration presents several challenges for registration methods due to its diverse nature. An interesting approach to relax this problem is to introduce prior knowledge about the correct alignment into the registration method. Prior information about the expected type of deformation or intensity mapping can be very helpful in solving the multimodal registration problem, and, so far, only a small amount of work has been performed along this direction. A detailed description of multimodal registration and state-of-the art methods is provided in chapter 2.

In conclusion, segmentation and registration are fundamental elements for medical image analysis tasks. In many applications, these approaches dictate the outcome of the entire analysis, since measurements and other processing steps are based on registered and segmented regions. Quantification algorithms are applied to segmented structures to extract essential diagnostic information such as shape, size, texture, angle, and motion. Moreover, both algorithms often play complementary roles in image processing. For example, if images are in correspondence and segmentation is known in one of them, it becomes easier to segment the other image. One example that greatly benefits from this complementary role play is cardiac segmentation from magnet resonance (MR) cine data as proposed in [Joll 10]. If segmentations are available in both images using the same set of segmentation classes, these corresponding classes can be used to identify the spatial transformation that would establish accurate correspondence. It is further possible to validate registration accuracy if the correct segmentation is known. In our clinical validation work, we demonstrate the accuracy

of our registration results by examining alignment with known segmentations and also show the added clinical value in knowing the correspondences between segmentations from multiple modalities.

In this dissertation, we focus on the introduction, analysis, and characterization of registration approaches that leverage prior knowledge models for the fully-automatic alignment of multiple complementary modalities and using a range of transformation models. We will present two notions of statistical intensity prior models that are intuitive and formulated in one unifying Bayesian framework, and we will show that these prior models promise to address a wide selection of the introduced clinical applications.

1.2 Medical Image Modalities

The proposed models presented in the following chapters are directly applicable to other research fields such as computer vision or image analysis, and they are designed to be applied to any type of image and modality combination. In the majority of this dissertation, however, we focus on the domain of medical imaging. Each type of medical image shows specific characteristics that play an important role in the clinical decision making. For example, cardiovascular function can be estimated from multiple modalities or a combination thereof. It appears that one logical future direction of clinical post-processing applications will be the combine-it-all strategy such that the physician can leverage all available information to make the most informed decision. Consequently, multimodal registration techniques will be required to robustly align images from a large spectrum of modalities. Figure 1.2 displays such a possible spectrum of modalities currently used for cardiovascular analysis. In order to allow the reader to understand the details of each modality and its characteristics, we briefly summarize the most prevalent image modalities in the medical imaging community. The information presented in the following sections is mostly gathered from [Kale 00, Kipp 04, Cher 03, http].

Tomographic Reconstruction

Tomographic images are 2-d representations of structures lying within a selected plane in a 3-d object. Tomographic reconstruction, first developed in diagnostic radiology and later extended to nuclear medicine, includes positron emission tomography, single photon emission computed tomography (SPECT), magnetic resonance tomography (MRT), and x-ray computed tomography in modern clinical systems. These systems use detectors placed or rotated around the object so that many different angular views, also known as *projections*, of the object are obtained. Mathematical algorithms then are used to *reconstruct* images from these projection data. The reconstruction of images from multiple projections of detected radionuclide emissions within the body is known as *emission computed tomography* (i.e. PET, SPECT), and reconstruction of images from transmitted emissions from an external source is known as *transmission computed tomography* (i.e. CT). For these modalities, the basic reconstruction algorithm is the same for both emission as well as transmission categories, but significant differences exist in the details of the implementations. MR relies on magnetic resonance effects

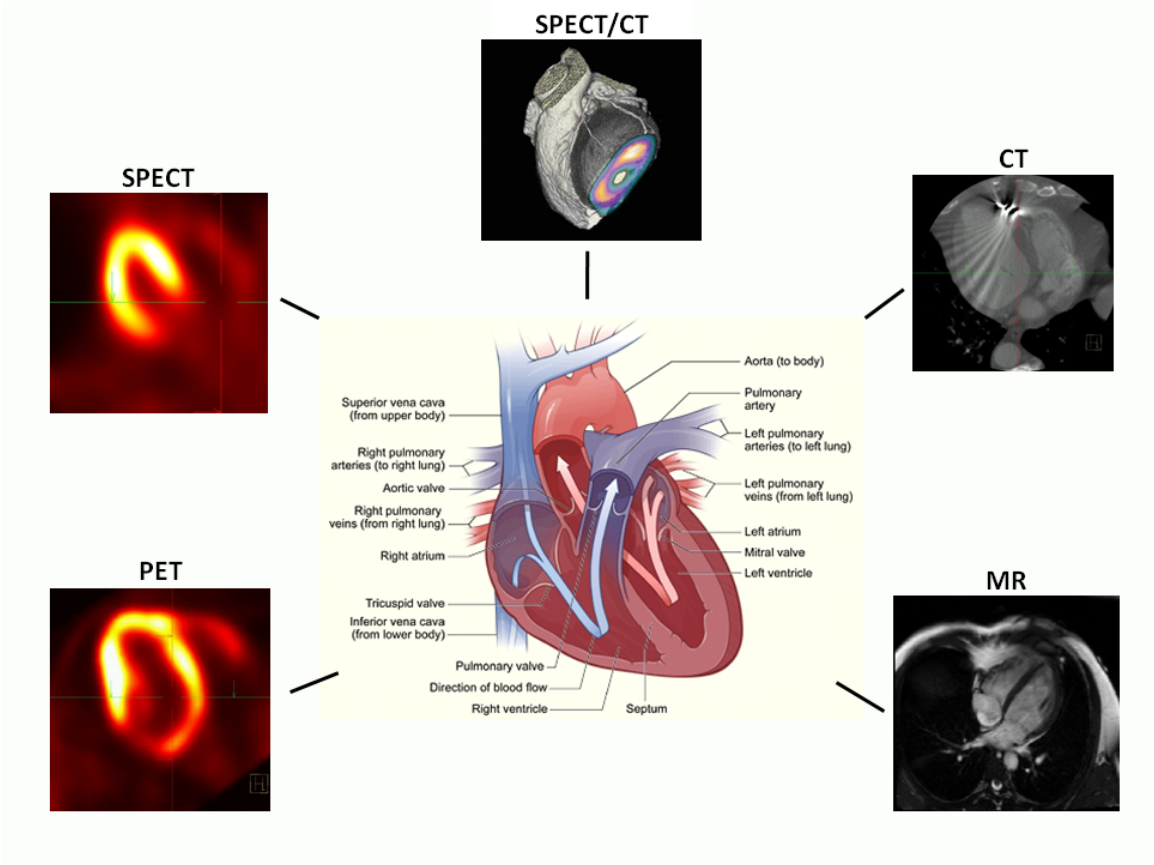


Figure 1.2: The electrical system of the heart sided by different clinical acquisitions involving the following examples of medical image modalities: SPECT/CT (fused rendering), CT, SPECT, PET, and MR.

inside the body but could be categorized as an emission tomography. However, the tomographic reconstruction in MRI as described in section 1.2.4 differs significantly from the other imaging modalities.

The following is a description of each individual image modality and its peculiarities that are directed at the tomographic reconstruction techniques.

1.2.1 Computed Tomography

The predominant image modality in our current work is computed tomography from x-ray images, which is generally referred to as CT. X-rays or Röntgen-rays, named after Wilhelm Conrad Röntgen, are a form of electromagnetic radiation that can penetrate solid objects and that is used to take images of the object's inside in diagnostic radiography. Since its introduction in the 1970s, CT has become an important tool in medical imaging to supplement X-rays and medical ultrasonography. The advantages of CT are a high contrast-to-noise ratio (CNR), image noise is predominantly influenced by physically generated quantum noise (i.e. fluctuations in the number of x-ray quanta registered by the detector), and that it results in high resolution images [Cher 03, Kale 00].

Independently of the parameters described in the previous paragraph, that exhibit an objectively measurable and quantifiable influence on the image quality, the CT imaging system may produce artificial structures that deviate from reality, i.e. also known as *artifacts*. These artifacts cannot always be determined objectively and it requires considerable experience on the part of the examiner, as well as a proper knowledge of the system's artifact behavior, to decide whether image information is artifacted or not. Important causes of artifacts in CT are patient/organ movement, beam hardening, scattered radiation, partial volume effects, sampling errors, exceeding the limits of the field of measurement, and metallic implants. Most of those causes can be addressed by system adaptations or stringent acquisition protocols. However, some of them such as patient/organ movement or metallic implants are simply unavoidable and could occur frequently in specific applications (e.g. pacemakers in cardiac perfusion imaging). The current recommendation for the examiner how to treat metallic implants is to select the slice orientation of the imaging system such that the implant is excluded from the scanned section as much as possible. These content extinguishing artifacts strongly deteriorate intensity-based registration algorithms, severely impact multimodal image alignment quality and therefore the results of the analysis.

The high-resolution CT images in medical imaging depict anatomical context in high quality. Therefore, it has been utilized in hybrid imaging systems to provide anatomical context for functional or nuclear imaging. We will introduce these systems and corresponding applications in chapter 5. Other popular applications of CT include CT angiography, dynamic (perfusion) CT, interventional CT, quantitative CT, and 4-d CT cardiac imaging.

1.2.2 Single Photon Emission Tomography

The approach of using a rotating gamma camera to acquire data for computed tomographic images, which is employed with radionuclides that emit single or multiple γ rays with no angular correlation, is known as *single photon emission computed tomography* or SPECT. Radionuclides are injected through a radiotracer into the patients body, and the scan has to be performed within the radiotracer's half life. The gamma camera acquires 2-d projection images by counting the activity of γ rays that reach its collimator detectors at equally spaced angular intervals around the patient. Typically, clinical SPECT images are reconstructed on a matrix of 64×64 or nowadays more commonly 128×128. These cross-sectional images are produced for all slices with the field-of-view (FOV) of the gamma camera, resulting in a stack of contiguous 2-d images forming a 3-d volume.

Ideally, the signal level for a voxel in a SPECT image would be linearly proportional to the amount of activity contained with the volume of tissue in the patient that corresponded to the location of that voxel. This would be useful not only for quantitative applications, such as perfusion studies, but also for visual interpretations of the images. In practice, however, this ideal result is not achieved because acquisition reality and idealized assumptions made for reconstruction do not meet. An extensive description of practical and theoretical considerations involved in SPECT imaging can be found in [Cher 03].

The quality of SPECT images heavily depends on the quality of attenuation and scatter corrections, partial-volume effects, spatial resolution, volume sensitivity, and most importantly on the fact that the underlying projection geometry is unknown. The limited resolution and low image quality in SPECT, for example as compared to PET, originates from the problem that it needs to be estimated with collimators. Collimators are designed as an array of tubes that end in a photon sensitive detector, i.e. once a photon hits a collimator tube it increases the count for this particular location. However, the angle at which a photon hits the collimator tubes cannot be determined, and, hence, the true originating location of that photon might never be known.

The key imaging feature of SPECT (and PET) is that radiotracers typically accumulate in specific organs or types of cells. This information is particularly important in the visualization of cancer where malignant cells may occur anywhere in the body. In the United States, SPECT imaging is most frequently used for studies of myocardial perfusion to assess coronary artery disease or heart muscle damage following myocardial infarction. It is common to perform cardiac perfusion studies both under resting conditions and also following a stress to the heart induced by exercise or by the injection of a drug that causes vasodilation. These are called *rest/stress studies*. Other important applications using SPECT are cerebral perfusion studies, tumor detection and localization in oncology, imaging of infection and inflammation, measurement of liver and kidney function, and using radiolabeled antibodies and peptides for tumor localization and ultimately for tumor therapy.

1.2.3 Positron Emission Tomography

A second form of tomographic nuclear medicine imaging, *positron emission tomography* or PET, uses radionuclides that decay by positron emission. This mode can be used only with positron-emitting radionuclides. The acquisition itself and general reconstruction algorithm is similar to that of SPECT, however, using different radiotracers and a key difference in the projection geometry calibration. PET detectors count annihilation photons that are produced when a positron interacts with an ordinary electron. Annihilation photons show back-to-back directional characteristics that are advantageous for the image reconstruction exploited by special annihilation-coincidence detector systems. PET has gained widespread clinical acceptance in recent years and is now firmly established in clinical nuclear medicine.

Due to the unique annihilation feature, PET systems, especially those employing multiple detector rings and multiring coincidence detection, have substantially higher detection efficiencies (by orders of magnitude) than what is achievable with typical SPECT systems. Therefore, PET images usually can be reconstructed with higher cut-off frequencies, and their final spatial resolution is generally superior to SPECT images. Despite a few additional correction strategies, attenuation correction is by far the largest single correction in PET, and image artifacts are of similar origin as in SPECT.

The major clinical applications for PET imaging are in oncology, neurology, and cardiovascular disease [Kipp 04]. The uptake of FDG-18 , ^{18}F-fluorodeoxy-glucose is the tracer of choice for most PET applications, reflects glucose metabolism that is regionally, i.e. in the affected tissue, altered by numerous pathologic conditions. A widespread application of PET imaging is the detection and staging of cancer. FDG-PET is, further, used diagnostically to evaluate myocardial viability and stratify patients with coronary artery disease with regard to bypass surgery.

1.2.4 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to visualize detailed internal structure and function of the body. MRI provides greater contrast between different soft tissues of the body than CT (introduced in section 1.2.1) does, making it especially useful in neurological (brain), musculoskeletal, cardiovascular, and oncological (cancer) imaging. Unlike CT, MRI uses no ionizing radiation. Rather, it uses a powerful magnetic field to align the nuclear magnetization of hydrogen nuclei atoms in water in the body. Radio frequency (RF) fields are used to systematically alter the alignment of this magnetization. This causes the hydrogen nuclei to produce a rotating magnetic field detectable by the scanner. This signal can be manipulated by additional magnetic fields to build up enough information to construct an image of the body. Magnetic resonance imaging is a relatively new technology with the first MR images being published approximately 30 years ago. By comparison, the first human X-ray image was taken in 1895.

Magnetic resonance imaging is based on the principles of nuclear magnetic resonance (NMR). It is known mostly to chemists that the term nuclear does not refer to radioactivity, and in order to avoid the negative connotations associated with the word nuclear for the general public, the term was dropped from the name of the imaging

technique in the late 1970's. Besides being mostly known as a tomographic imaging technique, MRI has throughout the years become a volume imaging method.

The imaging technique relies most frequently on the relaxation properties of excited hydrogen nuclei in the body's tissue. A superconducting external magnet in clinical MR systems generates a large magnetic field (i.e. typical magnetic field strengths are 1.5T and 3T (Tesla)) that aligns the spins of the atomic nuclei with non-zero spin numbers within the tissue of the to be imaged object parallel to the magnetic field - and produces a net magnetization. The magnetic dipole moment of the nuclei then experiences a phenomenon called precession due to the external magnetic field just like a spinning top interacts with gravity. The proton precessional frequency is determined from the Larmor equation, in which the frequency of precession is equal to a constant times the main magnetic field strength. The tissue is then exposed to a sequence of electromagnetic energy pulses (RF pulse). As energy is absorbed from the RF pulse, the net magnetization rotates away from the original longitudinal direction. The amount of rotation (termed the flip angle) depends on the strength and duration of the RF pulse. This transmitted RF pulse must be at the precessional frequency of the nuclei in order for resonance to occur and for efficient transfer of energy from the RF coil (i.e. RF pulse sender) to the nuclei. Depending on the type of RF pulse (i.e. 90° or 180°) the net magnetization gets rotated into the transverse plane or along the z-direction. The different RF pulses are important for the spin echo (SE) and gradient-recalled-echo (GRE) imaging techniques.

As the net magnetization rotates back to original alignment, the nuclei emit energy that can be measured and used to form the image. Due to its direction, the realignment with the external magnetic field is termed longitudinal relaxation or T1 relaxation. The rate at which this longitudinal magnetization grows back is different for protons associated with different tissues and is the fundamental source of contrast in T1-weighted images. T2-weighted imaging relies upon local dephasing of spins following the application of a transverse energy pulse (i.e. 90° RF pulse). The definition of T2 is the time that it takes for the transverse magnetization to decay to 37% of its original value. Both T1- and T2-weighted images are frequently acquired for most medical examinations. Often, a paramagnetic contrast agent (e.g. gadolinium) is administered, and both pre-contrast and post-contrast images are obtained. Contrast acquisitions can be useful to determine functional values such as myocardial blood flow through the left and right ventricle in myocardial perfusion imaging.

For further details, we refer the interested reader to a valuable introduction to the physics of an MRI scanner in [Pool 05] from the viewpoint of a radiologist.

1.3 Thesis Contributions

In this dissertation, we direct our attention to the problem of multimodal image registration due to its importance in medical image analysis and the complexity of this problem on routine clinical data. This complexity makes the robust and accurate, automatic estimation of non-rigid deformations an appealing problem in both theory and practice. The information gained from individual image modalities is unique and in the majority complementary to that of others. As a consequence, more clinical information and a higher diagnostic value for physician and patient is gained by

combining complementary modalities. In recent years, image fusion has been adopted for many clinical applications and systems to bring anatomical context into molecular, ultrasonic, magnetic-resonance, or low resolution morphological images. Accurate image registration is needed to ensure that fusion of multiple such modalities is correct as many analytic and diagnostic applications depend on the correspondence between the images.

The goal of this thesis is to derive statistical intensity prior models that are general and simple enough to allow for the seamless integration of prior knowledge from any number and type of image modality, and for the automatic processing of large numbers of clinical patient data while fulfilling the high clinical demands for speed, accuracy, and trivial human-computer interface interaction. The main contributions in this thesis are:

1. **Statistical intensity prior model:** A new statistical intensity prior model for multimodal image registration is proposed that combines prior knowledge derived from previously aligned image data with current observations. This integrated statistical intensity prior (SIP) model has three advantages compared to traditional registration methods:

   - It shows a larger capture range and allows for deformable registrations to be initialized far away from the optimal alignment.
   - It achieves accurate registration results by leveraging from both previous and current observations at almost no extra cost in computational performance.
   - This single prior mixture model is easy to add on to any existing intensity-based registration approach and is applicable to any modality combination allowing it to be used in a wide range of registration problems in computer vision and image processing.

   This method is applied to synthetic brain MR, and a large number of clinical PET/CT and SPECT/CT images.

2. **Generalized statistical intensity prior model:** Based on the integrated SIP model, a generalization (i.e. generalized SIP or GSIP) of the non-parametric prior is developed that spans an entire space of encoded prior knowledge. The GSIP model is given by a kernel density estimate on the space of joint intensity distributions computed from a representative set of previously aligned image pairs. The new prior allows for a comprehensive modeling of prior knowledge and an automatic selection of the best fitting prior distribution(s) for the current data set. A training step does not require any kind of supervision other than deciding that a given data set is correctly aligned. This method is applied to a large variety of multimodal medical images.

3. **Clinical validation and applications:** The proposed methods are systematically validated on extensive sets of clinical data and for different clinical applications. The validations are performed for both rigid and non-rigid transformation models. In addition, for the application of cardiac perfusion studies, a

new, clinically motivated metric for the validation of non-rigid image registration techniques is proposed. In summary, registration accuracy and robustness are significantly improved by employing statistical intensity prior models in more than 80 % of the cases.

1.4 Previous Publications

Portions of this thesis have been previously published. The integrated statistical intensity prior model was published for non-rigid and rigid transformations in [Guet 05, Guet 07] and applied to image guided surgery in [Liao 06]. The generalized statistical intensity prior model was published in [Crem 06b]. The implementation of both models on a Graphics Processing Unit (GPU) was published in [Vett 07, Fan 08], and the pre-clinical application of Ventricular Tachycardia using the integrated SIP registration was published in [Tian 08].

1.5 Thesis Organization

The thesis is organized as follows. In Chapter 2, we provide background materials on multimodal image registration and present an overview of related work on prior models. In Chapter 3, we develop the SIP non-parametric model and study its strengths and weaknesses on synthetic and simulated medical data. In Chapter 4, we develop the GSIP non-parametric model and apply it to medical and non-medical images. In Chapter 5, we present a systematic validation framework for multimodal registration for rigid as well as non-rigid transformation models, and validate SIP and GSIP on two clinical applications. Finally, we conclude the thesis in Chapter 6 with a summary and discussion of future directions.

Chapter 2

Background and Literature

Traditional image registration algorithms are typically designed using the following core components: interpolation, transformation, optimization, and similarity. Figure 2.1 displays the functional relationship of these core building blocks during the registration procedure. For better performance, i.e. increased robustness, handling of large deformations, and faster computational speed, the images to be registered are often processed through a multi-resolution image pyramid, i.e. resampling by n number of resolution levels. In each image resolution level, the registration problem can be solved by iterating through the following scenario. A similarity metric, that is been chosen to best describe the correct alignment of both images, is being evaluated between moving and fixed volume. An optimizer, then, uses the transformation model and somewhat intelligently determines the next best transformation. This transformation is applied to the moving image using an interpolator that approximates the true image function given the discrete nature of digitized images. The iteration scenario completes by evaluating the similarity between the newly transformed image and fixed image to verify whether an optimum has already been reached. After convergence the final transformation of the current resolution level is transformed to the next level, if necessary, and the registration procedure is repeated using the previously estimated transformation as initial transformation.

Although this is a strongly simplified description of medical image registration, it stressed the importance of all four core elements for the success of the registration algorithm. These building blocks should be formulated carefully with respect to the

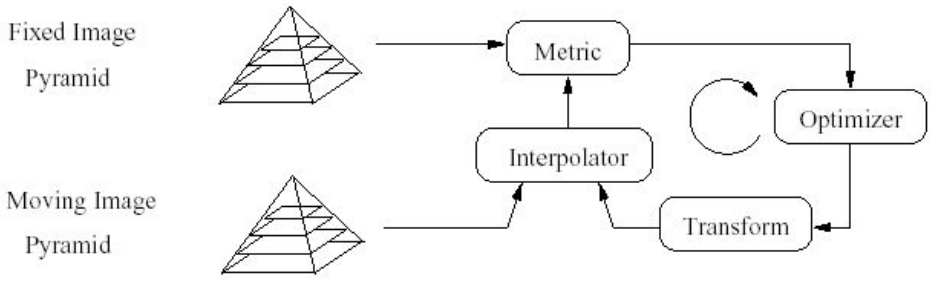


Figure 2.1: Building blocks of a typical registration algorithm.

underlying registration problem to be addressed. Hill and colleagues wrote an excellent introduction to the topic of medical image registration [Hajn 01].

2.1 Variational Image Registration

Multimodal image registration is the process of establishing correspondence between two or more images that have been acquired by multiple sensors. In the context of variational calculus, the image registration problem is formulated as a search problem in the domain of continous functions. Assuming the images acquired by the medical sensors are described by two functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ that are square-integrable and, hence, live in the infinite-dimensional $\mathbf{L}^2$ space, image registration is the search for a function $\boldsymbol{u} : \Omega \to \mathbb{R}^n$ that assigns each point $\boldsymbol{x}$ in Ω a displacement vector $\boldsymbol{u}(\boldsymbol{x}) \in \mathbb{R}^n$ aligning the two images accurately. Ω represents a bounded region of $\mathbb{R}^n$ and we further restrict ourselves to $n = 2, 3$ here. This function is searched in a set of admissible functions $\mathcal{F}$ such that it minimizes an energy functional of the form

$$E(\boldsymbol{u}) = \mathcal{I}_{\text{data}}(\boldsymbol{u}) + \alpha\, \mathcal{I}_{\text{smooth}}(\boldsymbol{u}). \tag{2.1}$$

Generally speaking, the set $\mathcal{F}$ is assumed to be a linear subspace of a Hilbert space H with the scalar product being denoted $\langle \cdot, \cdot \rangle_H$.

The term $\mathcal{I}_{\text{data}}(\boldsymbol{u})$ is designed to measure the *distance* between f_1 and the $\boldsymbol{u}$-deformed image $f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))$, essentially describing how similar f_1 and deformed f_2 are. The images f_1 and f_2 are further referred to as *reference* and *alignment* image, respectively. The term $\mathcal{I}_{smooth}(\boldsymbol{u})$ prevents fast variations of function $\boldsymbol{u}$ with α controlling the amount of this regularization. Then, the image registration problem is defined as the minimization of E with respect to $\boldsymbol{u}$ leading to a matching that represents the solution $\hat{\boldsymbol{u}}$ to the minimization problem:

$$\hat{\boldsymbol{u}} = \underset{\boldsymbol{u} \in \mathcal{F}}{\arg\min}\, E(\boldsymbol{u}) = \underset{\boldsymbol{u} \in \mathcal{F}}{\arg\min} \big(\mathcal{I}_{\text{data}}(\boldsymbol{u}) + \alpha\, \mathcal{I}_{\text{smooth}}(\boldsymbol{u}) \big). \tag{2.2}$$

2.1.1 Bayesian Formulation

The formulation of the image registration problem as a deterministic minimization of energies, as presented in (2.2), is paralleled by formulating the problem using the probabilistic framework of Bayesian inference. In this context, the problem of multimodal image registration can be solved by finding the most likely displacement field $\boldsymbol{u}$ given the two images $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))$. That is equivalent to maximizing the following posterior probability

$$\mathcal{P}\big(\boldsymbol{u} \,\big|\, f_1, f_2\big) = \frac{\mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}\big)\, \mathcal{P}\big(\boldsymbol{u}\,\big)}{\int_{\boldsymbol{u}} \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}\big)\, \mathcal{P}(\boldsymbol{u}) d\boldsymbol{u}} \tag{2.3}$$

given by the Bayes rule. After neglecting the factor in the above equation that does not depend on the displacement field $\boldsymbol{u}$ and thus do not affect the maximization, we receive the final expression of the conditional probability

$$\mathcal{P}\big(\boldsymbol{u} \,\big|\, f_1, f_2\big) \propto \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}\big)\, \mathcal{P}\big(\boldsymbol{u}\big). \tag{2.4}$$

The optimization problem in (2.3) separates into two factors that are interpreted as follows: The first factor represents the measurement likelihood, stating how likely the two images are given the correspondence induced by the displacement field $\boldsymbol{u}$. Images f_1 and f_2 are supposedly more "likely" the more similar they are made by the function $\boldsymbol{u}$. The second factor provides the a priori probability for the displacement field $\boldsymbol{u}$. The measurement likelihood can be regarded as a data term whereas the a priori probability can be understood as a geometric prior for the displacement field. This interpretation will reveal itself by looking at how these factors can be modeled.

We model the measurement likelihood for a general registration problem by stating that images f_1 and f_2 are more likely to be aligned, if they are more similar according to a specified criterion:

$$\mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}\big) \;=\; k \cdot \exp\Big(-\mathcal{I}_{\text{data}}\big(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u})\big)\Big) \tag{2.5}$$

where $\mathcal{I}_{\text{data}}$ needs to be strictly positive, i.e. $\mathcal{I}_{\text{data}}(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u})) \geq 0,\ \forall f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u})$ and $k > 0$. It evaluates the given data with respect to the registration problem and is chosen such that $\mathcal{I}_{\text{data}}(f_1, f_2, \boldsymbol{u}) \to \min \iff f_1(\boldsymbol{x}) = f_2(\boldsymbol{x}+\boldsymbol{u})$.

The geometric prior states an *a priori* likelihood for $\boldsymbol{u}$ describing which $\boldsymbol{u}$ is more or less likely. As described above, a common choice for non-rigid image registration formulations is to assume the underlying displacement field to be smooth:s

$$\mathcal{P}(\boldsymbol{u}) \;=\; \alpha \cdot \exp\big(-\mathcal{I}_{\text{smooth}}(\boldsymbol{u})\big), \tag{2.6}$$

where α determines the amount of smoothness that is imposed on the displacement field.

Maximizing the posterior distribution in (2.3) is equivalent to minimizing its negative log likelihood leading to the following minimization problem

$$E(\boldsymbol{u}) \;=\; -\log\big(\mathcal{P}\left(\boldsymbol{u}\,|f_1, f_2\right)\big) = \mathcal{I}_{\text{data}}(\boldsymbol{u}) \;+\; \alpha\,\mathcal{I}_{\text{smooth}}(\boldsymbol{u}) \tag{2.7}$$

$$\hat{\boldsymbol{u}} \;=\; \arg\min_{\boldsymbol{u}} E(\boldsymbol{u}), \tag{2.8}$$

with $\hat{\boldsymbol{u}}$ denoting that displacement field that corresponds to the Maximum A Posteori point of the posterior distribution in (2.3).

Comparing equations (2.1) and (2.7), it becomes clear that both problem formulations eventually lead to the same expression and are valid modeling assumptions. As a matter of fact, both frameworks have been successfully applied to numerous registration problems. Historically, the Bayesian framework has generally been used for parametric formulations due to its inherently natural handling of parameter optimization. Although it is not straightforward to use Bayesian modeling in a non-parametric setting, Bayesian inference presents an attractive theoretical framework in which our statistical prior formulations are nicely embedded. Besides providing a sound interpretation for the derivation of statistical prior models, the use of the Bayesian framework enabled us to formulate them in the first place. The detailed Bayesian formulations of the SIP models are presented in section 3.3 and 4.1.

2.1.2 Variational Formulation

Independent of the choice of the modeling framework, the calculus of variations can be used to determine the Euler-Lagrange equations of the specified energy functionals

E that reveals the minimizer $\hat{\boldsymbol{u}}$ in (2.8). Assuming that $E(\boldsymbol{u})$ is sufficiently regular, its first variation at $\boldsymbol{u} \in \mathcal{F}$ in the direction of $\tilde{\boldsymbol{u}} \in \mathcal{F}$ is defined by (see e.g. [Aube 02]):

$$\delta_{\tilde{\boldsymbol{u}}} E(\boldsymbol{u}) = \left. \frac{\partial E(\boldsymbol{u} + \epsilon \tilde{\boldsymbol{u}})}{\partial \epsilon} \right|_{\epsilon=0} \tag{2.9}$$

The gradient $\nabla_H E(\boldsymbol{u})$ of E is defined by requiring the following equality to hold for every $\tilde{\boldsymbol{u}} \in \mathcal{F}$:

$$\delta_{\tilde{\boldsymbol{u}}} E(\boldsymbol{u}) = \langle \nabla_H E(\boldsymbol{u}), \tilde{\boldsymbol{u}} \rangle_H \,. \tag{2.10}$$

If a minimizer $\hat{\boldsymbol{u}}$ of E exists, then the set of equations $\delta_{\tilde{\boldsymbol{u}}} E(\hat{\boldsymbol{u}}) = 0$ must hold for every $\tilde{\boldsymbol{u}} \in \mathcal{F}$. This is equivalent to $\nabla_H E(\hat{\boldsymbol{u}}) = 0$. These are called the Euler-Lagrange equations associated with the energy functional E. Due to an infinite-dimensional functional space where the optimization is carried out, there is usually, i.e. in non-academic scenarios, no closed-form solution available for solving the Euler-Lagrange equation. With the assumption, however, that the energy functional is locally convex, we can solve for the Euler-Lagrange equation by descending the gradient of the functional. If we have a suitable initial guess that is located in the vicinity of the global optimum, the gradient descent approach leads to an optimal solution. The solution of the Euler-Lagrange equation can be found by solving the following initial value problem:

$$\begin{cases} \frac{\partial \boldsymbol{u}}{\partial t} + \frac{\partial E(\boldsymbol{u})}{\partial \boldsymbol{u}} \;=\; 0, \\ \boldsymbol{u}_{t=0}(\cdot) \;=\; \boldsymbol{u}_0. \end{cases} \tag{2.11}$$

Provided the suitable initial guess u_0, the time-dependent, differentiable function $\boldsymbol{u} : [0, +\infty[\rightarrow L^2$ is being computed solving the partial differential equation above, and the asymptotic state when $t \rightarrow \infty$ is then chosen as the solution to the image registration problem given that $u_t \in \mathcal{F}$, $\forall t$. Consequently, the update rule in the k-th time step directly results from (2.11):

$$\boldsymbol{u}_{k+1}(\boldsymbol{x}) = \boldsymbol{u}_k(\boldsymbol{x}) - \tau_k \frac{\partial E(\boldsymbol{u}_k(\boldsymbol{x}))}{\partial \boldsymbol{u}_k(\boldsymbol{x})}, \tag{2.12}$$

where τ_k is the size of the applied time-step and $\boldsymbol{u}_k(\boldsymbol{x})$ the displacement field at time k for all locations $\boldsymbol{x} \in \Omega$.

The choice of a suitable smoothing term or geometric prior $\mathcal{I}_{\text{smooth}}$ is usually made based on the type of application. One example of a traditional smoothing term is the Tikhonov regularization [Tikh 77]

$$\mathcal{I}_{\text{smooth}} = \int |\nabla \boldsymbol{u}|^2 \, dx, \tag{2.13}$$

where $\nabla \boldsymbol{u} = \left(\frac{\partial \boldsymbol{u}}{\partial x_1}, \frac{\partial \boldsymbol{u}}{\partial x_2}, \ldots, \frac{\partial \boldsymbol{u}}{\partial x_n} \right)$, that has been pioneered by the seminal work of Horn and Schunck [Horn 81]. More sophisticated priors are conceivable, for example non-quadratic (robust) smoothness priors that allow for discontinuities in the estimated displacement fields (cf. [Brox 04]), an edge-based smoothness term that changes from an isotropic smoothing in homogeneous regions to anisotropic smoothing across contours of the image [Herm 02], a linear-elastic smoothing as described in [Chri 01], or a curvature regularization smoothness term as proposed in [Fisc 03].

In this thesis, we will in the majority focus on non-rigid transformations as defined by the displacement function $\boldsymbol{u} \in \mathbf{L}^2$. However, multimodal image registration in clinical practice, currently, relies on simpler, more restricted transformation models that are easier to compute, provide reasonably robust registration, but are not sufficient to describe motion patterns as they occur within the human body or between different subjects. The following section introduces the most common types of transformation models.

2.1.3 Types of Transformations

Let's recall that the images to be registered are defined by two functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$, where f_1 is the reference and f_2 the alignment image. The spatial transformation Φ_S describes a transformation that warps the alignment image onto the reference image, i.e. deforming the alignment image. The non-rigid transformation

$$\Phi_S(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}) \tag{2.14}$$

achieves this warping by assigning a displacement vector $\boldsymbol{u}(\boldsymbol{x})$ to each location $\boldsymbol{x} \in \mathbb{R}^n$ in the domain. During registration, the measurement likelihood function $\mathcal{I}_{\text{data}}$ is maximized with respect to a spatial transformation acting on the alignment image, i.e.:

$$\mathcal{I}_{\text{data}}(f_1(\boldsymbol{x}), f_2(\Phi_S(\boldsymbol{x}))) = \mathcal{I}_{\text{data}}(f_1(\boldsymbol{x}), f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))) \tag{2.15}$$

In this definition, Φ_S is an unconstrained, non-rigid transformation that is sometimes also referenced as non-parametric transformation model. Other types of transformations are combined in the class of parametric transformation models. In this class, two general types of transformations are usually distinguished: *rigid* and *non-rigid* transformations. Rigid transformations are defined as transformations that do not change the distance between any two points that undergo the transformation. In contrary, non-rigid transformations are capable of deforming the image locally.

Non-Rigid Transformation Models

Non-rigid transformations can be described by parametric transformation models with a high degree-of-freedom (DOF) such as transformations using radial basis functions (RBF), splines (i.e. Thin-Plate or B-Splines), or n-th order polynomials. Polynomial models are a linear combination of higher order terms to describe the non-rigid deformation, whereas RBFs and spline models describe the spatial transformation as a linear combination of basis functions θ_i, e.g.:

$$\Phi_S(x_1, x_2, x_3) = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} a_{00} & \dots & a_{0n} \\ a_{10} & \dots & a_{1n} \\ a_{20} & \dots & a_{2n} \end{pmatrix} \begin{pmatrix} \theta_1(x_1, x_2, x_3) \\ \vdots \\ \theta_n(x_1, x_2, x_3) \end{pmatrix}, \tag{2.16}$$

defined for 3-d transformations. As shown by the above equations such radial basis functions are spatially located in the image domain Ω describing the displacements at those locations. The values of the deformation in between the basis functions are interpolated depending on the type of basis function θ. The Thin-Plate spline (TPS) or B-Spline models are specific types of basis functions with the TPS incorporating geometric characteristics such as strain energy in the formulation.

Affine Transformation Models

The most restricted types of transformations are affine and rigid transformations since they are defining how a group of or all locations within the image domain are transformed by a small number of parameters. For example, the affine transformation model only allows for scale, shear, rotations, and translations within the linear transformation matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$:

$$\mathbf{\Phi_S}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} + \boldsymbol{t}, \tag{2.17}$$

with $\boldsymbol{t} \in \mathbb{R}^n,\ n = 2, 3$ describing the translations, and $\mathbf{A}$ defining shear, scale, and rotations.

Rigid Transformation Model

A transformation model is called rigid body transformation model, if the Euclidean distance between any two locations in the domain remains unchanged by the transformation. Therefore, the rigid transformation model allows only for rotations $\mathbf{R} \in \mathbb{R}^{n \times n}$, $\mathbf{R}$ being an orthonormal matrix with $\det \mathbf{R} = 1$, and translations $\boldsymbol{t}$:

$$\mathbf{\Phi_S}(\boldsymbol{x}) = \mathbf{R}\boldsymbol{x} + \boldsymbol{t}. \tag{2.18}$$

In the case of 3-d transformations, one way of describing the rotation matrix is to use rotation angles $\mathbf{R}(\alpha, \beta, \gamma)$ around the three coordinate axes:

$$\mathbf{R} = \begin{pmatrix} \cos\alpha\cos\beta & \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{pmatrix}.$$

A more robust and convenient mathematical notation for representing rotations in 3-d are unit quaternions. Compared to the Euler angles, as presented above, they are simpler to compose and avoid the problem of gimbal lock. See [Altm 86] for a detailed explanation of rotations and quaternions.

The advantage of affine and rigid transformation models are that they defined by only a small number of parameters. That provides a fast but simplified solution to the registration problem. Moreover, restricted parametric transformation models provide a greater numerical stability and appear more intuitive and understandable to the clinical end user. Therefore, affine and rigid models are prevalent and can be considered state-of-the-art methods in clinical applications despite their disability to accurately capture the underlying deformations.

The majority of this dissertation is concerned with the non-parametric registration of medical imaging data sets. For convenience, we will use the notation of the displacement field $\boldsymbol{u}$ rather than the more general transformation model Φ_S when refering to the transformation model.

2.1.4 Measures of Similarity

The central component of a registration algorithm is the evaluation of the image alignment quality. A set of mathematical functions called objective or similarity

functions is specifically designed to measure the quality of a particular alignment. Depending on the context, these functions are sometimes also referred to as cost or energy functions. During registration an objective function produces a numerical value that is used in the optimization procedure (see also Fig. 2.1). Similarity functions can be categorized in *feature-based* and *intensity-based* functions.

Feature-based similarity functions measure image similarity given a set of features in both images such as points, landmarks, edges, surfaces, saliency, and many more. The features are either automatically extracted or provided by the user. After feature correspondences between the two sets of features have been established, the matching problem reverts to generating a transformation from these correspondences, either directly or through interpolating using the underlying transformation type. An automatic way of establishing correspondences between features is to employ a feature-based registration algorithm (e.g. landmark-based, point-based, or surface-based registration) using either parametric or non-parametric transformation models [Fitz 98, Josh 00, Rohr 01, Hill 01, Guet 03].

Intensity-based similarity functions compute similarity by directly involving the images' intensities and are also refered to similarity measures. There are several popular intensity-based similarity metrics used in medical image registration, since they are very suitable for fully automatic registration tasks. The different measures assume each a specific relationship between the intensity values of the images to be compared. Intensity-based measures can be further categorized in two subclasses: direct and indirect similarity measures. Direct similarity measures compute similarity by directly operating on the image intensities. Indirect measures, however, are calculated on the basis of statistics such as comparing the intensity distributions of the images.

A simple direct similarity measure is the *Sum-of-Squared-Differences (SSD)* that assumes the image intensities to be constant over time, i.e. between the compared images. It can be written as

$$\mathcal{I}_{\mathrm{SSD}}(f_1(\boldsymbol{x}), f_2(\Phi_S(\boldsymbol{x}))) = \frac{1}{|\Omega|} \int_{\Omega} \left(f_1(\boldsymbol{x}) - f_2\left(\Phi_S(\boldsymbol{x})\right)\right)^2 d\boldsymbol{x} \tag{2.19}$$

This intensity constancy assumption may not always hold in practice due to image noise and artifacts. The quadratic influence of such imperfections on the similarity measure can be reduced if one applies the absolute distance instead. The *Sum-of-Absolute-Differences (SAD)* is defined as

$$\mathcal{I}_{\mathrm{SAD}}(f_1(\boldsymbol{x}), f_2(\Phi_S(\boldsymbol{x}))) = \frac{1}{|\Omega|} \int_{\Omega} \left|f_1(\boldsymbol{x}) - f_2\left(\Phi_S(\boldsymbol{x})\right)\right| .d\boldsymbol{x} \tag{2.20}$$

Both measure perform well on images that were acquired by similar sensors. However, even in the case of monomodal images, the constancy assumption might be violated due to structures moving in and out of the image plane, e.g. in cardiac cine MRI. A robust alternative in this case is provided by other direct measures such as *Cross-*

Correlation (CC), or *Local Cross-Correlation (LCC)*. The LCC measure in the discrete case is defined as

$$\mathcal{I}_{\mathrm{LCC}}(f_1(\boldsymbol{x}), f_2(\Phi_S(\boldsymbol{x}))) = \sum_{i,j}^{N_{nb}} \frac{\sum_{nb}(f_1(i,j)-\bar{f}_1)(f_2(i,j)-\bar{f}_2)}{\sqrt{\sum_{nb}(f_1(i,j)-\bar{f}_1)^2 \sum_{nb}(f_2(i,j)-\bar{f}_2)^2}}. \tag{2.21}$$

These correlation measures assume a linear relation between the images and have proven to robustly deal with images that show slight intensity variations such as perfusion or T1-/T2-weighted MRI imaging.

In case the images have been acquired by different sensors, the intensity constancy or linearity cannot be assumed anymore. Then, we need to revert to indirect statistical approaches to evaluate the image alignment. These approaches are based on information theory and statistics and we will refer to them as *statistical measures* in the remainder of this document. Probabilistic measures such as

- Maximum *Joint Entropy (JE)*, proposed by Woods *et al.*. [Wood 92] and Hill *et al.* [Hill 93],
- Maximum *Mutual Information (MI)*, independently proposed by Viola and Wells *et al.* [Viol 95] and by Collignon and colleagues [Coll 95],
- Maximum *Entropy Correlation Coefficient (ECC)*, suggested by Collignon [Coll 98] and Maes [Maes 97], and
- The generalized *Correlation-Ratio (CR)*, proposed by Roche *et al.* [Roch 98],

are the most influential in multimodal, medical image registration.

The key idea of the criterion of maximum mutual information, for example, is to find a deformation field that maximizes the statistical dependency between two images where the intensities are considered samples from two random variables:

$$\mathcal{I}_{\mathrm{MI}}\big(f_1(\boldsymbol{x}), f_2(\Phi_S(\boldsymbol{x}))\big) = \int_{\mathbb{R}^2} p_{\Phi_S}(i_1,i_2) \log \frac{p_{\Phi_S}(i_1,i_2)}{p_{f_1}(i_1)p_{f_2}(i_2)} di_1 di_2, \tag{2.22}$$

where $i_1 = f_1(\boldsymbol{x})$, $i_2 = f_2(\Phi_S(\boldsymbol{x}))$, and $p_{f_1}(i_1)$, $p_{f_2}(i_2)$, $p_{\Phi_S}(i_1,i_2)$ are the marginal and joint intensity distributions estimated from $f_1(\boldsymbol{x})$ and $f_2(\Phi_S(\boldsymbol{x}))$. Statistical dependency is measured by the Kullback-Leibler (KL) divergence between the joint intensity distribution of the two images and the product of the marginals. After the MI criterion was adopted for medical image registration, it was further modified with regards to normalization [Maes 97, Coll 98, Stud 99] and has proven to be a robust similarity measure for rigid multimodal registration problems.

Despite the success of the MI measure, it can be shown that it strongly depends on the overlap domain of the images. This is particularly disturbing in medical images where it is common to experience a large homogenous background that negatively influence the mutual information measure. This effect is amplified for the maximum joint entropy measure, and, hence, evident for MI. Studholme *et al.*circumvent this

problem by proposing a *normalized mutual information* measure [Stud 99] that is less sensitive to changes in the image overlap through dividing the marginal entropies by the joint entropy:

$$\mathcal{I}_{\mathrm{NMI}}(f_1, f_2) = \frac{\mathcal{H}(f_1) + \mathcal{H}(f_2)}{\mathcal{H}(f_1, f_2)}, \tag{2.23}$$

where $\mathcal{H}(f_1), \mathcal{H}(f_2)$ are the marginal entropies of f_1 and f_2, with the entropy being defined as

$$\mathcal{H} = \int_{\mathbb{R}} p(i) \log p(i) di \tag{2.24}$$

, and $\mathcal{H}(f_1, f_2)$ denotes the joint entropy

$$\mathcal{H} = \int_{\mathbb{R}^2} p(i_1, i_2) \log p(i_1, i_2) di_1 i_2. \tag{2.25}$$

They found a distinct improvement in the behavior of this normalized measure for rigid registration methods. Another form of normalized mutual information is the entropy correlation coefficient (ECC) [Maes 98] that is related to NMI in the following manner

$$\mathcal{I}_{\mathrm{ECC}}(f_1, f_2) = 2 - \frac{2}{\mathcal{I}_{\mathrm{NMI}}(f_1, f_2)} \tag{2.26}$$

For more information on probabilistic similarity measures in medical image registration we refer the interested reader to a number of survey papers [Brow 92, Main 98, Roch 00, Hill 01].

The registration of functional and morphological images, however, turns out to be rather challenging due to their imaging dichotomy. Routine clinical applications currently rely on rigid-body registration algorithms that optimize on low-level image information. Therefore, the established alignment arises purely from the matched volumes and strongly depends on the available image information.

The statistical intensity prior models that are introduced later are based on the same theoretical principles as statistical similarity measures leveraging on their strengths and improving on their weaknesses.

2.2 Related Work

Learning from previous experience has been playing a key role in the advancements of man kind and it has proven beneficial to apply previously gathered knowledge to current tasks. The field of *machine learning* is concerned with the research and development of algorithms that allow computers to evolve behaviors based on empirical data such as sensor data. The major focus of the research is to automatically learn to recognize complex patterns and make intelligent decisions based on the given data. In general, an extensive amount of data is required to learn such complex patterns. A slightly different approach is to utilize knowledge from previous observations enhance purely data driven image processing algorithms. In other words, prior knowledge is providing additional information through constraints or other means to algorithms optimizing on the current data. We refer to those approaches as *prior models* rather than *learning-based processing* to avoid confusion with machine learning. Though

concepts from machine learning have been applied to image registration, for example, by Zhou *et al.* [Zhou 06] who use classification of learned image patches to align ultrasound images, but the question of (clinical) practicality of such approaches always arises when compared to prior models due to an extensive and elaborate training step. The introduction of prior knowledge has greatly enhanced numerous purely low-level driven image processing algorithms and, hence, benefited many image analysis tasks. The reason is that prior knowledge constrains the solution space for challenging problems to a meaningful subset. For example, constraining image segmentation algorithms by an elaborate formulation of shape priors has significantly improved segmentation accuracy and robustness, especially on noisy and corrupted data as they appear in many practical applications [Rous 02, Chan 05, Crem 06a, Kohl 06]. In general, such methods are formulated so that the optimizing function includes a data term and a shape prior term modeling previously observed shapes.

Deformation Priors

In the context of image registration such prior knowledge is most closely related to *statistical deformation models (SDM)* that compute deformation priors from non-rigid registrations of medical image data and utilize statistics (e.g. principal component analysis) to encode the prior knowledge [Ruec 03, Xue 06]. The concept of SDMs is similar to active shape models, which capture statistical information about the shapes across a population, while utilizing information from the dense deformation field. Another direction of incorporating deformation priors in deformable registration algorithms is to make very specific assumptions on how to impose regularization on the deformation field $\boldsymbol{u}$. Beyond a number of fairly sophisticated regularity constraints such as non-quadratic smoothness priors [Weic 01], a linear-elastic smoothness prior [Chri 01, He 03], or a curvature regularization smoothness term [Fisc 03], Roth and Black [Roth 05] suggested to learn the statistics of optical flow from training sequences using Markov-Random-Field cliques and to subsequently impose these as smoothness priors for variational motion estimation.

Nevertheless, all these techniques implicitly or explicitly (e.g. through the need of segmentation) make specific assumptions about and are intertwined with the occurring shapes. Consequently, in data that includes undefined or highly varying shape boundaries and objects, deformation patterns may be irregular and therefore difficult to impose.

Intensity Priors

An intriguing and promising direction of research is to learn and impose statistical intensity priors not on the displacement field $\boldsymbol{u}$, but on the simultaneously estimated joint intensity distribution p_u (see eq. (3.1)), which characterizes the intensity transformation between two images. In contrary to deformation priors, this approach is independent of occurring shape patterns. Intensity priors are more readily obtained and can be applied to a wider range of problems than deformation-based prior knowledge, which makes it an appealing approach both in theory and practice.

In the context of medical image registration, the first pursuits using intensity priors involved ideas on optimizing similarity of the current observations to a single

instance of prior observations for rigid as well as non-rigid transformation models. Leventon and Grimson [Leve 98] suggested to rigidly register multimodal images by optimizing on the currently observed joint intensity distribution until it resembles a prior one. Chung et al. [Chun 02] found empirically that utilizing the Kullback-Leibler (KL) divergence for distance computation between distributions is superior to the log likelihood used by Leventon and Grimson. Zöllei et al. [Zoll 03] showed that Chung's method makes some implicit assumptions about the desired solution which do not always hold suggesting that the registration result may be biased towards the given prior. Hermosillo *et al.* [Herm 02] formulated the supervised learning approach of Leventon and Grimson using the maximum likelihood criterion for variational multimodal registration.

Although these approaches increase the capture range of multimodal registration methods, the registration accuracy clearly depends on the quality of the provided single instance of prior knowledge.

Generalized Prior Models

In [Guet 05], we further developed the idea of utilizing prior knowledge by proposing to compare to previously observed intensity mappings while simultaneously optimizing on the similarity of the currently matched image data. This concept is similar to the one used with shape priors where a data term is combined with a prior term. Our previous, rather ad-hoc, introduction of this idea in [Guet 05] finds further theoretical justification in a Bayesian inference formulation that is derived in chapter 3. The idea of leveraging prior and observed samples of the joint intensity distribution for registration has also been employed by Sabuncu *et al.* [Sabu 05, Sabu 08] who construct and optimize on an Euclidean Minimum Spanning Trees (EMST) that encode the image samples. However, the EMST gradients for non-rigid transformation models remain to be derived. Specialization with regards to certain clinical applications have been proposed by Gholipour *et al.* [Ghol 07] who employ the joint intensity distribution for the non-parametric registration of echo-planar and structural magnetic resonance brain images, and us [Guet 07] suggesting a weighing scheme on the prior distribution to favor intensity mappings that belong to a specific organ in rigid multimodal registration.

Despite the increased accuracy delivered by such methods, they are limited to a *single* prior joint intensity distribution or single prior data set. In [Crem 06b], we address this problem by formulating a non-parametric prior model sampling an entire space of joint intensity distributions for deformable image registration. The statistical foundation for such a general prior model lies in the theory of Gaussian processes: the basis for probability distributions on infinite-dimensional function spaces. Gaussian processes perform well on a large range of data sets, which is what we are interested in. Depending on the amount of available prior data, this non-parametric prior model allows for an accurate modeling of the space of admissible joint intensity distributions. An interesting comparable prior model was later formulated as a parametric prior of a collection of joint intensity mappings using Dirichlet priors, i.e. the distribution of distributions, for rigid and non-rigid image registration by Zöllei *et al.* [Zoll 06b, Zoll 07]. The core of this dissertation describes a statistical framework that unifies a formulation for optimizing on observed data, single instance prior, or an entire "library" of prior

knowledge in form of joint intensity distributions, i.e. extending our published work. We, further, investigate the relationship between these different prior models. In chapter 4, the non-parametric, generalized prior intensity model is presented that spans the space of all admissible joint intensity distributions.

Recently, plenty of interest has been generated in using prior models to address the challenging problems in non-rigid multimodal image registration. Retrieving complex non-rigid deformations accurately and robustly in multimodal data sets presents numerous challenges to registration algorithms. This is due to the fact that handling complex, non-rigid deformations involves estimating the optimal transformation in a suitable, infinite-dimensional functional space [Ruec 98, Gaen 98, Herm 02]. Depending on algorithm-related choices, the solution space could be filled with many locally optimal solutions. Therefore, the design of non-rigid, multimodal registration algorithms has attracted much research attention in recent years including general approaches as in [Gaen 98, Chef 01, Roge 03] or specific clinical applications such as [Cama 03, DAgo 03, Ghol 07]. The resulting algorithms, however, are staying largely unnoticed in routine clinical applications as they fail to suffice clinical demands for accuracy, robustness, and computation time. This leads to compromises in the employed registration such as reverting to simpler transformation models [Bank 09].

Interestingly, common clinical motion patterns are of non-rigid nature originating from patient, organ, and respiratory motion. Therefore, the third core component of this dissertation is concerned with systematic validation of registration algorithms on clinical data sets resulting in a concise presentation of the proposed prior models' practical value. This is, up to our knowledge, one of few validation studies that investigate the performance fully-automatic multimodal registration algorithms on a large set of routine clinical image data.

Chapter 3

Statistical Intensity Prior Model

It is known that traditional intensity-based registration methods have problems associated with initialization and poor performance on routine clinical image data that are artifacted or noisy. In this chapter, we present a new combined model of current and prior statistical observations that addresses both issues listed above. This model, which we call *integrated statistical intensity prior* model, is the limit case of a non-parametric prior model on the space of joint intensity distributions. The theoretical foundations for probability distributions on infinite-dimensional function spaces is provided by the theory of Gaussian processes. We will call the joint intensity distribution that is derived from a previously aligned image data set *statistical intensity prior* (SIP). The statistical intensity prior formulation is distinguished from nearly all previous prior formulations in that it is not making any modeling assumption on the underlying probability distributions of the prior knowledge. Because of this, we will specify a solution using the calculus of variations for the optimization in functional spaces.

The space of statistical intensity priors is infinite-dimensional in the number m of described prior data sets; however, in this chapter we assume only one prior joint intensity distribution. At first glance, it might seem contra-intuitive to distract our attention with a bound prior model. Nevertheless, deriving prior and posterior probability distributions from little prior knowledge is of significant practical importance. One of the biggest issues in clinical applications remains the availability of high-quality prior data sets for training. Previous work on prior modeling for registration [Sabu 05, Zoll 06a, Zoll 07] does not include discussions on practical aspects such as model performance for restricted availability of prior clinical data sets. Zöllei *et al.* [Zoll 07] presented a parametric prior model that uses the Dirichlet distribution as a conjugate prior on multinomial prior distributions. Furthermore, three interesting categories of prior confidence (i.e. no, medium, and full) are thoroughly analyzed w.r.t. the relationship of Dirichlet encoded priors and current statistical methods. Despite this stimulating analysis, it would be interesting to correlate these levels of confidence with clinical problems in order to provide the clinician user with an intuition of how to map confidence levels with regards to the available prior data.

The goal of this chapter is to derive a SIP model that improves on traditional multimodal registration techniques while utilizing limited amounts of prior knowledge. We further believe that a systematical validation of multimodal registration techniques

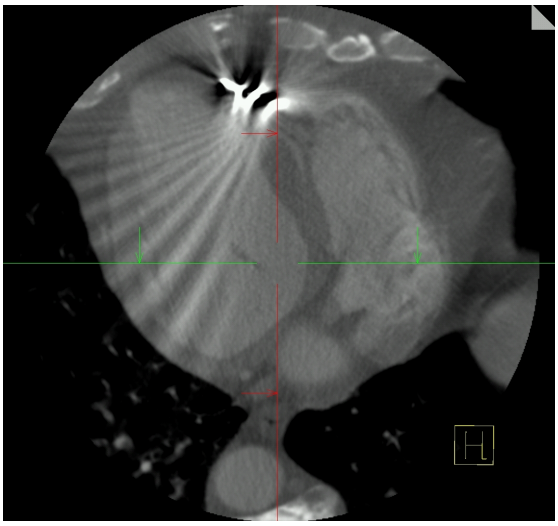
on complementary multimodal data sets (e.g. an anatomical / functional modality combination) has not been performed yet. The goal of this thesis is to derive a non-parametric prior on the infinite-dimensional function space of prior distributions and to show its applicability in clinical environments by systematically validating this model for multiple clinical applications. As it turns out in our model, the posterior distribution of the registration problem for the limit case $m = 1$ becomes a compromise between data and prior information resulting in an *integrated SIP* model.

Particular advantages of the integrated SIP over traditional intensity-based registration methods are an increased robustness to initialization (i.e. capture-range) and an improved performance on artifact and noisy image data. As we show in the following sections, the integrated prior allows for initializations far away from the correct solution and even for initializations that create ambiguous matching situations. The ambiguity is resolved by providing the respective statistical intensity prior. Image noise, artifacts, or morphological changes in the image data are not part of the expected joint distribution, and as a consequence they are masked out by the integrated SIP model to achieve the desired alignment. As opposed to learning based approaches to image registration that are required to train a large number of features, image data, or pooled samples in order to achieve accurate alignment, there is no cumbersome training step required for the integrated SIP model. As we will present in our experimental chapter 5, the proposed prior model serves as a practical, efficient, and robust model that constitutes a crucial building block of the comprehensive non-parametric prior on the space of statistical intensity priors.

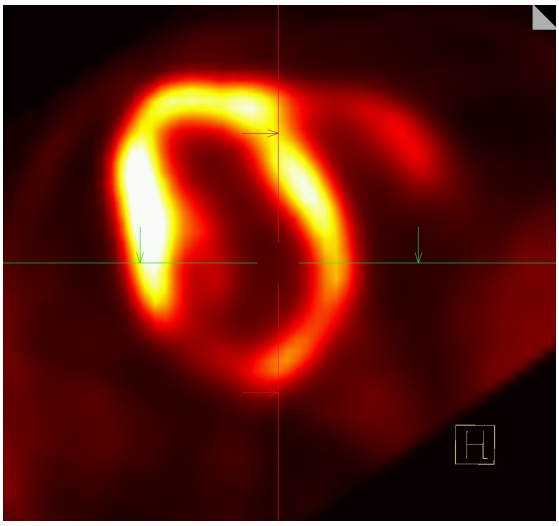
In medical image registration, the choice of the transformation model impacts the complexity of the solution space and, hence, the achievable quality of alignment. We will formulate the integrated statistical prior model for the most general case of dense deformations in order to benefit a wide range of clinical applications. However, the model is also applicable to constrained transformation models such as rigid transformations. For certain clinical applications, it may further be desirable to adapt the statistical intensity prior such that a constrained alignment of particular image regions (e.g. specific organs) can be ensured. In section 3.4, the derivation for such a constrained prior and an example application is presented.

3.1 Behavior of Traditional Registration Methods

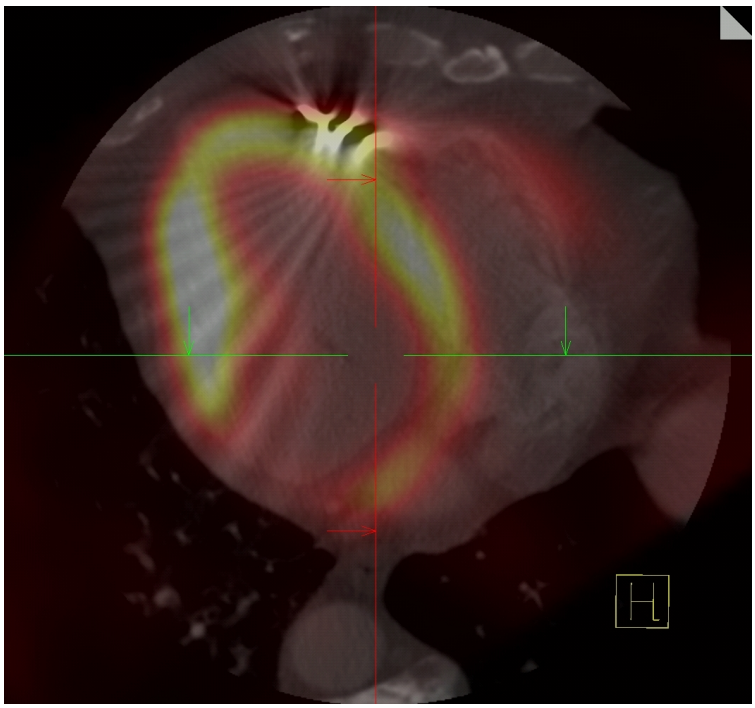
Although traditional registration methods have proven to be fairly robust in many clinical applications [Plui 03], they rely on purely data-driven criteria such as the previously introduced intensity-based similarity measures. As a consequence, registration results will deteriorate in case the low-level image information is corrupted due to noise, occlusion, or image reconstruction artifacts. In addition, certain multimodality combinations show little morphological and anatomical commonalities, which is also why the fusion of such modalities (e.g. PET/CT, SPECT/CT, PET/MR, or SPECT/MR) is of high clinical importance. Traditional similarity measures like normalized mutual information (NMI), entropy correlation coefficient (ECC) or correlation ratio (CR) are then computed from noisy estimates of the respective underlying probability distributions. Consequently, little work can be found that reports the accuracy of traditional registration methods on clinical data that is degenerate. Although such



(a) Computed Tomography



(b) Positron Emission Tomography



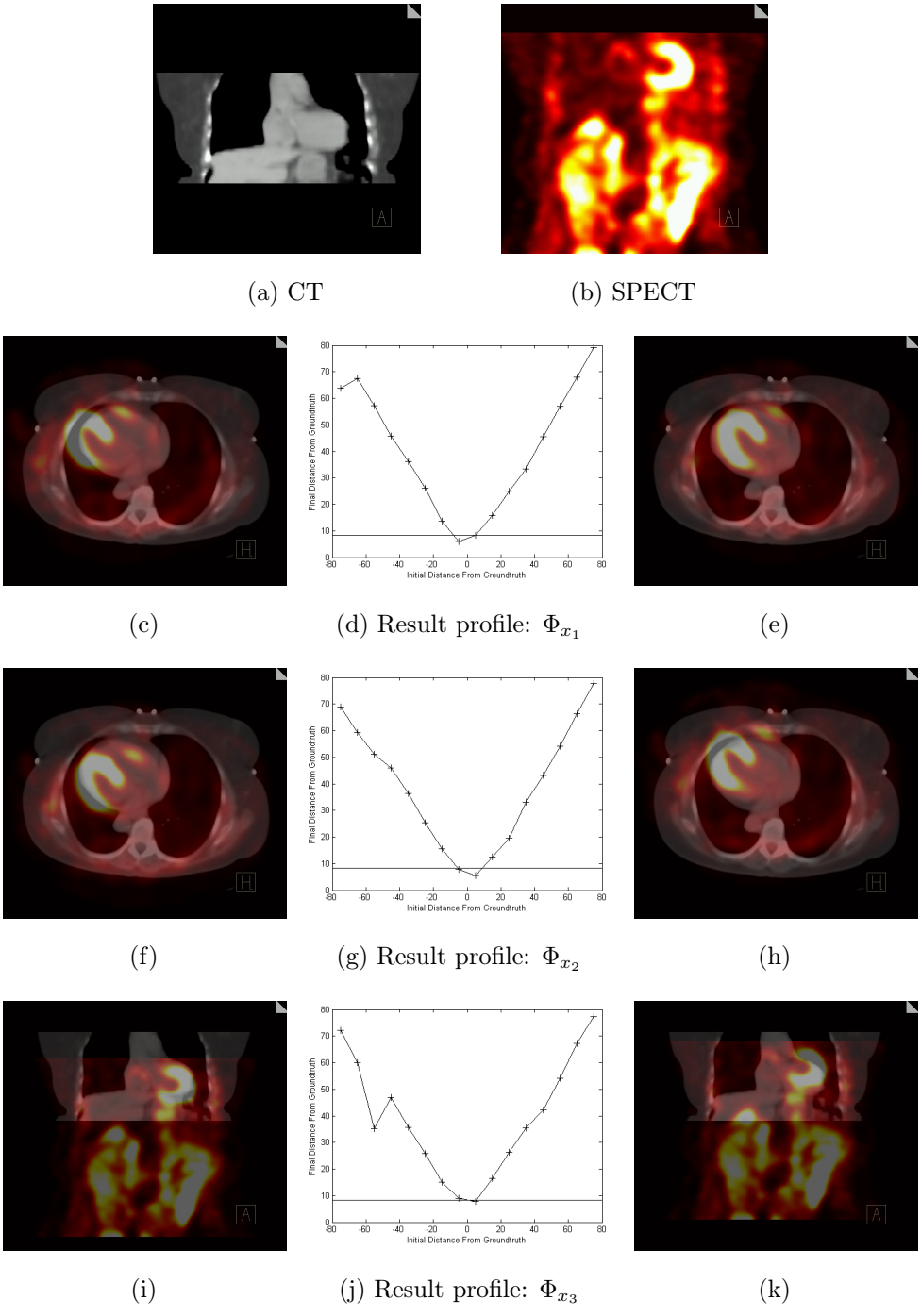
(c) PET/CT

Figure 3.1: (a)-(b) Clinical example of complementary image modalities showing a slice of the heart in head view. (c) Inaccurate convergence of traditional NMI registration due to a typical CT image reconstruction artifact.

data is usually treated as exceptional data within the technical community, in a clinical setting artifacted patient data is very much common. In chapter 5, we introduce two clinical applications that in the majority of the cases experience image artifacts. An example of such degenerate image data where where traditional NMI registration fails to properly align the images is given in Figure 3.1. Figure 3.1a shows a 256×256 pixels image slice of a CT heart acquisition including a bright and streaky windmill artifact centered in the cardiac wall. Such artifacts are resulting from CT reconstruction errors that occur at metallic medical implants such as implanted pacemakers or electrodes intended to support degenerate heart activity. Fig. 3.1b shows a 128×128 pixels image slice of a cardiac PET scan that visualizes cardiac function. In order to register these inherently complementary image information, one could revert to the previously introduced statistical similarity metrics such as NMI (defined in Eq. (2.23)). The result of rigidly registering PET and CT volume using NMI is shown in Fig. 3.1c.

The reason for the poor behavior of traditional data driven measures is revealed in Fig. 3.1c, where the strong activity from PET, represented as bright intensities, aligns with the image artifact in CT. The presence of image artifacts in the data may shift the location of the similarity measure's optimum. In addition, such degenerate data increases the number of locally optimal solutions with only one of them representing the desired clinical matching. Though many factors play an important role in the success of a registration method (e.g. choice of the optimization technique, interpolation strategy, multi-resolution approaches, etc.), it is usually inconvenient and at times impractical to determine the optimal set of parameters on degenerated data sets. As we can see from this examples, the performance of traditional similarity measures is strongly degraded on artifacted data.

Another key problem with traditional registration methods, the problem of limited capture range (i.e. the interval of parameters or distance away from the correct solution that does not contain any local optima other than the correct solution), is documented in Fig. 3.2. In this figure, we see the 2-d rendering of a clinical 3-d SPECT/CT data set at different spatial alignments. The visualized alignments describe the capture ranges for traditional mutual information-based registration technique tested by translational initializations Φ_S^0 between ± 50 mm away from the ground truth along the x_1, x_2, and x_3-dimensions. One can notice the limited range along each parameter dimension for this type of image data visualizing the challenging registration problem in routine clinical use. The capture range is directly correlated to the robustness of a registration method and, hence, to its clinical applicability. Its computation is rather difficult and may vary significantly within an application due to a large variety of patient data in clinical settings. The reason for the poor capture range on this data is revealed in Fig. 3.1 by examining the SPECT image data. Due to the nature of the SPECT imaging technique, the reconstructed images hold only few identifying structures (see section 1.2 for more details on SPECT imaging). Though the 3-d image data would allow for a robust estimation of probability distributions from the considered image intensities, the limited availability of identifying image objects in SPECT leads to a noisy description of the images' joint intensity relationships.



(a) CT (b) SPECT

(c) (d) Result profile: Φ_{x_1} (e)

(f) (g) Result profile: Φ_{x_2} (h)

(i) (j) Result profile: Φ_{x_3} (k)

Figure 3.2: Capture range visualization through a convergence experiment on translation parameters $\Phi_{x_{1,2,3}}$ using a 3-d clinical SPECT/CT data set with available ground truth (GT). Result profiles plot final vs. initial distance from GT. (c)-(k) The images next to these plots visualize the maximum ($\pm$) distance along dimension x_i from where registration still succeeds using ECC.

3.2 Statistical Intensity Priors

Intensity histograms have proven to be very robust and meaningful in solving selected computer vision problems. Swain and Ballard [Swai 91], for example, describe how object detection and identification can be performed through comparing to a model histogram that describes the searched object. Information such as color, brightness, and size of objects within an image can be captured by intensity histograms. A joint intensity histogram combines two 1-d intensity histograms and describes the frequencies of intensity pairs in the two volumes under registration defined at a specific alignment. Statistical similarity measures such as MI, its normalized versions, or correlation ratio are computed utilizing the joint histogram as a key element because it becomes distinct when the two matched images align.

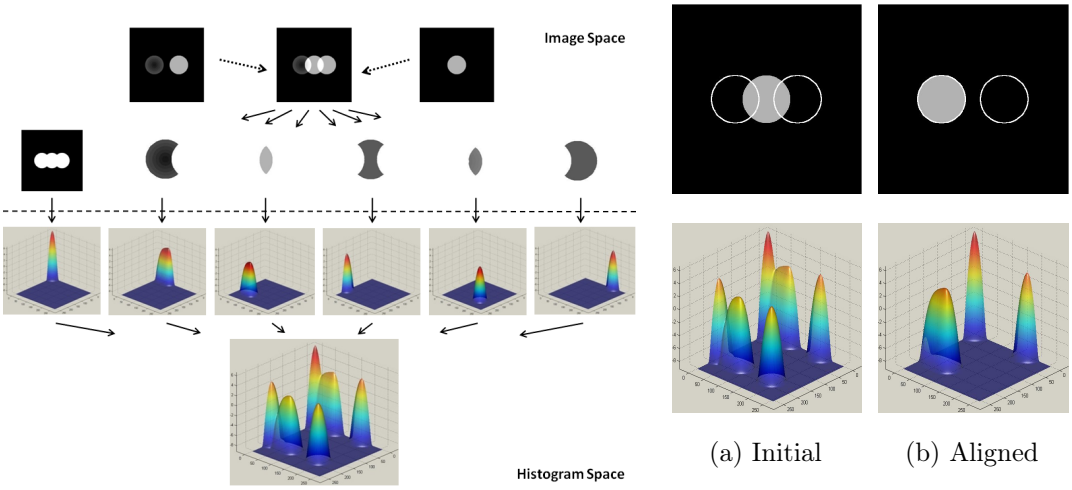
The joint histogram can be thought of as a collection of point samples from a joint probability density function (pdf) of the two given images. Assuming the two images are specified by functions f_1 and f_2, where f_2 is deformed by the displacement field $\boldsymbol{u}$. Then, the joint density can be inferred by the general representation of a normalized Parzen-Window [Parz 62] estimator:

$$p_{\boldsymbol{u}}(i_1, i_2) \equiv \frac{1}{|\Omega|} \int_{\Omega} G_{\sigma}(i_1 - f_1(\boldsymbol{x}), i_2 - f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})))\mathrm{d}\boldsymbol{x}, \tag{3.1}$$

with a two-dimensional Gaussian kernel G_{σ} with variance σ, and $\boldsymbol{u}(\boldsymbol{x})$ denotes the displacement field at location $\boldsymbol{x}$. Note that the Parzen estimator is not limited to Gaussian kernels. The density estimator for the marginal densities is defined analogous in 1-d.

Figure 3.3 explains how the intensity relation between the images is described by different contributions to a joint pdf or histogram of two images that align three disks. In addition, Figs. 3.3a and 3.3b show the joint histogram for two different, but specific example alignments. The first one matches the center of one disk with the middle point between the two other disks. By examining the image space of Fig. 3.3a it becomes clear that this initial alignment is ambiguous since one disk is at exactly equal distance to both disks in the second image. Though this example alignment is purely synthetic, it has practical value in simulating ambiguous matching situations that could occur in clinical practice. One can imagine such ambiguous alignment situations to occur in cancer treatment follow-up monitoring where lesions or tumorous tissue appear or (hopefully) disappear in between two acquisitions. The role of registration in follow-up studies is merely tumor growth quantification. Figure 3.3b describes an accurate matching of two equal sized disks. Note how distinguished the joint histogram becomes from the initial alignment once the image data is accurately matched.

Consequently, the joint intensity distribution can further be used as a fixed prior model that holds knowledge about the probability of occurring joint intensities at accurate alignment. Registration is then achieved by retrieving an optimal transformation such that the resulting joint intensity distribution matches the fixed prior distribution as described in [Leve 98, Herm 02, Chun 02, Zoll 03]. It has been shown that those models result in more robust registration methods with significantly larger capture ranges than statistical similarity measures deliver. While fixed joint intensity prior models increase robustness of multimodal registration techniques, the accuracy of such



(a) Initial (b) Aligned

Figure 3.3: Joint histogram contributions of two disk images at a particular alignment. The joint histogram becomes distinctive if two data sets are precisely aligned (a)-(b). The joint histograms (bottom parts) are visualized as 3D height maps in logarithmic scale.

purely prior knowledge driven methods is dependent on the accuracy of the statistical intensity prior. The computed registration can only result in an alignment as good as being described by the prior joint intensity model. Hence, it may introduce a bias that will be reflected in a degraded registration solution. Therefore, the applicability of such methods to routine clinical application is limited.

A powerful compromise to the abovementioned issues is the combination of statistical similarity measures and intensity priors as introduced in our early works [Guet 05] where we proposed a more robust and more accurate similarity criterion for nonrigid intensity-based multimodal image registration. At that time, we formulated this compromise using a variational energy minimization framework yielding promising preliminary results. Unfortunately, the theoretical foundations for such an integrated statistical intensity prior model were not clarified. In the following section, we will now present the Bayesian framework that provides the theoretical foundation for our original combined approach. It will further allow us to formulate a comprehensive statistical intensity prior model in later sections. The integrated SIP model generates an optimization function that ensures consistency with current and prior observations by granting flexible control between the two influences. It is not obvious how to find the appropriate balance between the two observations. However, we found in our experiments that the control parameter has an optimum value and in that way addresses the adaptability issue of fixed joint intensity models.

3.3 Consistency of Current with Prior Observations

Assume we are given a representative set of pre-registered image pairs $\{f_1^j, f_2^j\}_{j=1,\ldots,m}$, where $f_k^j : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. These image pairs may be obtained from various image modalities or body sections. Each registered image pair gives rise to a specific joint intensity distribution $p_j(i_1, i_2)$, stating which intensities i_1 and i_2 are likely to be in

correspondence for the given image pair. The goal of this thesis is to derive means to impose this knowledge during optimization in variational image registration algorithms. The matching problem, therefore, formulated as finding the most likely displacement field function $\boldsymbol{u}$ and joint intensity distribution $p_{\boldsymbol{u}}$, given the two images f_1 and f_2, under the constraint that the joint intensity distribution $p_{\boldsymbol{u}}$ is provided by the one arising from $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))$. That is we propose to maximize the conditional distribution

$$\begin{aligned}\mathcal{P}\big(\boldsymbol{u},\, p_{\boldsymbol{u}} \,\big|\, f_1, f_2, \{p_j\}\big) &= \frac{\mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\; \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big)}{\int_{\boldsymbol{u},p_{\boldsymbol{u}}} \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\; \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big) d\boldsymbol{u} dp_{\boldsymbol{u}}} \\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\; \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big) \\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}\big)\; \mathcal{P}\big(p_{\boldsymbol{u}} \big|\, \boldsymbol{u}\big)\; \mathcal{P}\big(\boldsymbol{u}\big)\; \mathcal{P}\big(p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big),\end{aligned} \tag{3.2}$$

with respect to the displacement field $\boldsymbol{u}$. Proportionality in the above expressions means that we have only neglected factors that do not depend on the displacement field $\boldsymbol{u}$ and thus do not affect the maximization. In the second step in (3.2), we have made the assumption that the prior decouples into a geometric prior $\mathcal{P}(\boldsymbol{u})$ on the displacement field and a prior on the joint intensity distribution $p_{\boldsymbol{u}}$. This latter term can be used to model intensity normalization that plays a major role in the success of a multimodal registration technique. For now, we can assume $\mathcal{P}\big(p_{\boldsymbol{u}} \big|\, \boldsymbol{u}\big)$ to be 1 because $p_{\boldsymbol{u}}$ is derived from $\boldsymbol{u}$ but fixed for a particular $\boldsymbol{u}$, and we will ignore this term. In section 3.3.2, it will become clear that this term bears a certain significance for clinical applications, and we will describe a respective model for it.

Consequently, the optimization problem in (3.2) separates into three factors that can be interpreted as follows: The first factor provides the measurement likelihood, stating how likely the two images are given the correspondence induced by the displacement field $\boldsymbol{u}$. The second factor in (3.2) indicates the a priori probability of a displacement field $\boldsymbol{u}$. And the last factor specifies how consistent the estimated joint intensity distribution is with respect to the previously learned ones. In the following sections, we will describe how these three expressions can be modeled.

Low Confidence Prior

In clinical applications, large amounts of prior knowledge may not always be available instantly and only build up over time. In this case, we can assume only a low confidence in our prior model. As the availability m of previously registered data sets increases, the confidence in the prior model increases.

Assuming there is only one representative pre-registered data set available (i.e. $m = 1$), then the set $\{p_j\}$ consists only of one learned probability distribution p^l and the respective prior probability can be formulated as:

$$\mathcal{P}\big(p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big) = \mathcal{P}\big(p_{\boldsymbol{u}} \;\big|\; p^l\big) \propto \exp\left(-\,\alpha_1\, \frac{\mathcal{I}_{\text{sip}}(p_{\boldsymbol{u}}, p^l)^2}{2\,\sigma_\ell^2}\right) \tag{3.3}$$

If we now consider the log likelihood of the posterior probability $\mathcal{P}\left(\boldsymbol{u}, p_{\boldsymbol{u}} \,|f_1, f_2\right)$, the following minimization problem will lead to the MAP point:

$$\begin{aligned} E(\boldsymbol{u}, p_{\boldsymbol{u}}, f_1, f_2) &= -\log\left(\mathcal{P}\left(\boldsymbol{u}, p_{\boldsymbol{u}} \,|f_1, f_2\right)\right) \\ &= \mathcal{I}_{\text{data}} + \alpha_1 \frac{\mathcal{I}_{\text{sip}}(p_{\boldsymbol{u}}, p^l)^2}{2\,\sigma_\ell^2} + \alpha_2\,\mathcal{I}_{\text{smooth}}(\boldsymbol{u}), \end{aligned} \tag{3.4}$$

where $\mathcal{I}_{\text{smooth}}$ and $\mathcal{I}_{\text{data}}$ refer to the smoothness and data term respectively. The parameters α_1 and α_2 determine the influence of prior knowledge and smoothness assumption, respectively. In our experiments, we show how a reliable value for α_1 can be found deterministically based on the available image information. Without supervised training, accurate (theoretical) derivation of parameter α_1 is difficult. In [Sabu 05], Sabuncu *et al.*propose to automatically estimate the influence of prior samples as a fraction between the amounts of prior and observed samples. As it appears, this computation is merely a shifting of the problem to the decision of how many samples to draw from the currently observed and from the prior distribution. Moreover, the proposed parameter computation has not been validated for feasibility on clinical, multimodal data. Nevertheless, we experimentally show that a robust value of α_1 can be determined to achieve robust and accurate image registration.

Formulation (3.4) is equivalent to our previously proposed method in [Guet 05] with $\mathcal{I}_{\text{data}}$ defined as the maximum mutual information similarity and $\mathcal{I}_{\text{smooth}}$ being the standard Tikhonov regularizer [Tikh 77]. Note that using a Gaussian kernel in (3.3) rather than an exponential one, as in our previous work, will lead to an additional factor of $\mathcal{I}_{\text{sip}}$ in the gradient expression of (3.4) that provides better convergence properties as the gradient goes to zero for $p_{\boldsymbol{u}} \to p^l$. We now realize that a minimum of (3.4) corresponds to the Maximum A Posteriori point of our Bayesian posterior probability formulation (3.2) for the nonrigid registration problem justifying an additive integration of the prior term in the variational optimization framework in [Guet 05]. It also provides an explanation why such an arrangement provides a favorable solution.

Similar to incorporating shape priors for robust segmentation, e.g. as presented in [Crem 06a], the above formulation (3.4) proposes to complement a traditional data term with an energy that measures the closeness to a given prior joint pdf. Our approach differs from previous approaches in that the prior is used as an additional constraint during the registration optimization process. Most prior-driven approaches that allow for small number of prior data sets have been using solely the distance to a prior as a similarity metric for registration [Leve 98, Chun 02, Gan 04, Herm 02] rather than integrating it into a combined model. As the statistical intensity prior integrates with the data term in (3.4) we will refer to it as *integrated statistical intensity prior* model. Slightly different approaches from formulation (3.4), but with similar strengths, have been proposed by Sabuncu *et al.*and Zoellei [Sabu 08, Zoll 06a] who pool together samples from prior and current observations. The pool of samples can be compared to our combined energy (3.4). Despite the promising preliminary registration results of these two methods, they remain largely unvalidated on routine clinical data.

Figure 3.4 gives a visual interpretation of the proposed integrated model for the previously mentioned ambiguous matching problem involving three disks. This figure shows the three spaces that are deeply intertwined during image registration, i.e the

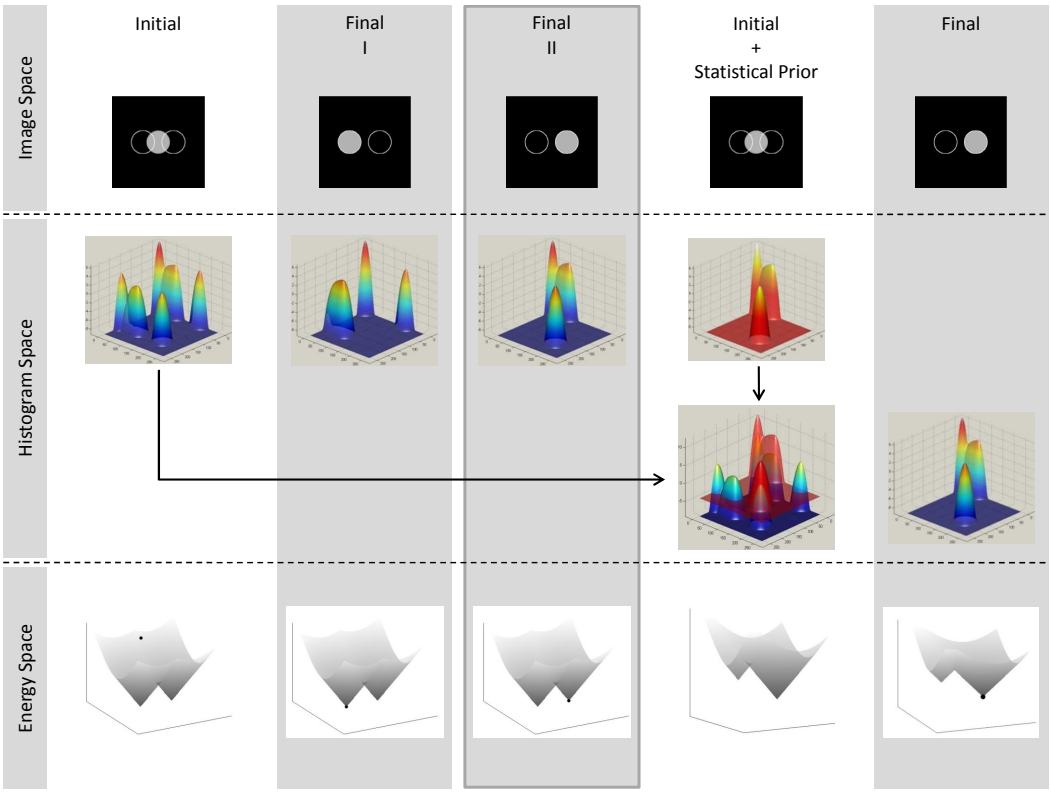


Figure 3.4: Registration problem (left column) that shows two equally optimal solutions for data-driven registration methods (2 middle columns). Adding statistical priors about the correct solution dissolves such ambiguities (right column).

image space, the histogram space, and the energy landscape. Each column in Fig. 3.4 represents one stage during registration of the images as described by the three different spaces. The top row visualizes the alignment images with overlayed object gradient edges of the reference image, the middle row plots the joint histograms of the alignment as 3-d height maps in log scale, and the bottom row prints an abstract but corresponding energy landscape. The black dot in this energy landscape represents the current alignment. For traditional methods, the schematic energy landscape is portrayed in the bottom row along the first three columns and visualizes the ambiguity of the matching problem by two dominant equal optima. Thus, traditional methods such as MI will either match the one disk in the alignment image with the left of the two reference disks (column "Final I") or with right one (column "Final II"). Suppose the matching with the right disk is clinically relevant. Using the joint histogram of the alignment ("Final II") as a prior will alter the energy landscape as shown in column "Initial + Statistical Prior". Note that in histogram space both prior and the data histogram are optimized upon resulting in the final alignment as displayed by the rightmost column in Fig 3.4.

The combined model has a number of favorable properties. First, this model leverages prior knowledge as an additional information channel during optimization.

Second, the flexible encoding using weights allows the prior to be turned on or off. In fact, using a multi-resolution scheme during registration, prior knowledge could only be used in coarse resolution levels. And third, in case no prior knowledge is available, the combined model reverts to pure data-driven optimization $\mathcal{I}_{\text{data}}$ corresponding to, for example, maximum mutual information.

The minimization of this low confidence prior model is achieved by descending the gradient of a combined cost function that optimizes the similarity of both volumes as well as the closeness of both volumes' joint intensity distribution to the statistical intensity prior. Minimization of energy (3.4) by gradient descent leads to a partial differential equation for $\boldsymbol{u}$ of the form:

$$\begin{aligned} \frac{\partial \boldsymbol{u}}{\partial t} &= -\frac{\partial E(\boldsymbol{u}, p_{\boldsymbol{u}}, f_1, f_2)}{\partial \boldsymbol{u}} \\ &= -\frac{\partial \mathcal{I}_{\text{data}}(\boldsymbol{u}, p_{\boldsymbol{u}}, f_1, f_2)}{\partial \boldsymbol{u}} - \alpha_1 \frac{\frac{1}{2}\,\partial \mathcal{I}_{\text{sip}}(p_{\boldsymbol{u}}, p^l)^2}{\partial \boldsymbol{u}} - \alpha_2 \frac{\partial \mathcal{I}_{\text{smooth}}(\boldsymbol{u})}{\partial \boldsymbol{u}}. \end{aligned} \tag{3.5}$$

The respective partial derivatives of the three terms are computed as follows.

3.3.1 The Variational Derivatives

Given the mathematical framework of calculus of variations, each gradient can be derived using the Parzen density estimator. The respective partial differential equation is:

$$\frac{\partial p(i_1, i_2)}{\partial \boldsymbol{u}} = \frac{1}{|\Omega|} \int_{\Omega} \partial_2 \mathrm{G}_\sigma\Big(i_1 - f_1(\boldsymbol{x}), i_2 - f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))\Big) \nabla f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})) \mathrm{d}\boldsymbol{x}, \tag{3.6}$$

where $\boldsymbol{x}$ is the N-dimensional spatial location, and $\partial_2 \mathrm{G}_\sigma(\ldots)$ denotes the partial derivative after G_σ's second variable. For example, $\partial_2 p(i_1, i_2) = \frac{\partial}{\partial i_2} p(i_1, i_2)$.

Statistical Intensity Prior Energy

The term $\mathcal{I}_{\text{sip}_{\text{KL}}}$ is designed to ensure the statistical consistency of the current joint pdf with the prior. This can be done by computing the Kullback-Leibler (KL) divergence between observed and prior pdf:

$$\mathcal{I}_{\text{sip}_{\text{KL}}}(\boldsymbol{u}) = \iint_{\mathbb{R}} p^o_{\boldsymbol{u}}(i_1, i_2) \ln \frac{p^o_{\boldsymbol{u}}(i_1, i_2)}{p^\ell(i_1, i_2)} \mathrm{d}i_1 \mathrm{d}i_2 \tag{3.7}$$

with $p^o_{\boldsymbol{u}}$ denoting the observed joint intensity distribution dependent on $\boldsymbol{u}$, and $p^\ell(i_1, i_2)$ representing the prior joint intensity distribution of a previously aligned data set. Equations (3.24) and (3.7) reveal the relationship of $\mathcal{I}_{\text{sip}_{\text{KL}}}$ and $\mathcal{I}_{\text{MI}}$. Mutual information is the KL divergence between observed joint pdf and the product of the observed marginals, whereas in $\mathcal{I}_{\text{sip}_{\text{KL}}}$ the product of the marginal densities is replaced by the training data.

In the calculus of variations, the Gateaux derivative yields the gradient for the statistical prior energy functional:

$$\left.\frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}(\boldsymbol{u}, f_1, f_2, \tilde{\boldsymbol{u}})}{\partial \epsilon}\right|_{\epsilon=0} = \frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}}{\partial \epsilon} = \frac{\partial}{\partial \epsilon} \iint\limits_{\mathbb{R}} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) \ln \frac{p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2)}{p^\ell(i_1, i_2)}\, di_1 di_2$$

$$\begin{aligned}
\ldots &= \iint\limits_{\mathbb{R}} \frac{\partial}{\partial \epsilon} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) \cdot \ln(p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2)) + p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) \cdot \frac{\partial}{\partial \epsilon} \ln(p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2)) \\
&\qquad - \frac{\partial}{\partial \epsilon} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) \cdot \ln(p^\ell(i_1, i_2))\, di_1 di_2 \\
&= \iint\limits_{\mathbb{R}} \left[1 + \ln\left(\frac{p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2)}{p^\ell(i_1, i_2)}\right)\right] \frac{\partial}{\partial \epsilon} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2)\, di_1 di_2 \qquad (3.8)
\end{aligned}$$

with

$$p^\ell(i_1, i_2) = \frac{1}{|\Omega|} \int\limits_{\Omega} \mathrm{G}_\gamma \left(f_1^\ell(\boldsymbol{x}) - i_1, f_2^\ell(\boldsymbol{x}) - i_2\right) d\boldsymbol{x} \qquad (3.9)$$

being derived from previously aligned image data, and hence not dependent on $\boldsymbol{u}$. The observed joint density is written as

$$p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) = \frac{1}{|\Omega|} \int\limits_{\Omega} \mathrm{G}_\sigma \left(f_1(\boldsymbol{x}) - i_1, f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}) + \epsilon\tilde{\boldsymbol{u}}(\boldsymbol{x})) - i_2\right) d\boldsymbol{x} \qquad (3.10)$$

and its partial derivative is

$$\begin{aligned}
\frac{\partial}{\partial \epsilon} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1, i_2) = \frac{1}{|\Omega|} \int\limits_{\Omega} &\partial_2 \mathrm{G}_\sigma \left(f_1(\boldsymbol{x}) - i_1, f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}) + \epsilon\tilde{\boldsymbol{u}}(\boldsymbol{x})) - i_2\right) \cdot \\
&\cdot \nabla f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}) + \epsilon\tilde{\boldsymbol{u}})\tilde{\boldsymbol{u}}(\boldsymbol{x})\, d\boldsymbol{x}, \qquad (3.11)
\end{aligned}$$

where ∂_2 denotes the partial derivative after the function's second variable. Replacing equation (3.11) in (3.8) and letting $\epsilon = 0$ yields

$$\begin{aligned}
\frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}}{\partial \epsilon} = \frac{1}{|\Omega|} \iint\limits_{\mathbb{R}} \int\limits_{\Omega} &\mathbf{L}^{\boldsymbol{u}}(i_1, i_2) \cdot \partial_2 \mathrm{G}_\sigma \left(f_1(\boldsymbol{x}) - i_1, f_2(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x}) - i_2\right) \cdot \\
&\cdot \nabla f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})) \cdot \tilde{\boldsymbol{u}}(\boldsymbol{x})\, dx di_1 di_2 \qquad (3.12)
\end{aligned}$$

with

$$\mathbf{L}^{\boldsymbol{u}}(i_1, i_2) = \left[1 + \ln\left(\frac{p^o_{\boldsymbol{u}}(i_1, i_2)}{p^\ell(i_1, i_2)}\right)\right].$$

Note that the following formula holds for partial differentiation and the convolution operator

$$\frac{\partial}{\partial x_i}(f * g)(\boldsymbol{x}) = \frac{\partial f}{\partial x_i} * g = f * \frac{\partial g}{\partial x_i} \qquad (3.13)$$

Eventually, we have

$$\frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}}{\partial \epsilon} = \frac{1}{|\Omega|} \int_{\Omega} \left[\frac{\partial}{\partial i_2} \mathbf{L}^{\boldsymbol{u}}(i_1, i_2) * \mathrm{G}_{\sigma} \right] (f_1(\boldsymbol{x}), f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))) \cdot \nabla f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})) \cdot \tilde{\boldsymbol{u}}(\boldsymbol{x})\, d\boldsymbol{x} \quad (3.14)$$

and

$$\frac{\partial}{\partial i_2} \mathbf{L}^{\boldsymbol{u}}(i_1, i_2) = \frac{\partial_2 p^0_{\boldsymbol{u}}(i_1, i_2)}{p^0_{\boldsymbol{u}}(i_1, i_2)} - \frac{\partial_2 p^{\ell}(i_1, i_2)}{p^{\ell}(i_1, i_2)}. \quad (3.15)$$

The gradient of the energy functional with respect to $\boldsymbol{u}$, i.e. $\nabla_{\boldsymbol{u}} \mathcal{I}$, comprising the Kullback-Leibler divergence between observed and prior joint intensity densities results from the following scalar product, as defined in section 2.1:

$$\left. \frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}(\boldsymbol{u}, f_1, f_2, \tilde{\boldsymbol{u}})}{\partial \epsilon} \right|_{\epsilon=0} = \langle \nabla_u \mathcal{I}_{\text{KL}}(\boldsymbol{u}), \tilde{\boldsymbol{u}} \rangle_H = 0. \quad (3.16)$$

This leads us to the partial derivative:

$$\frac{\partial \mathcal{I}_{\text{sip}_{\text{KL}}}(\boldsymbol{u})}{\partial \boldsymbol{u}} = -\frac{1}{|\Omega|} \left[\left(\frac{\partial_2 p^o_{\boldsymbol{u}}(i_1, i_2)}{p^o_{\boldsymbol{u}}(i_1, i_2)} - \frac{\partial_2 p^{\ell}(i_1, i_2)}{p^{\ell}(i_1, i_2)} \right) * \mathrm{G}_{\sigma} \right] (f_1(\boldsymbol{x}), f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})) \cdot \nabla f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})). \quad (3.17)$$

We can see that the term $\frac{\partial_2 p^o_{\boldsymbol{u}}(i_1,i_2)}{p^o_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 p^{\ell}(i_1,i_2)}{p^{\ell}(i_1,i_2)}$ serves as a comparison function. In fact, alignment is achieved by continuous adjustments of the joint intensity model until it resembles the learned joint pdf. Instead of taking the nearest most likely intensity correspondence into account, intensity correspondences learned from previous, correct alignments are considered.

Jensen-Shannon Divergence

The Kullback-Leibler divergence is a distance measure between distributions, however, it is not a distance metric. That can lead instabilities and incorrect estimations with respect to our gradient estimation for values of the joint distribution that are close to zero. The following chapter explains these issues on an academic example.

The Jensen-Shannon divergence has further desirable properties besides being symmetric and a metric. For example, it still relates to other information-theoretic functionals such as the KL-divergence or Mutual Information and hence it shares their mathematical properties and intuition. Furthermore, joint intensity distributions compared with the JS-divergence can be weighted and we introduce a variant of our prior model in the next section that makes use of this. The Jensen-Shannon divergence is defined as

$$\mathcal{I}_{\text{sip}_{\text{JS}}}\big(p^o_{\boldsymbol{u}} \,||\, p^{\ell}\big) = \frac{1}{2} \mathcal{I}_{\text{sip}_{\text{KL}}}\big(p^o_{\boldsymbol{u}} \,||\, \overline{p}_{\boldsymbol{u}}\big) + \frac{1}{2} \mathcal{I}_{\text{sip}_{\text{KL}}}\big(p^{\ell} \,||\, \overline{p}_{\boldsymbol{u}}\big), \quad (3.18)$$

where

$$\overline{p}_{\boldsymbol{u}} = \frac{p^o_{\boldsymbol{u}} + p^{\ell}}{2}. \quad (3.19)$$

Using the definitions and derivatives for the KL-divergence from above, we can derive the Jensen-Shannon gradient as follows

$$\begin{aligned}\left.\frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(p^o_{\boldsymbol{u}}, p^\ell, \tilde{\boldsymbol{u}})}{\partial \epsilon}\right|_{\epsilon=0} &= \frac{1}{2}\frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(p^o_{\boldsymbol{u}}, \overline{p}_{\boldsymbol{u}})}{\partial \epsilon} + \frac{1}{2}\frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(p^\ell, \overline{p}_{\boldsymbol{u}})}{\partial \epsilon} = \dots \\ &= \dots \\ &= \iint\limits_{\mathbb{R}} \left[\frac{1}{2}\ln\left(\frac{p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1,i_2)}{\overline{p}(i_1,i_2)}\right)\right] \frac{\partial}{\partial \epsilon} p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1,i_2)\, di_1 di_2 \end{aligned} \tag{3.20}$$

Setting

$$\mathbf{L}^{\boldsymbol{u}}(i_1,i_2) = \left[\frac{1}{2}\ln\left(\frac{p^o_{\boldsymbol{u}+\epsilon\tilde{\boldsymbol{u}}}(i_1,i_2)}{\overline{p}(i_1,i_2)}\right)\right]$$

the needed partial derivative can be computed as

$$\frac{\partial}{\partial i_2}\mathbf{L}^{\boldsymbol{u}}(i_1,i_2) = \frac{1}{2}\left[\frac{\partial_2 p^0_{\boldsymbol{u}}(i_1,i_2)}{p^0_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 \overline{p}_{\boldsymbol{u}}(i_1,i_2)}{\overline{p}_{\boldsymbol{u}}(i_1,i_2)}\right]. \tag{3.21}$$

Then, the variational gradient of the Jensen-Shannon divergence can be written as

$$\begin{aligned}\frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(\boldsymbol{u})}{\partial \boldsymbol{u}} = -\frac{1}{2\,|\Omega|}\left[\left(\frac{\partial_2 p^0_{\boldsymbol{u}}(i_1,i_2)}{p^0_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 \overline{p}_{\boldsymbol{u}}(i_1,i_2)}{\overline{p}_{\boldsymbol{u}}(i_1,i_2)}\right) * \mathrm{G}_\sigma\right](f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x})) \\ \cdot \nabla f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x})).\end{aligned} \tag{3.22}$$

Note how the comparison function changed as compared to (3.17). The term $\frac{\partial_2 p^0_{\boldsymbol{u}}(i_1,i_2)}{p^0_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 \overline{p}_{\boldsymbol{u}}(i_1,i_2)}{\overline{p}_{\boldsymbol{u}}(i_1,i_2)}$ shows us that the currently observed joint intensity model gets updated until it resembles the average joint pdf between observed and learned pdf. It is an intriguing observation that this variational gradient computation ensures the symmetry property of JS.

Mutual Information

To model the first factor in (3.2) we are free to choose any of the previously introduced data comparison or similarity metrics. However, by reverting to the well-known concept of maximal mutual information, we achieve a consistent intuition for our comparison measures since Mutual Information, Kullback-Leibler and Jensen-Shannon divergence are closely related information-theoretic similarity measures. The Mutual Information between two images f_1 and f_2 is maximized by stating that they are more likely to be aligned if the two images $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x}+\boldsymbol{u})$ are statistically more dependent:

$$\mathcal{P}(f_1, f_2 \,|\, \boldsymbol{u}, p_{\boldsymbol{u}}) \propto \exp\Big(\mathcal{I}_{\text{MI}}\big(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u})\big)\Big), \tag{3.23}$$

with the mutual information $\mathcal{I}_{\text{MI}}$ being defined as:

$$\mathcal{I}_{\text{MI}}\big(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))\big) = \int_{\mathbb{R}^2} p_{\boldsymbol{u}}(i_1,i_2)\log\frac{p_{\boldsymbol{u}}(i_1,i_2)}{p_{f_1}(i_1)p_{f_2}(i_2)}di_1 di_2, \tag{3.24}$$

where $i_1 = f_1(\boldsymbol{x})$, $i_2 = f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))$, and $p_{f_1}(i_1)$, $p_{f_2}(i_2)$, $p_{\boldsymbol{u}}(i_1, i_2)$ are the marginal and joint intensity distributions estimated from $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))$.

In the statistical inference (3.2) this constraint will favor displacement fields $\boldsymbol{u}$ that maximize the statistical dependency between two random variables in form of the given images. In order to retrieve the alignment that yields the maximum MI between reference and alignment function, we define energy term $\mathcal{I}_{\mathrm{MI}}$ as the negative MI in (3.24). The alignment function f_2 is related to the reference function f_1 through the displacement field $\boldsymbol{u}$. Similar to the statistical prior energy gradient, the computation of the variational derivative using the definitions from above yields the gradient of mutual information energy as follows:

$$\frac{\partial \mathcal{I}_{\mathrm{MI}}(\boldsymbol{u})}{\partial \boldsymbol{u}} = \frac{1}{|\Omega|}\left[\left(\frac{\partial_2 p^o_{\boldsymbol{u}}(i_1,i_2)}{p^o_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 p^o_{\boldsymbol{u}}(i_2)}{p^o_{\boldsymbol{u}}(i_2)}\right) * \mathrm{G}_\sigma\right](f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x})) \cdot \nabla f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x})), \quad (3.25)$$

where $*$ denotes the convolution operator, G_σ is a two-dimensional Gaussian with standard deviation σ, and $|\Omega|$ is a normalizing constant. Similar to (3.17), the term $\frac{\partial_2 p^o_{\boldsymbol{u}}(i_1,i_2)}{p^o_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 p^o_{\boldsymbol{u}}(i_2)}{p^o_{\boldsymbol{u}}(i_2)}$ serves as a comparison function. It implicitly takes the knowledge of the nearest most likely intensity correspondence into account. Hermosillo, Chefd'Hotel and Faugeras have found this in their work on multimodal image matching [Herm 02].

Smoothness Prior on the Displacement Field

The last factor in (3.2) allows to impose a prior on the displacement field $\boldsymbol{u}$ stating which displacement fields are *a priori* more or less likely. As proposed in [Roth 05], one could also learn such priors from training sequences of optic flow fields. Since our contribution is the statistical modeling of priors on the intensity transformation between the two images, we shall merely impose a common smoothness prior on the displacement field:

$$\mathcal{P}(\boldsymbol{u}) \propto \exp\left(-\alpha_2 \mathcal{I}_{\mathrm{smooth}}(\boldsymbol{u})\right). \quad (3.26)$$

In this work we use the smoothness prior that was pioneered in the seminal work of Horn and Schunck [Horn 81]:

$$\mathcal{I}_{\mathrm{smooth}}(\boldsymbol{u}) = \frac{1}{2}\int_\Omega |\nabla \boldsymbol{u}(\boldsymbol{x})|^2 \, d\boldsymbol{x}, \quad (3.27)$$

with $\boldsymbol{x} \in \mathbb{R}^n$. The gradient of the smoothness constraint leads to the diffusion term $\triangle \boldsymbol{u}$:

$$\frac{\partial \mathcal{I}_{\mathrm{smooth}}(\boldsymbol{u})}{\partial \boldsymbol{u}} = \mathrm{div}\left(\frac{\Phi'_{\mathrm{smooth}}(|\nabla \boldsymbol{u}|)}{|\nabla \boldsymbol{u}|}\nabla \boldsymbol{u}\right) = \mathrm{div}(\nabla \boldsymbol{u}) = \Delta \boldsymbol{u}, \quad (3.28)$$

where $\Phi_{\mathrm{smooth}}(|\nabla \boldsymbol{u}|) = \frac{1}{2}|\nabla \boldsymbol{u}(\boldsymbol{x})|^2$, and $\Delta \boldsymbol{u} = \sum_{i=1}^n \frac{\partial^2 \boldsymbol{u}}{\partial x_i^2}$ denotes the Laplace operator.

The choice of a suitable smoothing term or geometric prior $\mathcal{I}_{\mathrm{smooth}}$ is usually made based on the type of application. More sophisticated priors are conceivable, for example non-quadratic (robust) smoothness priors that allow for discontinuities in the estimated displacement fields (cf. [Brox 04]), an edge-based smoothness term that changes from an isotropic smoothing in homogeneous regions to anisotropic

smoothing along contours of the image [Herm 02], a linear-elastic smoothing as described in [Chri 01], or a curvature regularization smoothness term as proposed in [Fisc 03].

3.3.2 Intensity Normalization

Some types of imaging modalities such as emission imaging show incoherent intensity ranges across patient populations due to their dependence on various patient parameters. For example, molecular image intensities represent counts of activity that cannot be normalized as reliable as, for example, CT intensities that are based on Hounsfield units (HU). Instead, the uptake in PET images, can be roughly quantified using the standard uptake value (SUV) that utilizes information such as amount of injected radionuclide activity, patient weight and height, body surface area, exposure time, and acquisition protocol parameters. However, SUV does not result in intensity values that fall into a standardizable range such as HU. Depending on scanner type (i.e. hybrid or stand-alone), acquisition protocol (i.e. gated vs. non-gated, attenuation corrected (AC) vs. non-AC), and the patient itself, the acquired intensity range may vary from 4-bit to up to 16-bit. That may significantly affect the performance of prior models such as the integrated statistical intensity prior. However, we have observed that (i) this problem can be treated as an application data dependent scaling issue, and (ii) the intensity range remains constant within a particular setting (i.e. acquisition protocol, scanner type). Due to the occurrence of activation hot spots or image artifacts, that are usually very bright but small in size, intensity normalization based on minimum and maximum intensity is not feasible. Note that if range normalization is not handled carefully, neither data-driven nor prior influenced registration methods will succeed.

Therefore, it becomes necessary to optimize the image intensities for consistency across populations in order to ensure consistency with the learned distributions. In the previous chapter, we considered the conditional distribution $\mathcal{P}(p_{\boldsymbol{u}}|\boldsymbol{u})$ to be 1. It models the probability of a joint intensity distribution given the displacement field $\boldsymbol{u}$ by stating that a given joint intensity distribution is more likely when it resembles a representative reference distribution. This term is only of practical importance and for normalized modalities such as CT, it is not necessary due to a consistent representation of objects. In other words, $\mathcal{P}(p_{\boldsymbol{u}}|\boldsymbol{u})$ is modeled to incorporate an intensity normalization optimization. This allows us to estimate an optimal intensity transformation between the current and the prior data set using the joint intensity model at the same time as the registration is computed. We model the conditional probability as follows

$$\mathcal{P}(p_{\boldsymbol{u}}|\boldsymbol{u}) = \exp\left(-\alpha\, \mathcal{I}_{\text{inorm}}(p^o_{\boldsymbol{u}}, p^\ell, \Phi_{\text{i}})\right) \tag{3.29}$$

where $\mathcal{I}_{\text{inorm}}$ is an intensity normalization term, $p^o_{\boldsymbol{u}}$ and p^ℓ denote currently observed and statistical intensity prior distribution, respectively, and Φ_i defines an intensity transformation that estimates the intensity variation between patients describing a simple affine intensity model

$$\Phi_{\text{i}}(i) = s \cdot i + c. \tag{3.30}$$

Then, the optimal intensity transformation minimizes the following JS-divergence between the two distributions

$$\begin{aligned}\mathcal{I}_{\text{inorm}}(p^o_{\boldsymbol{u}}, p^\ell, \Phi_\text{i}) &= \mathcal{I}_{\text{sip}_\text{JS}}\left(p^o_{\boldsymbol{u}}\left(f_1, \Phi_\text{i}(f_2)\right) \,||\, p^\ell(f_1, f_2)\right) \\ &= \frac{1}{2}\int_{\mathbb{R}^2} p^o_{\boldsymbol{u}}(i_1, \Phi_\text{i}(i_2)) \,\ln\frac{p^o_{\boldsymbol{u}}(i_1, \Phi_\text{i}(i_2))}{\overline{p}_{\Phi_\text{i}}(i_1, i_2)} di_1 di_2 + \\ &\qquad \frac{1}{2}\int_{\mathbb{R}^2} \overline{p}_{\Phi_\text{i}}(i_1, i_2) \,\ln\frac{\overline{p}_{\Phi_\text{i}}(i_1, i_2)}{p^\ell(i_1, i_2)} di_1 di_2 \quad (3.31)\end{aligned}$$

with

$$\overline{p}_{\Phi_\text{i}}(i_1, i_2) = \frac{p^o_{\boldsymbol{u}}(i_1, \Phi_\text{i}(i_2)) \,+\, p^\ell(i_1, i_2)}{2}.$$

This intensity normalization model is derived for the clinical application of PET or SPECT/CT imaging using the assumption that CT intensities are normalized by Hounsfield units. Therefore, the linear intensity model is only applied to the molecular imaging modalities, and without loss of generality, we define f_2 to describe those modalities.

In order to deal with the intensity range scaling issue in molecular imaging, we are recovering the optimal scale s that maximizes the conditional probability (3.29). Thus,

$$\hat{s} = \arg\min_s \mathcal{I}_{\text{sip}_\text{JS}}\left(p^o_{\boldsymbol{u}}\left(f_1, \Phi_\text{i}(f_2)\right) \,||\, p^\ell(f_1, f_2)\right), \quad (3.32)$$

where f_2 corresponds to the molecular image intensities, and constant c is chosen to be zero. Finding the maximum of (3.29) is performed simultaneously to retrieving the optimal displacement field $\boldsymbol{u}$ by retrieving the MAP of (3.2). Therefore, our Bayesian formulation incorporates both spatial and intensity normalization in one framework. We are aware that far more elaborate intensity modeling techniques exist than our simple affine attempt (3.30). For example, Jäger *et al.* [Jage 09] non-rigidly register the joint intensity distribution of magnetic resonance images to a template image and achieve intensity normalization. The goal of this work, however, is to derive advanced intensity prior models in medical image registration.

In clinical practice, one could accelerate the optimization of (3.2) by performing the maximization of (3.29) beforehand and keep $\hat{s}$ fixed during the registration optimization.

3.4 Constrained Statistical Intensity Prior

For some clinical applications, the choice of a constrained rigid or affine transformation model is more appealing due to increased simplicity, robustness, and efficiency of the registration procedure. Furthermore, the compromised overall registration accuracy may be acceptable since the clinical focus is only directed to parts of the image or volume (e.g. cardiac imaging). The following section describes a constrained statistical intensity prior model designed to ensure consistency with intensity intervals of the learned joint intensity relationships. We observe during our experiments that subsets of joint intensity distributions often correspond to specific regions or organs of interest in the medical image domain. Therefore, constraining the previously introduced

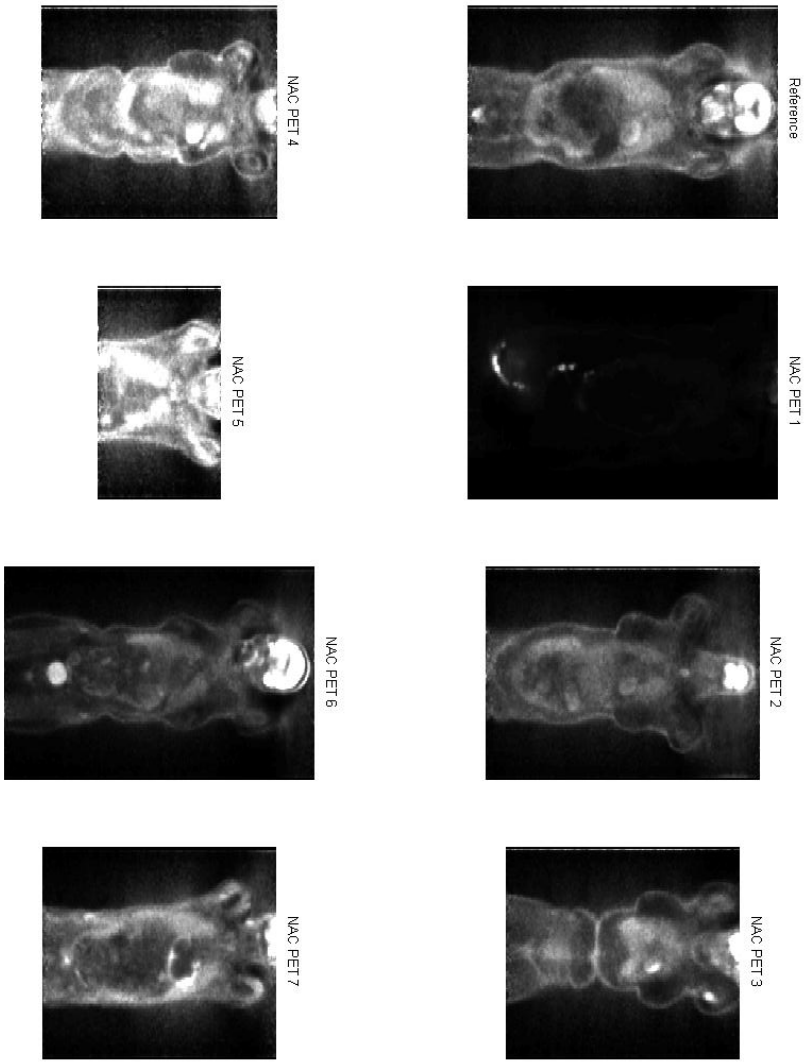


Figure 3.5: Eight frontal slices of non-attenuation corrected (NAC) PET images of PET/CT volumes acquired by hybrid scanners visualized using the same window/leveling function as in Fig. 3.6. This figure shows the unnormalized images.

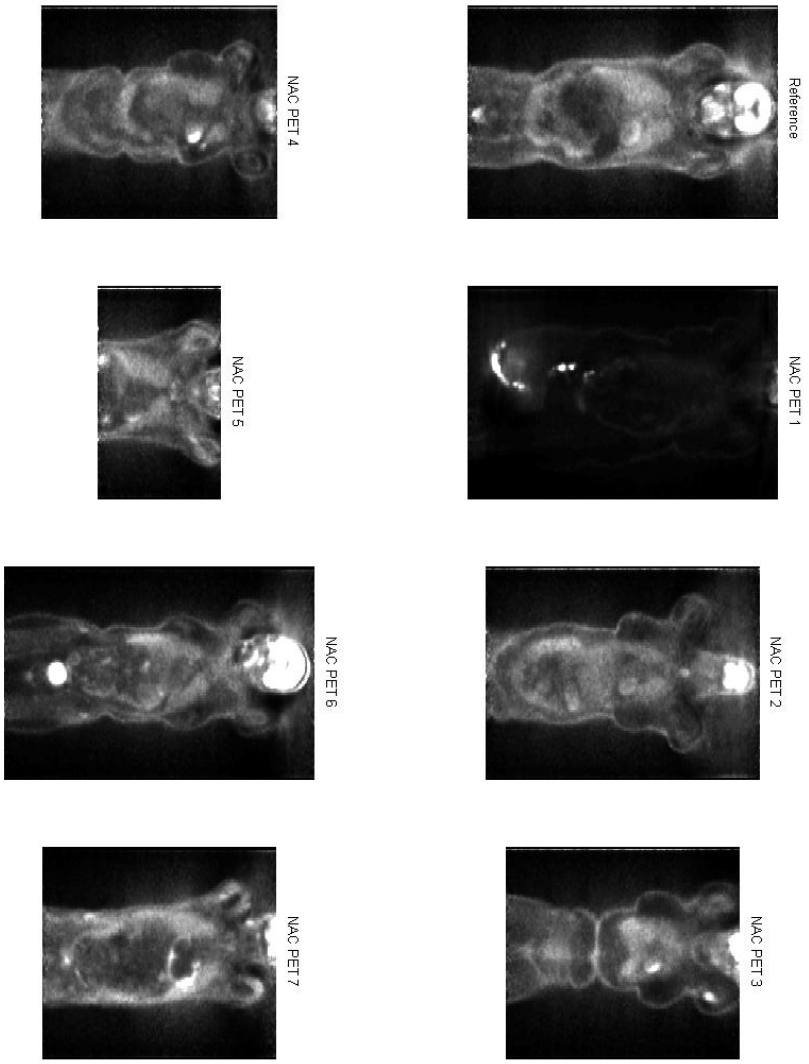
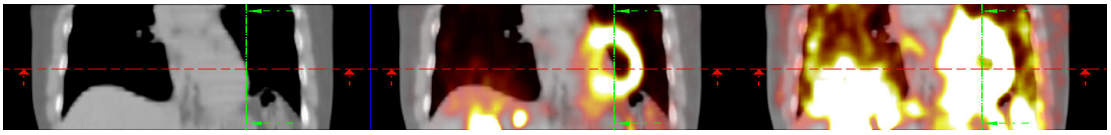


Figure 3.6: Eight frontal slices of non-attenuation corrected (NAC) PET images of PET/CT volumes acquired by hybrid scanners visualized using the same window/leveling function as in Fig.3.5. This figure visualizes the normalized version.



(a) CT (b) SPECT/CT W/L 1 (c) SPECT/CT W/L 2

Figure 3.7: Three anterior views of a misaligned cardiac SPECT/CT data set, the CT (a), and the SPECT overlayed on CT with two different window level settings 1, (b) and (c). The figure visualizes the challenging multi-modal registration problem.

integrated prior model can be interpreted as implicitly adding spatial information to the statistical intensity prior.

The following rigid registration method is designed specifically to address the clinical accuracy requirements of cardiac SPECT/CT imaging with the goal of providing a highly accurate registration approach that can be utilized as a fully-automatic method. Details on the clinical imaging technique and its dependency on accurate registration are provided in application section 5.3. The required accuracy in aligning cardiac SPECT/CT data is achieved by introducing a weighted intensity co-occurrence prior about an accurate alignment of just the mediastinum, as visualized in Fig. 3.10.

Two aspects are of importance for this application: (i) Given the noisy clinical data (Fig. 3.7), employment of the integrated SIP model is desirable to ensure robust and accurate alignment in general, and (ii) since misalignments within the heart region cause image artifacts, see section 5.3 for detailed description, the clinical accuracy requirement is mainly addressing the cardiac region. In the following, we derive a constrained SIP model considering those aspects. The achieved accuracy of the proposed approach is further compared to the accuracy of standard mutual information (MI) [Well 96, Maes 97] and the unconstrained integrated SIP with the focus of applicability. Achieving higher accuracy and robustness than MI in this application, the constrained statistical intensity prior model is not limited to cardiac SPECT/CT imaging.

3.4.1 Weighted Jensen-Shannon Divergence

In order to achieve the registration accuracy and robustness needed in CT-based AC for SPECT reconstruction, some open questions w.r.t. the integrated SIP model need to be answered. How does the choice of the α_1-parameter in (3.4) influence the registration of cardiac SPECT/CT and how should it be selected for this application? Secondly, is the Kullback-Leibler (KL) divergence a sufficient distance measure for ensuring consistency with learned joint pdfs? And the most intriguing open question: how well do the proposed schemes (3.4) and (3.35) generalize over a large pool of patients?

In contrary to the general SIP model (see chapter 4), the prior confidence value α_1 is difficult to derive for the integrated SIP model. Nevertheless, it is feasible to estimate a robust and meaningful α_1 based on empirical observations on synthetic as well as real clinical data. Synthetically, one can investigate the behavior of the

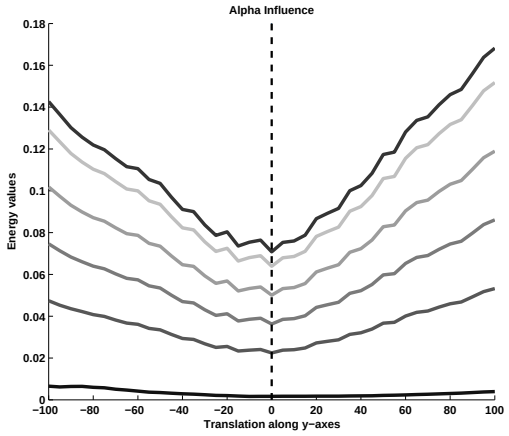


Figure 3.8: Influence of the α_1-parameter (eq. (3.4)) on the overall cost for a simple translation probing experiment in range [-100 mm; +100 mm] on SPECT/CT data. Values for α_1 are in range [0.6; 0.1] from top to bottom curve.

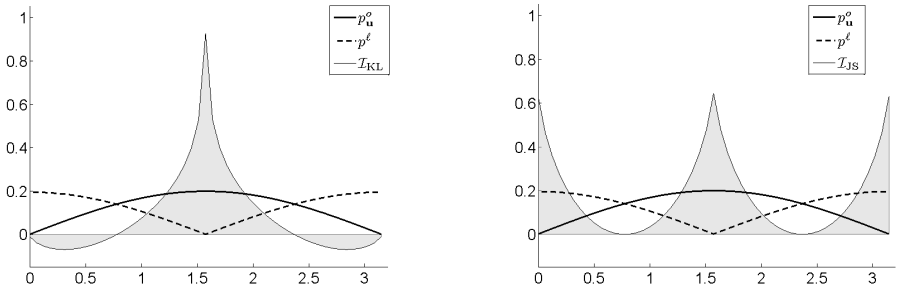
proposed energy (3.4) for a simple 1-d translation along an arbitrary coordinate axis passing through ground truth alignment. Figure 3.8a, for example, shows the results of such a synthetic experiment for different α_1 values. We note that a larger α_1, i.e. more prior influence, has a smoothing effect on the overall cost function. However, large values for α_1 also imply a strong tendency towards the prior that in return may limit the adaptability to a larger patient population. Therefore, an $\alpha_1 = 0.2$ is observed to be a good trade-off between the two forces. Decreasing the influence of MI allows to smooth out its local optima while still optimizing for the maximum mutual information that both images share.

Most previous work on prior based image registration [Chun 02, Soma 03, Guet 05, Crem 06b] utilize the KL-divergence to measure the (dis)similarity of probability distributions. In a discrete formulation, this can be written as:

$$\mathcal{I}_{\text{sip}_{\text{KL}}}\left(p^o_{\Phi_S}, p^\ell\right) = \text{KL}\left(p^o_{\Phi_S} \,||\, p^\ell\right) = \sum_{i,j} p^o_{\Phi_S}(i,j) \log\left(\frac{p^o_{\Phi_S}(i,j)}{p^\ell(i,j)}\right), \tag{3.33}$$

where $p^o_{\Phi_S}$ is the observed joint pdf of two volumes related to each other by rigid transformation Φ_S, and p^ℓ is the joint pdf learned from two previously aligned volumes. As previously discussed, the KL-divergence is not symmetric and Figure 3.9a illustrates the complication originating from this asymmetry when comparing two artificial distributions. We can observe that dissimilarities between the distributions may create opposite local contributions to KL dependent on the order of comparison. Consequently, local measures of the joint intensity distribution such as the previously derived gradients will be negatively influenced. The more appropriate statistical measure addressing this issue is provided by the Jensen-Shannon (JS) divergence that, in a discrete formulation, is defined as:

$$\mathcal{I}_{\text{sip}_{\text{JS}}}\left(p^o_{\Phi_S}, p^\ell\right) = \text{JS}\left(p^o_{\Phi_S} \,||\, p^\ell\right) = \frac{1}{2}\big(\text{KL}\left(p^o_{\Phi_S} \,||\, \overline{p}\right) + \text{KL}\left(p^\ell \,||\, \overline{p}\right)\big), \tag{3.34}$$



(a) Kullback-Leibler Divergence (b) Jensen-Shannon Divergence

Figure 3.9: Two plots of individual contributions from two artificial distributions for $p^o_{\boldsymbol{u}}$ and p^ℓ to (a) KL and (b) JS divergence. The area under the curve (gray) denotes the KL and JS value. Note that differences in the distribution have positive, limited, and comparable contributions to JS, but not to KL.

where $\overline{p} = \frac{p^o_{\Phi_S} + p^\ell}{2}$. Figure 3.9b nicely portraits the symmetry properties of JS-divergence. Besides correctly handling local measures, symmetry is of further importance when we want to emphasize on local contributions in the joint pdf originating from, for example, specific organs of interest.

Misalignments of the SPECT heart image into the lung region of CT attenuation map introduce artifacts that can lead to false diagnosis. Prior knowledge about the correct mapping within this area is important to ensure such a mapping in future registrations. An implicit property of the integrated SIP model is that stored prior information in the learned joint pdf is global and to some extent influenced by the size of the background. Local alignments are driven by the global matching especially if the transformation model is not capable of allowing local refinements such as rigid models. Thus, we propose a formulation of the prior model that utilizes local information stored in the learned joint pdf. In other words, the *constrained statistical intensity prior* formulation is given by:

$$\begin{aligned} \mathcal{I}_{\omega,\mathrm{sip_{JS}}}(p^o_{\Phi_S}, p^\ell) &= \omega \cdot \mathcal{I}_{\mathrm{sip_{JS}}}\left(p^o_{\Phi_S}, p^\ell\right) \\ &= \frac{1}{2}\sum_{\Omega} \omega(i,j)\left[p^o_{\Phi_S}(i,j)\ln\left(\frac{p^o_{\Phi_S}(i,j)}{\overline{p}\,(i,j)}\right) + \right. \\ &\qquad \left. p^\ell(i,j)\ln\left(\frac{p^\ell(i,j)}{\overline{p}\,(i,j)}\right)\right], \end{aligned} \tag{3.35}$$

where $\omega \in [0,1]^{|\Omega|\times|\Omega|}$, and $\cdot$ denotes the element-wise multiplication in the discrete case. The term ω will be chosen such that it penalizes organ specific intensity matchings that are inconsistent with a learned distribution. Hence, this penalty term introduces, to some extent, spatial information to intensity-based registration. Organ specific appearances in the joint intensity distribution can be estimated by either a segmentation or a manual outline of the organ of interest, see discussions below for details on the choice of ω. A crucial assumption of ω is that penalties need to be assigned comparably for differences between prior and joint pdf, see discussion KL

vs. JS above and Fig. 3.9. This requires a symmetric and strictly positive similarity measure on distributions.

The rigid transformation between the two data sets is obtained by minimizing the following expressions corresponding to the MAP point of the posterior distribution in (3.4) describing the limit case of the general SIP model. Instead of retrieving a displacement field $\boldsymbol{u}$, we are solving now for the parameters of transformation Φ_S. Further applying the intensity normalization step as described in section 3.3.2, we optimize a transformation $\hat{\Phi}$ such that

$$\hat{\Phi} \;=\; \underset{\Phi=\{\Phi_S,\Phi_i\}}{\arg\min} \left[(1-\alpha)\cdot \mathcal{I}^*_{MI}\left(p^o_\Phi\right) + (\alpha)\cdot \mathcal{I}_{\omega,\mathrm{sip_{JS}}}\left(p^o_\Phi, p^\ell_\Phi\right)\right] \tag{3.36}$$

where $\mathcal{I}^*_{MI} = -\mathcal{I}_{MI}$, and Φ is composed of a spatial rigid transformation Φ_S that aligns SPECT and CT volume and an intensity transformation Φ_i that warps p^ℓ to $p^o_{\Phi_S}$ to compensate for patient specific intensity variations. Φ_i is a 1-dimensional linear intensity transformation, as defined in (3.30), between the prior SPECT intensities and the intensity range of the SPECT volume to be registered. In our implementation, the two transformations are estimated sequentially but the framework above also allows for concurrent estimation. Eq.(3.36) is optimized in the parameter space of $\Phi = \{\Phi_S, \Phi_i\}$ by using either a gradient descent strategy or other strategies such as efficient line search, or hill-climbing optimizers.

Organ Specific Statistical Intensity Priors

The weighted Jensen-Shannon (wJS) divergence, defined in (3.35), is introduced to ensure an organ specific intensity co-occurrence. In order to derive a suitable ω for cardiac SPECT/CT registration, we segmented the heart in the SPECT volume using the method in [Kohl 06]. The penalty area of ω, i.e. white area in rightmost image of Fig. 3.10, is then generated by studying the joint pdfs for different alignments of the segmented heart with the CT, see two middle images in Fig. 3.10. In Fig. 3.10, the coordinate system is defined as follows: The origin is located in the lower left corner of each image, the horizontal and vertical axis refer to CT and SPECT volume intensities, respectively. Several interesting aspects are observed:

1. The joint intensity mappings corresponding to a segmented object in both modalities occur in a limited region within the joint pdf space. This is true for all possible spatial alignments of the two volumes (see two middle images in Fig. 3.10).

2. In order to ensure consistency with a learned distribution, the penalty term needs to cover all intensity pairs that the object may generate in the joint pdf. The reason is that a learned pdf not only states which intensities do match but also provides knowledge about which intensities do not match.

3. Evaluating a similarity measure on a subset of the joint pdfs eliminates unwanted influences from the unweighted learned distribution, e.g. background size dependency, influences of defining global structures, or bright image artifacts that usually occur outside the chosen subset.

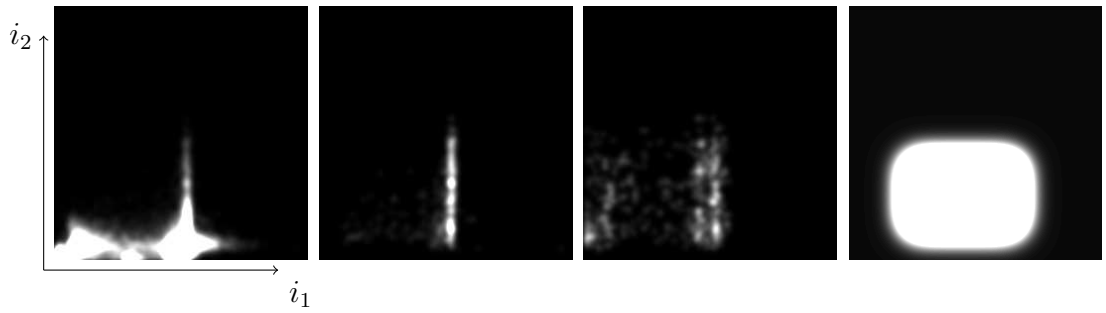


Figure 3.10: Observed joint pdfs of cardiac SPECT/CT data. From left to right, the distributions are displayed for the full volume overlap, the heart overlap, and the heart overlap at misalignment. The rightmost image presents the penalty term ω of eq. (3.35) that is generated from the observations made in two middle images.

Using the defined ω (i.e. rightmost image in Fig. 3.10), we apply the proposed constrained SIP model to a pool of cardiac SPECT/CT patients in the following demonstrations sections.

3.5 Integrated Statistical Intensity Prior: Demonstrations

Up to now, we have introduced a comprehensive statistical intensity prior model and several flavors of it. This section shows several examples of the integrated and constrained integrated SIP model on synthetic as well as clinical data and demonstrates several key properties of the statistical intensity priors. We, further, empirically estimate an optimal value of the confidence in the available prior data. The section separates in two parts describing observations for the rigid and non-rigid transformation models.

3.5.1 Robust and Accurate Parametric Registration

The robustness and accuracy of the integrated SIP for rigid transformation models is validated on a clinical problem and can be explored in chapter 5. In this section, we focus on synthetic and numerical experiments to gain some understanding in the behavior of the implied parameters.

3.5.2 Non-Rigid Registration Initialization and Convergence

In a significant amount of clinical image data ambiguous matching situations arise with only one clinically meaningful solution (e.g. in studies of disease or in certain populations such as the elderly). Moreover, there are plenty of challenges in clinical medical imaging data that pose limitations to intensity-based registration algorithms. To investigate the clinical potential of the proposed prior model under a variety of clinically realistic conditions, we demonstrate the model's ability to handle topological

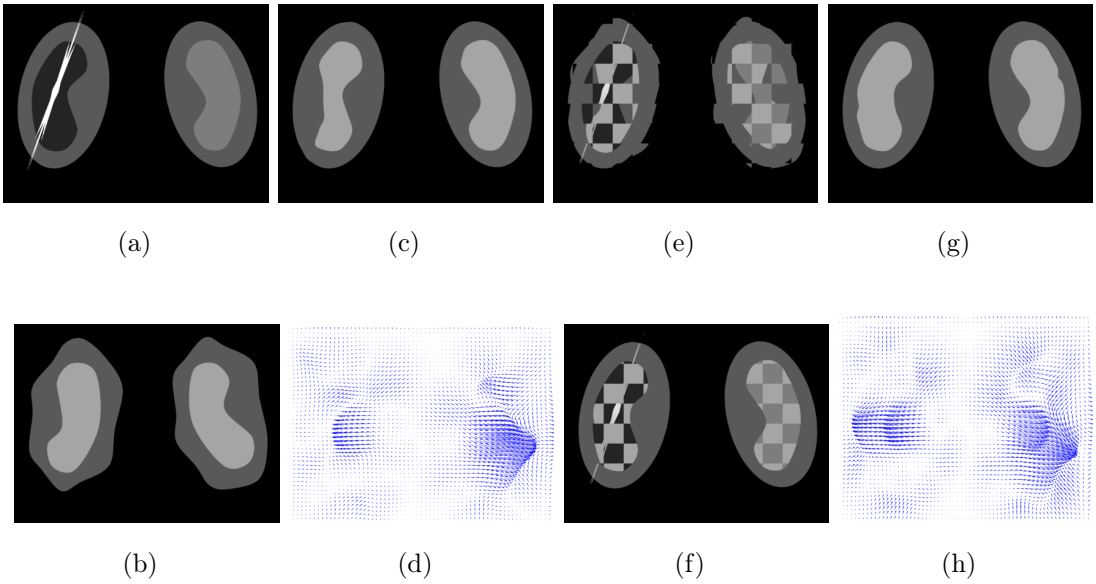


Figure 3.11: Mutual Information and Statistical Priors in the presence of image artifacts. (a) Reference image, (b) initial alignment image, (c) alignment after MI, (d) respective deformation field, (e) checkerboard visualization of (a) and (b), (g) alignment after SIP, (h) respective deformation field, and (f) checkerboard visualization of (a) and (g). Statistical intensity priors seems invariant to strong simulated artifacts.

changes and severe organ motion on simulated as well as synthetic and real clinical data sets.

Simulating Image Artifacts

This completely synthetic experiment investigates a registration problem that involves a simulated image reconstruction artifacts. The image of Fig. 3.11a contains a streak artifact of varying bright intensities as they are noticed in the reconstruction of metal implants in CT imaging. Data-driven image registration criteria such as MI tend to fail to correctly register images containing such streak-like artifacts (see Figs. 3.11b, 3.11c, and 3.11d). In this experiment, we use statistical intensity priors of rather sparse distributions describing the few intensity matchings of the individual components without the streaking artifact. Images 3.11f, 3.11g, and 3.11h illustrate the successful registration when using statistical intensity priors. The checkerboard visualization in image 3.11f nicely display the advantage of using statistical intensity priors. A checkerboard visualization shows how well two images match locally by alternating tiles of the two images like black and white alternate on a checkerboard. We observe statistical intensity priors to be more robust than MI with regards to bright image artifacts.

The subsequent experiments validate these preliminary and playful experiments (Figs. 3.4 and 3.11) more extensively and thoroughly on a variety of simulated and real clinical data sets.

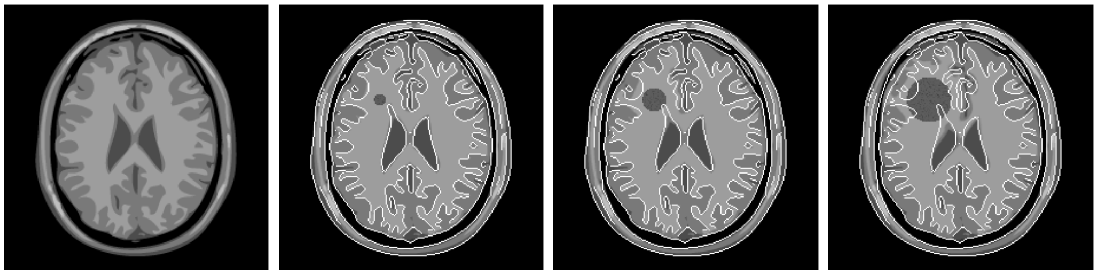


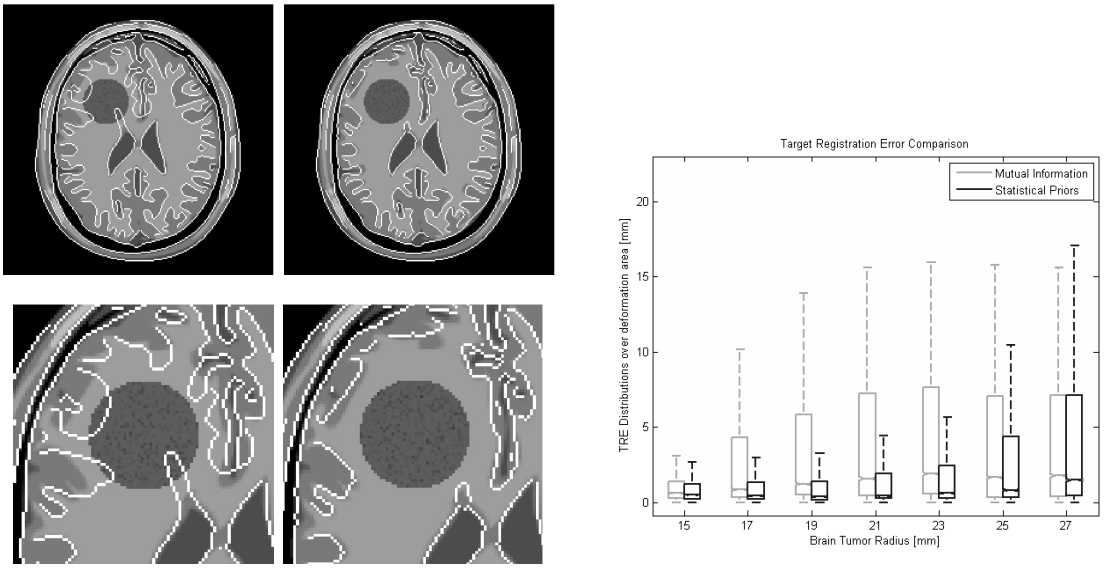
Figure 3.12: Examples of brain MR phantom images simulating expanding lesion tissue. A normal brain image (left) is overlayed as an edge map on phantom images including lesions of size $r = 5\,\text{mm}$, $r = 10\,\text{mm}$, and $r = 19\,\text{mm}$. Corresponding ground truth deformation fields are used to validate the registration algorithms.

Synthetic Brain MRI

To simulate changes in topology induced by expanding lesion tissue, we apply synthetic deformations to 2-D brain MR phantom images and add a simple lesion intensity model. The resulting images are typical in the clinical application of lesion follow-up (FU) studies where registration is being used to quantify lesion growth. Figure 3.12 shows sample images of synthetic lesion tissue visualizing this clinically relevant problem. It can be observed how neighboring tissue is being pushed away with increasing lesion size. The synthetic deformation is generated by locating eight radial basis functions (RBF) around a simple, circular lesion model that simulates lesion intensities by using an additive Gaussian noise model: $I_{\text{lesion}}(\boldsymbol{x}) = c + n(\boldsymbol{x})$, $\forall\, \boldsymbol{x}$ inside the lesion with c being a constant. The ground-truth deformation field $\boldsymbol{u}_{\text{gt}}(\boldsymbol{x})$ becomes a linear combination of a set of compactly supported RBFs irregularly spaced over the image domain: $\boldsymbol{u}_{\text{gt}}(\boldsymbol{x}) = \sum_{i=1}^{K} c_i \Phi(\boldsymbol{x} - \boldsymbol{x}_i)$. We use $\Phi(\boldsymbol{x}) = \phi\left(\frac{\|\boldsymbol{x}\|_2}{s}\right)$, $\boldsymbol{x} \in \mathbb{R}^2$, with a band-limited polynomial kernel $\phi(k) = (1-k)_+^4(3k^3 + 12k^2 + 16k + 4)$ for $k \geq 0$, where $(1-k)_+ = \max(1-k, 0)$, s is a scale for the basis function, and $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^2$, see [Rohd 03]. The parameters s and $\|\boldsymbol{x}\|_2$ are chosen such that deformation vectors are still larger than 0 at $3r$ away from the lesion center, with r being the lesion radius. In the following, we use radius r as a reference to the size of the local deformation induced by the lesion. We are aware that more elaborate brain deformation models exist (e.g. as proposed by Clatz *et al.* [Clat 05]). Nevertheless, our phantom model already points to strengths and weaknesses of the validated algorithms and serves well for the purpose of validating non-rigid registration algorithms.

Note that the ground truth deformation model differs from the dense deformation model that we employ during registration. Since in clinical applications the underlying type of deformation is unknown in most cases, a robust motion compensation algorithm needs to allow for a variety of deformation types. For registration accuracy validation, we compute the target registration error (TRE) that is generally defined using all pixels within a region of interest:

$$\text{TRE}(\boldsymbol{x}) = \left\| \boldsymbol{u}_{\text{gt}} \circ \hat{\boldsymbol{u}}^{-1}(\boldsymbol{x}) - \boldsymbol{x} \right\|_2 , \ \forall\, \boldsymbol{x} \in \text{ROI} \tag{3.37}$$



(a) NMI (b) SIP (c) TREs for Increasing Lesion Sizes

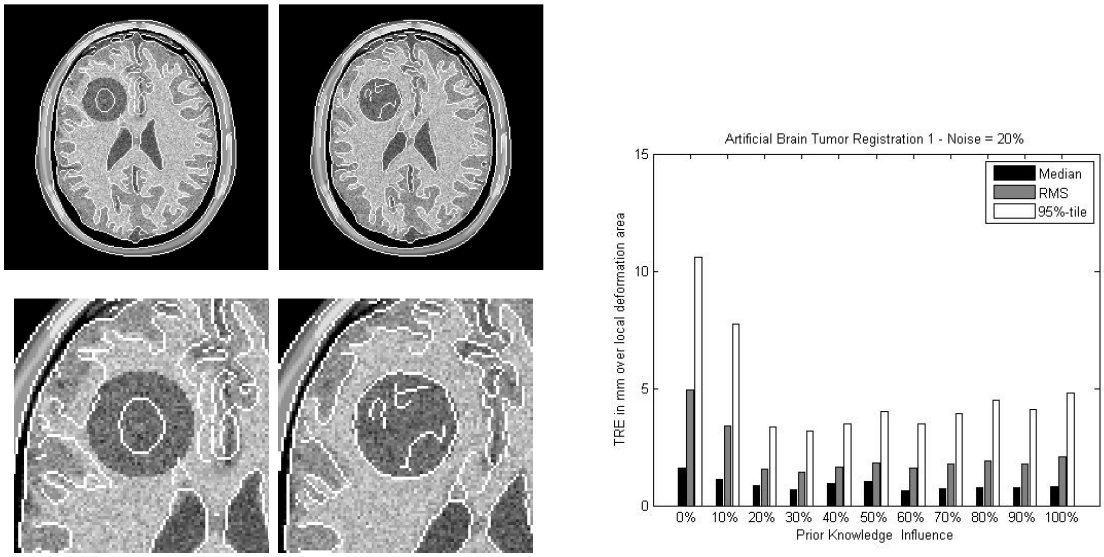
Figure 3.13: Synthetic brain MRI registration experiment on handling topological changes. Example registration results, (a) and (b), of a brain image without (edge map overlay) to an image with a lesion of size $r = 19\,\text{mm}$. While similar structures are mapped correctly using statistical priors, MI gets trapped in local optima caused by the appearing lesion. The proposed combined model using statistical intensity priors shows a larger capture range than MI reflected by significantly lower TREs for larger lesion sizes (c).

with $\boldsymbol{u}_{\text{gt}}$ and $\hat{\boldsymbol{u}}$ being the ground-truth and the retrieved deformation field, respectively. In order to avoid estimating the inverse of the ground-truth deformation field, we approximate the TRE as follows: $\text{TRE}(\boldsymbol{x}) \approx \|\boldsymbol{u}_{\text{gt}}(\boldsymbol{x}) - \hat{\boldsymbol{u}}(\boldsymbol{x})\|_2$.

The phantom allows us to quantify registration accuracy and robustness for varying lesion sizes representing varying deformations. We investigate algorithm performance, both qualitatively and quantitatively, w.r.t. local optima, image noise, and presence of topological changes. The two compared registration criteria are MI and the proposed combined statistical prior model that we will refer to as statistical intensity priors (SP). Moreover, we study the influence of the control parameter α_1 on retrieving the correct displacement field. Note that the proposed framework achieves registration based on maximization of MI when $\alpha_1 = 0$.

Accuracy and Robustness Validation

Aligning a normal brain image to brain images containing lesions of varying sizes (i.e. $r \in [15, 27]\,\text{mm}$) allowed the validation of the proposed method w.r.t. topological changes and indicate the extent of the method's capture range. The prior in this experiment is derived from the joint intensity histogram of the reference image with a copy of itself plus the lesion tissue and $\alpha_1 = 0.4$. Figure 3.13 shows sample MI and SP registration results superimposed as edge maps on the alignment image for $r = 19\text{mm}$ and plots the respective TREs for all validated lesion sizes. It can be observed that



(a) NMI (b) SIP (c) TREs for Increasing Prior Influence

Figure 3.14: Effect of prior influence on registration performance for noisy MR images simulating lesion follow-up studies. Alignment image, $r = 10\,\text{mm}$, is superimposed on reference image, $r = 19\,\text{mm}$, after registration using (a) MI and (b) 40% statistical prior influence. The TRE statistics (c) for increasing amounts of prior knowledge (note MI = 0%) suggest that there is an optimal choice for control parameter α_1 at about 20%-40% influence.

purely data-driven MI registration fails already for $r = 17\text{mm}$ (Fig. 3.13c) due to the existence of local optima that deteriorate MI to map local structures like the gray matter with lesion tissue (Fig. 3.13a). SP proofs to be more robust (Fig. 3.13b) and prevails with a much larger capture range (i.e. up to $r = 25\text{mm}$ as shown in Fig. 3.13c).

Parameter Robustness

Our method requires the selection of several parameters that affect the overall performance. We found in our experiments that the method is quite robust to variations in σ' and α_2. The control parameter α_1, however, can influence the registration result more significantly. Therefore, we conducted numerical and quantitative experiments in order to find the parameter's optimal values for a wide range of applications. Here, we registered two synthetic brain MR images that both include lesions (e.g. $r_{\text{align}} = 10\text{mm}$ and $r_{\text{ref}} = 19\text{mm}$). The images were further altered by additive Gaussian noise. Note that the prior model does not include knowledge about the added noise and is generated from two identical synthetic lesion images as shown in Fig. 3.12. Consistent with the previous experiment, MI based registration gets mislead to local optima whereas SP using $\alpha_1 = 0.4$ robustly aligns the noisy images (Figures 3.14a and 3.14b). Varying α_1, as plotted in Figure 3.14c, and comparing registration results shows an optimum range for α_1 in $[0.2, 0.4]$. Due to the difference of prior and observed distribution incurred by the added noise, registration accuracy decays as the influence of the prior

model increases. Nevertheless, any influence of the prior model already improves the registration accuracy of MI ($\alpha_1 = 0$).

The high accuracy of the integrated SIP model is mirrored in a root-mean-square (RMS) error of 2 mm to the ground-truth deformation in both experiments.

3.6 Summary

We have introduced and validated a new integrated statistical intensity prior model for both rigid and non-rigid transformation models. The introduced Bayesian framework allows to incorporate multiple complementing influences that have a strong impact for practical implementations and applications of the SIP model. We have shown that the limitation to only one instance of prior knowledge still allows for accurate and robust results in clinical applications addressing practical requirements such as robustness or shortcomings such as limited availability of accurate prior knowledge. In chapter 5, we will present a systematic validation of the introduced prior models on routine medical image data in the context of two clinical applications. On simulated data, prior models demonstrated an increased capture range, higher accuracy, and larger robustness when compared to standard intensity-based registration methods.

Chapter 4

Generalized Statistical Intensity Prior Model

In the previous chapter, we developed a new integrated statistical intensity prior model that combines samples from previously aligned and currently observed image data. The integrated SIP model has a large capture range, which means that the two volumes can be initially aligned far away from the optimal target alignment. The integrated SIP model further shows more robust and accurate results where traditional methods fail to align properly. It is, however, only capable of considering one instance of prior knowledge at a time, which is a practical limitation once a larger database of previously registered data is available.

In this chapter, we generalize the SIP formulation to a non-parametric modeling of the space of prior joint intensity distributions. The prior model is defined as a kernel density estimate on this infinite-dimensional function space that embeds the available prior distributions. The variational gradients derived from this new *generalized SIP* (GSIP) contribute to one total gradient embodying the contribution of all relevant joint prior distributions. Therefore, the accuracy of the model is improved by leveraging comprehensive prior knowledge, while maintaining other desirable properties of the integrated SIP, such as extended capture range and increased robustness. The integrated SIP is a special case of the GSIP. In order to compare between non-rigid registration using the GSIP and traditional registration methods, we performed a quantitative analysis on a series of medical and non-medical images and show corresponding results.

In the following derivation of a statistical framework, we assume that we are given an entire set of correctly registered pairs of image data sets. From these we can compute respective joint intensity distributions and construct a nonlinear statistical prior given by a kernel density estimate on the space of joint intensity distributions. It can be introduced into the registration process in the framework of Bayesian inference as described in the previous chapter. As a consequence, the subsequent image matching process is not only driven by a maximization of statistical dependence of the individual intensity distributions, but it will also favor matching results for which the resulting joint intensity distribution is statistically consistent with the set of learned joint intensity distributions. The following section introduces the generalization of the SIP as well as the concept of statistical consistency.

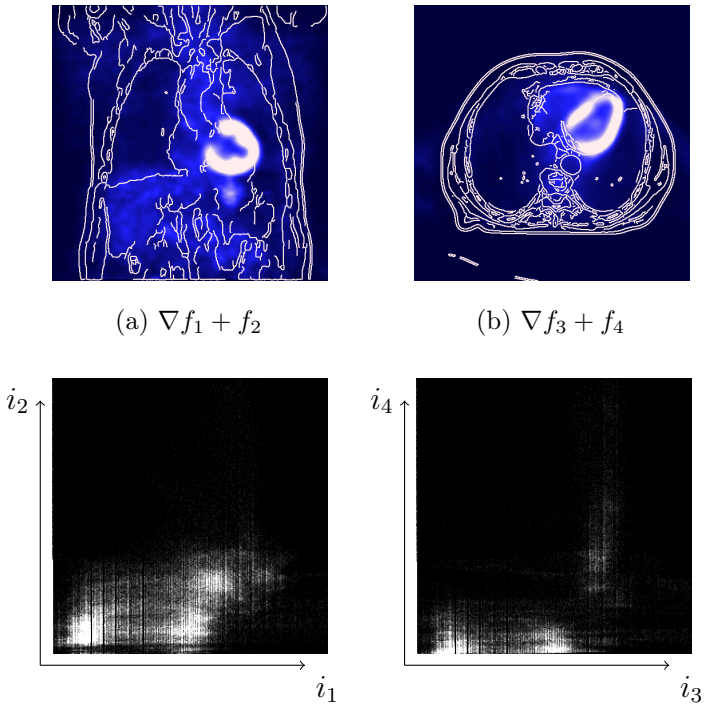


Figure 4.1: (a)-(b) Two PET slices overlayed with image edges from CT visualizing the image alignment in coronal and axial view. Bottom row shows corresponding joint intensity distributions. Bright colors indicate a higher count of PET/CT intensity pairs, whereas dark areas denote their absence in the medical data.

4.1 Generalized SIP

In the context of medical images, we found that a single joint intensity distribution is not sufficient to describe the variability of observed intensity correspondences: Given a set of pairs of matched images from different modalities, one finds great variations among the estimated joint distributions. Figure 4.1 shows coronary slices as obtained with a PET and a CT scanner, registrations of this pair of coronary slices and of a respective pair of axial slices. The inferred joint intensity histograms characterizing the intensity correspondence look quite different. Table 4.1 shows the KL distances between pairs of joint intensity distributions, each of which is estimated from a registered pair of medical images. These observed variations are due to different pairings of imaging modalities (PET, CT, SPECT), different acquisition protocols, or simply due to the selection of slices from different areas of the same scan, i.e. to a variation in the field of view.

Assume we are given a representative set of pre-registered images $\{f_1^j, f_2^j\}_{j=1,\dots,m}$, where $f_k^j : \Omega \subset \mathcal{R}^n \to \mathcal{R}$. These pairs of image data may be obtained from various image modalities and field-of-views. Each registered data set gives rise to a specific joint intensity distributions $p_j(i_1, i_2)$, stating which intensities i_1 and i_2 are likely to be in correspondence for the given image pair. The goal of this work is to derive means to impose this knowledge into variational image registration algorithms. Recall

KLD	PET/CT Whole Body	PET/CT Lungs	SPECT/CT Kidneys
PET/CT Whole Body	0.0	0.7740	3.9609
PET/CT Lungs	0.4871	0.0	3.8275
SPECT/CT Kidneys	2.6614	2.5604	0.0

Table 4.1: KL divergences for aligned, sample medical data showing the dissimilarity between joint intensity distributions (each of which was computed from the registration of respective image/volume pairs), as shown in Figure 4.1. The matching of slices requires different priors on the intensity correspondence, depending on which imaging modalities, which slice locations and which acquisition protocols are used.

the definition of the joint intensity distribution using a Parzen-window estimator and given a particular deformation field $\boldsymbol{u}$ as presented in (3.1):

$$p_{\boldsymbol{u}}(i_1, i_2) \equiv \frac{1}{|\Omega|} \int_{\Omega} G_{\sigma}(i_1 - f_1(\boldsymbol{x}), i_2 - f_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})))\mathrm{d}\boldsymbol{x},$$

Similar to (3.2), we formulate the registration problem as maximization of the conditional distribution

$$\begin{aligned} \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, f_1, f_2, \{p_j\}\big) &= \frac{\mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\, \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big)}{\int_{\boldsymbol{u}, p_{\boldsymbol{u}}} \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\, \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big) d\boldsymbol{u} dp_{\boldsymbol{u}}} \\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}\big)\, \mathcal{P}\big(\boldsymbol{u}, p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big) \\ &\propto \mathcal{P}\big(f_1, f_2 \,\big|\, \boldsymbol{u}, p_{\boldsymbol{u}}\big)\, \mathcal{P}\big(p_{\boldsymbol{u}} \big|\, \boldsymbol{u}\big)\, \mathcal{P}\big(\boldsymbol{u}\big)\, \mathcal{P}\big(p_{\boldsymbol{u}} \,\big|\, \{p_j\}\big), \end{aligned} \tag{4.1}$$

with respect to the displacement field $\boldsymbol{u}$. Again, proportionality in the above expressions means that we have only neglected factors that do not depend on the displacement field $\boldsymbol{u}$ and thus do not affect the maximization. The optimization problem (4.1) splits into four factors with the first three being already introduced in the previous chapter (i.e. the data, the intensity normalization, and the prior smoothness term).

After understanding the role that the prior term of the integrated SIP model plays in (3.2), we will now derive the fourth term of the final expression in (4.1) as a nonparametric prior on the space of joint intensity distributions.

Consistency with Learned Distributions

The theoretical basis of probability distributions on infinite-dimensional function spaces is provided by the theory of Gaussian processes [Rasm 05]. A Gaussian process is defined as a *collection of random variables, any finite number of which have a joint Gaussian distribution.* This definition automatically implies a *consistency* requirement, which is also known as the marginalization property. In other words, the examination

of a larger set of variables does not change the distribution of the smaller set. A Gaussian process is completely specified by its mean function and covariance function. Considering the joint intensity distributions as random variables, we propose to model the prior density as follows.

Given a set of joint intensity distributions $\{p_j\}_{j=1,\dots,m}$ obtained from a set of optimally registered image pairs, we can revert to concepts from kernel density estimation [Silv 92, Coma 02, Huan 04, Crem 06c] in order to derive the following prior on the space of joint intensity distributions:

$$\mathcal{P}(p_{\boldsymbol{u}} \mid \{p_j\}) \propto \frac{1}{m} \sum_{j=1}^{m} \exp\left(- \frac{\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}} || p_j)^2}{2\,\sigma_l^2} \right), \tag{4.2}$$

where

$$\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}} || p_j) = \frac{1}{2} \Big(\mathcal{I}_{\text{sip}_{\text{KL}}}(p^o_{\boldsymbol{u}} || \overline{p}) + \mathcal{I}_{\text{sip}_{\text{KL}}}(\overline{p} || p_j) \Big), \tag{4.3}$$

denotes the JS divergence measuring the dissimilarity between the intensity distribution $p_{\boldsymbol{u}}$ (induced by matching f_1 and f_2 under the displacement $\boldsymbol{u}$) and the previously learned joint distribution p_j using the following components:

$$\mathcal{I}_{\text{sip}_{\text{KL}}}(\overline{p} || p_j) = \iint_{\mathbb{R}} \overline{p}(i_1, i_2) \log \frac{\overline{p}(i_1, i_2)}{p_j(i_1, i_2)} \mathrm{d}i_1 \mathrm{d}i_2 \tag{4.4}$$

$$\overline{p} = \frac{p^o_{\boldsymbol{u}} + p_j}{2}. \tag{4.5}$$

In the optimization of (4.1), the distribution (4.2) therefore imposes statistical similarity between the inferred intensity correspondence $p_{\boldsymbol{u}}$ and the previously observed joint intensity distributions $\{p_j\}_{j=1,\dots,m}$. The kernel width σ_l in the density estimator is fixed to the average nearest neighbor distance computed for the set of joint intensity distributions $\{p_j\}$:

$$\sigma_l = \frac{1}{m} \sum_{i=1}^{m} \min_{j \neq i} \mathcal{I}_{\text{sip}_{\text{JS}}}(p_i || p_j). \tag{4.6}$$

The intuition behind this choice is that the width of the Gaussians is chosen such that on average the next prior joint intensity distribution is within one standard deviation. More sophisticated estimates, for example using cross validation, are conceivable and we refer the reader to [Silv 92].

In the context of Gaussian processes, equation (4.2) can be understood as the linear predictor using the *squared exponential* covariance function. It is interpreted as the linear combination of n kernel functions, each one centered on a training point (i.e. prior joint distribution). Now that the fourth factor in the inference problem (4.1) is specified, we can maximize this probability by minimizing its negative logarithm, which is given by an energy of the form:

$$E(\boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}) = \mathcal{I}_{prior}(p_{\boldsymbol{u}}, \{p_j\}) + \alpha_1 \mathcal{I}_{data}(p_{\boldsymbol{u}}) + \alpha_2 \mathcal{I}_{smooth}(\boldsymbol{u}), \tag{4.7}$$

where these three energies impose several constraints with α_1 and α_2 steering the influence of data and smoothness term. The energy $\mathcal{I}_{prior}$ guarantees that the joint intensity distribution induced by a displacement field $\boldsymbol{u}$ is consistent with previously

observed joint intensity distributions. According to (4.2), (3.23), and (3.27) the three energies are given by:

$$\mathcal{I}_{prior}(\boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\}) = -\log\left(\sum_{j=1}^{m} \exp\left(-\frac{\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)^2}{2\,\sigma_\ell^2}\right)\right), \tag{4.8}$$

$$\mathcal{I}_{data}(\boldsymbol{u}, p_{\boldsymbol{u}}) = -\mathcal{I}_{\text{MI}}\big((f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u})\big), \tag{4.9}$$

$$\mathcal{I}_{smooth}(\boldsymbol{u}) = \frac{1}{2}\int \sqrt{\sum_{i=1}^{n} \nabla \boldsymbol{u}_i(\boldsymbol{x})^2}\, d\boldsymbol{x}. \tag{4.10}$$

Minimization of the energy (4.7) by gradient descent leads to a partial differential equation for $\boldsymbol{u}$ as shown in (2.11) with individual gradients for each of the above energy terms. For the gradient derivations of $\mathcal{I}_{data}$ and $\mathcal{I}_{smooth}$, we refer the reader to the previous chapter, where the gradient of MI and the well known diffusion term $\Delta\boldsymbol{u}$ are presented for the two energies respectively. Given the previously introduced mathematical framework of variational calculus, we can derive the gradient of $\mathcal{I}_{prior}$ as follows:

$$\frac{\partial \mathcal{I}_{prior}(\boldsymbol{u}, p_{\boldsymbol{u}}, \{p_j\})}{\partial \boldsymbol{u}} = \sum_{j=1}^{m} \gamma_j \, \frac{\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)}{\sigma_\ell^2} \, \frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)}{\partial \boldsymbol{u}}, \tag{4.11}$$

with normalized weights:

$$\gamma_j = \frac{\hat{\gamma}_j}{\sum_i \hat{\gamma}_i} \tag{4.12}$$

$$\hat{\gamma}_j = \exp\left(-\frac{\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)^2}{2\,\sigma_\ell^2}\right). \tag{4.13}$$

We note that the gradient in (4.11) provides better convergence properties than using an exponential kernel in (4.2) as the gradient goes to zero for $p_{\boldsymbol{u}} \to p_j$ due to the additional term $\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)/\sigma_\ell^2$. In the previous chapter, we have derived the gradient of $\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}||p_j)$ with respect to the displacement field $\boldsymbol{u}$ as:

$$\frac{\partial \mathcal{I}_{\text{sip}_{\text{JS}}}(\boldsymbol{u})}{\partial \boldsymbol{u}} = -\frac{1}{2\,|\Omega|}\left[\left(\frac{\partial_2 p^0_{\boldsymbol{u}}(i_1,i_2)}{p^0_{\boldsymbol{u}}(i_1,i_2)} - \frac{\partial_2 \overline{p}_{\boldsymbol{u}}(i_1,i_2)}{\overline{p}_{\boldsymbol{u}}(i_1,i_2)}\right) * \mathrm{G}_\sigma\right](f_1(\boldsymbol{x}), f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x})) \cdot \nabla f_2(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))).$$

The interpretation of the additional term (4.11) in the evolution of the displacement field $\boldsymbol{u}$ is quite intuitive: It induces a change in the estimated displacement field $\boldsymbol{u}$ that aims at minimizing the JS-divergence $\mathcal{I}_{\text{sip}_{\text{JS}}}(p_{\boldsymbol{u}}, p_j)$, thereby drawing the current intensity distribution $p_{\boldsymbol{u}}$ toward the previously learned distributions $\{p_j\}$. More precisely, the energy gradient exerts a force on the estimated intensity distribution toward each learned intensity distribution p_j, which is modulated by a weight γ_j that decays exponentially with the distance between the intensity distributions — see equation (4.13). Thus this additional term comes into play only for those learned distributions

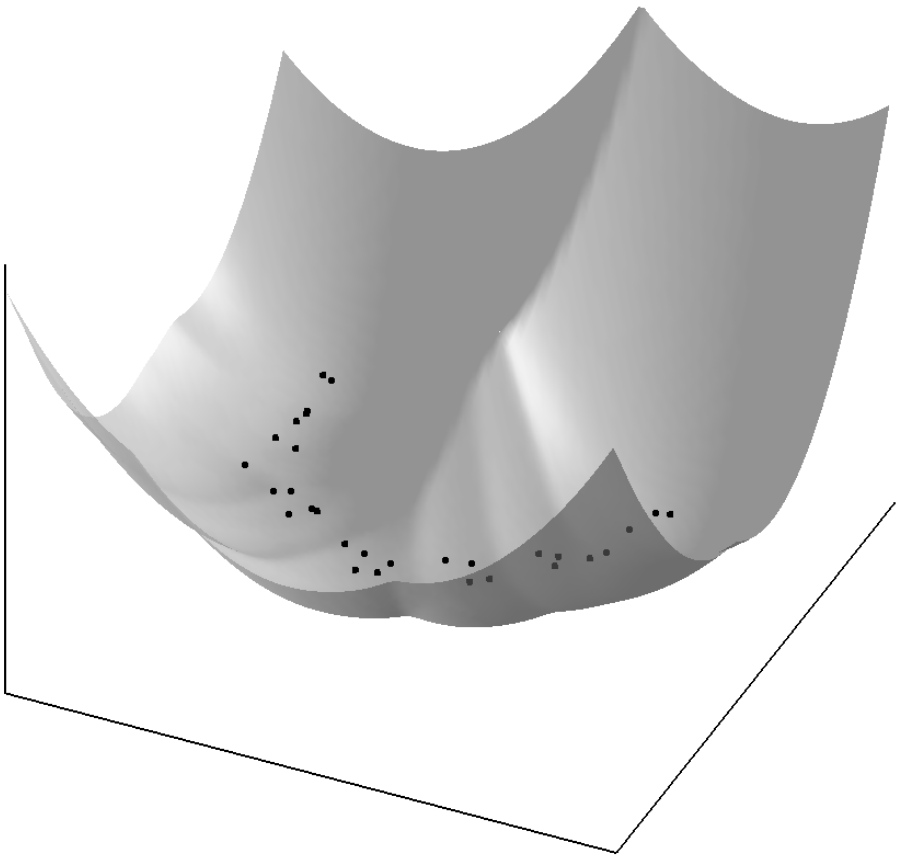


Figure 4.2: Schematic plot of energy (4.8). Each black point represents a joint intensity distribution. The energy (4.8) measures the dissimilarity between a given joint intensity distribution and the previously learned distributions.

that are most consistent with the currently estimated intensity distribution. And this is precisely the mechanism by which the algorithm "decides" which intensity distributions among the learned ones are to be used for a given registration task.

To further clarify this effect of the multimodal energy (4.8) we refer to the visualization in Figure 4.2: In this schematic drawing, each joint intensity distribution p_j is represented as a black 2D point. The energy (4.2) generated by all learned training points is shown as a shaded surface. It essentially extends the JS-divergence to a dissimilarity with respect to an entire set of joint intensity distributions. During the optimization process it constrains the displacement field such that the corresponding intensity distribution remains within the valleys of low energy. This ensures that the joint intensity distribution will favor similarity to previously learned intensity distributions during the optimization.

What does it mean that the current joint intensity distribution is forced to be similar to one or the other previously learned intensity distribution? To this end, let us consider the following very simple example. Assume we have learned two joint distributions, where the first one states that white pixels in image 1 are always associated with black pixels in image 2 and vice versa, while the second one states that the matching of white-to-white and black-to-black is most likely. Then enforcing similarity to one or the other by energy (4.8) has the following effect: If during optimization pairs of white pixels are associated through the displacement field, then this induces proximity to the second learned intensity distribution, and the prior will automatically enforce that black should also be associated with black – because a matching of white-to-white on one hand but black-to-white on the other is not consistent with any of the two learned intensity distributions. In other words: The matching of certain intensities will provide clues for the matching of others, as indicated by the learned joint distributions.

The above example illuminates the idea of imposing a prior on the space of joint intensity distributions. Note that this is fundamentally different from learning a *single* joint intensity distribution, as proposed for example by Leventon and Grimson [Leve 98]. Firstly, our method allows for a large variety of different intensity distributions. Secondly, the inherent selection mechanism allows the algorithm to infer statistical relations between matching of different intensity pairs, as in the simple example of two joint distributions discussed above.

4.2 Experiments and Results

In the following, we will evaluate the proposed statistical framework for image registration. In Section 4.2.1, a quantitative study on a SPECT - CT image pair shows that priors on the joint intensity distribution can improve the mutual-information-based registration process by increasing the basin of attraction and by shifting the location of the energy minimum to the correct one. In Section 4.2.2, a study on the registration of a PET - CT image pair shows that the proposed multimodal prior on the joint intensity distribution outperforms a simpler unimodal prior, because the multimodal one allows the registration process to "select" among appropriate joint distributions. Section 4.2.3 shows that the proposed prior allows to cope with partial occlusions in a face registration task.

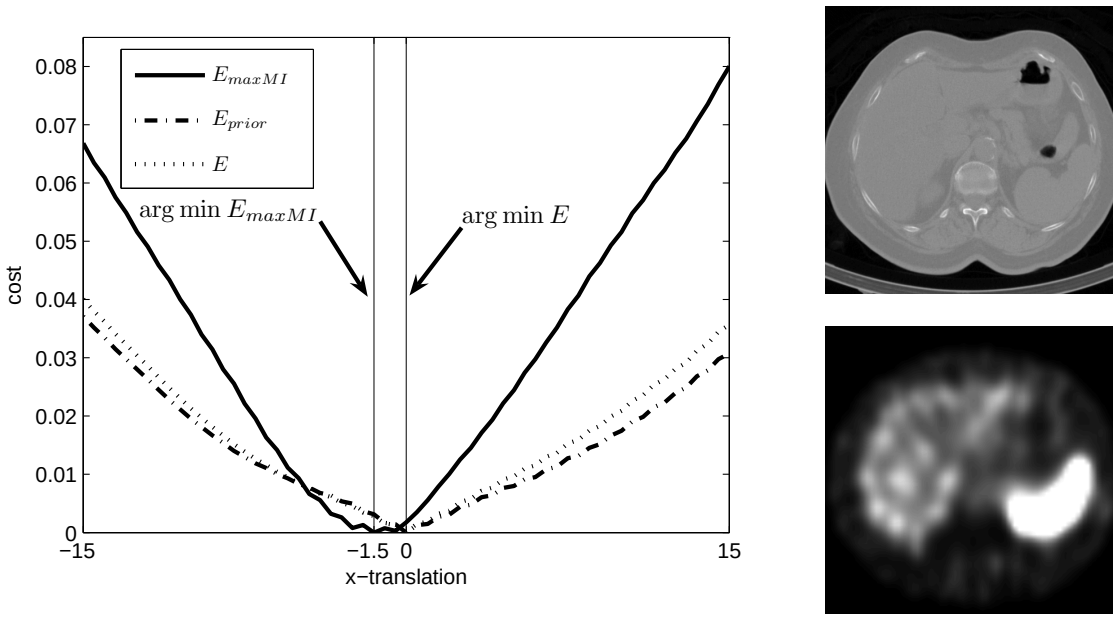


Figure 4.3: SPECT-CT registration performance analysis. SPECT slice was shifted horizontally within a range of $-15mm$ to $15mm$. Mutual information is noisy around the optimum and its minimum actually corresponds to an incorrect alignment. The integration of a prior on the joint intensity distribution provides for a larger basin of attraction and enables the estimation of the correct alignment.

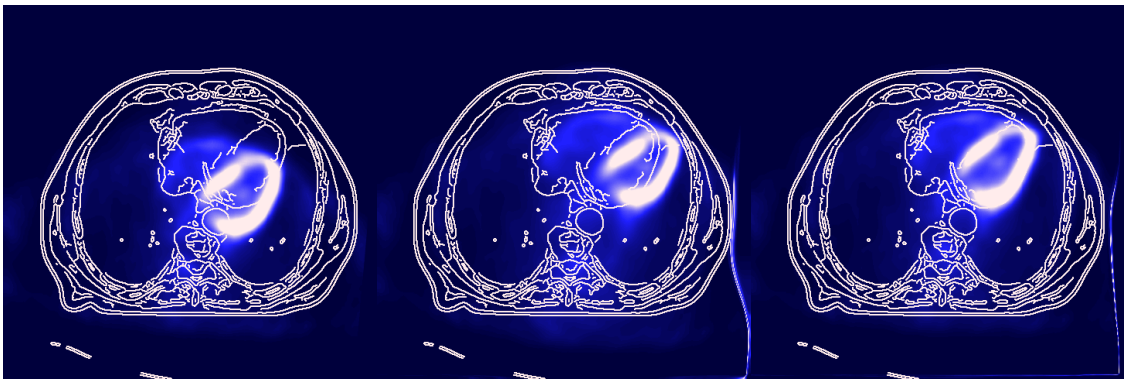
All implementations are done within a multi-resolution framework, giving computation times around 10 seconds for image pairs of size 450×450.

4.2.1 Quantitative Evaluation

Assume we are given a perfectly aligned image data set, such as the SPECT - CT image pair acquired by a Siemens Symbia T2 hybrid scanner in Fig. 4.3a. Now we use these data to study the performance of competing objective functions, e.g. $E_{maxMI} = \max(\mathcal{I}_{\mathrm{MI}}) - \mathcal{I}_{\mathrm{MI}}$ for MI, E_{prior} in (4.8), and the total energy E in (4.7). In this experiment, the SPECT slice was shifted horizontally, while the CT image remained fixed, and the respective values of all three objective functions are computed, see Fig. 4.3a. The energy plots show quantitatively that incorporating a prior (computed from the correctly aligned image pair) will lead to a superior registration algorithm. While this is only shown for the case of translation, one can expect similar improvements for non-rigid deformations.

4.2.2 PET/CT Medical Image Registration

Given several training image pairs, the proposed prior can incorporate a variety of joint intensity distributions. The following experiment will show that among this complementary information, the proposed algorithm selectively choses the intensity information appropriate for a specific registration task.



(a) (b) (c)

Figure 4.4: PET/CT registration. (a) deformed PET/CT, (b) registration result using average prior, (c) registration result using both priors. While using the averaged joint intensity distribution as a prior leads to misregistration, the proposed multi-modal prior on the joint intensity distribution allows for the correct registration.

The training data is composed of two sets of aligned PET/CT slices[1] acquired by a Siemens hybrid scanner, Figs 4.1a, 4.1b. The superimposed edge maps (white) illustrate structural information of the CT image and visualize the quality of alignment. Note the significant difference in Figure 4.1 between the two shown joint intensity distributions, which reflect a typical scenario as it occurs in clinical applications.

An artificial deformation is applied to the PET slices and compared to the recovered displacement fields of using only a single prior distribution vs. using two complimentary. The two priors represent the joint distribution of the axial and coronal PET - CT slices shown in Fig. 4.1. In the case of a single prior, the average of the two is used for fair comparison.

Figure 4.4 illustrates the advantage of using several prior distributions as opposed to only using one. The weighting factors of energy (4.7) are set with a preference towards the prior energy E_{prior}, i.e. α_1 is chosen to be small. The width σ is determined using equation (4.6), and α_2 is chosen to allow for a smooth displacement field.

The results of recovering the significant deformation between the PET and CT images (see Figure 4.4a) are shown in Figs. 4.4b and 4.4c. Using an average distribution misleads the algorithm and registration fails, see Fig. 4.4b. However, the proposed method can fully utilize the given priors and correctly "selects" the closest joint intensity distribution. As a result, the underlying deformation is fully recovered (Fig. 4.4c).

This experiment shows the strength of introducing a space of joint intensity distributions, while the algorithm is able to chose the best available prior information. In case no best information is available, the prior energy decays to zero, and performance will be at least as good as using a context-free similarity measure.

[1]PET and SPECT are nuclear imaging techniques which visualize centers of high activity in the human body, as described in sections 1.2.3 and 1.2.2.

4.2.3 Face Registration in the Presence of Occlusion

Non-rigid multi-modal registration can serve as a preprocessing step for face recognition, where facial and/or head motion must be recovered in order to establish correspondence. In the following experiment, we illustrate how prior knowledge on the joint intensity distribution improves the registration in the presence of lighting variation and occlusion. The experiment is to recover facial expressions and head movement between two images. The second image is taken under different lighting conditions with the person wearing sun glasses. The objective functions of comparison are (i) purely MI based registration and (ii) the proposed combined approach using prior knowledge in eq. (4.7). The first row of Fig. 4.5 shows the two pairs of manually registered training data used to construct the prior, i.e. $m = 2$. To compare the performance, the same images have been registered (i) by the context-free MI criterion and (ii) by additionally imposing a prior on the space of joint intensity distributions. The parameters used here are similar to previous experiment. The second row of Fig. 4.5 shows the reference and alignment images that are subject to registration. Those images show multi-modality due to a slight illumination change, but moreover due to the appearance of the sun glasses. Furthermore, Figure 4.5 illustrates the differences of learned data towards the current data.

There are two runs for each objective function that are being compared. Figure 4.6 shows the achieved results for pure MI and for imposing a space of prior information. Since the underlying transformation is unknown, the edge map of the alignment image is superimposed on the reference image for performance comparison. Column 4.6a shows the initial positions of the faces, column 4.6b shows the results using pure MI, and column 4.6c plots the results of the energy in equation (4.7). Comparing the edge maps it can be noticed that the proposed energy (4.7) is superior to using pure MI. The MI method matches the outline of the persons correctly but fails to match the glasses in the alignment image on the eye region of the reference image. The combined method, however, succeeds for both faces in establishing correspondence, see column 4.6c. Note that our method selects the prior intensity distribution, which corresponds best to the current input images.

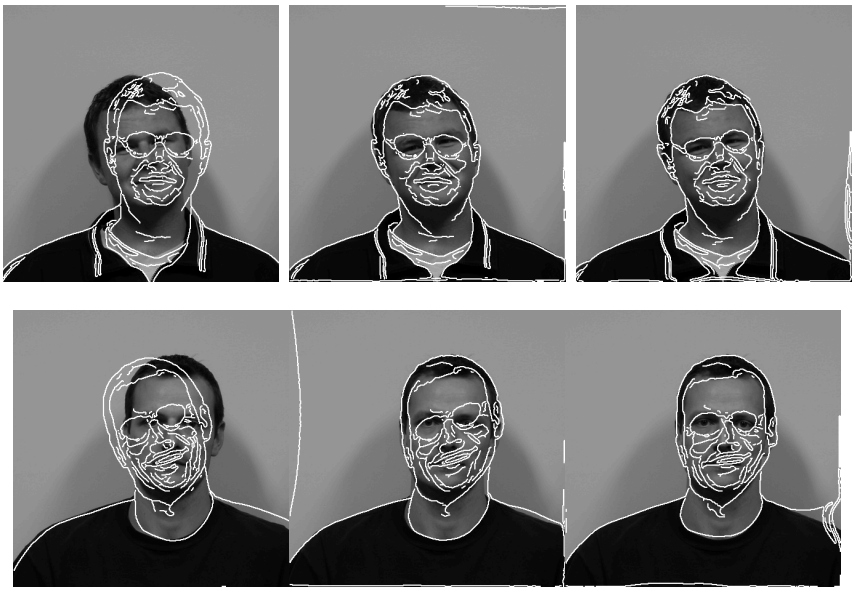
4.3 Summary

In this chapter, we proposed a multimodal prior on the joint intensity distribution in order to enhance image registration problems. While MI was shown to provide a powerful registration criterion, it remains a purely low-level criterion. Our formulation allows to enhance this existing registration method in order to integrate prior knowledge about likely intensity correspondences, which is statistically learned from multiple pairs of pre-registered training images. Experimental results on both medical and face images demonstrate that our approach outperforms purely MI based image registration.

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

(e) (f) (g) (h)

Figure 4.5: Face images used for training and registration. (a)-(d) training images, (e)-(h) reference and alignment images that are subject to registration. The latter pose a challenging registration task and slightly differ from the training data.



(a) (b) (c)

Figure 4.6: Face image registration results. Column (a) shows initial alignment of the two images, column (b) the final registration for pure MI-based energy, and column (c) illustrates the final registration using energy (4.7). The energy (4.7) shows to be superior to context-free MI energy by minimizing the distance towards previously learned intensity distributions.

Chapter 5

Medical Imaging Applications: Clinical Algorithm Assessment

5.1 Introduction

Medical image processing is widely used within the clinical environment. It has a strong influence on the clinical decision-making process towards diagnosis, surgery planning and guidance. As a natural consequence, high accuracy and robustness are expected from the implied imaging algorithms. However, sources of error are numerous in medical imaging applications with some being common to any processing method, such as partial voluming or intrinsic data variability due to patient movement during the acquisition, and with others being specific to the type of processing, i.e. image segmentation, registration, or classification.

In this chapter, we describe a systematic approach to validating medical image registration that follows the proposal for standardization of imaging algorithm validation in clinical environments by Jannin and colleagues [Jann 06]. In an attempt to standardize terminology and methodology that will make validation studies and results comparable, they propose a validation framework consisting of a reference-based validation procedure and a checklist designed for reporting reference-based validation studies. The subsequent sections briefly describe this framework and illustrate the application of the validation procedure to several routine clinical applications. This systematic validation of achievable registration accuracy and comparison to standard registration methods on more than 100 clinical PET/CT and SPECT/CT data sets is, according to the best of our knowledge, one of the largest registration validation study.

5.2 Systematic Validation of Registration Algorithms

Medical image processing algorithms are usually embedded in larger imaging systems and applications that help doctors reach a diagnosis or that support surgeons during interventional procedures. The algorithms became fundamental blocks of such systems and their performance may have a substantial impact on the performance of the larger

system or application and consequently on the clinical decision-making. Because of this central role the systematic analysis of image processing algorithms is nowadays inevitable to determine whether a theoretically sound method delivers the quality and accuracy required in clinical practice.

Such an assessment of performance is complex and multi-faceted. In order to manage the complexity of performance assessment of image processing algorithms (or any other technology or system), the terms verification, validation and evaluation, borrowed from software engineering, may be distinguished as follows:

(a) Verification: *System Built Correctly?*
Assessing whether the system is built according to the given specifications

(b) Validation: *Correct System Built?*
Assessing whether the system fulfills the purpose for which it is built

(c) Evaluation: *System Valuable?*
Assessing how the system is being accepted by the end-users.

In this work, we are concerned with the validation of image registration algorithms rather than verifying, validating, and evaluating the larger application or system in which they are embedded in. The performance of image registration algorithms depends on a number of factors, such as imaging modality, image content including degrading effects such as imaging artifacts, the class of spatial transformation used for registration, similarity measures, optimization strategies, and details of the implementation. This complex list of factors makes validation of any registration algorithm challenging and impedes a standardized comparison across results of the different algorithms.

There are several interesting activities within the medical imaging community to address these validation concerns by establishing and maintaining a gold standard for medical image registration techniques. Examples of those activities are the "Retrospective Image Registration and Evaluation Project" [West 97] and the "Non-Rigid Image Registration Evaluation Project (NIREP)" [Chri 06].

5.2.1 Algorithm Validation Framework

The model for defining and reporting a reference based validation protocol in image processing by Jannin and colleagues [Jann 06] has been widely accepted in the community. Their proposed validation framework, as shown in Figure 5.1, and their checklist for reporting the validation results serve as a strong standardizing ground work and provide insight into the validation process especially with respect to image registration algorithms. As a matter of fact, one of the major reasons why we adopt this model is that we strongly believe in the standardized validation and reporting of medical image processing algorithms.

Three main components of such a validation framework have been identified and are critical in designing any clinical validation studies:

I. Validation Objective/Hypothesis

II. Validation Datasets

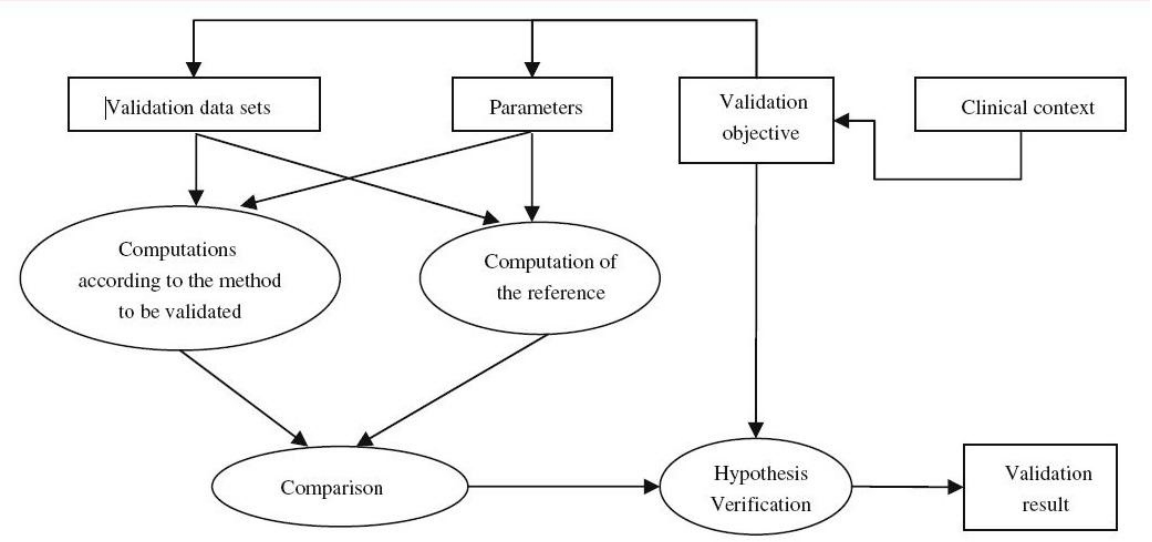


Figure 5.1: Functional diagram describing the main components of the reference-based validation framework proposed by Jannin *et al.* [Jann 06]

III. Validation Criteria

Let us examine Figure 5.1 for understanding the concepts of such validation studies. The crucial starting point of every study is the formulation of a validation objective. It is here where the clinical context determines the goal of the study (top right of Fig. 5.1). After a validation objective (i.e. the hypothesis) is formulated, the proposed validation framework is utilized to run the experiments on the specified validation datasets and using any given parameters. Results from the method under validation are compared with results from the reference method based on a computed validation criteria (left of Fig. 5.1). It will become clear in later sections that the choice of this validation criteria or metric is sometimes challenging and will possibly result in some sort of compromise between feasibility and accuracy. Finally, the comparative results are tested against the validation objective, and in order to conclude with the validation results, the hypothesis is either accepted or rejected.

The described components of the validation framework have been derived as a result of a literature analysis and discussions with experts from the medical imaging community [Jann 06].

5.2.2 Validation Criteria

As we have seen above, image processing methods can be validated according to specified performance criteria and specific validation objectives. In most validation studies found in the technical and clinical literature, validation criteria are assessed against a "gold standard" that is assumed to be close or equal to the "ground truth", e.g. [Guet 07, Jann 06, Spie 09, Joll 10]. In medical image registration, the ground truth is the geometrical transformation that correctly maps locations in one image to the anatomically corresponding locations in the other image. In image segmentation, the ground truth may be the correct anatomical labeling of each pixel or voxel of an image data set or the true structure boundaries.

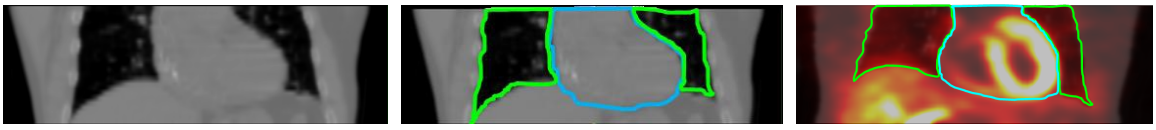
Unfortunately, it is sometimes tremendously challenging or even impossible to generate such ground truth (e.g. in the case of unconstrained motion patterns such as breathing or cardiac motion) from clinical data. A valid compromise and feasible alternative in this scenario is to substitute ground truth among the different processing types. For example, the ground truth generated to validate image segmentation can also serve as a ground truth for registration in applications where the medical decision-making process or algorithm's accuracy is based on correctly overlapping regions or organs. Although this strategy does not exhaustively describe the underlying ground truth, it enables reference-based validation studies where otherwise ground truth generation is impossible.

In summary, the ground truth or reference may originate from numerous sources. It can be an exact or approximate solution based on numerical simulations or physical experiments. It can also be a solution computed using one or several independent image processing methods. Finally, the ground truth can be an expert-generated solution or a solution that was generated using a priori knowledge about the ground truth.

Application Specific Validation Criteria

There could be several criteria employed for determining the performance of registration algorithms in systematic validation studies. In case sufficient ground truth is available, such a validation criteria may be expressed as the Euclidean distance to the reference data such as transformation parameters or mappings of specified anatomical locations (i.e. anatomical markers). For some clinical applications, however, such ground truth is not available or there may be more descriptive measures needed than the root-mean-square (RMS) Euclidean distance. In the case of cardiac studies using hybrid molecular imaging modalities (SPECT/CT, PET/CT), for example, segmentation based validation criteria express registration accuracy based on what is required for the intrinsic application (i.e. accurate alignment of the mediastinum). Some of those criteria measure a mismatch ratio [Chen 06] or a distance between the centers of gravity [Han 08] of dedicated segmented regions in both volumes of a multimodal (i.e. anatomical and molecular) image data set. An application specific validation criteria that is solely based on molecular image data is the SPECT (or PET) uptake measure that accumulates the uptake counts in clinically defined sectors of the heart (e.g. the AHA model [Cerq 02]). This criteria is also known as the quantitative polar plot measure after image reconstruction using a specific image alignment [Fric 04]. Image registration quality can then be validated by comparing the respective quantitative values for different alignments. It will become clear in subsequent section 5.3 how image alignment and reconstruction are related.

The three latter validation criteria for multimodal image alignment have been designed in view of a specific application (e.g. cardiovascular function assessment). Nevertheless, the existing validation criteria measuring cardiac alignment in multi-modal images lack meaningfulness, precision, and reproducibility. Therefore, we have developed a cardiac alignment index for SPECT or PET modalities that is designed to address these issues.



(a) CT (b) CT segmentations (c) Fused View

Figure 5.2: Visualization of the cardiac alignment metric computation. A cardiac CT volume (a) is segmented into lung and mediastinum region (b) and then fused with a SPECT volume (c). The ratio of how much SPECT uptake resides inside the lungs versus inside the mediastinum suggests how well the two volumes are aligned.

The Cardiac Alignment Index

The proposed validation criteria, referred to as the cardiac alignment index (CAI) in the following, is designed specifically for validation of image alignment in cardiac SPECT/CT or PET/CT multimodal imaging. It leverages the fact that SPECT or PET uptake should be maximal within the mediastinum and minimal inside the lungs in case SPECT (or PET) and CT volumes are precisely aligned. In case of misregistration between SPECT or PET and CT images, the uptake activity from liver or heart smears over the regions of lung creating problems with attenuation correction in PET or SPECT images. In the following, we derive the CAI on an example of SPECT/CT data and the derivation for PET/CT data is conferrable.

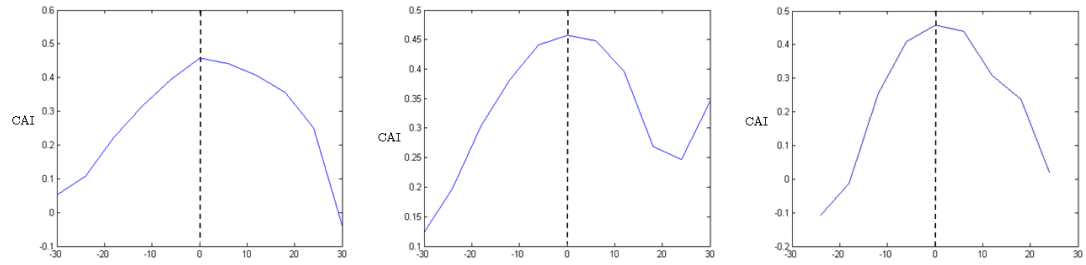
In order to compute the CAI, we first need to segment the mediastinum and lungs from CT either manually or automatically. Given the segmentations, see Fig. 5.2, and a spatially transformed SPECT function $f_{\text{spect}}(x)$, the CAI is defined as follows:

$$\text{CAI} = 1 - \frac{\frac{1}{|B_{\text{lungs}}|}\sum_{\Omega} B_{\text{lungs}}(x) f_{\text{spect}}(x + u(x))}{\frac{1}{|B_{\text{media}}|}\sum_{\Omega} B_{\text{media}}(x) f_{\text{spect}}(x + u(x))}, \qquad |B| \in \mathbb{N}^{+}, \tag{5.1}$$

where $|B|$ denotes the size of the respective segmentation, and $u : \Omega \mapsto \Omega$ describes the deformation field that maps SPECT and CT volumes. The CT segmentation masks for lungs and mediastinum are defined as binary masks B_{lungs} and B_{media} describing lung and mediastinum region, respectively:

$$B_{\text{organ}}(x) = \begin{cases} 1 & x \text{ is inside organ} \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

The masks can be generated using a automatic or semi-automatic image segmentation tool. The generation of our masks was performed using an automatic segmentation tool that offers an interactive, manual correction of the automatic result. We can observe that the CAI accumulates intensities that fall into lungs and mediastinum at a given alignment of CT and SPECT volume and represents the ratio of lung to mediastinum uptake. For changing SPECT functions f_{spect} dependent on the deformation field u, the CAI approaches 1, if the SPECT uptake inside the lungs approaches 0 and becomes infinite inside the mediastinum representing an accurate alignment of SPECT and CT volume. As opposed to other described cardiac alignment criteria such as



(a) T_x (b) T_y (c) T_z

Figure 5.3: The CAI criteria as a function of translation parameters T_x, T_y, T_z for a 3-d PET-CT data set. A distinct and unique optimum at accurate alignment can be observed.

quality control index (QCI) [Chen 06] or quantitative polar plot analysis [Fric 04], the CAI does not require segmentations from low resolution, noisy SPECT volumes. More importantly, it accurately captures all kinds of misalignments due to a unique optimum of uptake value at precise alignment, as shown in Fig. 5.3. The QCI, for example, cannot detect misalignments that shifted the left ventricle of the SPECT heart into the right ventricle of the CT heart. The accuracy of non-rigid registration algorithms in multimodal cardiac applications is objectively measured by employing an accurate, robust, and physician approved [Chin 08] CAI as a validation criteria.

5.3 Misregistration in Hybrid SPECT/CT Scanners

5.3.1 Problem Definition And Clinical Impact

The use of multi-modality imaging through hardware- or software-based solutions in clinical practice has become increasingly popular. For example, the, so called, hybrid SPECT/CT or PET/CT scanners combine low resolution molecular images with high-resolution CT providing anatomical context and accurate attenuation correction of SPECT or PET images by placing both (e.g. SPECT and CT) scanners physically next to each other.

The availability of those scanners generated a new research field of interesting clinical applications such as the accurate fusion of two complementary modalities for diagnostic purposes and CT based attenuation correction for enhancing image reconstruction of molecular image data. The latter is particularly interesting since it increases specificity of SPECT and PET by providing means for uniform attenuation correction [Zaid 03]. Image reconstruction algorithms in hybrid scanners can utilize the CT image as a highly accurate and detailed attenuation map for the attenuation correction step [Chen 06, Goet 07b]. In standalone PET or SPECT scanners, the attenuation map needs to be generated through a second scan, i.e. the transmission scan using Ge^{bd} or Cs^{137} [Cher 03]. This transmission scan requires a significant amount of scan time to acquire a low resolution map of anatomical structures matching the radionuclide distribution for AC that is filled with a substantial amount of noise.

In this section, we are investigating the strengths and weaknesses of the novel application of CT based AC. Moreover, it will become clear that the usage of high resolution CT images for molecular image reconstruction needs to be done carefully in order to prevent the generation of misleading image artifacts. As a matter of fact, this clinical risk of misregistration induces artifacts is significant enough to motivate the clinical research community to frequently report on technological issues involving hybrid scanners. Moreover, the American Society of Nuclear Cardiology (ASNC) updated their *Imaging Guidelines for Nuclear Cardiology Procedures* in 2007 to include the following statement:

> It has been made mandatory to check Registration and Attenuation Correction Accuracy as Quality Control (QC) procedures in hybrid SPECT/CT and PET/CT scanners [Hell 04, Imag 07].

Assuming this misregistration problem in hybrid scanners is clinically significant and established, the question of the potential clinical impact resulting from such a technological failure may be severe. It has been reported, for example, in the application of quantitative cardiac SPECT/CT analysis that spurious perfusion defect artifacts are introduced in the CT-based attenuation correction (AC) images of the SPECT acquisition due to misalignments. The misalignments falsify the uptake values that are utilized for diagnosis in myocardial perfusion imaging [Fric 04, Krit 05]. Therefore, an accurate registration between the two modalities is imperative to ensure the diagnostic confidence of physicians.

As a technical solution to the guidelines by the ASNC, registration as a preprocessing step to CT-based AC SPECT could be used as quality control to ensure that the modalities are well aligned. Therefore, such a registration tool cannot involve user interaction or correction. Certain sources of misregistration could be partly addressed through a stringent acquisition protocol that, nevertheless, is both prone to errors and complicated to use in clinical practice. A more efficient and promising approach of ensuring alignment is to utilize a fully automatic registration technique that is highly accurate and robust.

In order to establish an unbiased summary of literature evidence on the topic of misregistration in hybrid SPECT/CT scanners we performed a systematic literature review and report it in [Chin 08]. The following paragraph summarizes the reported findings of magnitude, etiology, and clinical impact of SPECT/CT misregistration.

The exhaustive search in multiple clinical publication databases yielded 22 out of 121 articles on hybrid SPECT/CT scanners that performed studies on non-phantom data. In order to infer the problem magnitude, the statistical analysis of a final seven studies that quantitatively report misregistration in hybrid SPECT/CT scanners was performed. The result yielded that the weighted average of misregistration occurrence that is equal to or greater than one SPECT voxel (i.e. approximately 4.79 mm) appears to be 41.8% of the total clinical cases with the 95% confidence interval ranging from 32.3% to 63.0%. Moreover, it was discovered that the mean magnitude of occurring misregistration is 8.6 ± 4.2 mm [Chin 08].

The problem etiology or reported underlying causes for such strong misregistrations are mainly due to the sequential nature of image acquisition in hybrid scanners [Chen 06], varying patient positioning between scans, and occurring respiratory, cardiac,

or patient motion [Tong 06]. As a consequence, the clinical impact of such misalignment is rather severe. It appears as the danger of a physician incorrectly reporting SPECT starting at misalignments of one SPECT voxel. This is due to the creation of artifactual perfusion defects that reduces diagnostic specificity of hybrid SPECT/CT imaging [Fric 04] and erroneous SPECT uptake quantification that questions the reliability of SPECT imaging in disease monitoring and the assessment of treatment response [Tong 05].

5.3.2 Clinical Assessment of Parametric Registration

In section 5.2 we introduced a validation framework that is composed of three main components: (i) validation hypothesis, (ii) validation data sets, and (iii) validation criteria. We are using this validation framework by Jannin *et al.* [Jann 06] in order to evaluate several suited algorithms for the clinical application of CT based SPECT AC (Figure 5.4).

Our validation hypothesis is formulated as: "In hybrid CT based SPECT AC, the best suited multimodal registration algorithm provides accurate results in the registration of CT and SPECT with a translational error of less than 4.79 mm (i.e. size of an isotropic voxel in our data) when compared to the physician aligned ground truth". We know from our literature review that the creation of artifacts is prevented during CT based SPECT AC, if a sub-voxel accurate alignment is ensured [Cahi 07, Goet 07b, Goet 07a]. The hypothesis is tested for three different parametric (i.e. rigid) registration algorithms: normalized mutual information (NMI) (i.e. $\alpha = 0$ in eq. (3.36)), the integrated statistic intensity prior model (JS) (i.e. replacing eq. (3.35) by (3.34) in (3.36)), and the constrained, or weighted, integrated SIP model (wJS) (i.e. eq. (3.36)). The validation data sets are comprised of 15 different cardiac SPECT/CT scans that were acquired by a Siemens Symbia T6 scanner. The field-of-view (FOV) for the SPECT data ($128 \times 128 \times 128$, $4.79 \times 4.79 \times 4.79\,\mathrm{mm}^3$) includes lungs, heart, and abdomen, whereas the FOV for CT data ($512 \times 512 \times 25$, $0.97 \times 0.97 \times 5\,\mathrm{mm}^3$) includes only heart and lungs. An expert physician manually aligned all data sets for a precise match of the heart region and we will use these alignments as the physician defined ground-truth (i.e. our BRONZE standard) to which all three algorithms are compared to. Furthermore, 15 different priors were generated from the expert aligned ground truth and employed during the validation. The validation criteria or metric will simply be the Euclidean distance to the established ground truth alignment.

Figure 5.4 summarizes the setup of the cardiac SPECT/CT validation framework visualizing how the validation studies (i.e. robustness study, prior sensitivity study, and capture range study) are performed for the different algorithms and later compared against the expert generated ground truth. A validation study is defined as follows: for all data sets, a significant number of registrations is performed per data set with different initial transformations away from ground truth alignment. For each registration result, the Euclidean distance is computed between the obtained and the ground truth alignment. Error mean and standard deviation of all registrations are then compared between the three registration methods under validation.

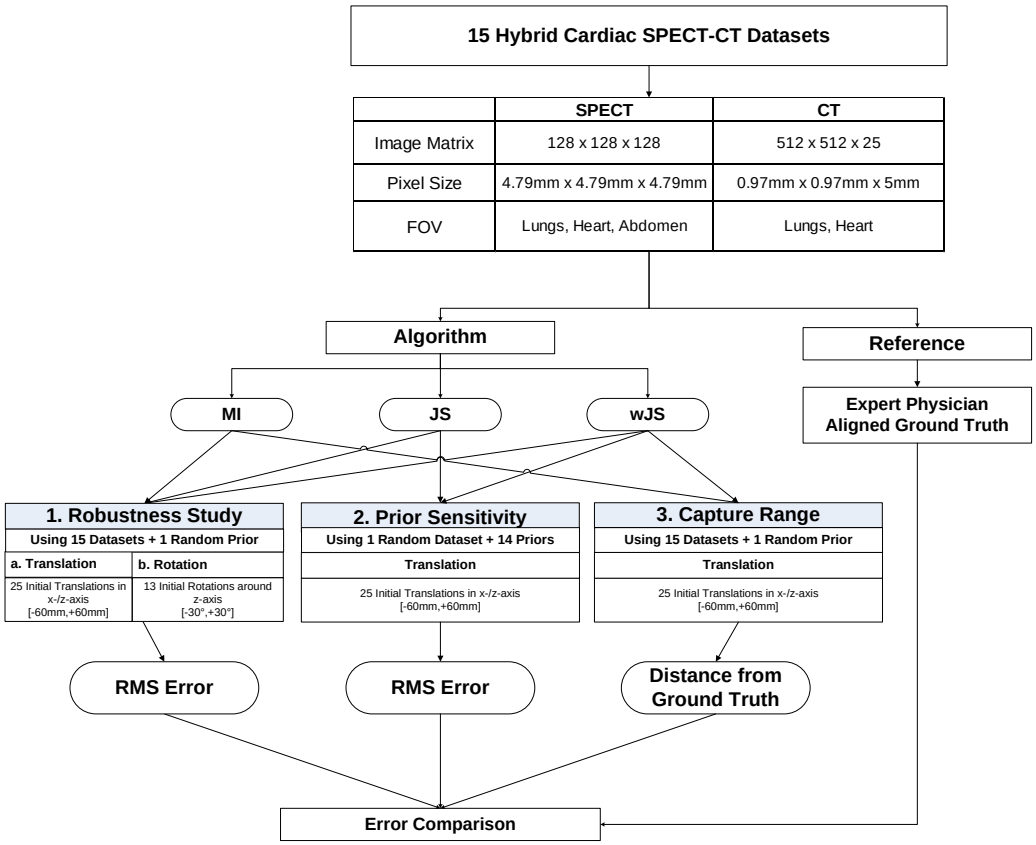
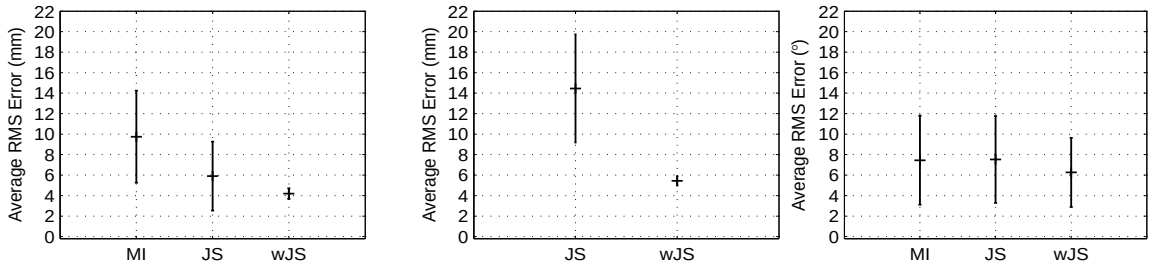


Figure 5.4: Our validation framework for parametric registration algorithms using ground truth on 15 clinical cardiac SPECT/CT data sets. The diagram visualizes multiple validation scenarios for the stated validation hypothesis.



(a) Robustness Test (b) Prior Sensitivity Test (c) Robustness Test

Figure 5.5: Comparison of error distributions for normalized MI, JS, and wJS over validation runs for translation (a) and (b), and rotation parameters (c). Integrating prior and data observations in combination with the newly weighted scheme, wJS, not only yields the best results but also generalizes well over multiple patients.

Validation Studies

We designed three different validation studies, as shown in Figure 5.4, for the following reasons. The robustness study was designed to find the most robust method across a large range of initial transformations from physician defined ground-truth using those 15 data sets. In this study, the prior based methods use a randomly chosen prior from among the generated priors. The second study, i.e. the prior sensitivity test, aims at determining the sensitivity of the integrated SIP models towards the chosen prior. Sensitivity is measured as distance to ground truth for all initializations and all priors. Needless to say that this study is only performed for the prior based algorithms. The third study validates the capture range for each method separately.

The initializations range from $-60\,\text{mm}$ to $+60\,\text{mm}$ away from our BRONZE standard in steps of 30 mm in x_1-/x_3- or in x_2-/x_3- direction resulting in 25 initializations for translation and from $-30°$ to $+30°$ in steps of $5°$ around the x_3-axes for 13 initial rotations. We evaluated translation and rotation initializations separately. The chosen α value in equation (3.36) is fixed to 0.2 for wJS and to 0.75 for JS for all experiments.

In eq. (3.36), the intensity transformation T_I is also estimated during optimization procedure. All 15 data sets showed minimal differences in the scaling parameter, i.e. it varied between 0.95 and 1.014, and no translational component was observed.

Validation Results

The results of the robustness and prior sensitivity tests are presented in Figs. 5.5a and 5.5b showing the mean registration error with respect to translation parameters. Figure 5.5c displays the mean angular registration errors for the robustness test. It can be observed that both JS and wJS are more accurate on average than NMI, and wJS additionally shows a small mean registration error and standard deviation. The constrained SIP model, wJS, outperforms NMI and the integrated SIP model (JS) with a mean translation error of $4.19 \pm 0.5\,\text{mm}$. Note that the error is smaller than a SPECT voxel (i.e. isotropic voxel in our data) showing sub-voxel accuracy for the constrained SIP model wJS. Normalized MI and JS show a mean error of more than

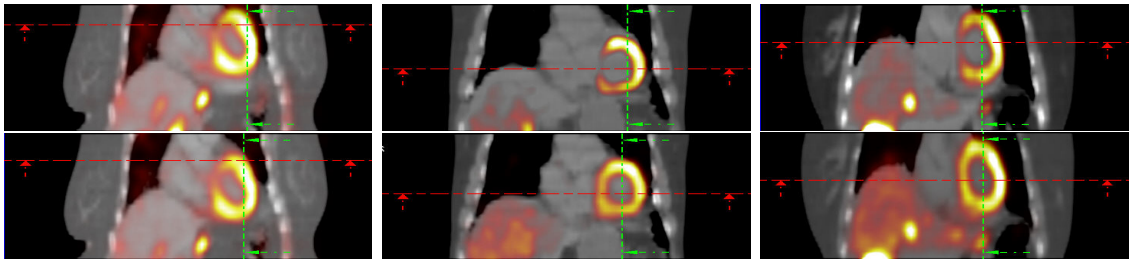


Figure 5.6: Registration results for 3 out of 15 patients. The top row shows the NMI result and the bottom row denotes the wJS results. The images illustrate the deviations from the optimum for NMI registration and high accuracy achieved by wJS approach.

2 [9.74 ± 4.49 mm] and more than 1 [5.9 ± 3.36 mm] voxel(s), respectively. We further noticed that optimization of NMI is attracted to local optima and that the global optimum for NMI deviates from the correct alignment if bright artifacts occur in CT data.[1] The SIP models do not deteriorate in those artifact data sets.

The angular errors further confirm the superiority of wJS over NMI, see Fig. 5.5c. The generally high observed mean angular error for wJS (6.2°), JS (7.5°) and for NMI (7.4°) is probably due to the little structural information apparent in SPECT. The angular error is an accumulation of errors from all three axes. In addition, Fig. 5.6 shows three visual results comparing NMI and wJS presenting the high accuracy achieved by wJS.

Our systematic validation studies show that the constrained statistical intensity prior (section 3.4) fulfills the clinical demands for registration accuracy of maximum 1 voxel mis-alignment in CT-based AC for cardiac SPECT, as mentioned in [Fric 04, Krit 05], and suggest the feasibility to use the approach for automated registration in hybrid scanners. Therefore, our validation hypothesis can be accepted for algorithm wJS only.

5.3.3 Clinical Assessment of Non-Parametric Registration

We have seen that parametric registration using statistical intensity priors yields accurate and robust results. However, in a number of clinical applications, it is not sufficient to use a parametric transformation model for registration due to its limited flexibility in dealing with complex motion patterns such as breathing or cardiac motion. Strongly misaligned local regions after successful parametric registration are still common in those applications. Figures 5.7a and 5.7d show an optimal rigid alignment of PET/CT data and significant misalignments resulting from breathing motion are still observed. As a consequence, there is a need for non-parametric algorithms that provide and guarantee required clinical accuracy. Furthermore, it can be shown that these local misalignments have a severe impact in numerous clinical applications. For example, it has been pointed out by the clinical literature review in section 5.3.1 that a slight (local) misregistration of more than 1 voxel can cause

[1] The observation was made during visual inspection of those data sets where the validation studies resulted in high errors.

Table 5.1: Phantom and clinical data sets used in the experiments discussed in sections 3.5.2 and 5.3.3.

Name	Modality	Size	Voxels [mm]	Scanner Type	Patients
P1	PET	$128^2 \times 172$	$5.15^2 \times 5.15$	Siemens ECAT PET	1
	CT	$512^2 \times 770$	$0.65^2 \times 0.5$	Siemens Somatom	
P2	PET (NAC)	$128^2 \times 303$	$5.20^2 \times 2.43$	Siemens Biograph	1
	CT	$512^2 \times 307$	$0.98^2 \times 2.4$	(hybrid)	
S1	SPECT	$128^2 \times 128$	$4.79^2 \times 4.79$	Siemens Symbia	24
	CT	$512^2 \times 28$	$0.98^2 \times 5.0$	(hybrid)	
S2	SPECT	$128^2 \times 128$	$4.79^2 \times 4.79$	Siemens Symbia	58
	CT	$512^2 \times 28$	$0.98^2 \times 5.0$	(hybrid)	

severe image artifacts in the application of CT based attenuation correction of hybrid scanner data.

For the clinical assessment of non-parametric registration the previous validation framework needs to be adapted to this different approach of solving the same clinical problem. More specifically validation hypothesis and criteria will be newly formulated. In the previous example, ground truth for clinical assessment of parametric registration was generated by manually translating and rotating the two volumes into correct alignment. That leads to some inevitable compromises in case the underlying transformation is of non-rigid nature. Therefore, the observer-dependent ground truth may vary and may not be expressive enough for the decision whether a non-parametric registration algorithm prevents the creation of perfusion artifacts. One of the many challenges in generating ground truth deformations is the lack of convenient manual tools that make it feasible and effortless to delineate the correct transformation between misaligned volumes. Besides such challenges and trade-offs, there is no feasible metric to directly compare the manual parametric ground truth alignment with the non-parametric result from the algorithms. The previously introduced application specific CAI measure solves this issue for the application of CT based AC in hybrid SPECT imaging. Section 5.2.2 describes in detail how the CAI, through the use of domain knowledge, indirectly compares the computed to the initial alignment and, if available, to the ground truth of the data set. This ensures a high accuracy validation.

The following section describes the validation strategy for non-rigid registration algorithms in an application specific manner. Having seen validation on phantom data and deformations in sections 3.5 and 4.2, we, now, validate the two different deformable registration algorithms, namely variational registration (i) using NMI and (ii) using statistical intensity prior models on a large number of clinical data sets.

The goals of the following studies are to reveal the performance of non-parametric registration using SIP models on clinical patient data and how it compares to standard methods, to visually present the need for non-parametric registration, and to verify the choice of the prior influence parameter.

Data Sets

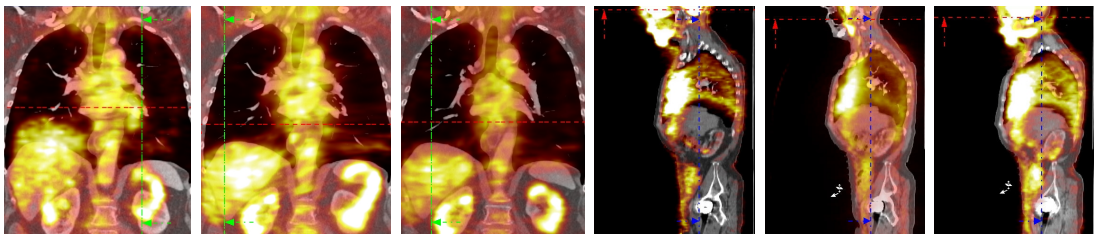
The validation data sets are manifold and constitute 86 patients in total. Table 5.1 provides an overview of each of the four collections of data sets that are used in detail.

The following can be stated about the contents of the data sets: P1 and P2 data sets are from PET/CT scanners with the former being CT based attenuation corrected PET and the latter non-attenuation corrected. The PET/CT images were acquired with a FOV extending from neck to the upper thigh region. Note the strong effect of AC on image quality by comparing Figs. 5.7a-5.7c to Figs. 5.7d-5.7f. Data sets S1 and S2 are both SPECT/CT data sets with a specific field-of-view (FOV) that includes mainly mediastinum (i.e. heart region) and lungs. Both sets are acquired using the same scanner type but at different hospital sites and with different imaging protocols. Data set S1 consists of 24 patients that were scanned using a *stress* protocol (i.e. the patient performed extensive physical activity shortly before the image scan), and it includes some of the abovementioned image artifacts such as windmill reconstruction artifacts due to metal implants in CT images, low or no uptake scar regions and extra-cardiac hot spots in SPECT images. The degree of misalignment in S1 was classified as mild to moderate on visual assessment by an expert physician. Data set S2 contains 58 patients also imaged using the *cardiac stress imaging* protocol. In addition, the CT was acquired using a breathing protocol called *breath-hold inspiration* and the SPECT data was acquired using ECG gating. Thus the amount of misalignment and the SPECT uptake values apparent in S2 are significantly larger than in S1.

We applied non-rigid registration using MI only and using the integrated SIP model to those data sets. Initially, data set P1 has been rigidly aligned, whereas data sets P2, S1, and S2 were used as acquired by the hybrid scanner. The total registration time per study varied between 40 and 250 secs averaging at about 93 secs on an Intel© Core 2 CPU with 2.33-GHz PC running Windows XP with 2GB RAM. The highly deviating computational times result from the choice of our convergence criteria for the gradient descent. It is either of the following three: a stable cost function value, a cost function value close to zero, or the maximum iteration number.

Qualitative Results

Data sets P1 and P2 illustrate the need for accurate alignment of PET/CT images in clinical practice, especially in presence of breathing motion during image acquisition (e.g. Fig. 5.7). The statistical prior for the corrected data (i.e. Figs. 5.7a-5.7c) was chosen from a similar PET/CT scan of a different patient that has been rigidly aligned by a physician. Since there is no prior data available for the uncorrected data set P2 (i.e. Figs. 5.7d-5.7f), we artificially created a prior using the following strategy. Assuming the original alignment is already close to the correct solution, the prior of this alignment may as well serve as a *self*-constraint for registration of extremely noisy images (e.g. Fig. 5.7d). The statistical intensity prior model allows to use a small weight on the prior essentially making it less of a constraint and assigning a higher importance to the observed data (i.e. here $\alpha_1 = 0.1$). For both data sets, the SIP model achieves a visually better alignment. Similar visual assessments can be made for patients selected from SPECT/CT data set S2 (Fig.5.8). Here, the statistical prior was generated from one randomly selected data set that a physician has classified as accurately aligned. The top row of Fig. 5.8 shows that the quality of cardiac SPECT/CT registration can be compromised when using mutual information alone. The reason for the failure of the data driven algorithm in both data sets may not be obvious at first. In Figures 5.7a-5.7c and the top of Figure 5.8, the mismatch can be



(a) Before (b) MI (c) MI + Prior (d) Before (e) MI (f) MI + Prior

Figure 5.7: Fused views of attenuation corrected (AC) and non-AC PET/CT data registered using mutual information (MI) and our proposed integrated SIP model. Adding prior knowledge increases registration quality where purely data-driven registration fails.

attributed to local optima of the global data-driven similarity measure. Especially at the lung-mediastinum border the matching of bright SPECT and dark CT intensities seems to pose a local optima for MI. In contrary, accuracy and robustness on the same data sets can be improved significantly by using a prior joint intensity distribution about the expected alignment as a support for MI (Fig. 5.8 bottom row). Note that the statistical intensity prior is also a global measure.

Quantitative Results

In general, validation of non-rigid registration algorithms on clinical data sets is very challenging due to either lack of ground-truth information or an appropriate gold standard. However, those algorithms can still be assessed by task-specific validation, e.g. using criteria that depend on an application, [Schn 01, Jann 02, Jann 06]. In this work, we quantitatively assess non-rigid registration algorithms for the application of cardiac perfusion imaging using hybrid SPECT/CT. We employ a reproducible, observer invariant, alignment metric that has been proposed specifically for quantitative validation of image registration algorithms in SPECT/CT and PET/CT cardiac data [Chin 08]. The proposed metric, referred to as the cardiac alignment index (CAI) in the following, leverages the fact that SPECT uptake should be maximal within the mediastinum and minimal inside the lungs when SPECT and CT volumes are precisely aligned.

Based on the CAI value of the original alignment it can be determined whether a method improves the alignment or not. We define registration success as an increased CAI value after registration as compared to before. Consequently, success is directly connected to reducing the number and degree of artifacts generated by methods that build on correct alignment (e.g. CT-based attenuation correction as described above). We perform our validation experiments separately on data sets S1 and S2 as they originate from different sites. The selected statistical priors for S1 and S2 respectively were generated from one physician aligned data set within each set and used throughout the experiments.

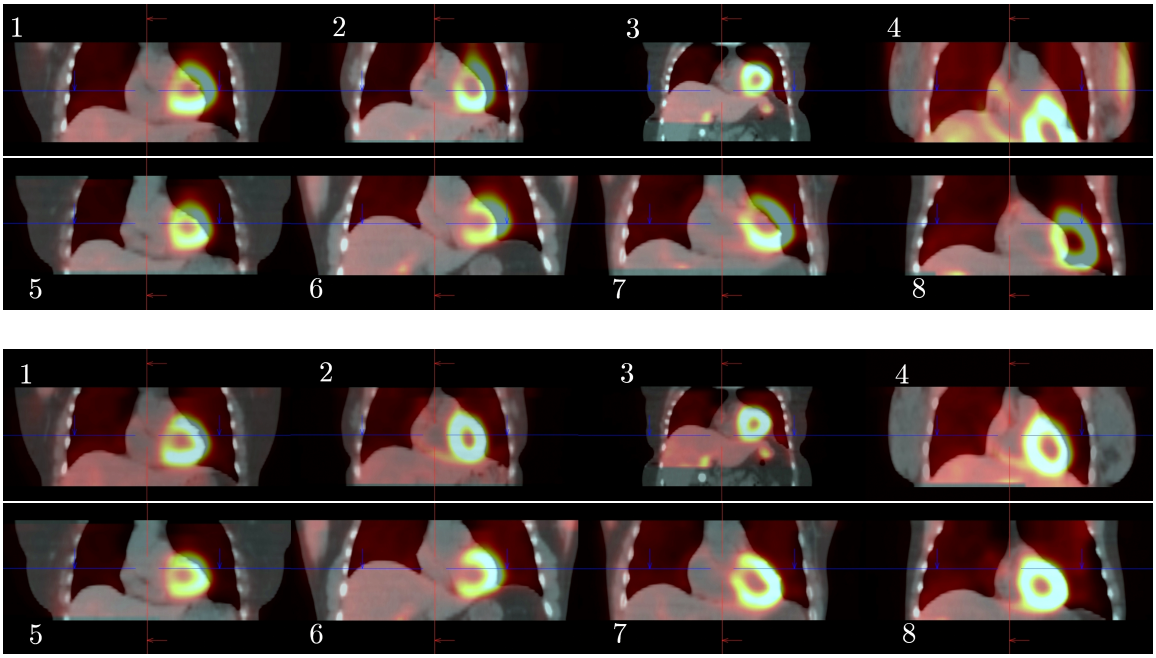


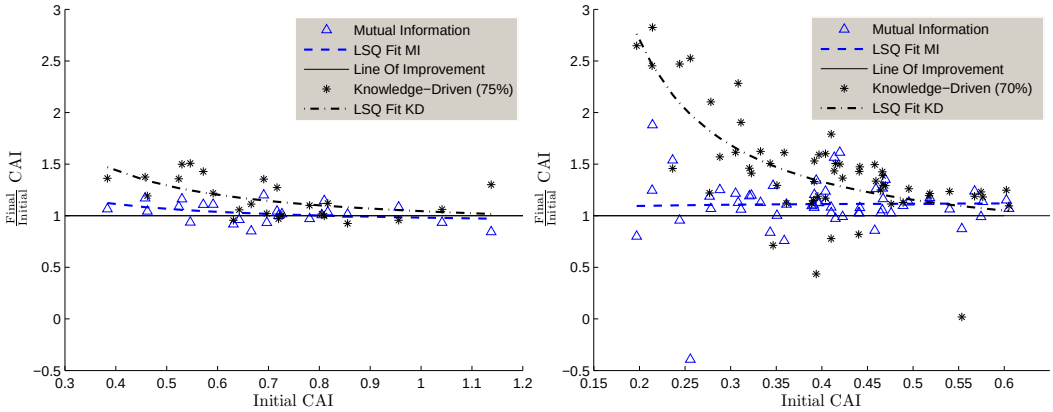
Figure 5.8: Top: Fused anterior views of eight cardiac SPECT/CT data sets after non-rigid registration using mutual information (MI). Most of the registrations failed due to convergence to a local optimum of MI. Bottom: The fused views of the same subjects as on top after non-rigid registration using a statistical intensity prior to support MI. The data sets are now successfully registered through the use of prior knowledge about correct alignment.

The results are summarized in Table 5.2 and grouped by successful registrations, failed registrations, and highest CAI value achieved in percent of the total number of patients within each set. Absolute numbers of data sets are given in parentheses. While employing the statistical intensity prior model the registration consistently improves in more and deteriorates in less data sets than without using any prior knowledge. It further achieves the most accurate results in the majority of patients (i.e. 83% and 76% in the rightmost column).

For interpretation of the CAI in clinical validation, the relative change of CAI after the registration was performed is more descriptive than the absolute achieved CAI value. From this relative change an improvement ratio (i.e. $ratio = \frac{\text{Final}}{\text{Initial}}$ CAI) can be computed to determine how strongly an algorithm improved the initial alignment. In

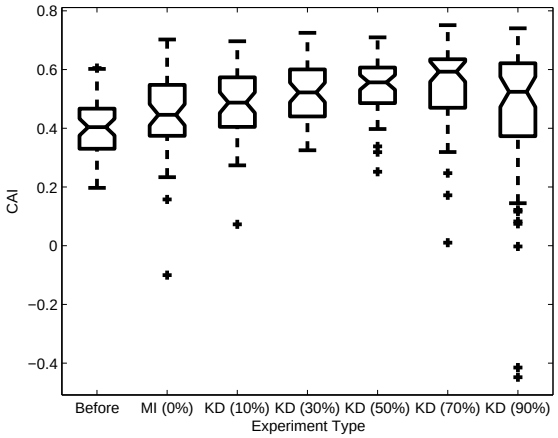
Data Sets	Success		Failures		Highest CAI	
	MI	SP	MI	SP	MI	SP
S1 (24)	67% (16)	80% (20)	23% (8)	16% (4)	12.5% (3)	83% (20)
S2 (58)	81% (47)	95% (55)	17% (10)	3% (2)	22% (13)	76% (44)
Total (82)	77% (63)	**92%** (75)	22% (18)	**7%**(6)	20% (16)	**78%** (64)

Table 5.2: Registration success and failures (as defined in the text) for data sets S1 and S2 evaluated using the CAI. Numbers in brackets refer to absolute number of data sets.



(a) Site I

(b) Site II



(c) Site II

Figure 5.9: Validation studies using Cardiac Alignment Index (CAI) for 82 cardiac SPECT/CT patients from two different hospital sites. The final over initial CAI ratio plotted against the initial CAI value, (a) and (b), represents the degree of improvement per registration method. The least-square fit shows that SIP models outperforms purely data-driven MI significantly for strongly misaligned data sets. Statistics over accuracies achieved by increasing prior influence, (c), illustrate the balance between closeness to prior data and observed data. The determined optimal value in 82 clinical data sets is observed at 30% (i.e. $\alpha_1 = 0.3$).

the case that no ground truth is available this measure is expressive for determining the more accurate technology. Figures 5.9a and 5.9b plot the improvement ratio over the initial CAI for both methods in S1 and S2. For a robust and accurate registration method, one could expect large improvement ratios for small initial CAI values and smaller ratios for initial values that are closer to 1.

A general trend of the results from the different algorithms can be visualized by a least-square-fit to an exponential curve. We discovered that the SIP model consistently outperforms data-driven MI w.r.t. robustness on data sets S1 and S2(Figs. 5.9a and 5.9b). Moreover, we performed an experimental validation of the α_1-parameter heuristically on data set S2. Figure 5.9c summarizes the registration of 58 patients comparing the achieved CAI values as box plot statistics for increasing prior influence. In this Figure, we observe that although the median CAI increases with higher α_1 values, the number of outliers among the data sets also rises. The optimal α_1 that achieves the most accurate results with the smallest variation for all data sets in S2 is found to be 0.3 (i.e. 30% prior influence) and is consistent with our previous heuristical findings. The second observation from Figure 5.9c is the SIP model's superiority over standard NMI approach on those 58 patient data sets. Although NMI raises the median CAI value, it also increases the CAI variation and is hence a less precise and less accurate algorithm. Those findings confirm our expectation that the α_1 parameter, supposedly, should not be the dominant factor for the statistical intensity prior model in (3.4).

5.4 Imaging Guidance for Ventricular Tachycardia Ablation Therapy

Physicians traditionally rely on manual comparison of images in clinical applications that support potentially life-saving interventions. This can lead up to a point where he or she has to form, in his mind, 3-d organ visualizations and relationships from multiple two-dimensional modalities using his carefully trained medical association of anatomy or disease. This is a highly complicated task that requires years of training and experience and is still prone to errors, because the image data used during the surgery is incomplete as visualized by the left image of Fig. 5.10. The medical imaging community currently investigates exciting topics that aim at taking this burden from the physicians and present an accurate high-end visualization and fusion of complimentary information to help in planning of, or navigation during these types of surgery. In Fig. 5.10, we provide an example of a left-ventricle segmentation from PET and a 4-chamber segmentation from CT scan. The fused visualization is made possible by establishing correspondence between PET and CT volume.

5.4.1 Description of VT Ablation and Current Clinical Solution

Ventricular tachycardia (VT) is an electrical abnormality in the cardiac conduction system that originates in one of the ventricles of the heart. This is a potentially life-threatening arrhythmia because it may lead to ventricular fibrillation (VF), asystole,

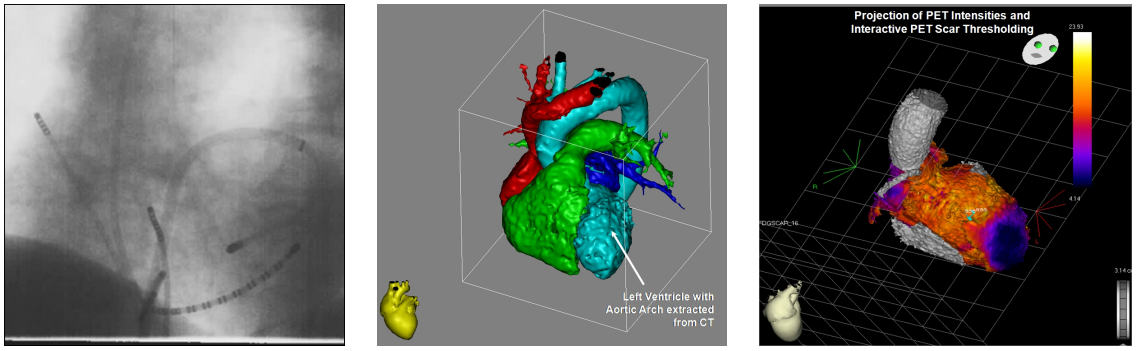


Figure 5.10: Typical images available during interventional VT ablation therapy. Original fluoroscopic x-ray image (left) that was previously used to perform ablation therapy on the heart. Images currently employed during ablation showing (middle) 3-d segmentation of the heart from CT and (right) PET intensity information projected onto the segmentation from CT.

and sudden death. Therapy may be directed toward two directions. Either terminating an episode of the arrhythmia or suppressing a future episode from occurring. The treatment for stable VT needs to be tailored to the specific patient, with regard to how well the individual tolerates such episodes of ventricular tachycardia, how frequently episodes occur, and so on.

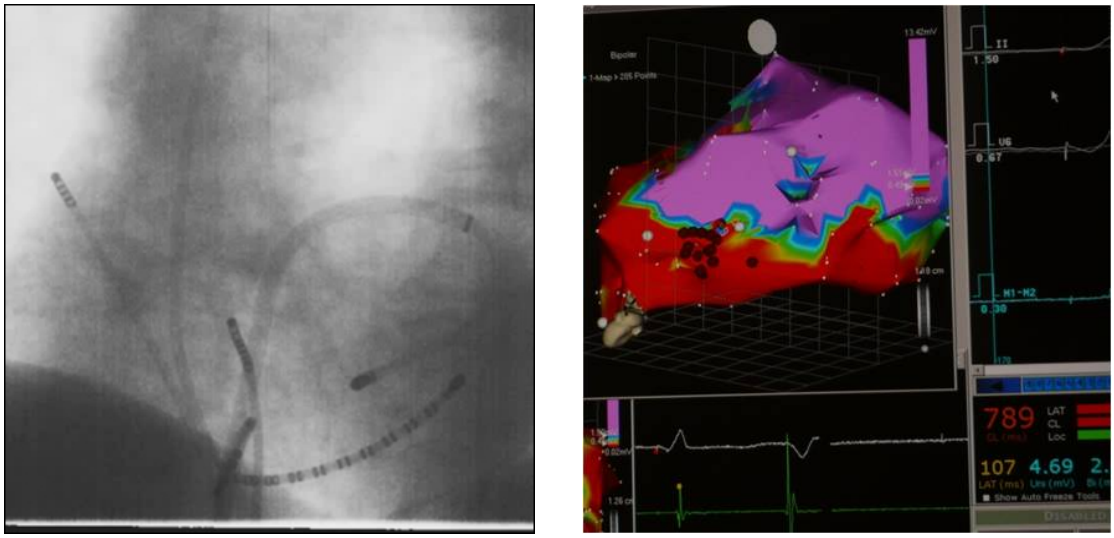
Therefore, treatment for frequently occurring arrhythmias are cardioverter defibrillators (ICD) that are implanted into one of the ventricles of the patient. An ICD is the most effective prevention against sudden cardiac death. Due to the prohibitive cost of ICDs, they are not routinely placed in all individuals.

Indications for ICD placement in the setting of ARVD include:

- Cardiac arrest due to VT or VF
- Symptomatic VT that is not inducible during programmed stimulation
- Failed programmed stimulation-guided drug therapy
- Severe RV involvement with poor tolerance of VT
- Sudden death of immediate family member
- Since ICDs are typically placed via a transvenous approach into the right ventricle, there are complications associated with ICD placement and follow-up.

After a successful implantation, the progressive nature of the disease may lead to fibro-fatty replacement of the myocardium at the site of lead placement. This may lead to under-sensing of the individual's electrical activity (potentially causing inability to sense VT or VF), and inability to pace the ventricle. As a consequence, an increasing number of patients presenting presented with frequent and appropriate implantable cardioverter-defibrillator firings require VT ablation despite optimal medical therapy.

It has been found that anatomically based "substrate modification" strategies have to be employed in more than 85% of the patients due to hemodynamic intolerance



(a) Fluoroscopic X-Ray (b) Voltage Map

Figure 5.11: Fluoroscopic x-ray image of catheters ablating the heart and an endocardial voltage map visualizing voltages across the ventricle using a color-map.

of the VT [Marc 00]. The basic idea of those approaches is to perform a linear ablation along lines across and along the myocardial scar and its border zone in order to interrupt conducting channels of surviving myocardium. For these procedures endocardial voltage mapping is frequently used to define the scar substrate.

Endocardial voltage mapping aims at generating an accurate voltage map of the endocardium by dense sampling of the ventricles using a steerable catheter with a distal tip electrode and ring electrode attached to measure the voltage currents. The 4mm tip and 2mm ring electrode are sender and receiver electrodes respectively at a inter-electrode distance of 4mm. Access to the ventricle is achieved by an invasive trans-aortic approach [Fuku 08]. Due to the size of the electrodes and the ventricles, this procedure usually requires several hours of invasive intervention that bears several risks for the patient. The physician needs to carefully operate the catheter such that no surrounding tissue is irreversibly injured by the measurement. Nevertheless, the resulting endocardial voltage map accurately defines healthy, scar (dead), and borderline scar tissues by applying standard clinical voltage criteria (i.e. scar: < 0.5 mV, abnormal: $0.5 - 1.5$ mV, and normal: > 1.5 mV myocardium) and it is considered the current clinical gold standard in VT ablation therapy [Tian 08]. The mapping information is indispensable during VT ablation of surviving myocardium that causes the abovementioned complications. The imaging feature important to the physician is the scar borderzone tissue.

Current clinical systems display the endocardial voltage map alongside fluoroscopic X-ray images that visualize the catheter very well. Neither the map nor the X-ray images allow the display of 3-d left ventricle (LV) anatomy and embedded 3-d scar reconstructions. Thus, the physician has to use the clinical knowledge of anatomy in his or her head to create a visual understanding of physical anatomical and scar locations on the voltage map. Also, the available information about border zone and

high-resolution anatomic scar features are not displayed on current systems and hence not utilized for the surgical procedure. Figure 5.11 shows a typical example of image information available to physicians in current clinical systems. The information is available as 2-D views with no anatomical guidance, it is not combined with voltage map, and scar information is not reconstructed. For successful ablation, the physician needs to combine the voltage information (Fig. 5.11b) with anatomical knowledge of the heart in his/her head to steer the ablation catheter (Fig. 5.11a) to the scar border zone.

Although the endocardial voltage map is the current gold standard for defining scar substrate, it has limited ability to detect intra-myocardial or epicardial scar, and in addition, suboptimal catheter contact can result in falsely low voltage measurements [Tian 08]. All these rather significant downsides of the current mapping systems lead to a number of technological advancements and active research in VT ablation therapy to better combine all available information.

5.4.2 VT Ablation Therapy: Technological Advancements

The technological advancements for VT ablation therapy have been in two major directions. The first one are advancements in the ability to track and navigate the ablation catheter using new magnetic technologies that allow remote tracking and navigation. The other direction is exploring the potentials of 3-d image guidance for VT ablation which will be the focus here. Image guidance in VT ablation is current being explored using magnet resonance (MR) guidance (i.e. real-time MR or delayed enhancement MR guided ablation), ultrasound, and PET/CT guidance. A rather recent statement from within the VT community shows that clinical interest in image guidance is strong.

> Advancements in imaging technology help the electro-physiologists to ablate not only anatomical lesions but also tissue and molecular targets. *(International Symposium on Ventricular Arrhythmia, Philadelphia, 2007)*

Clinical research in this area has further shown that current imaging technologies are capable to deliver new methodologies that improve on the downsides of endocardial voltage mapping. For example, Dickfeld and colleagues [Dick 08] reported that current image integration software can be used to register and display PET/CT derived myocardial scar as an area of absent voxels ("hole in the wall") within the reconstructed left ventricular wall to indicate scar location and size.

The goal of Tian and colleagues [Tian 08] is to move a newly developed, pre-clinical image integration module, that allows a widespread, clinical use of image-reconstructed three-dimensional scar maps from PET/CT to guide VT ablations, into clinical practice by comparing it to the current clinical gold standard. Their study was performed on 10 consecutive ICD patients with an ischemic cardiomyopathy scheduled to undergo VT ablation.

The integration platform allows the simultaneous visualization of multiple imaging data sets and of the 3-d scar embedded into 3-d left ventricle myocardium. This enables the physician to directly identify the substrate tissue with the LV anatomy and secondly, the use of PET 3-d reconstructions at multiple metabolic thresholds

allows the characterization of the scar border zone targeted during VT ablation. Last, but not least important, the registration of the high-resolution CT images enables the simultaneous display of the detailed substrate anatomy supplementing the intra-procedural guidance during VT ablations. The image guidance platform is composed of the following steps:

1. Pre-interventional image acquisition using either a hybrid or two standalone PET and CT scanners.

2. Multimodal image registration to ensure proper alignment of cardiac PET/CT data that will later provide more anatomical details for the scar reconstruction and visualization.

3. Segmentation of the left myocardium in the PET image data and estimation of scar border zone by multiple thresholding of metabolic activity within the segmented LV.

4. Four chamber segmentation of the contrast enhanced CT images including the left ventricle and the aortic arch.

5. The segmentation meshes from CT and PET can now be investigated through a fused visualization of four chambers, LV PET, and LV PET scar and border zone.

6. Clinical integration software provides further visualizations of segmentations or PET intensity projections on the CT segmentation to assists in therapy planning and in anatomical locating during interventions.

Figure 5.12 visualizes preliminary results of the image guided integration platform by showing different visualizations of the available segmentations. The fused visualization was made possible by successful application of the constrained statistical intensity prior model to PET/CT image data (i.e. step 2. above). We will realize in later sections that the successful application of the constrained model to noisy SPECT/CT data [Guet 07] translates well to PET/CT scans of patients scheduled to undergo VT ablation. Subsequent sections report in more detail the challenges that registration technologies are faced with in such data and presents our systematic validation procedures. Similarly, the robust segmentation of the left myocardium from noisy 4-d SPECT sequences using a prior shape model as introduced by Kohlberger and colleagues [Kohl 06] is well designed to robustly segment the left myocardium from PET. This approach uses a dynamic shape model of the myocardium and fits it to the available image data. It robustly segments the myocardium and the scar plus the scar border zone in step 3. of the image guidance platform.

Comparing Figures 5.11 and 5.12 the technological advancements and advantages are clearly visible. In conclusion, the fusion of multimodality imaging sets derived from PET/CT allow an accurate, simultaneous display of LV anatomy and myocardial scar and can assist in the further metabolic and anatomic characterization of the VT substrate. This suggests a possible role in facilitating a novel image-guided approach to substrate-based VT ablations [Tian 08].

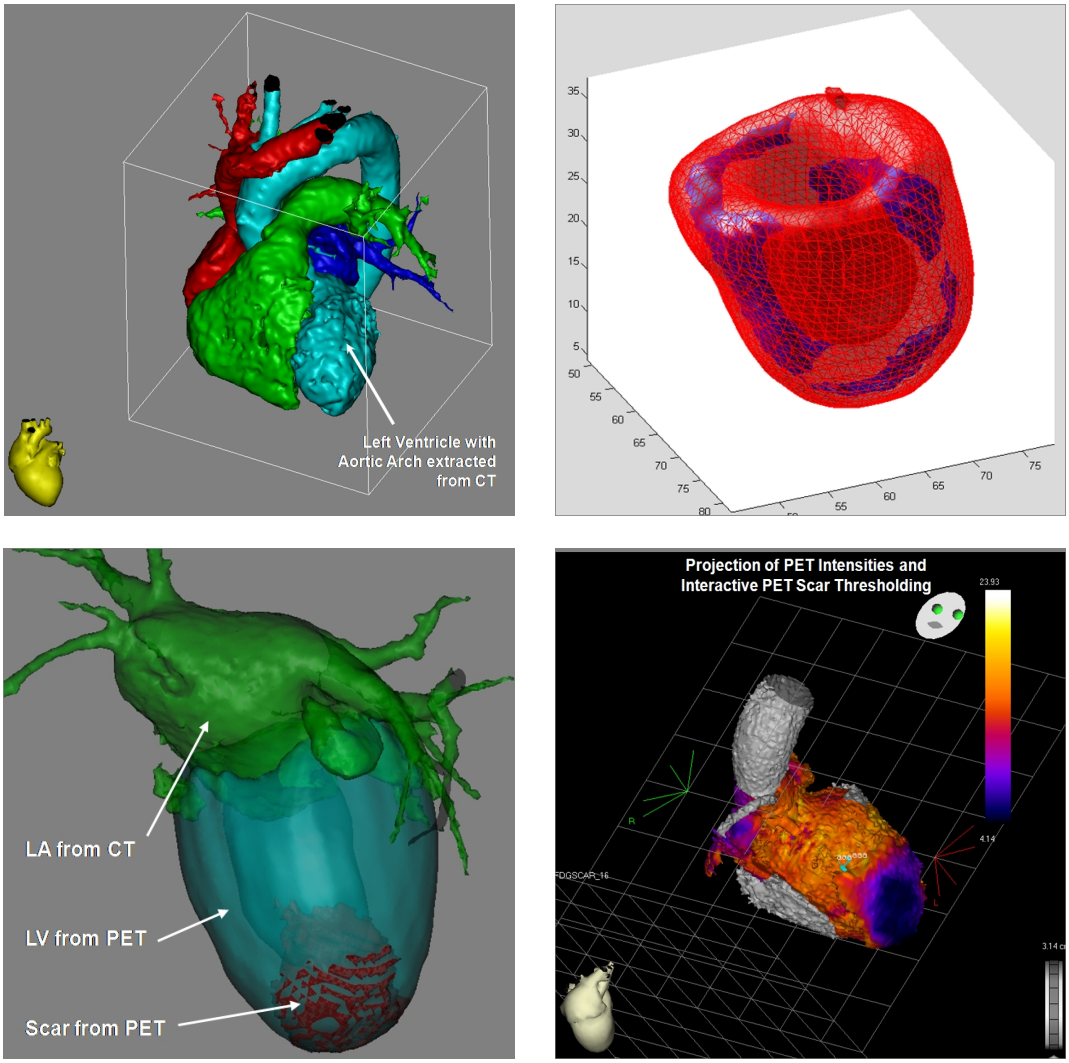


Figure 5.12: Fused visualization (bottom left) of segmented chambers (top left) and LV (top right) from CT and PET respectively assist the physician in locating the scar within the anatomy. Projections of PET intensities onto CT segmentation (bottom right) directly correlate anatomy and function for scar border zone visualization in VT ablation therapy. (Images are courtesy of Tian *et al.* [Tian 08], Kohlberger *et al.* [Kohl 06], and Biosense Webster, Israel.)

5.4.3 Challenges for and Clinical Validation of Registration Technologies

A key element in the previous description of image-guided VT ablation is the registration of multimodal image data sets. The accurate registration of PET and CT data allow the simultaneous display of 3-d scar embedded into 3-d LV myocardium and of the detailed substrate anatomy derived from high-resolution CT images. Physicians rely on the accuracy of potential registration technologies and base their decisions on the results derived from the registration. The biggest challenge in designing a robust registration technology is the quality of the available image data. As previously described, patients that undergo VT ablation are usually equipped with an implantable defibrillator or ICD. Such devices create streaky, windmill-like, reconstruction artifacts in the CT image generating images with a larger-than-expected noise level. Such artifacts are categorized as patient-based. Even though parametric registration approaches seem clinically acceptable, the quality of image data is unacceptable for standard multimodal registration approaches such as NMI.

In this chapter, we will investigate the clinical feasibility of using statistical intensity priors for the multimodal image registration in image-guided VT ablation. For that purpose, we leverage the validation framework for the clinical assessment of registration technology introduced in section 5.2. The following components of the validation framework have been identified for this application:

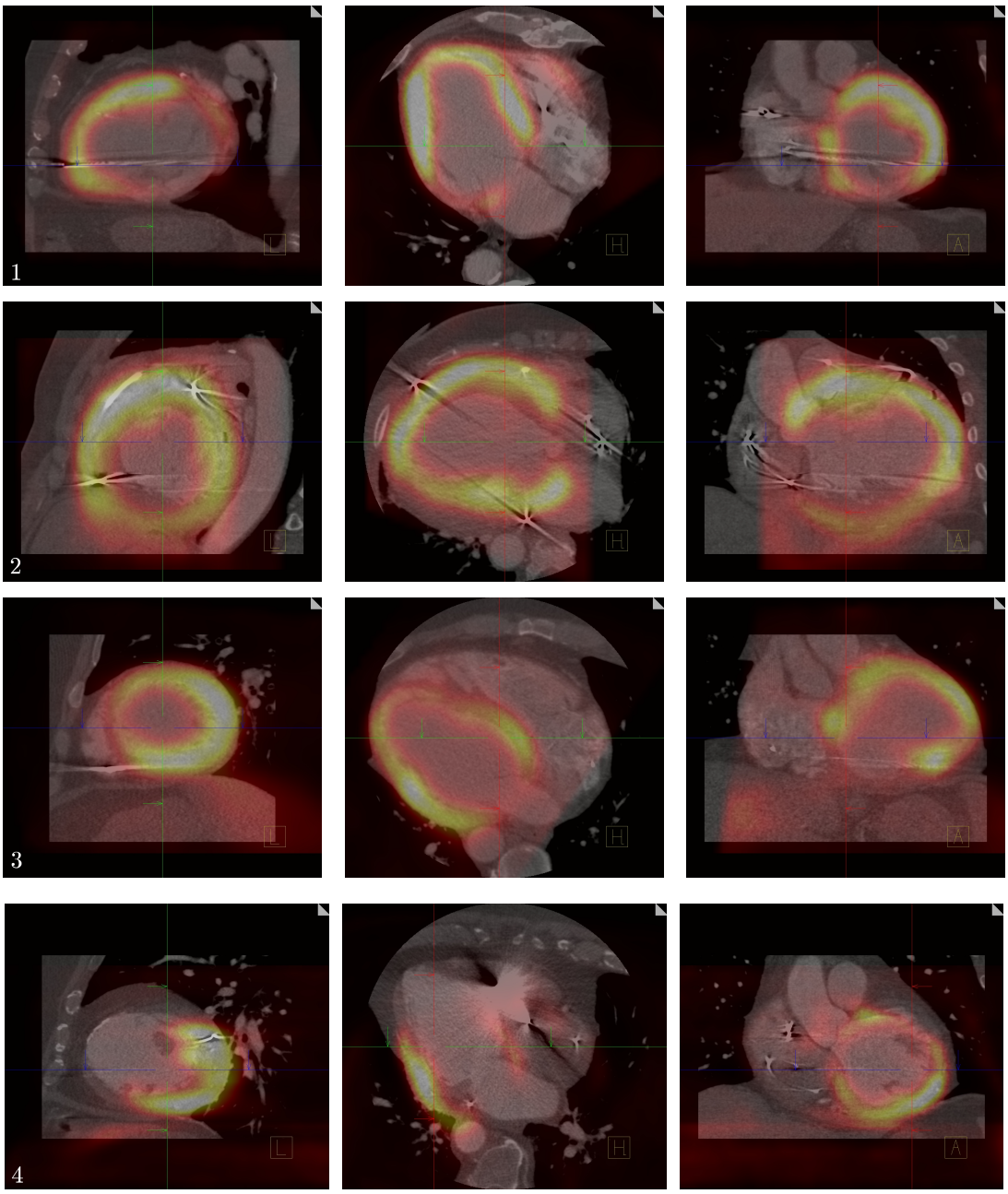
Validation Hypothesis In image-guided VT ablation therapy, the best suited multimodal registration algorithm provides accurate results in the registration of CT and PET with a translational error of less than the size of the PET voxel (i.e. sub-voxel accurate) when compared to the clinical gold standard.

Validation Data Sets PET/CT data sets of four patients scheduled for VT ablation and part of the clinical study in [Tian 08]. The image data is described in detail below.

Validation Criteria Visual assessment of the converged registration results of the different algorithms. Clinical ground truth was unfortunately not available.

The currently available clinical image data is given by four pathological PET/CT data sets of size $512 \times 512 \times 344$ (CT) and $144 \times 144 \times 45$ (PET). The slice number varies between the data sets and the size of a voxel is $0.39\text{mm} \times 0.39\text{mm} \times 0.45\text{mm}$ for CT and $4.0\text{mm} \times 4.0\text{mm} \times 4.0\text{mm}$ for PET. It can be observed in Figure 5.13 how challenging the data sets in VT ablation therapy are for multimodal registration technologies. Several bright image reconstruction artifacts due to implanted ICDs, missing image information in PET due to lack of metabolic activity in the left ventricle, and a strong difference in resolution between PET and CT are common in VT patient data sets.

Few image commonalities between diseased PET and CT, large amount of image artifacts, and low image quality as observed are the main contributors to the failure of standard data-driven registration techniques such as NMI. On the other hand, the integrated SIP model, as introduced in chapter 3, promises to cope with such challenges. Fig. 5.13 plots the registration results as achieved by the SIP model, and



(a) Left View (b) Head View (c) Anterior View

Figure 5.13: Three orthogonal views of registered and fused PET/CT data sets from four patients subject to VT ablation therapy. Registration is achieved by the integrated SIP model using a prior from patient three. Note the inherently challenging nature of the data, the reason why registration using NMI fails for all four data sets.

we note that the standard NMI approach did not yield any successful registration on these data. As a matter of fact, NMI registration failed so strongly for data sets two, three, and four that the influence of NMI during optimization (e.g. see eq. (3.36)) was removed, i.e. the α-parameter was chosen to be 1.0, in order to achieve a reasonable registration. Due to the fact that optimization was performed with respect to only the prior observations, the registration results for data sets two and four are reasonable, however, showing small room for improvement. Note that in this experiment the prior was generated from one of the data sets, i.e. patient three. For data set one, we could choose the optimal α-parameter, i.e. $\alpha = 0.3$, as determined by our previous experiments (e.g. see Fig. 5.9), and the visual result looks highly accurate. Following the results of our previous experiments, the combination of both previous and current observations promises to be a robust and accurate model for multimodal registration.

Data set one is the data set that was used throughout this dissertation (e.g. see Figure 3.1), and we now realized how highly accurate registration is achieved despite the strong image reconstruction artifacts. According to visual assessment, our hypothesis of these small use case validation runs is accepted for the integrated SIP model. This model delivers a robust and accurate solution where common approaches fail to properly align the data sets.

5.5 Summary

In this chapter, we have introduced a systematic validation framework for the clinical assessment of multimodal registration technologies. Moreover, we have presented several validation criteria for parametric and non-parametric image registration that measure the success or failure of such methods in multimodal data sets. Using these criteria, we employed the validation framework to assess the performance of statistical intensity prior models in two major clinical applications, i.e. CT based attenuation correction in hybrid scanner systems and image-guidance in VT ablation therapy. After performing several clinical validation studies on more than 82 patient data sets total, it can be seen that a consistent selection of α_1 (i.e. prior influence parameter) can be made without degrading registration quality. And since the generation of prior knowledge is effortless, a wide range of applications can be addressed by the SIP model. Note, it is one of the first glances at the performance of registration algorithms using prior knowledge on a larger set of clinical image data. The statistical intensity prior models nicely outperformed traditional methods in a consistent way because they leverage on more information as purely data-driven approaches. The flexibility and generality of the introduced models proof advantageous on noisy, degenerate patient data, an extreme scenario for clinical image data.

Chapter 6

Conclusion and Future Work

This thesis addressed the problem of establishing correspondence between multimodal medical images. We introduced two intensity prior models and validated their performance on numerous synthetic and clinical examples. Based on their performance, we showed the added clinical value by these methods on two specific, routine clinical applications. This chapter summarizes the achieved results and presents an outlook into future directions.

6.1 Statistical Intensity Prior Model

In chapter 3, we first introduce a non-parametric statistical intensity prior (SIP) model that allows to simultaneously consider prior knowledge and current observations. Though introduced through the use of variational calculus, this intensity prior is independent of the transformation model. The statistical intensity prior can be used as additional force for a registration algorithm to robustly align medical images. The design goal for this model was to introduce a simple, elegant, but efficient prior model that is best suited as robust support for existing algorithms and that can be integrated effortlessly into existing systems.

6.1.1 Main Results

1. Traditional intensity-based registration methods deteriorate on noisy, sparse, or artifacted multimodal medical image data, while SIP models, which also utilize prior knowledge about the correct intensity mapping, do not.

2. Unlike traditional registration measures, the SIP model shows a large capture range on a variety of normal and diseased multimodal medical data sets. For some modalities, a simple intensity correction model can be used as part of the SIP model.

3. Artificial constraints on possible intensity mapping resulting from a restricted transformation model can be bypassed by evaluating only on intensity mappings of interest in the SIP model. These mappings can be specific to a particular organ.

6.1.2 Future Work

Future investigations into the general use of SIP models are guaranteed. In particular, it would help to fully understand the SIP parameter α_1, perhaps finding a way to learn it from the training data such that it becomes optimal for a particular modality combination. Finally, an investigation into other data terms might be useful in further understanding the full practical potential of SIP models.

6.2 Generalized Statistical Intensity Prior Model

In chapter 4, we developed a generalization of the SIP formulation called general statistical intensity prior (GSIP) that allows to incorporate any amount of prior knowledge. The new formulation of a multimodal prior can produce a more accurate, unbiased prior model that allows for an automatic adjustment of confidence in the prior. We further note the applicability of GSIP to medical and non-medical data in our experimental results.

6.2.1 Main Results

1. The formulation of a non-parametric GSIP model allows for an accurate and flexible description of the underlying manifold of joint intensity probability distributions.

2. The GSIP model formulation, further, eliminates the need for a deterministic specification of intensity prior influence through an automatic estimation of the current confidence into the prior model. This parameter is well derived from available prior data.

3. Simulations show that the formulation of a multimodal prior allows for utilizing prior knowledge from multiple intensity-based sources, and maintains other desirable properties such as an extended capture range.

4. On a series of experiments, we compared the performance of traditional MI-based, unimodal prior-based, and multimodal prior-based registration algorithms. The GSIP remains the only algorithm that delivers accurate registration on occluded image data.

5. Simulations results show that the multimodal prior is more flexible than a unimodal prior demonstrating an increased robustness and an elimination of a bias towards the prior model.

6.2.2 Future Work

In future research, we would like to study the effect of large numbers of prior samples on the performance of GSIP and when it will become redundant to have an additional data term included in the formulation. Also, further investigation in different deformation priors and possible finding of its mathematical relationship to intensity priors may shed new light on the research of statistical prior model formulations. The current

formulation is based on kernel density estimation. It would be interesting to investigate the possibility of extending or connecting its formulation to traditional unsupervised approaches in machine learning. Finally, drawing connections between GSIP with other applications in image processing and computer vision might provide some new insight or even new solutions to existing problems.

6.3 Clinical Validation and Applications

In chapter 5, we described a systematic validation framework for the clinical assessment of multimodal image registration algorithms. This framework allows to evaluate an algorithm with respect to other algorithms as well as clinical ground truth, and it provides an understanding of the added clinical value by a particular multimodal registration algorithm. We, further, successfully applied SIP and GSIP models to routine clinical application of *CT-based Attenuation Correction in Hybrid Scanners* and to a pre-clinical application of *Image Guidance in Ventricular Tachycardia Ablation Therapy* adding tangible clinical value by providing robust and accurate correspondences in these multimodal medical images.

6.3.1 Main Results

1. The systematic literature review and utilization of a systematic validation framework not only provides thorough validation of SIP and GSIP, but also defines an added clinical value by those approaches.

2. The introduction of an application-specific validation metric, called *Cardiac Alignment Index*, allows for the validation of registration algorithms independent of the transformation model. We also demonstrate the need for non-rigid or deformable registration algorithms in such clinical applications.

3. Both SIP and GSIP systematically outperform traditional statistical registration methods on a wide collection of one hundred clinical data sets by decreasing registration failure rates to 7% and increasing registration accuracy ratings from 20% (MI) to 78% (SIP).

4. On routine clinical data, statistical intensity prior models remain within-voxel accurate and, therefore, eliminate image reconstruction artifacts in SPECT data induced by misaligned hybrid scanner data. These results suggest the feasibility of both models for automated multimodal image registration.

5. Besides an implementation on the GPU of SIP and GSIP that achieves interactive performance of non-rigid registration of standard clinical data sets, the implementation of both models on a standard Windows system achieves clinically acceptable computational performance between 40 and 250 seconds averaging around 93 seconds.

6.3.2 Future Work

In future work, we would like to investigate the performance of SIP and GSIP on other modality combinations including but not limited to MR/PET, MR/US, or CT/US. It would further be interesting to explore the benefit of GSIP in general, clinical applications such as MR perfusion studies, neurological studies using fMRI data, Oncology lesion follow-up studies, and so on. The full clinical potential has, by far, not been explored yet. We would like to further increase the accuracy ratings by utilizing GSIP instead of SIP in more applications.

6.4 Outlook

The main goal of the work presented in this thesis has been to contribute to robustly establishing correspondence in data sets from multiple modalities through the development of efficient statistical intensity prior models. Furthermore, it has been made our first priority to investigate and proof the feasibility of those methods in clinical registration problems involving challenging, clinical data from CT, MRI, PET, and SPECT images. The research presented herein can hopefully deepen the understanding of statistical intensity prior models, the relationship with currently observed data, and points out new directions for designing better intensity prior models and extending their application into other areas of medical image analysis and computer vision.

The majority of current clinical applications leverage the pair-wise registration of multimodal data sets. As a consequence of increasing amounts of medical data and the abilities to fuse any type of image modality, future research directions will be directed towards robustly registering of multiple multimodality data sets. This could be in the context of multimodal population registration approaches or the sequential pair-wise multimodal registration depending on the goals of the application. The physician, however, is eager to leverage all available clinical information such that he or she can make the most informed decision on the final diagnosis. The methods and their validation presented in this thesis presumably offer a starting point for these future registration needs.

Furthermore, automation of registration procedures is of high demand due to the tremendous amount of clinical data that makes a manual interaction a very inefficient tool to revert to for doctors and technicians.

Bibliography

[Altm 86] S. L. Altmann. *Rotations, Quaternions, and Double Groups*. Clarendon Press, Oxford University Press, 1986. 18

[Aube 02] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Vol. 147 of *Applied Mathematical Sciences*, Springer-Verlag, 2002. 16

[Bank 09] I. N. Bankman, Ed. *Handbook of Medical Image Processing and Analysis*. Academic Press, Elsevier, San Diego, CA, USA, 2nd Ed., 2009. 2, 24

[Bari 93] C. Barillot, D. Lemoine, L. L. Briquer, F. Lachmann, and B. Gibaud. "Data fusion in medical imaging: merging multimodal and multipatient images, identification of structures and 3D display aspects.". *European Journal of Radiology*, Vol. 17, No. 1, pp. 22–27, Jun 1993. 1

[Bosc 03] M. Bosc, F. Heitz, J. P. Armspach, I. Namer, D. Gounot, and L. Rumbach. "Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution.". *Neuroimage*, Vol. 20, No. 2, pp. 643–656, Oct 2003. 1

[Brow 92] L. G. Brown. "A survey of image registration techniques". *ACM Comput. Surv.*, Vol. 24, No. 4, pp. 325–376, 1992. 21

[Brox 04] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. "High Accuracy Optical Flow Estimation Based on a Theory for Warping". In: *Proceedings of European Conference on Computer Vision*, pp. 25–36, Springer, 2004. 16, 39

[Cahi 07] J. Cahill, J. Kritzman, E. Ficaro, and J. Corbett. "False Positive SPECT/CT Myocardial Perfusion Studies: Contribution of Respiratory Motion, Misregistration, and Other CT-related Artifacts When Using CT-based Attenuation Correction". *Journal of Nuclear Cardiology*, Vol. 14, No. 4, p. S106, July July 2007. 74

[Cama 03] O. Camara, O. Colliot, G. Delso, and I. Bloch. "3D nonlinear PET-CT image registration algorithm with constrained Free-Form Deformations". In: *Proceedings of Conference on Visualization, Imaging, and Image Processing*, September 2003. 24

[Cerq 02] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, M. S. Verani, A. H. A. W. G. on Myocardial Segmentation, and R. for Cardiac Imaging. "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association.". *Journal of Cardiovascular Imaging*, Vol. 18, No. 1, pp. 539–542, Feb 2002. 70

[Chan 05] T. Chan and W. Zhu. "Level Set Based Shape Prior Segmentation". In: *Proceedings of Computer Vision and Pattern Recognition Conference*, pp. 1164–1170, 2005. 22

[Chef 01] C. Chefd'Hotel, G. Hermosillo, and O. Faugeras. "A Variational Approach to Multi-Modal Image Matching". In: *Proceedings of Workshop on Variational and Level Set Methods*, IEEE Computer Society, 2001. 24

[Chen 06] J. Chen, S. Caputlu-Wilson, H. Shi, J. Galt, T. Faber, and E. Garcia. "Automated quality control of emission-transmission misalignment for attenuation correction in myocardial perfusion imaging with SPECT-CT systems". *Journal of Nuclear Cardiology*, Vol. 13, No. 1, pp. 43–49, February 2006. 70, 72, 73

[Cher 03] S. R. Cherry, J. A. Sorenson, and M. E. Phelps. *Physics in Nuclear Medicine*. Saunders, 3rd Ed., 2003. 4, 6, 7, 72

[Chin 08] P. Chinnadurai. *Knowledge-Driven Image Registration for Accurate Attenuation Correction in Hybrid SPECT/CT Scanners - A Clinical Validation Study*. Master's thesis, Indian Institute of Technology, Kharagpur, India, May 2008. 72, 73, 80

[Chri 01] G. E. Christensen and J. He. "Consistent Nonlinear Elastic Image Registration". In: *MMBIA '01: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'01)*, p. 37, IEEE Computer Society, Washington, DC, USA, 2001. 16, 22, 40

[Chri 06] G. E. Christensen, X. Geng, J. G. Kuhl, J. Bruss, T. J. Grabowski, I. A. Pirwani, M. W. Vannier, J. S. Allen, and H. Damasio. "Introduction to the Non-rigid Image Registration Evaluation Project (NIREP)". In: *WBIR*, pp. 128–135, 2006. 68

[Chrs 04] R. Chrástek, M. Skokan, L. Kubecka, M. Wolf, K. Donath, J. Jan, G. Michelson, and H. Niemann. "Multimodal retinal image registration for optic disk segmentation.". *Methods of Information in Medicine*, Vol. 43, No. 4, pp. 336–342, 2004. 1

[Chun 02] A. C. S. Chung, W. M. Wells, A. Norbash, and E. Grimson. "Multimodal Image Registration by Minimising Kullback-Leibler Distance.". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 525–532, 2002. 23, 30, 33, 45

[Clat 05] O. Clatz, M. Sermesant, P.-Y. Bondiau, H. Delingette, S. K. Warfield, G. Malandain, and N. Ayache. "Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation". *IEEE Transactions on Medical Imaging*, Vol. 24, No. 10, pp. 1334–1346, October 2005. 50

[Coll 95] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. "Automated multi-modality image registration based on information theory". In: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 263–274, 1995. 20

[Coll 98] A. Collignon. *Multi-modality medical image registration by maximization of mutual information*. PhD thesis, Catholic University of Leuven, Leuven, Belgium, 1998. 20

[Coma 02] D. Comaniciu and P. Meer. "Mean Shift: A Robust Approach Toward Feature Space Analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603–619, 2002. 58

[Crem 06a] D. Cremers. "Dynamical statistical shape priors for level set-based tracking". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 8, pp. 1262–1273, Aug. 2006. 22, 33

[Crem 06b] D. Cremers, C. Guetter, and C. Xu. "Nonparametric Priors on the Space of Joint Intensity Distributions for Non-Rigid Multi-Modal Image Registration". In: *Proceedings of Computer Vision and Pattern Recognition Conference*, pp. 1777–1783, 2006. 11, 23, 45

[Crem 06c] D. Cremers, S. J. Osher, and S. Soatto. "Kernel Density Estimation and Intrinsic Alignment for Shape Priors in Level Set Segmentation". *International Journal of Computer Vision*, Vol. 69, No. 3, pp. 335–351, 2006. 58

[DAgo 03] E. D'Agostino, J. Modersitzki, F. Maes, D. Vandermeulen, B. Fischer, and P. Suetens. "Free-Form Registration Using Mutual Information and Curvature Regularization". In: *Proceedings of Workshop on Biomedical Image Registration*, pp. 11–20, 2003. 24

[Denz 07] J. Denzler, J. Hornegger, J. Kittler, and C. R. M. Jr., Eds. *Sensor Data and Information Fusion in Computer Vision and Medicine, 30.07. - 04.08.2006*, Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2007. 1

[Dick 08] T. Dickfeld, P. Lei, V. Dilsizian, J. Jeudy, J. Dong, A. Voudouris, R. Peters, M. Saba, R. Shekhar, and S. Shorofsky. "Integration of three-dimensional scar maps for ventricular tachycardia ablation with positron emission tomography-computed tomography.". *Journal of Cardiovascular Imaging*, Vol. 1, No. 1, pp. 73–82, Jan 2008. 86

[Dros 07] M. Droske and M. Rumpf. "Multiscale joint segmentation and registration of image morphology.". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 12, pp. 2181–2194, Dec 2007. 1

[Fan 08] Z. Fan, C. Vetter, C. Guetter, D. Yu, R. Westermann, A. Kaufman, and C. Xu. "Optimized GPU implementation of learning-based non-rigid multi-modal registration". In: *Proceedings of SPIE Medical Imaging*, pp. 69142Y–69142Y–10, 2008. 11

[Fisc 03] B. Fischer and J. Modersitzki. "Curvature based image registration". *Journal of Mathematical Imaging and Vision*, Vol. 18, No. 1, 2003. 16, 22, 40

[Fish 01] J. W. Fisher, III, J. W. F. Iii, W. T. Freeman, T. Darrell, and P. Viola. "Learning Joint Statistical Models for Audio-Visual Fusion and Segregation". 2001. 1

[Fitz 98] J. M. Fitzpatrick, J. B. West, and C. R. Maurer. "Predicting error in rigid-body point-based registration.". *IEEE Transactions on Medical Imaging*, Vol. 17, No. 5, pp. 694–702, Oct 1998. 19

[Fric 04] H. Fricke, E. Fricke, R. Weise, A. Kammeier, O. Lindner, and W. Burchert. "A Method to Remove Artifacts in Attenuation-Corrected Myocardial Perfusion SPECT Introduced by Misalignment Between Emission Scan

and CT-Derived Attenuation Maps". *Journal of Nuclear Medicine*, Vol. 45, No. 10, pp. 1619–1625, October 2004. 70, 72, 73, 74, 77

[Fuku 08] K. Fukuzawa, A. Yoshida, S. Kubo, T. Takano, K. Kiuchi, G. Kanda, K. Takami, H. Kumagai, S. Torii, M. Takami, M. Yokoyama, and K. ichi Hirata. "Endocardial substrate mapping for monomorphic ventricular tachycardia ablation in ischemic and non-ischemic cardiomyopathy.". *Kobe J. Med. Sci.*, Vol. 54, No. 2, pp. E122–E135, 2008. 85

[Gaen 98] T. Gaens, F. Maes, D. Vandermeulen, and P. Suetens. "Non-rigid Multimodal Image Registration Using Mutual Information". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 1099–1106, 1998. 24

[Gan 04] R. Gan, J. Wu, A. C. S. Chung, S. C. H. Yu, and W. M. Wells. "Multiresolution Image Registration Based on Kullback-Leibler Distance.". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 599–606, 2004. 33

[Ghol 07] A. Gholipour, N. Kehtarnavaz, R. Briggs, and K. Gopinath. "Kullback-Leibler Distance Optimization for Non-rigid Registration of Echo-Planar to Structural Magnetic Resonance Brain Images". In: *Proceedings of International Conference on Image Processing*, pp. 221–224, 2007. 23, 24

[Goet 07a] S. Goetze, T. L. Brown, W. C. Lavely, Z. Zhang, and F. M. Bengel. "Attenuation Correction in Myocardial Perfusion SPECT/CT: Effects of Misregistration and Value of Reregistration". *Journal of Nuclear Medicine*, Vol. 48, No. 7, p. 1090, June 2007. 74

[Goet 07b] S. Goetze and R. L. Wahl. "Prevalence of misregistration between SPECT and CT for attenuation-corrected myocardial perfusion SPECT.". *Journal of Nuclear Cardiology*, Vol. 14, No. 2, pp. 200–206, April 2007. 72, 74

[Grim 96] W. L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Perez, W. M. Wells, and R. Kikinis. "An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization.". *IEEE Transactions on Medical Imaging*, Vol. 15, No. 2, pp. 129–140, 1996. 1

[Guet 03] C. Guetter. *A Variational Approach To Particle Tracking Velocimetry*. Master's thesis, University of Mannheim, Germany, 2003. 19

[Guet 05] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. "Learning Based Non-Rigid Multi-Modal Image Registration Using Kullback-Leibler Divergence". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 255–262, 2005. 11, 23, 31, 33, 45

[Guet 07] C. Guetter, M. Wacker, C. Xu, and J. Hornegger. "Registration of Cardiac SPECT/CT Data through Weighted Intensity Co-Occurrence Priors". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 725–733, 2007. 3, 11, 23, 69, 87

[Hajn 01] J. V. Hajnal, D. L. Hill, and D. J. Hawkes. *Medical Image Registration. Biomedical Engineering Series*, CRC Press LLC, Hammersmith Hospital, London, UK, June 2001. 14

[Han 08] J. Han, H. Köstler, C. Bennewitz, T. Kuwert, and J. Hornegger. "Computer-aided evaluation of anatomical accuracy of image fusion between X-ray CT and SPECT.". *Comput Med Imaging Graph*, Vol. 32, No. 5, pp. 388–395, Jul 2008. 70

[He 03] J. He and G. E. Christensen. "Large Deformation Inverse Consistent Elastic Image Registration". In: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 438–449, 2003. 22

[Hell 04] G. Heller, J. Links, T. Bateman, J. Ziffer, E. Ficaro, M. Cohen, and R. Hendel. "American Society of Nuclear Cardiology and Society of Nuclear Medicine joint position statement: attenuation correction of myocardial perfusion SPECT scintigraphy". *Journal of Nuclear Cardiology*, Vol. 11, No. 2, pp. 229–229, 2004. 73

[Herm 02] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras. "Variational Methods for Multimodal Image Matching". *International Journal of Computer Vision*, Vol. 50, No. 3, pp. 329–343, December 2002. 16, 23, 24, 30, 33, 39, 40

[Hill 01] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. "Medical image registration.". *Physics in Medicine and Biology*, Vol. 46, No. 3, pp. R1–45, Mar 2001. 19, 21

[Hill 93] 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: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 182–196, Springer-Verlag, London, UK, 1993. 20

[Horn 81] B. Horn and B. Schunck. "Determining Optical Flow". *Artificial Intelligence*, Vol. 17, pp. 185–203, 1981. 16, 39

[http] "http://en.wikipedia.org/". 4

[Huan 04] X. Huang, D. Metaxas, and T. Chen. "MetaMorphs: Deformable Shape and Texture Models". In: *Proceedings of Computer Vision and Pattern Recognition Conference*, pp. 496–503, 2004. 58

[Imag 07] "Imaging Guidelines for Nuclear Cardiology Procedures: SPECT and Instrumentation". *Journal of Nuclear Cardiology*, Vol. 14, p. 913, 2007. 73

[Jage 09] F. Jäger and J. Hornegger. "Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging". *IEEE Transactions on Medical Imaging*, Vol. 28, No. 1, pp. 137–150, 2009. 41

[Jann 02] P. Jannin, J. M. Fitzpatrick, D. J. Hawkes, X. Pennec, R. Shahidi, and M. W. Vannier. "Validation of medical image processing in image-guided therapy.". *IEEE Transactions on Medical Imaging*, Vol. 21, No. 12, pp. 1445–1449, December 2002. 80

[Jann 06] P. Jannin, C. Grova, and C. Maurer. "Model for defining and reporting reference-based validation protocols in medical image processing". *Journal of Computer Assisted Radiology and Surgery*, Vol. 1, No. 2, pp. 63–73, August 2006. xiii, 67, 68, 69, 74, 80

[Joll 10] M.-P. Jolly, C. Guetter, and J. Guehring. "Cardiac Segmentation in MR Cine Data Using Inverse Consistent Deformable Registration". In: *Proceedings of International Symposium on Biomedical Imaging: from Nano to Macro*, pp. 484–487, 2010. 3, 69

[Josh 00] S. C. Joshi and M. I. Miller. "Landmark matching via large deformation diffeomorphisms.". *IEEE Transactions on Image Processing*, Vol. 9, No. 8, pp. 1357–1370, 2000. 19

[Kale 00] W. A. Kalender. *Computed Tomography.* Publicis MCD Verlag, 2000. 4, 6

[Kipp 04] M. S. Kipper and M. Tartar. *Clinical Atlas of PET: with Imaging Correlation.* Saunders, 1st Ed., 2004. 4, 8

[Kohl 06] T. Kohlberger, D. Cremers, M. Rousson, R. Ramaraj, and G. Funka-Lea. "4D Shape Priors for a Level Set Segmentation of the Left Myocardium in SPECT Sequences". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 92–100, 2006. 22, 47, 87, 88

[Krit 05] J. N. Kritzman, S. Dey, J. Corbett, and E. P. Ficaro. "Changes in normal cardiac intensity distribution due to translation differences between CT and SPECT for a hybrid imaging system". *Journal of Nuclear Cardiology*, Vol. 12, No. 4, pp. S122–S123, July 2005. 73, 77

[Lee 05] Z. Lee, D. B. Sodee, M. Resnick, and G. T. Maclennan. "Multimodal and three-dimensional imaging of prostate cancer.". *Computerized Medical Imaging and Graphics*, Vol. 29, No. 6, pp. 477–486, Sep 2005. 1

[Leve 98] M. Leventon and E. Grimson. "Multi-Modal Volume Registration Using Joint Intensity Distributions". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 1057–1066, 1998. 23, 30, 33, 61

[Liao 06] R. Liao, C. Guetter, C. Xu, Y. Sun, A. Khamene, and F. Sauer. "Learning-Based 2D/3D Rigid Registration Using Jensen-Shannon Divergence for Image-Guided Surgery". In: *Proceedings of Workshop on Medical Imaging and Augmented Reality*, pp. 228–235, 2006. 11

[Loat 93] H. Loats. "CT and SPECT image registration and fusion for spatial localization of metastatic processes using radiolabeled monoclonals.". *Journal of Nuclear Medicine*, Vol. 34, No. 3 Suppl, pp. 562–566, Mar 1993. 1

[Maes 97] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. "Multimodality Image Registration by Maximization of Mutual Information". *IEEE Transactions on Medical Imaging*, Vol. 16, No. 2, pp. 187–198, April 1997. 20, 44

[Maes 98] F. Maes. *Segmentation and registration of multimodal medical images: From theory, implementation and validation to a useful tool in clinical practice.* PhD thesis, Catholic University of Leuven, 1998. 21

[Main 98] J. Maintz and M. Viergever. "A survey of medical image registration". *Medical Image Analysis*, Vol. 2, No. 1, pp. 1–36, 1998. 21

[Marc 00] F. E. Marchlinski, D. J. Callans, C. D. Gottlieb, and E. Zado. "Linear ablation lesions for control of unmappable ventricular tachycardia in patients with ischemic and nonischemic cardiomyopathy.". *Circulation*, Vol. 101, No. 11, pp. 1288–1296, Mar 2000. 85

[Parz 62] E. Parzen. "On estimation of a probability density function and mode". *The Annals of Mathematical Statistics*, Vol. 33, pp. 1065–1076, September 1962. 30

[Plui 03] J. Pluim, J. Maintz, and M. Viergever. "Mutual-information-based registration of medical images: a survey". *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 986–1004, August 2003. 26

[Pool 05] R. A. Pooley. "AAPM/RSNA physics tutorial for residents: fundamental physics of MR imaging.". *Radiographics*, Vol. 25, No. 4, pp. 1087–1099, 2005. 9

[Rasm 05] C.-E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2005. 57

[Rizz 97] G. Rizzo, P. Scifo, M. C. Gilardi, V. Bettinardi, F. Grassi, S. Cerutti, and F. Fazio. "Matching a computerized brain atlas to multimodal medical images.". *NeuroImage*, Vol. 6, No. 1, pp. 59–69, July 1997. 1, 3

[Roch 00] A. Roche, G. Malandain, and N. Ayache. "Unifying Maximum Likelihood Approaches in Medical Image Registration". *Journal of Imaging Systems and Technology*, Vol. 11, pp. 71–80, 2000. 21

[Roch 98] A. Roche, G. Malandain, X. Pennec, and N. Ayache. "The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 1115–1124, 1998. 20

[Roge 03] P. Rogelj, S. Kovacic, and J. C. Gee. "Point similarity measures for non-rigid registration of multi-modal data". *Computer Vision and Image Understanding*, Vol. 92, No. 1, pp. 112–140, 2003. 24

[Rohd 03] G. K. Rohde, A. Aldroubi, and B. M. Dawant. "The Adaptive Bases Algorithm for Intensity-Based Nonrigid Image Registration.". *IEEE Transactions on Medical Imaging*, Vol. 22, No. 11, pp. 1470–1479, November 2003. 50

[Rohr 01] K. Rohr. *Landmark-Based Image Analysis: Using Geometric and Intensity Models*. Kluwer Academic Publishers, Norwell, MA, USA, 2001. 19

[Roth 05] S. Roth and M. Black. "On the Spatial Statistics of Optical Flow". In: *Proceedings of International Conference on Computer Vision*, pp. 42–49, 2005. 22, 39

[Rous 02] M. Rousson and N. Paragios. "Shape Priors for Level Set Representations". In: *Proceedings of European Conference on Computer Vision*, pp. 78–92, Springer-Verlag, London, UK, 2002. 22

[Ruec 03] D. Rueckert, A. F. Frangi, and J. A. Schnabel. "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration.". *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 1014–1025, Aug 2003. 22

[Ruec 98] D. Rueckert, C. Hayes, C. Studholme, P. E. Summers, M. O. Leach, and D. J. Hawkes. "Non-rigid Registration of Breast MR Images Using Mutual Information". In: *Proceedings of Medical Image Computing and Computer Assisted Intervention*, pp. 1144–1152, 1998. 24

[Sabu 05] M. Sabuncu and P. Ramadge. "Graph theoretic image registration using prior examples". In: *Proceedings of European Signal Processing Conference*, 2005. 23, 25, 33

[Sabu 08] M. R. Sabuncu and P. Ramadge. "Using spanning graphs for efficient image registration.". *IEEE Transactions on Image Processing*, Vol. 17, No. 5, pp. 788–797, May 2008. 23, 33

[Schn 01] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, M. O. Leach, C. Hayes, A. Degenhard, R. Hose, D. L. G. Hill, and D. J. Hawkes. "Validation of Non-rigid Registration Using Finite Element Methods". In: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 344–357, 2001. 80

[Silv 92] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1992. 58

[Soma 03] S. Soman, A. C. S. Chung, W. E. L. Grimson, and W. M. W. III. "Rigid Registration of Echoplanar and Conventional Magnetic Resonance Images by Minimizing the Kullback-Leibler Distance". In: *Proceedings of Workshop on Biomedical Image Registration*, pp. 181–190, 2003. 45

[Spie 09] M. Spiegel, D. A. Hahn, V. Daum, J. Wasza, and J. Hornegger. "Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration.". *Comput Med Imaging Graph*, Vol. 33, No. 1, pp. 29–39, Jan 2009. 69

[Stud 99] C. Studholme, D. L. G. Hill, and D. J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment". *Pattern Recognition*, Vol. 32, No. 1, pp. 71–86, January 1999. 20, 21

[Swai 91] M. J. Swain and D. H. Ballard. "Color indexing". *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 11–32, November 1991. 30

[Tian 08] J. Tian, M. F. Smith, P. Chinnadurai, V. Dilsizian, A. Turgeman, A. Abbo, K. Gajera, C. Xu, D. Plotnick, R. Peters, M. Saba, S. Shorofsky, and T. Dickfeld. "Clinical Application of PET/CT Fusion Imaging for Three-Dimensional Myocardial Scar and Left Ventricular Anatomy during Ventricular Tachycardia Ablation". *Journal of Cardiovascular Electrophysiology*, Vol. 20, No. 6, pp. 597–605, December 2008. 1, 11, 85, 86, 87, 88, 89

[Tikh 77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed Problems*. Winston and Sons (Wiley), 1977. 16, 33

[Tong 05] C. M. Tonge, M. Manoharan, R. S. Lawson, R. A. Shields, and M. C. Prescott. "Attenuation correction of myocardial SPECT studies using low resolution computed tomography images". *Nucear Medicine Communications*, Vol. 26, No. 3, pp. 231–237, 2005. 74

[Tong 06] C. M. Tonge, G. Ellul, M. Pandit, R. S. Lawson, R. A. Shields, P. Arumugam, and M. C. Prescott. "The value of registration correction in the attenuation correction of myocardial SPECT studies using low resolution computed tomography images". *Nucear Medicine Communications*, Vol. 27, No. 11, pp. 843–852, 2006. 74

[Vett 07] C. Vetter, C. Guetter, C. Xu, and R. Westermann. "Non-Rigid Multi-Modal Registration on the GPU". In: *Proceedings of SPIE Medical Imaging*, pp. 651228–651236, 2007. 11

[Viol 95] P. Viola and W. M. Wells. "Alignment by maximization of mutual information". In: *Proceedings of International Conference on Computer Vision*, pp. 16–23, 1995. 20

[Vogt 06] S. Vogt, A. Khamene, and F. Sauer. "Reality Augmentation for Medical Procedures: System Architecture, Single Camera Marker Tracking, and System Evaluation". *International Journal of Computer Vision*, Vol. 70, No. 2, pp. 179–190, 2006. 1

[Weic 01] J. Weickert and C. Schnörr. "A theoretical framework for convex regularizers in PDE–based computation of image motion". *International Journal of Computer Vision*, Vol. 45, No. 3, pp. 245–264, 2001. 22

[Wein 08] W. Wein, S. Brunke, A. Khamene, M. Callstrom, and N. Navab. "Automatic CT-Ultrasound Registration for Diagnostic Imaging and Image-guided Intervention". *Medical Image Analysis*, Vol. 12, pp. 577–585, Oct. 2008. 1

[Well 96] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. "Multi-Modal Volume Registration by Maximization of Mutual Information". *Medical Image Analysis*, Vol. 1, No. 1, pp. 35–51, 1996. 44

[West 97] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. Hill, D. J. Hawkes, C. Studholme, J. B. Maintz, M. A. Viergever, G. Malandain, and R. P. Woods. "Comparison and evaluation of retrospective intermodality brain image registration techniques.". *J Comput Assist Tomogr*, Vol. 21, No. 4, pp. 554–566, 1997. 68

[Wood 92] R. P. Woods, S. R. Cherry, and J. C. Mazziotta. "Rapid automated algorithm for aligning and reslicing PET images.". *Journal of Computer Assisted Tomography*, Vol. 16, No. 4, pp. 620–633, 1992. 20

[Xue 06] Z. Xue, D. Shen, and C. Davatzikos. "Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping.". *Medical Image Analysis*, Vol. 10, No. 5, pp. 740–751, Oct 2006. 22

[Zaid 03] H. Zaidi and B. Hasegawa. "Determination of the attenuation map in emission tomography.". *Journal of Nuclear Medicine*, Vol. 44, No. 2, pp. 291–315, Feb 2003. 72

[Zhou 06] S. K. Zhou, B. Georgescu, D. Comaniciu, and J. Shao. "BoostMotion: Boosting a Discriminative Similarity Function for Motion Estimation". In: *Proceedings of Computer Vision and Pattern Recognition Conference*, pp. 1761–1768, 2006. 22

[Zoll 03] L. Zöllei, J. W. Fisher, and W. M. Wells. "A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration.". In: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 366–377, 2003. 23, 30

[Zoll 06a] L. Zöllei. *A Unified Information Theoretic Framework for Pair- and Group-wise Registration of Medical Images.* PhD thesis, Massachusetts Institute of Technology, 2006. 25, 33

[Zoll 06b] L. Zöllei and W. M. Wells. "Multi-modal Image Registration Using Dirichlet-Encoded Prior Information.". In: *Proceedings of Workshop on Biomedical Image Registration*, pp. 34–42, 2006. 23

[Zoll 07] L. Zöllei, M. Jenkinson, S. Timoner, and W. M. Wells. "A marginalized MAP approach and EM optimization for pair-wise registration.". In: *Proceedings of International Conference on Information Processing in Medical Imaging*, pp. 662–674, 2007. 23, 25