

DISS. ETH NO. 21402

Liver Motion Tracking in Ultrasound Sequences for Tumor Therapy

A dissertation submitted to
ETH ZÜRICH

for the degree of
Doctor of Sciences (Dr. sc. ETH Zürich)

presented by

Valeria De Luca

MSc in Medical Engineering, Università di Roma Tor Vergata, Italy
born December 25, 1984
citizen of Italy

accepted on the recommendation of

Prof. Dr. Gábor Székely, ETH Zürich, examiner
Prof. Dr. Tim Salcudean, University of British Columbia, co-examiner
Dr. Christine Tanner, ETH Zürich, co-examiner

2013

Abstract

Respiratory motion is a complicating factor for cancer treatments. In order to take full advantage of the recent technical developments in radiation therapy, where highly conformal radiation dose distributions are deposited into the tissues, accurate and real-time respiratory motion compensation is necessary. In this thesis surrogates of the abdominal motion are obtained from tracking points of interest in 2D ultrasound sequences of the liver under free breathing.

This thesis provides a collection of intensity-based methods for vessel tracking in long ultrasound sequences. The most promising approach is based on exploiting and learning the repetition in the images in case of approximately periodic motion, such as breathing.

In addition, I describe the integration of the presented tracking results with a model-based framework for predicting the motion of the treatment target in a novel hybrid MR- and ultrasound-guided high intensity focused ultrasound therapy. Finally, I examined an alternative multimodal approach for predicting the displacements of points of interest, based on transferring information of the breathing motion across modalities.

Sommario

Il movimento respiratorio è un fattore complicante nel trattamento dei tumori. Al fine di trarre il massimo vantaggio dai recenti sviluppi tecnologici nel campo della radioterapia, grazie ai quali dosi radianti altamente focalizzate sono depositate nei tessuti, è necessaria una compensazione accurata e in tempo reale del movimento respiratorio. In questa tesi i surrogati del movimento addominale sono ottenuti dal tracciamento di punti d'interesse in sequenze ecografiche bi-dimensionali epatiche sotto respirazione libera.

Questa tesi fornisce un insieme di metodi, basati sull'intensità delle immagini, per il tracciamento di lunghe sequenze ecografiche. L'approccio più promettente apprende da e sfrutta la ripetizione delle immagini nel caso di moto approssimativamente periodico, come la respirazione.

Inoltre descrivo l'integrazione di risultati del tracciamento presentati con un framework basato su un modello statistico, per predire il movimento del target del trattamento in un nuovo sistema di ultrasuono ad alta intensità focalizzato, guidato da immagini ibride ecografiche e di risonanza magnetica. Infine esamino un approccio multi-modale alternativo per predire gli spostamenti di punti di interesse, basato sul trasferimento di informazioni del movimento respiratorio tra le diverse modalità d'immagine.

Acknowledgment

The work on this thesis was funded by the Swiss National Science Foundation, project CRSII2 127549.

This thesis would not have been possible without the support of a number of people and I would like to extend my gratitude to each person who has helped me on this path.

First and foremost, I would like to thank Professor Gabor Szekely for giving me the opportunity to be a member of the Computer Vision Laboratory at ETH Zurich, and for the freedom to explore such an interesting research field. I would also like to extend my gratitude to Dr. Christine Tanner, for shaping the way I approach research, for guiding me through the last three years, and for teaching me how to conduct research.

A portion of this thesis was the result of collaboration with the University of Basel and the University Hospital of Geneva. My thanks go to all the people that collaborated in the project, especially to Prof. Philippe Cattin, Dr. Frank Preiswerk, Dr. Rares Salomir, and Dr. Lorena Petrusca.

I would particularly like to thank Dr. Helmut Grabner for his creative collaboration, and Michael Tschannen, whose hard work became an integral part of my own research.

To my colleagues at the Computer Vision Laboratory, especially to my office mates and office neighbors, thank you for the friendly and constructive atmosphere.

Lastly, I would like to thank my family (mamy, papino and bro), close friends (especially Zvia, Ferda and Sylvia) and Henning: their support, encouragement and patience during these last three years made the completion of this thesis possible.

Contents

1	Introduction	1
1.1	Application scenario	3
1.2	Contributions and organization	4
2	Related works	7
2.1	Main components in ultrasound tracking	8
2.1.1	Matching criteria	9
2.1.2	Transformation models	10
2.1.3	Optimization Schemes	11
2.1.4	Pre-processing	13
2.2	Tracking methods	13
2.2.1	Optical flow	13
2.2.2	Speckle tracking	16
2.2.3	Intensity-based registration	19
2.2.4	Feature tracking	21
2.2.5	Hybrid image registration	26
2.3	Validation	29
2.4	Conclusion	31
3	Intensity-based tracking for long sequences	40
3.1	Ultrasound sequences	40
3.2	Evaluation method	41
3.3	Affine registration	43
3.3.1	Registration details	43
3.3.2	Computational complexity	44
3.4	Scale-adaptive block-matching	44
3.4.1	Block configuration	45
3.4.2	Displacement calculation	45
3.4.3	Reference frame definition	46
3.4.4	Computational complexity	47
3.5	Results	48
3.6	Conclusion	49

4	Learning-based real-time tracking	52
4.1	Training phase	53
4.2	Real-time application phase	54
4.2.1	Learning-based affine registration	54
4.2.2	Learning-based block-matching algorithm	55
4.2.3	Computational complexity	56
4.3	Results	57
4.4	Conclusion	59
5	Simultaneous MR- and ultrasound-guided radiation therapy	65
5.1	Hybrid ultrasound/MR imaging system	65
5.1.1	MR-compatible ultrasound imaging	66
5.1.2	Simultaneous motion monitoring of the liver	68
5.1.3	Images co-registration	70
5.2	Model-guided respiratory motion prediction	71
5.2.1	Statistical model of the respiratory motion of the liver	72
5.2.2	Ultrasound tracking	72
5.2.3	Spatial and temporal prediction	74
5.3	Results	75
5.4	Conclusion	76
6	Alternative liver motion prediction	78
6.1	Material	79
6.2	Method	80
6.2.1	Dimensionality reduction	80
6.2.2	Selection of common cause signals	83
6.2.3	Prediction of feature location	83
6.3	Prediction evaluation	84
6.4	Results	84
6.5	Conclusion	86
7	Conclusion	88
7.1	Contributions	88
7.2	Future work	89
	Bibliography	92

List of abbreviations

- (M)TE (Mean) tracking error, page 41
- K -NN K -nearest neighbors, page 54
- BMA Block-matching algorithm, page 44
- HIFU High intensity focused ultrasound, page 1
- LB-Affine Learning-based affine registration, page 57
- LB-BMA Learning-based block-matching algorithm, page 55
- LB-BMA _{B} Learning-based block-matching algorithm based on the breathing signal, page 57
- LB-BMA _{B} (10C) Learning-based block-matching algorithm based on the breathing signal and a training set covering 10 breathing cycles, page 58
- LB-BMA₉₅ Learning-based block-matching algorithm based on 95% of the cumulative energy of the sorted eigenvalues, page 57
- LB-BMA₉₅(10C) Learning-based block-matching algorithm based on 95% of the cumulative energy of the sorted eigenvalues and a training set covering 10 breathing cycles, page 57
- MR(I) Magnetic resonance (imaging), page 2
- NCC Normalized cross correlation, page 43
- PCA Principal component analysis, page 53
- RF Radio frequency, page 41
- RMS Root mean square, page 68
- SA-BMA Scale-adaptive block-matching algorithm, page 45
- SFA Slow feature analysis, page 80
- STD Standard deviation, page 41

1

Introduction

Respiratory motion is a complicating factor for the treatment of non-resectable tumors located in the thorax and abdomen [Keall *et al.* 2006, Verma *et al.* 2011]. An established method for cancer treatment is radiation therapy, which uses ionization radiation to destroy tumor cells. During radiation therapy, a lethal dose should be delivered to the cancerous tissue, while sparing the healthy one. Therefore, accurate target localization and dose delivery are the main challenges in radiation therapy. The motion of the tumor caused by the patient breathing can compromise the effectiveness of the treatment.

Recent technical developments in radiation therapy, such as intensity modulated radiation therapy (IMRT), intensity modulated proton therapy (IMPT), and high intensity focused ultrasound (HIFU), provide the capability to deposit highly conformal radiation dose distributions into the tissues. However, the highly localized treatment makes these techniques sensitive to organ motion. Therefore compensation for any movement in the treatment region, like respiratory motion in the abdomen, is vital [Keall *et al.* 2006, Shirato *et al.* 2007]. This requires an accuracy in the range of millimeters and real-time feedback.

Unfortunately, observation and quantification of the motion of all structures of interest in real-time during therapy is often impossible. Instead, several treatment methods have been proposed to deal with respiratory motion. An obvious solution is to ask the patient to hold his/her breath while the therapy beam is on. In gated treatments the beam is only turned on during a certain phase of the breathing cycle, e.g. end exhalation. Although the aforementioned approaches compensate to some extent for breathing motion, they require reproducibility of the organ position for the selected breathing phase and a longer treatment time. Breath-holding also requires patient compliance. In addition, they compensate only for breathing motion, but not for other motion modes, e.g. organ drift. Thus, they are always sufficiently accurate within first 3 minutes after the patient set up, but deviated significantly for most subjects after 20 minutes [von Siebenthal *et al.* 2007c].

It would be desirable to keep the target and the treatment beam aligned throughout the entire breathing cycle. This technique is commonly referred to as tracking. A common approach is based on surrogate measures of the target motion and a model which relates target and surrogate motion [Keall *et al.* 2006]. Such surrogates include external signals such as recordings from a breathing bellow or a spirometer, and internal measurements from fast imaging (called image guidance). Studies have shown that for the abdomen internal surrogates correlate better with the target motion during breathing than external respiratory signals [Keall *et al.* 2006]. In addition, studies of the liver motion showed that an indirect strategy, based on a statistical motion model in combination with tracking a few surrogate markers inside the liver, allowed to accurately estimate drifts of the organ [von Siebenthal *et al.* 2007b, McClelland *et al.* 2011]. Thus, tracking internal landmarks has great potential for motion management.

Image guided therapies use image information gathered during therapy for adjusting the treatment plan. Image guidance can be achieved using different imaging modalities, such as fluoroscopy, computer tomography (CT), magnetic resonance imaging (MRI) or ultrasound. The latter represents the only technique capable of imaging and displaying soft tissue deformations and organ displacements in real-time, together with being non-ionizing and cheap.

Ultrasound imaging is a well established and widely used medical imaging technique. Echographic systems are inexpensive, easily accessible, relatively small and mobile, and do not require dedicated rooms. Ultrasound is a safe image modality as it is relying on non-ionizing radiation. In addition, ultrasound imaging is real-time and can generate image sequences at a high frame rate. Most available systems are using 2D probes. As an emerging technology, 3D ultrasound systems allow real-time volumetric imaging, which is particularly appealing for direct visualization of 3D anatomy and motion analysis. As ultrasound has a high temporal resolution, is non-invasive and low-cost, it is commonly used in clinical practice for diagnosis and follow-up, especially for cardiac screening (echocardiography), obstetrics, vascular imaging, abdominal imaging and breast cancer detection.

However, ultrasound has significant drawbacks compared to tomographic image modalities, such as MRI and CT, for imaging the anatomy. The generated images are difficult to understand and interpret. They are often limited by small acoustic windows and their quality is affected by several types of noise and artifacts. These include shadows and mirroring, which are mainly due to highly attenuating structures like bones and/or strongly reflecting interfaces. These artifacts are useful for some applications to understand the tissue composition, e.g. shadowing below a breast lesion. However, they are not stable in the presence of motion, as they depend on the tissue composition along the ultrasound beam. Therefore they represent an obstacle in estimating tissue motion.

Ultrasound tracking is the task of following (moving) structures on the ultrasound sequence over time. To achieve this, the spatial correspondence of a structure appearing on different frames has to be determined. The resulting displacement vectors of consecutive frames can be concatenated to provide the motion trajectory of a structure. Tracking features on images is a type of image registration with the focus on extracting quickly the motion of some structures on image sequences, rather than spatially transforming whole images to bring them into alignment. In the clinic, tracking information provides a feedback of the current position to allow for motion compensation in therapy and surgery guidance, for conformal treatment systems and quantitative physiopathological analysis.

The capability of imaging and displaying soft tissue deformations and organ displacements in real-time, makes ultrasound an appealing choice for medical applications which require tracking and analysis of tissue motion. These include blood flow measurement, strain rate imaging, ventricular deformation analysis, and motion compensation in image-guided intervention and therapy. The latter application is the focus of this thesis.

1.1 Application scenario

HIFU is a therapy procedure that applies high-intensity focused sonic energy to locally heat (up to 85°C) and destroy tumor tissue through ablation. It is a high precision (with a focus in the order of millimeters) and minimally or non-invasive method. The focusing of the ultrasound beam is mostly achieved electronically. The beam is steered to the target location by dynamically adjusting the relative phase of the phased array of the therapeutic transducer [Kennedy *et al.* 2003, Amin *et al.* 2008, Zhang and Wang 2010, Wijlemans *et al.* 2012].

Clinical HIFU procedures are usually guided by an imaging procedure, to enable treatment planning and targeting before applying the therapeutic thermal dose. Most commonly are MR-guided HIFU (MRgHIFU) [Ries *et al.* 2010, Holbrook *et al.* 2010, Arnold *et al.* 2011, Wijlemans *et al.* 2012] and ultrasound-guided HIFU (USgHIFU) [Pernot *et al.* 2003, Pernot *et al.* 2004, Marquet *et al.* 2006, Wu *et al.* 2005].

The ideal image-guided HIFU treatment would integrate two imaging modalities, such as MRI and ultrasound, acquired simultaneously, that can provide complementary and synergistic information. MRI can visualize the tumor and quantitatively measure the temperature elevation in and around it. Ultrasound imaging can be used for tracking of specific anatomical landmarks in the images, in order to give real-time information on the tissue motion. The motion tracking enables to synchronize the HIFU beam with the respiratory cycle and to lock it onto the target. In such scenario

the treatment time can be reduced and healthy tissue spared [Petrusca *et al.* 2012, Petrusca *et al.* 2013].

The proposed application scenario for a future HIFU treatment session is as follows:

1. **Pre-treatment phase**, where a static diagnostic 3D MRI of the patient is used to establish correspondence with the statistical motion model of the organ under treatment by segmenting the organ and selecting a few natural landmarks, which are important to predict the organ motion
2. **Set up phase**, where the ultrasound imaging plane is co-registered with the coordinate system of the MRgHIFU device and with the pre-treatment 3D MRI. An initial 2D ultrasound sequence is acquired to train the tracking algorithm
3. **Treatment phase**, where 2D ultrasound images are continuously acquired. Tracking is performed on these images in real-time in order to obtain the position of the surrogate markers. Based on the latter, the statistical motion model predicts the organ location and thus the tumor position for the treatment time

For validating the proposed motion compensation framework, simultaneous acquisition of MRI and ultrasound sequences is performed.

1.2 Contributions and organization

The main aim of this project is to develop and evaluate a liver motion prediction scheme for HIFU treatments. The primary focus of this thesis is to develop and validate accurate, robust and real-time tracking algorithms for motion compensation in 2D ultrasound sequences of the liver, acquired for several minutes and under free-breathing. The specific contributions and the organization of this thesis are as follows.

Chapter 2, *Related works*, provides an extensive literature review of the tracking methods developed for ultrasound time sequences. The tracking problem and its key components are defined. Existing tracking methods are categorized with regard to their motion estimation approach. Their main components, strength and weakness are described and their clinical applicability is discussed. Highly relevant approaches on 2D ultrasound tracking of the liver are discussed in more details. This chapter is based on the survey paper [De Luca *et al.* 2013a].

In Chapter 3, *Intensity-based tracking for long sequences*, I select two promising intensity-based tracking methods, namely a local affine registration [De Luca *et al.* 2012, De Luca *et al.* 2013b, Preiswerk *et al.* 2013] and a scale-adaptive block

matching algorithm. The latter combines several block-matching components and includes a novel adaptation of the block size to the feature scale. This work was originally presented in [De Luca *et al.* 2013b]. Their performance in long ultrasound sequences of the liver is assessed. In addition, I describe the characteristics of the acquired ultrasound sequences and the validation procedure, which is based on manual annotation of vessels centers for ground truth definition. The described data and the validation procedure are used to estimate the tracking accuracy of all evaluated tracking methods in this thesis.

Chapter 4, *Learning-based real-time tracking*, presents two novel learning-based approaches that enable robust and real-time tracking for repetitive motion scenarios. This work was originally presented in [De Luca *et al.* 2012, De Luca *et al.* 2013b]. The first method speeds up affine registration by exploiting the redundancy within the images in the case of repetitive movement, such as breathing. During an initial training phase the images are registered and the relationship between the image appearance and the displacements is learned. For each image in the real-time application phase, the most similar images in the training set are selected for predicting the associated displacements. I incorporate a mechanism to cope with unseen variations in the images during the application phase, which allows for tracking of non-periodic motions. [De Luca *et al.* 2012]. This method is adaptive, but requires time consuming affine registration for adaptation. Hence a second method is proposed, where fast and robust block matching is used for adaptation. In details, it is based on a scale-adaptive block-matching and temporal realignment driven by the image appearance learned from an initial training phase. I take advantage of the approximate periodicity of the breathing motion for learning image appearance and corresponding motion behavior (extracted by accurate but slow image registration) to allow frequent temporal realignment of the block-matching algorithm for drift-free real-time tracking [De Luca *et al.* 2013b].

Chapter 5, *Simultaneous MR- and ultrasound-guided radiation therapy*, describes the proposed application scenario, where the ultrasound system is modified to be MR-compatible. This allows for simultaneous acquisitions of ultrasound and MR images before and during treatment. Spatial-temporal prediction is based on the motion information from ultrasound, while the motion information from both co-registered modalities allows for validation. In details, accurate tracking surrogates of the liver motion from the ultrasound images is used to predict the tumor position, in combination with a statistical motion model of the liver. The work of this chapter is based on [Petrusca *et al.* 2013, Preiswerk *et al.* 2013].

In Chapter 6, I present a novel and unconventional approach for multi-modal motion prediction. It exploits the temporal commonality of multi-modal images acquired from the same organ at different locations during free-breathing. Strikingly there is no need for capturing the same region by the modalities. The method is based on extracting a low dimensional description of the image sequences, using slow feature

analysis, selecting the common breathing signal for both modalities and finding the most similar sub-sequences for predicting image feature location. The approach was evaluated for 3 volunteers on sequences of 2D MRI and 2D ultrasound images of the liver. Simultaneous acquisition of these images allowed for quantitative evaluation. The work of this chapter was originally presented in [De Luca *et al.* 2011].

Finally, Chapter 7 concludes this thesis, where I summarize and discuss the presented work, and outline future research directions.

2

Related works

Ultrasound tracking is the task of following (moving) structures on the ultrasound sequence over time. To achieve this, the spatial correspondence of a structure appearing on different frames has to be determined. The resulting displacement vectors of consecutive frames can be concatenated to provide the motion trajectory of a structure. Tracking features on images is a type of image registration with the focus on extracting quickly the motion of some structures on image sequences, rather than spatially transforming whole images to bring them into alignment. In the clinic, tracking information provides a feedback of the current position to allow for motion compensation in therapy and surgery guidance, for conformal treatment systems and quantitative physiopathological analysis.

This Chapter outlines the state-of-the-art of tracking ultrasound time sequences, classified according to methodology and clinical applications. I consider only sequences of 2D and 3D images, based on radio frequency (RF) and enveloped (B-mode) data, continuously acquired at the same spatial location. Image classification, compounding, segmentation, registration of non-consecutive images and multimodal approaches are not reviewed.

The Chapter is structured as follows. Section 2.1 defines the tracking problem in ultrasound sequences and gives an overview of the main components in ultrasound tracking, i.e. the data term, the regularization term and the optimization strategy. In Section 2.2 I list the main tracking strategies and classify them according to the used image features. Therefore I discriminate between speckle tracking, intensity-based registration, feature tracking, and hybrid approaches. For each strategy I focus on its main components, advantages and limitations, reported performance and clinical application. In Section 2.3 I discuss the importance of ultrasound tracking validation and describe the most common performance measures. I also comment on the state-of-the-art results. Section 2.4 concludes the Chapter with a global assessment of the current situation and presents possible future research directions.

2.1 Main components in ultrasound tracking

In a clinical scenario, ultrasound sequences are acquired for estimating the motion of anatomical structures. The movement of a structure is described by motion vectors, obtained by finding the spatial correspondence of the structure in consecutive images. Image registration is the process of determining the spatial correspondence between images. Given two images, a reference (or fixed) and a source (or moving) image, image registration spatially transforms the source image in order to align it with the reference image.

The methods for estimating the motion vectors can be categorized into direct and indirect schemes. Direct methods are intensity-based. This means that the optimization criterion is derived from the intensities of the corresponding pixels after applying a spatial transformation. Intensity-based methods spatially transform the source image in order to align it with the reference image, see Figure 2.1. Indirect methods first extract a set of features, e.g. points, local intensity statistics or edges, from the images and then match these features across the images to find spatial correspondences. From these sparse correspondences, a continuous transformation is then estimated to map the moving image to the fixed image. This estimation can also serve as regularization and needs to address the issue of outliers. Tracking focuses in comparison to image registration on estimating the motion of some image structures, rather than aligning the images. Beside spatial accuracy, tracking algorithms need to provide fast motion extraction which is temporally consistent. Both motion estimation schemes have strength and weaknesses. The main advantage of the direct method is that it does not require the extraction of features. Yet it is difficult to define an image similarity measure which is robust to changes in image appearance and large displacements. Contrarily, indirect approaches are usually faster and better suited for large displacements and intensity variations. The main drawback is the feature detection. In fact, there may be only a few features extracted and matched correctly, due to the presence of noise and artifacts in the images.

Image registration is an ill-posed problem, as it requires one displacement vector per voxel for transforming the moving image, while the image provides insufficient information in regions without point features. To address this so-called aperture problem it is necessary to constrain the solution. Strategies for this include using a transformation model with fewer parameters; actively changing the displacement field (e.g. filtering the displacement field to make it divergence-free); or penalizing unrealistic properties of the displacement fields, like volume changes and non smoothing deformations, by adding a regularization term to the cost function.

The main components of image registration are the matching criterion, the transformation model and the optimization scheme, which will be discussed next. Many

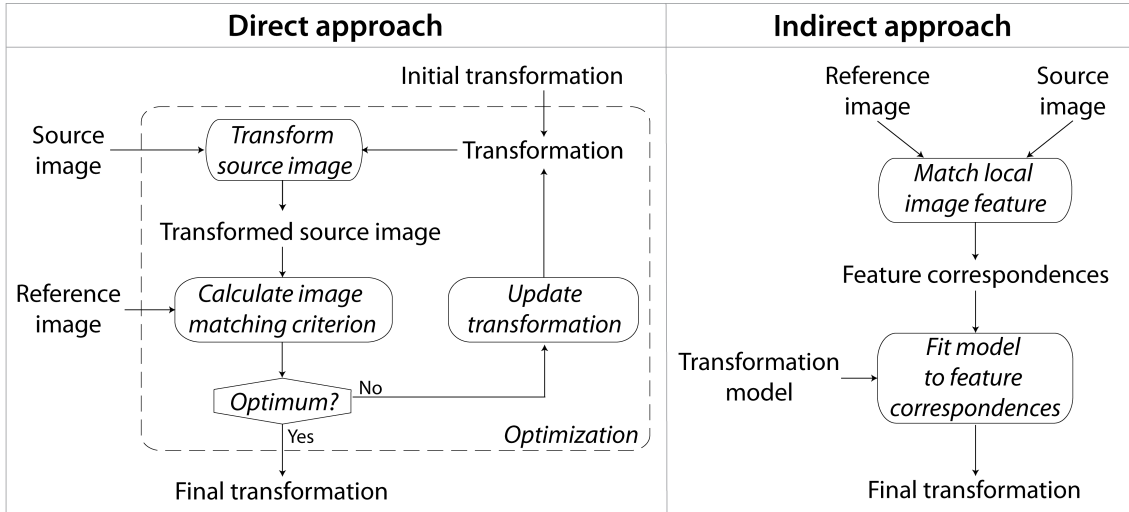


Figure 2.1: Illustration of direct (left) and indirect (right) motion estimation approach.

reviews in the field of image registration and object tracking are currently available [Maintz and Viergever 1998, Hill *et al.* 2001, Zitova and Flusser 2003, Yilmaz *et al.* 2006, Sotiras *et al.* 2012]. In this Chapter I focus on strategies specifically developed for tracking ultrasound sequences.

2.1.1 Matching criteria

In intensity-based methods, the data term is formulated using an image similarity metrics. Ideally, the similarity measure should have a single global optimum when the images are in spatial correspondence. However, noise, artifacts and missing contrast in the images make it difficult to establish the matching between the images. Intensity-based similarities are usually computed over regions of interest (ROIs) of the images or over small blocks (e.g. block-matching algorithm). A common similarity measure used in ultrasound tracking is based on the difference between the intensities in the images, such as the sum of squared difference (SSD) or sum of absolute difference (SAD). The SSD formulation is based on the assumption that additive, uncorrelated Gaussian noise is predominantly responsible for the difference between intensities. In ultrasound monomodal applications this assumption does not always hold, due to the different sources of noise and artifacts. Widely used approaches are correlation techniques, e.g. cross correlation (CC) and normalized cross correlation (NCC), that assume a linear relationship between the intensity distributions. NCC has proved to be robust and able to deal with outliers. As the intensity relationship is not necessarily linear, more sophisticated approaches have also been investigated, based on the joint probability between the image intensities.

The most popular example is mutual information (MI), which is based on the relationship of joint versus marginal entropy. MI was initially developed for multimodal applications to cope with the same structure having different appearance in the different modalities. In the context of ultrasound image matching, its performance is unlikely to be superior to NCC, since shadows and other artifacts in the images often inconsistently affect different image structures. An ultrasound-specific similarity measure based on a maximum likelihood (ML) framework, has been proposed by [Cohen and Dinstein 2002]. The likelihood function is based on the assumption that both images are contaminated by Rayleigh distributed multiplicative noise.

In feature-based approaches the data term is usually formulated as the Euclidean or Mahalanobis distance between the descriptors of the matched features. In dynamic contour tracking, deformable models are applied dynamically to the image sequence. An external energy is used to pull the contour of the previous frame towards matching the current image features, like large image gradients [Cohen and Cohen 1993, Blake and Isard 1998], local region-based information [Paragios and Deriche 2002], local phase, or combinations of these. In feature-based tracking there are often false feature correspondences, due to erroneous extraction and/or matching of the features. Therefore a more robust way of formulating the data term is by using a probabilistic approach based on M-estimators [Blake and Isard 1998]. Examples are least squares and different ML estimators [Jacob *et al.* 2002, Comaniciu *et al.* 2004], which are commonly adopted in Bayesian tracking frameworks.

2.1.2 Transformation models

Transformation models define the space of possible spatial transformations. They are very useful for constraining the registration problem and act as interpolators for information-deprived regions. Their choice is also driven by a compromise between computational efficiency and adequate motion description. The transformations can be classified according to the involved degrees of freedom (e.g. rigid, affine, deformable) and their spatial coverage (local or global).

There are several ways for enforcing regularization of the dense correspondences in a registration problem. In a straightforward approach, one can reduce the degrees of freedom of the estimated transformation, like only allowing translations. More flexible transformations interpolate point correspondences by global or local functions (e.g. thin-plate spline, B-spline), where the correspondences come from feature-matching or are determined as transformation parameters during intensity-based registration [Rueckert *et al.* 1999]. In a more explicit way, the transformation model is regularized by an additional term in the cost function, which penalizes some transformation properties like irregularity. Smoothing the deformation field in a post-processing step is another strategy for regularization. For time sequences,

temporal smoothness can usually also be expected and hence temporal regularization has been applied.

2.1.3 Optimization Schemes

The optimization task in image registration is generally to find the transformation parameters which provide the global optimum of the matching criteria. Such a solution can be found via exhaustive search or optimization strategies. Exhaustive search evaluates the goal function on a full grid defined by a sufficiently fine quantization of the transformation parameters. Although exhaustive search is computational expensive, it is simple to implement and thus still applied in cases where the transformation is limited to translations of a few points within a small neighborhood and for few spatial points, e.g. block-, points- and contour-matching algorithms.

In many cases the registration problem is characterized by a large number of degrees of freedom and therefore optimization algorithms are indispensable for reaching the solution. Optimization strategies can be classified in continuous and discrete methods [Sotiras *et al.* 2012]. Continuous optimization methods treat real value variables, while discrete ones take values from a discrete set. Discrete search spaces are often fully assessed, while continuous methods explore the search space from initial values. The latter is computationally expensive if local optima are to be avoided, but can provide precise results. A third class of optimizers is composed by heuristic methods. While continuous and discrete methods are limited with respect to the nature of the cost function, heuristic methods handle a wide range of problems and solution space. However, they do not provide any guarantee on the optimality of the solution [Sotiras *et al.* 2012].

The most common optimizers used in ultrasound tracking are continuous methods, such as gradient descent (GD) [Snyman 2005], Powell's conjugate direction [Press *et al.* 1992], Levenberg-Marquardt (LM) [Marquardt 1963, Press *et al.* 1992] and quasi-Newton methods [Nocedal and Wright 1999]. The simplest and most used optimizer is GD, where the optimization takes steps in the direction of the negative of the gradient of the objective function. A variation of this method, called stochastic gradient descent (SGD), considers an approximation of the gradient [Bottou 2004]. However GD suffers from slow convergence. More efficient are conjugate direction search methods, with Powell's method [Powell 1964] being applicable to none differential functions (as no gradients are used) while conjugated GD [Hestenes and Stiefel 1952] converges faster.

A common quasi-Newton optimizer used in ultrasound registration is the limited memory Broyden-Fletcher-Goldfarb-Shannon optimization routine with simple bounds (LBFGSB) [Byrd *et al.* 1995]. It gives a good performance for the optimization of a

large amount of parameters, such as in B-spline transformations, while also eliminating the need for storing the inverse of the Hessian during the optimization routine [Heyde *et al.* 2012].

The Gauss-Newton (GN) algorithm is only used to minimize sum of squared function values [Björck 1996, Nocedal and Wright 1999], like SSD and squared regularization terms. Its advantage is that it does not require the computation of second derivatives. A related approach is the LM algorithm, which uses a simplified Hessian matrix [Pennec *et al.* 2003]. It is slower but more robust than the GN algorithm.

[Mignotte *et al.* 2001] implemented an optimization technique based on the Genetic Algorithm (GA) with an elitist strategy, called hybrid GA. GA is a robust stochastic search and global optimization procedure which mimics the evolution of natural systems [Goldberg 1989]. Its hybrid implementation combines the genetic search with a local optimization technique. In details, the elite-preservation strategy [Goldberg 1989] is used to select the individuals of the next generation, based on their higher fitness. Then a percentage of the best individuals are used to initialize a gradient ascent.

An discrete optimizer is graph-cuts [Tang and Hamarneh 2010, Ishikawa 2009], which is based on the max-flow min-cut theorem to define a minimal cut of the graph. The energy function is usually formulated as a likelihood function, Gaussian mixture model (GMM) or Markov random field (MRF). A technique invented specifically for MRF optimization is the Iterated Conditional Modes (ICM) [Geman and Geman 1984]. ICM is a deterministic algorithm that proceeds first by choosing an initial configuration for the variables. Then, it iterates over each node in the graph and calculates the value that minimizes the energy given the current values for all the variables in its neighborhood. The algorithm is guaranteed to converge, and may be terminated according to a chosen criterion of convergence [Geman and Geman 1984]. ICM was used for block-matching by [Yeung *et al.* 1998b].

An example of a heuristic search method used in ultrasound tracking is the Nelder-Mead method [Nelder and Mead 1965, McKinnon 1999, Shekhar and Zagrodsky 2002]. This method does not need to calculate derivatives, can make good improvements for relative few function evaluations, and is simple to understand. However, it does not guarantee convergence and hence can make negligible progress with many iterations. Running the algorithm several times with different starting estimates is used to overcome this problem.

Multiresolution (or coarse-to-fine) approaches have been implemented to increase the probability of finding the global optimum in the parameter space and to make the registration procedure faster and more robust. The images are first registered at a lower resolution for larger expected displacements. The resulting transformation

is applied to the next resolution level. The process is repeated until the highest resolution level is reached.

2.1.4 Pre-processing

The high noise level in ultrasound images is a main challenge for tracking. Ultrasound images often contain speckle patterns, which are a form of multiplicative, locally correlated noise [Yu and Acton 2002]. In many applications, speckles represent the main feature to track, see Sec. 2.2.2. On the contrary, in region- and feature-based registration the presence of speckles can be counterproductive. Therefore, a pre-processing filtering step is useful to improve the tracking performance, at the expense of higher computational costs. Speckle reduction techniques have been listed in [Noble *et al.* 2011]. They include multiplicative models, adaptive filters, partial differential equation- and wavelet-based methods, and non-local-means filter [Noble *et al.* 2011]. In ultrasound tracking, examples of common speckle suppression are anisotropic diffusion [El-Sharkawy *et al.* 2001, Yu and Acton 2002], Gaussian filter [Foughi and Abolmaesumi 2005b, Foughi *et al.* 2006b, Foughi *et al.* 2006a, Leung *et al.* 2009] or median filter [Shekhar and Zagrodsky 2002, Shekhar *et al.* 2004].

2.2 Tracking methods

This Section reviews the current state of tracking ultrasound image sequences. The aim is to provide an overview of the existing tracking methods, describe their critical aspects, and discuss their strengths and limitations. Techniques like speckle tracking or optical flow extensions incorporate knowledge of the ultrasound physics, while registration techniques are often intensity-based. Many tracking approaches consist of extracting features, e.g. boundaries, and tracking them in each frame. Other methods incorporate *a priori* information, such as the likelihood of the shape, deformations or appearance; or the relationship between image patterns and motion vectors, learned from previously analyzed sequences.

2.2.1 Optical flow

Tissue motion in ultrasound images has been estimated by a method called optical flow (OF). Tracking with OF computes a dense displacement field based on the assumption that the image intensities remain constant over time. OF methods successfully applied to ultrasound sequences can be classified into three groups [Barron *et al.* 1994, Boukerroui *et al.* 2003, Fleet and Weiss 2005]:

1. Differential techniques
2. Phase-based techniques
3. Block-matching methods

Differential techniques

Differential techniques compute the motion velocity from spatio-temporal derivatives of the pixel intensities. Typical examples are the Lucas-Kanade [Lucas and Kanade 1981], Horn-Schunck [Horn and Schunck 1981] and Nagel [Nagel 1987] implementations, which are compared on simulated ultrasound data in [Baraldi *et al.* 1996]. These methods give reliable results on good quality images, but they are highly sensitive to noise because of the numerical differentiation. The estimation of the dense displacement field is achieved by minimizing a voxel-based energy function that combines an intensity constraint and a spatio-temporal smoothness constraint of neighboring displacements [Meunier 1998, Pellot-Barakat *et al.* 2004, Tavakoli *et al.* 2008, Fehrenbach *et al.* 2010, Mukherjee *et al.* 2011]. In many cases the optimal solution is computed following a coarse-to-fine approach [Pratikakis *et al.* 2003, Pellot-Barakat *et al.* 2004, Fehrenbach *et al.* 2010, Mukherjee *et al.* 2011].

Differential methods produce inaccurate results where the assumption of intensity constancy is violated, which is likely to happen in ultrasound sequences, e.g. due to artifacts. To avoid this problem, a MRF model of the OF has been proposed [Pellot-Barakat *et al.* 2004]. [Brox *et al.* 2004] recently introduced an OF method which additionally includes a gradient constancy assumption and avoids linearization of the data term. This method leads to a more accurate and less noise-sensitive performance when applied to intraoperative 4D transesophageal echocardiographical images [Mukherjee *et al.* 2011, Mukherjee *et al.* 2012].

Phase-based techniques

Phase-based techniques rely on an initial decomposition of the image into band-pass channels, and assume the conservation of the phase in each channel. The velocity vectors are defined in terms of the phase behavior of the filter outputs. The main advantage offered by the phase is that it is amplitude-invariant, approximately linear over relatively large spatial domains, and stable with respect to affine transformations in consecutive frames [Fleet and Jepson 1990, Fleet and Jepson 1993]. A phase-based scheme was proposed by [Mulet-Parada and Noble 2000] for cardiac wall velocity estimation. The use of the phase information makes the motion estimation robust to attenuation artifacts and therefore theoretically suitable for ultrasound images. However, the performance of such approaches depends on the filter design and the filter response is only optimal for a limited velocity range.

Block-matching methods

Block-matching methods compute the local displacements from the translations which provide the best match of image regions in two consecutive frames. Block-matching algorithms have been (and are still) extensively researched and developed to estimate the motion of speckle patterns in ultrasound image sequences. Such techniques are usually referred to as speckle tracking and are discussed in more detail in Section 2.2.2. In this Section I review only block-matching of image regions relying on features rather than homogeneous regions containing only speckles.

The optimal match is usually found by maximizing a correlation measure [Stoitsis *et al.* 2005, Duan *et al.* 2005b], by minimizing image dissimilarity, like SSD [Mikic *et al.* 1998, Meairs and Hennerici 1999], or by the Cohen and Dinstein ML estimator [Boukerroui *et al.* 2003]. Compared to differential techniques, correlation-based block-matching methods are less sensitive to noise and fast motion, but they assume that the displacements within the local region are the same [Duan *et al.* 2005b]. Region-matching algorithms have been developed following Anandan's algorithm [Anandan 1989, Bergen *et al.* 1992], which is based on a Laplacian pyramid [Burt and Adelson 1983] that allows the computation of large displacements and enhances edges, and a coarse-to-fine SSD-based matching strategy. Such an approach is applied to characterize the plaque motion in carotid artery stenosis [Meairs and Hennerici 1999]. Another matching technique is based on Singh's framework [Singh 1990, Singh and Allen 1992] and tested on ultrasound sequences by [Boukerroui *et al.* 2003, Mikic *et al.* 1998, Duan *et al.* 2005b, Duan *et al.* 2005a, Duan *et al.* 2006, Duan *et al.* 2007]. Singh's matching consists of two stages. In the first stage, the SSD is computed with three temporally adjacent bandpass filtered images. Then, the individual squared differences are transformed into probability distributions in a coarse-to-fine strategy (as in [Anandan 1989]). A confidence measure, based on the eigenvalues of the covariant matrix calculated from the distributions, is included to the matching. In the second stage, weighted least squares velocity estimates are propagated using neighborhood constraints.

The performance of differential OF and block-matching methods was compared for carotid artery images [Stoitsis *et al.* 2005].

OF methods are in general computationally expensive for dense flow field estimation. Therefore the motion estimation is often computed only at some locations, such as selected points [Veronesi *et al.* 2006], extracted vessels [Demi *et al.* 2008, Meairs and Hennerici 1999] or endocardial walls [Duan *et al.* 2005b, Duan *et al.* 2005a, Duan *et al.* 2006]. In other cases the OF hypothesis is used to guide active contours [Mikic *et al.* 1998, Akgul *et al.* 1998]. Table 2.1 summarizes the published works on OF techniques.

2.2.2 Speckle tracking

Speckle tracking has been used in several clinical applications. Well established fields are blood flow assessment and elastography. The latter generates displacement fields resulting from known, applied boundary conditions (forces or displacements). This quantitative information is used to estimate the (relative) tissue elasticity. Ultrasound elastography is typically used for the analysis of the myocardial deformation.

Speckle

Speckle is an inherent characteristic of ultrasound images. It is a random, deterministic interference pattern generated by the constructive and destructive superposition of echo signals reflected from very small scatters [Burckhardt 1978, Bashford and von Ramm 1996]. The interference pattern does not correspond to the imaged anatomy and hence can limit the quality of the ultrasound images, which prevents for example discerning fine structures. However, speckles can be useful in the identification of tissue motion. The interference pattern changes only gradually when the tissue deforms or moves slightly [Meunier 1998]. For such scenarios, the interframe motion can be recovered by locating the same unique pattern in consecutive images even if the tissue itself has otherwise a homogeneous appearance.

The speckle pattern is observable in RF data [O'Donnell *et al.* 1994, Friemel *et al.* 1995, Bashford and von Ramm 1996, Lubinski *et al.* 1999, Morsy and Von Ramm 1999, Kaluzynski *et al.* 2001, Revell *et al.* 2003, Pernot *et al.* 2003, Pernot *et al.* 2004, Revell *et al.* 2005, Chen *et al.* 2005, Yoshikawa *et al.* 2005, Marquet *et al.* 2006, Jiang and Hall 2007, Kuo and von Ramm 2008, Jiang and Hall 2009, Harris *et al.* 2007, Harris *et al.* 2010, Harris *et al.* 2011, Byram *et al.* 2010, Jia *et al.* 2010, Lediju *et al.* 2010, Lediju Bell *et al.* 2012, Compas *et al.* 2011, Compas *et al.* 2012], as well as in envelope-detected and scan converted images [Chen *et al.* 1992, Chen *et al.* 1994, Chen *et al.* 1995, Meunier 1998, Yeung *et al.* 1998b, El-Sharkawy *et al.* 2001, Cohen and Dinstein 2002, Golemati *et al.* 2003, Yu *et al.* 2006, Hsu *et al.* 2005, Notomi *et al.* 2005, Lin *et al.* 2007, Basarab *et al.* 2007, Crosby *et al.* 2009, Korstanje *et al.* 2010, Touil *et al.* 2010], and bandpass energy data [Yeung *et al.* 1998a]. [Yu *et al.* 2006] have presented a comparison of the reliability of different features for speckle tracking, such as envelope-detected, RF and bandpass filtered signals.

Speckle tracking

Speckle tracking assumes that a given region of interest in the image has a unique speckle pattern due to the acoustic properties of the surrounding tissue. Therefore,

the tissue displacement is determined by locating the same unique pattern in the consecutive frames [Hsu *et al.* 2005]. The tracking algorithm is performed by block-matching methods as described above.

In particular, matching is performed using SSD [Yeung *et al.* 1998b, Yeung *et al.* 1998a, Yu *et al.* 2003], SAD [Friemel *et al.* 1995, El-Sharkawy *et al.* 2001, Crosby *et al.* 2009, Touil *et al.* 2010] or in most cases (N)CC as matching criteria. In details, the CC of 1D RF signals was computed in [Lubinski *et al.* 1999, Pernot *et al.* 2003, Pernot *et al.* 2004, Marquet *et al.* 2006]. 2D correlation based algorithms were developed by [Chen *et al.* 1994, Chen *et al.* 1995, Friemel *et al.* 1995, Lubinski *et al.* 1999, Golemati *et al.* 2003, Yoshikawa *et al.* 2005, Hsu *et al.* 2005, Pinton *et al.* 2006, Basarab *et al.* 2007, Jiang and Hall 2007, Jiang and Hall 2009, Korstanje *et al.* 2010, Compas *et al.* 2011], and successively extended for 3D images [Morsy and Von Ramm 1999, Kuo and von Ramm 2008, Harris *et al.* 2007, Harris *et al.* 2010, Harris *et al.* 2011, Jia *et al.* 2010, Compas *et al.* 2012]. Phase-sensitive NCC was originally introduced by [Wear and Popp 1987] and applied in speckle tracking by [O'Donnell *et al.* 1994, Kaluzynski *et al.* 2001, Chen *et al.* 2005, Byram *et al.* 2010, Lediju *et al.* 2010, Lediju Bell *et al.* 2012] to counteract the decorrelation due to variations of the amplitude and phase of the scattered waves. Others proposed motion estimation with a new ML estimator, which takes account of the image formulation by using the assumption that both images are contaminated by Rayleigh distributed, multiplicative noise [Cohen and Dinstein 2002, Lin *et al.* 2007, Revell *et al.* 2003, Revell *et al.* 2005].

Different matching strategies have been suggested. The simple, single-step block-matching method was adapted to improve efficiency and accuracy by multiresolution [Yeung *et al.* 1998b, Revell *et al.* 2005, Lin *et al.* 2007, Basarab *et al.* 2007] and deformable block-matching algorithms [Yeung *et al.* 1998a, Basarab *et al.* 2007, Touil *et al.* 2010]. In addition, speckle tracking has been combined with shape tracking [Pirat *et al.* 2008, Compas *et al.* 2011, Compas *et al.* 2012], to obtain complementary displacement values for the target tissue (via speckle tracking) and its boundary (via shape tracking). In [Kuo and von Ramm 2008] the tracking accuracy is improved by using statistical analysis of the tracked features to reject false matches.

The main contributions in speckle tracking are listed in Table 2.2.

Challenges in speckle tracking

The accuracy of the motion estimates depends on the stability of the speckle pattern, which is assumed in speckle tracking. However in practice, the patterns in two consecutive images are not identical, as the electronic noise distribution varies from frame to frame. In addition, artifacts and motion ambiguities appear in regions of

weak acoustic scatterers, image saturation and signal reverberation. Furthermore real tissues undergo seldom pure translational motion, but also rotation, stretching, shearing or even non-linear deformations, which all cause speckle decorrelation. These changes of the speckle pattern appearance are the main limitation of the accuracy in speckle tracking algorithms.

Tissue motion is not confined to a 2D plane, so it is necessary to track displacements in 3D. Due to technological limitations, initial research in speckle tracking was conducted in 2D ultrasound sequences [Chen *et al.* 1992, Chen *et al.* 1994, Chen *et al.* 1995, O'Donnell *et al.* 1994, Friemel *et al.* 1995, Yeung *et al.* 1998b, Yeung *et al.* 1998a, Lubinski *et al.* 1999]. The still limited availability of 3D ultrasound scanners makes the processing of 2D sequences still an active and important research topic [Jiang and Hall 2009, Touil *et al.* 2010, Korstanje *et al.* 2010, Compas *et al.* 2011]. However, speckle decorrelation due to out-of-plane motion remains a important issue. The development of real-time 3D scans allows researchers to introduce 3D tracking algorithms and to better understand speckle decorrelation. [Chen *et al.* 2005] showed that 3D correlation-based tracking is more accurate than 2D tracking for elastography, since it overcomes this out-of-plane decorrelation. The feasibility of speckle tracking in 3D has been studied in phantoms and simulated volume sequences [Harris *et al.* 2007, Harris *et al.* 2010, Harris *et al.* 2011]. This includes an extensive study on speckle decorrelation and understanding the dependence of the tracking performance on the temporal and spatial resolution of the images. Currently only few 3D speckle tracking studies have been carried out on human volumetric data, namely on the heart [Crosby *et al.* 2009, Byram *et al.* 2010, Compas *et al.* 2012] and on the liver [Harris *et al.* 2010, Lediju *et al.* 2010]. The 3D motion field can also be estimated from biplane sequences [Yoshikawa *et al.* 2005], or in real-time from 1D RF signals acquired during HIFU treatments [Pernot *et al.* 2003, Pernot *et al.* 2004, Marquet *et al.* 2006]. The latter is based on receiving several 1D RF echoes from different subapertures of the HIFU device. The axial displacement is then estimated for each subaperture on successive RF signals, and corresponds to the 3D displacement vector along the beam axis of the corresponding subaperture.

The goal of speckle tracking is to perform the 3D displacement estimation in real-time. Real-time 2D elastography, based on 2D block-matching, is already commercially available [Hsu *et al.* 2005, Pinton *et al.* 2006]. Another fast 2D solution requires the use of parallel computing [Jiang and Hall 2007]. While being more robust against speckle decorrelation, 3D correlation-based speckle tracking is computationally expensive. However, the high computational cost can be reduced by introducing better optimization schemes, such as ICM [Yeung *et al.* 1998b], conjugate GD [Yeung *et al.* 1998a], dynamic programming [Chen *et al.* 2005, Jiang and Hall 2009], together with customized hardware solutions [Chen *et al.* 2005, Harris *et al.* 2007, Kuo and von Ramm 2008]. Speed improvements were also re-

ported in [Kuo and von Ramm 2008] by performing the tracking calculations in the native spherical coordinate system of the 3D data and parallelizing the computations.

2.2.3 Intensity-based registration

Only few articles focus on the registration of ultrasound time sequences. Most consider only two images, eluding the tracking problem, and therefore they are not reviewed. Tracking applications include cardiac motion estimation [Ledesma-Carbayo *et al.* 2001, Ledesma-Carbayo *et al.* 2005, Shekhar and Zagrodsky 2002, Heyde *et al.* 2012], cardiac strain estimation (stress echocardiography) [Shekhar *et al.* 2004, Elen *et al.* 2008], carotid artery examinations [Salcudean *et al.* 1999, Metz *et al.* 2011], brain deformations during neuro-surgical interventions [Penec *et al.* 2003], breast cancer diagnosis [Chen *et al.* 2006], and respiratory motion in the abdomen [Nakamoto *et al.* 2007, Wein *et al.* 2008].

Image registration methods can be divided into direct methods, based on measuring the similarity between pixel or voxel intensities, and indirect approaches, based on image features (e.g. points, edges, surfaces). In this Section, I discuss intensity-based methods, leaving feature-based approaches to Section 2.2.4. Intensity-based methods use the image content of the region of interest, operating directly on the image gray values, without data reduction or segmentation. However, they suffer from high computational costs, especially in the 3D case. Registration algorithms can be grouped according to the three key aspects described in Section 2.1.

Matching criteria

Registration methods based on voxel-wise intensity similarity can be divided into correlation methods [Chen *et al.* 1991, Salcudean *et al.* 1999, Nakamoto *et al.* 2007, Wein *et al.* 2008], intensity difference methods, e.g. SSD [Penec *et al.* 2003, Ledesma-Carbayo *et al.* 2001, Ledesma-Carbayo *et al.* 2005, Chen *et al.* 2006, Heyde *et al.* 2012], and entropy-based similarity measures, e.g. MI [Krucker *et al.* 2002, Shekhar and Zagrodsky 2002, Shekhar *et al.* 2004, Elen *et al.* 2008]. In addition, the region to be registered can be selected by prior segmentation [Shekhar and Zagrodsky 2002, Chen *et al.* 2006].

Transformation models

Depending on the mechanical properties of the tissue and the application, tissue motion in ultrasound images was modeled using a translational [Salcudean *et al.* 1999], rigid [Chen *et al.* 1991] or non-rigid transformation. Rigid transformations

were also used when a neglectable amount of non-rigid deformation was expected between frames. In many cases, rigid models are applied locally to image subregions, and the global motion is obtained by stitching [Chen *et al.* 2006] or by fitting a non-linear function, e.g. thin-plate splines [Krucker *et al.* 2002]. A broadly used linear transformation is affine mapping [Shekhar and Zagrodsky 2002, Shekhar *et al.* 2004, Wein *et al.* 2008].

While rigid transformations are computationally relative inexpensive as they deal few parameters, they are hardly applicable for soft tissue motion. In practice, more degrees of freedom are required. A popular choice are free-form deformations [Rueckert *et al.* 1999, Pennec *et al.* 2003], which consist of a regular grid of control points whose displacements are interpolated by B-splines [Ledesma-Carbayo *et al.* 2001, Nakamoto *et al.* 2007, Elen *et al.* 2008, Metz *et al.* 2011, Heyde *et al.* 2012]. Their advantage lies in the local support of the B-splines and the possibility to appropriately select the number of control points. Several methods have been developed which regularize also for temporal smoothness of the deformation [Ledesma-Carbayo *et al.* 2005, Nakamoto *et al.* 2007, Elen *et al.* 2008, Metz *et al.* 2011].

Optimization strategies

Common optimizers are GD strategy [Ledesma-Carbayo *et al.* 2005, Nakamoto *et al.* 2007], adaptive stochastic GD [Metz *et al.* 2011], the direct Powell-Brent search [Wein *et al.* 2008], the simplex Nelder-Mead method [Shekhar and Zagrodsky 2002, Shekhar *et al.* 2004], the iterative LM method [Ledesma-Carbayo *et al.* 2001, Pennec *et al.* 2003], and quasi-Newton methods such as LBFGSB [Elen *et al.* 2008, Heyde *et al.* 2012]. A multiresolution strategy is often applied [Ledesma-Carbayo *et al.* 2001, Ledesma-Carbayo *et al.* 2005, Nakamoto *et al.* 2007, Metz *et al.* 2011] to improve the registration performance, in terms of accuracy, speed and robustness.

For the registration of images in ultrasound sequences, two different strategies are employed. The first strategy is the pairwise or frame-to-frame method. It can be applied to consecutive pairs of images [Ledesma-Carbayo *et al.* 2001, Krucker *et al.* 2002, Shekhar *et al.* 2004, Heyde *et al.* 2012]. The displacement field with respect to the reference frame is then calculated by sequential composition of the transformations. It suffers from error accumulation and hence is generally only suitable for short time sequences. Such systematic errors can be reduced by considering a spatio-temporal registration model [Ledesma-Carbayo *et al.* 2005] or an non-iterative pairwise strategy, where all images are directly registered to a selected reference image of the sequence [Pennec *et al.* 2003, Nakamoto *et al.* 2007, Elen *et al.* 2008]. In this case, the reference image must be of high quality such that

only this reference and an initialization from the previous transformation are necessary for registering the whole sequence. A different pairwise approach considers the first frame of the sequence as moving image to deform with respect to the following images [Chen *et al.* 2006]. The second strategy is the groupwise registration [Metz *et al.* 2011] (`elastix`¹), which uses an implicit reference frame created from all images, and therefore avoids any bias from selecting the reference image. Its main limitation is that such a registration framework has a higher computational complexity than the pairwise approach and becomes computationally very expensive for long sequences. In particular, for any real-time application, the strategy needs to change, as updating the reference image for each incoming frame is impractical.

Table 2.3 provides an overview of the discussed intensity-based registration algorithms for ultrasound time sequences.

2.2.4 Feature tracking

Feature tracking in ultrasound sequences is an active research area. Clinical applications focus especially on the left ventricle (LV) motion and the deformation of the myocardial borders [Herlin and Ayache 1992, Chalana *et al.* 1996, Tseng *et al.* 1996, Papademetris *et al.* 1999, Malassiotis and Srinivasan 1999, Mignotte *et al.* 2001, Jacob *et al.* 2002, Comaniciu *et al.* 2004, Yang *et al.* 2008, Orderud *et al.* 2007, Nascimento and Marques 2008, Häme *et al.* 2012]. Other applications include liver motion estimation [Zhang *et al.* 2010, Schneider *et al.* 2012], vessel tracking [Das and Banerjee 2004, Guerrero *et al.* 2007], pelvic structure response to muscle activities [Peng *et al.* 2006], boundary tracking of the anal canal [Xiao *et al.* 2007], and tongue tracking in speech sequences [Akgul *et al.* 2000, Li *et al.* 2005, Tang and Hamarneh 2010].

The main purpose in identifying features in a sequence of moving images is to determine image structures which can reliably be tracked. Rather than inspecting the entire image, as done for example in OF, feature tracking focuses only on these image structures. In the case of ultrasound images, used structures are the interfaces between different tissues or organs, like the walls of the heart chambers and vessels. Feature tracking in ultrasound images is especially challenging due to weak echoes, poor signal-to-noise ratio (SNR), signal dropouts, artifacts and noise. Additionally, ultrasound images do not necessarily contain complete or closed boundaries. To cope with these challenges, several frameworks for feature detection and tracking have been developed. The most successful ones have been those that use prior knowledge of the object shape, motion or image appearance, such as Bayesian approaches [Mignotte *et al.* 2001, Jacob *et al.* 1999, Jacob *et al.* 2002, Guerrero *et al.* 2007, Orderud *et al.* 2007, Angelova *et al.* 2007, Yang *et al.* 2008, Nascimento and

¹<http://elastix.isi.uu.nl/>

Marques 2008, Zhang *et al.* 2010], active contours [Malassiotis and Strintzis 1999, Xiao *et al.* 2007] and level sets [Håme *et al.* 2012, Noble *et al.* 2011]. The task of tracking the motion of structures in ultrasound sequences can be divided into two parts, which will be described next. Firstly, the features to track need to be defined in terms of their representation, and detect them in the images. Secondly, the tracking problem, i.e. how to track particular features over time, needs to be formulated. Brief reviews on related works in this field are included in [Noble and Boukerroui 2006, Jacob *et al.* 2002, Comaniciu *et al.* 2004, Xiao *et al.* 2007, Nascimento and Marques 2008, Tang and Hamarneh 2010].

Features representation and detection

In ultrasound tracking the most common features are geometrically represented by single points or contours, which describe curve or surface structures.

Point features in 3D volumes are detected based on local extrema of the response of the Laplacian of Gaussian [Schneider *et al.* 2012]. Rotationally invariant feature descriptors are constructed by sparsely sampling the volume on a rectilinear grid at and around the extrema points. These descriptors, tested on images of porcine heart and liver phantoms, are computationally inexpensive and provide sufficient registration accuracy [Schneider *et al.* 2012].

Contour features, representing curves or surfaces on image boundaries, are used in most works. Contours are usually assumed to be smooth and deformable models are often used for determining their position. The deformation of these models should satisfy some problem-specific constraints related to the object to model. The contour representation can be explicit or implicit. In the explicit representation, the contour is defined by a set of points and an interpolation function (e.g. a spline function) between them. In the implicit representation (e.g. level set), the points lie on a spatial grid and the evolving contour is described by some properties of an underlying embedding function.

In 2D images, a very common explicit deformable model is based on active contours, also called snakes. Snakes were introduced by [Kass *et al.* 1988, Cohen and Cohen 1993] and used to detect boundaries in ultrasound images by [Herlin and Ayache 1992, Chalana *et al.* 1996, Malassiotis and Strintzis 1999, Das and Banerjee 2004, Zhang *et al.* 2010], [Akgul *et al.* 2000] (**tonTrack** algorithm) and [Li *et al.* 2005] (**EdgeTrak** algorithm²). Active contours are energy-minimizing curves or surfaces which balance between external energies (to pull the contour toward matching the image features) and internal energies (to keep the contour smooth). The main limitation of snake-based approaches is that they demand good initialization and their accuracy is compromised by high noise level, which makes their usage for ultrasound

²<https://www.eecis.udel.edu/wiki/vims/index.php/Main/EdgeTrak>

images very difficult without a good deformation model. When dealing with 3D images, the contours of the heart were represented in a parametric and differentiable way by using a continuous distance transform neural network (CDTNN) [Tseng *et al.* 1996], or by employing a 3D B-spline surface model [Papademetris *et al.* 1999, Orderud *et al.* 2007].

In contrast to the explicit representation, the implicit one is topologically flexible, allowing the contours to break apart, develop holes, or join without the need for reparametrization. An example of an implicit contour description is the level set method, used for the tracking of echocardiographic volumes in [Häme *et al.* 2012]. The level set method allows the treatment of surfaces and images in any dimension.

Deformable models rely in most cases on the information provided by the object boundary. Several strategies are used for detecting edges in ultrasound images. The gradient can be extracted by computing the magnitude of the spatial image gradient [Herlin and Ayache 1992, Chalana *et al.* 1996, Akgul *et al.* 2000, Li *et al.* 2005, Xiao *et al.* 2007]. A segmentation software platform [Papademetris *et al.* 1998], originally developed for MR images, was used to segment ultrasound cardiac volumes in [Papademetris *et al.* 1999]. A simple edge detector based on variable gradient strength [Rabben *et al.* 2000] is used to determine the position of the strongest edge along each normal [Orderud 2006, Orderud *et al.* 2007]. Boundaries were also extracted from binary masks, which were generated using Otsu's method [Peng *et al.* 2006] or by thresholding noise-filtered images using a histogram-based method [Das and Banerjee 2004]. [Malassiotis and Strintzis 1999] extracted edges by applying the Sobel operator and then used these as input for a generalized Hough transform to find an elliptical approximation of the LV boundary at the first frame of the sequence.

Some manual inputs are often required to obtain the initial contour. These come from manually tracing the boundary of the structure of interest [Tseng *et al.* 1996, Zhang *et al.* 2010, Häme *et al.* 2012] or from manual initialization of a few points on it [Akgul *et al.* 2000, Peng *et al.* 2006] or inside the object to be segmented [Guerrero *et al.* 2007, Angelova *et al.* 2007, Nascimento and Marques 2008]. [Abolmaesumi *et al.* 2000, Guerrero *et al.* 2007] used a probabilistic edge detection function to find contour points along uniformly distributed rays originating from a seed point inside a vessel. In [Nascimento and Marques 2008] the initial estimate of the object boundary from the internal points is based on the shape probabilistic data association filter (SPDAF) method [Nascimento and Marques 2004]. The detection of the edge points is then similar to Active Contours [Blake and Isard 1998, chap. 5]. In details, the predicted contour is represented by equally spaced samples and the intensity transitions are detected along the direction orthogonal to the contour. A Monte Carlo algorithm for extracting contours was also proposed [Angelova *et al.* 2007]. The initial manual tracing of the endocardial and epicardial contours at end of systole and end of diastole is used to train the CDTNN [Tseng *et al.* 1996].

The low SNR and the poor contrast in ultrasound images make edge detection by gradient-based operators a challenge. Depending on the orientation of the tissue and the effects of shadows and artifacts, it is difficult to define a global edge threshold that works well for different image regions and across images [Jacob *et al.* 1999]. Therefore [Tang and Hamarneh 2010] calculated instead the gradients of local phase features [Leung *et al.* 2009], which can be interpreted as a qualitative description of salient image regions, such as edges or ridges, that are invariant to changes in illumination or image contrast. In [Comaniciu *et al.* 2004] the feature detection is based on a combination of spatio-temporal noise reduction filtering (2D least mean squares filter [Evans and Nixon 1996] and integrated backscatter boundary enhancement) and phase congruency [Kovesi 2000]. [Jacob *et al.* 2002] developed a wavelet-based ridge detection algorithm, where the Coifman wavelet series was used to dyadically decompose the intensity profiles. The latter is reconstructed with a best-basis algorithm.

Contour-based features can provide reliable information for the shape of the target and the tracking of its boundary. However, they lack regional information from the image and may lose their stability when the boundary information is not strong enough [Xiao *et al.* 2007]. To deal with noise and unrelated edges in images, [Mignotte *et al.* 2001] proposed a Bayesian segmentation framework for detecting the endocardial boundaries. They considered two classes, blood and muscle, and assigned to each pixel a membership likelihood for these two classes based on the intensity value. A likelihood energy term segments the image in a ML sense [Mignotte *et al.* 2001]. More robust approaches fuse boundary- and region-based information. [Li *et al.* 2005] incorporates in the external energy formulation the edge gradient and the intensity information in local regions around each snake element. [Xiao *et al.* 2007] uses the sum of the image gradient magnitude and the region information, described by a GMM, as external energy.

Matching a deformable model to a given image is defined as an energy minimization problem, where external and internal forces should be balanced. The external forces push the model to match the image features (described above) well. The internal energy is used to describe prior knowledge, like regularity, continuity or shape of the contour. It helps the estimation of the contour position where image edges are noisy, weak or absent. In the simpler methods, internal forces are based on general smoothness and continuity of the curve [Herlin and Ayache 1992, Chalana *et al.* 1996, Malassiotis and Strintzis 1999, Akgul *et al.* 2000, Das and Banerjee 2004, Li *et al.* 2005, Xiao *et al.* 2007, Tang and Hamarneh 2010]. In more robust approaches, *a priori* knowledge about the likely shapes and deformations of the objects is used. The shape prior captures statistically the variability of the expected shape. The shape-space is learned via Principal Component Analysis (PCA) [Malassiotis and Strintzis 1999, Jacob *et al.* 1998, Jacob *et al.* 1999, Jacob *et al.* 2002, Yang *et al.*

2008] or strongly-adapted PCA [Comaniciu *et al.* 2004] in order to constrain the deformation of the model.

Feature Tracking

Several strategies have been proposed for tracking features in ultrasound sequences. First I describe how contours are tracked before summarizing the tracking of other features.

A simple approach, designed for tracking fast moving contours, extracts the contour independently on each frame [Herlin and Ayache 1992, Papademetris *et al.* 1999, Peng *et al.* 2006]. Iterative variations of this strategy use the optimal contour of the previous frame for initialization [Herlin and Ayache 1992, Mignotte *et al.* 2001, Das and Banerjee 2004, Li *et al.* 2005, Xiao *et al.* 2007]. In the level set formulation, the tracking method finds strong edges which are parallel to the boundary of the previous frame, computes the new deformation at these locations, and then propagates the deformation to the rest of the image volume [Häme *et al.* 2012]. In the case of snakes, the iterative procedure is very sensitive to noise and does not guarantee convergence in the presence of large motion or shadows, which are common in ultrasound images [Malassiotis and Strintzis 1999]. Therefore knowledge about the contour dynamics have been learned and exploited to constrain the deformation. Temporal constraints to the motion are included as regularization terms in the energy minimization problem [Chalana *et al.* 1996, Malassiotis and Strintzis 1999, Akgul *et al.* 2000, Mignotte *et al.* 2001, Xiao *et al.* 2007, Tang and Hamarneh 2010].

Explicit transformation models can be applied to the contour representation, such as rigid [Xiao *et al.* 2007], rigid and scaling [Zhang *et al.* 2010], or affine transformations [Tseng *et al.* 1996]. [Guerrero *et al.* 2007] assumed that the vein center moves with constant velocity. In [Mignotte *et al.* 2001] first a local translation is applied to the contour points and then globally refined using a deformable template. [Jacob *et al.* 1998, Jacob *et al.* 1999, Jacob *et al.* 2002] restricted the model to linear transformations. They introduced a second-order autoregressive motion model for describing the temporal properties of the contour in addition to a shape model. More sophisticated approaches differentiate the system dynamics and the shape constraints by considering the global pose of the contour and the local shape deformation independently [Papademetris *et al.* 1999, Comaniciu *et al.* 2004, Orderud *et al.* 2007, Nascimento and Marques 2008]. This separation is intended to ease modeling, since changes in shape are often parametrized differently from the deformation associated with global position, size and orientation [Orderud *et al.* 2007]. [Yang *et al.* 2008] proposed an automatic one-step forward prediction algorithm to generate motion priors by using manifold learning (ISOMAP).

Motion estimation has been computed using the non-parametric variable-bandwidth density-based fusion (VBDF) estimator at different scales [Comaniciu *et al.* 2004]. At each spatial coordinate of a neighborhood, the motion of control points is firstly estimated as the most significant mode at a large scale. Then, the detected mode is tracked across finer scales. At each scale the mode detections is initialized with the converged location from the previous scale by relying on least squares estimates [Comaniciu 2003]. To avoid error accumulation, the motion of the control points is compared in each frame to a model extracted from the first frame.

A graph-based approach was proposed by [Tang and Hamarneh 2010]. Here, the tongue contour tracking is formulated as a unsupervised graph labeling problem. Each vertex of the contour is labeled with a displacement vector describing its motion. The final displacement labels are those minimizing a multilabel MRF energy function composed of data likelihood and spatio-temporal regularization terms.

In a Bayesian framework, boundary tracking is usually formulated as an estimation of the *posterior* probability based on the detected contours in all past image frames. The model parameters are estimated at each frame by alternating prediction and updating steps. Popular techniques include Kalman filtering (assuming Gaussian distributions) [Malassiotis and Strintzis 1999, Jacob *et al.* 1999, Jacob *et al.* 2002, Guerrero *et al.* 2007, Orderud *et al.* 2007, Yang *et al.* 2008], particle filtering [Angelova *et al.* 2007, Yang *et al.* 2008], conditional density propagation tracking [Zhang *et al.* 2010], and multiple model data association tracking (based on Kalman filtering) [Nascimento and Marques 2008]. The statistical nature of the Bayesian approach makes the tracking process robust against noise and fast enough to achieve real-time capability. However, Bayesian tracking often requires training data.

[Schneider *et al.* 2012] suggested to match point features in two steps. Firstly to find a rough correspondence, a rigid transformation is assumed and points are matched symmetrically by minimizing the Euclidean distance of their feature descriptor vectors. Secondly, the RANSAC algorithm is applied to eliminate outliers and increase robustness of the tracking. The registration of consecutive images follows a group-wise strategy [Wachinger *et al.* 2008].

Many of the developed feature tracking algorithms are real-time [Comaniciu *et al.* 2004, Guerrero *et al.* 2007, Orderud *et al.* 2007, Zhang *et al.* 2010, Schneider *et al.* 2012]. Relevant works on ultrasound feature tracking are listed in Table 2.4.

2.2.5 Hybrid image registration

Intensity-based methods are popular in registration, but they are often insufficient for tracking ultrasound time sequences. Although they operate directly on the image gray values, without data reduction or feature extraction, they are sensitive

to intensity variations and have problems with matching fine structures. In addition the low SNR poses a challenge. In 2D sequences topological changes may occur as anatomical structures move in the out-of-plane direction. Feature-based methods are potentially fast, but they often require *a priori* information and good initialization. Artifacts, such as shadows, might be the cause of feature disappearance. In recent years, researchers have been focusing their attention in developing so-called hybrid methods, which combine intensity- and feature-based registration approaches. The main advantage of hybrid methods is that they use complementary information derived from the images in order to increase the robustness and accuracy of the tracking framework. Publications based on hybrid ultrasound tracking are listed in Table 2.5 and briefly described next. Ultrasound tracking performance can be generally improved by using a hybrid approach.

Attribute vector

[Foroughi and Abolmaesumi 2005a, Foroughi and Abolmaesumi 2005b] proposed a variation to the Hierarchical Attribute Matching Mechanism for Elastic Registration (HAMMER) [Shen 2004]. This method was tested on 3D sequences of the liver [Foroughi and Abolmaesumi 2005b, Foroughi *et al.* 2006b, Foroughi *et al.* 2006a] and 2D liver and carotid artery images [Leung *et al.* 2009]. The method starts with the extraction of image features, so-called attributed vectors, which are employed for establishing correspondence during tracking and are computed at each image pixel. The attributes are the intensity of the pixel, the magnitude of the gradient, the variance of the magnitudes of gradients and the second order derivative, i.e. the Laplacian of the Gaussian (LoG) [Foroughi and Abolmaesumi 2005a, Foroughi and Abolmaesumi 2005b]. The attribute vector is the set of the normalized attributes at two scales. In [Foroughi *et al.* 2006a, Leung *et al.* 2009] the attribute vector is reduced to three elements: voxel intensity, gradient magnitude and LoG. The intensity guarantees that bright voxels will not be registered to dark ones and vice versa. The gradient magnitude and LoG emphasize boundaries.

The feature-matching is performed on a set of leading points, which are selected based on having a high importance. The point importance is calculated for each attribute vector as a weighted sum of its components [Foroughi and Abolmaesumi 2005a]. The best match between leading points is computed by minimizing an energy function based on the similarity of the attribute vectors. However, due to shadowing effects in ultrasound images, it is possible that a feature disappears. To counter this phenomenon, a threshold for accepting matches of leading points is introduced. Displacements are calculated only from the leading points correspondences. Gaussian interpolation of these displacements provides a dense displacement field [Foroughi and Abolmaesumi 2005b]. This algorithm does not require boundaries extraction nor a numerical optimization procedure and hence is fast. However, with a current

registration speed of 5.5 frames per second (2D images [Leung *et al.* 2009]), the performance is still far from real-time.

Phase and intensity

The local phase (LP) feature was developed by [Morrone and Owens 1987]. LP can be seen as an amplitude-weighted phase of a bandpass filtered Fourier component of a signal [Mellor and Brady 2004] and is estimated in 2D images with the monogenic signal, which is an isotropic extension of the analytic signal [Felsberg and Sommer 2000, Woo *et al.* 2009, Cifor *et al.* 2012]. Theoretically, LP measures are invariant to image brightness and contrast, and thus suitable for ultrasound images [Mulet-Parada and Noble 2000, Cifor *et al.* 2012, Woo *et al.* 2009]. Yet the phase estimation depends on the 2D bandpass filter, which shows a more complex behavior than the 1D case and can introduce blurring in the images. This blurring is not affine invariant and therefore limits registration accuracy [Mellor and Brady 2004].

[Woo *et al.* 2009] incorporated intensity and LP information in the data term. The tracking problem is defined in a variational framework (see Sec. 2.2.1) to find the dense displacement field while enforcing a diffeomorphic transformation between the images. They also compared different similarity measures, such as SSD, NCC and MI.

[Cifor *et al.* 2012] proposed a hybrid feature-based Log-Demons registration method. The registration is based on multichannel Log-Demons [Vercauteren *et al.* 2008]. The channels represent the features (intensity, LP and phase congruency) [Kovesi 2000] and demons forces are calculated for each channel. Additionally, regional spatial correspondences are determined by a block-matching scheme based on the squared distance of the selected features. The deformation field is obtained from a weighted sum of the voxel-based updates for each image feature and the regional correspondences. Evaluated on 2D images, this method showed robustness and accuracy in the presence of shadows, topological changes and variable visibility of structures of interest.

Speckle and boundaries

Speckle and contour tracking are very common methods used for ultrasound tracking. However, both have their weakness. Speckle tracking suffers from the instability of the speckle patterns, while contours are often difficult to segment in noisy ultrasound images. A combination of the two methods could improve the robustness and accuracy of the tracker.

[Pirat *et al.* 2008] used Velocity Vector Imaging (VVI)³ on echocardiographic sequences of dogs. This tracking software incorporates speckle and endocardial border tracking. Off-line processing and required manual delineation of the initial LV contour limit this product.

To achieve robust ultrasound tracking, multiple information, including speckle patterns, boundary detection and motion prediction have been fused in [Wang *et al.* 2010] to handle noisy and missing data. This learning-based method automatically estimates the 3D displacements of the myocardium. After automatic detection of the LV boundaries using Marginal Space Learning [Zheng *et al.* 2008], the dense motion tracking employs a multiple cues Bayesian framework. There the likelihood term is computed from both cues, the boundary detection and the speckle block-matching, to maximize accuracy and robustness.

2.3 Validation

A registration method can be accepted only after a careful and appropriate validation. This is not a trivial task and has been discussed in several review papers [Maintz and Viergever 1998, Hill *et al.* 2001, Zitova and Flusser 2003]. Generic validation criteria for medical image processing are also described in [Buvat *et al.* 1999, Jannin *et al.* 2006] and adapted for ultrasound tracking by [Leung *et al.* 2009].

A validation framework should indicate, and possibly quantify, the precision and accuracy, robustness, reliability, computational complexity and clinical use of a specific tracking approach [Maintz and Viergever 1998]. Validation is also fundamental to compare different methods.

Precision measures the repeatability or reproducibility of the tracking results for the same or very similar input. Tests include tracking of the same data while using different initializations of the method and tracking data with similar, but changed noise patterns. The standard deviation or variance of the results are used as a measure for imprecision. The property of most concern for the clinicians is accuracy [Maintz and Viergever 1998]. Ideally it measures the closeness of the tracking results and the actual (true) values. Accuracy can be assessed in a qualitative or quantitative way. Quantitative accuracy can only be established if the true spatial correspondences are known (ground truth) or can be estimated in comparison to the performance of an accepted method (gold standard). Yet, in real ultrasound sequences a ground truth is rarely available and a gold standard by manual annotations is time consuming and not without error. Hence the first assessment is generally done in a qualitative

³syngo[®] Velocity Vector Imaging technology, Siemens Medical Solutions, Ultrasound Division, Mountain View, California

manner, e.g. by visual inspection [Fitzpatrick *et al.* 1998]. This method relies on experts to judge the results, ideally in a blinded observer study.

In many cases the evaluation is done by considering the improvement of the similarity between the source and target images due to registration. However, this does not constitute an assessment of the registration performance, as there is no guarantee that the global optimum of the similarity measure provides the maximum registration accuracy.

Synthetic images and simulated deformation fields applied to real images have the advantage that the applied displacements are known *a priori*. A drawback can be limited realism due to simplistic or biased simulations. Phantom studies are also very common as they rely on realistic image formation. However, compression and rotation causes non-rigid deformations that are not always easy to calibrate and the image content might be unrealistic.

Quantitative measures of the registration accuracy are calculated over regions of interest (obtained after segmentation), contours, or points. Gold standard annotation is preferably performed by multiple expert observers to enable assessment of interobserver variability and improve annotation accuracy. Assessment of region tracking is based on region overlap ratio, volume difference [Pennec *et al.* 2003] or the Hammoude distance [Nascimento and Marques 2008]. Often performance is only assessed for a subset of images or in specific motion positions, e.g. end of systole/diastole for cardiac sequences, or end exhalation/inhalation for respiratory motion. The mean registration error (MRE) is often calculated from the difference between the resulting and the ground truth displacement of a set of points. This difference is formulated as the mean geometric (Euclidean) distance (MED) between points, centroids, landmarks, or contour points. Common error measures are the mean squared error (MSE), the root mean square error (RMSE) and mean absolute difference (MAD). In case of contours, the Hausdorff distance [Nascimento and Marques 2008] or the point-to-surface distance [Orderud *et al.* 2007, Yang *et al.* 2008, Wang *et al.* 2010] is used. Statistical tests include linear regression [Jacob *et al.* 2002, Veronesi *et al.* 2006], the Bland-Altman test [Jacob *et al.* 2002, Veronesi *et al.* 2006, Orderud *et al.* 2007] and the Pearson's test [Chalana *et al.* 1996, Stoitsis *et al.* 2005] for comparing the resulting segmented regions with the (average) reference ones.

To evaluate the reproducibility of stochastic tracking methods, [Zhang *et al.* 2010] compared several runs of the algorithm on the same dataset. The variance of the results represents then the imprecision of the method.

Apart from accuracy, many researchers propose to measure robustness by the percentage of registrations achieving a certain accuracy [Cohen and Dinstein 2002, Mignotte *et al.* 2001, Demi *et al.* 2008].

The evaluation strategies and results of the main ultrasound tracking contributions, considered in this Chapter, are listed in Tables 2.1 to 2.5.

2.4 Conclusion

Ultrasound imaging is a widely available medical image modality and, thanks to its high temporal resolution, is suitable for temporal analysis, such as motion estimation of the organs under investigation. In this Chapter, I have presented and discussed the main ultrasound tracking strategies and listed the most important contributions to the topic.

Ultrasound images are often characterized by low SNR and affected by noise, shadows and artifacts, which limit the performance of straight-forward tracking algorithms. Therefore new approaches need to be found. There is a rising interest in using complementary information and incorporating *a priori* knowledge. The former is represented by hybrid methods, which consider in the data term the contribution of different features, e.g. intensity, phase, gradient, to overcome the limitations of single cues. *A priori* information is used for feature detection (e.g. shape model) as well as for transformation estimation (e.g. motion model).

An advantage of ultrasound imaging is the high frame rate, also for 3D acquisitions. Common frame rates range from 15 to 70 Hz for 2D images and from 5 to 15 Hz for consecutive 3D volumes. For such temporal resolution, the motion between frames is generally relatively small. In this case the local motion could be approximated by linear deformations. However spatio-temporal regularization has proved to be necessary, in order to avoid error accumulation.

Many studies have been conducted on artificial or phantom data, while only a fraction of algorithms have been successfully tested on real clinical data. The research community still lacks a publicly available database of real ultrasound sequences with gold standard information and a common validation framework. These are both critical for comparing different methods and establishing precisely the state-of-the-art performance of ultrasound tracking. In addition, routine clinical usage of ultrasound tracking requires extensive validation on real patient data, which is nevertheless still missing in many contributions.

Ultrasound tracking of the liver was performed by intensity-based methods in most of the cases, see Tables 2.1 to 2.5. Yet, quantitative evaluation of tracking the human liver under free breathing was reported only by [Harris *et al.* 2010, Lediju *et al.* 2010, Lediju Bell *et al.* 2012, Foroughi *et al.* 2006b, Foroughi *et al.* 2006a] for 3D sequences and by [Zhang *et al.* 2010, Cifor *et al.* 2012] for 2D. Intensity-based and hybrid approaches achieved good accuracy (approximately 1.4 mm mean tracking error [Harris *et al.* 2010] and 91% mean overlap ratio [Cifor *et al.* 2012]), yet they

were tested for sequences of short duration (less than a minute) and miss real-time performance. The latter was only achieved by the feature-based approach proposed by [Zhang *et al.* 2010].

For ultrasound tracking to be clinically useful it should be accurate, robust, automatic (or with minimal human intervention) and real-time. The current state-of-the-art accuracy for ultrasound tracking is in the range of 1-2 mm. A critical aspect of tracking is error accumulation and robustness over long sequences. More effort should be devoted to solve this problem and more robust and less noise sensitive algorithms need to be developed. Real-time performance is required for applications such as image guided interventions and therapy. While several contributions are already real-time, most are not. Speed improvements should not only rely on hardware acceleration, but also try to reduce the computational burden of the method.

In summary, ultrasound tracking is an active and promising field of research, with some methods having already successfully made the transition to clinical practice, while most still need improvements in accuracy, robustness and speed as well as proper validation to be clinically applicable. Advancement and availability of 4D ultrasound scanners will shape tracking developments in the near future.

In this Chapter, I listed the main components (Section 2.1) and the main contributions (Section 2.2) in ultrasound tracking. In the following Chapter, I will describe the ultrasound sequences of the liver acquired from healthy volunteers under free breathing, and the tracking validation scheme. I will introduce two intensity-based approaches that were developed for tracking sequences of the liver and showed promising results, namely affine registration and a scale-adapted block-matching algorithm.

Table 2.1: Summary of the main contributions in **optical flow** (Sec. 2.2.1), sorted by year of appearance, stating reference, image modality (**Im.**), key registration components, validation framework and clinical application (**Appl.**).

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Baraldi [1996]	2D	(wL)Sq (dI/dt)	SmT(t)	-	QN(*)	MVE \pm SD	0.3 \pm 0.4pix, 0.1 \pm 0.2rad	1	Synth	-
						QL	1.6 \pm 1.6pix, 0.9 \pm 0.8rad	1	Sim	Heart
						-	-	1(11)	Human	
Meunier [1998]	3D	LSq (dI/dt)	Aff, SmT(t)	-	QN	MVE	0.0%(transl), 0.1%(AXrot), 1.8%(Lrot), 0.4%(def)	3	Synth	-
Mikic [1998]	2D	SSD (I)	SmT(+t)	GS	QN	MAD \pm SD	1.5 \pm 0.3mm 1.7 \pm 0.6mm 1.8 \pm 0.3mm	3 3 2	Human	LV MV AR
						QL	-	11(250tot)	Human	Tongue
						-	-			
Akgul [1998]	2D	En (I,G)	SmT(+t)	DP	QL	-	-	11(250tot)	Human	Tongue
Meairs [1999]	3D	SSD (I)	SmT(+t)	GN	QN	TestApp	-	45(16-25)	Human	CA
Mulet-Parada [2000]	2D	FA (Ph)	SmT(t)	-	QL	-	-	1(32)	Human	Heart
Boukerroui [2003]	2D	ML (I)	Transl, SmT(t)	GS	QL	-	-	1(100)	Human	Heart
Pratikakis [2003]	3D	SQE (dI/dt)	SmT	MG	QN(*)	MAnGE \pm SD	13.8 \pm 24.0 $^\circ$	1(2)	Sim	Brain
						MSE	9.7vox \sim 35vox	1(2)	Aphan	
Pellot-Barakat [2004]	2D	SAD (I)	SmT(t)	ICM	QN(*)	TestApp	-	1(73) 2	Phan Human	- Breast
Stoitsis [2005]	2D	NNC, SSD (I)	Transl, SmT(t)	-	QN(*)	Pearson	0.3(0.7) AX(R)displ 0.2(0.7) AX(R)vel	10(75)	Human	CA
Duan [2005, 2006,2007]	3D	NCC (I)	Transl, TPS	-	QL QN	-	-	1(16)	Human	Heart
						MME (MAnGE)	\sim 0mm (0 $^\circ$)transl, <0.8mm (25 $^\circ$)rot, <2mm (105 $^\circ$)def	3	Synth	
Veronesi [2006]	3D	SSD (dI/dt)	Transl	-	QN	LReg BA	r=0.99p<0.05 -0.2mm bias	10(4hb)	Human	LV
Demi [2008]	2D	ErF ($I, dI/dt$)	SmT(t)	-	QN(*)	Fail	10%	15(50)	Human	CA
Tavakoli [2008]	2D	wLSq (dI/dt)	Aff(+t), SmT(t)	GS	QN(*)	MAnGE(var)	2.6(1.4) $^\circ$	1	Synth	Heart
						MME(var)	3.7(1.1)%	1	Hphan	
Fehrenbach [2010]	2D	SSD (I)	Aff, gTransl	GN	QN QL	MAE \pm SD	0.4 \pm 0.1pix	1(100)	Synth	-
						-	-	1(750)	Animal	Liver
Mukherjee [2011,2012]	3D	L_2 (I)	SmT(+t)	QP	QL QN(*)	-	-	28(30-50)	Human	Heart
						MAnGE \pm SD	<0.1 \pm 0.7 $^\circ$ (transl), <1.5 \pm 7.9 $^\circ$ (rot), <2.4 \pm 0.9 $^\circ$ (def) <20.1 \pm 29.3 $^\circ$ (transl)	3 1	Sim Phan	

^a **Matching criteria.** (w)LSq: (weighted) least squares, L_i : L_i norm, En: energy function, NCC: normalized cross correlation, SSD: sum or squared differences, MI: mutual information, FA: feature asymmetry, ML: maximum likelihood estimator, SQE: semi-quadratic estimator, ErF: error function. dI/dt : image flow, I: intensity, Ph: phase, G: gradient.

^b **Transformation Models.** (g)Transl: (global) translation, Aff: affine, SmT: smoothing/regularization term, TPS: thin-plate spline. (t): temporal, (+t): spatio-temporal

^c **Optimization strategy.** GS: Gauss-Seidel, GN: Gauss-Newton, DP: dynamic programming, QP: quadratic programming, ICM: iterative conditional modes, MG: multiresolution/multigrid minimization

^d **Evaluation.** QN: quantitative, QL: qualitative. (*): comparative study

^e MRE: mean registration error, MVE: mean velocity error, MAD: mean absolute difference, MAE: mean absolute error, MSE: mean squared error, MME: mean magnitude error, MAnGE: mean angular error, SD: standard deviation, var: variance, Max: maximum. Fail: percentage of failure cases. Pearson: Pearson's test, BA: Bland-Altman test, LReg: linear regression. TestApp: test on application

^f Values rounded to the first decimal digit. AX: axial, L: lateral, R: radial. transl: translation, rot: rotation, def: deformation/compression, displ: displacement, vel: velocity

^g Number of sequences (number of images per sequence). tot: total number of images, hb: heart beats (cardiac cycles)

^h Human: patient data, Animal: animal *in vivo* data, Phan: controlled motion of tissue-mimicking phantom, A(H)phan: controlled motion of animal (human) cadaver organs/tissues, Sim: simulated deformations applied to a real image/volume, Synth: computer-generated synthetic sequence

ⁱ LV: left ventricle, CA: carotid artery, MV: mitral valve, AR: aortic root

Table 2.2: Summary of the main contributions in **speckle tracking** (Sec. 2.2.2), sorted by year of appearance, stating reference, image modality (**Im.**), key registration components, validation framework and clinical application (**Appl.**).

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Chen[1992, 1994,1995]	2D Bm	(N)CC (I)	Transl	-	QN	MED	~0.2mm (AX), ~1.1mm (L)	4	Phan	Tissues
O'Donnell [1994]	2D RF	NCC (Ph)	Transl	-	QL	-	-	1(21)	Phan	-
Friemel [1995]	2D RF	(N)CC, SAD (I)	Transl	-	QL	-	-	-	Sim	-
Bashford [1996]	3D RF	LMaxSim (I,dl/dt)	Transl	-	QN	MAE	~40%	2(16)	Phan	Blood
Yeung [1998b]	2D Bm	SSD (I)	Transl, SmT(t)	ICM	QN(*)	MSE	1.0pix(transl), 0.7pix(rot), 0.7pix(def), 0.2pix(shear)	1	Phan	-
					QL	-	-	1	Human	Muscle
Yeung [1998a]	2D Bm	SSD (I)	locDef, SmT(+t)	CGD	QN	MSE	0.7-1.2pix (vibr), 0.7pix(rot), 0.6pix (def) 0.1pix(transl), 0.0pix(def)	1	Synth	-
					QL	-	-	1	Human	Muscle
Morsy [1999]	3D RF	NCC (I)	Transl	-	QN	STD	~±0.1mm(AX), 0.3mm(L)	-	Phan	Muscle
						MTE	~0.2mm(L)	-	-	-
Lubinski [1999]	1,2D RF	NCC (I,Ph)	Transl	-	QL	-	-	2	Phan	-
El-Sharkawy [2001]	2D Bm	SAD (I)	Transl	-	QL	-	-	1	Human	Liver
Kaluzynski [2001]	2D RF	NCC (I,Ph)	Transl	-	QN	TestApp	-	1(166)	Phan	Heart
Cohen [2002]	2D Bm	ML (I)	Transl	-	QN(*)	PCC,MSE	83%,0.3pix	1(2)	Synth	-
					QL(*)	TestApp	-	1(200)	Human	Fetus
Revell [2003,2005]	2D RF	NCC (F), ML (I)	Transl, SmT(t)	-	QN(*)	DFD	41, 26, 184 139.5 218	3 2(30) 14	Phan Hphan Human	- Ten
						-	-	2	Phan	Heart
						-	-	-	-	-
Pernot [2003,2004]	1D RF	CC (I)	Transl, 3Dtriang	-	QN	DispE	<1.2mm <6%	1(10s) 1	Synth Phan	-
					QL	TestApp	-	3	Aphan	Liver
Golemati [2003]	2D Bm	NCC (I)	Transl	-	QN	TestApp	-	8(85)	Human	CA
Yoshikawa [2005]	2.5D RF	CC (I)	Transl	-	QN	MTE	0.1-0.5mm	1	Phan	-
					QL	-	-	1	Human	Liver

table 2.2 - (Continue)

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Chen [2005]	3D RF	NCC (I,Ph)	Transl	DP	QN(*)	RMS	0.7 μ m(AX), <15 μ m(L)	1	Phan	-
							7.0 μ m(AX), 25.3 μ m(L), 37.7 μ m(EL)	1(<1s)	Synth	LV
Hsu[2005]	2D Bm	NCC (I)	Transl	-	QL	-	-	1(60s)	Phan	Liver
Yu [2006]	3D RF, Bm	CC (I)	Transl, Rot, axDef	-	QN(*)	MTE	<~3.5vox(def), <~8.5vox(Lrot)	2	Synth	-
						TestApp	-	1	Phan	
					QL(*)	-	-	1	Aphan	LV
								1	Humam	
Marquet [2006]	1D RF	CC (I)	Transl, 3Dtriang	-	QL	-	-	3(30-120s)	Animal	Liver
Pinton [2006]	2D RF	NCC (Ph)	Transl, SmT	PC	QN(*)	MTE \pm SD	<~0.7 \pm 17 μ m	1	Synth	-
					QL(*)	-	-	1	Aphan	Liver
Lin[2007]	2D Bm	ML (I)	Transl, Interp	-	QN(*)	MAngE	<6 $^{\circ}$	2	Synth	-
					QL	-	-	4	Human	Muscle Ten Liver LV
Basarab [2007]	2D Bm	NCC (I)	Transl, Interp	-	QL(*)	-	-	2(2)	Phan	-
Jiang [2007]	2D RF	SSD (I)	Transl, Interp	PC	QN	DispE var	<10 μ m ²	1	Synth	-
					QL	-	-	1	Phan	-
								2(60,54)	Human	Breast Thyr
								1(60)		
Kuo[2008]	3D RF	NCC (I)	Transl, dirF	PC	QN(*)	TrD	4.2 \pm 1.2mm vs 4mm (AZtransl), 9.9 \pm 1.1mm vs 10mm (AXtransl)	3	Phan	-
Jiang [2009]	2D RF	NCC (I)	Transl, SmT	DP	QL(*)	-	-	2(25)	Human	Breast Thyr
								3(19-24)	Phan	Uterus
								1		
Harris [2007,2010]	3D RF	NCC (I)	Transl	-	QN	MTE \pm SD	<0.4mm(transl)	36(37-65)	Phan	Liver
						RMS \pm SD	0.8 \pm 0.5mm(sin)	3(50)		
						MAD \pm SD	1.7 \pm 1.5mm	12(15-50s)	Human	
						1 \pm 0.7mm	6(20s)			
Crosby [2009]	3D Bm	SAD, NCC (I)	SmT	-	QN	TestApp	-	1(21) 3(1s)	Synth Human	LV
Byram [2010]	3D RF 3D Bm	NCC (Ph)	Transl, Interp	-	QN	DispE	<0.2mm	3(32)	Phan	-
						MAD	0.5cm/s(AX), 2.5cm/s(L), 2.5cm/s(EL)	4(1s)	Human	Heart
						TestApp	1cm/s(AX), 3cm/s(L), 2.5cm/s(EL)	1(5s)		
Jia [2010]	3D RF	NCC (I,Ph)	Transl, Interp	DP	QL	-	-	1	Animal	Heart
Touil [2010]	2D Bm	SAD (I)	Transl, locDef, Interp		QN(*)	MaxMTE	1.2mm(transl), 0.3mm(def)	2	Sim	Heart
					QL	-	-	1	Human	
Korstanje [2010]	2D Bm	NCC (I)	Transl, Interp	-	QN	RMS	1.3%	36(100-1000)	Aphan	Ten
						(MTE \pm SD)	(0.1 \pm 0.1mm)	3(100-1000)	Hphan	
							1.0%	2(100-1000)	Human	
							1.6%	2(100-1000)		
						(0.3 \pm 0.6mm)				
Lediju [2010,2012]	3D RF	NCC (Ph)	Transl, Interp	-	QN	RMS	1mm(CM), 2-6mm(RM)	6(~144)	Human	Liver

table 2.2 - (Continue)

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Compas [2011]	2D RF	NCC (I,Ph), EuclD ($I, dI/dt$, curv), NCC(pt)	Transl, SmT+ TPS, RBF	DA	QN	TestApp	-	6	Animal	LV
Compas [2012]	3D RF	NCC (I,Ph), SqD(curv), NCC(pt)	Transl+ Interp, TPS, RBF+ Interp(+t)	DP,	QL	-	-	5	Animal	LV

* RF: radio frequency data, Bm: B-mode images

^a **Matching criteria.** (N)CC: (normalized) cross correlation, SAD: sum of absolute distances, SSD: sum of squared differences, SqD: square of differences, ML: maximum likelihood estimator, LMaxSim: local maxima similarity, EuclD: Euclidean distance. I: intensity, Ph: phase, F: frequency spectrum, PBE: pass-band energy, dI/dt : image flow, curv: curvature, pt: point coordinates

^b **Transformation Models.** Transl: translation, Rot: rotation, Aff: affine, locDef: local deformation, axDef: axial deformation, 3Dtriang: 3D triangulation, SmT: smoothing/regularization term, Interp: interpolation, RBF: radial basis function, TPS: thin-plate spline, dirF: directional filtering. (t): temporal, (+t): spatio-temporal

^c **Optimization strategy.** ICM: iterated conditional models, CGD: conjugated gradient descent, DA: deterministic annealing technique, DP: dynamic programming, PC: parallel computing

^d **Evaluation.** QN: quantitative, QL: qualitative. (*): comparative study

^e MTE: mean tracking error (also called mean displacement error), MSE: mean surface distance, MAnGE: mean angular error, MED: mean Euclidean distance, RMS: root mean squared error, MAD: mean absolute difference, MAE: mean absolute error, DispE: displacement estimated error, TrD: tracked distance versus actual distance. SD: standard deviation, var: variance, Max: maximum. DFD: displaced frame differencing, PCC: percentage of correct classification. TestApp: test on application

^f Values rounded to the first decimal digit. AX: axial, L: lateral, EL: elevation, AZ: azimuth. transl: translation, rot: rotation, def: deformation, sin: sinusoidal motion, CM: cardiac motion, RM: respiratory motion.

^g Number of sequences (number of images per sequence). s: seconds

^h Human: patient data, Phan: controlled motion of tissue-mimicking phantom, A(H)phan: controlled motion of animal(human) cadaver organs/tissues, Sim: simulated deformations applied to a real image/volume, Synth: computer-generated synthetic sequence

ⁱ LV: left ventricle, CA: carotid artery, Ten: tendon, Thyr: thyroid

Table 2.3: Summary of the main contributions in **intensity-based registration** (Sec. 2.2.3), sorted by year of appearance, stating reference, image modality (**Im.**), key registration components, validation framework and clinical application (**Appl.**).

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Chen[1991]	2D	NCC	Rigid	-	QN	MRE	16%	6	Phan	-
Salcudean [1999]	2D	NCC	Transl	-	QL	-	-	1(~30s)	Human	CA
Ledesma -Carbayo [2001]	2D	SSD	Bsp	LM	QN QL	MSE -	1.3mm -	1 2	Sim Human	LV
Krucker [2002]	3D	MI	locTransl, gTPS	-	QN	MED±SD MRE±SD	0.2±0.1mm ~0.6±0.3mm ~0.3±0.2mm	2(20) 2 2	Sim Human Phan	Breast
Shekhar [2002,2004]	3D	MI	Rigid,Aff	NM	QN	MED	2.1mm,4.1mm	5(12-22)	Human	LV
Pennec [2003]	3D	SSD	SoG	LM	QL QN	- MAVD	- 0.1cm ³	1(10) 1(3)	Phan Animal	Brain
Ledesma -Carbayo [2005]	2D	SSD	Bsp(+t)	GD	QN(*) QN	MED TestApp	0.7pix 1.3pix -	3(32) 1(32) 24	Sim Synth Human	LV
Chen[2006]	2D	SSD	locRigid, Interp	-	QN	TestApp	-	100(60)	Human	Breast
Nakamoto [2007]	2D	NCC	Bsp	GD	QN	2D-MED(Max) 3D-MED(Max)	0.4mm(1.0mm) 1.1mm(2.0mm)	2(~14s)	Animal	Liver
Elen[2008]	3D	MI	Bsp(+t)	LBFGSB	QN	MAD±SD	0.7±0.4mm	12(14-20)	Synth	LV
Wein[2008]	2D	LNCC	Aff	PB	QL	-	-	4(~420-950)	Human	Liver

table 2.3 - (Continue)

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Metz[2011]	2D	Var	Bsp(+t)	SGD	QL	-	-	1	Human	CA
Heyde [2012]	2D	SSD	Bsp	LFBGSB	QN(*)	MED TestApp	~2.4pix -	16	Animal	LV

^a **Matching criteria.** (L)NCC: (local) normalized cross correlation, SSD: sum or squared differences, MI: mutual information, Var: variance

^b **Transformation Models.** (loc)Transl: (local) translation, Aff: affine, gTPS: global thin-plate spline, Bsp: B-spline. SoG: sum of Gaussians, Interp: interpolation. (+t): spatio-temporal

^c **Optimization strategy.** LM: Levenberg-Marquardt, (S)GD: (stochastic) gradient descent, NM: Nelder-Mead, LFBGSB: limited memory Broyden-Fletcher-Goldfarb-Shannon optimization routine with simple bounds, PB: Powell-Brent

^d **Evaluation.** QN: quantitative, QL: qualitative evaluation. (*): comparative study

^e MTE: mean tracking error, MAVD: mean absolute volume difference, MSE: mean surface distance, MED: mean Euclidean distance, MAD: mean absolute difference, SD: standard deviation, Max: maximum. TestApp: test on application

^f Values rounded to the first decimal digit

^g Number of sequences (number of images per sequence). s: seconds

^h Human: patient data, Animal: animal *in vivo* data, Phan: controlled motion of tissue-mimicking phantom, Sim: simulated deformations applied to a real image/volume, Synth: computer-generated synthetic sequence

ⁱ LV: left ventricle, CA: carotid artery

Table 2.4: Summary of the main contributions in **feature tracking** (Sec. 2.2.4), sorted by year of appearance, stating reference, image modality (**Im.**), key registration components, validation framework and clinical application (**Appl.**).

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Herlin[1992]	2D	En (G)	SmT	-	QL	-	-	4(38)	Human	Heart
Tseng[1996]	3D	L_2 (pt)	Aff	GN	QN	MAD \pm SD	1.3 ± 1.1 mm	1	Human	Heart
Chalana [1996]	2D	En (GaussG)	SmT(+t)	-	QN QN(*)	MAD \pm SD Pearson	3.6 ± 1.7 mm 0.93	44(25)	Human	Heart
Papademetris [1999]	3D	En (pt)	locDef, gPose, SmT	-	QN	TestApp	-	3(ES-ED)	Animal	Heart
Malassiotis [1999]	2D	En (G)	SmT(+t), Sh.prior	-	QL	-	-	1(95)	Human	LV
Jacob[1999]	2D	ML (Lph,pt)	Aff,LMM, Sh.prior	-	QL(*)	-	-	2(10s)	Human	LV
Akgul[2000]	2D	En (G)	Bsp(+t)	GD	QL	-	-	1(78)	Human	Tongue
Mignotte [2001]	2D	ML (I)	locTransl, gDef,SmT	HGA	QN	PCC	79%	1	Human	Heart
Jacob[2002]	2D	ML (LPH,pt)	Bsp,LMM Sh.prior, DC	-	QN	TestApp LReg BA MCD \pm SD	- $r^2=0.96$ -0.84cm ² bias - 1.3 ± 1.5 cm ² (epiC), 0.2 ± 2.6 cm ² (endoC)	1(2.5s) 9(27tot)	Human	LV
Das [2004]	2D	En (G)	SmT	-	QL	-	-	1(150)	Human	LLA
Comaniciu [2004]	2D	MahalDist (I)	locDef, gPose, Sh.prior, const.v.	-	QN(*)	MSSD \pm SD MAD \pm SD	8.3 ± 14.3 pix 1.7 ± 1.6 pix	32(18-90)	Human	Heart
Li [2005]	2D	En (I,G)	Bsp,SmT	DP	QN	MSD	~0.8mm	3(33-67)	Human	Tongue
Peng [2006]	2D	Otsu (I)	SmT(+t)	-	QN	TestApp	-	3(8,25,11s)	Human	Pelvis
Xiao [2007]	2D	En (I,G)	SmT(+t)	-	QN	MTE	~0.1pix	1(12)	Human	AnC
Guerrero [2007]	2D	SAD (I)	Bsp, const.v.	-	QN	MED \pm SD	~3.8 \pm 2.6pix	2(100)	Phan	Vessel
Orderud [2007]	3D	SSD (I)	locDef, gPose,Bsp	-	QN	BA MPTSD	4.1ml bias 2.7mm	21(~20)	Synth	LV

table 2.4 - (Continue)

Reference	Im.	Match. ^a	Transf. ^b	Opt. ^c	Eval. ^d	Performance		Validation		Appl. ⁱ
						Measure ^e	Value ^f	No. ^g	Type ^h	
Yang [2008]	3D	L_2 (SF)	M.prior, TPS	-	QN(*)	MPTSD (var,Max)	1.3mm (1.1mm,9.8mm)	67(11-25)	Human	LV
Nascimento [2008]	2D	L_2 (I), EuclD (pt)	Transl, Bspl, ModeEl	-	QN(*)	HD MED HamD	20.8pix 4.8pix 0.2pix	2(215,470)	Human	Heart
Zhang [2010]	2D	DenF (I,pt)	Rigid, Scal	-	QN(*)	var	<1.6mm	1(909)	Human	Liver
Tang [2010]	2D	En (LPhG)	Rigid, SmT(+t)	GC	QN(*)	MED±SD	4.5±1.6mm	8(~13s)	Human	Tongue
Håme [2012]	3D	LevelSet (G)	Band mask	-	QN	OE±SD VD±SD MSuD±SD (RMS±SD, Max±SD)	20.4±3.1% 11.1±6.6% 1.1±0.2mm (1.5±0.3mm, 5.1±0.9mm)	10	Human	Heart
Schneider [2012]	3D	EuclD (LoG)	Rigid, sym.match	-	QN	MRE±SD RMS TestApp	0.3±0.1mm, 0.4±0.0° ~0.4mm,0.8° -	1(437) 34(1hb) 15(50-80)	Aphan Human Aphan	Heart Liver

^a **Matching criteria.** En: energy function, L_i : L_i norm, ML: maximum likelihood estimator, Otsu: Otsu's segmentation, SAD: sum of absolute distances, SSD: sum of squared differences, EuclD: Euclidean distance, MahalDist: Mahalanobis distance, DenF: density function [Blake and Isard 1998]. I: intensity, (Gauss)G: (Gaussian of the) gradient, LPh(G): local phase (gradient), LoG: Laplacian of Gaussians, SF: steerable feature [Zheng et al. 2008], pt: point coordinates

^b **Transformation Models.** (loc)Transl: (local) translation, Rot: rotation, Scal: scaling, Aff: affine, (loc,g)Def: (local, global) deformation, gPose: global pose, Bspl: B-spline, TPS: thin-plate spline, sym.match: symmetric match, SmT: smoothing/regularization term, cont.v.: constant velocity, Sh., M.prior: shape, motion prior, LMM: linear motion model, DC: distance constrain, ModeEl: mode elimination. (+t): spatio-temporal

^c **Optimization strategy.** GN: Gauss-Newton, GD: gradient descent, HGA: hybrid genetic algorithm, DP: Dynamic programming, GC: graph-cuts

^d **Evaluation.** QN: quantitative, QL: qualitative. (*): comparative study

^e MRE: mean registration error, MAD: mean absolute difference, MS(S)D: mean sum of squared distances, MED: mean Euclidean distance, MCD: mean contour difference, MPTSD: mean point-to-surface distance, HD: Hausdorff distance, HamD: Hamoude distance, RMS: root mean square, OE: volumetric overlap error, VD: relative absolute volume difference, MSuD: symmetric surface distance, SD: standard deviation, var: variance, Max: maximum. PCC: percentage of correct classification, LReg: linear regression, BA: Bland-Altman test, Pearson: Pearson's test. TestApp: test on application

^f Values rounded to the first decimal digit

^g Number of sequences (number of images per sequence). ES: end of systole, ED: end of diastole, s: second, tot: total number of images for QN evaluation.

^h Human: patient data, Phan: controlled motion of tissue-mimicking phantom, Aphan: controlled motion of animal organs, Synth: computer-generated synthetic sequence

ⁱ LV: left ventricle, LLA: lower limb artery, AnC: anal canal

Table 2.5: Summary of the main contributions in **hybrid registration** (Sec. 2.2.5), sorted by year of appearance, stating reference, image modality (**Im.**), key registration components, validation framework and clinical application (**Appl.**).

Reference	Im.	Match. ^a	Transf. ^b	Eval. ^c	Performance		Validation		Appl. ^h
					Measure ^d	Value ^e	No. ^f	Type ^g	
Foroughi [2005,2006]	3D	L_2 (I,G,LoG)	Transl, GInterp	QN	MRE	~1.4mm ~1.3mm	30(2)	Human Sim	Liver
Pirat [2008]	2D	VVI*(I,G)	VVI*	QN	TestApp	-	7(~50)	Animal	Heart
Leung [2009]	2D	L_2 (I,G,LoG)	Transl, GInterp	QN	MRE	0.9mm 0.1mm	20(20-40) 20(20-40)	Sim Sim	Liver CA
				QL	-	-	10(50)	Human	Liver
				QN	MOvR±SD	93±1.5%	10(50)	Human	CA
Woo [2009]	2D	L_1 (I,LP)	SmT	QN(*) QL(*)	MRE±SD -	3.8±6.8pix -	1(2) 1(2) 1(2)	Synth Animal Human	- Heart
Wang [2010]	3D	ML (I,SF)	Markov Motion	QN QL(*)	MPTSD±SD TestApp	2.7±2.6mm -	264	Human	LV
Cifor [2012]	2D	L_2 (I,LP,PC)	locTransl, SmT	QN(*)	MOvR±SD	90.8±5%	8(22-71)	Human	Liver

- ^a **Matching criteria.** L_i : L_i norm, ML: maximum likelihood estimator. I: intensity, G: gradient, LoG: Laplacian of Gaussians, LP: local phase, SF: steerable feature [Zheng *et al.* 2008], PC: phase congruency
- ^b **Transformation models.** (loc)Transl: (local) translation, GInterp: Gaussian interpolation, SmT = smoothing/regularization term
- * VVI: syngo[®] Velocity Vector Imaging technology, Siemens Medical Solutions, Ultrasound Division, Mountain View, California
- ^c **Evaluation.** QN: quantitative, QL: qualitative. (*): comparative study
- ^d MRE: mean registration error, MOvR: mean overlap ratio, MPTSD: mean point-to-surface distance, SD: standard deviation. TestApp: test on application
- ^e Values rounded to the first decimal digit
- ^f Number of sequences (number of images per sequence)
- ^g Human: patient data, Animal: animal *in vivo* data, Sim: simulated deformations applied on a real image/volume, Synth: computer-generated synthetic sequence
- ^h CA: carotid artery, LV: left ventricle

3

Intensity-based tracking for long sequences

In Chapter 2 a description of the relevant approaches for analyzing 2D sequences of the liver under free breathing has been presented. Examples of these approaches are explained in details in this Chapter. Their tracking performance is evaluated for long 2D ultrasound image sequences, which are first described in Section 3.1. Section 3.2 describes the evaluation scheme used to validate and quantify the tracking accuracy. This evaluation scheme will be adopted also in Chapter 4. I will adapt two intensity-based methods, local affine registration (see Section 3.3) and a novel scale-based block-matching algorithm (see Section 3.4), for the tracking of long sequences. The computational complexity of each strategy is calculated. Finally, I will show and discuss the results of the aforementioned methods.

3.1 Ultrasound sequences

Ultrasound sequences of the liver of 9 volunteers during free breathing were acquired at the Geneva University Hospital [Petrusca *et al.* 2013, Petrusca *et al.* 2011]. An Acuson clinical ultrasound scanner (Antares; Siemens Medical Solutions, Mountain View, CA), modified to be MR-compatible, generated real time second harmonic images (center frequency: 1.8-2.2 MHz). Ultrasound and MR images were simultaneously acquired. The ultrasound images were exported on-the-fly using a frame grabber device [Petrusca *et al.* 2013]. The image sequences consist of 2D slices of the liver acquired at a fixed location (longitudinal or intercostal plane) over 5-10 min, in order to evaluate ultrasound tracking performance for ultrasound- and MR-guided treatment scenarios. More details regarding the simultaneous acquisition of ultrasound and MR images will be provided in Chapter 5. The sequences have a temporal and spatial resolution of 14-25 Hz and 0.3-0.7 mm, respectively. Their lengths are 5 min 21 sec (1 sequence), 5 min 28 sec (7 sequences) and 10 min

08 sec (1 sequence). The number of frames per acquisition ranges between 2650 and 14516.

The main difficulties related to these images are, in addition to a low SNR, small acoustic windows and shadows due to the presence of ribs, and radio frequency (RF) interferences from the MR scanner. The mentioned artifacts are visible in Figure 3.1. During the acquisition of four sequences, few images (0.04% - 0.2%) were skipped by the frame grabber device. These skipped frames never occurred within the first minute of acquisition, i.e. always after at least 1100 frames.

In the rest of this thesis I will use the following notation. 2D ultrasound images, characterized by D pixels, are acquired at a frame rate f Hz, resulting in a temporal sequence of T images $I(t_i)$ at time $t_i = t_0 + i/f$ for $0 \leq i \leq T - 1$. The first frame of each of the 9 sequences is shown in Figure 3.1.

3.2 Evaluation method

The methods were tested for a total of 25 vessels in 9 sequences, see Figure 3.1. I qualitatively assessed by visual inspection the tracking results for all vessels. I quantitatively evaluated the tracking error for the 15 vessels, which the observer was confident to be able to reliably annotate. I randomly selected 10% of the images from the real-time application phase and manually annotated the position (denoted as \bar{P}_v) corresponding to the selected landmark $P_v(t_0)$.

For the annotated frame (\hat{t}), I calculated the tracking error

$$\text{TE}_v(\hat{t}) = \|P_v(\hat{t}) - \bar{P}_v(\hat{t})\|. \quad (3.1)$$

I summarize the results by the mean (MTE), standard deviation (STD) and 95th percentile of all $\text{TE}_v(\hat{t})$, considering all the landmarks as a single distribution. I also computed the mean tracking error for each landmark v (MTE_v) and report the range for the 15 vessels. I included the motion magnitude of the vessels, defined as

$$\|P_v(t_0) - \bar{P}_v(\hat{t})\|. \quad (3.2)$$

I estimated the inter-observer variability of the results. Two additional experts annotated 3% of randomly selected images from the real-time application phase. I then defined as ground truth the mean position over the 3 annotations and calculated the tracking error as before.

The paired Wilcoxon signed-rank test was used for assessing if median results were statistically significantly different at the 0.001 level. The Wilcoxon test was chosen because the distribution of TE_v for all annotated images and vessels was not normal and not transformable to a normal distribution by means of a logarithmic transformation [Siegel 1956].

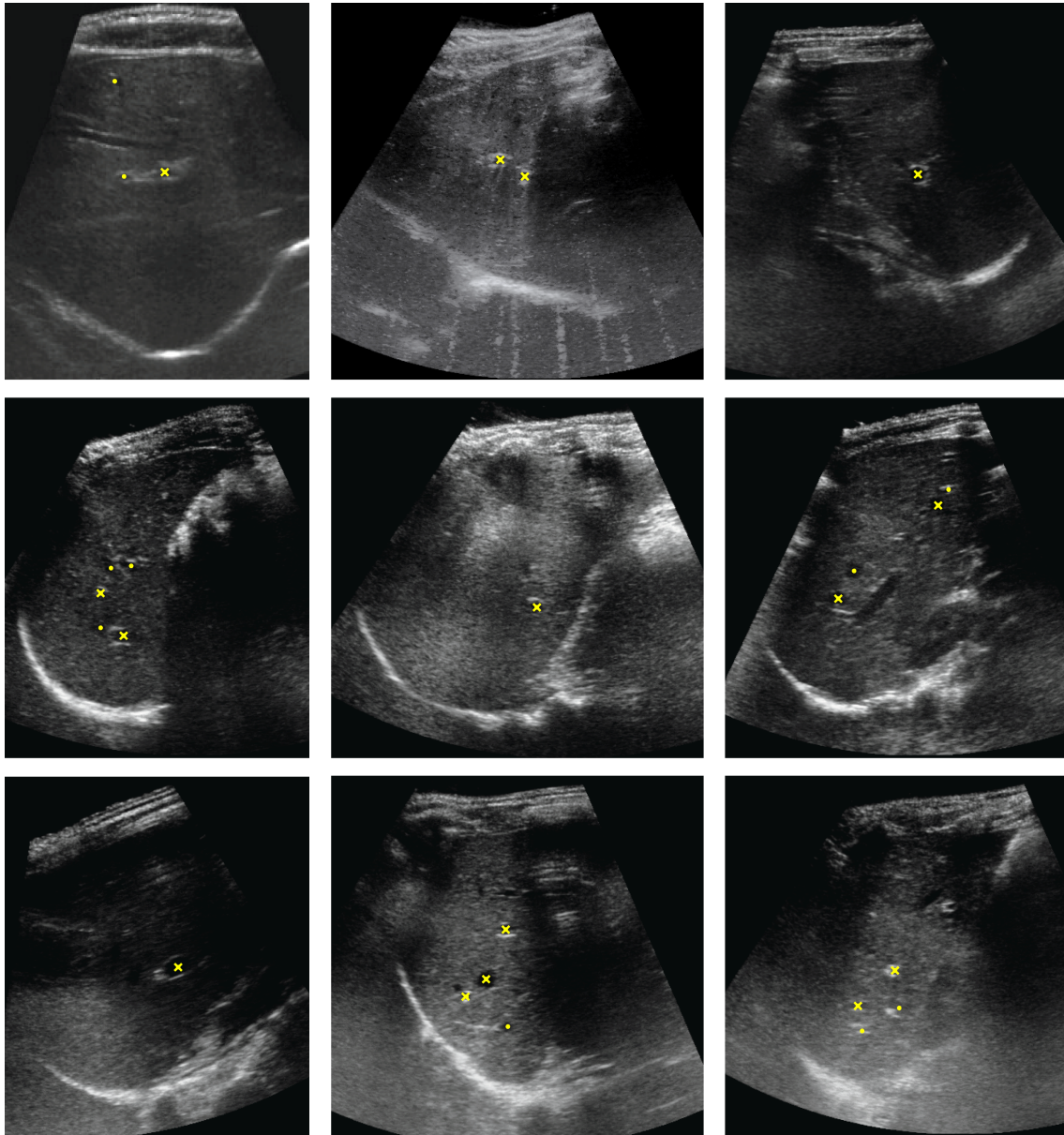


Figure 3.1: First frame $I(t_0)$ of the 9 sequences and manual annotation of the tracked vessel centers $P_v(t_0)$, $v \in [1, \dots, 25]$. Quantitative evaluation was based on the 15 P_v marked by 'x'. Artifacts visible in these images include MR-RF interferences (first row, second image) and small acoustic windows (last row).

3.3 Affine registration

This section describes how features on the ultrasound sequences are tracked by employing affine registration. The method is first described in details and followed by the analysis of its computational complexity. Results of its tracking performance are given in Section 3.5.

In short, the images $I(t_i)$ are registered to the first image $I(t_0)$, in order to obtain spatial correspondence at each time t_i . The position of features to track, e.g. $P_v(t_0)$ for vessel v , are manually selected in $I(t_0)$, see Figure 3.1. The proposed registration optimizes the parameters of an affine transformation with respect to normalized cross correlation (NCC). The transformation is defined over a manually selected region around $P_v(t_0)$.

3.3.1 Registration details

In the 2D domain, the affine transformation $\mathbf{T}(\mathbf{x})$, at position $\mathbf{x} = [x; y]$, is defined by 6 parameters, 4 representing implicitly the combination of rotation, anisotropic scaling and shearing ($a_{m,n}$, $m, n \in [1, 2]$), and 2 translation coefficients (τ_m , $m \in [1, 2]$):

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \quad (3.3)$$

The registration task is then to find these 6 free parameters Θ such that the images are aligned, which is indirectly quantified by an image similarity measure. This can be formulated as the optimal transformation \mathbf{T}^{opt} :

$$\mathbf{T}^{opt} = \operatorname{argmin}_{\mathbf{T} \in \Theta} C(\mathbf{T}), \quad (3.4)$$

where the cost function $C(\mathbf{T})$ measures the dissimilarity between the image region A and the transformed image region $B^{\mathbf{T}}$, i.e. $C(\mathbf{T}_v) = -\rho(A, B^{\mathbf{T}})$, where ρ is a similarity measure. NCC is chosen as similarity measure and for each tracked region v , the NCC is given by:

$$\rho(A_v, B_v^{\mathbf{T}}) = \frac{\sum_{n=1}^W (A_v[n] - \bar{A}_v)(B_v^{\mathbf{T}}[n] - \bar{B}_v^{\mathbf{T}})}{\sqrt{\sum_{n=1}^W (A_v[n] - \bar{A}_v)^2} \sqrt{\sum_{n=1}^W (B_v^{\mathbf{T}}[n] - \bar{B}_v^{\mathbf{T}})^2}}, \quad (3.5)$$

where $A_v[n]$ and $B_v^{\mathbf{T}}[n]$ denote the intensity of A_v and $B_v^{\mathbf{T}}$ at pixel n . A_v is a manually selected region around $P_v(t_0)$ of size $W < D$ pixels, and B_v the corresponding region in the transformed image $I(t_i)^{\mathbf{T}}$ of same size. \bar{A}_v and $\bar{B}_v^{\mathbf{T}}$ are the mean intensity values of A_v and $B_v^{\mathbf{T}}$, respectively.

I use a Gradient Descent (GD) optimizer to solve Eq. (3.4). At each iteration step, GD advances the parameters in the search direction of the gradient. To improve

accuracy and robustness, a multiresolution approach with 2 levels is chosen. In addition, registration is initialized by the result from $I(t_{i-1})$ to $I(t_0)$. This is based on the assumption that the main motion is continuous. Such initialization allows for a faster convergence of the solution, removing the initial major components of misalignment between $I(t_i)$ and $I(t_0)$.

As mentioned in Section 3.1, few frames in the sequences were skipped by the frame grabber device. In such cases the frame-to-frame displacement can be particularly high and the registration can fail. Therefore, the images immediately following the drop out were manually identified and the registration initialized by the results from a similar image in the first breathing cycle.

The registration algorithm was implemented using the Insight Segmentation and Registration Toolkit (ITK) [Ibanez *et al.* 2005].

3.3.2 Computational complexity

For image registration, the computational complexity per image region v depends on I , E , V and R , where:

- I is the average number of iterations (maximum set to 300)
- E is the number of image similarity calculations per iteration and transformation parameter. Assuming numerical gradient calculation, $E = 2$
- V is the number of calculations for the image similarity per registration parameters. Assuming the fixed image region of W pixels, the number of computations required to calculate the NCC are W for the numerator, $2W + 3$ for the denominator and one division (see Eq. (3.5)). The affine transformation is characterized by n^2W calculations, where n is the image dimension ($n = 2$ for 2D). Therefore the number of main computations is given by $V = W + 2W + n^2W = 7W$.
- R is the number of transformation parameters. It is equal to $(n + 1)n$ for a n D affine transformation, i.e. 6 parameters for 2D affine registration

For 2D images, the number of computations required by the affine registration is $84IW$, with $I \ll W$. Therefore the complexity can be approximated by $O(W)$.

3.4 Scale-adaptive block-matching

Block-matching algorithms (BMAs) compute the local displacements from interpolating the translations which provide the best match of image regions in two

consecutive frames. Many BMAs have been proposed (e.g. [Boukerroui *et al.* 2003, Revell *et al.* 2005, Lin *et al.* 2007, Byram *et al.* 2010, Harris *et al.* 2010]), yet their performance has so far only been assessed on relative short sequences (<1 min). Therapy guidance requires the tracking of long sequences, which poses a special challenge for BMA due to its iterative nature. Furthermore the tested sequences suffer from noise, interferences, low SNR and frame drop outs. In order to create a robust framework for feature tracking in long sequences, I propose an algorithm, which combines several block-matching components and includes a novel adaptation of the block size to the feature scale.

The key components of the proposed scale-adaptive block-matching method (SA-BMA) are a novel adaptation of the block size to the feature scale and the new combination of the interpolation function from [Lin *et al.* 2007] and the temporal realignment from [Revell *et al.* 2005].

3.4.1 Block configuration

Traditionally the size of the blocks is chosen empirically [Morsy and Von Ramm 1999, Harris *et al.* 2007] or is equal to the size of the ultrasound speckle element [Kaluzynski *et al.* 2001]. In the presented work, I adapt the block size to the feature size in order to ensure that every block contains a part of the feature, which limits the aperture problem and avoids ambiguous matches due to homogeneous blocks.

As in the previous Section, the position of features to track, e.g. $P_v(t_0)$ for vessel v , are manually selected in the first frame $I(t_0)$, see Figure 3.1. A region of interest $ROI_v(t_0)$ around feature v is selected. Such region is generally smaller than the affine registration region (see Section 3.3.1). BM is performed for each $ROI_v(t_0)$, which covers a $M \times N$ grid of equally sized squares (called blocks) $B_{i,v}$ of size Δb_v with center points $G_{i,v}$, $i \in [1, \dots, MN]$. Size Δb_v is determined as a function of the feature size. As vessel cross sections can be approximated by an ellipse, I search for blob-like features centered at P_v . A scale-space approach (local maxima of a Difference of Gaussian (DoG)) [Lindeberg 1998, Schneider *et al.* 2012] is used to detect the most likely blob in $ROI_v(t_0)$. The size of the blob is derived from the Gaussian kernel size s that minimizes the DoG, where s relates to the minor semi-axis r_v of an ellipse fitted to the vessel cut by $r_v = \sqrt{2}s$. Δb_v is obtained by rounding r_v to the next greater integer value.

3.4.2 Displacement calculation

I compute the motion field in each ROI_v by determining the displacement at $G_{i,v}$ via block-matching, and use weighted interpolation [Lin *et al.* 2007] to obtain the

displacement of P_v . At time step t^* the displacement of $G_{i,v}(t^{ref})$ in the reference frame t^{ref} to $G_{i,v}(t^*)$, denoted as $\mathbf{d}_{G_{i,v}}(t^*)$, is determined by the displacement \mathbf{v} which maximized the NCC between $B_{i,v}(t^{ref})$ and the block from $I(t^*)$ centered at $G_{i,v}(t^{ref}) + \mathbf{v}$, see Eq. (3.5). The values of \mathbf{v} are restricted to cover only a certain search region, which is of size $\Delta b_v + 2\beta_v$, where β_v is determined by the maximum expected frame-to-frame displacement of the vessel of interest.

The reference frame is generally the previous frame ($t^* - 1$). Other strategies for t^{ref} are described in Section 3.4.3. The displacement of the tracked point from t^{ref} to t^* ($\mathbf{d}_v(t^*)$) is deduced from the single block displacements $\mathbf{d}_{G_{i,v}}(t^*)$ by weighted interpolation:

$$\mathbf{d}_v(t^*) = \sum_{\hat{i}} w_{\hat{i}} \mathbf{d}_{G_{i,v}}(t^*), \quad (3.6)$$

where $w_{\hat{i}}$ are the weights and $\hat{i} = \{i | Q(i, t^*) = 1\}$. In details, $Q(i, t^*)$ is the neighborhood filtering mask for ROI_v , which is defined by:

$$Q(i, t^*) = \begin{cases} 1 & \text{for the 9 } G_{i,v}(t^{ref}) \text{ closest to } P_v(t^{ref}) \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

I consider the weights $w_{\hat{i}}$ [Lin *et al.* 2007]:

$$w_{\hat{i}} = 0.5 \frac{1}{D_{\hat{i}}^2 + 1} \frac{1}{\sum_{\hat{i}} \frac{1}{D_{\hat{i}}^2 + 1}} + 0.5 \frac{\alpha_{\hat{i}}}{\sum_{\hat{i}} \alpha_{\hat{i}}}, \quad (3.8)$$

with $D_{\hat{i}}$ the Euclidean distance from $G_{i,v}(t^{ref})$ to $P_v(t^{ref})$, and $\alpha_{\hat{i}} = \sigma_{\hat{i}}^2 / \mu_{\hat{i}}$ the ratio between the variance ($\sigma_{\hat{i}}^2$) and the mean ($\mu_{\hat{i}}$) of the pixel intensities in $B_{i,v}(t^*)$. This interpolation scheme has the advantage that it incorporates regularization (first term) and takes into account the relative image content (second term) [Lin *et al.* 2007]. The position of the tracked point is $P_v(t^*) = P_v(t^{ref}) + \mathbf{d}_v(t^*)$.

3.4.3 Reference frame definition

Block-matching methods can generally only cope with small deformations and appearance changes, as they are based on the translations of local regions. Hence block-matching is applied to temporally consecutive frames (i.e. $t^{ref} = t^* - 1$) for tracking. However, this strategy is subject to error accumulation leading to drift. Such errors are particularly relevant in long sequences. Yet the approximate periodic nature of respiratory motion provides frequently frames which are similar to the initial frame and the block-matching is again applicable for aligning these [Revell *et al.* 2005]. In order to reduce drift, the current image region $ROI_v(t^*)$ is realigned with the initial frame $ROI_v(t_0)$ when the correlation between the two images is high. Errors occur also due to the quantization of $\mathbf{d}_{G_{i,v}}$. To reduce the quantization error,

the reference frame remains unchanged when the current displacement $\mathbf{d}_v(t^*)$ is very small. In details, I introduce the following strategy:

```

if  $\rho(ROI_v(t_0), ROI_v(t^*)) > \theta_{NCC,v}$  then  $t^{ref} = t_0$ 
else if  $\|\mathbf{d}_v(t^*)\| \leq \epsilon_d$  then  $t^{ref} = t_{prev}^{ref}$ 
else  $t^{ref} = t^* - 1$  end

```

where $\rho(ROI_v(t_0), ROI_v(t^*))$ is the NCC between image region $ROI_v(t_0)$ and $ROI_v(t^*)$ (see Eq. (3.5)), $\theta_{NCC,v}$ is the 84th percentile of the NCC values, with respect to $ROI_v(t_0)$, gathered from an initial subset of the sequence, $\mathbf{d}_v(t^*)$ is the displacement of the tracked point in the current frame t^* (see Eq. (3.6)), $\epsilon_d = 0.01$ pixel, and t_{prev}^{ref} denotes t^{ref} from the previous image pair.

3.4.4 Computational complexity

In this thesis, I consider the case of the exhaustive search block-matching. The drawback of this algorithm, also known as full search, is that it is the most computationally expensive BMA [Lin and Tai 1997]. The computational cost can be easily reduced by using faster matching algorithms [Lin and Tai 1997], such as hierarchical search [Lin et al. 2007]. For each image and tracked vessel v , the complexity of the proposed algorithm depends on the following steps:

- The $K = 9$ closest grid points $G_i(t^{ref})$ to $P(t^{ref})$ are selected using K-nearest neighbor (K-NN) search with $KMN\log(MN)$ computations, where MN is the number of all grid points G_i . Hence the complexity is $O(MN\log(MN))$.
- The complexity of the BMA for K grid points is given by the calculation of the NCC at each searching position. In details, the complexity of each NCC calculation is $O(B)$ (see Section 3.3.2), where $B = \Delta b^2$ is the number of pixels per block. The number of iterations in the search area is $(2\beta + 1)^n$. Therefore BMA requires $KnB(2\beta + 1)^n$ computations, with $n = 2$ for 2D images
- The displacement of the vessel is calculated via interpolation of K blocks, see Eq. (3.6). The complexity of the interpolation is dominated by the complexity of calculating α in Eq. (3.8), which is $O(B)$

This results in a total of $9MN\log(MN) + 18B(2\beta + 1)^2 + 9B$ calculations. With $\beta < B$, $B \approx MN$ and $\beta^2 > \log(B)$, the complexity is $O(B\beta^2)$.

3.5 Results

I compared the performance of the affine registration (see Section 3.3) and SA-BMA (see Section 3.4). As baseline BMA, I modified the SA-BMA to have fixed block size of $\Delta b_v = 16$.

I tracked a total of ~ 50000 frames, acquired over a total of ~ 50 min. Δb_v is in the range of $[4, 22]$ pixels and the size of the tracked vessels varies from 2 to 9 mm. β_v varies from 3 to 10 pixels (from 1 to 4 mm), while the total number of grid points MN ranges in $[30, 520]$. Manual initialization of the affine registration because of skipped frames was necessary for a total of 36 images, i.e. $\sim 0.007\%$.

Results for the approaches described in this Chapter, affine registration and SA-BMA, are listed in Table 3.1. The best performance is achieved by affine registration with a MTE of 0.90 mm. The introduction of scale-adapted blocks improved the tracking performance by 35% compared to standard SMA, reducing the MTE from 3.22 mm (BMA) to 2.19 mm (SA-BMA). In Figure 3.3 I illustrate the benefit of the SA-BMA for the worst case of BMA.

Similar results were achieved for the inter-observer data set (Table 3.2). The inter-observer MTE \pm STD is 0.32 ± 0.19 mm and the range of MTE_v over all observers is $[0.16, 0.70]$ mm. Manual annotation is therefore substantially better than any of the proposed method, i.e. MTE is reduced by 64% and 85% with respect to affine registration (MTE=0.90 mm) and SA-BMA (2.18 mm), respectively. Figure 3.2 summarizes the TE results for each of the 15 vessels and the 3 proposed methods (i.e. BMA, SA-BMA and affine registration), by showing the minimum, 25th percentile, median, 95th percentile and maximum of the TE_v distribution for all images annotated by 3 observers. It can be observed that the affine registration clearly provides the lowest errors apart from vessel 4. SA-BMA has lower errors than BMA for all but two vessels (number 2 and 14). BMA has a total failure case (vessel 7 illustrated in Figure 3.3), where the MTE_7 is higher than the mean vessel motion (12.16 vs. 3.00 mm). The median TE_v of BMA and SA-BMA, and of SA-BMA and affine registration were statistically significantly different at the 0.001 level (Wilcoxon signed-rank test).

Table 3.1: Tracking results (in mm) for the different methods w.r.t. manual annotation from one observer (10%, ~ 7500 images). Best results are in bold face.

	MTE \pm STD (95 th TE)	range MTE _v
VesselMotion	5.17 \pm 3.21 (10.59)	[2.81, 11.48]
Affine	0.90 \pm 0.37 (1.51)	[0.34, 1.82]
BMA	3.22 \pm 2.26 (7.24)	[1.25, 12.35]
SA-BMA	2.19 \pm 1.46 (4.90)	[1.20, 5.79]

Table 3.2: Tracking results (in mm) for the different methods w.r.t. manual annotation from three observers (3%, ~ 2500 images). Best results are in bold face.

	MTE \pm STD (95 th TE)	range MTE _v
VesselMotion	5.22 \pm 3.23 (10.57)	[3.00, 11.30]
Affine	0.90 \pm 0.35 (1.45)	[0.35, 1.90]
BMA	3.20 \pm 2.26 (7.17)	[1.26, 12.16]
SA-BMA	2.18 \pm 1.45 (4.83)	[1.22, 5.78]

3.6 Conclusion

In this Chapter, I compared two intensity based frameworks for vessel tracking in long ultrasound sequences.

The methods were evaluated on long 2D sequences of the liver of 9 volunteers under free breathing. Vessel were tracked for 5-10 min. To my knowledge, this is the first evaluation for tracking such long ultrasound sequences.

Using affine registration, I achieved a mean accuracy of 0.90 mm. Such performance improves the state-of-the-art in 2D ultrasound tracking of the human liver (1.6 mm [Zhang *et al.* 2010]). The affine registration is extremely accurate. However, its computational complexity is linear with the size or the region to track. In addition, the number of iterations required by the optimizer might increase with larger displacements (from 50 to 300 iterations). Another drawback of the proposed strategy is that it requires manual reinitialization when frame drop outs occur.

The proposed BMA is computationally faster, as $B, \beta \ll W$, and potentially real-time [Hsu *et al.* 2005, Pinton *et al.* 2006]. This tracking strategy is only partially affected by frame drop outs, as an adequate search area is used and a more sophisticated temporal realignment strategy is introduced. However, the accuracy of BMAs is lower. The failure of standard BMA in long sequences might be due to an inappropriate block size, changes in the image similarity values and error accumulation.

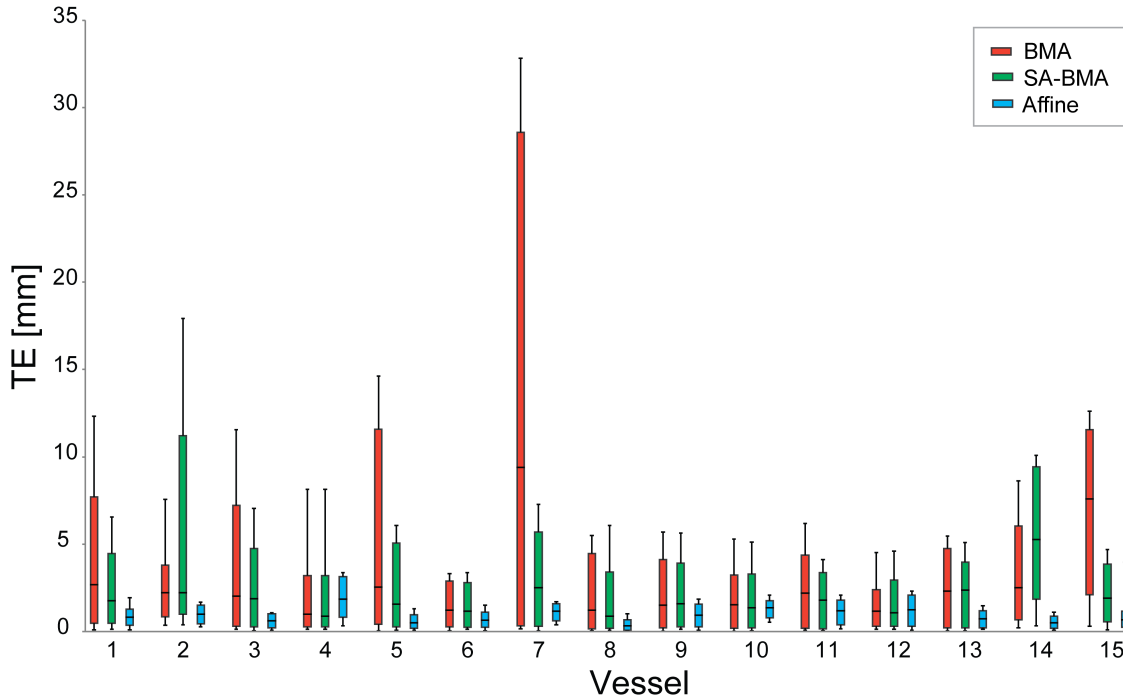


Figure 3.2: Summary of the tracking error distribution (TE) in mm for each of the 15 vessels and 3 methods, BMA (red), SA-BMA (green) and affine registration (blue). The error bars indicate the minimum and maximum error, while the colored bars define the 25th percentile, median and 95th percentile of the TE. Affine registration provides a relative robust performance for all vessels, while BMA (SA-BMA) has several (a few) very poor results (median > 2.5 mm, 95% > 10 mm).

The introduction of scale-adaptive blocks was significant for the improvement of the results. While adaption to the feature size reduces the error caused by ambiguous matches, the use of NCC for measuring the opportunity of temporal realignment can be misleading. Even with adaptation to the individual ultrasound sequence, temporal realignment of the tracking was often too sparse.

In the following Chapter, I propose a learning based approach to achieve accurate, robust and fast tracking. The main idea is to exploit the repetitive nature of respiration and to learn the relationship between image appearance and registration results.

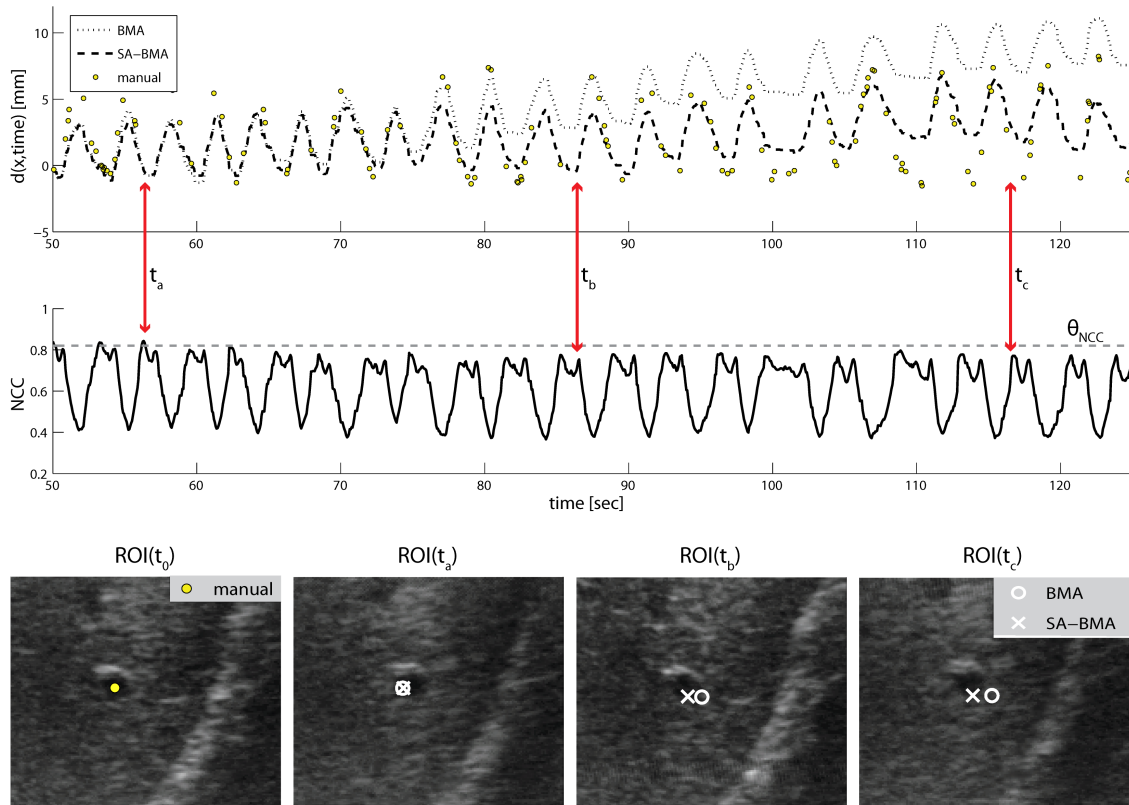


Figure 3.3: Comparison of the block-matching tracking performance for a failure sequence of BMA ($MTE_v=12.4$ mm). (Top) Main motion component of manual annotation and \mathbf{d}_v from 2 methods for a temporal subset. (Middle) Corresponding NCC to first image. (Bottom, left to right) First image with annotation ($P_v(t_0)$), image with tracking results at last realignment (t_a) of (SA-)BMA, at $t_b = t_a + 30$ s and $t_c = t_a + 60$ s. Drift occurs in a significant (moderate) way for BMA (SA-BMA) for $t > t_a$.

4

Learning-based real-time tracking

In the previous Chapter, two intensity-based approaches for ultrasound tracking, namely a local affine registration and a novel scale-adaptive block-matching algorithm (SA-BMA), were presented, evaluated and discussed. While the first method is extremely accurate, it suffers from high computational costs. Contrary, BMAs can achieve real-time performance, but their accuracy on long sequences is lower.

In image-guided therapy, the compensation for any organ movement should be highly accurate and precise, and be achieved in real-time, i.e. faster than the image acquisition rate. Several techniques have been proposed to handle real-time respiratory organ movement in image-guided applications. Most of the current tracking systems are based on fluoroscopy images [Shirato *et al.* 2006] with the major disadvantages of higher dose to the patient and the need to implant fiducial markers. [Ries *et al.* 2010] proposed tracking of 2D MRIs based on a Kalman predictor combined with a compensation for through-plane motion. Experiments carried out on images of the kidney under regular respiratory motion due to mechanical ventilation showed a tracking precision (STD) of ± 1.1 mm and processing time of 40 ± 8 ms. A 2D real-time ultrasound tracking method based on active contours and conditional density propagation was proposed in [Zhang *et al.* 2010] and evaluated on a short sequence (10 breathing cycles). Results showed a precision (variance) of 1.6 mm and a processing time of 5 ms per frame.

While feature-based tracking methods are potentially fast, they might suffer from bad initialization, feature disappearances and feature location inaccuracies. Intensity-based methods do not require feature extraction and hence can work in regions with weak features. They usually need no pre-processing, but are generally computationally expensive. During treatment, images are acquired continuously over several minutes and, as seen in the previous Chapter, their temporal realignment is crucial to ensure robust tracking and to avoid error accumulation.

In this Chapter, a learning-based tracking framework, which combines the strengths of the aforementioned methods to achieve robust, accurate and fast vessel tracking in long ultrasound sequences, is introduced. It exploits the repetition in the images

in case of organs subject to repetitive movement, such as breathing, and is divided into an initial training phase (see Section 4.1) and a real-time application phase (see Section 4.2). In the application phase the redundancy in the data is investigated for accelerating affine registration (see Section 4.2.1) or for initializing the SA-BMA (see Section 4.2.2). Experiments and results are described in Section 4.3. The Chapter is concluded by Section 4.4 with a discussion on the performances and advantages of the proposed tracking methods and on potentials areas of improvement.

4.1 Training phase

During the training phase, a sequence covering 10 breathing cycles is acquired, resulting in T_{10C} images $I(t_i)$, with $t_i \in [t_0, \dots, T_{10C}]$.

PCA

In order to store the image appearance efficiently, I use a preprocessing step to reduce the dimensionality of the data. A large number of dimensionality reduction methods have been proposed and compared [Van der Maaten *et al.* 2009]. On real data, the performance of linear methods was not inferior to that of nonlinear ones. As linear methods are computationally less expensive, I choose Principal Component Analysis (PCA) for embedding $I(t_i) \in \mathbb{R}^D$ into a low-dimensional data representation $\mathbf{S}(t_i) = [s_1(t_i); \dots; s_L(t_i)] \in \mathbb{R}^L$, with $L \ll D$.

Each image $I(t_i)$ is reshaped into a single column vector $\mathbf{x}(t_i) \in \mathbb{R}^D$. I calculate the mean with respect to time ($\bar{\mathbf{x}}$) and the covariance matrix $\mathbf{C} = \hat{\mathbf{X}}^T \hat{\mathbf{X}}$, where the $(i+1)$ -th column of $\hat{\mathbf{X}}$ is equal to $\hat{\mathbf{x}} = \mathbf{x}(t_i) - \bar{\mathbf{x}}$. I then solve the eigenproblem

$$\mathbf{C}\mathbf{w}_j = \lambda_j \mathbf{w}_j, \quad (4.1)$$

$\forall j \in [1, \dots, T_{tr}]$ and $T_{tr} \leq T_{10C}$ being the number of training images. The eigenproblem in (4.1) is rearranged such that the eigenvalues λ_j are sorted ($\lambda_j \geq \lambda_{j+1}$), with \mathbf{w}_j being the corresponding eigenvector. For each j -th eigendirection I calculate the data projection $p_j = (\mathbf{w}_j)^T \hat{\mathbf{x}}$, so that the original image sequence can be reconstructed by

$$\mathbf{x} = \sum_{j=1}^{T_{tr}} p_j \mathbf{w}_j + \bar{\mathbf{x}}. \quad (4.2)$$

The most reconstructive information is captured by the eigenvectors associated with the largest eigenvalues and data reduction is achieved by considering only the first L eigenvectors, i.e. $s_j(t_i) = p_j(t_i)$ for $j \in [1, \dots, L]$. The selection of L is based on the energy accumulated in the first L eigenvectors (calculated by the normalized

cumulative sum of the eigenvalues λ_1 to λ_L) as this describes the reconstruction accuracy. I choose L such that the cumulative energy just exceeds 95%.

In addition, I extract the breathing signal \mathbf{s}_B from the PCA space \mathbf{S} , by selecting the \mathbf{s}_j that captures the main respiratory motion. In details, I compute the Fast Fourier Transform (FFT) of each \mathbf{s}_j and choose the one that provides the maximum of the power spectral density within a frequency range of 0.15-0.4 Hz (2.5-6 s), which corresponds to common breathing.

Image Registration

The images $I(t_i)$, with $t_i \in [t_1, \dots, T_{tr}]$, are registered to the first image $I(t_0)$ to obtain spatial correspondence at each time t_i . This enables the extraction of trajectories for individual points, e.g. vessel centers P_v . The registration optimizes the parameters of an affine transformation with respect to NCC over a manually selected region around $P_v(t_0)$, and is initialized by the result from registering $I(t_{i-1})$ to $I(t_0)$, as described in Section 3.3.

The low-dimensional representation \mathbf{S} of the images and the corresponding registration results (e.g. spatial locations for points of interest P_v) are stored at each time step $t_i \in [t_0, \dots, T_{tr}]$.

4.2 Real-time application phase

During a real-time procedure, new images are continuously acquired. Given the current image $I(t^*)$, I want to find the most similar images in the training database for obtaining a prediction of the current displacements for point of interest P_v (Section 4.2.1), or for temporal realignment of BMA (Section 4.2.2). For this, I first project the new image into the PCA space:

$$\mathbf{S}(t^*) = \sum_{j=1}^L p_j \mathbf{w}_j (\mathbf{x}(t^*) - \bar{\mathbf{x}}). \quad (4.3)$$

4.2.1 Learning-based affine registration

I determine the K -nearest neighbors (K -NN) in the training database to $\mathbf{S}(t^*)$. In details, I find $t_{m_k} \in [t_0, \dots, T_{tr}]$ with the K smallest Euclidean distances

$$d_k(t^*) = \|\mathbf{S}(t^*) - \mathbf{S}(t_{m_k})\|_2 = \sqrt{\sum_{j=1}^L [s_j(t^*) - s_j(t_{m_k})]^2}, \quad (4.4)$$

with $1 \leq k \leq K$. The distances are sorted so that $d_k \leq d_{k+1}$. Having determined the K -NN PCA representations from the training set, I retrieve the K associated positions $P_v(t_{m_k})$ per point of interest $P_v(t_0)$.

The prediction of the vessel center v at time t^* is obtained by weighted linear interpolation of $P_v(t_{m_k})$:

$$\hat{P}_v(t^*) = \sum_{k=1}^K \left(1 - \frac{d_k(t^*)}{\sum_k d_k(t^*)} \right) P_v(t_{m_k}). \quad (4.5)$$

Outliers

Completely repetitive motion is very unlikely. Therefore a mechanism to cope with previously unseen variations (outliers) and adapt to these (e.g. drift of exhale position or deeper inhalation) is incorporated. In order to detect such outliers I establish a threshold on the image similarity, which is based on the statistics of the training data. Specifically, for each training image I find its nearest neighbor to the remaining training samples and calculate the minimum Euclidean distance $d_1(t_i)$, for $t_i \in [t_0, \dots, T_{tr}]$, similar to (4.4). The threshold θ_{out} is then set to the 95th percentile of the distribution of the minimum distances. During the application phase, if $d_K(t^*) > \theta_{out}$ then $I(t^*)$ is categorized as outlier and affine registration is computed as described in Section 3.3. Finally I update the training database by adding the registration result and the PCA description of the outlier, and by recalculating θ_{out} . Summarizing, the tracking strategy is as follows:

```

if  $d_K(t^*) < \theta_{out}$  then do prediction, see Eq. (4.5)
else do affine registration and update S end

```

4.2.2 Learning-based block-matching algorithm

Alternatively, the training information can be used to allow frequent temporal re-alignment of the scale-adaptive block-matching algorithm (SA-BMA), previously described in Section 3.4, in order improve its robustness and accuracy.

Depending on the similarity of $\mathbf{S}(t^*)$ to the training data, the initial frame and the previous frame, a reference frame is chosen. The logic proposed for the learning-based block-matching (LB-BMA) is as follows:

```

outlierFlag = false
if  $\|\mathbf{S}(t^*) - \mathbf{S}(t_0)\|_2 < \theta_0$  then  $t^{ref} = t_0$ 
else if  $d_1(t^*) < \theta_{out}$  then  $t^{ref} = t_{m_1}$ 
else if  $\|\mathbf{S}(t^*) - \mathbf{S}(t^* - 1)\|_2 < \theta_{out}$  then  $t^{ref} = t^* - 1$ 
else outlierFlag = true end
if (outlierFlag == false) then do SA-BMA
else do affine registration and update  $\mathbf{S}$  end

```

The threshold θ_0 is the 5th percentile of the Euclidean distance between $\mathbf{S}(t_0)$ and $\mathbf{S}(t_i) \forall t_0 < t_i \leq T_{tr}$. $d_1(t^*)$ is the smallest Euclidean distance between $\mathbf{S}(t^*)$ and the training set \mathbf{S} , see Eq. (4.4). θ_{out} is the 95th percentile of the distribution of the minimum Euclidean distances between the $\mathbf{S}(t_i)$ in \mathbf{S} , see Section 4.2.1.

4.2.3 Computational complexity

During the application phase, vessels of images that are not an outlier can be tracked by predicting the displacements from (a) the training data (see Section 4.2.1), or (b) by the learning-based SA-BMA (see Section 4.2.2). The computational complexity of the tracking methods is:

- Projection of the new image into the training PCA space: $O(LD)$, where L is the number of selected eigenvectors and D the number of image pixels
- Method (a): Computation of the point displacement
 - K-NN search: let $T_{tr} \leq T$ be the size of the search space (i.e. the number of training images), then the K nearest neighbors can be found in $KT_{tr} \log T_{tr}$ calculations
 - Prediction via interpolation: for each 2D point, the complexity depends on $2K$
- Method (b): BMA requires $9MN \log(MN) + 18B(2\beta + 1)^2 + 9B$ calculations, see Section 3.4.4

For v points per image ($1 \leq v \leq 5$), method (a) requires a total of $LD + KT_{tr} \log T_{tr} + 2Kv$ operations. As $K, v < L \ll D$ and $T_{tr} \ll D$, the main complexity is linear in D and can be approximated by $O(LD)$. The complexity of method (b) is based on $LD + 9MN \log(MN) + 18B(2\beta + 1)^2 + 9B$, and with $B \approx MN$ and $B\beta \ll D$ is again $O(LD)$.

4.3 Results

The training phase has one parameter, i.e. the number of training images (T_{tr}). In order to get a good trade-off between the size and the information content of the training set, I apply the L-curve method to determine T_{tr} . Specifically, I calculate the threshold value θ_{out} (see Section 4.2.1) for increasing number of training images. Then I fit a low-degree (degree=4) polynomial curve to the data and finally find the point on the curve which is closest to the origin. The number of training images associated with this point are used as T_{tr} . An example of L-curve is shown in Figure 4.1a. T_{tr} is in the range of [52, 193] for all datasets, while T_{10C} varies from 519 to 966 frames.

The PCA space, derived from the T_{tr} training images, is characterized by L in the range of [12, 52], vs. $D > 23000$.

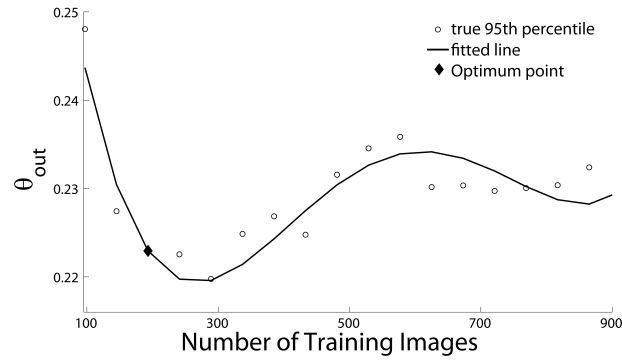
The learning-based affine registration (LB-Affine) has an additional parameters, i.e. the number of nearest neighbors (K), which is determined using the remaining training images (from $T_{tr} + 1$ to T_{10C}). I calculate the prediction error for the images in the remaining training set, defined by the Euclidean distance between the vessel position from Eq. (4.5) and the one from affine registration. I allow $K = 1, 2, \dots, 20$, and then choose the K that corresponds to the minimum mean error. The chosen K varies in [3, 8]. Figure 4.1b shows an example of the mean prediction error as a function of K for a representative sequence.

I firstly evaluated the registration error for the images acquired during the training phase (from t_0 to T_{10C}). The affine registration achieves an accuracy of 0.63 ± 0.36 mm (1.30 mm) on average (MTE \pm STD (95th percentile of TE)), with a MTE_v range of [0.42, 0.84] mm.

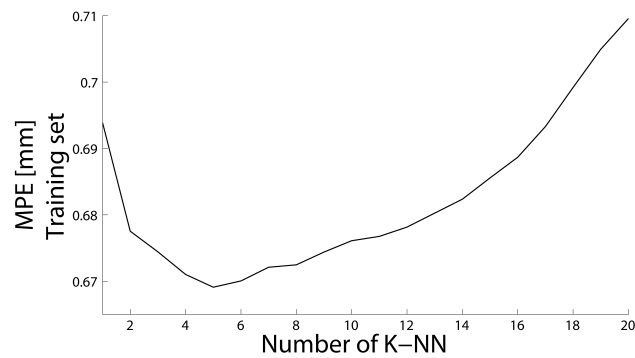
As in the previous Chapter, I tracked the same ~ 50000 frames, acquired over a total of ~ 50 min. For comparison, I computed the motion magnitude of the landmarks, see Eq. (3.2). I compared two versions of the LB-BMA, called LB-BMA₉₅ and LB-BMA_B, considering \mathbf{S} and \mathbf{s}_B respectively. Additionally, LB-BMA was tested using $T_{tr} = T_{10C}$ (LB-BMA(10C)). Then the PCA space is characterized by L in the range of [86, 287].

Results for all methods are shown in Tables 4.1 and 4.2. The highest accuracy is achieved by LB-Affine with a MTE of 0.86 mm. The best block-matching performance is provided by LB-BMA₉₅(10C) with a MTE of 0.96 mm. Using a larger training set improved the MTE by 44% (36%) for LB-BMA₉₅ (LB-BMA_B). In addition, the improvement with respect to (SA-)BMA is substantial (see Section 3.5).

The MTE of the 3 observers annotation is 0.32 mm (more details in Section 3.5), which improves the accuracy with respect to the proposed automatic methods by 62% and 67% for LB-Affine (MTE=0.84 mm) and LB-BMA₉₅(10C) (0.97 mm),



(a) Training phase L-curve



(b) Mean prediction error (in mm) vs. number of nearest neighbors

Figure 4.1: Example of parameter optimization. (a) The number of training images T_{tr} is selected to provide a good trade-off between density and size of the training set using the L-curve method. (b) K is chosen so that the mean prediction error (MPE) evaluated over the remaining training images (from $T_{tr} + 1$ to T_{10C}) is minimum.

respectively, see Table 4.2. Figure 4.2 compares the results of individual vessels for LB-BMA_B(10C), LB-BMA₉₅(10C) and LB-Affine, showing the minimum, maximum, median, 25th and 95th percentile of the TE_v distribution of all images annotated by 3 observers for each vessel. It can be observed that using only the breathing signal (LB-BMA_B(10C)) leads to comparative high 95th percentile (>5 mm) for 8 vessels. The corresponding median values are not that different, suggesting that the method can perform well, but is not very reliable. When using more PCA components (LB-BMA₉₅(10C)) only one 95th percentile is that high and for LB-Affine all of them are below 3 mm.

For the inter-observer data set, the statistical significance of the median TE_v of SA-BMA and LB-BMA₉₅(10C) and LB-Affine was tested using the paired Wilcoxon signed-rank test. The probabilities (p) of equal medians are as follows:

SA-BMA	vs.	LB-BMA ₉₅ (10C)	p<0.001
SA-BMA	vs.	LB-Affine	p<0.001
LB-BMA ₉₅ (10C)	vs.	LB-Affine	p=0.23

No other statistical tests were performed.

Figure 4.3 illustrates the benefit of a learning-based approach (LB-BMA₉₅(10C)) for the worst case of BMA (see Section sec:res). Figure 4.4 shows a case of skipped frames during acquisition and the advantage of a learning-based approach (LB-BMA₉₅(10C)) over SA-BMA.

The percentages of images similar to the training phase and the outliers during the application phase are reported in Table 4.3. The percentage of required registrations during the application phase for LB-Affine and all 25 tracked vessels is on average 4.7% and always below 21%. For LB-BMA₉₅ (LB-BMA_B) t^{ref} is picked from the training set for 72.0% (99.0%) of the frames, while less than 0.3% requires affine registration (see Table 4.3). For the remaining frames $t^{ref} = t^* - 1$. For LB-BMA₉₅(10C) (LB-BMA_B(10C)) the mean percentage of detected outliers is 1.4% (1.0%). Figure 4.5 compares the percentage of required affine registrations during application phase for LB-BMA_B(10C), LB-BMA₉₅(10C) and LB-Affine for each of the 15 manually annotated frames. Generally, no correlation between the tracking error (Figure 4.2) and the amount of affine registrations (Figure 4.5) per vessel can be observed. The affine registration percentage is particularly high for LB-Affine for vessel 4 (20.8%). However, its accuracy (see Figure 4.2) is slightly worse than LB-BMA₉₅(10C) for the same vessel, which requires only 1.3% of affine registrations. LB-BMA₉₅(10C) requires for one case 8.2% affine registrations, while otherwise these are always below 2.1%.

The average computational time for each PCA projection (see Eq. 4.3) was 1.4 ms for LB-Affine and LB-BMA, and 13.0 ms for LB-BMA(10C). The average time needed to calculate the motion prediction of the tracked vessel per frame in the application phase was 0.06 ms for LB-Affine and 100 ms for block-matching methods. The latter had a range of [30, 350] ms. These measurements exclude outliers, i.e. the images that required affine registration, which is computed in approximately 800-2500 ms per image region. These results were obtained using non-optimized Matlab software and no GPU parallel computing (single PC with Intel®Core™i7-920 at 2.66 GHz processor and 8 GB RAM).

4.4 Conclusion

I proposed a novel, simple and robust framework for vessel tracking in long ultrasound sequences. The method is based on learning the relationship between image appearance and feature displacements, in order to speed-up image registration or

Table 4.1: Tracking results (in mm) for the different methods w.r.t. manual annotation from one observer (10%, ~ 7500 images). Best results are in bold face.

	MTE \pm STD (95 th TE)	range MTE _v
VesselMotion	5.17 \pm 3.21 (10.59)	[2.81, 11.48]
LB-Affine	0.86 \pm 0.46 (1.68)	[0.65, 1.03]
LB-BMA _B	1.93 \pm 1.83 (6.12)	[1.55, 2.34]
LB-BMA ₉₅	1.70 \pm 1.26 (4.23)	[1.38, 2.28]
LB-BMA _B (10C)	1.24 \pm 1.41 (3.81)	[1.04, 1.49]
LB-BMA ₉₅ (10C)	0.96 \pm 0.64 (2.26)	[0.38 , 2.34]

Table 4.2: Tracking results (in mm) for the different methods w.r.t. manual annotation from three observers (3%, ~ 2500 images). Best results are in bold face.

	MTE \pm STD (95 th TE)	range MTE _v
VesselMotion	5.22 \pm 3.23 (10.57)	[3.00, 11.30]
LB-Affine	0.84 \pm 0.44 (1.60)	[0.64, 1.00]
LB-BMA _B	1.96 \pm 1.94 (6.57)	[1.60, 2.36]
LB-BMA ₉₅	1.69 \pm 1.24 (4.20)	[1.38, 2.21]
LB-BMA _B (10C)	1.21 \pm 1.39 (3.67)	[0.98, 1.49]
LB-BMA ₉₅ (10C)	0.97 \pm 0.65 (2.20)	[0.36 , 2.24]

Table 4.3: Comparison of the partitioning performance of the different methods during the application phase for all 25 vessels. **Training**: percentage of interpolations/initializations from images in the training set. **Registration**: percentage of images categorized as outliers and hence requiring affine registration.

	Training %		Registration %	
	mean	range	mean	range
LB-Affine	95.3	[79.2, 99.2]	4.7	[0.8, 20.8]
LB-BMA _B	99.6	[90.6, 100]	0.0	[0, 0.3]
LB-BMA ₉₅	73.5	[49.5, 94.7]	0.1	[0, 0.3]
LB-BMA _B (10C)	99.0	[95.8, 99.7]	1.0	[0.3, 3.8]
LB-BMA ₉₅ (10C)	68.3	[27.2, 94.1]	1.4	[0.3, 8.2]

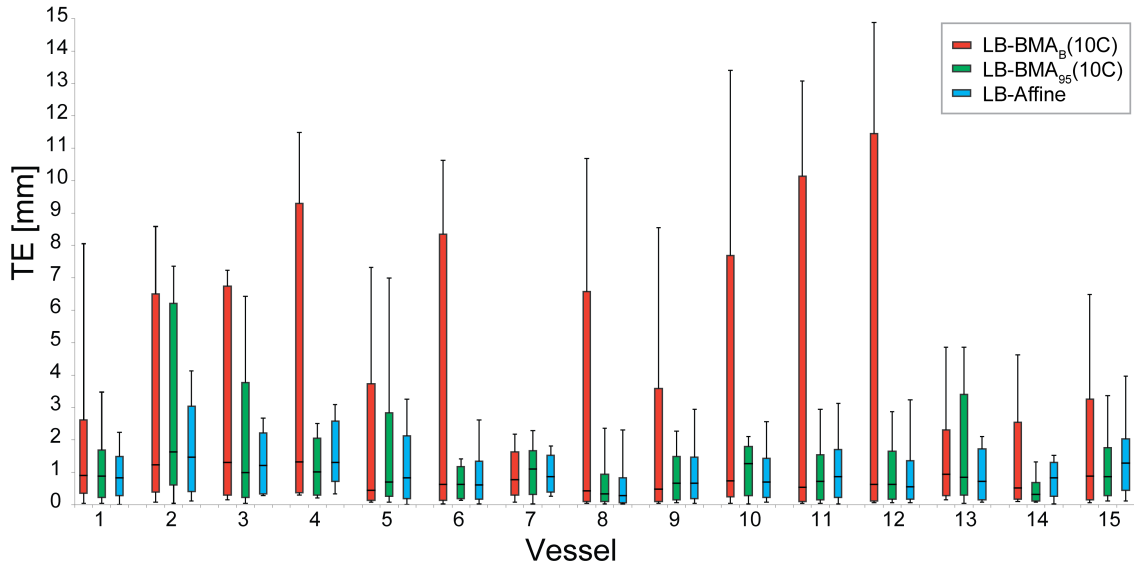


Figure 4.2: Summary of the tracking error distribution (TE) in mm for each of the 15 vessels for LB-BMA_B(10C) (red), LB-BMA₉₅(10C) (green) and LB-Affine (blue). The error bars show the minimum and maximum error, while the colored bars define the 25th percentile, median and 95th percentile of the TE.

to allow frequent reinitialization of a scale-adaptive block-matching algorithm. The method exploits the redundancy of the images in a repetitive motion scenario while still being able to adapt to irregularities.

The method was evaluated on long ultrasound sequences of the liver of 9 volunteers under free breathing and achieved a mean accuracy of 1 mm for tracking vessels for 5-10 min. Compared to non-learning intensity-based approaches (see Chapter 3), the proposed method is automatic, robust to interference, noise (see Figure 3.1), and frame drop outs (see Figure 4.4).

The temporal realignment of the tracking was often too sparse for (SA-)BMA (see Section 3.4). In contrast, the learning based approach enables more frequent realignments to relevant images by exploiting the repetition in the images and learning the main variation in image appearance. This allows to cope with and adapt to previously unseen variations by detecting these, registering the current image and updating the training set.

Tests on respiratory liver motion show that on average only 1-5% of the images required registration during the application phase, despite acquisitions up to 10 min and irregular changes in exhale position (drift) [von Siebenthal *et al.* 2007a]. Furthermore, the training images are acquired in a relatively short time (<1 min).

Reducing computational costs by using only the breathing signal for measuring image similarity increased the LB-BMA mean errors (1.70 vs. 1.93 mm and 0.96 vs. 1.24 mm).

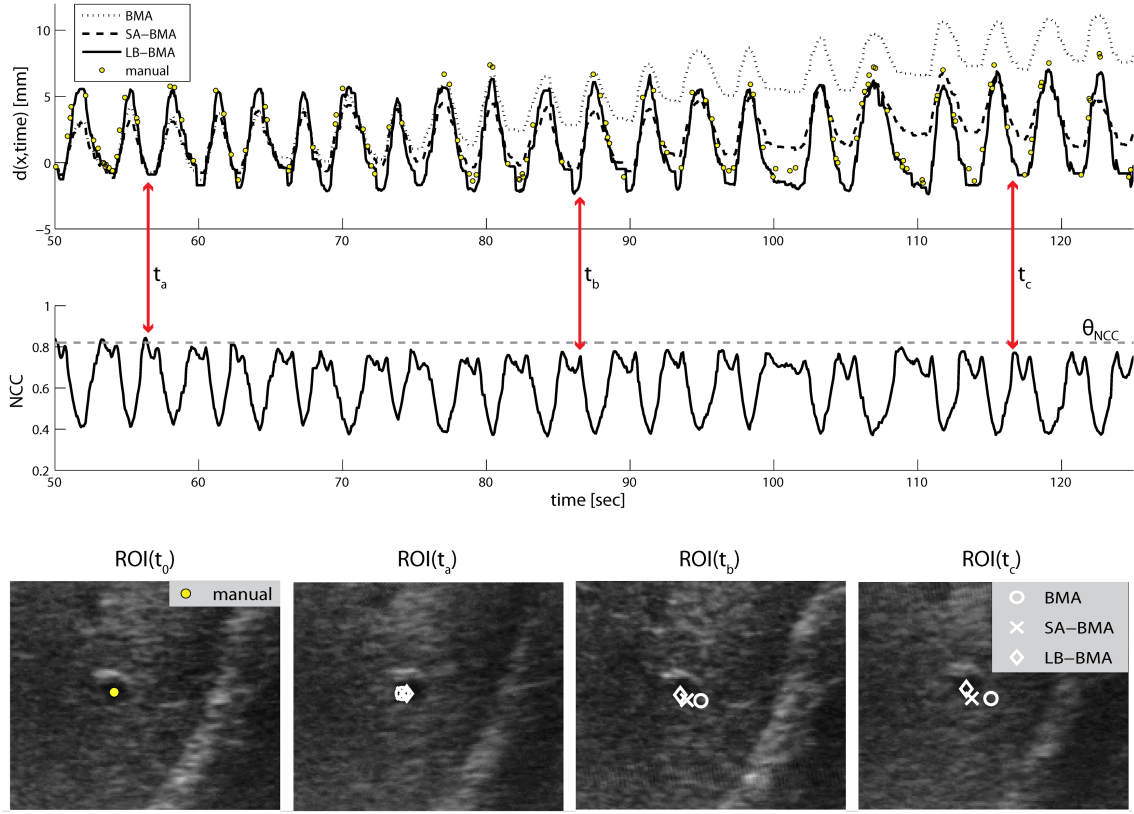


Figure 4.3: Comparison of the block-matching tracking performance for a failure sequence of BMA ($MTE_v=12.4$ mm, see Table 3.1). (Top) Main motion component of manual annotation and displacement \mathbf{d}_v from 3 methods for a temporal subset. (Middle) Corresponding NCC to first image. (Bottom, left to right) First image with annotation ($P_v(t_0)$), image with tracking results at last realignment (t_a) of (SA-)BMA, at $t_b = t_a + 30$ s and $t_c = t_a + 60$ s. Drift occurs in a significant (moderate) way for BMA (SA-BMA) for $t > t_a$, while LB-BMA remains robust.

While affine registration performed well on the training set, it was only applied to outliers during the real-time application phase due to its computational complexity (see Section 3.3.2).

The ultrasound tracking should provide continuous feedback on the displacements of points of interest in time. Therefore the main requirements for the tracking algorithm are both accuracy and speed. LB-Affine achieved on average the highest accuracy and the computation of the predicted displacements is with 1.5 ms real-time, i.e. faster than the image acquisition time (40-70 ms, 14-25 Hz). Yet, the number of affine registrations during the application phase is still too high to obtain continuous tracking of the images. The mean accuracy of $LB-BMA_{95}(10C)$ is also 1 mm. Even if the average computational time of 113 ms is higher, the percentage of registrations needed is lower (1.4%). Therefore the displacements of points of interest

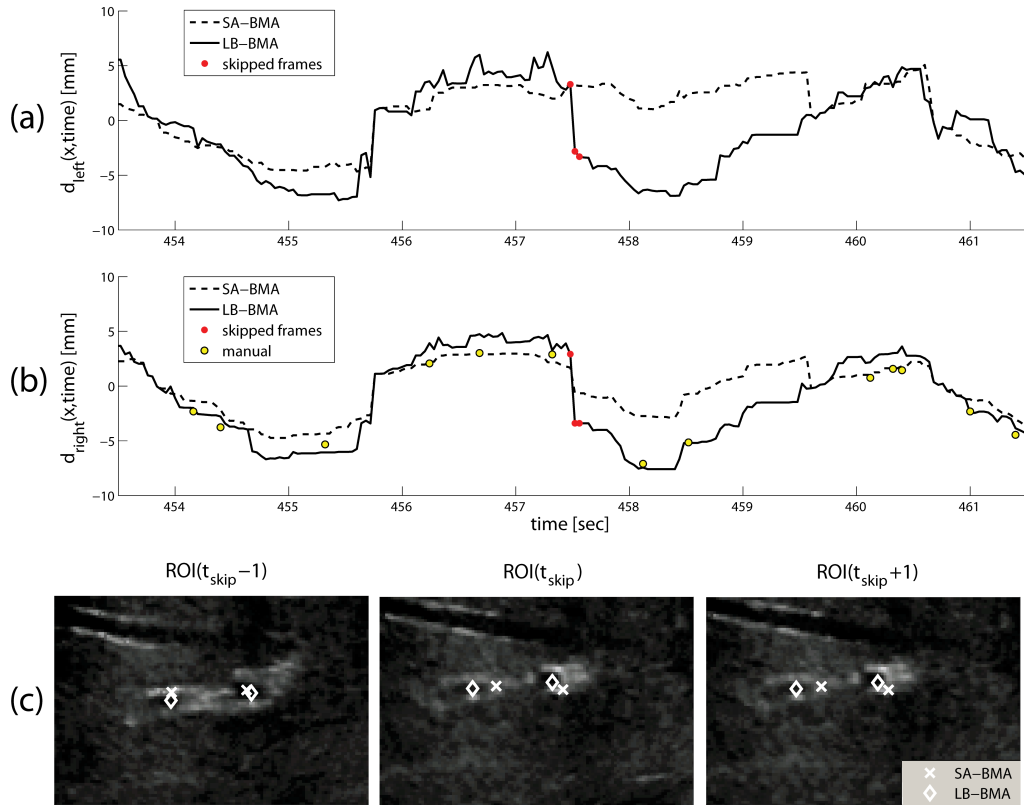


Figure 4.4: Example of frames skipped during acquisition for one volunteer. (a,b) Main motion component of displacements \mathbf{d}_{left} and \mathbf{d}_{right} of the (a) left and (b) right vessel from SA- and LB-BMA for a temporal subset around the frames shown in (c). MTE_{right} is 2.04 mm and 0.92 mm for SA- and LB-BMA, respectively. The left vessel was not quantitatively evaluated. (c) Image region with tracking results before ($t_{skip} - 1$) and after (t_{skip} and $t_{skip} + 1$) skipping happened, showing that motion compensation is insufficient for SA-BMA when dropouts occur, while LB-BMA remains robust.

can be computed without interruptions. Additionally, real-time implementation of BMA is possible [Hsu *et al.* 2005, Pinton *et al.* 2006].

I expect the proposed method to be generally beneficial for real-time tracking of mainly repetitive motion (e.g. respiratory, cardiac). In particular, it will substantially improve the real-time capability of intensity-based 3D image registration, which has a considerably higher computational complexity. The temporal reinitialization strategy could also be used to speed-up the training phase by searching for the most similar previous image for initialization of affine registration, instead of using the previous image ($t - 1$).

In this Chapter, I proposed two learning-based methods for real-time, robust and accurate tracking of long ultrasound sequences of the liver. The achieved accu-

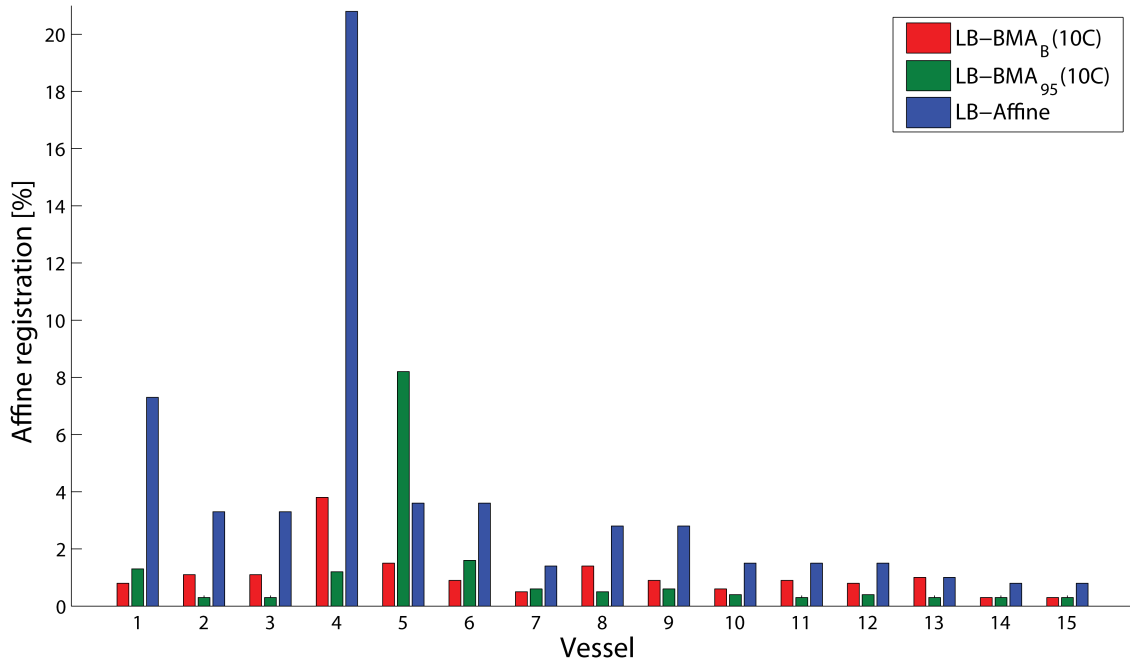


Figure 4.5: Percentages of affine registration required during application phase by LB-BMA_B(10C), LB-BMA₉₅(10C) and LB-Affine for each of the 15 manual annotated vessels.

racy and robustness of the proposed tracking method for long and very difficult ultrasound sequences makes it applicable for real-time ultrasound guidance during radiation therapy under free-breathing. In the following Chapter, I will describe a novel treatment scenario, where MR and ultrasound images are acquired simultaneously (for validation) during MRgHIFU. The motion vectors obtained from tracking ultrasound images of the liver are used as internal surrogates for a model-guided spatio-temporal prediction of the tumor position in the liver during treatment.

5

Simultaneous MR- and ultrasound-guided radiation therapy

In Chapter 4, I introduced a robust and fast algorithm for accurate tracking of anatomical landmarks, such as vessel centers, of the liver under free breathing in long ultrasound sequences. In this Chapter, I describe the clinical scenario where this ultrasound tracking method is used. In Section 5.1, a novel hybrid ultrasound and MR imaging system for guidance in HIFU treatments of the upper abdomen, subject of respiratory motion, is introduced. During treatment the direct observation of the target motion is often impossible. Therefore a novel strategy for its motion compensation is proposed. Simultaneously acquired MR and ultrasound images are co-registered in the pre-treatment phase (see Section 5.1.3). Tracking of ultrasound images during the application phase generates surrogate signals of the organ motion. These surrogates are combined with a statistical motion model of the liver (see Section 5.2.1) in order to obtain spatio-temporal motion prediction of the target (see Section 5.2.3). Results showing the applicability of such system in clinical practice are presented in Section 5.3 and discussed in Section 5.4.

5.1 Hybrid ultrasound/MR imaging system

Hybrid imaging systems integrate two modalities, taking advantages of their strengths, in order to gain more information on the investigated anatomy. Hybrid ultrasound/MRI systems were developed only recently [Curiel *et al.* 2007, Tang *et al.* 2008, Feinberg *et al.* 2010, Viallon *et al.* 2010]. The combination of ultrasound and MR imaging can improve guidance and assessment in thermal therapies, such as HIFU. Their successful combination is mainly due to the complementary information which the two modalities provide.

Ultrasound and MR imaging are both non-ionizing techniques, but are characterized by different spatial and temporal resolutions, and are also characterized by a difference in sensitivity when measuring tissue properties. Ultrasound imaging has the advantages of being portable, inexpensive and free of geometrical distortion. Due to its fast imaging rate, ultrasound is a valuable tool for motion tracking [Pernot *et al.* 2004]. In HIFU treatments, ultrasound can also directly visualize acoustic obstacles and sonosensitive microbubbles [Dindyal and Kyriakides 2011], and characterize the HIFU thrombolysis [Wright *et al.* 2012]. Yet currently ultrasound cannot provide robust thermometry information. MRI offers multiplanar views, good soft tissue contrast, and techniques for real-time thermometry during treatment [Ishihara *et al.* 1995, Holbrook *et al.* 2010]. Therefore it is widely used for guidance in HIFU treatments [Hynynen 2010]. However, its lower temporal resolution is a limit for fast motion compensation.

5.1.1 MR-compatible ultrasound imaging

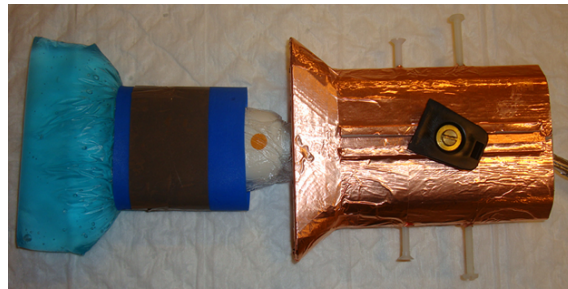
In the proposed hybrid system developed at the Radiology Department of the University Hospital of Geneva [Petrusca *et al.* 2011, Petrusca *et al.* 2013], ultrasound and MR images are acquired simultaneously. MR-compatible ultrasound imaging was achieved using an Acuson clinical ultrasound scanner (Antares; Siemens Medical Solutions, Mountain View, CA) and a CH4-1 transducer (256 phase array transducer, 1.8-4.0 MHz frequency bandwidth, multifocal option), which was modified by the manufacturer to avoid magnetic materials. The ultrasound scanner was integrated with 1.5 T Espree and 3 T Trio clinical MR scanners (Magnetom; Siemens, Erlangen, Germany).

In order to obtain MR-compatible ultrasound imaging, additional electromagnetic (EM) shielding was required. A customized holder for the ultrasound probe was produced using the stereo-lithography (electrical insulator resin). The probe head was placed centrally inside the holder, allowing an internal gap of 3-4 cm from the holder edge. To prevent noise at the RF detection range of the MR scanner, the holder and the transmission line (a 7 m-long cable) were entirely shielded using cooper and aluminum coating. The transmission line was further coated by a flexible plastic tube (electrical insulator). The frontal edge of 2 mm thickness of the holder was left non-metalized, in order to avoid electrical contact with the skin.

The ultrasound probe was accommodated in a thin plastic bag, shaped to match the cavity of the holder and filled with standard ultrasonic gel, free of air bubbles (Aquasonic, Parker Laboratories, Fairfield, NJ), see Figure 5.1a. This configuration ensures careful acoustic coupling with the volunteer skin and respiratory motion decoupling. In addition, it reduces the risk of residual susceptibility artifact from the probe and avoids electrical contact with the patient skin.

The shielded transducer was attached to an orbital ring using an articulated handler, see Figure 5.1b, to obtain a fixed position of the ultrasound probe in the MR environment. Both the orbital ring and the handler were MR-compatible. The orbital ring was positioned on the MR bed approximately at the isocenter of the scanner. To facilitate the adjustment of the ultrasound probe position, the proposed setup allowed for 5 degrees of freedom: head/feet translation of the ring on the MR bed, circular revolution of the handler on the perimeter of the ring, 2 rotations in the handler, and translation of the ultrasound holder parallel to the base of the handler, to control the distance between the ultrasound transducer and the patient skin.

During image acquisition, the ultrasound imaging probe operated inside the magnet bore, while the ultrasound scanner operated outside the magnet room. Therefore, the transmission line was passed through the Faraday cage via waveguide and the EM shielding structures were grounded to the cage.



(a) The ultrasound transducer is accommodated in a plastic bag filled with ultrasonic gel and in a EM-shielded holder.



(b) The ultrasound holder is attached to the MR-compatible orbital ring using an articulated handler.

Figure 5.1: MR-compatible ultrasound system. Adapted from [Petrusca *et al.* 2013].

Mutual influence of the two modalities

In vitro tests on a static gel phantom were performed to assess the mutual influence of the two modalities. MR images were acquired according to clinical T1 VIBE (T1-weighted fat-saturated volume interpolated gradient echo) and 4D imaging protocols [von Siebenthal *et al.* 2007a]. Simultaneously, the ultrasound scanner was turned on and off. The SNR of the MR images with and without ultrasound, was computed. Main results are summarized in Section 5.3, while more details and further tests can be found in [Petrusca *et al.* 2013].

The influence of the MR system on the ultrasound images was evaluated on a gel phantom with echogenic inclusions. Ultrasound images were recorded while pausing the MR scanner, performing automatic shimming within a volume of interest, or running a 4D sequence. Two successive ultrasound images acquired for a given status of the MR system were subtracted. The root mean square (RMS) of the difference image was then computed within an angular sector, set at a radial distance of approximately 8 cm from the transducer and including more than 4000 pixels.

5.1.2 Simultaneous motion monitoring of the liver

Simultaneous MR and ultrasound temporal sequences of the liver of healthy volunteers were acquired. The imaging setup with the volunteer positioned on the MR bed is illustrated in Figure 5.2a. We acquire data of 6 subjects, consisting of 2 female and 4 male subjects between 28 and 45 years of age, under free breathing.

Pre-treatment scan

Firstly, we acquired for each subject a 2D ultrasound image and a 3D MRI during breath-hold. These images were co-registered to allow initial spatial correspondence between the two image modalities, see Section 5.1.3. During image acquisition, all volunteers are instructed to suspend respiration at end exhalation. 2D second harmonic ultrasound images are characterized by 14-17 Hz temporal and 0.28-0.4 mm spatial resolution, depth range of 18-21 cm, single focus mode, center frequency of 2.2 MHz, 20 dB output gain and 45-60 dB dynamic window. All the ultrasound images in this Chapter have the aforementioned characteristics. MR images were acquired on a 1.5 T scanner. A standard six-element phased-array body coil was used for signal acquisition, see Figure 5.2. A high resolution abdominal image was obtained from T1 VIBE scan (TE/TR equal to 1.96/4.36 ms, flip angle of 10° , spatial resolution 1.25 mm x 1.25 mm x 2 mm).



(a) Volunteer positioning on the MR bed with the ultrasound imaging setup.

(b) Ultrasound subcostal longitudinal scan.

Figure 5.2: Example of imaging setup.

Simultaneous time sequences

Ultrasound and MR images were simultaneously acquired to obtain the ground-truth motion for validation. Data from both modalities was recorded in 3 to 5 blocks for a duration of approximately 5 min 30 s each, with a break of 1 min between blocks.

2D second harmonic ultrasound sequences were acquired to observe the liver motion of each subject. The images were exported on-the-fly from the ultrasound reconstructor to an external PC for processing using a Video Graphics Array-Universal Serial Bus grabber (native resolution, 1280x1024 exported 1:1 uncompressed, 30 Hz). Subcostal/intercostal oblique/longitudinal scans were chosen as a compromise between maximizing the size of the acoustic window (narrowed by the presence of ribs) and minimizing the of the out-of-plane motion, see Figure 5.2b.

MR images were generated with a balanced steady-state free precession (bSSFP) sequence. This sequence allows for high vessel-tissue contrast, necessary for the detection of the organ position. The main acquisition parameters are as follows: TE/TR 1.38/3 ms, flip angle 78° , in-plane resolution 1.82 mm isotropic, slice thickness 6 mm, slice gap -2 mm, 30 slice positions, image matrix 192x192, parallel imaging acceleration factor 2.

The acquisition sequence of the MR images consisted of alternating between 2D sagittal navigator slices (at a fixed spatial position) and 2D sagittal data slices (at

different locations to cover the right liver lobe equidistantly) with a frequency of 2.6 Hz. 4D MR images can be created by sorting the data slices based on the similarity of the enclosing navigator slices [von Siebenthal *et al.* 2007a]. In details, non-rigid 2D registration of the liver in all navigator images [Rueckert *et al.* 1999] is performed, resulting in a 2D displacement field for each navigator. Since every data slice is enclosed by a preceding and a subsequent navigator, data slices belonging to the similar positions of the liver can be retrieved by comparing the deformation fields of their enclosing navigators. 2D frames are then retrospectively stacked to create time-resolved 3D images [von Siebenthal *et al.* 2007a, Preiswerk *et al.* 2013].

The multimodal acquisitions started at approximately the same time point (0.1 s offset), triggered by optical pulses generated by the MR scanner, which unfroze the ultrasound acquisition.

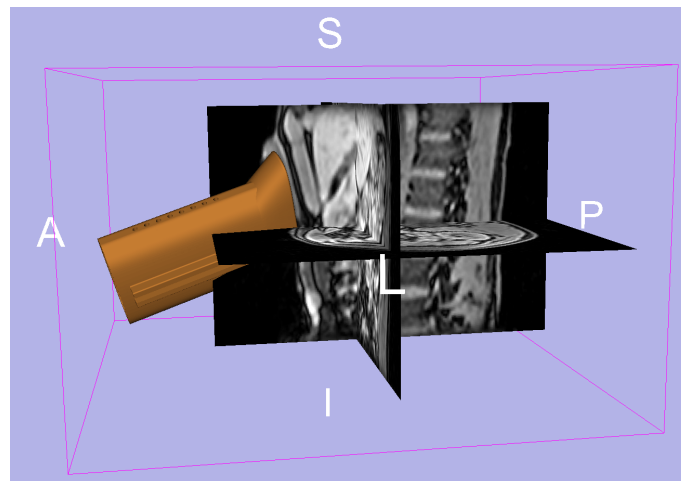
5.1.3 Images co-registration

The advantage of a simultaneous hybrid image acquisition is that the images can be inherently spatially and temporally registered [Curiel *et al.* 2007]. Establishing spatial correspondence in such a imaging setup is easier as the motion of the volunteer does not need to be compensated for. Hence only the spatial correspondence between the two image coordinate systems need to be determined. Multimodal registration of ultrasound and MR images from separate acquisitions at different times was reported by [Porter *et al.* 2001, Penney *et al.* 2004, Huang *et al.* 2005, Curiel *et al.* 2007, Milko *et al.* 2009, Chandrana *et al.* 2011, Mercier *et al.* 2012].

The spatial correspondence between the 2D ultrasound imaging plane and the 3D MRI reference frame is retrospectively established in the pre-treatment phase from the breath-hold images, see Figure 5.3. The shielded ultrasound probe is not directly visualized in MR images. Yet, the resin-made wall of the transducer holder and the external part of the coupling gel-bag are visible in the 3D MRI. I manually localized the ultrasound holder in the 3D MR volume, see Figure 5.3a. The position of the ultrasound plane with respect to the holder is known by design and hence can easily be related to the MRI, resulting in a 3D rotation $\mathbf{R}_{3D} \in \mathbb{R}^{3 \times 3}$. A final in-plane rigid transformation aligns manually annotated anatomical landmarks in the two modalities. The rigid transformation is composed of rotation $\mathbf{R}_{2D} \in \mathbb{R}^{2 \times 2}$ and translation $\mathbf{T}_{2D} \in \mathbb{R}^{2 \times 1}$. Finally, 2D points \mathbf{P} in the ultrasound image are transformed to 3D physical positions in the 3D MRI coordinate system by:

$$\mathbf{P}_{3D} = \mathbf{R}_{3D} \mathbf{T}_{\uparrow 3D} (\mathbf{R}_{2D} \cdot \mathbf{P} + \mathbf{T}_{2D}) = \mathbf{M} \cdot \mathbf{P}, \quad (5.1)$$

where $\mathbf{T}_{\uparrow 3D} \in \mathbb{R}^{3 \times 2}$ maps 2D into 3D coordinates.



(a) Registration of the ultrasound holder model in the reference frame of the 3D MRI.



(b) 2D MRI.

(c) Ultrasound image.

(d) Fused image.

Figure 5.3: Example of co-registration of the pre-treatment breath-hold images. (b) Interpolated 2D MRI in the corresponding ultrasound imaging plane (c). (d) Overlay of the registered 2D MRI and the gradient magnitude ultrasound image. The gradient of the ultrasound image was only computed for visual purposes.

5.2 Model-guided respiratory motion prediction

In our proposed work, a population-based statistical motion model, aligned to a pre-operative 3D scan of the subject, is used together with a Bayesian reconstruction algorithm to compute a spatio-temporal prediction of the organ position based on partial information obtained from tracking ultrasound images [Preiswerk *et al.* 2013].

5.2.1 Statistical model of the respiratory motion of the liver

To model and predict the respiratory liver motion in our experiments, the population-based statistical motion model from [Preiswerk *et al.* 2012] was used. This model consists of liver motion data from 4D MRIs of 20 healthy individuals and does not include the 6 datasets acquired for this study.

To build a population-based motion model, mechanical correspondence is established along the involved datasets. This correspondence is achieved by a semi-automatic framework and results in a master shape for each subject, consisting of 900 corresponding surface points. Correspondence within the organ is established computing the mean shape of all the master shapes and embed it in an isotropic grid of 10 mm spatial resolution. The grid is then warped to all master shapes.

The motion model is computed using PCA of the motion fields extracted from the registration results at the corresponding grid locations. For subject s at time t , we consider the relative offset of the liver $\mathbf{l}_s(t)$ to the master exhalation state $\check{\mathbf{l}}_s$:

$$\mathbf{x}_s(t) = \mathbf{l}_s(t) - \check{\mathbf{l}}_s. \quad (5.2)$$

We compute the mean motion vector over the n considered respiratory states for all subjects $\bar{\mathbf{x}} = 1/n \sum_{i=1}^n \mathbf{x}_i$ and the covariance matrix

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (5.3)$$

Singular value decomposition (SVD) of Σ yields to

$$\Sigma = \mathbf{U}\Lambda\mathbf{U}^T, \quad (5.4)$$

where \mathbf{U} is the matrix of the eigenvectors and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{n-1})$ is the matrix of the corresponding eigenvalues.

The projection of the liver data into the model space is

$$\mathbf{c}_s(t) = \text{diag}(\sigma_i^{-1})\mathbf{U}^T(\mathbf{x}_s(t) - \bar{\mathbf{x}}), \quad (5.5)$$

with $\sigma_i = \sqrt{\lambda_i}$, sorted in descending order.

5.2.2 Ultrasound tracking

During the application phase, accurate tracking of points of interest in ultrasound sequences was achieved using learning-based affine registration, described in Chapter 4. The tracked points $\mathbf{P}_p(t)$ are distributed on the diaphragm, liver surface and vessel centers, see Figure 5.4.

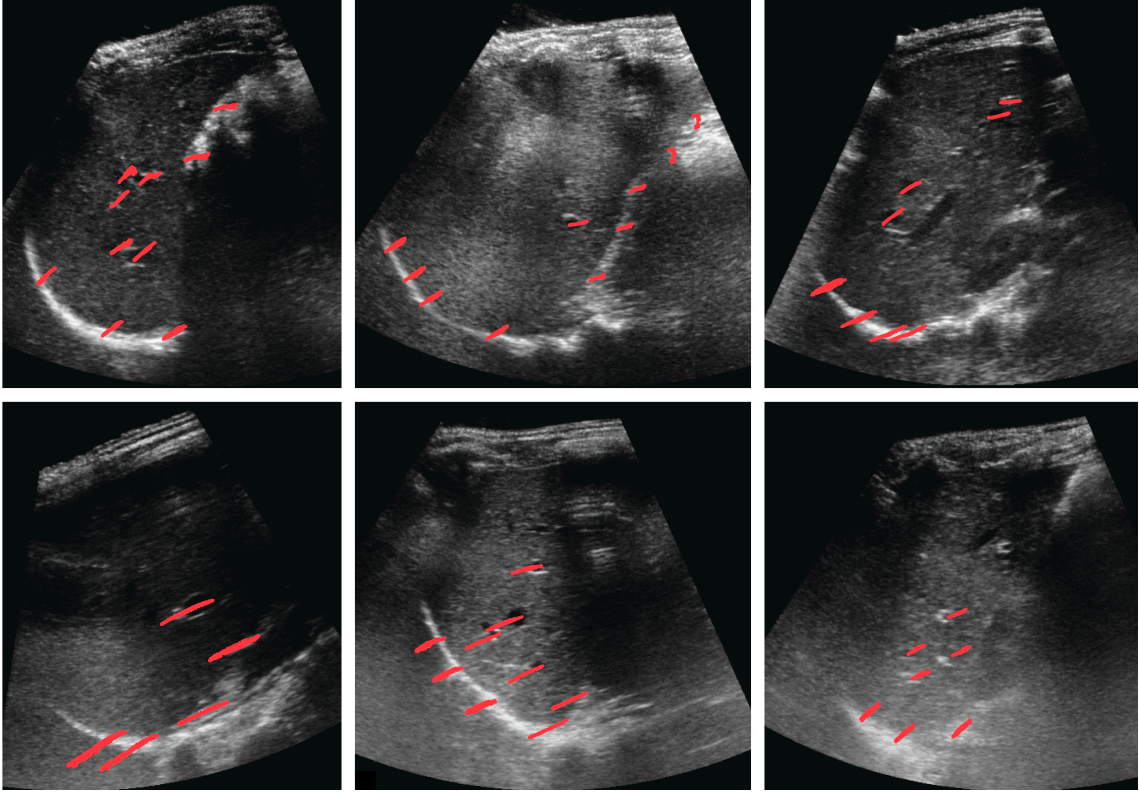


Figure 5.4: First frame $I(t_0)$ of the 6 ultrasound sequences with overlaid tracking results (red) for five breathing cycles, $\mathbf{P}_p(t) \forall t \in [t_0, \dots, T_{5C}]$. Tracked points include vessel centers within the liver and points along the liver surface (mostly on the diaphragm). Adapted from [Preiswerk *et al.* 2013].

A typical exhalation position $\check{\mathbf{P}}_p$ of each tracked fiducial p in the ultrasound sequence is chosen and assigned to its nearest grid point k in the liver. As the organ motion is observed in 2D images, the tracked points are mapped to the 3D MRI coordinate frame (see Eq. 5.1):

$$\mathbf{P}_{p,3D}(t) = \mathbf{M}\mathbf{P}_p(t). \quad (5.6)$$

Given the displacement vector $\Delta\mathbf{P}_{p,3D}(t) = \mathbf{P}_{p,3D}(t) - \check{\mathbf{P}}_{p,3D}(t)$ and the mean motion $\bar{\mathbf{x}}_k$ of grid point k , the point displacement is projected onto the mean motion direction assigned to k in order to correct for out-of-plane motion:

$$\Delta\hat{\mathbf{P}}_{p,3D}(t) = \cos(\alpha) \frac{\|\Delta\mathbf{P}_{p,3D}(t)\| \bar{\mathbf{x}}_k}{\|\bar{\mathbf{x}}_k\|}, \quad (5.7)$$

with α being the angle between $\Delta\mathbf{P}_{p,3D}(t)$ and $\bar{\mathbf{x}}_k$. The ultrasound motion vectors $\Delta\hat{\mathbf{P}}_{p,3D}(t)$ are used as sparse observations for the prediction of the complete motion of the liver.

5.2.3 Spatial and temporal prediction

Spatial prediction

At each time step t , we estimate the motion vector $\mathbf{x}_s(t)$ from the sparse observations $\mathbf{r} := \Delta\hat{\mathbf{P}}_{s,3D}(t)$. This is achieved following the Bayesian approach of [Blanz and Vetter 2002], where the most probable model reconstruction from partial data is computed through a maximum *a posteriori* estimation of the model coefficients $\tilde{\mathbf{c}}$. More details are available in [Preiswerk *et al.* 2013, Blanz and Vetter 2002]. In short, this amounts to minimizing the cost function

$$E = \|\mathbf{Q}\tilde{\mathbf{c}} - \mathbf{r}\| + \eta \|\tilde{\mathbf{c}}\|, \quad (5.8)$$

where $\mathbf{Q} = \mathbf{L}\mathbf{U}\text{diag}(\sigma_i)$, $\mathbf{r} = \mathbf{L}\mathbf{x}$, \mathbf{L} is a linear transformation relating the motion vector to the partial observation \mathbf{r} , and η is a weight factor. After SVD $\mathbf{Q} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, with $\mathbf{W} = \text{diag}(w_i)$ and $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$, and the solution to the Bayes problem is given by [Blanz and Vetter 2002]:

$$\tilde{\mathbf{c}}_s(t) = \mathbf{V} \cdot \text{diag}\left(\frac{w_i}{w_i^2 + \eta}\right) \cdot \mathbf{U}^T \cdot \Delta\hat{\mathbf{P}}_{s,3D}(t). \quad (5.9)$$

Finally, the deviation of the liver shape $\mathbf{l}_s(t)$ from its exhalation position $\check{\mathbf{l}}_s$ is computed as follows:

$$\mathbf{l}_s(t) = \check{\mathbf{l}}_s + \text{diag}(\sigma_i)\mathbf{U}\tilde{\mathbf{c}}_s(t) + \bar{\mathbf{x}}. \quad (5.10)$$

For efficiency, the data is reconstructed considering only the first eigenvalues that correspond to the 95% of the cumulative sum of λ_i .

Temporal prediction

The treatment system has a lag time to process input signals and adjust the treatment beam to the current target position. Therefore at time t we need to estimate the target position in the future $t + \Delta t$. Firstly, the temporal prediction of the partial observation is computed (from $\Delta\hat{\mathbf{P}}_{s,3D}(t)$ to $\Delta\tilde{\mathbf{P}}_{s,3D}(t + \Delta t)$). Then the spatial prediction of the motion vector $\mathbf{x}_s(t + \Delta t)$ is computed from the estimated $\Delta\tilde{\mathbf{P}}_{s,3D}(t + \Delta t)$ instead of $\Delta\hat{\mathbf{P}}_{s,3D}(t)$, as described in the previous section.

Various methods for temporal prediction of motion traces have been proposed [Verma *et al.* 2011]. Here, an artificial neural network (ANN) [Sharp *et al.* 2004] is used to generate the output $\Delta\tilde{\mathbf{P}}_{s,3D}(t + \Delta t)$. The ANN uses an input signal of 3 states, corresponding to the past two time states and the current state [$\Delta\hat{\mathbf{P}}_{s,3D}(t - 2)$, $\Delta\hat{\mathbf{P}}_{s,3D}(t - 1)$, $\Delta\hat{\mathbf{P}}_{s,3D}(t)$], and one hidden layer of 10 neurons [Preiswerk *et al.* 2013]. A schematic representation of the ANN is illustrated in Figure 5.5. The input

time states I_i , $i = [1 \dots 3]$, corresponding to $\Delta\hat{\mathbf{P}}_{s,3D}(t-3+i)$, are connected to each node of the hidden layer

$$H_j = \frac{1}{1 + e^{-\sum_{i=1}^3 w_{i,j} I_i}}, \quad (5.11)$$

with $j = [1 \dots 10]$. Then, the output status, corresponding to $\Delta\tilde{\mathbf{P}}_{s,3D}(t + \Delta t)$, is computed from the hidden layer $O = \sum_{j=1}^{10} v_j H_j$. The weights $w_{i,j}$ and v_j of the network are trained using the Levenberg-Marquardt backpropagation algorithm [Hagan and Menhaj 1994] on the 2D ultrasound tracking results $\mathbf{P}_{p,train}$ from a training phase of 1 min, prior to the simultaneous MR/ultrasound sequence.

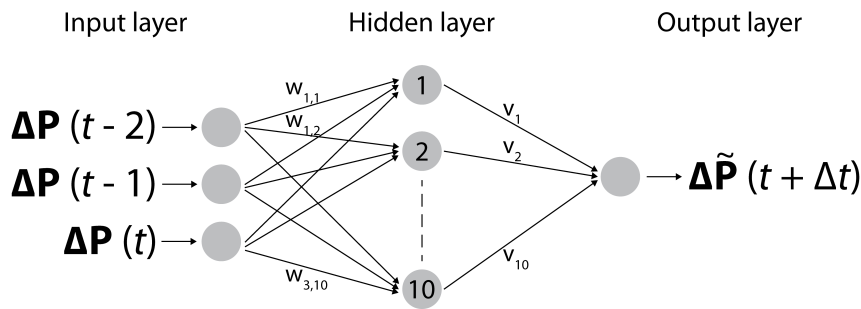


Figure 5.5: Schematic overview of the artificial neural network with trained weights w and v for generating the temporal prediction. Adapted from [Preiswerk *et al.* 2013].

Evaluation

Quantitative evaluation of the prediction was done with respect to 3D motion vectors from the ground truth $\mathbf{x}_s(t)$. In details, at each time t of the reconstructed sequence, The distance between the predicted displacement and the ground truth is computed at each model point k . The prediction error is therefore defined as the mean distance for all subject over the complete sequences. We also report the range of the mean distance resulted from each subject.

5.3 Results

No significant mutual interferences were detected during the simultaneous acquisition of ultrasound and MRI, with MR sequences that are commonly used in clinical routine, see Section 5.1.1. The SNR of the MR images dropped from 64.08 (ultrasound system turned off) to 58.60 (on) for the T1 VIBE, and from 44.08 (ultrasound off) to 44.85 (on) for the 4D sequence, corresponding to a SNR loss of 8.55% and 1.75%, respectively. The ultrasound image quality was visually invariant to the MR

system status and to the position of the ultrasound transducer inside or outside the magnet bore. The electronic noise RMS in the selected ROI of the second harmonic ultrasound images increased by 1.5% during MRI shimming and by 4.8% during 4D MRI. Further results are provided in [Petrusca *et al.* 2013].

Visual assessment of the multimodal registration confirmed accurate spatial matching between the 2D ultrasound and the corresponding MRI plane acquired during breath-hold. An example of the co-registration is shown in Figure 5.3.

A total of 49 points for the 6 ultrasound sequences were tracked, see Figure 5.4. Quantitative evaluation was performed on 11 points (vessel centers) including all the 6 sequences, see Section 3.2 and Figure 3.1 (last two rows). Ultrasound tracking achieved an accuracy of 0.80 ± 0.37 mm (1.40 mm) on average (MTE \pm STD (95th percentile of TE)), with a MTE_p range of [0.34, 1.24] mm.

For spatio-temporal prediction, Δt was set to 200 ms. The regularization factor η in Eq. 5.9 was experimentally determined and set to 1.0 (3.0) for spatial (spatio-temporal) prediction. Both spatial and spatio-temporal prediction were evaluated for the first block of simultaneously acquired images (\sim 5 min 30 sec). Results are summarized in Table 5.1. For comparison, both the mean breathing amplitude and the reconstruction error are listed. In the latter the spatial prediction is computed replacing $\Delta \hat{\mathbf{P}}_{p,3D}(t)$ with the actual 3D motion vectors from the ground truth $\mathbf{x}_s(t)$. On average, the use of 2D ultrasound surrogates for 3D spatial prediction increases the error by 56%, i.e. by 0.9 mm. While for 5 subjects the mean prediction errors are acceptable (<3.0 mm spatial, 3.3 mm spatio-temporal), subject number 4 shows 3.3 mm (4.9 mm) average spatial (spatio-temporal) error, and 2.3 mm mean reconstruction error, with respect to a mean motion amplitude of 7.7 mm, see Table 5.1. More detailed results can be found in [Preiswerk *et al.* 2013].

5.4 Conclusion

Dynamic acquisition of simultaneous ultrasound and MR images in the abdomen was achieved by integrating clinical devices and was used for validating a novel method to predict respiratory liver motion. This method computes a spatial or a spatio-temporal prediction of the organ motion only from sparse points tracked in the ultrasound images. A potential application is the ultrasound/MR hybrid guidance of HIFU therapies in the liver. The proposed approach is completely non-invasive and predicts the motion of the liver accurately within clinical requirements [Kennedy *et al.* 2004, Shirato *et al.* 2007] over long time scales.

Further improvements could be envisioned from enhanced handling of patient motion. Even if a comfortable position is ensured, subjects often move slightly over the course of the imaging/treatment session. Therefore, both the subject and the

Table 5.1: Mean prediction error (in mm) for 5.5 min continuous **spatial** motion prediction (forth column), **spatio-temporal** motion prediction (last column), w.r.t. breathing amplitude (second column) and reconstruction error, i.e. spatial motion prediction from 3D ground truth (third column).

Subject	Motion amplitude	Reconstruction error	Spatial prediction	Spatio-temporal prediction
1	6.7	1.2	2.0	2.5
2	9.5	1.6	2.9	2.9
3	10.4	1.3	2.1	2.6
4	7.7	2.3	3.3	4.9
5	6.7	1.5	1.8	2.2
6	15.9	1.9	3.0	3.3
Mean	9.5	1.6	2.5	3.1

transducer can be tracked to detect deviations from their initial positions, e.g. using optical markers. Additionally, the co-registration of MR and ultrasound images should not be limited to the pre-treatment breath-hold phase, but could be automated when the transducer is tracked.

In this Chapter, I described the system developed for acquiring simultaneous ultrasound and MR images for hybrid MR/USgHIFU treatment. In addition, I showed the feasibility of using the tracking results from the ultrasound images for predicting the motion of the region under treatment. In Chapter 6, I describe an alternative and unconventional method for transferring information between MR and ultrasound images for predicting the breathing motion. This method does not require multimodal image registration and is based on slow feature analysis to capture the breathing patterns in each image modality.

6

Alternative liver motion prediction

In Chapter 5, I described a novel clinical scenario for HIFU treatment where ultrasound and MR images are acquired simultaneously prior to and during treatment. Tracking structures on the ultrasound sequences (Chapters 3 and 4) is used for providing surrogate measures of the tumor motion, which often cannot be observed in real-time during therapy. Tumor motion prediction is based on a statistical motion model of the liver.

Specifically, a statistical motion model of the liver [von Siebenthal *et al.* 2007b, Preiswerk *et al.* 2013] was used to learn the motion pattern of the whole organ. The tumor motion is then predicted from the motion of a few landmarks in the organ. Vessel motion information was obtained from tracking ultrasound images, as described in Chapter 4. This approach requires that spatial correspondence between the surrogate and the organ motion model can be established, e.g. registration of the ultrasound and MR images.

Relating the ultrasound structures back to the pre-therapeutic MRI is not an easy task, especially for 2D images. A main challenge is to define an effective image similarity measure. For liver ultrasound images, approaches include gradient images [Milko *et al.* 2009], matching of extracted vessels [Porter *et al.*, Penney *et al.* 2004], simulations of ultrasound images from CT images [Wein *et al.* 2007, King *et al.* 2010] and hybrids [Lange *et al.* 2009]. Assuming the same breathing phase, alignments are often optimized with respect to a rigid transformation. The reported registration errors were on average in the range of 3.7 mm.

In this Chapter, I propose an unconventional approach for transferring information between multimodal images. It exploits the temporal commonality of multimodal images acquired from the same organ during free-breathing. Based on the hypothesis of the presence of a global temporal pattern in the data, it is possible to relate multimodal images of a moving organ without the need for explicit image registration. For a clinical setup without simultaneous MR and ultrasound acquisitions, I propose to acquire a short pre-therapeutic 4D MRI, and record ultrasound images

of the same organ during treatment (details in Section 6.1). The organ under investigation is then the same before and during therapy, while the position of the image acquisition and modality have changed, see Figure 6.1. Instead of multimodal image registration of two static images, I propose a novel approach, which is based on the observation, that the appearance changes of the organ in time are mainly due to a common cause, i.e. the breathing.

The method, described in Section 6.2, is based on extracting a low dimensional description of the image sequences, selecting the breathing signal for both modalities and finding the most similar sub-sequences for predicting image feature location. The approach was evaluated for 3 volunteers on sequences of 2D MRI and 2D ultrasound images of the liver acquired at different locations. Simultaneous acquisition of these images allowed for quantitative evaluation of the predicted MRI feature location with respect to ground truth data). The evaluation protocol and results are presented in Sections 6.3 and 6.4, respectively.

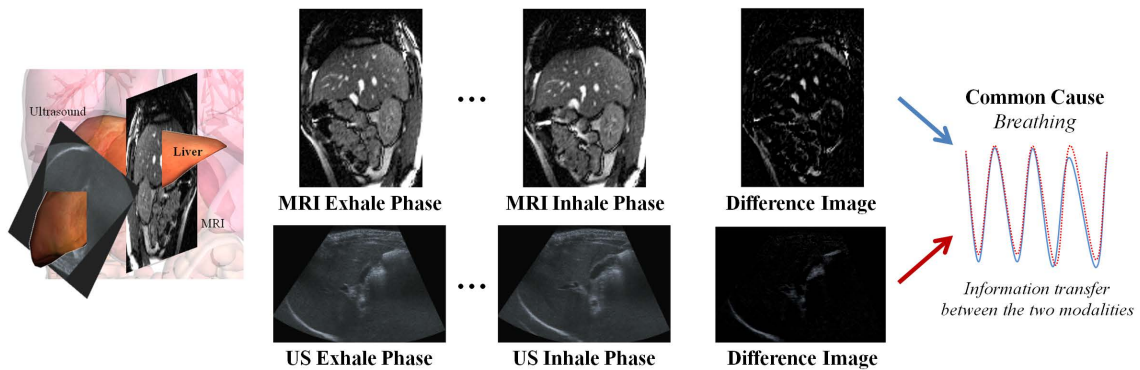


Figure 6.1: **Left:** MR and ultrasound images of the liver are acquired at different spatial locations (e.g. sagittal MRI, right subcostal oblique ultrasound). **Center:** typical liver images from simultaneously acquired 2D MR (top) and ultrasound (bottom) in end exhalation and end inhalation position are depicted. **Right:** from the images sequences, a description of the breathing as common cause is extracted, which is invariant of the image modality and slice location, in order to transfer image information between the two modalities.

6.1 Material

Ultrasound and MR images of the liver were simultaneously acquired for 3 volunteers during free-breathing at the Radiology Department of Geneva University Hospital [Petrusca *et al.* 2011]. This was accomplished by modifying the ultrasound equipment to be MR compatible (see Section 5.1). The sequence of 2D ultrasound images obtained from real time ultrasound second harmonic imaging (center frequency =

2.2 MHz) had a temporal resolution of 25 Hz and a spatial resolution of 0.6 mm. The acquisition sequence of MR images consisted of alternating between 2D navigator slices (at a fixed spatial position) and 2D data slices (at different locations to cover the liver) with a frequency of 2.45 Hz for the first two volunteers and 2.86 Hz for the third volunteer. From the aforementioned acquisition sequence, 4D MR images can be created by sorting the data slices based on the similarity of the enclosing navigator slices [von Siebenthal *et al.* 2007a], see Section 5.1.2. This allows for processing 2D navigators and extending the results to 3D MR volumes. MR navigator images had a spatial resolution of 2.34, 2.42 and 1.72 mm for the first, second and third volunteer respectively.

6.2 Method

The proposed method (illustrated in Figure 6.2) consists of 3 steps, namely

Dimensionality Reduction. Given an image sequence I^m from modality $m \in \{MRI, US\}$, I^m is embedded into a low-dimensional data representation \mathbf{S}^m using dimensionality reduction techniques.

Selection of Common Cause Signals. The components from \mathbf{S}^{MRI} and \mathbf{S}^{US} characterized by a global repetitive pattern related to breathing are selected.

Prediction of Feature Location. The prediction of image feature locations in one modality is achieved by evaluating the similarity between the breathing signals and selecting the associated images.

6.2.1 Dimensionality reduction

I aim to find an invariant description of the breathing signal with respect to the image modality and independently from an exact spatial correspondence. I assume that the breathing signal lies in a low-dimension manifold extracted from the data. Numerous methods are available for dimensionality reduction [Van der Maaten *et al.* 2009]. However, it is often unclear which method is applicable for a specific problem. I explored the following dimensionality reduction methods: Principle Component Analysis (PCA), Slow Feature Analysis (SFA) [Wiskott and Sejnowski 2002], Isomaps [Tenenbaum *et al.* 2000] and Laplacian eigenmaps [Belkin and Niyogi 2003]. Especially, SFA was proposed to capture the invariant temporal structures in the data by extracting slowly to quickly changing components [Wiskott and Sejnowski 2002].

For each modality $m \in \{MRI, US\}$, I acquire images characterized by D^m pixels at a frequency rate of f^m . I consider a time sequence of T^m images $I^m(\tau_i^m)$, with $\tau_i^m = \tau_0^m + i/f^m$ for $0 \leq i \leq L^m - 1$. Dimensionality reduction methods transform the initial dataset $I^m(\tau_i^m) \in \mathbb{R}^{D^m}$ into a new dataset $\mathbf{S}^m(\tau_i^m) = [\mathbf{s}_1^m(\tau_i^m); \dots; \mathbf{s}_d^m(\tau_i^m)] \in \mathbb{R}^d$, with $d \ll D^m$. I used the dimensionality reduction toolbox from Laurens van der Maaten [Van der Maaten *et al.* 2009] for manifold learning of Isomaps and the Laplacian eigenmaps, and Matlab for PCA and SFA.

PCA

Each image $I^m(\tau_i^m)$ is reshaped into a single column vector $\mathbf{x}^m(\tau_i^m) \in \mathbb{R}^{D^m}$. I calculate the mean with respect to time ($\bar{\mathbf{x}}^m$) and the covariance matrix $\mathbf{C}^{\hat{\mathbf{X}}} = \hat{\mathbf{X}}^m \hat{\mathbf{X}}^m$, where the i -th column of $\hat{\mathbf{X}}$ is equal to $\hat{\mathbf{x}}^m(\tau_i^m) = \mathbf{x}^m(\tau_i^m) - \bar{\mathbf{x}}$. I then solve the eigenproblem

$$\mathbf{C}^{\hat{\mathbf{X}}} \mathbf{w}_j^{PCA} = \lambda_j^{PCA} \mathbf{C}^{\hat{\mathbf{X}}} \mathbf{w}_j^{PCA}, \quad (6.1)$$

$\forall j \in [1, \dots, T^m]$. λ_j^{PCA} is the i -th largest eigenvalue and \mathbf{w}_j^{PCA} the corresponding eigenvector. In each j -th eigendirection I calculate the data projection $\mathbf{p}_j^{PCA} = (\mathbf{w}_j^{PCA})^T \hat{\mathbf{X}}^m$. I consider a set of d projections so that $\mathbf{s}_j^m(\tau_i^m) = \mathbf{p}_j^{PCA}(\tau_i^m)$ for $j \in [1, \dots, d]$.

SFA

I perform SFA on the PCA projections in order to extract signals with increasing temporal frequency from $\mathbf{P}^{PCA} = [\mathbf{p}_1^{PCA}; \dots; \mathbf{p}_J^{PCA}] \in \mathbb{R}^{J \times T^m}$, with $J \ll T^m$. Let $\hat{\mathbf{P}}^{PCA}$ be the zero-mean data matrix. I solve the generalized eigenproblem

$$\dot{\mathbf{C}}^{\hat{\mathbf{P}}^{PCA}} \mathbf{w}_j^{SFA} = \lambda_j^{SFA} \mathbf{C}^{\hat{\mathbf{P}}^{PCA}} \mathbf{w}_j^{SFA}, \quad (6.2)$$

where $\mathbf{C}^{\hat{\mathbf{P}}^{PCA}}$ and $\dot{\mathbf{C}}^{\hat{\mathbf{P}}^{PCA}}$ are the covariance matrix of the dataset and the temporal difference, respectively. The slowest components in $\hat{\mathbf{P}}^{PCA}$ are the projections onto the eigenvectors \mathbf{w}_j^{SFA} associated with the smallest λ_j^{SFA} and $\mathbf{s}_j^m(\tau_i^m) = (\mathbf{w}_j^{SFA})^T \hat{\mathbf{P}}^{PCA,m}$, $\forall j \in [1, \dots, d]$.

Isomap

Isomap is a nonlinear dimensionality reduction technique that attempts to preserve global properties of the original data in the low-dimensional representation, such as the pairwise geodesic distances between centered datapoints $\hat{\mathbf{x}}^m(\tau_i^m)$.

Every datapoint $\hat{\mathbf{x}}^m(\tau_i^m)$ is connected in a neighborhood graph G to its K -nearest neighbors (K -NN) in $\hat{\mathbf{X}}$ based on pairwise Euclidean distances

$d_{\hat{\mathbf{x}}^m}(i, j) = \|\hat{\mathbf{x}}^m(\tau_i^m) - \hat{\mathbf{x}}^m(\tau_j^m)\|$. As in [Van der Maaten *et al.* 2009], I choose $K = 12$. The shortest paths in G are computed using Dijkstra’s shortest-path algorithm [Dijkstra 1959], resulting in a matrix \mathbf{D}_G that contains the shortest path distances between all T^m datapoints. The low-dimensional embedding \mathbf{P}^{ISO} is constructed by minimizing the following function:

$$\phi(\mathbf{P}^{ISO}) = \|\varphi(\mathbf{D}_G) - \varphi(\mathbf{D}_{\mathbf{P}^{ISO}})\|_2 \quad (6.3)$$

where $D_{\mathbf{P}^{ISO} i, j} = d_{\mathbf{P}^{ISO}}(i, j)$ and $d_{\mathbf{P}^{ISO}}(i, j) = \|\mathbf{p}^{ISO}(\tau_i^m) - \mathbf{p}^{ISO}(\tau_j^m)\|$ are the Euclidean distances between the low-dimensional representations $\mathbf{p}^{ISO}(\tau_i^m)$ and $\mathbf{p}^{ISO}(\tau_j^m)$ of the paired datapoints $\hat{\mathbf{x}}^m(\tau_i^m)$ and $\hat{\mathbf{x}}^m(\tau_j^m)$, respectively. The operator $\varphi(\mathbf{D})$ is [Tenenbaum *et al.* 2000]

$$\varphi(\mathbf{D}) = -\mathbf{H}\mathbf{D}^2\mathbf{H}/2, \quad (6.4)$$

where \mathbf{D}^2 is the matrix of the squared distances, with $D_{i, j}^2 = d^2(i, j)$, and \mathbf{H} is the centering matrix (i.e. ensuring that the mean of $\varphi(\mathbf{D})$ is zero) with components $H_{i, j} = \delta_{i, j} - 1/T^m$. More details can be found in [Tenenbaum *et al.* 2000]. The low-dimensional embedding \mathbf{P}^{ISO} , which minimizes (6.3), is determined using eigen-decomposition of $\varphi(\mathbf{D}_G)$, as in (6.1) for the covariance matrix $\mathbf{C}^{\hat{\mathbf{x}}}$. Let λ_p^{ISO} be the p -th eigenvalue (sorted in descending order) and \mathbf{w}_p^{ISO} the corresponding eigenvector. I consider a set of d projections so that $\mathbf{s}_p^m(\tau_i^m) = \mathbf{p}_p^{ISO}(\tau_i^m) = (\mathbf{w}_p^{ISO})^T \hat{\mathbf{x}}(\tau_i^m)$ for $p \in [1, \dots, d]$.

Laplacian eigenmaps

Laplacian eigenmaps preserve local properties of the manifold. For all centered image vectors $\hat{\mathbf{x}}^m(\tau_i^m)$, $j \in [1, \dots, T^m]$, the Laplacian eigenmaps algorithm constructs a neighborhood graph G , in which each $\hat{\mathbf{x}}^m(\tau_i^m)$ is connected to its K -NN in $\hat{\mathbf{X}}$ in a weighted manner. As for Isomaps, I select $K = 12$. The weights between $\hat{\mathbf{x}}^m(\tau_i^m)$ and its K neighbors $\hat{\mathbf{x}}^m(\tau_j^m)$ are defined by

$$w_{i, j} = e^{-\|\hat{\mathbf{x}}^m(\tau_i^m) - \hat{\mathbf{x}}^m(\tau_j^m)\|^2/2}, \quad (6.5)$$

resulting in a sparse adjacency matrix \mathbf{W} , with $W_{i, j} = w_{i, j}$, as weights decay exponentially and weights outside the neighborhood are zero. The low-dimensional representation \mathbf{P}^{LE} is obtained by minimizing the following cost function:

$$\phi(\mathbf{P}^{LE}) = \sum_{i, j} (\mathbf{p}^{LE}(\tau_i^m) - \mathbf{p}^{LE}(\tau_j^m))^2 w_{i, j}. \quad (6.6)$$

Let \mathbf{M} be a diagonal matrix with $M_{i, i} = \sum_j w_{j, i}$ (approximating the number of edges attached), then the Laplacian matrix is given $\mathbf{L} = \mathbf{M} - \mathbf{W}$. It can be proved that $\phi(\mathbf{P}^{LE}) = 2(\mathbf{P}^{LE})^T \mathbf{L} \mathbf{P}^{LE}$, see [Belkin and Niyogi 2003, Van der Maaten *et al.*

2009]. The minimization problem of $\phi(\mathbf{P}^{LE})$ is then reduced to the solution of the generalized eigenproblem (similar to (6.2))

$$\mathbf{L}\mathbf{w}_p^{LE} = \lambda_p^{LE}\mathbf{M}\mathbf{w}_p^{LE}. \quad (6.7)$$

The p -th data projection into the low-dimensional representation is given by $\mathbf{p}_p^{LE} = (\mathbf{w}_p^{LE})^T \hat{\mathbf{x}}^m$. I select the first d components for $\mathbf{s}_p^m(\tau_i^m) = \mathbf{p}_p^{LE}(\tau_i^m)$ for $p \in [1, \dots, d]$.

For all the aforementioned strategies, the individual components $\mathbf{s}_i^m(\tau_i^m)$ are then normalized to zero mean and a standard deviation of one. The signals are further normalized to a common sampling frequency $f = \min\{f^{MRI}, f^{US}\}$ (in the specific case $f = f^{MRI}$). I denote the normalized low-dimensional descriptions as $\hat{\mathbf{S}}^m(t) = [\hat{\mathbf{s}}_1^m(t); \dots; \hat{\mathbf{s}}_d^m(t)]$, with $t = \tau_0^m + i/f$ and $0 \leq i \leq T^m - 1$. Figure 6.2a shows examples of $\hat{\mathbf{S}}^{US}$ and $\hat{\mathbf{S}}^{MRI}$.

6.2.2 Selection of common cause signals

I aim to select corresponding individual components in $\hat{\mathbf{S}}^{US}$ and $\hat{\mathbf{S}}^{MRI}$. Relying on the statistical assumption that a similar breathing pattern is observed in the two modalities, I want to find the components which indicate the breathing as common cause. Hence, I employ frequency analysis (i.e. Fast Fourier Transform) to the 1D eigenmodes of both modalities. Then I choose the signals characterized by a power spectral density maximum at a frequency in the range of 0.15-0.40 Hz (common breathing frequency window). I denote the selected breathing signals as $\hat{\mathbf{s}}_B^m(t)$ for modality $m \in \{US, MRI\}$.

6.2.3 Prediction of feature location

After normalization and selection of the breathing signals, I look for their similarity. I select the most similar short sequences in $\hat{\mathbf{S}}_B^{MRI}$ from a finite number of observations in $\hat{\mathbf{S}}_B^{US}$, in order to predict MR navigators. My observation consists of a short sequence $[\hat{\mathbf{s}}_B^{US}(t - (N - 1)/f), \dots, \hat{\mathbf{s}}_B^{US}(t/f)]$ from the selected signal $\hat{\mathbf{s}}_B^{US}$, in order to take into account the breathing history (e.g. exhalation, inhalation). I calculate the K -NN in $\hat{\mathbf{S}}_B^{MRI}$ to this sequence. In detail, I find t_k^{MRI} so that, for a given t^{US} , the distance

$$d_k = \sqrt{\sum_{n=0}^{N-1} [\hat{\mathbf{s}}_B^{US}(t^{US} - n/f) - \hat{\mathbf{s}}_B^{MRI}(t_k^{MRI} - n/f)]^2}, \quad (6.8)$$

$\forall 1 \leq k \leq K$ is minimized, see Figure 6.2b. I considered $N = 4$ and $K = 3$. Having found the K -NN MR short sequences from the ultrasound sequence, I selected the associated MR navigator images, see Figure 6.2c.

I evaluated this approach on vessel center locations on the MR navigators. Specifically, vessels with cross-sectional cuts were semi-automatically segmented from MR navigator images¹ and their center locations $\mathbf{v}(t)$ extracted. For each of the N frames, I linearly interpolate the K center locations, to get the prediction of the vessel center position (see Figure 6.2d):

$$\bar{\mathbf{v}}^{MRI}(t^{US} - n/f) = \sum_{k=1}^K \left(1 - \frac{d_k}{\sum_k d_k}\right) \mathbf{v}^{MRI}(t_k^{MRI} - n/f). \quad (6.9)$$

6.3 Prediction evaluation

I used data from 3 volunteers (Vol.1 - Vol.3). For each volunteer I considered one ultrasound sequence and two MRI sequences and differentiate between two setups. For *Seq1* the ultrasound and MR sequences were acquired simultaneously. The MR images of this sequence define the ground truth (GT) data, to validate the method. For *Seq2* I use MR images acquired minutes later and the ultrasounds from *Seq1*. This setup was used to evaluate the performance for an independent MR set, similarly to the targeted clinical application.

I also assessed the effect of replacing $\hat{\mathbf{s}}_B^{US}$ with the main diaphragm motion (first PCA eigenmode of a point displacement at the diaphragm). The diaphragm displacement was obtained from affine registration of ultrasound liver image regions covering the diaphragm. For registration details, refer to Section 3.3.

I quantified the prediction error by computing the mean of the Euclidean distance between the ground truth vessel center location \mathbf{v}^{GT} and the corresponding location on the predicted MR images $\bar{\mathbf{v}}^{MRI}$, see Figure 6.2d:

$$\bar{E}_{t^{US}} = \frac{1}{N} \sum_{n=0}^{N-1} \left\| \mathbf{v}^{GT}(t^{US} - n/f) - \bar{\mathbf{v}}^{MRI}(t^{US} - n/f) \right\|. \quad (6.10)$$

I summarized the results by the mean and standard deviation (STD) of $\bar{E}_{t^{US}} \forall t^{US}$.

6.4 Results

Table 6.1 lists the mean prediction error \bar{E} obtained from each volunteer and tested method. It can be observed that SFA achieved on average lowest errors closely

¹Sagittal navigator images were acquired. Hence out-of-plane motion is expected to be minor (2 mm) compared to the captured in-plane motion (15 and 8 mm) [Nguyen *et al.* 2008].

followed by PCA. Average errors increased by 1 mm (0.6 mm) for diaphragm tracking for *Seq1* (*Seq2*). When assuming no respiratory motion (using the mean position of \mathbf{v}^{GT} over *Seq1* as prediction), I obtained an average prediction error over all 3 volunteers of 5.94 ± 2.11 mm (7.49 ± 3.55 mm) for *Seq1* (*Seq2*).

For the best approach (SFA), I evaluated the linear dependence between the selected breathing signals $\hat{\mathbf{s}}_B^{US}$ and $\hat{\mathbf{s}}_B^{MRI}$, and between $\hat{\mathbf{s}}_B^{MRI}$ and \mathbf{v}^{GT} . I compared the correlation with that of the common prediction approach, i.e. diaphragm tracking to \mathbf{v}^{GT} , see Table 6.2. SFA provides on average higher correlations, supporting the better performance of SFA over diaphragm tracking for *Seq1* (Table 6.1).

Table 6.1: Mean \pm STD of the prediction error (in mm), see Eq. (6.10) for all the proposed approaches, namely Isomap (ISO), Laplacian eigenmaps (LE), Principal Component Analysis (PCA) and Slow Feature Analysis (SFA). Best results are in bold face and second best are underlined.

		Proposed approach				
		Tracking	ISO	LE	PCA	SFA
<i>Seq1</i>	Vol.1	3.91 ± 1.94	2.86 ± 1.66	3.73 ± 1.13	<u>1.89 ± 0.98</u>	1.72 ± 0.93
	Vol.2	3.00 ± 1.97	4.36 ± 2.52	3.08 ± 1.76	<u>2.98 ± 1.72</u>	2.89 ± 1.63
	Vol.3	4.02 ± 2.85	4.64 ± 3.45	<u>3.39 ± 2.45</u>	3.57 ± 2.45	3.35 ± 2.37
	Mean	3.65 ± 2.25	3.96 ± 2.54	3.40 ± 1.78	<u>2.81 ± 1.72</u>	2.65 ± 1.65
<i>Seq2</i>	Vol.1	5.87 ± 3.60	5.14 ± 2.38	6.05 ± 2.59	<u>3.58 ± 1.55</u>	3.36 ± 1.50
	Vol.2	<u>3.99 ± 1.33</u>	3.31 ± 1.73	6.40 ± 3.41	4.84 ± 2.64	4.60 ± 2.57
	Vol.3	<u>4.55 ± 2.69</u>	6.98 ± 5.07	10.83 ± 6.77	4.32 ± 2.64	4.67 ± 2.88
	Mean	4.80 ± 2.54	5.14 ± 2.73	7.76 ± 4.26	<u>4.25 ± 2.28</u>	4.21 ± 2.31

Table 6.2: Correlation between extracted signals and \mathbf{v}^{GT} .

		SFA		
		Tracking - \mathbf{v}^{GT}	$\hat{\mathbf{s}}^{US}$ - $\hat{\mathbf{s}}^{MRI}$	$\hat{\mathbf{s}}^{MRI}$ - \mathbf{v}^{GT}
Vol.1	30.7 %	92.3 %	87.6 %	
Vol.2	84.4 %	79.6 %	94.8 %	
Vol.3	85.1 %	91.7 %	87.6 %	
Mean	66.7 %	87.9 %	90.0 %	

6.5 Conclusion

In this Chapter, I proposed an unconventional method for relating multimodal images sequences. It is based on representing the acquired data in a low-dimensional embedding, extracting common cause signals (e.g. breathing) from both image sequences, finding the most similar sub-sequences of these signals, and using the associated images in order to predict the location of image features. The method generates the prediction in a completely unsupervised manner.

Using this approach, I predicted the location of anatomical landmarks by relying on a temporal pre-therapeutical MRI sequence based on the observation of ultrasound images acquired during therapy with an average accuracy of 4.2 mm. The achieved performance is comparable to state-of-the-art methods (3.7 mm for multimodal registration [Penney *et al.* 2004, Wein *et al.* 2007, Lange *et al.* 2009], 1.4 mm for ultrasound tracking of the liver [Foroughi *et al.* 2006a, Harris *et al.* 2010]) while being less complex. Lowest mean errors were achieved when employing slow feature analysis. This demonstrates the advantage of explicitly using the temporal information stored in the data and supports previous investigations [Wiskott and Sejnowski 2002].

After these encouraging results, the method could be tested for predicting the displacement of landmarks in ultrasound images acquired during treatment from pre-therapeutic data, as in the application scenario described in Chapter 5. Potential improvements could be investigated when combining the presented approach with statistical breathing models [von Siebenthal *et al.* 2007b, Preiswerk *et al.* 2013]. Additionally, the method could be applied to 3D landmark prediction by using 4D MR techniques [von Siebenthal *et al.* 2007a].

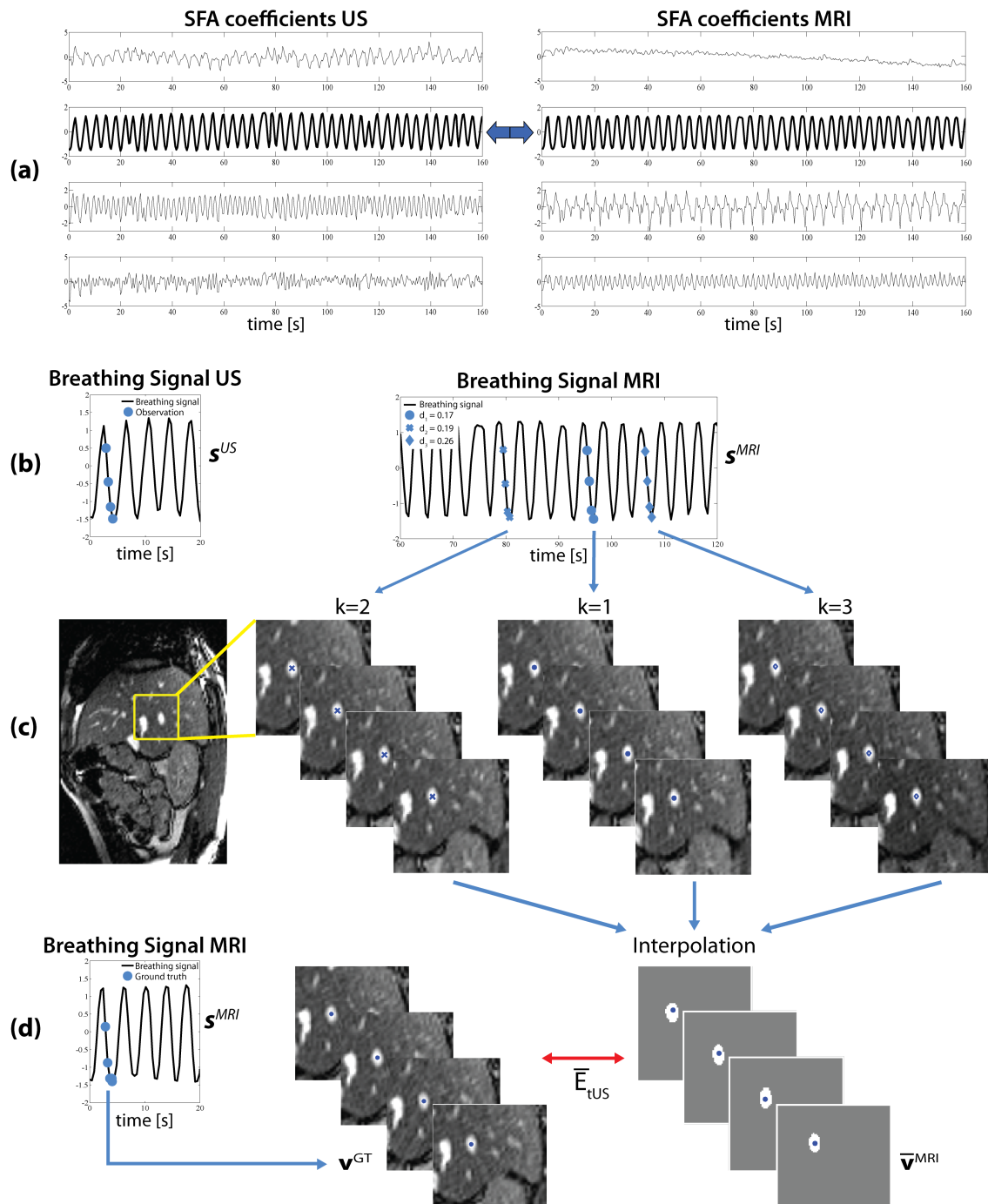


Figure 6.2: (a) Low-dimensional representation of the images and selection of the breathing signals ($\hat{\mathbf{s}}_B^{US}$, $\hat{\mathbf{s}}_B^{MRI}$). (b) From the observation of a short sequence of $\hat{\mathbf{s}}_B^{US}$, I find the most similar sequences in $\hat{\mathbf{s}}_B^{MRI}$. (c) I select the associated MR images and extract image information (vessel center locations). (d) For evaluation, I interpolate the vessel center locations ($\bar{\mathbf{v}}^{MRI}$) and calculate the prediction error \bar{E}_{tUS} w.r.t. ground truth (\mathbf{v}^{GT}).

7

Conclusion

In this thesis I proposed a novel framework for tracking ultrasound sequences of the liver under free breathing. The approximate periodicity of the breathing motion is exploited for learning the image appearance and corresponding motion behavior to allow fast and accurate computation of the displacement of points of interest. All the presented approaches were tested on long sequences (5 to 10 min) and validated with respect to multiple observers' manual annotation on a subset of images. An accuracy of below 1 mm was achieved. Applicability of the tracking method for a novel hybrid MR/USgHIFU treatment scenario was also tested. There model-based spatio-temporal prediction of the liver motion was driven by the ultrasound tracking results as surrogates. The specific contributions of each Chapter are described next. Thoughts on future research avenues conclude this Chapter.

7.1 Contributions

In Chapter 2, *Related works*, I provided an extensive literature review of the tracking methods developed for ultrasound time sequences. I have discussed the main strategies and listed the most important contributions to the topic. While only few methods have been already successfully applied in clinical practice, most still need improvements in accuracy, robustness and speed as well as proper validation. The most promising methods (intensity-based) for tracking long ultrasound sequences were selected for further investigations.

In Chapter 3, *Intensity-based tracking for long sequences*, I compared two intensity-based frameworks (affine registration and a scale-adaptive block-matching algorithm) for vessel tracking in long ultrasound sequences (5-10 min each) of the liver of 9 volunteers under free breathing. To my knowledge, this is the first evaluation for tracking such long ultrasound sequences. Using affine registration, I achieved a mean accuracy of 0.90 mm. Such performance improves the state-of-the-art in 2D ultrasound tracking of the human liver. However, it requires manual reinitialization when frame drop outs occur and has high computational costs. The proposed

block-matching algorithm is computationally faster and potentially real-time. Yet its accuracy is lower (2.18 mm), due to a suboptimal strategy for temporal realignment of the tracking.

In Chapter 4, *Learning-based real-time tracking*, I proposed a novel, simple and robust learning-base framework for tracking ultrasound sequences. The method learns the relationship between image appearance and feature displacements, in order to speed-up image registration or to allow frequent reinitialization of the scale-adaptive block-matching algorithm. The method exploits the redundancy of the images in a repetitive motion scenario while still adapting to irregularities. The method was evaluated on the long ultrasound sequences of the liver described in Chapter 3. A mean tracking accuracy of 0.84 mm and 0.97 mm was achieved for the learning-based affine registration and block-matching algorithm, respectively. The latter copes better with larger image variations and ensures an overall faster performance, as fewer registrations due to outliers are required.

In Chapter 5, *Simultaneous MR- and ultrasound-guided radiation therapy*, I show the feasibility of using the ultrasound tracking results as input for a model-based spatio-temporal prediction of the target motion during a novel hybrid MR/USgHIFU treatment system. Results show a clinically acceptable accuracy for the spatial prediction (2.5 mm), when compared to the clinical margins of error, which are generally included during treatment planning. Yet, a real-time implementation for the prediction is missing.

In Chapter 6, *Alternative liver motion prediction*, I explored the possibility of transferring motion information between image modalities, without using image registration, in order to predict the displacements of points of interest in one modality from the observation of the other. The method represents the acquired images in a low-dimensional embedding via slow feature analysis, extracts the breathing signal from both image sequences, finds the most similar sub-sequences of these signals, and uses the associated images to predict the location of image features in a completely unsupervised manner. The location of anatomical landmarks is predicted from a pre-therapeutical MRI sequence and from the observation of ultrasound images acquired during therapy. An average accuracy of 4.21 mm is obtained, comparable to state-of-the-art methods for multimodal registration and ultrasound tracking of the liver, while being less complex.

7.2 Future work

Real-time

The accuracy and robustness of learning-based tracking method over long ultrasound sequences is already sufficient for the clinical application. Yet real-time per-

formance during the whole application phase is missing. This problem could be solved by optimizing the current implementation for run-time. Preliminary tests on the block-matching algorithm support this expectation. Additionally, methods that better adapt to and compensate for image variations should be preferred, as fewer image registration are required, which are computationally expensive and hence require interruption of the treatment.

Learning

In the training phase of the proposed learning-based tracking, the displacements of points of interest are computed by affine registration. In the current strategy, the registration of each image to the initial frame is initialized by the results from the previous frame. Unfortunately, frame drop outs might occur in an unpredictable way and manual initialization of the registration is necessary. Online learning of the image appearance during the training phase could be used to find the most similar images, and therefore to automatically select the best transformations for initializing the registration algorithm.

The image appearance is learned using PCA. Other unsupervised learning techniques, such as non-linear dimensionality reduction methods, could be tested. Bayesian frameworks have been successfully used in feature tracking, but were not considered in this thesis, mainly due to their higher complexity.

The presented work is generally beneficial for tracking approximately repetitive motions. While in this thesis I focused on respiratory motion in the liver, another potential application could be tracking cardiac sequences. The proposed learning-based tracking could be applied to other image modalities. Experiments were already successfully carried out for tracking 2D MR sequences of the liver [De Luca *et al.* 2012]. Finally this method could be especially beneficial for 3D tracking, considerably reducing its computational cost.

Hybrid tracking strategy

In Chapter 4, I proposed two learning-based tracking strategies. While the learning-based affine registration achieves the best accuracy, the learning-based scale-adaptive block matching algorithm can better compensate for outliers. In future work these two strategies could be fused. In detail, the inlier treatment of the learning-based affine registration (interpolation of associated displacements) could be combined with the outlier strategy of the learning-based block matching (block-matching with previous frame if similar enough). This strategy could allow for prediction of the displacements of points of interest without interruptions.

Diaphragm tracking

In this thesis I mainly focused on the tracking of vessels, apart from including diaphragm tracking using learning-based affine registration for liver motion prediction (see Chapter 5). However, a dominant image structure, generally at least partially visible on liver ultrasound images and correlated with respiratory motion is the diaphragm. Yet local intensity-based tracking methods might fail due to the aperture problem. Therefore larger regions for matching are necessary to achieve globally consistent alignments. At the same time the computational cost increases.

A possible approach, building upon to the learning-based block-matching algorithm presented in Chapter 4, could include a more sophisticated training phase, where the relationship between the displacement of multiple features in the images, e.g. vessels and points on the diaphragm, is additionally learned. This information could be used during the application phase to allow for directional search of the best match of points on the diaphragm, where a bigger area is necessary to avoid the aperture problem. In detail, the displacement vectors of diaphragm points could be searched in two directions, one perpendicular to the diaphragm line and the other given by the main motion direction. The latter could be obtained by computing the displacement vectors of the other features in the same image and of all features in the most similar images in the training set, and by interpolating along these vectors.

Feature-based tracking methods, such as contour tracking, are generally faster. Yet robust detection and extraction of the contour-based features is difficult. Contours can provide reliable information for the shape of the diaphragm, but lack regional information from the image and may lose stability when the boundary information is not strong enough, e.g. in the presence of noise, shadows and acoustic occlusions. Therefore I would recommend to follow an hybrid approach that fuses boundary- and region intensity-based information.

MR/USgHIFU

Multimodal registration involving ultrasound and MR images has not been studied in this thesis. During hybrid ultrasound and MR guidance, the exact spatial relationship of the two image modalities with respect to the 3D MRI coordinate system should be established automatically. Additional sensors for tracking devices in the MR scanner (e.g. Hall sensors) would facilitate this automatic registration and help detect changes in the ultrasound probe position. Moreover, patients might move during treatment. Hence co-registration should not be limited to the pre-treatment phase.

Bibliography

- [Abolmaesumi *et al.* 2000] P. Abolmaesumi, M. R. Sirouspour, and S. E. Salcudean. Real-time extraction of carotid artery contours from ultrasound images. In *Proc IEEE Int Symp Comput Based Med Syst*, pages 181–186, 2000. [2.2.4](#)
- [Akgul *et al.* 1998] Y. S. Akgul, C. Kambhamettu, and M. Stone. Extraction and tracking of the tongue surface from ultrasound image sequences. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pages 298–303, 1998. [2.2.1](#)
- [Akgul *et al.* 2000] Y. S. Akgul, C. Kambhamettu, and M. Stone. A task-specific contour tracker for ultrasound. In *Proc Workshop Math Methods Biomed Image Analysis*, pages 135–142, 2000. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Amin *et al.* 2008] V. Amin, L. Wu, T. Long, R. Roberts, S. McClure, and T. Ryken. Therapy planning and monitoring of tissue ablation by high intensity focused ultrasound (HIFU) using imaging and simulation. In *Conf Proc IEEE Eng Med Biol Soc*, pages 4471–4471, 2008. [1.1](#)
- [Anandan 1989] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int J Comput Vis*, 2:283–310, 1989. [2.2.1](#)
- [Angelova *et al.* 2007] D. Angelova, P. Konstantinova, and L. Mihaylova. Contour Tracking in 2D Images Using Particle Filtering. In *Proc Int Radar Symp*, 2007. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Arnold *et al.* 2011] P. Arnold, F. Preiswerk, B. Fasel, R. Salomir, K. Scheffler, and P. C. Cattin. 3D Organ Motion Prediction for MR-Guided High Intensity Focused Ultrasound. In *Med Image Comput Comput Assist Interv*, volume 6892 of *LNCS*, pages 623–630. Springer, 2011. [1.1](#)
- [Baraldi *et al.* 1996] P. Baraldi, A. Sarti, C. Lamberti, A. Prandini, and F. Sgallari. Evaluation of differential optical flow techniques on synthesized echo images. *IEEE Trans Biomed Eng*, 43(3):259–272, 1996. [2.2.1](#)
- [Barron *et al.* 1994] J. R. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Int J Comput Vis*, 12:43–77, 1994. [2.2.1](#)
- [Basarab *et al.* 2007] A. Basarab, W. Aoudi, H. Liebgott, D. Vray, and P. Delachartre. Parametric Deformable Block Matching for Ultrasound Imaging. In *Proc Int Conf Image Proc*, volume 2, pages 429–432, 2007. [2.2.2](#), [2.2.2](#)

- [Bashford and von Ramm 1996] G. R. Bashford and O. T. von Ramm. Ultrasound three-dimensional velocity measurements by feature tracking. *IEEE Trans Ultrason Ferroelectr Freq Control*, 43(3):376–384, 1996. [2.2.2](#)
- [Belkin and Niyogi 2003] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*, 15(6):1373–1396, 2003. [6.2.1](#), [6.2.1](#)
- [Bergen *et al.* 1992] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In G. Sandini, editor, *Comput Vis ECCV*, volume 588 of *LNCS*, pages 237–252. Springer, 1992. [2.2.1](#)
- [Björck 1996] A. Björck. *Numerical methods for least squares problems*. SIAM, 1996. [2.1.3](#)
- [Blake and Isard 1998] A. Blake and M. Isard. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer-Verlag, 1st edition, 1998. [2.1.1](#), [2.2.4](#), [2.4](#)
- [Blanz and Vetter 2002] V. Blanz and T. Vetter. Reconstructing the Complete 3D Shape of Faces from Partial Information. *Informationstechnik und Technische Informatik*, 44(6):295–302, 2002. [5.2.3](#), [5.2.3](#)
- [Bottou 2004] L. Bottou. Stochastic learning. In O. Bousquet and U. von Luxburg, editors, *Advanced Lectures on Machine Learning*, volume 3176 of *LNAI*, pages 146–168. Springer-Verlag, 2004. [2.1.3](#)
- [Boukerroui *et al.* 2003] D. Boukerroui, J. A. Noble, and M. Brady. Velocity Estimation in Ultrasound Images: A Block Matching Approach. In C. Taylor and J. A. Noble, editors, *Inf Process Med Imaging*, volume 2732 of *LNCS*, pages 586–598. Springer, 2003. [2.2.1](#), [2.2.1](#), [3.4](#)
- [Brox *et al.* 2004] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High Accuracy Optical Flow Estimation Based on a Theory for Warping. In T. Pajdla and J. Matas, editors, *Comput Vis ECCV*, volume 3024 of *LNCS*, pages 25–36. Springer, 2004. [2.2.1](#)
- [Burckhardt 1978] C. B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Trans Son Ultrason*, 25(1):1–6, 1978. [2.2.2](#)
- [Burt and Adelson 1983] E. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans Commun*, 31:532–540, 1983. [2.2.1](#)
- [Buvat *et al.* 1999] I. Buvat, V. Chameroy, F. Aubry, M. Pelegrini, G. El Fakhri, C. Huguenin, H. Benali, A. Todd-Pokropek, and R. Di Paola. The need to develop guidelines for the evaluation of medical image processing procedures. In *Proc SPIE*, volume 3661, pages 1466–1477, 1999. [2.3](#)
- [Byram *et al.* 2010] B. Byram, G. Holley, D. Giannantonio, and G. Trahey. 3-D phantom and in vivo cardiac speckle tracking using a matrix array and raw echo

- data. *IEEE Trans Ultrason Ferroelectr Freq Control*, 57(4):839–854, 2010. [2.2.2](#), [2.2.2](#), [2.2.2](#), [3.4](#)
- [Byrd *et al.* 1995] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J Sci Comput*, 16(5):1190–1208, 1995. [2.1.3](#)
- [Chalana *et al.* 1996] V. Chalana, D. T. Linker, D. R. Haynor, and Y. Kim. A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Trans Med Imaging*, 15(3):290–298, 1996. [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#)
- [Chandrana *et al.* 2011] C. Chandrana, P. Bevan, J. Hudson, I. Pang, P. Burns, D. Plewes, and R. Chopra. Development of a platform for co-registered ultrasound and MR contrast imaging in vivo. *Phys Med Biol*, 56(3):861–877, 2011. [5.1.3](#)
- [Chen *et al.* 1991] E. J. Chen, I. A. Hein, J. B. Fowlkes, R. S. Adler, P. L. Carson, and W. D. J. O’Brien. A comparison of the motion tracking of 2-D ultrasonic B-mode tissue images with a calibrated phantom. In *Proc IEEE Ultrason Symp*, volume 2, pages 1211–1214, 1991. [2.2.3](#), [2.2.3](#)
- [Chen *et al.* 1992] E. J. Chen, W. K. Jenkins, and W. D. J. O’Brien. The accuracy and precision of estimating tissue displacements from ultrasonic images. In *Proc IEEE Ultrason Symp*, volume 2, pages 1069–1072, 1992. [2.2.2](#), [2.2.2](#)
- [Chen *et al.* 1994] E. J. Chen, W. K. Jenkins, and W. D. J. O’Brien. The impact of various imaging parameters on ultrasonic displacement and velocity estimates. *IEEE Trans Ultrason Ferroelectr Freq Control*, 41(3):293–301, 1994. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Chen *et al.* 1995] E. J. Chen, W. K. Jenkins, and W. D. J. O’Brien. Performance of ultrasonic speckle tracking in various tissues. *J Acoust Soc Am*, 98(3):1273–1278, 1995. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Chen *et al.* 2005] X. Chen, H. Xie, R. Erkamp, K. Kim, C. Jai, J. M. Rubin, and M. O’Donnell. 3-D Correlation-Based Speckle Tracking. *Ultrasonic Imaging*, 27(1):21–36, 2005. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Chen *et al.* 2006] C. J. Chen, R. F. Chang, W. K. Moon, D. R. Chen, and H. K. Wu. 2-D ultrasound strain images for breast cancer diagnosis using nonrigid subregion registration. *Ultrasound Med Biol*, 32(6):837–846, 2006. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Cifor *et al.* 2012] A. Cifor, L. Risser, D. Chung, E. M. Anderson, and J. A. Schnabel. Hybrid feature-based Log-Demons registration for tumour tracking in 2-D liver ultrasound images. In *Proc IEEE Int Symp Biomed Imaging*, pages 724–727, 2012. [2.2.5](#), [2.4](#)
- [Cohen and Cohen 1993] L. D. Cohen and I. Cohen. Finite Element Methods for Active Contour Models and Balloons for 2D and 3D Images. *IEEE Trans Pattern Anal Mach Intell*, 15(11):1131–1147, 1993. [2.1.1](#), [2.2.4](#)

- [Cohen and Dinstein 2002] B. Cohen and I. Dinstein. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognit*, 35(2):455–463, 2002. [2.1.1](#), [2.2.2](#), [2.2.2](#), [2.3](#)
- [Comaniciu *et al.* 2004] D. Comaniciu, X. S. Zhou, and S. Krishnan. Robust real-time myocardial border tracking for echocardiography: an information fusion approach. *IEEE Trans Med Imaging*, 23(7):849–860, 2004. [2.1.1](#), [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Comaniciu 2003] D. Comaniciu. Nonparametric information fusion for motion estimation. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, volume 1, pages 59–66, 2003. [2.2.4](#)
- [Compas *et al.* 2011] C. B. Compas, B. A. Lin, S. Sampath, A. J. Sinusas, and J. S. Duncan. Combining shape and speckle tracking for deformation analysis in echocardiography using radial basis functions. In *Proc IEEE Int Symp Biomed Imaging*, pages 1322–1325, 2011. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Compas *et al.* 2012] C. B. Compas, E. Y. Wong, X. Huang, S. Sampath, B. A. Lin, X. Papademetris, K. Thiele, D. P. Dione, A. J. Sinusas, M. O’Donnell, and J. S. Duncan. A combined shape tracking and speckle tracking approach for 4D deformation analysis in echocardiography. In *Proc IEEE Int Symp Biomed Imaging*, pages 458–461, 2012. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Crosby *et al.* 2009] J. Crosby, B. H. Amundsen, T. Hergum, E. W. Remme, S. Langeland, and H. Torp. 3-D Speckle Tracking for Assessment of Regional Left Ventricular Function. *Ultrasound Med Biol*, 35(3):458–471, 2009. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Curiel *et al.* 2007] L. Curiel, R. Chopra, and K. Hynynen. Progress in Multimodality Imaging: Truly Simultaneous Ultrasound and Magnetic Resonance Imaging. *IEEE Trans Med Imag*, 26(12):1740–1746, 2007. [5.1](#), [5.1.3](#)
- [Das and Banerjee 2004] B. Das and S. Banerjee. Inertial snake for contour detection in ultrasonography images. *IEE Proc Vis Image Sign*, 151(3):235–240, 2004. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [De Luca *et al.* 2011] V. De Luca, H. Grabner, L. Petrusca, R. Salomir, G. Szekely, and C. Tanner. Keep Breathing! Common Motion Helps Multi-modal Mapping. In *Med Image Comput Comput Assist Interv*, volume 6891 of *LNCS*, pages 597–604. Springer, 2011. [1.2](#)
- [De Luca *et al.* 2012] V. De Luca, C. Tanner, and G. Szekely. Speeding-up Image Registration for Repetitive Motion Scenarios. In *Proc IEEE Int Symp Biomed Imaging*, pages 1355–1358, 2012. [1.2](#), [7.2](#)
- [De Luca *et al.* 2013a] V. De Luca, G. Szekely, and C. Tanner. Ultrasound Tracking: a Survey. *Med Image Anal*, 2013. submitted. [1.2](#)
- [De Luca *et al.* 2013b] V. De Luca, M. Tschannen, G. Szekely, and C. Tanner. A Learning-based Approach for Fast and Robust Vessel Tracking in Long Ultrasound

- Sequences. In *Med Image Comput Comput Assist Interv*, volume 8149 of *LNCS*, pages 518–525. Springer, 2013. [1.2](#)
- [Demi *et al.* 2008] M. Demi, E. Bianchini, F. Faita, and V. Gemignani. Contour tracking on ultrasound sequences of vascular images. *Pattern Recognition and Image Analysis*, 18:606–612, 2008. [2.2.1](#), [2.3](#)
- [Dijkstra 1959] E. W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1:269–271, 1959. [6.2.1](#)
- [Dindyal and Kyriakides 2011] S. Dindyal and C. Kyriakides. Ultrasound microbubble contrast and current clinical applications. *Recent Pat Cardiovasc Drug Discov*, 6(1):27–41, 2011. [5.1](#)
- [Duan *et al.* 2005a] Q. Duan, E. D. Angelini, S. L. Herz, O. Gerard, P. Allain, C. M. Ingrassia, K. D. Costa, J. W. Holmes, S. Homma, and A. F. Laine. Tracking of LV Endocardial Surface on Real-Time Three-Dimensional Ultrasound with Optical Flow. 3504:434–445, 2005. [2.2.1](#)
- [Duan *et al.* 2005b] Q. Duan, E. D. Angelini, S. L. Herz, C. M. Ingrassia, O. Gerard, K. D. Costa, J. W. Holmes, S. Homma, and A. F. Laine. Dynamic Cardiac Information From Optical Flow Using Four Dimensional Ultrasound. In *Conf Proc IEEE Eng Med Biol Soc*, pages 4465–4468, 2005. [2.2.1](#)
- [Duan *et al.* 2006] Q. Duan, E. D. Angelini, O. Gerard, S. Homma, and A. F. Laine. Comparing optical-flow based methods for quantification of myocardial deformations on RT3D ultrasound. In *Proc IEEE Int Symp Biomed Imaging*, pages 173–176, 2006. [2.2.1](#)
- [Duan *et al.* 2007] Q. Duan, E. D. Angelini, S. Homma, and A. F. Laine. Validation of optical-flow for quantification of myocardial deformations on simulated RT3D ultrasound. In *Proc IEEE Int Symp Biomed Imaging*, pages 944–947, 2007. [2.2.1](#)
- [El-Sharkawy *et al.* 2001] A. E. M. El-Sharkawy, K. Z. Abd-Elmonien, A. B. M. Youssef, and Y. M. Kadah. Improved Ultrasound Speckle Motion Tracking Using Nonlinear Diffusion Filtering. volume 4325, pages 453–461, 2001. [2.1.4](#), [2.2.2](#), [2.2.2](#)
- [Elen *et al.* 2008] A. Elen, H. F. Choi, D. Loeckx, H. Gao, P. Claus, P. Suetens, F. Maes, and J. D’hooge. Three-Dimensional Cardiac Strain Estimation Using Spatio-Temporal Elastic Registration of Ultrasound Images: A Feasibility Study. *IEEE Trans Med Imaging*, 27(11):1580–1591, 2008. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Evans and Nixon 1996] A. N. Evans and M. S. Nixon. Biased motion-adaptive temporal filtering for speckle reduction in echocardiography. *IEEE Trans Med Imaging*, 15(1):39–50, 1996. [2.2.4](#)
- [Fehrenbach *et al.* 2010] J. Fehrenbach, M. Masmoudi, and D. Melodelima. Low dimensional optimization for in vivo real-time porcine liver motion estimation using ultrasound imaging. *Ultrasonics*, 50(1):44–51, 2010. [2.2.1](#)

- [Feinberg *et al.* 2010] D. A. Feinberg, D. Giese, D. A. Bongers, S. Ramanna, M. Zaitsev, M. Markl, and M. Günther. Hybrid ultrasound MRI for improved cardiac imaging and real-time respiration control. *Magn Reson Med*, 63(2):290–296, 2010. [5.1](#)
- [Felsberg and Sommer 2000] M. Felsberg and G. Sommer. A New Extension of Linear Signal Processing for Estimating Local Properties and Detecting Features. In *DAGM Symp*, pages 195–202. Springer-Verlag, 2000. [2.2.5](#)
- [Fitzpatrick *et al.* 1998] J. M. Fitzpatrick, D. L. G. Hill, Y. Shyr, J. West, C. Studholme, and C. R. J. Maurer. Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain. *IEEE Trans Med Imaging*, 17(4):571–585, 1998. [2.3](#)
- [Fleet and Jepson 1990] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *Int J Comput Vis*, 5(1):77–104, 1990. [2.2.1](#)
- [Fleet and Jepson 1993] D. J. Fleet and A. D. Jepson. Stability of phase information. *IEEE Trans Pattern Anal Mach Intell*, 15(12):1253–1268, 1993. [2.2.1](#)
- [Fleet and Weiss 2005] D. J. Fleet and Y. Weiss. Optical Flow Estimation. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Mathematical models for Computer Vision: The Handbook*, chapter 15, pages 239–258. Springer, 2005. [2.2.1](#)
- [Foroughi and Abolmaesumi 2005a] P. Foroughi and P. Abolmaesumi. A modified HAMMER algorithm for deformable registration of ultrasound images. In *Proc CARS*, volume 1281, pages 236–241, 2005. [2.2.5](#)
- [Foroughi and Abolmaesumi 2005b] P. Foroughi and P. Abolmaesumi. Elastic registration of 3D ultrasound images. In *Med Image Comput Comput Assist Interv*, volume 3749 of *LNCS*, pages 83–90. Springer, 2005. [2.1.4](#), [2.2.5](#)
- [Foroughi *et al.* 2006a] P. Foroughi, P. Abolmaesumi, and K. Hashtrudi-Zaad. Intra-subject elastic registration of 3D ultrasound images. *Med Image Anal*, 10(5):713–725, 2006. [2.1.4](#), [2.2.5](#), [2.4](#), [6.5](#)
- [Foroughi *et al.* 2006b] P. Foroughi, P. Abolmaesumi, and K. Hashtrudi-Zaad. Towards Real-time Registration of 4D Ultrasound Images. In *Conf Proc IEEE Eng Med Biol Soc*, pages 404–407, 2006. [2.1.4](#), [2.2.5](#), [2.4](#)
- [Friemel *et al.* 1995] B. H. Friemel, L. N. Bohs, and G. E. Trahey. Relative performance of two-dimensional speckle-tracking techniques: normalized correlation, non-normalized correlation and sum-absolute-difference. In *Proc IEEE Ultrason Symp*, volume 2, pages 1481–1484, 1995. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Geman and Geman 1984] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of image sequences. *IEEE Trans Pattern Anal Mach Intell*, 6:721–741, 1984. [2.1.3](#)

- [Goldberg 1989] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1st edition, 1989. [2.1.3](#)
- [Golemati *et al.* 2003] S. Golemati, A. Sassano, M. J. Lever, A. A. Bharath, S. Dhanjil, and A. N. Nicolaides. Carotid artery wall motion estimated from B-mode ultrasound using region tracking and block matching. *Ultrasound Med Biol*, 29(3):387–399, 2003. [2.2.2](#), [2.2.2](#)
- [Guerrero *et al.* 2007] J. Guerrero, S. E. Salcudean, J. A. McEwen, B. A. Masri, and S. Nicolaou. Real-Time Vessel Segmentation and Tracking for Ultrasound Imaging Applications. *IEEE Trans Med Imaging*, 26(8):1079–1090, 2007. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Hagan and Menhaj 1994] M. T. Hagan and M. B. Menhaj. Training feedforward networks with the Marquardt algorithm. *IEEE Trans Neural Netw*, 5(6):989–993, 1994. [5.2.3](#)
- [Häme *et al.* 2012] Y. Häme, V. Gamarnik, K. M. Parker, J. W. Holmes, and A. F. Laine. Level set-based tracking of the endocardium without a shape prior from 3D ultrasound images. In *Proc IEEE Int Symp Biomed Imaging*, pages 466–469, 2012. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Harris *et al.* 2007] E. J. Harris, N. R. Miller, J. C. Bamber, P. M. Evans, and J. R. N. Symonds-Taylor. Performance of ultrasound based measurement of 3D displacement using a curvilinear probe for organ motion tracking. *Phys Med Biol*, 52(18):5683–5703, 2007. [2.2.2](#), [2.2.2](#), [2.2.2](#), [3.4.1](#)
- [Harris *et al.* 2010] E. J. Harris, N. R. Miller, J. C. Bamber, J. R. N. Symonds-Taylor, and P. M. Evans. Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound. *Phys Med Biol*, 55(12):3363–3380, 2010. [2.2.2](#), [2.2.2](#), [2.2.2](#), [2.4](#), [3.4](#), [6.5](#)
- [Harris *et al.* 2011] E. J. Harris, N. R. Miller, J. C. Bamber, J. R. N. Symonds-Taylor, and P. M. Evans. The effect of object speed and direction on the performance of 3D speckle tracking using a 3D swept-volume ultrasound probe. *Phys Med Biol*, 56(22):7127–7143, 2011. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Herlin and Ayache 1992] I. Herlin and N. Ayache. Features extraction and analysis methods for sequences of ultrasound images. In G. Sandini, editor, *Comput Vis ECCV*, volume 588 of *LNCS*, pages 43–57. Springer, 1992. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Hestenes and Stiefel 1952] M. R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *J Res Natl Bur Stand*, 49(6):409–436, 1952. [2.1.3](#)
- [Heyde *et al.* 2012] B. Heyde, P. Claus, R. Jasaityte, D. Barbosa, S. Bouchez, M. Vandenheuvell, P. Wouters, F. Maes, and J. D’hooge. Motion and deformation estimation of cardiac ultrasound sequences using an anatomical B-spline

- transformation model. In *Proc IEEE Int Symp Biomed Imaging*, pages 266–269, 2012. 2.1.3, 2.2.3, 2.2.3, 2.2.3, 2.2.3
- [Hill *et al.* 2001] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys Med Biol*, 46(3):R1–45, 2001. 2.1, 2.3
- [Holbrook *et al.* 2010] A. B. Holbrook, J. M. Santos, E. Kaye, V. Rieke, and K. B. Pauly. Real-time mr thermometry for monitoring hifu ablations of the liver. *Magn Reson Med*, 63(2):365–373, 2010. 1.1, 5.1
- [Horn and Schunck 1981] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artif Intell*, 17:185–203, 1981. 2.2.1
- [Hsu *et al.* 2005] A. Hsu, N. R. Miller, P. M. Evans, J. C. Bamber, and S. Webb. Feasibility of using ultrasound for real-time tracking during radiotherapy. *Med Phys*, 32(6):1500–1512, 2005. 2.2.2, 2.2.2, 2.2.2, 3.6, 4.4
- [Huang *et al.* 2005] X. Huang, N. Hill, J. Ren, G. Guiraudon, D. Boughner, and T. M. Peters. Dynamic 3D ultrasound and MR image registration of the beating heart. *Med Image Comput Comput Assist Interv*, 8(2):171–178, 2005. 5.1.3
- [Hynynen 2010] K. Hynynen. MRI-guided focused ultrasound treatments. *Ultrasonics*, 50(2):221–229, 2010. 5.1
- [Ibanez *et al.* 2005] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, <http://www.itk.org/ItkSoftwareGuide.pdf>, 2nd edition, 2005. 3.3.1
- [Ishihara *et al.* 1995] Y. Ishihara, A. Calderon, H. Watanabe, K. Okamoto, Y. Suzuki, K. Kuroda, and Y. Suzuki. A precise and fast temperature mapping using water proton chemical shift. *Magn Reson Med*, 34(6):814–823, 1995. 5.1
- [Ishikawa 2009] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pages 2993–3000, 2009. 2.1.3
- [Jacob *et al.* 1998] G. Jacob, J. A. Noble, and A. Blake. Robust contour tracking in echocardiographic sequences. In *Proc IEEE Int Conf Comput Vis*, pages 408–413, 1998. 2.2.4, 2.2.4
- [Jacob *et al.* 1999] G. Jacob, J. A. Noble, M. Mulet-Parada, and A. Blake. Evaluating a robust contour tracker on echocardiographic sequences. *Med Image Anal*, 3(1):63–75, 1999. 2.2.4, 2.2.4, 2.2.4
- [Jacob *et al.* 2002] G. Jacob, J. A. Noble, C. Behrenbruch, A. D. Kelion, and A. P. Banning. A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. *IEEE Trans Med Imaging*, 21(3):226–238, 2002. 2.1.1, 2.2.4, 2.2.4, 2.2.4, 2.3

- [Jannin *et al.* 2006] P. Jannin, C. Grova, and C. R. J. Maurer. Model for defining and reporting reference-based validation protocols in medical image processing. *Int J Comput Assist Radiol Surg*, 1:63–73, 2006. [2.3](#)
- [Jia *et al.* 2010] C. Jia, P. Yan, A. J. Sinusas, D. P. Dione, B. A. Lin, Q. Wei, K. Thiele, T. J. Kolias, J. M. Rubin, L. Huang, J. S. Duncan, and M. O’Donnell. 3D Elasticity imaging using principal stretches on an open-chest dog heart. In *Proc IEEE Ultrason Symp*, pages 583–586, 2010. [2.2.2](#), [2.2.2](#)
- [Jiang and Hall 2007] J. Jiang and T. J. Hall. A parallelizable real-time motion tracking algorithm with applications to ultrasonic strain imaging. *Phys Med Biol*, 52(13):3773–3790, 2007. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Jiang and Hall 2009] J. Jiang and T. J. Hall. A Generalized Speckle Tracking Algorithm for Ultrasonic Strain Imaging Using Dynamic Programming. *Ultrasound Med Biol*, 35(11):1863–1879, 2009. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Kaluzynski *et al.* 2001] K. Kaluzynski, X. Chen, S. Y. Emelianov, A. R. Skovoroda, and M. O’Donnell. Strain rate imaging using two-dimensional speckle tracking. *IEEE Trans Ultrason Ferroelectr Freq Control*, 48(4):1111–1123, 2001. [2.2.2](#), [2.2.2](#), [3.4.1](#)
- [Kass *et al.* 1988] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int J Comput Vis*, 1(4):321–331, 1988. [2.2.4](#)
- [Keall *et al.* 2006] P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. R. Ramsey, M. B. Van Herk, S. S. Vedam, J. W. Wong, and E. Yorke. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med Phys*, 33(10):3874–3900, 2006. [1](#)
- [Kennedy *et al.* 2003] J. E. Kennedy, G. R. Haar, and D. Cranston. High intensity focused ultrasound: surgery of the future? *Br J Radiol*, 76(909):590–599, 2003. [1.1](#)
- [Kennedy *et al.* 2004] J. E. Kennedy, F. Wu, G. R. ter Haar, F. V. Gleeson, R. R. Phillips, M. Middleton, and D. Cranston. High-intensity focused ultrasound for the treatment of liver tumours. *Ultrasonics*, 42(19):931–935, 2004. [5.4](#)
- [King *et al.* 2010] A. P. King, K. S. Rhode, Y. Ma, C. Yao, C. Jansen, R. Razavi, and G. P. Penney. Registering preprocedure volumetric images with intraprocedure 3-D ultrasound using an ultrasound imaging model. *IEEE Trans Med Imaging*, 29(3):924–937, 2010. [6](#)
- [Korstanje *et al.* 2010] J. W. H. Korstanje, R. W. Selles, H. J. Stam, S. E. R. Hovius, and J. G. Bosch. Development and validation of ultrasound speckle tracking to quantify tendon displacement. *J Biomech*, 43(7):1373–1379, 2010. [2.2.2](#), [2.2.2](#), [2.2.2](#)

- [Kovesi 2000] P. Kovesi. Phase congruency: A low-level image invariant. *Psychol Res*, 64:136–148, 2000. [2.2.4](#), [2.2.5](#)
- [Krucker *et al.* 2002] J. F. Krucker, G. L. LeCarpentier, J. B. Fowlkes, and P. L. Carson. Rapid elastic image registration for 3-D ultrasound. *IEEE Trans Med Imaging*, 21(11):1384–1394, 2002. [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Kuo and von Ramm 2008] J. Kuo and O. T. von Ramm. Three-dimensional motion measurements using feature tracking. *IEEE Trans Ultrason Ferroelectr Freq Control*, 55(4):800–810, 2008. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Lange *et al.* 2009] T. Lange, N. Papenberg, S. Heldmann, J. Modersitzki, B. Fischer, H. Lamecker, and P. Schlag. 3D ultrasound-CT registration of the liver using combined landmark-intensity information. *Int J Comput Assist Radiol Surg*, 4(1):79–88, 2009. [6](#), [6.5](#)
- [Ledesma-Carbayo *et al.* 2001] M. J. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, and M. Unser. Cardiac Motion Analysis from Ultrasound Sequences Using Non-Rigid Registration. In W. Niessen and M. Viergever, editors, *Med Image Comput Comput Assist Interv*, volume 2208 of *LNCS*, pages 889–896. Springer, 2001. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Ledesma-Carbayo *et al.* 2005] M. J. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, M. Suhling, P. Hunziker, and M. Unser. Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *IEEE Trans Med Imaging*, 24(9):1113–1126, 2005. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Lediju Bell *et al.* 2012] M. A. Lediju Bell, B. C. Byram, E. J. Harris, P. M. Evans, and J. C. Bamber. In vivo liver tracking with a high volume rate 4D ultrasound scanner and a 2D matrix array probe. *Phys Med Biol*, 57(5):1359–1374, 2012. [2.2.2](#), [2.2.2](#), [2.4](#)
- [Lediju *et al.* 2010] M. A. Lediju, B. C. Byram, E. J. Harris, P. M. Evans, and J. C. Bamber. 3D Liver tracking using a matrix array: Implications for ultrasonic guidance of IMRT. In *Proc IEEE Ultrason Symp*, pages 1628–1631, 2010. [2.2.2](#), [2.2.2](#), [2.2.2](#), [2.4](#)
- [Leung *et al.* 2009] C. Leung, K. Hashtrudi-Zaad, P. Foroughi, and P. Abolmaesumi. A Real-Time Intrasubject Elastic Registration Algorithm for Dynamic 2-D Ultrasound Images. *Ultrasound Med Biol*, 35(7):1159–1176, 2009. [2.1.4](#), [2.2.4](#), [2.2.5](#), [2.3](#)
- [Li *et al.* 2005] M. Li, C. Kambhamettu, and M. Stone. Automatic contour tracking in ultrasound images. *Clin Linguist Phon*, 19(6-7):545–554, 2005. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Lin and Tai 1997] Y. C. Lin and S. C. Tai. Fast full-search block-matching algorithm for motion-compensated video compression. *IEEE Trans Commun*, 45(5):527–531, 1997. [3.4.4](#)

- [Lin *et al.* 2007] C. H. Lin, M. C. J. Lin, and Y. N. Sun. Ultrasound motion estimation using a hierarchical feature weighting algorithm. *Comput Med Imaging Graph*, 31(3):178–190, 2007. [2.2.2](#), [2.2.2](#), [3.4](#), [3.4.2](#), [3.4.2](#), [3.4.2](#), [3.4.4](#)
- [Lindeberg 1998] T. Lindeberg. Feature detection with automatic scale selection. *Int J Comput Vis*, 30:79–116, 1998. [3.4.1](#)
- [Lubinski *et al.* 1999] M. A. Lubinski, S. Y. Emelianov, and M. O’Donnell. Speckle tracking methods for ultrasonic elasticity imaging using short-time correlation. *IEEE Trans Ultrason Ferroelectr Freq Control*, 46(1):82–96, 1999. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Lucas and Kanade 1981] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings of Imaging Understanding Workshop*, pages 121–130, 1981. [2.2.1](#)
- [Maintz and Viergever 1998] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Med Image Anal*, 2(1):1–36, 1998. [2.1](#), [2.3](#)
- [Malassiotis and Strintzis 1999] S. Malassiotis and M. G. Strintzis. Tracking the left ventricle in echocardiographic images by learning heart dynamics. *IEEE Trans Med Imaging*, 18(3):282–290, 1999. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Marquardt 1963] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J Appl Math*, 11(2):431441, 1963. [2.1.3](#)
- [Marquet *et al.* 2006] F. Marquet, M. Pernot, J. F. Aubry, M. Tanter, G. Montaldo, and M. Fink. In-vivo non-invasive motion tracking and correction in High Intensity Focused Ultrasound therapy. In *Conf Proc IEEE Eng Med Biol Soc*, pages 688–691, 2006. [1.1](#), [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [McClelland *et al.* 2011] J. R. McClelland, S. Hughes, M. Modat, A. Qureshi, S. Ahmad, D. B. Landau, S. Ourselin, and D. J. Hawkes. Inter-fraction variations in respiratory motion models. *Phys Med Biol*, 56(1):251–272, 2011. [1](#)
- [McKinnon 1999] K. I. M. McKinnon. Convergence of the Nelder-Mead simplex method to a non-stationary point. *SIAM J Optim*, 9:148158, 1999. [2.1.3](#)
- [Meairs and Hennerici 1999] S. Meairs and M. Hennerici. Four-Dimensional Ultrasonographic Characterization of Plaque Surface Motion in Patients With Symptomatic and Asymptomatic Carotid Artery Stenosis. *Stroke*, 30(9):1807–1813, 1999. [2.2.1](#)
- [Mellor and Brady 2004] M. Mellor and M. Brady. Non-rigid Multimodal Image Registration Using Local Phase. In C. Barillot, D. Haynor, and P. Hellier, editors, *Med Image Comput Comput Assist Interv*, volume 3216 of *LNCS*, pages 789–796. Springer, 2004. [2.2.5](#)
- [Mercier *et al.* 2012] L. Mercier, V. Fonov, C. Haegelen, R. F. Del Maestro, K. Petrecca, and D. L. Collins. Comparing two approaches to rigid registration of

- three-dimensional ultrasound and magnetic resonance images for neurosurgery. *Int J Comput Assist Radiol Surg*, 7(1):125–136, 2012. [5.1.3](#)
- [Metz *et al.* 2011] C. T. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen. Nonrigid registration of dynamic medical imaging data using $nD + t$ B-splines and a groupwise optimization approach. *Med Image Anal*, 15(2):238–249, 2011. [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Meunier 1998] J. Meunier. Tissue motion assessment from 3D echographic speckle tracking. *Phys Med Biol*, 43(5):1241–1254, 1998. [2.2.1](#), [2.2.2](#)
- [Mignotte *et al.* 2001] M. Mignotte, J. Meunier, and J. C. Tardif. Endocardial Boundary Estimation and Tracking in Echocardiographic Images using Deformable Template and Markov Random Fields. *Pattern Anal Appl*, 4:256–271, 2001. [2.1.3](#), [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#)
- [Mikic *et al.* 1998] I. Mikic, S. Krucinski, and J. D. Thomas. Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. *IEEE Trans Med Imaging*, 17(2):274–284, 1998. [2.2.1](#)
- [Milko *et al.* 2009] S. Milko, E. L. Melvaer, E. Samset, and T. Kadir. Evaluation of bivariate correlation ratio similarity metric for rigid registration of US/MR images of the liver. *Int J Comput Assist Radiol Surg*, 4(2):147–155, 2009. [5.1.3](#), [6](#)
- [Morrone and Owens 1987] M. C. Morrone and R. A. Owens. Feature detection from local energy. *Pattern Recognit Lett*, 6(5):303–313, 1987. [2.2.5](#)
- [Morsy and Von Ramm 1999] A. A. Morsy and O. T. Von Ramm. FLASH correlation: a new method for 3-D ultrasound tissue motion tracking and blood velocity estimation. *IEEE Trans Ultrason Ferroelectr Freq Control*, 46(3):728–736, 1999. [2.2.2](#), [2.2.2](#), [3.4.1](#)
- [Mukherjee *et al.* 2011] R. Mukherjee, C. Sprouse, T. Abraham, B. Hoffmann, E. McVeigh, D. Yuh, and P. Burlina. Myocardial motion computation in 4D ultrasound. In *Proc IEEE Int Symp Biomed Imaging*, pages 1070–1073, 2011. [2.2.1](#)
- [Mukherjee *et al.* 2012] R. Mukherjee, A. Pinheiro, J. Gammie, D. Yuh, T. Abraham, E. McVeigh, and P. Burlina. Dense myocardial motion from 4D ultrasound: Comparative performance evaluation. In *Proc IEEE Int Symp Biomed Imaging*, pages 278–281, 2012. [2.2.1](#)
- [Mulet-Parada and Noble 2000] M. Mulet-Parada and J. A. Noble. 2D+T acoustic boundary detection in echocardiography. *Med Image Anal*, 4(1):21–30, 2000. [2.2.1](#), [2.2.5](#)
- [Nagel 1987] H. H. Nagel. On the estimation of optical flow: relations between different approaches and some new results. *Artif Intell*, 33:298–324, 1987. [2.2.1](#)

- [Nakamoto *et al.* 2007] M. Nakamoto, H. Hirayama, Y. Sato, K. Konishi, Y. Kakeji, M. Hashizume, and S. Tamura. Recovery of respiratory motion and deformation of the liver using laparoscopic freehand 3D ultrasound system. *Med Image Anal*, 11(5):429–442, 2007. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Nascimento and Marques 2004] J. C. Nascimento and J. S. Marques. Robust shape tracking in the presence of cluttered background. *IEEE Trans Multimed*, 6(6):852–861, 2004. [2.2.4](#)
- [Nascimento and Marques 2008] J. C. Nascimento and J. S. Marques. Robust Shape Tracking With Multiple Models in Ultrasound Images. *IEEE Trans Image Process*, 17(3):392–406, 2008. [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#)
- [Nelder and Mead 1965] J. A. Nelder and R. Mead. A simplex method for function minimization. *Comp J*, 7:308313, 1965. [2.1.3](#)
- [Nguyen *et al.* 2008] T. N. Nguyen, J. L. Moseley, L. A. Dawson, D. A. Jaffray, and K. K. Brock. Adapting population liver motion models for individualized online image-guided therapy. In *Conf Proc IEEE Eng Med Biol Soc*, pages 3945–3948, 2008. [1](#)
- [Noble and Boukerroui 2006] J. A. Noble and D. Boukerroui. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging*, 25(8):987–1010, 2006. [2.2.4](#)
- [Noble *et al.* 2011] J. A. Noble, N. Navab, and H. Becher. Ultrasonic image analysis and image-guided interventions. *Interface Focus*, 1:673–685, 2011. [2.1.4](#), [2.2.4](#)
- [Nocedal and Wright 1999] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 1999. [2.1.3](#)
- [Notomi *et al.* 2005] Y. Notomi, P. Lysyansky, R. M. Setser, T. Shiota, Z. B. Popovic, M. G. Martin-Miklovic, J. A. Weaver, S. J. Oryszak, N. L. Greenberg, R. D. White, and J. D. Thomas. Measurement of Ventricular Torsion by Two-Dimensional Ultrasound Speckle Tracking Imaging. *J Am Coll Cardiol*, 45(12):2034–2041, 2005. [2.2.2](#)
- [O’Donnell *et al.* 1994] M. O’Donnell, A. R. Skovoroda, B. M. Shapo, and S. Y. Emelianov. Internal displacement and strain imaging using ultrasonic speckle tracking. *IEEE Trans Ultrason Ferroelectr Freq Control*, 41(3):314–325, 1994. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Orderud *et al.* 2007] F. Orderud, J. Hansgard, and S. I. Rabben. Real-Time Tracking of the Left Ventricle in 3D Echocardiography Using a State Estimation Approach. In N. Ayache, S. Ourselin, and A. Maeder, editors, *Med Image Comput Comput Assist Interv*, volume 4791 of *LNCS*, pages 858–865. Springer, 2007. [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#)
- [Orderud 2006] F. Orderud. A Framework for real-time left ventricular tracking in 3D+T echocardiography, using nonlinear deformable contours and kalman filter based tracking. In *Comput Cardiol*, pages 125–128, 2006. [2.2.4](#)

- [Papademetris *et al.* 1998] X. Papademetris, J. Rambo, D. P. Dione, A. J. Sinusas, and J. S. Duncan. Visually interactive cine-3D segmentation of cardiac MR images. *Supplementary to J Am Coll Cardiol*, 31(2), 1998. [2.2.4](#)
- [Papademetris *et al.* 1999] X. Papademetris, A. J. Sinusas, D. P. Dione, and J. S. Duncan. 3D Cardiac Deformation from Ultrasound Images. In *Med Image Comput Comput Assist Interv*, volume 1679 of *LNCS*, pages 420–429. Springer, 1999. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Paragios and Deriche 2002] N. Paragios and R. Deriche. Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation. *Int J Comput Vis*, 46(3):223–247, 2002. [2.1.1](#)
- [Pellet-Barakat *et al.* 2004] C. Pellet-Barakat, F. Frouin, M. F. Insana, and A. Herment. Ultrasound elastography based on multiscale estimations of regularized displacement fields. *IEEE Trans Med Imaging*, 23(2):153–163, 2004. [2.2.1](#)
- [Peng *et al.* 2006] Q. Peng, R. Jones, and C. Constantinou. 2D Ultrasound Image Processing in Identifying Responses of Urogenital Structures to Pelvic Floor Muscle Activity. *Ann Biomed Eng*, 34:477–493, 2006. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Pennec *et al.* 2003] X. Pennec, P. Cachier, and N. Ayache. Tracking brain deformations in time-sequences of 3D US images. *Pattern Recognit Lett*, 24:801–813, 2003. [2.1.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.3](#)
- [Penney *et al.* 2004] G. P. Penney, J. M. Blackall, M. S. Hamady, T. Sabharwal, A. Adam, and D. J. Hawkes. Registration of freehand 3D ultrasound and magnetic resonance liver images. *Med Image Anal*, 8(1):81–91, 2004. [5.1.3](#), [6](#), [6.5](#)
- [Pernot *et al.* 2003] M. Pernot, M. Tanter, and M. Fink. Feasibility of real-time motion correction for H.I.F.U applications. In *Proc IEEE Ultrason Symp*, volume 1, pages 998–1001, 2003. [1.1](#), [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Pernot *et al.* 2004] M. Pernot, M. Tanter, and M. Fink. 3-D real-time motion correction in high-intensity focused ultrasound therapy. *Ultrasound Med Biol*, 30(9):1239–1249, 2004. [1.1](#), [2.2.2](#), [2.2.2](#), [2.2.2](#), [5.1](#)
- [Petrusca *et al.* 2011] L. Petrusca, V. De Luca, P. Arnold, Z. Celicanin, T. Goget, V. Auboiroux, M. Viallon, F. Santini, S. Terraz, K. Scheffler, C. Tanner, P. Cattin, and R. Salomir. Ultrasound/MR hybrid imaging: truly simultaneous motion monitoring in the abdomen and image co-registration. In *Proc Int Soc Magn Reson Med*, volume 19, 2011. [3.1](#), [5.1.1](#), [6.1](#)
- [Petrusca *et al.* 2012] L. Petrusca, M. Viallon, S. Terraz, V. De Luca, Z. Celicanin, V. Auboiroux, S. Brunke, P. Cattin, and R. Salomir. *Simultaneous Ultrasound Imaging and MRI Acquisition*, pages 457–470. Medical Radiology. Springer, 2012. [1.1](#)
- [Petrusca *et al.* 2013] L. Petrusca, P. Cattin, V. De Luca, F. Preiswerk, Z. Celicanin, V. Auboiroux, M. Viallon, P. Arnold, F. Santini, S. Terraz, K. Scheffler, C. D.

- Becker, and R. Salomir. Hybrid Ultrasound/Magnetic Resonance Simultaneous Acquisition and Image Fusion for Motion Monitoring in the Upper Abdomen. *Invest Radiol*, 48(5):333–340, 2013. [1.1](#), [1.2](#), [3.1](#), [5.1.1](#), [5.1](#), [5.1.1](#), [5.3](#)
- [Pinton *et al.* 2006] G. F. Pinton, J. J. Dahl, and G. E. Trahey. Rapid tracking of small displacements with ultrasound. *IEEE Trans Ultrason Ferroelectr Freq Control*, 53(6):1103–1117, 2006. [2.2.2](#), [2.2.2](#), [3.6](#), [4.4](#)
- [Pirat *et al.* 2008] B. Pirat, D. S. Khoury, C. J. Hartley, L. Tiller, L. Rao, D. G. Schulz, S. F. Nagueh, and W. A. Zoghbi. A Novel Feature-Tracking Echocardiographic Method for the Quantitation of Regional Myocardial Function: Validation in an Animal Model of Ischemia-Reperfusion. *J Am Coll Cardiol*, 51(6):651–659, 2008. [2.2.2](#), [2.2.5](#)
- [Porter *et al.*] B. C. Porter, D. J. Rubens, J. G. Strang, J. Smith, S. Totterman, and K. J. Parker. Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. *IEEE Trans Med Imaging*, 20(4):354–359. [6](#)
- [Porter *et al.* 2001] B. C. Porter, D. J. Rubens, J. G. Strang, J. Smith, S. Totterman, and K. J. Parker. Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers. *IEEE Trans Med Imag*, 20(4):354–359, 2001. [5.1.3](#)
- [Powell 1964] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comp J*, 7:155162, 1964. [2.1.3](#)
- [Pratikakis *et al.* 2003] I. Pratikakis, C. Barillot, P. Hellier, and E. Memin. Robust multiscale deformable registration of 3D ultrasound images. *Int J Image Graph*, 3:547–565, 2003. [2.2.1](#)
- [Preiswerk *et al.* 2012] F. Preiswerk, P. Arnold, B. Fasel, and P. C. Cattin. Robust tumour tracking from 2D imaging using a population-based statistical motion model. In *Proc Workshop Math Methods Biomed Image Analysis*, pages 209–214, 2012. [5.2.1](#)
- [Preiswerk *et al.* 2013] F. Preiswerk, V. De Luca, P. Arnold, Z. Celicanin, L. Petrusca, C. Tanner, R. Salomir, and P. C. Cattin. Model-Guided Respiratory Organ Motion Prediction of the Liver from 2D Ultrasound. *Med Image Anal*, 2013. submitted. [1.2](#), [5.1.2](#), [5.2](#), [5.4](#), [5.2.3](#), [5.2.3](#), [5.5](#), [5.3](#), [6](#), [6.5](#)
- [Press *et al.* 1992] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992. [2.1.3](#)
- [Rabben *et al.* 2000] S. I. Rabben, A. H. Torp, A. Stoylen, S. Slordahl, K. Bjornstad, B. O. Haugen, and B. Angelsen. Semiautomatic contour detection in ultrasound M-mode images. *Ultrasound Med Biol*, 26(2):287–296, 2000. [2.2.4](#)

- [Revell *et al.* 2003] J. Revell, M. Mirmehdi, and D. McNally. Ultrasound Speckle Tracking For Strain Estimation. Technical Report CSTR-04-005, Department of Computer Science, University of Bristol, 2003. [2.2.2](#), [2.2.2](#)
- [Revell *et al.* 2005] J. Revell, M. Mirmehdi, and D. McNally. Computer Vision Elastography: Speckle Adaptive Motion Estimation for Elastography Using Ultrasound Sequences. *IEEE Trans Med Imaging*, 24(6):755–766, 2005. [2.2.2](#), [2.2.2](#), [3.4](#), [3.4.3](#)
- [Ries *et al.* 2010] M. Ries, B. D. de Senneville, S. Roujol, Y. Berber, B. Quesson, and C. Moonen. Real-time 3D target tracking in MRI guided focused ultrasound ablations in moving tissues. *Magn Reson Med*, 64(6):1704–1712, 2010. [1.1](#), [4](#)
- [Rueckert *et al.* 1999] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*, 18(8):712–721, 1999. [2.1.2](#), [2.2.3](#), [5.1.2](#)
- [Salcudean *et al.* 1999] S. Salcudean, G. Bell, S. Bachmann, W. Zhu, P. Abolmaesumi, and P. Lawrence. Robot-Assisted Diagnostic Ultrasound Design and Feasibility Experiments. In C. Taylor and A. Colchester, editors, *Med Image Comput Comput Assist Interv*, volume 1679 of *LNCS*, pages 1062–1071. Springer, 1999. [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Schneider *et al.* 2012] R. J. Schneider, D. P. Perrin, N. V. Vasilyev, G. R. Marx, P. J. del Nido, and R. D. Howe. Real-time image-based rigid registration of three-dimensional ultrasound. *Med Image Anal*, 16:402–414, 2012. [2.2.4](#), [2.2.4](#), [2.2.4](#), [3.4.1](#)
- [Sharp *et al.* 2004] G. C. Sharp, S. B. Jiang, S. Shimizu, and H. Shirato. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol*, 49:425–440, 2004. [5.2.3](#)
- [Shekhar and Zagrodsky 2002] R. Shekhar and V. Zagrodsky. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Trans Med Imaging*, 21(1):9–22, 2002. [2.1.3](#), [2.1.4](#), [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Shekhar *et al.* 2004] R. Shekhar, V. Zagrodsky, M. J. Garcia, and J. D. Thomas. Registration of real-time 3-D ultrasound images of the heart for novel 3-D stress echocardiography. *IEEE Trans Med Imaging*, 23(9):1141–1149, 2004. [2.1.4](#), [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Shen 2004] D. Shen. Image Registration by Hierarchical Matching of Local Spatial Intensity Histograms. In C. Barillot, D. Haynor, and P. Hellier, editors, *Med Image Comput Comput Assist Interv*, volume 3216 of *LNCS*, pages 582–590. Springer, 2004. [2.2.5](#)
- [Shirato *et al.* 2006] H. Shirato, K. Suzuki, G. C. Sharp, K. Fujita, R. Onimaru, M. Fujino, N. Kato, Y. Osaka, R. Kinoshita, H. Taguchi, S. Onodera, and

- K. Miyasaka. Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy. *Int J Radiat Oncol Biol Phys*, 64(4):1229–1236, 2006. [4](#)
- [Shirato *et al.* 2007] H. Shirato, S. Shimizu, K. Kitamura, and R. Onimaru. Organ motion in image-guided radiotherapy: lessons from real-time tumor-tracking radiotherapy. *Int J Clin Oncol*, 12:8–16, 2007. [1](#), [5.4](#)
- [Siegel 1956] S. Siegel. *Non-parametric statistics for the behavioral sciences*, pages 75–83. McGraw-Hill, 1956. [3.2](#)
- [Singh and Allen 1992] A. Singh and P. Allen. Image-flow computation: An estimation-theoretic framework and a unified perspective. *CVGIP. Image Understanding*, 56(2):152–177, 1992. [2.2.1](#)
- [Singh 1990] A. Singh. An estimation-theoretic framework for image-flow computation. In *Proc IEEE Int Conf Comput Vis*, pages 168–177, 1990. [2.2.1](#)
- [Snyman 2005] J. A. Snyman. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*. Springer, 2005. [2.1.3](#)
- [Sotiras *et al.* 2012] A. Sotiras, D. Christos, and N. Paragios. Deformable Medical Image Registration: A Survey. Rapport de recherche RR-7919, INRIA, 2012. [2.1](#), [2.1.3](#)
- [Stoitsis *et al.* 2005] J. Stoitsis, S. Golemati, A. K. Dimopoulos, and K. S. Nikita. Analysis and quantification of arterial wall motion from B-mode ultrasound images - comparison of block-matching and optical flow. In *Conf Proc IEEE Eng Med Biol Soc*, pages 4469–4472, 2005. [2.2.1](#), [2.3](#)
- [Tang and Hamarneh 2010] L. Tang and G. Hamarneh. Graph-based tracking of the tongue contour in ultrasound sequences with adaptive temporal regularization. In *Conf Comput Vis Pattern Recognit Workshops*, pages 154–161, 2010. [2.1.3](#), [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Tang *et al.* 2008] A. M. Tang, D. F. Kacher, E. Y. Lam, K. K. Wong, F. A. Jolesz, and E. S. Yang. Simultaneous Ultrasound and MRI System for Breast Biopsy: Compatibility Assessment and Demonstration in a Dual Modality Phantom. *IEEE Trans Med Imag*, 27(2):247–254, 2008. [5.1](#)
- [Tavakoli *et al.* 2008] V. Tavakoli, M. S. Nambakhsh, N. Sahba, and A. Makinian. A new variational technique for combining affine registration and optical flow in echocardiography images. In *Conf Proc IEEE Eng Med Biol Soc*, pages 205–208, 2008. [2.2.1](#)
- [Tenenbaum *et al.* 2000] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. [6.2.1](#), [6.2.1](#), [6.2.1](#)

- [Touil *et al.* 2010] B. Touil, A. Basarab, P. Delachartre, O. Bernard, and D. Fribolet. Analysis of motion tracking in echocardiographic image sequences: Influence of system geometry and point-spread function. *Ultrasonics*, 50(3):373–386, 2010. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Tseng *et al.* 1996] Y. H. Tseng, J. N. Hwang, and F. H. Sheehan. 3-D heart contour delineation and motion tracking of ultrasound images using continuous distance transform neural networks. In *Proc IEEE Workshop Neural Networks Signal Process*, pages 361–370, 1996. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Van der Maaten *et al.* 2009] L. J. P. Van der Maaten, E. O. Postma, and H. J. Van den Herik. Dimensionality reduction: A comparative review. *TiCC TR*, 2009. [4.1](#), [6.2.1](#), [6.2.1](#), [6.2.1](#)
- [Vercauteren *et al.* 2008] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In *Med Image Comput Comput Assist Interv*, volume 5241 of *LNCS*, pages 754–761. Springer, 2008. [2.2.5](#)
- [Verma *et al.* 2011] P. S. Verma, H. Wu, M. P. Langer, I. J. Das, and G. A. Sandison. Survey: Real-Time Tumor Motion Prediction for Image-Guided Radiation Treatment. *Comput Sci Eng*, 13(5):24–35, 2011. [1](#), [5.2.3](#)
- [Veronesi *et al.* 2006] F. Veronesi, C. Corsi, E. G. Caiani, A. Sarti, and C. Lamberti. Tracking of left ventricular long axis from real-time three-dimensional echocardiography using optical flow techniques. *IEEE Trans Inf Technol Biomed*, 10(1):174–181, 2006. [2.2.1](#), [2.3](#)
- [Viallon *et al.* 2010] M. Viallon, S. Terraz, J. Roland, E. Dumont, C. D. Becker, and R. Salomir. Observation and correction of transient cavitation-induced PRFS thermometry artifacts during radiofrequency ablation, using simultaneous ultrasound/MR imaging. *Med Phys*, 37(4):1491–1506, 2010. [5.1](#)
- [von Siebenthal *et al.* 2007a] M. von Siebenthal, G. Szekely, U. Gamper, P. Boesiger, A. Lomax, and P. Cattin. 4D MR imaging of respiratory organ motion and its variability. *Phys Med Biol*, 52:1547–1564, 2007. [4.4](#), [5.1.1](#), [5.1.2](#), [6.1](#), [6.5](#)
- [von Siebenthal *et al.* 2007b] M. von Siebenthal, G. Szekely, A. Lomax, and P. Cattin. Inter-Subject Modelling of Liver Deformation During Radiation Therapy. In *Med Image Comput Comput Assist Interv*, volume 4791 of *LNCS*, pages 659–666. Springer, 2007. [1](#), [6](#), [6.5](#)
- [von Siebenthal *et al.* 2007c] M. von Siebenthal, G. Szekely, A. J. Lomax, and P. C. Cattin. Systematic errors in respiratory gating due to intrafraction deformations of the liver. *Med Phys*, 34(9):3620–3629, 2007. [1](#)
- [Wachinger *et al.* 2008] C. Wachinger, W. Wein, and N. Navab. Registration Strategies and Similarity Measures for Three-dimensional Ultrasound Mosaicing. *Acad Radiol*, 15(11):1404–1415, 2008. [2.2.4](#)

- [Wang *et al.* 2010] Y. Wang, B. Georgescu, D. Comaniciu, and H. Houle. Learning-based 3D myocardial motion flowestimation using high frame rate volumetric ultrasound data, 2010. [2.2.5](#), [2.3](#)
- [Wear and Popp 1987] K. A. Wear and R. L. Popp. Theoretical Analysis of a Technique for the Characterization of Myocardium Contraction Based Upon Temporal Correlation of Ultrasonic Echoes. *IEEE Trans Ultrason Ferroelectr Freq Control*, 34(3):368–375, 1987. [2.2.2](#)
- [Wein *et al.* 2007] W. Wein, B. Röper, and N. Navab. Integrating Diagnostic B-Mode Ultrasonography Into CT-Based Radiation Treatment Planning. *IEEE Trans Med Imaging*, 26(6):866–879, 2007. [6](#), [6.5](#)
- [Wein *et al.* 2008] W. Wein, J. Z. Cheng, and A. Khamene. Ultrasound based Respiratory Motion Compensation in the Abdomen. In *MICCAI Workshop on Image Guidance and Computer Assistance for Soft-Tissue Interventions*, 2008. [2.2.3](#), [2.2.3](#), [2.2.3](#), [2.2.3](#)
- [Wijlemans *et al.* 2012] J. W. Wijlemans, L. W. Bartels, R. Deckers, M. Ries, W. P. Mali, C. T. Moonen, and M. A. van den Bosch. Magnetic resonance-guided high-intensity focused ultrasound (MR-HIFU) ablation of liver tumours. *Cancer Imaging*, 12(2):387–394, 2012. [1.1](#)
- [Wiskott and Sejnowski 2002] L. Wiskott and T. J. Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *NC*, 14(4):715–770, 2002. [6.2.1](#), [6.5](#)
- [Woo *et al.* 2009] J. Woo, B. W. Hong, C. H. Hu, K. Shung, C. C. Kuo, and P. Slomka. Non-Rigid Ultrasound Image Registration Based on Intensity and Local Phase Information. *J Signal Process Syst*, 54:33–43, 2009. [2.2.5](#)
- [Wright *et al.* 2012] C. Wright, K. Hynynen, and D. Goertz. In vitro and in vivo high-intensity focused ultrasound thrombolysis. *Invest Radiol*, 47(4):217–225, 2012. [5.1](#)
- [Wu *et al.* 2005] F. Wu, Z. B. Wnag, H. Zhu, W. Z. Chen, J. Z. Zou, J. Bai, K. Q. Li, C. B. Jin, F. L. Xie, and H. B. Su. Feasibility of US-guided High-Intensity Focused Ultrasound Treatment in Patients with Advanced Pancreatic Cancer: Initial Experience. *Radiology*, 236:10341040, 2005. [1.1](#)
- [Xiao *et al.* 2007] D. Xiao, W. S. Ng, C. B. Tsang, and U. R. Abeyratne. A region and gradient based active contour model and its application in boundary tracking on anal canal ultrasound images. *Pattern Recognit*, 40(12):3522–3539, 2007. [2.2.4](#), [2.2.4](#), [2.2.4](#)
- [Yang *et al.* 2008] L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu. 3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers. In *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pages 1–8, 2008. [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#)

- [Yeung *et al.* 1998a] F. Yeung, S. F. Levinson, D. Fu, and K. J. Parker. Feature-adaptive motion tracking of ultrasound image sequences using a deformable mesh. *IEEE Trans Med Imaging*, 17(6):945–956, 1998. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Yeung *et al.* 1998b] F. Yeung, S. F. Levinson, and K. J. Parker. Multilevel and Motion Model-Based Ultrasonic Speckle Tracking Algorithms. *Ultrasound Med Biol*, 24(3):427–441, 1998. [2.1.3](#), [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Yilmaz *et al.* 2006] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys*, 38(4), 2006. [2.1](#)
- [Yoshikawa *et al.* 2005] H. Yoshikawa, T. Azuma, K. Kawabata, Y. Taniguchi, and S. Umemura. Three-Dimensional Tracking Method of Tissue Motion with Biplane Images. *Jpn J Appl Phys*, 44(6B):4561–4566, 2005. [2.2.2](#), [2.2.2](#), [2.2.2](#)
- [Yu and Acton 2002] Y. Yu and S. T. Acton. Speckle reducing anisotropic diffusion. *IEEE Trans Image Process*, 11(11):1260–1270, 2002. [2.1.4](#)
- [Yu *et al.* 2003] W. Yu, N. Lin, P. Yan, K. Purushothaman, A. Sinusas, K. Thiele, and J. S. Duncan. Motion Analysis of 3D Ultrasound Texture Patterns. In *Functional Imaging and Modeling of the Heart*, volume 2674 of *LNCS*, pages 253–261, 2003. [2.2.2](#)
- [Yu *et al.* 2006] W. Yu, P. Yan, A. J. Sinusas, K. Thiele, and J. S. Duncan. Towards pointwise motion tracking in echocardiographic image sequences Comparing the reliability of different features for speckle tracking. *Med Image Anal*, 10(4):495–508, 2006. [2.2.2](#)
- [Zhang and Wang 2010] L. Zhang and Z. B. Wang. High-intensity focused ultrasound tumor ablation: Review of ten years of clinical experience. *Front Med China*, 4:294–302, 2010. [1.1](#)
- [Zhang *et al.* 2010] X. Zhang, M. Günther, and A. Bongers. Real-Time Organ Tracking in Ultrasound Imaging Using Active Contours and Conditional Density Propagation. In H. Liao, P. Edwards, X. Pan, Y. Fan, and G. Z. Yang, editors, *Medical Imaging and Augmented Reality*, volume 6326 of *LNCS*, pages 286–294. Springer, 2010. [2.2.4](#), [2.2.4](#), [2.2.4](#), [2.3](#), [2.4](#), [3.6](#), [4](#)
- [Zheng *et al.* 2008] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. *IEEE Trans Med Imaging*, 27(11):1668–1681, 2008. [2.2.5](#), [2.4](#), [2.4](#)
- [Zitova and Flusser 2003] B. Zitova and J. Flusser. Image registration methods: a survey. *Image Vis Comput*, 21:977–1000, 2003. [2.1](#), [2.3](#)

List of publications

Journal Publications

1. L. Petrusca, P. Cattin, V. De Luca, F. Preiswerk, Z. Celicanin, V. Auboiroux, M. Viallon, P. Arnold, F. Santini, S. Terraz, K. Scheffler, C. D. Becker, and R. Salomir. Hybrid Ultrasound/Magnetic Resonance Simultaneous Acquisition and Image Fusion for Motion Monitoring in the Upper Abdomen. *Invest Radiol*, 48(5):333-340, 2013.
2. V. De Luca, G. Szekely, and C. Tanner. Ultrasound Tracking: a Survey. *Med Image Anal*, 2013. submitted.
3. F. Preiswerk, V. De Luca, P. Arnold, Z. Celicanin, L. Petrusca, C. Tanner, R. Salomir, and P. C. Cattin. Model-Guided Respiratory Organ Motion Prediction of the Liver from 2D Ultrasound. *Med Image Anal*, 2013. submitted.

Conference Proceedings

1. L. Petrusca, V. De Luca, P. Arnold, Z. Celicanin, T. Goget, V. Auboiroux, M. Viallon, F. Santini, S. Terraz, K. Scheffler, C. Tanner, P. Cattin, and R. Salomir. Ultrasound/MR hybrid imaging: truly simultaneous motion monitoring in the abdomen and image co-registration. In *Proc Int Soc Magn Reson Med*, volume 19, 2011.
2. V. De Luca, H. Grabner, L. Petrusca, R. Salomir, G. Szekely, and C. Tanner. Keep Breathing! Common Motion Helps Multi-modal Mapping. In *Med Image Comput Comput Assist Interv*, volume 6891 of *LNCS*, pages 597-604. Springer, 2011.
3. V. De Luca, C. Tanner, and G. Szekely. Speeding-up Image Registration for Repetitive Motion Scenarios. In *Proc IEEE Int Symp Biomed Imaging*, pages 1355-1358, 2012.

4. V. De Luca, M. Tschannen, G. Szekely, and C. Tanner. A Learning-based Approach for Fast and Robust Vessel Tracking in Long Ultrasound Sequences. In *Med Image Comput Comput Assist Interv*, volume 8149 of *LNCS*, pages 518-525. Springer, 2013.

Book chapters

1. L. Petrusca, M. Viallon, S. Terraz, V. De Luca, Z. Celicanin, V. Auboiroux, S. Brunke, P. Cattin, and R. Salomir. Simultaneous Ultrasound Imaging and MRI Acquisition, pages 457-470. *Medical Radiology*. Springer, 2012.

Curriculum Vitae

Personal Data

Name Valeria De Luca
Date of birth 25th December 1984
Place of birth Chieti (CH), Italy
Citizenship Italian

Education

2010 – 2013 *ETH Zurich, Computer Vision Laboratory, Switzerland*
Doctoral studies
2006 – 2009 *University of Roma Tor Vergata, Italy*
Master of Science in Medical Engineering
2003 – 2006 *University of Roma Tor Vergata, Italy*
Bachelor in Medical Engineering
1998 – 2003 *Liceo Scientifico F.Masci, Chieti, Italy*
High school Diploma in scientific studies

Work Experience

2010 – 2013 *ETH Zurich, Computer Vision Laboratory, Switzerland*
Research assistant
2008 – 2010 *Duke University Medical Center, Durham, NC, USA*
Visiting graduate student and researcher at the
Hyperthermia Laboratory of the Radiation Oncology Department
2006 *Policlinico Gemelli, Roma, Italy*
Researcher at the Ophthalmology Department