UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXVIII CICLO DEL DOTTORATO DI RICERCA IN FISICA

Borsa di Dottorato FSE S.H.A.R.M.

Automated Deformable Registration of Breast Images:

towards a software-assisted multimodal breast image reading

Settore scientifico-disciplinare: FIS/07 FISICA APPLICATA

DOTTORANDA
MATILDE COSTA

COORDINATORE
PROF. PAOLO CAMERINI

SUPERVISORE DI TESI
PROF. RENATA LONGO

CO-SUPERVISORE DI TESI
DR. SILVIA PESENTE

ANNO ACCADEMICO 2014 / 2015

Abstract

Early diagnosis is an important challenge for the effectiveness of breast cancer treatment. If the cancer is detected in the early stages, it can be more effectively treated, and a less invasive treatment can be achievable. To this aim an accurate routine screening plays a key role. Modern breast diagnosis is based on the image analysis; nowadays Mammography is the gold standard technique for breast screening. However, recent studies have shown that the combined use of this imaging modality with Ultrasound examination leads to an improved diagnostic accuracy. Therefore, an effective first level breast image exam combines both Mammography and Ultrasound, moreover the careful comparison of prior mammograms is mandatory. Breast imaging is rapidly evolving toward 3D examination: Digital Breast Tomosynthesis and Automated Breast Ultrasound scanners represent the most innovative and promising technologies. However, the 3D modalities increase dramatically the number of images to evaluate and for these new techniques are not yet available adequate visualization tools that allow to properly develop the potentials of the 3D imaging. Moreover the breast is a highly deformable structure, and this greatly complicates the visual comparison of imaging modalities. Due primarily to the challenges posed by the relevant non-rigid deformations, the great variability in the breast anatomy and the lack of rigid structures, the development of automated methods which enable registration of information within and across breast imaging modalities remains an active research field, which has still to translate suitable methods into clinical practice.

The present PhD research is a pioneering work in this field, especially in 3D US registration. It focus on the feasibility of the most modern registration techniques, properly customized for breast imaging. The aim is to provide automated spatial correlation tools between two 3D breast images, addressed to support physicians in the image analysis and comparison.

Contents

Chapter 1

Introduction and outline of the thesis

Contents

This thesis involves concepts from two different scientific fields. On one hand, methods and algorithmic tools from image processing are examined and designed. On the other hand, technical issues coming from the medical field, and more specifically from cancer diagnosis context, need to be studied and understood. In this first chapter an overview of the clinical context is presented to introduce the motivation and main challenges of the work. Firstly, the clinical problem and the current incidence of the disease on the population are reported. Then, a background of the breast screening program and imaging techniques is proposed, in order to explain the clinical needs faced with this work.

1.1 Breast cancer and early detection

Cancer is a disease that affects more and more people of all ages, being actually one of the most important causes of human deaths. It is a degeneration of

the normal function of a cell. Due to some genetic abnormalities, cancer cells have uncontrolled growth and can affect healthy tissues. A malignant tumor is a group of cancer cells that can grow into (i.e. invade) surrounding tissues, and typically form a solid lesion, or spread (i.e. metastasize) to distant areas of the body. Breast cancer is a malignant tumor that starts in the cells of breast. It represents an important issue for medical research, because it is a disease with a high incident rate and it causes the most cancer deaths in women all over the world (T. Chen et al. 2012). The International Agency for Research on Cancer (IARC), the specialized cancer agency of the World Health Organization, in 2013 released the latest data on cancer incidence, mortality, and prevalence worldwide (Ferlay et al. 2013). In this report, the Agency highlights the priority should be given to cancer prevention and control measures for breast and cervical cancers and states that breast cancer is the second most common cancer overall. Breast cancer represents about 12% of all new cancer cases and 25% of all cancers in women (*World Cancer Research Fund International*). In 2012, 1.7 million women were diagnosed with breast cancer. Since the 2008 estimates, breast cancer incidence has increased by more than 20% and it has been increasing in most regions of the world. The highest incidence of breast cancer was in Northern America and Oceania; however, Belgium was the country with the highest rate of breast cancer, followed by Denmark and France. Italy covered the thirteenth rank in this breast cancer statistic. The lowest incidence is in Asia and Africa. Meanwhile incidence rates remain highest in more developed regions, mortality is relatively much higher in less developed countries due to a lack of early detection and access to treatment facilities. For example, in western Europe, breast cancer incidence has reached more than 90 new cases per 100,000 women annually, compared with 30 per 100,000 in eastern Africa. In contrast, breast cancer mortality rates in these two regions are almost identical, at about 15 per 100,000, resulting in a poorer survival in eastern Africa. An important trend shown by statistics (referring to the population of the United States) is that the death rates have been decreasing in the last decades; this decrease is thought to be the result of treatment advances, earlier detection through screening, and increased awareness (*Breastcancer.org*). Due to the invasive nature of the disease, the detection in its early stage is crucial to reduce the mortality for breast cancer. This increases the effectiveness of the treatment (Sivaramakrishna et al. 1997) and avoids cancer growth and extent in multi foci (multifocal breast cancer), or in more sites inside the breast (multicentric breast cancer), or even worse its spread from the original site to other parts of the body (metastatic breast cancer). The larger the tumor size at detection time, the smaller is the survival probability of the patient (Ducan et al. 1976) and additionally more invasive is the treatment. If the breast cancer is in an early stage, breast conservative surgery is possible against mastectomies or other aggressive treatment like chemotherapy, that is an important issue for the psychological implications in cancer patients. Research has proved that

breast conserving treatment combined with radiotherapy works as well as mastectomy at treating early breast cancer (*Cancer Research UK*). The stage of the breast cancer is usually expressed as a number on a scale of 0 through IV, with stage 0 describing non-invasive cancers that remain within their original location and stage IV describing invasive cancers that have spread outside the breast to other parts of the body. The goal of screening is the detection of breast cancer when it is stage 0, or stage I (i.e. cancers < 2 cm in size, invasive, without spread to axillary lymph nodes). In particular, the best prognosis is seen with cancers smaller than 1 cm in size (Tabar et al. 2000, R. Smith et al. 2004). Due to the unknown cause of breast cancer and the multiple advantages of the early diagnosis, it is very important to convince the women of the core role of regular breast screening. As explained in the following, X-ray mammography is the standard imaging method for breast cancer screening. Mammography screening has been shown in clinical trials to reduce breast cancer mortality by 25% to 30% for women in the 50 to 70 age group (Tabar et al. 1995, Tabar et al. 2001). Additionally, it is proved that the reduction in mortality attributable to widespread mammographic screening is almost entirely due to early detection (Paci et al. 2002). A mammogram can find breast changes that could be cancer years before physical symptoms appear. Results from many decades of research show that women who have regular mammograms are more likely to have breast cancer found early, less likely to need aggressive treatment, and more likely to be cured.

1.2 Breast imaging techniques

Breast cancer screening refers to exams used to check a woman's breasts for cancer, before symptoms of the disease appear. The goal of screening tests is to find diseases at an early stage, when treatment is most likely to be successful. Size and stage of breast cancer are some of the most important factors in predicting the prognosis of a woman, and thus are crucial to the survival of the patient. A number of screening tests are employed, including clinical and self breast examination (i.e. palpation) to feel for lumps or other changes, and imaging techniques used to screen for tumors and other nonpalpable abnormalities.

Sensitivity and specificity of a test

Commonly, the diagnostic accuracy of an imaging technique is evaluated by means of the sensitivity and the specificity parameters. Sensitivity, also called the true positive rate, stands for the ability of a test to correctly identify the tumor when it is present (true positive). It measures the percentage of positives (people with disease) that are correctly identified as such. Therefore, a test with 100% sensitivity, correctly identifies all patients with the disease; a test with a lower sensitivity, for instance 80%, detects 80% of sick patients

(true positive) but 20% of patients with disease go undetected (false negative result of the test). Specificity, also called the true negative rate, refers to the ability to correctly identify the nature of a suspicious mass, distinguishing benign tumor from malignant (true negative). It measures the percentage of negatives (healthy people without the disease) that are correctly identified as such. Therefore, a test with 100% specificity, correctly identifies all patients without the disease. A test with 80% specificity correctly reports a negative test result for 80% of patients without the disease (true negatives), but 20% patients without the disease are incorrectly identified as positive (false positives). Hence, high sensitivity is important for a screening test to identify the disease, avoiding false negative result that may delay the medical care. But if the test is not very specific, patient without disease could undergo the treatment. There is usually a trade-off between these parameters in the choice of the imaging technique. A test used in a screening program must have good sensitivity in addition to acceptable specificity, avoiding as more false positives as possible. A brief description of conventional imaging modalities used for breast cancer diagnosis is now presented.

Imaging techniques

Different types of breast cancer exist and the examination through breast palpation is often not sufficient to detect the cancer; some fibroadenomas are so small that they cannot be felt by physician. The modern breast diagnosis is based on image analysis. Medical imaging techniques measure some physical or chemical reaction of the patient's body after using a specific stimulus (e.g. X-rays, electromagnetic pulses, radioactive tracers, etc.), and image the different responses of the tissues. Depending on the physical principle, each technique is thus sensitive to specific tissue properties. In order to be suitable for a diagnostic examination the imaging technique need to observe the following requirements: fast image acquisition; be a non-invasive exam (in term of effects on patient's health); be a reliable exam, with high sensitivity and acceptable specificity. The most common imaging modalities currently used for breast cancer diagnosis are X-ray, Magnetic Resonance Imaging (MRI), and Ultrasound (US). Each imaging technique has its own advantages and weaknesses in term of diagnostic accuracy. Hence, there is not a unique solution proper for all patients (that differ for age, history, physiological state, breast anatomy, etc.). In addition, breast cancer is a heterogeneous disease, and consequently, a single screening modality has limitations in imaging all the different subtypes. A multimodality approach that combines techniques providing complementary pathological information, should be used to diagnose asymptomatic women (Lander et al. 2011). The conventional system for breast X-ray imaging is Mammography (MG), that provides a 2D projection image (called mammogram) of the breast compressed between two flat paddles. The mammogram describes the degree of X-rays attenuation inside the

breast, where image intensity is related to the tissue's density. The greater the difference between the densities, the greater the image contrast. MG is still the imaging modality most widely accepted as gold standard for breast cancer screening. Although the sensitivity of MG in women with fatty breast tissue (less than 50% glandular) can be as high as 97% (Kolb et al. 2002), ability of mammography to detect breast cancer dramatically decreases in women with dense breasts, falling to as low as 48% (Inciardi 2012). A dense tissue typically appearing bright in mammograms would hide structural abnormalities such as cancerous tumors; furthermore most cancers and dense breast tissues have similar appearance on mammogram. it results in an increased number of false negatives (missed cancers) with screening MG in women with dense breast tissue. It has been reported that mammography alone can miss between 37 and 70 percent of breast cancers within dense breast population (Jacob 2012). Therefore, MG alone cannot provide a reliable cancer diagnosis of women with dense breasts and there is a significant clinical need for an alternative or supplemental method. Many studies have shown that US and MRI can help to detect breast cancers that may not be visible with MG. MRI is a non-ionizing technique; it uses the magnetic properties of spinning hydrogen atoms and radiowaves, instead of X-rays, to produce high-resolution 3D images of the breast anatomy. Typically MRI makes use of a contrast agent to enhance the details of breast tissue and to provide dynamic information on tissue vascularization. Therefore, MRI can be used especially to differentiate tissues with similar density. It is typically used with success in the diagnosis of high-risk women (i.e. women with a lifetime risk of over 20%, women who have a positive test result to the so-called BRCA genetic mutation, or have multiple first-degree relatives with a history of premenopausal breast cancer). It is also effective for women with dense breast, however, its application for breast examination is limited for the high cost of the exam and for the use of the contrast agent. US uses high-frequency sound waves that allow a real-time characterization of breast lesions. It has the potential to be an ideal diagnostic tool because less expensive compared with the other modalities, well-tolerated and safe. Breast US is commonly used as an adjunct to MG in patients with palpable masses or symptomatic breast disease (Guo et al. 2006, Hooley et al. 2013) or to evaluate the mammographic findings for its ability to differentiate cystic to solid lesions. It is shown that combining MG and US cancer detection is improved by about 2.9% (Girardi et al. 2013); nevertheless, US may increase false-positive biopsies. Moreover, breast density does not seem to influence ultrasound accuracy, therefore US can be used as supplemental diagnostic tool to detect cancers missed by MG in dense breasts (Berg et al. 2008, Kaplan 2001, Berg, Z. Zhang, et al. 2012, Inciardi 2012). US breast examination is also considered in women at very high risk for breast cancer who cannot tolerate breast MRI. In common practice US consists of a freehand transducer that enables to collect images corresponding to a breast section, captured in real-time during the breast examination. When the physician

performs the targeted US, he has to direct the US transducer to the specific parts of the breast of concern, where the suspicious mammographic findings are potentially located. Hence, the physician normally reviews the mammogram before US to know what to search and where to search the lesion. During the examination, it is important to recognize and correlate the area that is being examined with US with the lesions detected in MG. The correlation is complicated by the fact that the breast is scanned in different positions and that both modalities provide only 2D images. In addition, US captures only a section of breast, where breast structures are not represented at the same scale of analysis of MG (see figure 1.1). Therefore, US examination requires to have a solid knowledge of breast anatomy, a good understanding of the imaging techniques and strong interpretive US skills. A relevant shortcoming of
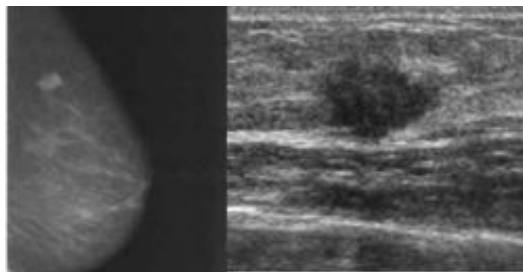

Figure 1.1: example of the appearance of a lesion on a mammogram, on the left, and on a US capture, on the right. The different scale representation of the lesion between the two modalities combined with the limited field-of-view of the US image, make the correlation a very difficult task. The correlation is dependent to the skill of physician who interprets the MG and performs the US examination. Note that in the US image any information about both probe orientation and lesion location with respect to nipple position or surrounding breast structures are lost and the same image plane at the same location is not reproducible (Lin et al. 2011).

US is that only a 2D cross-sectional image of breast with limited field-of-view can be captured. The lack of any spatial reference with respect to the whole volume disallows to localize the image plane, and consequently, to correlate it with other images of the same patient. A retrospective study with images previously collected, especially if acquired with a different modality, such as MG, is not feasible.

1.3 3D breast imaging

From the previous considerations it emerges that the comparison between breast examinations is a highly operator-dependent procedure and a retrospective study of US images is not currently feasible. An important step towards the overcoming of this limitations is represented by the technological

advances that enable the collection of 3D images of breast. In the last decades, many efforts have been made by industries for moving from 2D to 3D imaging modalities. Recently, two innovative technologies have been introduced as promising breast diagnostic techniques, the X-ray Digital Breast Tomosynthesis (DBT) and the Automated Breast Ultrasound (ABUS) scanner. DBT is an apparatus similar to the MG. The breast is positioned in the same way as in a conventional MG, but the X-ray tube moves in an arc around the breast taking multiple X-ray projections at different angles. A 3D dataset is then reconstructed from this sequence of projections. The main advantage of DBT is that lesion conspicuity is improved because overlaying tissue structures are resolved. ABUS consists of a linear high-resolution US transducer placed over the breast that automatically captures a spatial sequence of two-dimensional cross-sectional slices to reconstruct a volumetric information of the breast. It allows the collection of 3D high-quality images, with a standardized and less operator-dependent procedure. Besides an improved diagnostic accuracy, these systems represent also a plausible alternative to overcome some of the before mentioned limitations of the conventional systems: 3D imaging preserves the actual characterization of both tissues and abnormalities and their location inside the breast; moreover two 3D scans of the same region are expected to represent all the structures in that region in both images. This is crucial for a spatial correlation between two different exams of the same patient.

1.4 Clinical needs

For a more accurate diagnosis, physicians compare different examinations for each patient. Normally, two different types of image comparison could be made: a cross-modality comparison between DTB and US images, as well as the comparison of images collected with same modality but at different times. Indeed, worldwide accepted indications (D'Orsi et al. 2013) for breast cancer diagnosis recommend the comparison of the examination with prior ones for an accurate and complete diagnosis. From the previous considerations it is clear that an accurate image comparison is feasible only taking advantage of the novel three-dimensional techniques. The comparison of two volumes, however, is not straightforward and it is time-consuming if manually performed by physicians. Currently the comparison between two different exams performed by physicians is only qualitative (i.e. visual), and it requires a lot of experience due to the complex anatomy of the breast and its non-rigid behavior that can induce relevant local deformations.

1.5 Image registration

Image registration is the mathematical tool that allows to determine a spatial mapping function between two volumes. It is useful when different images of the same patient are represented within different coordinate systems compared to the reference of the patient or/and use different image intensities to represent the underlying anatomy. There are rigid methods that establish a global transformation model between the reference systems of the two volumes, and more complex non-rigid methods aimed to evaluate pointwise correlations between the volumes to handle local deformations. The deformable methods are those more appropriate for breast images. The registration is an ill-posed problem that requires to include a priori knowledge and constraints to solve it. These constraints have to be specific for the application.

1.5.1 Challenges in breast applications

Image registration plays an important role in breast imaging, where reproducible imaging and patient positioning are very difficult due to the breast morphology and the high degree of its mobility. It has the task of finding corresponding locations between two breast images not always straightforward visible owing to relevant local deformations and the low image contrast between breast tissues. Deformations, resulting from breast motion, compression or size change must be compensated during the registration. Hereafter, a list of specific features of the breast is reported, that make breast image registration a very challenge task. This includes also the technical issues to be taken into account in the development of the registration methods.

Key issues

Highly deformable anatomy
Breast consists entirely of soft tissues. Thus it has an inherent non-rigid behavior and can move freely because not constrained by any surrounding organs. In case of highly deformable anatomies, the registration method has to handle large deformations avoiding discontinuity of the registration field and unrealistic distortions of structures. Usually smooth transformations are preferred and large displacements, like deformations shown by tissues during breast compression, are uneasy to be estimated. Moreover, breast shows discontinuity in breast tissue deformations similar to sliding motion between adipose and glandular tissue, as well as respect to chest wall, especially under the pressure of the transducer during ultrasound acquisitions. This corresponds to discontinuity in the transformations that generally is not allowed in most registration methods. In addition, due to the large breast motion an accurate and robust pre-alignment procedure is required and can be the key issue for the success of the method.

Heterogeneous composition
The inhomogeneous, anisotropic nature of the soft tissues within the breast, as well as their low contrast characterization make the image interpretation more complex.

Various imaging conditions
Breast imaging modalities typically investigate the breast with patient lying in different positions: for instance, during X-ray imaging patient is standing in front of the apparatus and breast is pulled out and squeezed between two pads, in US she lies in supine position and breast is flattened and undergoes the US probe pressure (figure 1.2). Therefore, geometric deformations can appear between breast images as:

- local deformations due to different breast orientation and amount of compression, as soft nature of breast tissue makes the imaging conditions less reproducible;

- relevant both global and local deformations due to different patient positioning when images are collected with different modalities.
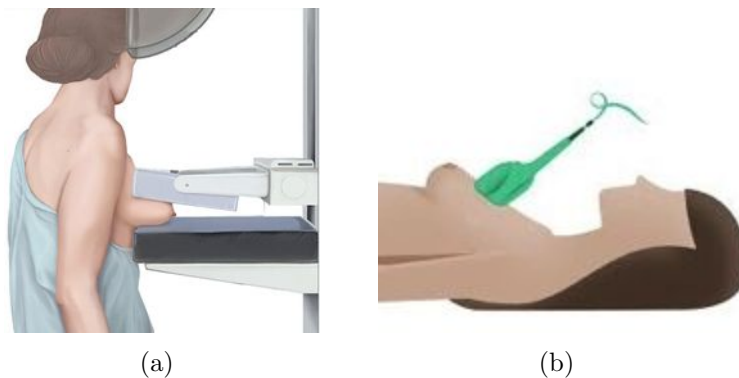

(a)   (b)

Figure 1.2: breast position during MG (on the left) and US scanning (on the right).

Physiological changes
Breast is an organ with a high anatomical variability, both in an individual over time and even inter-subject. In a patient breast anatomy can change over time for different physiological factors, like hormonal cycle as well as the weight loss. These temporal changes are quite difficult to predict or model. Therefore, the appearance of breast tissue of two images collected at different times could be significantly different, making the registration more complicated. Further, the wide variability over the population does not contribute to identify common consistent features inside breast (i.e allowing a reliable classification of their shape and appearance) that could be taken as reference for the registration task.

Lack of a priori information

Normally the registration methods take advantage of a priori information on the deformation model (i.e. on both the biomechanical properties of the organ and the imaging technique) underlying the differences shown by the images to be registered. As it will be explained later, it results from the fact the registration is an ill-posed problem. Therefore, in the registration method additional constraints have to be integrated to solve the mathematical problem. Especially for highly deformable anatomies constrained methods are usually proposed, and the constrains have to be carefully established to avoid undesired results. In the context of medical imaging, a lot of supplementary information is usually used for improving the registration result. Among them, manual or automatic organs segmentations can be incorporated, as well as landmarks, rigidity constraints and biomechanical models. However, breast lacks of these features: there are not structures with rigid behavior like bones, the contouring of clinically relevant structures is not a common practice in breast screening, and breast lacks of distinctive features to be used as reliable fiducial markers. Normally for breast the nipple is the unique reliable reference feature. Moreover, due to its highly deformable behavior and its heterogeneity an accurate biomechanical breast modeling is not an easy task.

Therefore, from these considerations can be concluded that the direct application to breast images of general-purposes methods is not practicable and the need of a customized solution emerges.

1.5.2 State-of-the-art of breast image registration

The development of innovative and sophisticated registration algorithms is an active research field, that can count on the contribution of a wide scientific community. Nevertheless the deformable registration is still an unresolved issue for most applications, including breast imaging. Several reviews may be found that report the most popular methods proposed in literature for a wide range of applications (Maintz et al. 1998, Zitova et al. 2003, Crum et al. 2003). Survey papers have been published also in the field of medical image registration (Hill et al. 2001, Sotiras et al. 2013); however, very few reviews are dedicated to breast application (Sivaramakrishna 2005, Guo et al. 2006). Several articles propose methods using deformable breast models based on a finite-element method (FEM) to solve both intra- and inter-modality registration tasks (Coman et al. 2004, Ruiter et al. 2004, Unlu et al. 2005, Hopp et al. 2012). A recent project led by a group of researchers of the Institute Fraunhofer MEVIS (*Fraunhofer MEVIS*) is dedicated to study registration solutions for multimodal breast imaging. Their aim is to map all breast images into a unified breast model and synchronizes corresponding positions between different images in real-time. In a recent work (Georgii et al. 2013) the authors

presented a method to compute the spatial correlation between the two DBT projections of the breast (i.e. MLO and CC). Recently, some studies have been carried out to assess the clinical impact of the use of 3D automated US in conjunction with DBT (Padilla et al. 2013, X. Zhang et al. 2014), but, on my knowledge, there are not published works proposing a method for the registration of 3D-US and DBT images.

1.6 Purpose of the PhD work

This work is aimed to study the feasibility of a customization of the general deformable registration methods to the specific case of breast images: the evaluation should estimate if it is possible to derive reliable and accurate registration algorithms able to overcome the serious issues that such a highly deformable organ poses, taking advantage of the advanced 3D acquisition techniques mentioned in the previous sections, together with novel or customized approaches to image registration. The main goal is to give support to physicians in the comparison task of two breast volumes through software tools able to automatically estimate pointwise correlations among data.

The specific tasks faced within the project are:

(a) **temporal comparison of two examinations**: it means the comparison of two volumes of a breast collected using the same modality, for example using DBT, but at different times. The goal is to remove image differences introduced by different breast positioning, allowing to highlight only the changes in breast tissue over time. These changes often reflect suspicious changing of the normal breast tissue structures resulting in the appearance of suspicious masses or some architectural distortions of tissue. Therefore, it is important to evaluate these abnormalities comparing with prior assessment as they can be new symptoms inherently suspicious or suspicious because more extensive or with more irregular shape than before. However, corresponding locations between two 3D images could be not always straightforward visible due to local displacements introduced by the imaging technique (i.e. different breast positioning, breast compression, etc.) as well as physiological changes of breast tissues (i.e. changing in parenchyma's pattern, appearance of pathology, etc.). Moreover in case of breast dense tissue the overall image contrast decreases. Therefore, manually comparison is not straightforward and, additionally, takes time requiring to scroll the high number of slices of the two volumes. The automated correlation is aimed to make faster the image comparison task and more reliable the identification of corresponding regions between the exams;

(b) **comparison of two examinations from different viewpoints**: it means the comparison of two or more volumes of a breast collected from

different viewpoints as the sensor has a limited field-of-view to image the entire volume. This regards the particular case of the US volumes collected with the ABUS (briefly introduced in the previous section of imaging techniques, and explained in more detail in the second chapter). This type of scanner requires the acquisition of at least three scans of each breast at different viewpoint (i.e. collected positioning the US probe onto different breast sections) to be sure of investigating the whole area of interest of more likely cancer. In this case each view shows a section of the breast (more details are reported in the second chapter). For a complete diagnosis, the physician should analyze at least three volumes comprising high number of slices. In addition, among these three different scans relevant deformations are impressed under the compression of the US probe. These deformations make the comparison of the sections a complex task, increasing also the time required for the interpretation. In this case, an automated correlation of the structures in the overlap area among two adjacent views can be helpful for a more reliable and saving time analysis;

(c) **cross-modality comparison of two examinations**: it refers to the comparison of two breast examinations collected using different imaging modalities, such as DBT and 3D US images. It is aimed to take advantage from the combination of different information to improve the diagnostic accuracy. In this case it is a very complex procedure due to the completely different patient positioning and the different physical principles of the two modalities (that implies a different image interpretation), but it is helpful to make use of complementary anatomical information. The visual correlation of these two examinations requires specific skills and a wide experience, enforced even by the huge variability of the breast tissue among patients. In this case a simultaneous navigation inside these volumes, automatically spatially correlated, could be useful to support comparison of the two exams.

1.7 European Social Fund fellowship and collaborations

My PhD scholarship has been co-financed by the University of Trieste and the Friuli-Venezia Giulia Region that participated to the S.H.A.R.M. (Supporting Human Assets in Research and Mobility) project under the European Social Fund 2007/2013 (Operational Programme Objective 2 - Regional Competitiveness and Employment of the ESF 2007-2013 - Axis IV Human Capital). The aim of this European project was to encourage the collaboration between the academic environment and local small-medium enterprises (SMEs). In accordance with this perspective, during the PhD project the collaborations

with the R&D group of Tecnologie Avanzate TA Srl and Datamind Srl have been established. Both SMEs are settled at the science and technological park in Udine. Tecnologie Avanzate TA Srl is a leading company on the national territory for the distribution of novel high-tech medical devices, specific for imaging, quality assurance and treatment applications, offering also technical assistance on their products. The company is also active in the field of the archiving and communication systems of medical data, developing advanced tools for reporting, visualization and segmentation process. Datamind Srl is a software house with strong experience in image processing and computer vision with a wide range of applications. In addition to the mentioned collaborations, my research work was performed in close collaboration with the physicians of the Radiology Department of A.O.U. Ospedali Riuniti of Trieste and the Clinical Center Michelangelo in Florence specialized in prevention of breast cancer, who have offered their clinical expertise in breast diagnostic examination. For the design and building of some custom-made hardware components used for the project, I could counted on the collaboration of the INFN mechanical workshop of Trieste and the electronics laboratory of the Physics department of the University of Trieste.

1.8   Organization of the thesis

The document is structured as follows: a first part (chapters 2 and 3) that describes the state-of-the art in the two main fields treated in this work, breast imaging and image registration; and a second part (chapters 4-6) dedicated to the original work of analysis, design, development and evaluation of the framework for breast registration.

In Chapter 2 the two reference breast imaging techniques are treated extensively describing their acquisition process and their physical principles. A special attention is given to the improvements in their 3D configuration.

In Chapter 3 an overview of the image registration theory together is presented with the most commonly used tools and methods for both rigid and non-rigid registration. It also introduces the formalism used in the subsequent chapters.

Chapter 4 describes an experimental setup built for the reconstruction of 3D US images used for preliminary studies in a controlled environment for the image registration techniques to be integrated in the developed framework. Additionally, the prototype, making use of conventional 2D US probes, suggests a possible way that can be further investigated to apply the developed registration algorithms even when an advanced 3D US scanner is not available.

Chapter 5 is focused on the architectural design of the whole framework and on its modularity that allows an easy derivation, given a reference clinical task, of an associated specialized registration algorithm. The last part of the

chapter describes and analyzes a series of classical and novel techniques (integrated in the framework) inherently bound to the registration process whose pertinence with a specific problem can be tested and compared.

In Chapter 6 the derivation of three different algorithms for the three different clinical tasks are taken into consideration in this work. This demonstrates the ability of the framework to be customized for the breast registration problems, regardless of the modality used and the type of deformation shown by breast between the two volumes.

Finally, the results are discussed and future research directions are given.

Chapter 2

Breast imaging techniques

Contents

This chapter focuses on the overview of the breast imaging techniques used in this work. The basic principles underlying the image formation, the apparatus description as well as their clinical application are explained. The attention is addressed to the novel imaging systems that allow to collect three-dimensional images of breast, i.e. X-ray Digital Breast Tomosynthesis and 3D Automated Breast Ultrasound scanners. The chapter starts with a brief description of breast anatomy useful to interpret the breast images treated in next sections.

2.1   Introduction to breast anatomy

Breast is the site of the mammary gland and it is an organ entirely composed of soft tissue, comprising glandular, connective and fat tissues, surrounded by skin. There are not muscles in this organ, but breasts overlay the pectoralis major muscles that cover the chest wall. At the front of the chest, the breast tissue can extend from the clavicle to the middle of the sternum. At the sides of the chest, the breast tissue can extend into the axilla. Breast heterogeneous

anatomy consists of ducts, lobes, lymphatics, blood vessels and nerves (see figure 2.1). The glandular tissue (also known as parenchyma, or simply breast tissue) is arranged radially about the nipple and separated from the skin by a variable layer of subcutaneous fat. Typically each breast has 15 to 20 lobes, and each lobe has many lobules, which end in dozens of tiny bulbs that can produce milk. The lobes, lobules, and bulbs are all linked by thin tubes called ducts. These ducts lead to the nipple in the center of the area of skin called areola. The spaces between lobules and ducts are filled with fat. The structural integrity of the breast is maintained by the connective tissue called Cooper's ligaments. Fat tissue (also known as adipose or fatty tissue) is present also behind the retromammary zone, in the prepectoral space. Each breast also contains blood vessels and vessels that carry lymph (or lymphatic) system. The lymph vessels lead to small lymph nodes, and clusters of which are distributed under the arm (axillary nodes), above or below the clavicle (supraclavicular or infraclavicular nodes), and inside the chest (internal mammary nodes), as well as in many other parts of the body. The breast can vary greatly in form, size, and composition, and is generally symmetrical in shape. The anatomy varies in an individual over time, owing to the mammary gland function regulated by hormones.
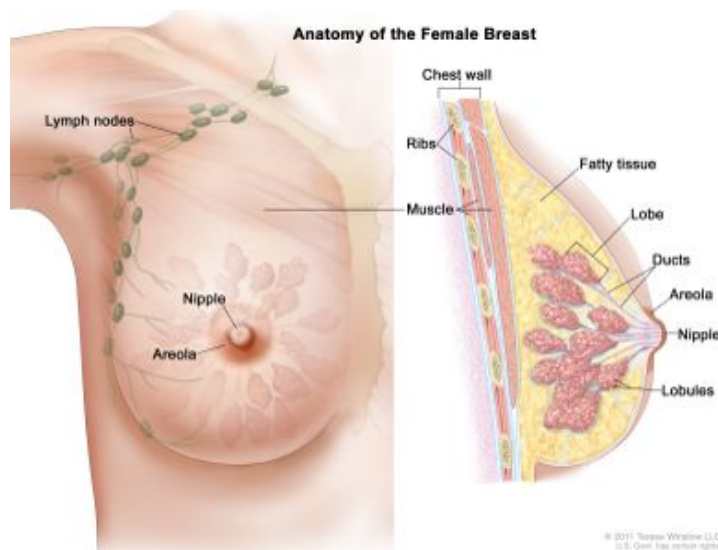

Figure 2.1: anatomy of the female breast: on the left a front view of breast; on the right a cross-section scheme of the mammary gland. The nipple, lymph nodes, lobes, lobules, ducts, and other parts inside of the breast are shown (*National Cancer Institute*).

Breast density and cancer risk

The relative amounts of glandular, connective and adipose tissue inside the breast establish the breast density. The amount of each type of tissue varies

over women. Dense breast is defined as having a higher percentage of glandular tissue within breast relative to fat: if more than 50% of breast is made up of glandular tissue, then the breast is classified as dense. Approximately 40% of women have dense breast tissue (Pisano et al. 2005). An evidence of the link between the percentage of glandular tissue and other parameters such as age of patient or breast size is not yet proved. Certainly, the breast density varies over a lifetime of patient, depending on the hormone activity. Generally breasts become less dense with increasing age: younger women have denser breasts and post menopausal women have progressively an increase of adipose tissue at the expense of a decrease of the amount of glandular tissue. (Checka et al. 2012) reports that over 74% of women in their 40s, 57% of women in their 50s, 44% of women in their 60s and 36% of women in their 70s have dense breast tissue. Other factors can play a role in density changing; for instance, women on estrogen therapy have denser breasts and women with significant mass loss have denser breasts compared with images done previously. Dense breasts require particular attention in screening, for the following reasons:

- breast density is an independent risk factor at any age: having dense breasts can increase a woman's risk to develop cancer by 6 times (Boyd et al. 2007). Furthermore, cancer developed in dense breasts is typically larger, of a higher grade and more frequently node-positive (found lymph glands) with poorer prognosis and a higher chance of returning and spreading than in fatty breasts (Boyd et al. 2007, Yaghjyan et al. 2011);

- dense breast tissue can compromise cancer detection on mammographic imaging (i.e. the sensitivity of the exam, as it will be explained later), because it can potentially hide breast cancers: glandular tissue has often low contrast against cancer tissue in the images, making harder to detect breast cancer.

Hence, determining the breast density of each patient is very important, in order to define a proper screening program. And, consequently, the need of establishing a common procedure to classify the density emerges.

2.2 Digital Breast Tomosynthesis

In this section a brief description of the X-ray image production is presented. Then, the main features of MG and DBT systems are reported to introduce the advantages of collecting 3D breast data.

2.2.1 X-ray imaging

It is an imaging technique that uses the penetrating ability of ionizing radiation (i.e. X-ray photons that carry enough energy to ionize atoms and break

up molecular bonds) to produce an image of internal structures of the human body. The image results from the different attenuation properties of the soft tissues. X-rays used in diagnostic radiology have energies above 5–10 keV.

Basic principles

X-ray radiation for medical imaging is typically produced by an X-ray tube by focusing and accelerating a beam of high-energy electrons towards a metal target. The X-rays are produced when the electrons hits and interacts with a metal target (called anode). The type of metal involved depends on the application. In medical X-ray tubes the target is usually tungsten, but sometimes molybdenum is used in MG when softer X-rays are needed. When the electrons hit the target, X-rays are created by two different atomic processes: characteristic X-ray emission and Bremsstrahlung. Hence, a typical X-rays spectrum consists of a broad continuous spectrum, generated when the incident high-energy electrons are slowed down rapidly in the target and release X-rays (Bremsstrahlung production), and peaks superimposed on the continuous spectrum, relating to the characteristic radiation for the target element. The voltage used to accelerate the electrons determines the maximum energy of the X rays, while tube charge determines the amount of radiation generated in the tube. Choice of the emitted radiation energies depends on the application and the tissues being imaged. For instance, figure 2.2 shows as low energies are preferred for breast imaging, where the contrast between adipose and glandular tissues is higher. Too high radiation energies will result in poor images since the radiation cannot be readily attenuated, however too low energies will increase the radiation dose to the patient since they are totally absorbed by the body, without contributing to the image (as it is explained in the following paragraph). The lowest energy of the photons reaching the patient is determined by the filtration of the beam. An X-ray filter consists of a thin metal foil, usually aluminum or copper, is usually placed over the window of the X-ray tube, in order to absorb the low energy part of the spectrum; this effect is called beam hardening, since it shifts the center of the spectrum towards harder, i.e. higher energy, X-rays. That means average photon energies in filtered spectra are higher than in the unfiltered ones. The selection of appropriate thickness of the filter is a balance between image quality and absorbed dose in the patient.

Interaction of X-rays with matter
Attenuation

Due to their high energy X-rays can penetrate into the material optically opaque. However, the number of photons (or the intensity of the beam) of a primary beam decreases exponentially as it passes through the irradiated object, and it is proportional to the thickness of the material and to the number of incident photons. This attenuation is described in a first approximation

(for a monochromatic X-ray beam) by the Beer-Lambert law:

$$I(x) = I_0 e^{-\frac{\mu(E)}{\rho}\rho x} \qquad (2.1)$$

where I_0 is the incident beam intensity, I(x) the intensity after passing through a thickness x of homogeneous material; constant of proportionality $\mu(E)$ is called linear attenuation coefficient and depends on the density and the atomic number of the material passed through, and on the energy of the beam. Typically the attenuation is expressed in terms of the mass attenuation coefficient $\mu(E)/\rho$, where ρ is the density of the material. It is usually measured in cm^2/g. X-rays are more attenuated by denser and thicker materials. Generally μ decreases with increasing energy. It expresses the probability per unit length that a photon of a given energy will interact during its passage through the material and it is the sum of the interaction processes. Figure 2.2 shows the measured linear X-ray attenuation coefficients of breast glandular tissue, breast fat and infiltration breast tumor plotted versus X-ray energy. It can be noted that there is a low contrast between the glandular and tumor tissue, meanwhile the difference between the attenuation coefficients of the fat and glandular tissues increases with decreasing energy. For X rays in the
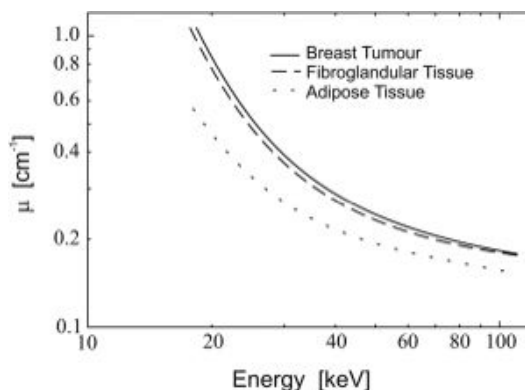

Figure 2.2: Linear X-ray attenuation coefficients of breast adipose and glandular tissues versus X-ray energy (Yaffe 2008).

diagnostic energy range, the main interaction processes that cause the beam attenuation are absorption (photoelectric effect) or scattering (Compton or Rayleigh scatterings). The X-rays responsible to image formation are the transmitted photons (i.e. primary photons emerging from the X-ray tube). A photon that is absorbed does not contribute to the image. A photon that is scattered changes its direction of motion and may lose some of its energy. This can contribute to image, but decrease the image quality (i.e. it contributes to increase the scatter to primary ratio (Bushberg et al. 2002)). The strength of these interactions depend on the energy of the X-rays and the atomic number of the material. Photoelectric absorption is the dominant interaction mechanism in the soft X-ray regime and for the lower hard X-ray energies. At higher

energies, Compton scattering dominates. For water and soft tissues absorption dominates up to about 26 keV photon energy (figure 2.3(a)), where Compton scattering takes over. For higher atomic number substances this limit is higher (figure 2.3(b)).

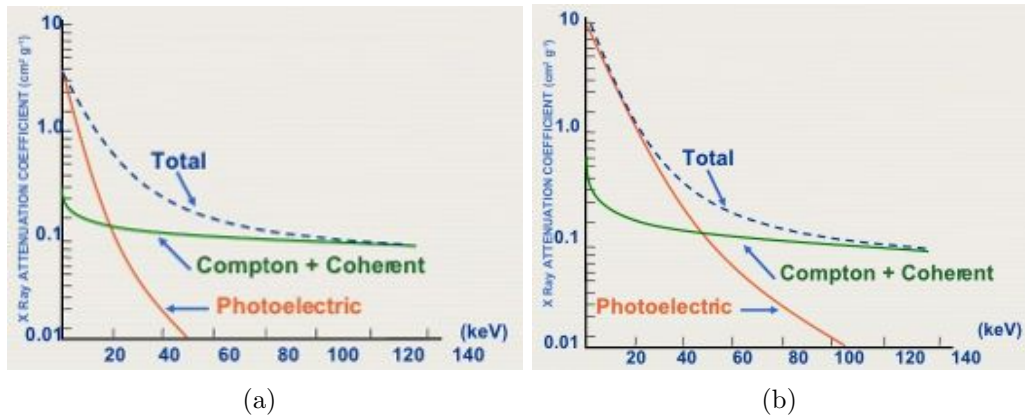(a)   (b)

Figure 2.3: Contribution of photoelectric and Compton interactions to attenuation of X-ray in water (a) and bone (b) as function of the X-rays energies (in the range of energies used for diagnostic and interventional radiology). These plots show that for bones the photoelectric effect occurs more likely, meanwhile for water Compton scattering dominates.

- Photoelectric absorption
  In a photoelectric absorption event, photon transfers all its energy to the electron with which it interacts, thus ionizing the atom to which the electron was bound and producing a photoelectron. The vacancy in the electron shell is filled by an electron from an outer shell and a new photon so-called characteristic photon is created. The probability for absorption varies rapidly with the atomic number of the atom. It decreases with increasing photon energy until the energy exceeds the binding energy of the electrons in a shell. The high atomic number in bones together with their high density explains why they are shown up so clearly on medical radiographs.

- Compton scattering
  Compton scattering is an inelastic scattering of the X-ray photon by an outer shell electron. After the collision the photon deviates from the incident direction and transfers part of its energy to the scattering electron (also called secondary electron), thereby ionizing the atom. The scattered photon can go in any direction, but a direction similar to the original direction is a bit more likely, especially for high-energy X-rays. The probability of Compton scattering depends on the incident photon energy and on the density of the material, but it is approximately independent of the atomic number.

- Rayleigh scattering
  It is an elastic scattering mechanism, where photons scatter without losing energy. This is important only at low photon energies.

Image formation

An X-ray image is obtained by placing the part of the patient to be investigated in front of an X-ray detector and then illuminating it with an X-ray beam. When a uniform beam passes through the body, it interacts with the tissues, producing a variable transmitted X-ray beam that is dependent on the attenuation along the beam paths. The transmitted fraction are then captured by the detector. This produces a projection image of the object, that screens anatomical information obtained measuring the radiation absorption across the object. In this image materials with different absorption coefficient can be distinguished, and the image contrast increases as the attenuations of materials differ. Object contrast is proportional to the difference between the numbers of photons transmitted through the patient's body at two neighboring areas. It depends on differences in thickness, density, and atomic composition along X-rays passing the body at different positions. Hence, there is a low contrast between muscle and soft tissue, meanwhile high contrast between air and tissues. The image shows the patient anatomy with varying grey-levels: dark values refer to tissue where X-rays could easily pass through the patient (e.g. lungs) and bright values represent tissue with high X-ray absorption (e.g. bones). Hence, tissues of high density and/or high atomic number cause more X-ray beam attenuation and are shown as brighter on a radiograph. Less dense tissues cause less attenuation of the X-ray beam, and appear darker on radiographs than tissues of higher density. That also means that structures behind dense materials are less illuminated by the beam and thus do not visible on image. Another parameters that can affect the image quality is the scattered radiation. Diagnostic relevant information is contained in primary radiation (i.e. that has not interacted passing through the matter), nevertheless, scattered radiation fraction (i.e. photons that have interacted with matter and have a different direction with respect to primary radiation) may be higher and is superimposed to the image, decreasing image quality.

2.2.2 2D imaging: Digital Mammography

Digital mammography is an X-ray imaging system specific for breast acquisitions. The apparatus consists of (figure 2.4):

- a supporting plane at the height of the patient's chest comprising the detector, and on which is positioned the breast and then compressed by means of a parallel plate of radiolucent material;

- an X-ray tube whose focal spot is typically distant about 60 cm from the detector, and radiates the breast from the top (Bushberg et al. 2002).
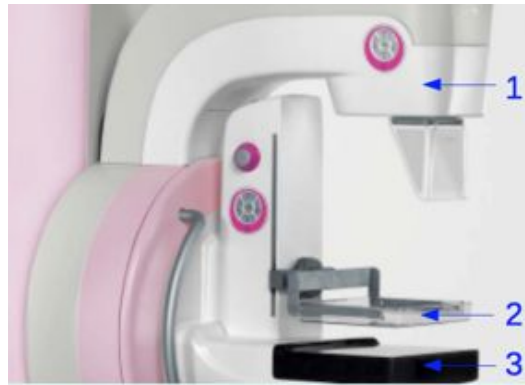
Figure 2.4: the main components of a standard MG apparatus are: an X-ray tube (1); a compression paddle (2); and a detector (3).

During the exposure, the patient stands up in front of the apparatus. This tube has an anode of molybdenum, or of rhodium in the case of particularly dense breast. Typically low energy X-rays are used, by setting the tube voltage below 30 kVp; indeed, at low energies the difference between the attenuation coefficients of the tissues increases, and consequently even the contrast increases. The parallel-plate compression is mainly used to flatten the breast tissues, reducing their superimposition (breast is flatten until an average thickness of 5 cm). Moreover this evens out the thickness of breast to increase image quality by reducing the thickness of tissue that x-rays must penetrate, decreasing the amount of scattered radiation that degrades image quality, and reducing the required radiation dose. The compression also is aimed to immobilize the breast during exposure that can take few seconds, preventing motion blur. Due to the compression of the breast, this examination is painful for some patients. With this system a 2D projection of breast (also called mammogram) is collected, where the entire breast (from nipple to chestwall) is included in the field-of-view of the image. The image that is produced is a two-dimensional representation of three-dimensional space. Each pixel is therefore an average of the information obtained through the full thickness of the breast. The overlapping of tissue structure affects the visibility of abnormalities. The examinations is normally performed by a technician, meanwhile the image are read and interpreted by a physician. A complete breast screening comprises two different views of each breast taken from different angles: one is the Cranio-Caudal (CC), i.e. head-to-foot view, and the other one is the Medio-Lateral Oblique view (MLO) (figure 2.5). This in order to partially overcome the shortcoming of tissues superimposition that can hide the behind structures. Indeed MG is characterized by lower specificity in case of breast dense tissue that can hide cancer. During the acquisition of the CC view, the patient stands in front of the apparatus and the breast is imaged from nipple to chest wall. It must show the medial part as well the external lateral portion of the breast as much as possible. The acquisition of

the MLO view is carried out rotating X-ray source together with detector at an angle of 45° with respect to the CC plane, and the patient stands on the side in order to radiate the breast up to include on projection both pectoral muscle and axilla. The limitation of this technique is that the two-dimensional
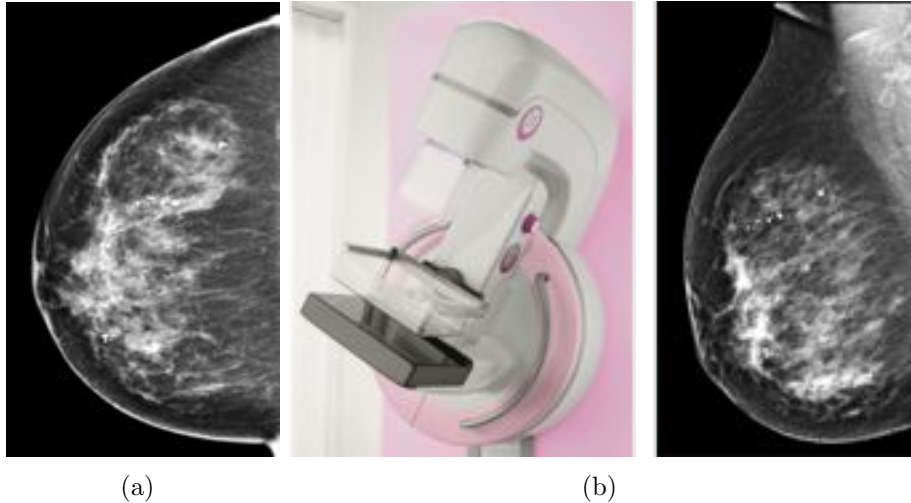

(a)   (b)

Figure 2.5: (a) example of the CC view of MG. (b) a typical MLO view, where the breast is compressed at a different angle of the MG apparatus as shown in the picture.

nature of MG results in tissue superposition, which can create two problems: dense glandular tissue located above and/or below a lesion of interest can reduce the visibility of the lesion (reducing sensitivity), or two or more normal features that are only vertically separated can appear to be the projection of a lesion of interest (reducing specificity) (Sechopoulos 2013). Moreover, the ability of MG to detect breast cancer decreases significantly in women with dense breasts.

Radiographic appearance of breast anatomy
In a typical mammogram, the whole breast is imaged, and glandular tissue appears bright as it is radiodense, while fat tissue appears darker because radiolucent (figure 2.6). Lesions usually appear brighter than the surrounding tissue, because they are denser than fat. Problems emerge with dense tissue, where lesions and breast tissue have low contrast and dense tissue appears as a wide bright area that can mask underlying structures.

Role in breast cancer screening
MG is the standard imaging modality for breast cancer screening. The goal of MG is the early detection of breast cancer on asymptomatic women, typically through detection of characteristic masses and/or microcalcifications. MG is also the reference modality for the assessment of breast density. Breast den-
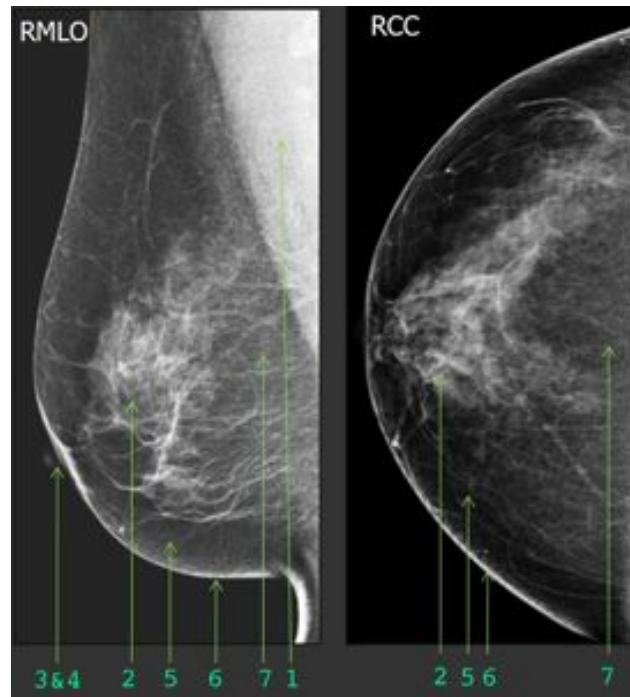
Figure 2.6: Normal breast anatomy: 1. pectoralis; 2. glandular tissue; 3. nipple; 4. areola; 5. subcutaneous fat; 6. skin; 7. retromammary fat (Hong et al. 2012).

sity is based on how fibrous and glandular tissues are distributed in the breast, compared to the fraction of breast that is made up fat tissue. This distribution can be qualitatively assessed by radiologists evaluating the appearance of breast on mammogram. Hence, density is expressed as a percentage (visually estimated) of the mammogram occupied by attenuating tissue. The American College of Radiology (ACR) has defined a breast composition classification (Breast Imaging-Reporting and Data System (BI-RADS) density assessment (D'Orsi et al. 2013)). As dense breasts are linked to a higher risk of breast cancer and to a decrease of sensitivity of MG, for each category an indication on the possibility that a lesion could be obscured by normal tissue have also been reported. Figure 2.7 shows examples of how the breast density appears on a mammogram, corresponding to these BI-RADS categories.

2.2.3 3D imaging: Digital Breast Tomosynthesis

DBT is an evolution of MG technique, allowing a volumetric reconstruction of the whole breast from a finite number of two-dimensional projections collected at different angles (within a narrow angular range) around the breast. Each projection is comparable to a mammogram, but acquired at low dose. It is introduced to overcome the loss of information in the third dimension and the shortcoming of overlapping structures. The apparatus is similar to MG, but in
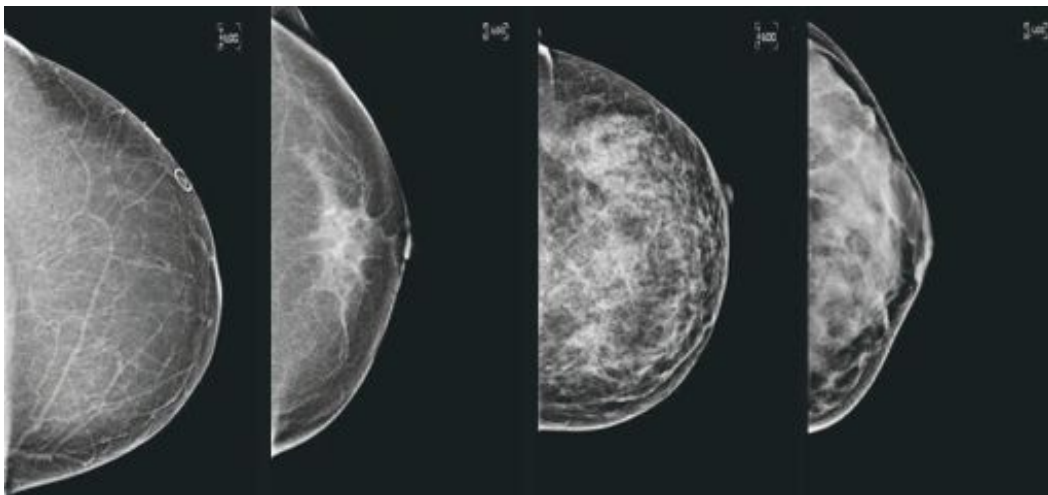
Figure 2.7: Breast composition illustrations. from left to right: almost entirely fatty breast ($< 25\%$ glandular tissue); breast with scattered glandular densities (25-50% glandular tissue); heterogeneously dense breast, which may obscure small masses (50-75% glandular tissue); and extremely dense breast ($>75\%$ glandular tissue), which lowers the sensitivity of mammography (*GE Healthcare*).

this case X-ray source can move covering an arc with respect to breast and detector (figure 2.8). The series of projections is then processed by a reconstruction algorithm which uses the different location in the projections of the same tissues to compute their vertical position, thereby estimating the 3D distribution of the tissues. Various reconstruction algorithms can be used, including shift-and-add, matrix inversion, filtered back projection, maximum likelihood reconstruction, and simultaneous algebraic reconstruction technique. Certain reconstruction methods may be better for masses and other methods better for calcifications (Helvie 2010). For instance, the most simple method, the shift-and-add algorithm, consists of a shifting and adding of the projections to bring structures of a given plane in focus. By varying the amount of shifting, planes at different depths inside breast thickness can be rendered in focus (Dobbins et al. 2003). Hence, each reconstructed slice contains not only the objects of interest in the focal plane, but also objects from every other plane blurred out and superimposed on the plane of interest (out-of-plane artifact). After the reconstruction algorithm a series of high-resolution parallel sliced images are obtained. The number of slices varies with thickness of the breast and the type of reconstruction method.

Resolution
Due to the limited angle of the projection acquisitions, DBT images is characterized by anisotropic spatial resolution, with very high spatial resolution in the planes parallel to the detector (in-plane resolution), the same resolution

Figure 2.8: sketch of the DBT rotating gantry with respect to the breast position.

of as a mammogram, and a considerably lower resolution in the perpendicular direction (in-depth resolution).

System description
Several commercial DBT systems are available on the market. Currently, most DBT systems consist of the same basic components as digital MG systems: a detector, a compression plate, and an X-ray tube mounted on an arm. The additional features that distinguish a DBT system from a MG are the ability of the X-ray tube to rotate around a point close to or on the detector and having a detector with fast readout. The systems differ for additional modifications implemented by manufacturers to optimize their system and for the choice of the detector's type, X-ray tube target and filtration, angular range, number of projections, as well as reconstruction method. The DBT images used in this thesis are collected with Hologic Selenia Dimensions system and High Definition Breast Tomosynthesis produced by Siemens Healthcare.

DBT volumes
The resulting 3D images consist of a series of high resolution projection images at different depths inside the breast (see figure 2.9). The number of slices depends on the thickness of the compressed breast. Typically a frame every 1 mm is reconstructed. Thus, an average size of a volume is 55-60 frames. The plane resolution depends on the equipment. Typically a resolution of $0.11 \times 0.11 mm^2$ is achieved (A. Smith 2011).

Advantages and limitations
The main advantages offered by DBT over MG (Guo et al. 2006) are:

slice # 1   slice # 20   slice # 30   slice # 50

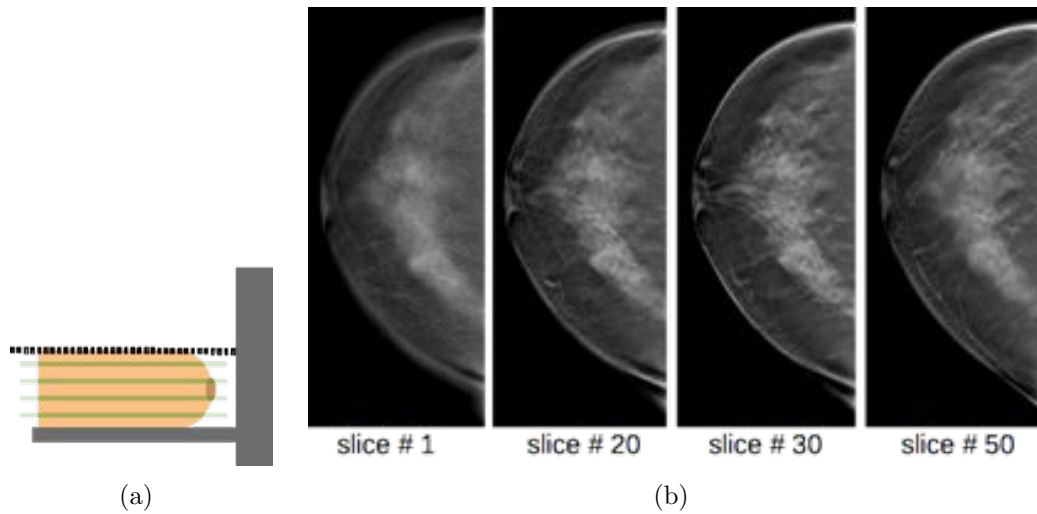(a)   (b)

Figure 2.9: example of a reconstructed DBT volume, where the projection images reported in (b) correspond to the green lines drawn in (a) at different depths inside the breast.

- the reconstructed 3D images preserve depth information, allowing a 3D lesion localization and overcoming the superimposition issue of normal structures (in figure 2.10 is reported an example in which localization and size information of a mass are preserved in DBT images on the contrary of the tissues overlapping resulting in the MG projection, and even superimposition with surrounding breast tissue is resolved). These are important improvements for a better morphological analysis of masses and architectural distortions, that could improve detection and characterization of breast lesions especially in women with dense breasts. As a consequence a possible reduction of false negatives and false positives recalls could result;

- improved contrast is provided. While for MG the dynamic range is used to represent all the structures superimposed in a single projection of the whole volume, on the contrary, in a DBT volume the structures are imaged in different planes and thus the dynamic range is assigned to view image variations in a single slice. It results in a higher overall image contrast between breast structures.

Several early experimental clinical DBT studies have shown good patient acceptance, physician preference for DBT images, improvement in sensitivity, improvement in characterization, and often longer physician reading times (Lo et al. 2006, Helvie, Roubidoux, et al. 2007, Gur et al. 2009, Helvie, Chan, et al. 2009, Teertstra et al. 2010, Dang et al. 2014, Rafferty et al. 2013).

A shortcoming of DBT images are the presence of out-of-plane objects, that

slice # 27   slice # 36   slice # 45
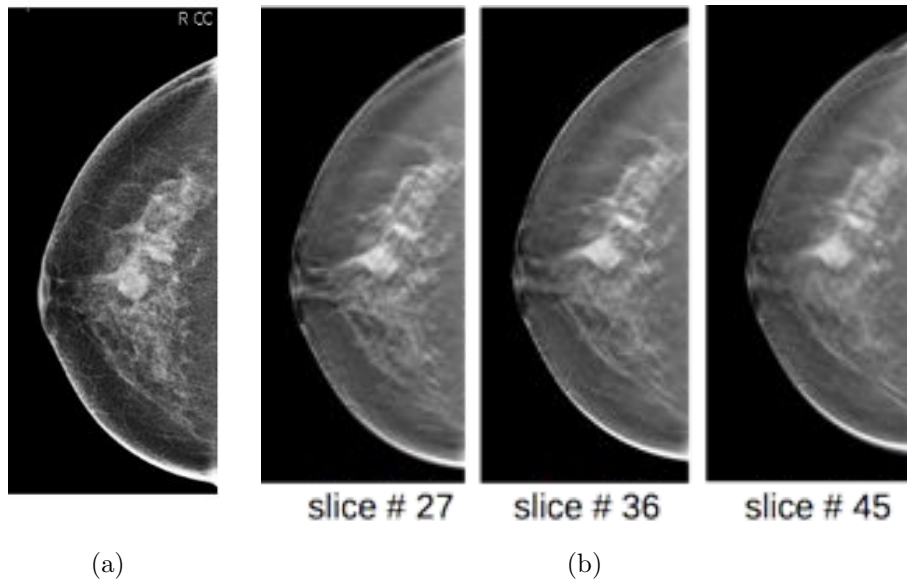
(a)   (b)

Figure 2.10: comparison between MG and DBT images of the same patient: the MG projection (a) and three DBT reconstructed projections at different depths (b). It can be noted that in each DBT slice only the in focus structures corresponding to the reconstructed plane are reported. This allows to distinguish the lesions to the normal breast tissue, meanwhile in the mammogram the same structures are overlapped in a single projection.

causes blur effect. This artifact is mainly due to the limited angular aperture and to the type of the reconstruction algorithm used.

2.3   3D Breast Ultrasound Imaging

In this section the second imaging modality used in this thesis is described: the Ultrasound imaging technique (US). After a brief introduction of the basic principle of US, a description of the important role played by this modality in breast US examination and of the handheld technique commonly used are presented. This preliminary section aims to introduce modern Automated Breast Ultrasound (ABUS) scanners for the acquisition of 3D US images of breasts. This represents the state-of-the-art available on the market.

2.3.1   Ultrasound imaging

US, also called sonography, uses high-frequency sound waves, that can be used to investigate the internal structures of the human body. The waves are transmitted by the contact between the transducer and the area of interest and can travel into the human body. Therefore, an image of the anatomy of the body can be formed. Medical image formation relies on the property of tissues to be

transparent to US and the property of ultrasonic waves to undergo reflection and refraction phenomena at the interface between tissues having different acoustic properties. Hence, US has the ability to detect and localize boundaries between organs and any small structures or inhomogeneities (such as tissue abnormalities) able to produce reflected US waves, and to characterize tissues based on the properties of these echoes.

Basic principles (Bushberg et al. 2002, Middleton et al. 2004, Kremkau 2010) Ultrasonic waves are transmitted through the matter by high-frequency particle vibrations. Thus US can travel through gases, liquids, and solids, but not in vacuum. A US measure consists in the generation of US wave, its propagation through a medium and in the interpretation of the response subsequent to the interaction between US and matter. The probe is the component of the US equipment that is placed in direct contact with the patient's body. In medical US, the probe works both as a generator and a detector of US waves (i.e. it is operating in pulse-echo mode), hence it sends US pulse into tissue and then receives echoes back. A modern US probe comprises an array of transducers, consisting of piezoelectric crystals, that convert electrical pulses into mechanical oscillations which generate the sound waves. Then the returned echoes are converted back into electrical pulses and are further processed in order to form the US image presented on the screen. The diameter of the disc of piezoelectric material determines the geometry of the ultrasonic beam, and its thickness the frequency of resonance. The beam can then be focused (by a lens or other technologies) to give it a specific size and shape at various depths within the body. According to the number and layout of the transducers, different shapes of US probe exist (see figure 2.11).
A broad classification divides the probes into:

- linear type: where a linear sequence of transducers produces a linear beam (i.e. with a rectangular field-of-view) that results in a rectangular image as wide as the width of the beam. The transducers are activated sequentially, and each crystal element generates a scan line of image information. The larger the number of scan lines, the better the lateral resolution (i.e. spatial resolution in the direction of the length of the array). It is often used for US of superficial structures, so the operating frequency ranges from 3 to 12 MHz;

- convex or curved type: it consists of an array of transducers mounted on an arc-shape probe and the resulting beam has a wider field-of-view, with a sector shape, that produces a trapezoidal image. The shape and size of the beam vary with distance from the transducer, that causes the lack of lateral resolution at greater depths. The operating frequency ranges from 1 to 5 MHz;

- phased or sector type: A phased array system may have the same geometrical configuration as either a linear or curvilinear array, but the

procedure of activating the crystal elements is different. Instead of being activated sequentially, in a phased array all transducers are excited together. For each pulse, the system produces only one scan line over the whole area of the array. The US beam has a sector shape, almost triangular, and produces a fan like image that is narrow near the transducer and increase in width with deeper penetration. It is use for small acoustic window. It is useful when scanning between the ribs as it fits in the intercostal space. The operating frequency ranges from 1 to 5 MHz.

The choice of the probe depends on the application. For breast imaging a linear probe is typically used. Note that the term transducer is commonly used also to refer to the US probe. Hence, in the following these two term will be used interchangeably.
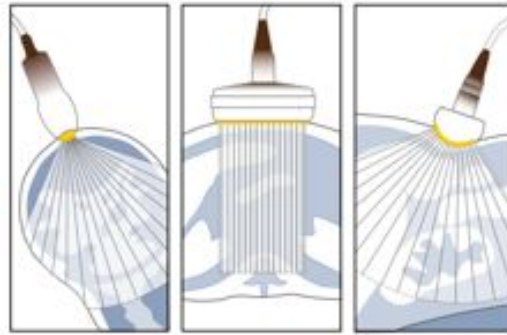

Figure 2.11: from left the following probe types are drawn with the corresponding US beam: phased-array, linear and convex probes.

Image formation depends on the following main properties of the US waves:

- Frequency
  One of the most significant characteristics of US is its frequency, which is the rate at which the sound source vibrate. Thus the frequency of sound is determined by the transducer's characteristics. US refers to a sound wave with a frequency greater than the upper limit of audible frequencies for human, which is generally over 20 kHz. The frequencies used in medical imaging range from 2 MHz to approximately 15 MHz. Therefore, frequencies typically range from 2.5 MHz for deep abdomen, obstetric and gynaecological imaging to 15.0 MHz for superficial structures and musculoskeletal imaging. Breast US requires a high frequency transducer (8-15 MHz); a lower frequency transducer may be required for large breasts and the axilla. Normally, in medical US pulsed US beams are used, that do not have a single frequency; the transducers are characterized by more than one operating frequency, i.e. a bandwidth of frequency, and commonly its central value is the reference frequency. The frequency of the transducer is chosen based on several factors, including size of objects to be detected, depth of penetration, and tissue

composition. It must be carefully selected to provide a proper balance between image detail and depth of penetration. In general, high frequency pulses produce higher quality images (i.e. short wavelengths are required to produce good anatomical details), but cannot penetrate very far into the body (the explanation will be discussed later). Therefore, the higher the frequency, the smaller the detail that can be detected, but the shorter the depth of penetration of US. This explains why high frequencies are used for superficial body structures and low frequencies are used for those that are deeper. After introducing even the other properties of US, it will be clearer the criteria for choosing the proper frequency.

- Propagation speed
  The speed of sound varies by the medium it travels through and the sound waves travel faster through solids, followed by liquids and gases. Therefore, in the body, US travels fastest through bone tissue and slowest through air. Even the type of sound waves depends on the medium. Sound travels through the gases and liquids as longitudinal waves (i.e. particles vibrate along the wave propagation direction). Through solids it can be transmitted as both longitudinal waves and shear (or transverse) waves (i.e. particles oscillations are transverse to the wave propagation direction), each is traveling with different speed. Hence, in general, the US waves traveling into the soft tissues are longitudinal waves, instead bones can transmit both longitudinal and transversal oscillations. For a longitudinal wave, the propagation speed, denoted by the letter c (m/s), is related to the density ρ (kg/m^3) of the medium and the coefficient of stiffness K (or the modulus of bulk elasticity for gases) (Pa, corresponding to kg/ms^2), by the following relationship:

$$c = (K/\rho)^{1/2} \tag{2.2}$$

  Thus the speed of sound increases with the stiffness (i.e. the resistance of an elastic body to deformation by an applied force) of the material, and decreases with the density. Note that the speed of sound is substantially independent to frequency. The speed of sound in the most of soft tissues has an average value around 1540 m/s, in the air is about 340 m/s, in fat 1450 m/s, in water is 1480 m/s, and in bone as high as about 4080 m/s.

- Acoustic impedance
  Acoustic impedance (Z) is a physical property of a material that describes how much resistance a US beam encounters as it passes through it. It is of considerable importance in characterizing the propagation of US waves at the interface between two tissues. It depends on speed of sound (c, in m/s) and density of the medium (ρ, in kg/m^3) by this

relationship:

$$Z = \rho c \tag{2.3}$$

It means that a denser medium offers greater impedance. So, as the density of a tissue increases, the impedance increases. Similarly, but less intuitively, if the speed of sound increases, then impedance increases. The SI unit for acoustic impedance is the *Rayl*, corresponding to kg/m^2s. As reference, air has low acoustical impedance ($0.0004 \times 10^6 kg/m^2s$), water has higher impedance than air ($1.5 \times 10^6 kg/m^2s$), and bone have higher impedance than water ($7.8 \times 10^6 kg/m^2s$). Meanwhile fat has an acoustic impedance quite less than that of soft tissues and water ($1.3 \times 10^6 kg/m^2s$).

Interaction of US with matter

As US pulse passes through matter, such as human tissue, it interacts in several different ways. Some of these interactions are necessary to form an US image, whereas others absorb much of the US energy or produce artifacts and are generally undesirable in diagnostic examinations. The ability to interpret the results of US examination depends on a thorough understanding of these US interactions. When a beam of US pulses is passed into a body, several things happen. Most of the US energy is absorbed and the beam is attenuated. This is undesirable and does not contribute to the formation of an image (like in X-ray imaging). Some of the pulses will be reflected by internal body structures and send echoes back to the surface where they are collected by the transducer and used to form the image. Therefore, the general US image is a display of structures or reflecting surfaces in the body that produce echoes.

- Attenuation and penetration

As the US pulse moves through matter, it is gradually attenuated, i.e. it continuously loses energy. Two main mechanisms contribute to US attenuation: absorption and scattering. The US attenuation generally depends on two factors: the material through which US is passing, and the frequency of the US. US energy is more likely to be lost in gases, while it travels through liquids or solids more efficiently. And moreover, those with a higher frequency show a higher attenuation factor. Attenuation is linearly dependent on frequency. Attenuation rate is usually measured in decibels per centimeter of tissue and it is expressed in terms of the attenuation coefficient of the specific tissue type (measured in dB/cm MHz). Bone has the higher attenuation coefficient (20 dB/cm MHz), followed by air (12 dB/cm MHz). Soft tissues has an average attenuation coefficient of 0.9 dB/cm MHz, meanwhile water produces the least attenuation 2×10^{-3} dB/cm MHz). This means that water is a very good conductor of US. This explains why a gel is placed between the transducer and the skin to eliminate air for the best sound conduction. Water within the body, such as in cysts and the bladder, forms windows through which underlying structures can be easily imaged. Compared to the soft tissues of

the body, bone has a relatively high attenuation rate. Bone is a good absorber of US and this implies that shields some parts of the body against easy access by US.

The attenuation influences the distance traveled by the US through the tissue. Penetration is the term used to describe the distance from the sound source at which 50% of the original energy remains. There is an inverse relationship between absorption and penetration. As tissues absorb energy from a sound wave, a reduced amount of energy remains to be carried forward by the wave, thus the US penetration is decreased. Therefore, a high frequency wave is associated with high attenuation thus limiting tissue penetration, whereas a low frequency wave is associated with low tissue attenuation and deep tissue penetration. This is an important factor for the selection of the appropriate frequency of US, because the frequency controls the depth of penetration.

- Reflection and refraction

Generally, when an ultrasonic wave passes from one medium to another, some energy is reflected (reflected echoes) and the remaining energy is transmitted. The reflection of US pulse occurs at the interface between two dissimilar materials. In order to form a reflection interface, the two materials must differ in terms of acoustic impedance, and the reflection is proportional to the difference in impedance between the tissues. The amount of reflection that occurs in a perpendicular direction is expressed by:

$$Reflection fraction = [(Z_2 - Z_1)/(Z_2 + Z_1)]^2 \qquad (2.4)$$

where Z_1 and Z_2 represent the impedance in tissue 1 and tissue 2, respectively. So there is no reflection for $Z_1 = Z_2$ (i.e. echoes are not produced if there is no difference in a tissue or between tissues), very low reflection fraction when $Z_1 \sim Z_2$, and total reflection for $Z_1 >> Z_2$ (or $Z_1 << Z_2$) (i.e. sound is completely reflected, resulting in total acoustic shadowing, such as occurs behind bones). The larger the difference in acoustic impedance between two tissues, the more sound will be reflected back to the transducer and the less sound carries on traveling through the tissue.

The human body consists of a variety of tissues and organs with acoustic impedance different from each other. Hence, US repeatedly encounters boundaries and striking these tissues are reflected differently. The acoustic properties of skin, fat, blood vessels, and muscle are similar. When US encounters boundaries between acoustically similar tissue, such as adipose and muscle, the amount of reflection is insignificant. On the contrary, at interfaces between soft tissue and materials such as bone, calculi, and gas, strong reflections are produced. Since within the body the impedances (except for air and bone) are very similar, now it is clearer why US can easily penetrate into the body, but the reflected echoes are weak. Because the impedance characteristics of metal and air are so different, the amount of reflection at a metal-air interface

is about 99%, which means that the amount of US transmitted from a metal transducer to air is negligible (and the same for the interface air-skin, where the impedance of air is very different to that of soft tissue). This is another reason for using a coupling medium between the transducer and the skin during US scan. Meanwhile, at a tissue-bone interface, about 15% of incident energy is reflected. It explains why US is limited in evaluating structures encased in bone like brain anatomy. When US is reflected from an interface between two different soft tissue, as well as at a fat-muscle interface, or water-fat, less than 1% of sound is reflected.

When US pulse is not perpendicular to a surface but passes through an interface at a relatively small angle (between the beam direction and interface surface), the penetrating pulse (i.e. transmitted across the boundary) direction will be shifted by the refraction process. This can produce certain artifacts. Meanwhile the reflected wave is directed away from the boundary on a new path (i.e. reflected away from the transducer) that has same angle but is a mirror image of the incident wave. Refraction is proportional to the difference in acoustic impedance of the boundary materials and to the incident angle of the wave. Therefore, refraction at boundaries formed by layers of skin, fat, blood, or muscle is very small. At tissue-air boundaries, however, because the impedance characteristics of tissue and air are so different, the transmitted wave changes direction by 90% (i.e. total internal reflection). This means that the wave traveling along the boundary is totally reflected instead of crossing it.
Refraction follows the Snell's law:

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{c_1}{c_2} \tag{2.5}$$

where θ_1 and θ_2 are measured with respect to the normal at the refractive boundary, and θ_1 corresponds to the incident angle and θ_2 to refraction angle. c_1 e c_2 represent the speed of sound, respectively, in the tissue 1 and tissue 2.

Image formation
The basic US image (or sonogram), called B-mode (brightness) image, is a display of structures or reflecting surfaces (interface//or boundary) in the body that produce echoes. These echoes contain both spatial and contrast information of the objects need to form an image. In fact, the location in the horizontal direction is determined by the position of the beam. In the depth direction, it is determined by the time required for the pulse to travel from source to the reflecting site and for the echo pulse to return back (time of flight), that means the transducer can measure the delay between the emission of the US pulse and the detection of the reflected echo. From this measure and the knowledge of the speed of sound, the location of the reflecting interface (i.e. its distance from the source) can be derived. The brightness

(i.e. the image gray-level value) depends on the amplitude of the echo (with respect to the amplitude of the incident wave), which is related to the difference in acoustic impedance at the interface. Hence, echoes show up as bright (white) spots in the image. Most anatomical areas have inhomogeneous composition, that results in the well-known gray and white background. In the contrary, liquid-based masses, such as a cyst, is dark in the image, since there are no reflecting surfaces within a fluid. However, these structures present bright boundaries owing to usually high difference in acoustic impedance with respect to the surrounding tissue (as explained before). Using appropriate grayscale (i.e. dynamic range), tissues that give very weak signals can be better distinguished (i.e. the image contrast can be increased). Generally, in US the term hypoechoic or hyporeflective is used for low intensity regions, and hyperechoic or hyperreflective for high intensity regions.

The spatial resolution of US image is determined by the cross sectional dimensions of the ultrasonic beam and by the duration of the ultrasonic pulse (i.e. spatial pulse length). Generally it is divided in: axial (or longitudinal) resolution, that is the ability to distinguish objects spatially close along the direction of propagation of the beam, and depends on the duration of the ultrasonic pulse (i.e. spatial pulse length); lateral resolution, that defines the resolution in the scan plane direction, it is dependent on both the transducer element height and width, and thus it is determined by the cross sectional dimensions of the ultrasonic beam. As the beam size varies with depth, potentially resulting in a depth-dependent image quality.

US artifacts

When forming a B-mode image, the imaging system makes a number of assumptions about US propagation in tissue (Feldman et al. 2009). These include:

1. the speed of sound in human tissue is constant,

2. the US beam and its echo travel in a straight path,

3. the attenuation in tissue is constant, and

4. the pulse travels only to target that are on the beam axis and back to the transducer (i.e. after a single reflection). That means that the depth of an object is directly related to the amount of time for a US pulse to return to the transducer as an echo.

Significant variations from these conditions in the target tissues are likely to give rise to visible image artifacts. Most artifacts may be grouped into speed-of-sound artifacts, attenuation artifacts or reflection artifacts according (i.e. errors inherent to the US beam characteristics and the presence of multiple echo paths) to which of the above conditions is violated (Meire et al. 2001).

The ability to recognize US artifacts is important as artifacts have the potential to interfere with image interpretation. Some artifacts may cause structures to appear in an image that are not present anatomically or a structure that is present anatomically may be missing from the image. US artifacts may also show structures as present but incorrect in location, size, or brightness (Kremkau 1998. The most frequent US artifacts include the so-called reverberation, mirror image, speed displacement, attenuation errors, shadowing, and increased through-transmission artifacts. Artifactual echoes, including speckle and other spurious noise, as well as posterior acoustic patterns, including posterior enhancement (characteristic of simple cysts) and posterior acoustic shadowing (characteristic of some solid masses) are artifacts encountered routinely in clinical practice.

Role in breast cancer diagnosis
The ability of US in the discrimination of low contrast tissues is widely accepted, especially for the examination of substantially superficial organs and entirely composed of soft tissues, such as breast. In the introduction its usefulness in case of dense breast tissue has been already highlighted. In breast imaging US is widely used for the following advantages: it is almost free-cost technique and widely available; it is safe and less invasive than other options (like X-ray imaging or MRI). In fact, it does not use ionizing radiations for investigation, it is painless exam and it is free of any other adverse side effects, so it is well tolerated by all patients (Mendelson et al. 1995, H. Wang et al. 2012). It is used even for pregnant women and for women at very high risk for breast cancer who cannot tolerate breast MRI, as well as those women with dense breast tissue (i.e. who have higher radiation-induced cancer risk (Vachon et al. 2007)). US also allows real-time imaging (essential in case of biopsy with US guidance). These are all important factors that could make US a valuable tool for breast cancer diagnosis.

In common practice, US is used as supplemental tool or as a secondary diagnostic technique (i.e. recall exam in order to characterize mammographic findings). Breast US is generally not used as a screening tool because it does not always detect some early signs of cancer such as microcalcifications, and moreover it may miss small lumps or solid tumors that are commonly detected with MG. In current indications it is recommended as primary screening technique only in case of young women (usually under 30 years of age) or pregnant patients who are symptomatic. However, the use of US examination as primary screening test is currently under investigation (Berg, Bandos, et al. 2016). Latest studies have demonstrated that cancer detection rate with US is comparable with MG, with a greater proportion of invasive and node-negative cancers among US detections, even if false positives are more common with US examination.

Sonographic appearance of breast anatomy

A typical breast US image shows up the following structures (figure 2.12): the skin, approximately 2 mm thick with an hyperechoic appearance; the premammary layer of subcutaneous fat, anterior to the brighter and heterogeneous premammary fascia, and Cooper's ligaments shown as hyperechoic bands coursing through the subcutaneous fat; the mammary zone composed by glandular tissue, where the majority of breast cancers detectable by US are located; the retromammary fascia and fat, posterior to the mammary layer; and the most posterior structures that are the pectoralis muscle and then the ribs. The sonographic appearance of malignant tumor is most often that of a solid hypoechoic mass (compared with breast fat), i.e. cysts, tumors, and distortions appear dark on the scan.
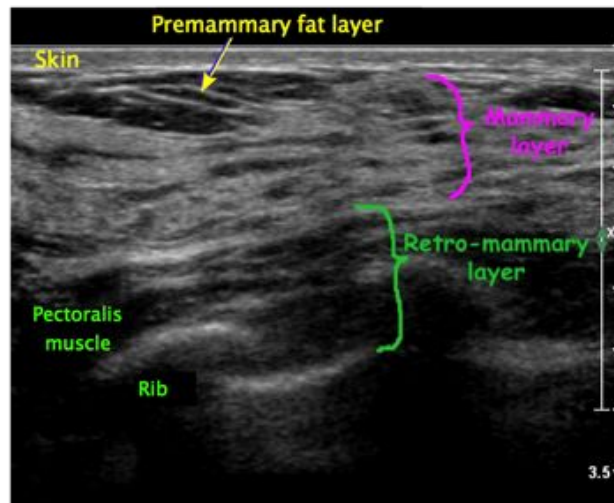

Figure 2.12: US capture of the normal breast anatomy, showing: the pre-mammary zone, comprising skin and overlying breast fat; the mammary zone composed by glandular tissue; the retro-mammary zone, including predominantly fat and the muscles of the chest wall.

2.3.2 2D imaging: Handheld Ultrasound

In a standard US examination the breast is scanned using a conventional 2-dimensional transducer. The US is not performed by a technician as a mammographic acquisition, but it is a real-time imaging examination performed by physicians. During the examination, the patient lies in a supine position, and the physician can easily investigate the whole breast, comprising the axillary region, moving freely the transducer around the breast (handheld US scanning), and bi-dimensional images of specific areas of concern are collected (US guidelines included in the ACR BI-RADS Atlas (D'Orsi et al. 2013) recommend 5 images for a diagnostic US negative exam). Typically a broadband linear array transducer is used, with a bandwidth of frequencies

of 5-12 MHz, for high resolution superficial applications. However, in women with small breast (with breast thickness < 3 cm) or when performing targeted US to evaluate a superficial lesion, a linear 5-17MHz transducer may be used. Hands-on real-time scanning is a more flexible examination allowing the physician to control pressure and orientation of the probe (to overcome anisotropy and edge refraction from vessels, Cooper's ligaments, edge of cysts) to sharpen up the edges of a lesion. As a consequence, a deeper study of lesions conspicuity and margins analysis is possible. Real-time scanning also allows the operator to assess lesion mobility, location, and relationship to adjacent structures (Hooley et al. 2013).

However, as the image quality is highly dependent on the probe pressure, patient positioning and angle of insonation, the detection accuracy is related to the acquisition skill and experience of the physician. (Berg et al. 2006) reported radiologist performance regarding the detection, description, and interpretation of breast lesions using handheld breast sonography. In their study, the detection varied according to the lesion size, with only an approximately 44% detection rate for lesions 3.1 to 5 mm, a 53% detection rate for lesions 5.1 to 7.0 mm, a less than 70% detection rate for lesions 7.1 to 9.0 mm, and reliable detection (97%) only when the mean lesion diameter was greater than 11.0 mm.

Hence, despite the multiple advantages, the major shortcomings of hand-held US (HHUS) can be summarized as follow:

- high degree of subjectivity: HHUS scanning technique requires specific technical skill and the US result is highly dependent on the experience and subjective judgments by the operator (Wojcinski et al. 2011);

- time-consuming: it needs long examination time, i.e. around 30 minutes for a typical scan;

- operator dependence: that means considerable user variability and low uniformity in the quality of examinations;

- HHUS provides only partial 2D view of the breast, that do not allow a retrospective comparison with other examinations;

- lack of reproducibility: that prevents the possibility of a second evaluation. In fact it is difficult to localize a 2D image plane and reproduce the same image plane at the same location (Lin et al. 2011, H. Wang et al. 2012, due to the lack of position information relative to the entire breast.

As such, HHUS is not clinically accepted to be used as a population screening tool for breast cancers (Buchberger et al. 1999, Kopans 1999, Wojcinski et al. 2011).

2.3.3 3D imaging: Automated Breast Ultrasound

Due to multiple advantages of this modality, recently many efforts have been employed from industries to develop newer technologies for automatic and repeatable acquisitions and for providing 3D images of whole breast volume. Automated Breast US (ABUS) scanners have been introduced to provide a potentially alternative to conventional handheld screening breast US.

Two different types of automated scanners have been currently proposed (Chou et al. 2007:

- Automated whole-breast ultrasound (AWBUS), where patient lies in a prone position and a tomographic acquisition of the whole breast (freely suspended in a warm water bath) is captured through a ring array of transducers. The existing systems include: SoftVue^{TM} system produced by Delphinus Medical Technologies, and SonixEmbrace^{TM} system developed by Ultrasonix.

- ABUS scanner where the patient is in a supine position and the breast is automatically scanned by a wide footprint linear transducer. The ABUS systems currently available on the market are: ACUSON S2000^{TM} Automated Breast Volume Scanner (ABVS) designed by Siemens Healthcare, Invenia^{TM} ABUS by GE Healthcare, and Adjunctive AWBUS system developed by SonoCiné Inc.

The datasets used in this thesis refer to the Siemens ABVS system. Therefore, the description of the apparatus will be focused on this system (note that some general considerations about the equipment and the scanning procedure are valid for this system as well as the other similar systems, independently to the manufacturer).

The possibility to obtain an automated acquisition of 3D breast images is the major interesting feature of this type of novel technology. Like conventional HHUS, ABUS uses high-frequency sound waves targeted at the breast examination to collect a set of closely spaced 2D images, from which a 3D high-resolution volume of the breast can be reconstructed.

The main goals pointed by ABUS systems are:

- decrease of the physician's examination time per case;

- produce a standardized, high quality examination that improves the conspicuity of cancers, and at the same time decreases the operator dependency of the examination result.

Clinical indications
ABUS is indicated for use as an adjunct to MG to increase breast cancer detection, explicitly for breast cancer screening in asymptomatic women for

whom screening MG findings are normal or benign (BI-RADS assessment category 1 or 2), with dense breast tissue (BI-RADS composition C or D), and have not had previous clinical breast intervention. The device is not intended to be used as a replacement for screening MG.

System description

ABVS apparatus is equipped with a scan station, an examination table, and a workstation to review the 3D images (see figure 2.13). The scan station includes a flexible arm that ends with the transducer enclosed in a case, and a touchscreen monitor that provides a real-time preview of the breast scan state. A mesh membrane is fixed at the bottom of the transducer's case (see figure 2.14) to aid in acoustic coupling, compression and stabilization of the probe over the breast. The US probe consists of a wide-frequency-bandwidth linear transducer (5–14 MHz using a 9-MHz center frequency), with a large footprint, with preset depth and gain. With this frequency bandwidth a thickness up to 6 cm is investigable. The probe face is 15 cm wide, and a maximum total field-of-view of up to $15 \times 17 cm^2$ can be covered (see figure 2.14). Technical properties of the transducer are summarized in the (table 2.1, *Siemens Healthcare*). The system also provides conventional transducers for HHUS scanning that can be combined with the automated scan.

The scan is performed with the patient lying on the back on the examina-

Figure 2.13: picture of tha Siemens ABVS equipment: (1) scan station; (2) table for patient scanning; (3) flexible arm; (4) scan head assembly for the 3D imaging.

tion table. The exam can be performed by a technician, who applies a layer of gel over the breast to be examined, and then manually places the scan head on the breast, positioning the transducer over the nipple (figure 2.15). Before scanning, the breast has to be firmly compressed for breast immobiliza-
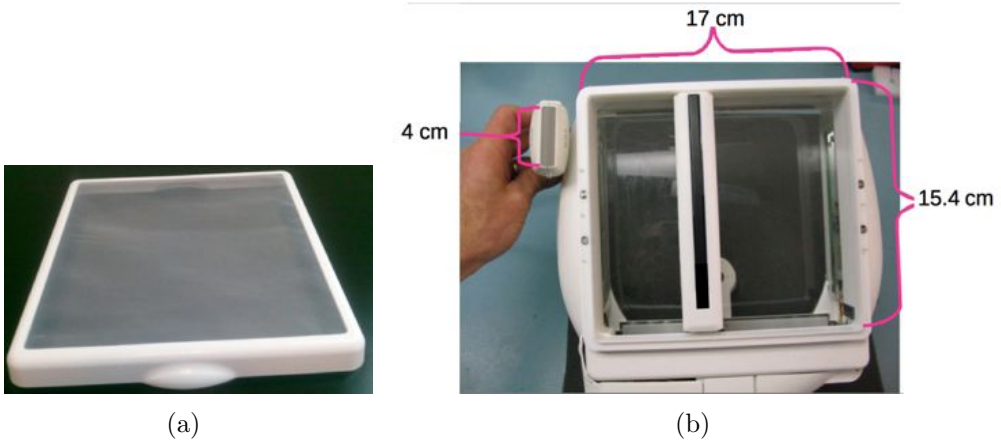
<center>(a)   (b)</center>

Figure 2.14: (a) mesh membrane to aid in acoustic coupling, compression and stabilization; (b) comparison between the size of the ABVS and the HHUS probes.

Parameter	Value
frequency bandwidth	5.0-14.0 MHz
selectable 2D frequencies	7.0, 8.0, 9.0, 10.0, and 11.0 MHz
footprint	$154mm \times 168mm$
array length	154 mm
maximum field-of-view	$154mm \times 168mm$
maximum display depth	60 mm
line spacing	0.175 mm or 0.125 mm
resolution	axial = 0.09 mm
	lateral = 0.16 mm
	sagittal = 0.44 mm

Table 2.1: technical specification of the 14L5BV transducer.

tion during the images acquisition avoiding motion artifacts and for ensuring a good acoustic contact between the transducer and the breast. The compression is automatically applied by the system; there are up to 3 levels of compressions. In case of patient's pain, the operator can manually act to change the level of compression. The system automatically scan the entire field-of-view covered by the transducer in a standard manner (from bottom to top side of breast) with optimized imaging settings (depth, frequency, focal zone placement and overall gain) on the estimated size of the breast and the scan view (i.e. transducer position with respect to the nipple and chest wall). As the transducer moves over the breast volume it acquires consecutive transverse images at regular intervals. With a single pass of the US beam, a 3D reconstructed image can be formed. The volume is displayed through a multiplanar reconstruction, where the coronal, sagittal, and transverse planes are reconstructed and can be independently examined (figure 2.16). The ability

to generate images from the coronal plane (i.e. from the nipple to the chest wall) represents a unique feature provided by ABUS, which is not achievable on conventional 2D US. A typical full examination consists in the acquisi-

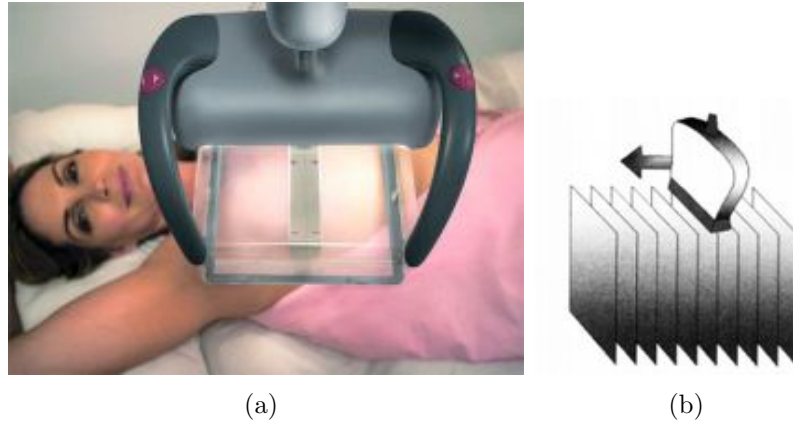(a)   (b)

Figure 2.15: Initial position of the ABUS probe to be placed over the nipple as reference. Then the scan starts from the bottom to the top of the breast. The ABUS probe automatically collects a sequence of equally and closely spaced 2D images. By knowing the spatial interval along the scanning motion, a volumetric information can be digitally reconstructed.

tion of three automated scans per breast in the anteroposterior (AP view) and both oblique positions (i.e. medial, MED, and lateral, LAT, views), for covering the entire breast volume. Three different views of each breast are acquired, by moving the probe in adjacent positions (i.e. the scan head is positioned on three different orientations), covering the area from the axilla to the sternum (figure 2.17). Occasionally, additional views were required for larger breasts and with the scans centered on either a palpable abnormality or axillary lymph nodes.

A single scan requires an acquisition time of approximately 1 minute, and the total acquisition time per patient, including the setup time, is 15 minutes. Interpretation can take between 5 and 15 min, depending on the experience of the radiologist and the complexity of the case (H. Wang et al. 2012). A study on the interpretation times (Chae et al. 2015) reports that the mean interpretation times for the 3 radiologists were 4 ± 2 minutes for the coronal view and 6 ± 2 minutes for the transverse view. Although the difference in the interpretation times was significant for the coronal and transverse views, both were dramatically shorter than the 19 minutes reported for handheld sonography (Berg et al. 2008). These results suggest that ABUS system is a time-efficient tool that could be effectively used in screening practice.

Breast US volumes
A typical dataset consists of 320 high resolution axial 2D images collected at
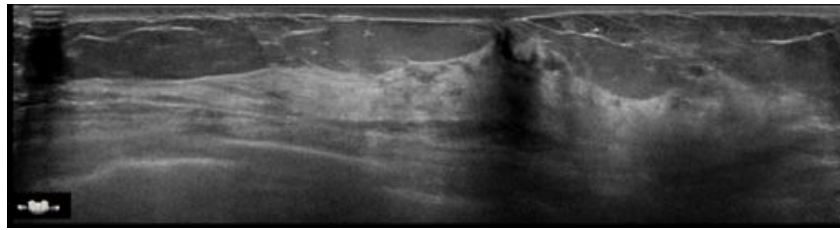
(a)
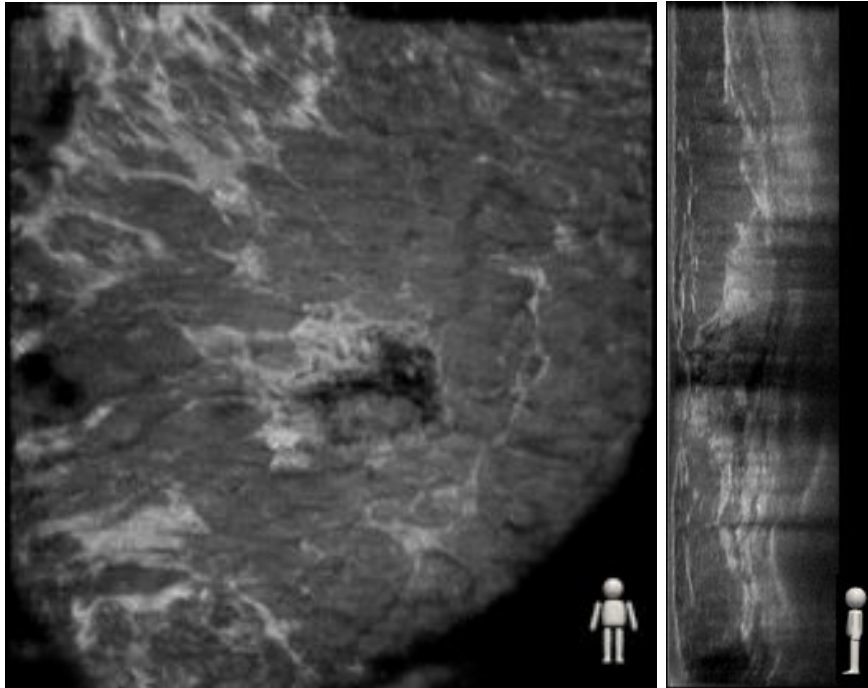
(b)   (c)

Figure 2.16: From the volumetric dataset, three orthogonal views are available: the transverse, coronal and sagittal planes. It allows breast to be viewed and examined in multiple orientations. The transverse planes, i.e. the scanned slices, can be scrolled to visualize the breast axial section from bottom to top of breast. The coronal sections begin at the level of the skin and are layered upon each other up to the chest wall.

regular slice intervals of 0.5 mm. The reconstructed frame resolution depends on the acquisition parameters automatically selected by the system depending on the breast size. A typical resulting voxel size is $0.21 \times 0.11 \times 0.52 mm^3$ a maximum volume up to $15.4 \times 16.8 \times 6 cm^3$ can be captured.

Advantages and limitations
Several recent studies have investigated the role and impact of this new technology (KM Kelly, Dean, Lee, et al. 2010, Chang et al. 2011, K Kelly et al. 2011, Wojcinski et al. 2011, Isobe et al. 2011, Shin et al. 2011, V. Giuliano
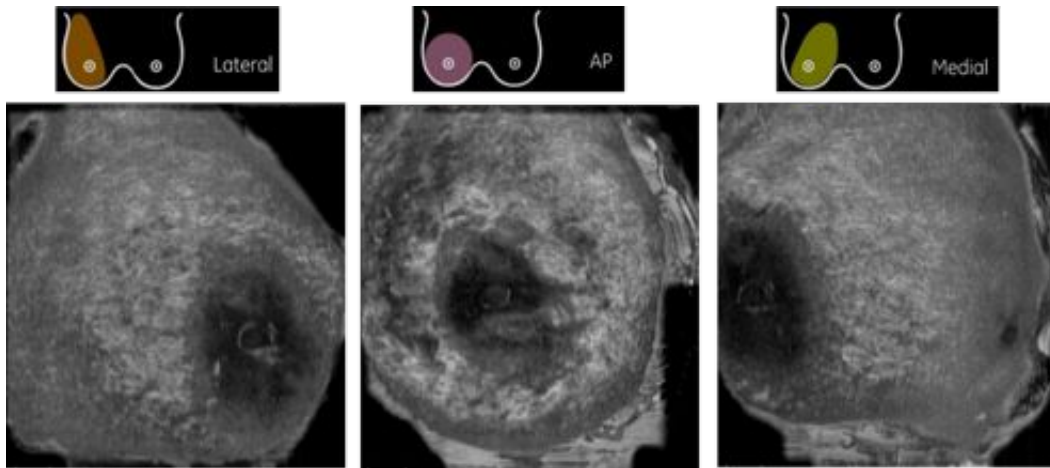
Figure 2.17: example of coronal planes extracted from three volumes of a breast collected at different probe's position: lateral (LAT, on the left), anteroposterior (AP, in the middle), and medial (MED, on the right) views. Using the areola area (dark area in the images) as reference, it can be noted the relevant deformation induced by probe pressure.

et al. 2012, Kim et al. 2014). ABUS reflects a promising modality in breast imaging, however appears to be on a par with HHUS in terms of diagnostic quality, showing a comparable diagnostic performance (KM Kelly, Dean, Lee, et al. 2010, Stoblen et al. 2011).
The main proved advantages provided by ABUS are (KM Kelly et al. 2010):

- 3D imaging allows multiplanar analysis (proving unique transversal and coronal views), thus a more accurate characterization of lesions is achievable, with an improved analysis of their shape and margins due to the volumetric information and the multiple scanning orientations. In particular, spiculated margins are often observed in the coronal view, whereas they cannot be well documented or are absent in the transverse view (see figure). A spiculated margin has been known as the highest discriminating feature for characterizing malignant breast lesions. (Stavros et al. 1995, Rotten et al. 1999) This additional value of the coronal view has been reported (Cho et al. 2006).

- Furthermore, the standardized scanning of 3D images allows reproducibility (i.e. consistency from one exam to another) and retrospectively analysis. Although three scans are required to cover the entire breast, the large transducer's field of view allows the nipple to be included in each scan, providing a reliable reference point.

- The standardization of the acquisition process allows technician to perform the scanning in place of physician, and thus image interpretation

to be separated from the acquisition process As the transducer automatically scans the breast, image quality is less operator dependent.

- Studies have demonstrated a substantial reduction in examination time, with greater sensitivity and fewer false-positive (K Kelly et al. 2011). When used in addition to MG, ABUS can improve breast cancer detection by 55% over MG alone (Inciardi 2012, Brem et al. 2015).

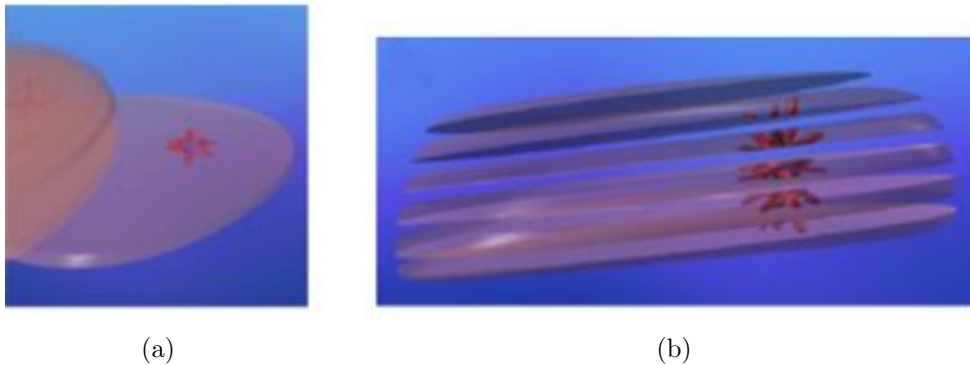

(a)   (b)

Figure 2.18: schematic representation of how the reconstructed volumetric information can improve conspicuity of lesions.

However, limitations of ABUS include that the device cannot provide a unique scan of the entire breast, especially for large breast, it requires a time-consuming review of a large number of images by the physician, and the technology lacks real-time capability. As a consequence, the sonographer cannot clear an area affected by an artifact (even if the three different scans required for each breast can overcome some artifacts due to the different probe positioning), and moreover the need for a second US examination to re-evaluate indeterminate findings implies the recall of patients. In addition, ABUS is still a high cost technology, and consequently this type of system is not yet widespread in Italy.

During the scanning, some critical issues require particular attention as can affect image quality. These include: the need to avoid dark US absorption areas through the volume due to an improper acoustic coupling between probe and transducer (figure 2.19(b)), motion artifacts due to an improper stabilization of breast during scanning (figure 2.19(a)), and the need to include in the transducer's field-of-view an overlap area among scans to ensure that the entire breast has been scanned.
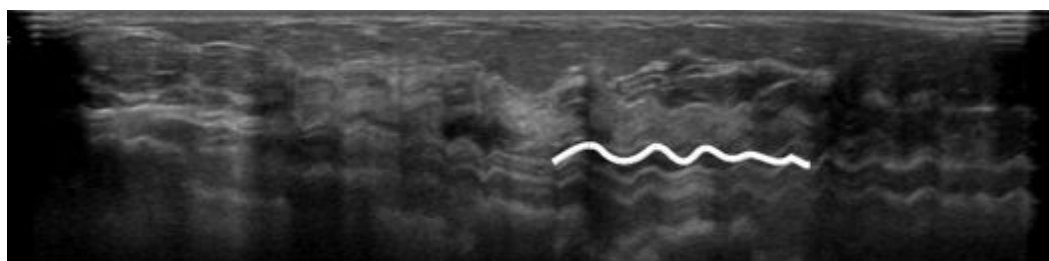
(a)
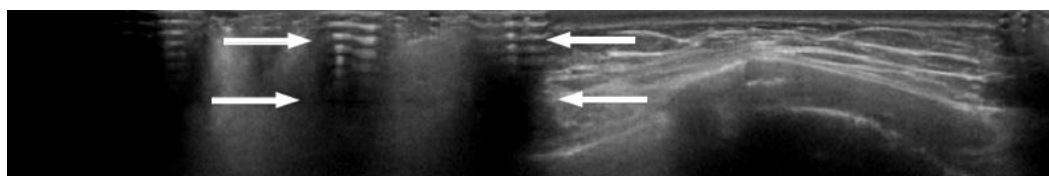
(b)

Figure 2.19: Typical ABUS image artifacts. a) sagittal view with wave pattern due to respiration or other patient motion. b): transversal view with air artifact due to lack of contact between probe and breast.

Chapter 3

Image registration theory

Contents

In this chapter, some basic mathematical concepts regarding image registration theory are presented, as well as the terminology that will be used in the following. Image registration is the process of estimating the spatial transformation between the coordinates of two images of the same or similar objects. Different mathematical implementations exist to solve the registration mathematical problem, depending on the degrees of freedom of the involved transformation model and the known a priori information. In the case of rigid registration method, the task is to find the optimal parameters of a transformation matrix that globally best matches the two images. In the case of non-rigid registration the task consists in estimating a dense vector field that should map each location of one image onto its corresponding location of the second one. The displacement field can be evaluated by the optimization of an energy functional (also known as objective function or cost function). In this chapter an overview of the state-of-the-art solutions are presented. Note that I will often refer to images and pixels, but the same concepts can be easily extended to volumes and voxels.

3.1 Introduction to image registration theory

Image registration is the process of determining a spatial mapping function that correlates corresponding points between two images. It has a vast range of applications and the reasons of registering two images can be multiple. Mainly, the registration task consists of estimating the optimal spatial transformation so that the images are matched either for representing data in a common reference system or for exposing the differences of concern between the images. Therefore, the resulting correspondences can be used to change the appearance (by rotating, translating, stretching etc.) of one image so it more closely resembles another or to automatically correlate corresponding points, so the pair can be directly compared, combined or analyzed. Image registration is an active research field and in the last years there is growing interest in medical imaging applications.

Mono- and multi-modal registration

The registration methods are classified into two main categories. When the images are taken with the same sensor, the registration is known as monomodal or intra-modality registration; otherwise, when the aim is to compare images collected using different imaging techniques, the registration refers to multi-modal methods, also called inter-modality or cross-modality registration. In the monomodal comparison significant intensity or contrast variations are not expected, rather geometrical differences due to displacements and deformations of the objects. For medical images, in the case of acquisitions at different times the aim is to help to relate clinically relevant information useful for monitoring changes in an individual over time. When the sensor has a limited field-of-view or the quality of investigation and the amount of information depend on the orientation of the sensor, it could be helpful to acquire more images of the same subject from different viewpoints. The aim of registering these images is to improve the visualization and the understanding of the subject's anatomy. In this case differences in image contrast among acquisitions are expected. In the multimodal comparison the images may differ for image intensities, geometric distortions caused by different patient positioning, and even type of information, often complementary among images, depending on the imaging technique. Due to the different physical phenomena that are measured by the different modalities, there is no functional relation between the intensity maps of corresponding anatomies. In addition, the two images may be represented in different coordinate systems. Multimodal comparison represents a very complex task with a lot of degrees of freedom. The purpose is in fact to take advantage of the integration of more information of the same subject for a more complete analysis. The correlation between the two coordinate systems and the integration of information are the tasks of image registration. The geometry of the anatomy is not the only interesting property. Indeed, tissues can look similar but have different functionalities.

Such functional information can improve the characterization of the tissues (and help in the definition of clinical volumes).

Rigid and non-rigid transformation model

The transformation model that maps the two images can include a global transformation function between the coordinate systems or a local transformation function that takes into account even relative displacements between objects. In the first case, the model can involve any combinations of translation, rotation, scaling or even shearing motion (affine transformations) between the coordinate systems, and the transformation model is called rigid. This type of transformation accounts for the overall motion of the object, that means the distance between any two points in the first image is preserved when mapped onto the second image. When the matching function between the two images involves also relative motions and local deformations of the objects it is called nonrigid or deformable transformation model. In this case correspondence between structures in two images cannot be achieved without localized stretching of the images. The term deformable denotes that the images are associated through a dense vector field, representing a spatially varying deformation model. For some medical applications a rigid body approximation can be used. For example, brain images of the same subject, even if acquired with different modalities, often present relatively little change in brain shape or position within the skull over relatively short periods between scans. In most cases, a rigid transformation is not adequate for representing the actual spatial transformations. In particular, soft tissues and other deformable organs can vary in shape as well as in position relative to surrounding structures. Generally, the patient motion due to different positioning could be coarsely corrected applying a global alignment of the images; meanwhile internal motions of the organs after different positioning or different breathing cycle, and pathological changes between scans cause non-rigid deformations, which need to be compensated using deformable models. In the case of breast images, since the breast is entirely composed of soft tissue, it easily deforms during different acquisitions, thus its behavior cannot be modeled simply as rigid and a non rigid registration method is required. Today the most interesting and challenging works in registration involve the development of application-specific non-rigid matching techniques. These are known to produce promising and satisfactory results; however, there are several problems with their applications. One is the fact that these approaches are computationally intensive and time demanding. Frequently the algorithms can take many hours to register images, which makes them unsuitable for interactive use, allowing only off-line interpretation. The recent techniques of computer assisted surgery or real time serial registration of intraoperative images require fast and sophisticated implementations of the algorithms and high performance workstations. Another point is that these techniques are sensitive to initial positioning of

the images to be matched. If the initial global alignment is not enough accurate getting closer corresponding points, the deformable matching procedure may perform poorly, i.e. could not converge to the wanted result. Therefore, a satisfying pre-registration step can be a key issue. A third obstacle to a widespread clinical use of non-rigid registration is the difficulty in validating the results, as explained below. For all these points deformable registration has been, along with organ segmentation, one of the main challenges in modern medical image analysis. Accurate non-rigid motion estimation is still an open issue, especially in highly deformable anatomies.

Ill-posed nature of image registration

The computation of the transformation function is complicated by the inherently ill-posed nature of the registration problem. According to Hadamard's definition, a well-posed problem satisfies the following properties: existence, uniqueness, and stability of the solution. Registration is usually an underdetermined inverse problem (i.e. more unknowns than equations), meaning that information is not enough to uniquely determine the solution and to constrain this solution being stable (i.e. the solution's behavior changes with continuity with initial conditions).
This implies that:

- many transformations can produce very similar results reaching a similar good match between the images (Rohlfing 2006) (i.e. there are not enough constraints from images only to define a unique optimal solution);

- optimal matching function does not guarantee a reasonable solution reflecting real non-rigid deformations (i.e. there are not enough constraints from images to constrain the solution being physically meaningful). Constraints have to be included for disallowing unrealistic folding or tearing of the structures. It is generally very complex to incorporate all appropriate physical and mathematical constraints within the objective function of an inverse problem;

- still worse, different methods lead to different solutions in function of the internal modeling of the transformation (Rohlfing 2006). For example, if geometrical features are used to match images, there will be many different possible deformation fields which can align those features but which behave differently away from those features. Similarly there will also be many possible deformation fields that can result in voxel intensities appearing to be well matched between images. The estimated solution strongly depends on the method and on the constraints;

- for most medical images, registration is affected by the problem of local minima, especially when images are noisy: the estimated solution is un-

stable if falls into local minima, instead of converges to local maximum. The local minima are one of the causes of the distortions mentioned above.

In addition, another important issue that need to be addressed is the criterion to validate the resulting solution. This has to evaluate how the match is accurate, how the estimated solution is good among all the possibilities and it has to establish if the solution is optimal and adequate for a particular application. In addition, in case of non-rigid transformations the validation is not an easy task. In fact, the validation criterion has to take into account that dense fields, defined in each location of the images, are estimated. Then a pointwise validation could be proper in these applications, but it is an impracticable solution. This issue of current importance is still unsolved. All these considerations explain the current emphasis on the development and validation of application-specific non-rigid techniques.

3.2   Definitions

In this section a description of a common terminology and basic concepts of image registration is presented.

Reference and template
When two images are being registered, one is conventionally regarded as static and defines a frame of reference and the other is transformed to bring corresponding features into alignment. The static image is variously known as the reference, target or fixed image. The image undergoing transformation is also known as the template, source, update or floating image. Given the two images, the goal of the image registration is to find a spatial transformation, such that the deformed template matches the reference image according to a suitable distance measure between image features. The criterion used to register the images is known as the similarity measure or the objective or cost function.

Displacement field and reverse mapping
The result of deformable registration is a vector field called displacement or deformation field. It is a dense vector field, because defined for each location of the reference domain. The displacement field describes the mapping function of corresponding locations between the reference and the template. Hence, it describes how to spatially transform the template in order to obtain a deformed image (i.e. resampled to the space of the reference image) similar to reference. When the vector field defines the transformation function that maps points from the template to the resulting points of the transformed template, the function refers to a forward mapping. If the template is considered as input space and the transformed template is the destination or output space, in the forward mapping procedure, the function hereafter denoted by

T maps each input image pixel (u_k, v_k) of the input space to the destination point (x_k, y_k) of the output space (see figure 3.1). However, forward mapping
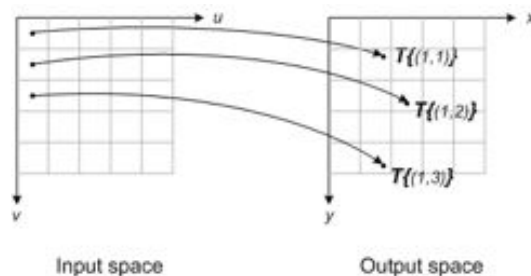

Figure 3.1: diagram illustrating the forward mapping procedure, where the transformation T describes the destination (x_k, y_k) for every location (u_k, v_k) in the input space, i.e. $(x_k, y_k) = T(u_k, v_k)$.

has two main disadvantages: gaps and overlaps. This means that it could happen that some output pixels do not correspond to any input image pixels (i.e. their value is undefined) or some output pixels corresponds to more than one input image pixel. In both cases, it is challenging to figure out a reasonable way to set the value of those output pixels. In the case of reverse or backward mapping (see figure 3.2), the displacement field is defined for each location on the destination image (i.e. transformed template), and it describes the corresponding location in the source image (i.e. template). In this case the value of each pixel of the transformed template is determined. The advantage of this approach is that the field is directly defined on the sampling grid of the destination image. This approach completely avoids problems with gaps and overlaps. Only an interpolation method is needed to establish a rule for determining the output pixel value (x_k, y_k) when the input location (u_k, v_k) does not coincide to the center of an input grid pixel. If the mapping function is invertible, the forward and reverse mapping are the same, and the reverse mapping corresponds to the inverse spatial transformation used in the forward mapping; otherwise, the reverse matching is more convenient. In this thesis the displacement field refers to a reverse mapping estimation. Note that, the term transformation (or warping) will be usually used to denote the application of a transformation model or a vector field on an image, while deformation denotes the internal non-rigid motion of a body in a physical point of view, and displacement field is the dense vector field representing such non-rigid transformations.

Field inverse consistency
An important property of a displacement field is its symmetry, also called field inverse consistency, that guarantees the same result of the image registration independently on which image is chosen as reference and template. In general the resulting transformation is asymmetric, especially for deformable registration, as only one of the images is transformed during the process. Indeed,
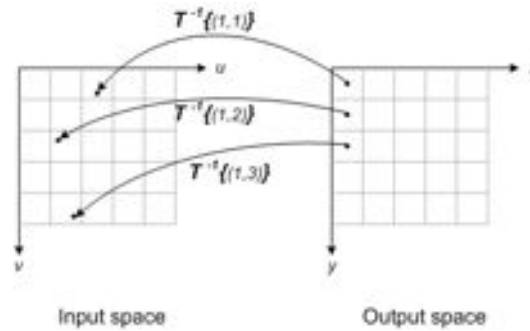
Figure 3.2: diagram illustrating the reverse mapping procedure, where the inverse spatial transformation T^{-1} is applied to determine the corresponding the location (u_k, v_k) in the input space for each destination pixel (x_k, y_k), i.e. $(u_k, v_k) = T^{-1}(x_k, y_k)$. Using an interpolation procedure, an approximate value for the destination pixel (x_k, y_k) is set from the input pixels nearest to input location (u_k, v_k).

the registration from template to reference will not necessarily lead to the inverse transformation as the registration from reference to template. As a consequence, interchanging the order of the input images, there is no guarantee that each point of the reference can be related to a corresponding position in the template image using the inverse transformation of the resulting displacement field that maps each point of the template to reference. An inverse consistent registration method (that will be explained in the next section) can be implemented to ensure the symmetry of the resulting displacement field. It is achieved constraining the displacement field estimated from reference to template images to be consistent and coherent with the displacement field estimated from template to reference during the image registration procedure.

Feature vector and feature descriptor
Feature is a widely used term in image processing and it may have different meanings depending on the application. A feature represents a relevant information in an image: it may refer to a distinctive local structure, like an interest point, edge, corner or blob, as well as it may refer to a measurable property that characterizes a structure, like its gray-level intensity value, orientation and scale. In the first case, the feature represents a structure or a whole object in an image and can be the result of a feature detection applied to the image. A feature detector represents a method aimed at identifying interesting points, i.e. features, in the image. In the second case, a feature represents a property used to characterize a pixel. If more features are used to describe a pixel, a feature vector is associated to the pixel. A feature vector is an n-dimensional vector that contains information describing a structure's important characteristics. Given a pixel, the feature vector describes the image structure in a neighborhood, i.e. a local image patch, around the pixel.

The most basic vector is a set of numbers representing the intensity values and its size corresponds to the number of pixels of the image patch. Otherwise, a set of local image features can be extracted, where each feature can identify the pixel location, the scale and the orientation in the local image neighborhood, or other properties. Thus, the vector is a list of features and its size depends on the number of features used to characterized the pixel. A feature descriptor typically indicates a method used to extract a specific set of information from an image, and the output of a feature descriptor is a feature vector corresponding to each pixel. Its role is to characterize the local image appearance around a pixel using the most representative features. The choice of features to be determined depends on the specific problem.

3.3 Components of registration methods

This section provides a summary of the state-of-the-art techniques proposed in literature (Zitova et al. 2003, Crum et al. 2004, Modersitzki 2004, Goshtasby 2005, Sotiras et al. 2012, Sotiras et al. 2013) and introduces the basic concepts useful to describe the nonparametric formulation of an image registration problem, that will be used in this thesis.

Given a reference image denoted by R and a template image denoted by T, both defined in the image domain $\Omega \subset \mathbb{R}^d \to \mathbb{R}$, $d = 2, 3$, the goal of registration is to estimate a displacement field $\mathbf{u} : \Omega \to \mathbb{R}^d$ such that the transformed template T' image is similar to the reference image R. For a particular point $\mathbf{x} \in \Omega$, the quantity $R(\mathbf{x})$ is the image intensity value at the spatial position $\mathbf{x}$ of the reference and $T(\mathbf{x})$ refers to the intensity at the position x of the template. Then the purpose of the registration is to determine a displacement field $\mathbf{u}$ such that $T(\mathbf{x} + \mathbf{u}(\mathbf{x})) = T'(\mathbf{x}) = R(\mathbf{x})$ or similar to $R(\mathbf{x})$. The optimal displacement field is often achieved by an energy optimization problem, that will be presented in detail in the following. Transformations used in non-rigid registration range from smooth local variation described by a limited number of parameters to dense displacement fields defined at each voxel. Depending on the type of transformation a different formulation of the energy functional results. Every image registration algorithm relies on defining three components, that are selected according to the application: a similarity measure, a transformation model, and an optimization strategy.

3.3.1 Similarity measure

The similarity measure, also called metric or matching criterion, determines the quality of the match between images. The similarity is evaluated between feature vectors extracted from both the images. The aim of registration is to compute a displacement field such that these differences are as small as possible. Different similarity criteria exist, depending on the type of features

used for the matching. Registration can be based on geometric information, i.e. a small set of interest points detected from the images, or onto measures computed directly on the image intensity values. Hence, generally registration methods are classified into: landmark-based, intensity-based, and hybrid approaches. Therefore, the choice of the metric depends on the imaging modalities and on the structures represented in the images to be registered.

Landmark-based registration
This approach exploits geometric information, i.e. anatomical information in medical images, that usually are sparsely distributed throughout the images. The landmarks are assumed to be placed in salient image locations which are considered to correspond to meaningful anatomical locations. The most common features include control points, edges, contours, surfaces, and so on, which firstly need to be extracted and then are matched to their counterparts found in the second image. These correspondences define the transformation. Also fiducial markers can be used to guide image registration. In this case, the metric can be simply the sum of distances between geometrical features identified in both images, and the transformation results from the minimization of this metric. After establishing explicit correspondences between the pairs of landmarks, interpolation is used to infer the values of the displacement filed throughout the rest of the image volume in a way consistent with the matched landmarks. The use of such structural information ensures that the mapping has biological validity. However, transformations that give a good features similarity, often produce tearing as result of correspondence problems. Moreover, if the features form a sparse set of correspondences, the interpolation procedure results in a decrease in accuracy as the distance from the landmarks increases. Two advantages of geometric registration are that it is independent to image contrast variations and it is robust with respect to the initial conditions and the existence of large deformations. The solution of the registration problem is relatively straightforward once landmarks have been extracted. On the other hand, it requires good landmark detection. This approach, in fact, relies on the accurate definition of these geometrical features, which is possible when identifiable anatomical structures are visible in both images, or when their manual segmentation is available. Their automated identification may be complex in practice as it requires robust and accurate feature detection algorithms. Geometry-based metrics are therefore more appropriate when manual pre-processing is feasible. For all these reasons in case of breast image registration this approach could be not advantageous.

Intensity-based registration
In this case the feature is the grayscale value of the pixel, and the image intensities are directly exploited to compute the transformation. This approach, in fact, uses mathematical or statistical criteria to describe and match intensity patterns, but does not use anatomical knowledge. The intensity pattern in

a location $\mathbf{x}$ is commonly extracted from a neighborhood of $\mathbf{x}$, $N(\mathbf{x}) \subset \Omega$, that, in most cases, takes the form of a representative rectangular patch, or sub-image, centered in $\mathbf{x}$.

These methods define a measure that evaluates the similarity of intensity patterns around two locations (one from the reference and one from the template) and adjust the transformation until the summation over the whole image of the similarity measures between corresponding locations is maximized. It is convenient to define a function $\Psi : \Omega \to \mathbb{R}$ associated to a similarity metric as follows: given the reference image R, the template image T and the displacement field $\mathbf{u}$, $\Psi(\mathbf{x}; \mathbf{u}, R, T)$ is the function that measures the similarity of each location $\mathbf{x}$ in R with its corresponding location in T. The quantity to be maximized can be then expressed as:

$$\sum_{\mathbf{x} \in R} (\Psi(\mathbf{x}; \mathbf{u}, R, T)). \tag{3.1}$$

When compared to the geometric methods, this approach has the potential to better represent the accuracy of the dense deformation field. Nonetheless, it comes at the cost of increased computational expense.

Popular similarity measures include the sum of absolute sifferences, the sum of squared differences, cross Correlation, and mutual information. These matching criteria can be presented according to whether they tackle mono-modal or multi-modal registration problems.

Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD)

These are the simplest similarity measures; they assume that the image patches have identical intensity patterns except for Gaussian noise. Using these metrics the similarity is maximized when the pointwise difference between two patches is as small as possible. SAD metric is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = - \sum_{\mathbf{y} \in N(\mathbf{x})} |T(\mathbf{y} + \mathbf{u}(\mathbf{x})) - R(\mathbf{y})| \tag{3.2}$$

while SSD it is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = - \sum_{\mathbf{y} \in N(\mathbf{x})} (T(\mathbf{y} + \mathbf{u}(\mathbf{x})) - R(\mathbf{y}))^2 \tag{3.3}$$

The minus sign before the summation indicates simply that SAD and SSD are actually measures of dissimilarity and should be minimized to maximize the similarity. Both are based on an intensity conservation hypothesis and are not invariant to rotation and scale; hence, for these two similarity measures to be successful, the assumption of comparable intensity between the same anatomical structures is required. Therefore, their use is restricted to monomodal registration tasks, where the images share the same imaging modality.

Cross Correlation (CC)

If a linear relationship between the intensities of two images is assumed, the optimal criterion is the cross correlation. CC value is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = \sum_{\mathbf{y} \in N(\mathbf{x})} T(\mathbf{y} + \mathbf{u}(\mathbf{x})) \cdot R(\mathbf{y}) \tag{3.4}$$

It is usually computed on the zero-mean version of the patches (subtracting each patch for its average intensity) and normalized dividing by the square root of the auto correlation of both patches (Ibanez et al. 2003); in this form is known as Normalized Cross Correlation (NCC).
NCC value is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = \frac{\sum_{\mathbf{y} \in N(\mathbf{x})} (T(\mathbf{y} + \mathbf{u}(\mathbf{x})) - \bar{T}) \cdot (R(\mathbf{y}) - \bar{R})}{\sqrt{\sum_{\mathbf{y} \in N(\mathbf{x})} (T(\mathbf{y} + \mathbf{u}(\mathbf{x})) - \bar{T})^2 \cdot \sum_{\mathbf{y} \in N(\mathbf{x})} (R(\mathbf{y}) - \bar{R})^2}} \tag{3.5}$$

where:

$\bar{T} = \frac{\sum_{\mathbf{y} \in N(\mathbf{x})} T(\mathbf{y} + \mathbf{u}(\mathbf{x}))}{|N|}$ is the average of the neighborhood of $\mathbf{x} + \mathbf{u}(\mathbf{x})$ in T, and
$\bar{R} = \frac{\sum_{\mathbf{y} \in N(\mathbf{x})} R(\mathbf{y})}{|N|}$ is the average of the neighborhood of $\mathbf{x}$ in R.

Multi-modal registration is more challenging, because the choice or the definition of an appropriate matching criterion that could extract discriminative information, visible in both modalities, is a harder task. To solve this problem, the similarity measure needs to define weaker relationships between intensity that could highlight the underlying structures, and at the same time be less affected by the different intensity characteristics of the imaging modalities. Most of the proposed approaches make use of information theoretic measures, related by some statistical or functional relationship, like: entropy, joint entropy, mutual information, and normalized mutual information.

Mutual Information (MI)

This is a metric derived from the information theory and measures the statistical dependency of two random variables. It is based on Shannon entropy and aims to find a statistical intensity relationship across images that maximizes the amount of shared information between two images. MI does not assume any relationship between the image intensities, but it only assumes that a probabilistic relationship between pixel intensities is maximized at registration. Owing to its statistical nature, this metric is invariant to intensity and is appropriate for the multimodal registration, where intensity changes are assumed.

If I is the intensity domain of the image R and T, the MI value is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = H(N_R(\mathbf{x})) + H(N_T(\mathbf{x} + \mathbf{u}(\mathbf{x}))) - JH(N_R(\mathbf{x}), N_T(\mathbf{x} + \mathbf{u}(\mathbf{x}))) \tag{3.6}$$

with $H(N_R(\mathbf{x}))$ the entropy of the patch neighborhood of $\mathbf{x}$ in R:

$$H(N_R(\mathbf{x})) = \sum_{i \in I} P(i, N_R(\mathbf{x})) \cdot log(P(i, N_R(\mathbf{x}))) \tag{3.7}$$

$H(N_T(\mathbf{x} + \mathbf{u}(\mathbf{x})))$ the entropy of the patch neighborhood of $\mathbf{x} + \mathbf{u}(\mathbf{x})$ in T:

$$H(N_T(\mathbf{u}(\mathbf{x}))) = \sum_{i \in I} P(j, N_T(\mathbf{u}(\mathbf{x}))) \cdot log(P(j, N_T(\mathbf{u}(\mathbf{x})))) \tag{3.8}$$

and $JH(N_R(\mathbf{x}), N_T(\mathbf{x} + \mathbf{u}(\mathbf{x})))$ the joint entropy of the patch neighborhood of $\mathbf{x}$ in R with the patch neighborhood of $\mathbf{x} + \mathbf{u}(\mathbf{x})$ in T:

$$JH(N_R(\mathbf{x}), N_T(\mathbf{x} + \mathbf{u}(\mathbf{x}))) = \sum_{i \in I} \sum_{j \in I} JP(i, j, N_R(\mathbf{x}), N_T(\mathbf{x} + \mathbf{u}(\mathbf{x}))) \\ \cdot log(JP(i, j, N_R(\mathbf{x}), N_T(\mathbf{u}(\mathbf{x})))) \tag{3.9}$$

where $P(i, N_R(\mathbf{x}))$ is the probability to find the intensity value i in the neighborhood of $\mathbf{x}$ in R, $P(j, N_T(\mathbf{x} + \mathbf{u}(\mathbf{x})))$ is the probability to find the intensity value j in the neighborhood of $\mathbf{x}+\mathbf{u}(\mathbf{x})$ in T, and $JP(i, j, N_R(\mathbf{x}), N_T(\mathbf{x}+\mathbf{u}(\mathbf{x})))$ is the joint probability to find the intensity values i and j in corresponding locations in the neighborhoods of $\mathbf{x}$ in R and of $\mathbf{x} + \mathbf{u}(\mathbf{x})$ in T respectively. These probabilities are commonly represented by the normalized intensity histograms extracted from the patch they refer to.

Since MI does not consider the spatial information of a particular pixel within a patch, MI is not invariant to changes in image overlap. Thus, in certain cases it may be possible that MI is maximized when the images get misaligned. Therefore, the **Normalized Mutual Information (NMI)** has been proposed to minimize the overlap problem. It is based on the assumption that a lower entropy of the joint intensity distribution corresponds to a better alignment.

The NMI value is expressed as:

$$\Psi(\mathbf{x}; \mathbf{u}, R, T) = \frac{H(N_R(\mathbf{x}) + H(N_T(\mathbf{x} + \mathbf{u}(\mathbf{x})}{JH(N_R(\mathbf{x}), N_T(\mathbf{x} + \mathbf{u}(\mathbf{x})))} \tag{3.10}$$

Although this formulation can solve the image-overlap issue, the lack of structural information is a serious drawback for most deformable registration tasks; Different approaches have been proposed to include spatial context information into MI: Pluim (Pluim et al. 2000) combined MI with image gradient

information but clearly with this approach the modality invariance of the metric is quite completely lost as the image gradient is a modality dependent feature; other approaches (Studholme et al. 1999, Loeckx et al. 2010),like the so called Conditional Mutual Information (CMI), increase the number of dimensions of the statistical variables used to compute the joint probabilities; this is commonly implemented adding a third or forth channel in the joint probability histogram that includes structural information. An obvious drawback of these approaches is their extremely high computational complexity that makes them unsuited for a full resolution deformable field evaluation.

Hybrid registration
This approach exploits both the sparse geometric correspondences along with an intensity-based criterion to improve the estimation of the spatial transformation. Combining geometric features and intensity features in registration should result in more robust methods.

3.3.2   Transformation model

The transformation model defines how the template image can be warped in order to match the reference. It should model the deformation relating the two images. Hence, the choice of the class of geometric transformations implies an assumption, i.e. knowledge of a priori information, regarding the nature of the deformation to be recovered. Different types of transformations exist, depending on the number of degrees of freedom of the model to be represented. As mentioned before in the introduction, rigid and non-rigid models are possible. The registration methods can be broadly classified into parametric and non-parametric, depending on the method used to estimate the transformation model.

Parametric methods aim at optimizing the parameters of a transformation model in low-dimensional (e.g for rigid transformation) or high-dimensional spaces. The number of parameters that the registration needs to estimate corresponds to the degrees of freedom of the expected deformation model: the more the number of parameters, the more flexible is the transformation model and the more complex and time demanding is the computation. Instead non-parametric methods act directly on the displacement field (Holden 2008) and represent a multidimensional optimization problem. In the first case the a priori information is directly included into the model used to described the non-rigid transformation (implicit regularization achieved by the parametrization of the displacement field), while in the latter, it has to be included in the optimization process by using proper regularization technique (explicit regularization, making use of constraints that the solution must satisfy).

Hereafter an overview of the existing techniques for both these approaches is presented.

Parametric registration

Rigid and affine transformations
If the registration model is rigid, only a few parameters are needed to define the global transformation. The rigid transformation corresponds to a problem with 6 degrees of freedom; it is completely described by 6 parameters, 3 for the translation vector and 3 for the rotation vector. The affine transformation can require up to 12 parameters, including also 3 parameters for scaling and 3 for shearing. The transformation matrix is then applied to the whole image.

Parametric deformable registration
When local deformations are included, the non-rigid transformation needs more parameters to be accurately modeled. A simple way of increasing the number of degrees of freedom is to define the transformation as a combination of k pre-defined transformations, or to define the transformation by interpolation of the displacement computed in only l pre-defined control points (i.e. through a definition of a mesh). The parametric registration consists then in searching for the parameters (k parameters in the first case, and $3l$ in the second case) that optimize the objective function. The most common methods are the radial basis functions (Fornefett et al. 1999), the thin-plate splines (Rohr et al. 1996), the B-spline based free form deformation (Rueckert et al. 1999), and the mesh-based finite element method (Ferrant et al. 2000). These families of interpolation strategies use corresponding control points, defined in the reference and the template image, and a basis function to define correspondences away from these points. These points can be organized using a regular grid (like in the B-spline method) or using a mesh (like in the finite element method). Therefore, the displacement of these points are taken as parameters for defining the global transformation. Hence, the displacement is evaluated in this restricted set of locations in the image (that can be displaced from their original location), and the displacement of all voxels can then be computed using interpolation between these points. Parametric methods may be computationally expensive, and can lead to some problems related to optimization in high-dimensional spaces when the objects to be registered are highly deformable. Moreover, the choice of good control points is not always an easy task. Indeed, in order to keep as few parameters as possible for the optimization, the number of control points must be as low as possible, without decreasing the accuracy of the transformation. Therefore, the points should be located where there is the most useful information in the images (e.g. on high contrast regions, edges, etc.) in order to get a good matching after registration. This makes such methods potentially difficult to fully automate in practice.

- Radial basis functions (RBFs)
  RBF is one of the most important families of interpolation strategies, where the value of the resulting mapping function at an interpolation point has the form of a linear combination of radial symmetric functions (plus a linear term) of its distance from the known control points. A characteristic of this method is that a deformation field can be interpolated from a set of irregularly placed known samples, but the displacement determined at one point influences the values of all other points in the whole image domain. It means that each control point can vary locally, but has a globally influence. Thus, an interpolation in sparsely populated areas is feasible, but on the other hand, this behavior can be a disadvantage when seeking local transformations. It limits the ability to model complex and localized deformations. Hence, in order to counter it, sufficient landmarks are required in the regions of interest. Nevertheless, as the number of control points increases, the computational cost associated with moving a single point rises steeply.

- Thin-plate splines (TPS)
  TPS is an interpolation method that uses a spline function, expressed as a weighted sum of radially symmetric logarithmic functions, to define correspondences away from the control points. Also in this case, each control point belonging to a thin-plate spline has a global influence on the transformation. Hence, also a thin-plate spline is not suitable for the registration of images with local geometric differences. TPS represents the equation of a plate of infinite extent deforming under loads centered at the control points. The plate deflects under the imposition of loads whose values correspond to the weight of the spline function. Good accuracy is achieved at and near the control points, errors are large away from the control points, and this can be attributed to the logarithmic basis functions, which are radially symmetric, used to define the transformation. When the arrangement of the points is nonsymmetric, large errors are obtained in areas where large gaps exist between control points.

- B-spline based free-form deformations (FFDs)
  FFD is one of the most common types of transformation models used in medical image registration. It consists of a regular grid (superimposed on the image) that gets deformed under the influence of the control points free to move independent on its neighbouring grid points. This gives an optimization problem having several millions of variables. The dense deformation field is given as a summation of cubic-B splines. By contrast the thin-plate splines, B-splines are only defined in the vicinity of each control point; perturbing the position of one control point only affects the transformation in the neighbourhood of the point. Thus, B-splines are locally controlled and computationally efficient. Neverthe-

less, the local flexibility and the computational complexity of the local motion model are related to the resolution of the B-spline control points mesh. More control points may improve the registration accuracy, but the computation time will also increase dramatically. Main disadvantage of B-spline based non-rigid registration techniques is that special measures are sometimes required to prevent folding of the deformation field and these measures become more difficult to enforce at finer resolutions. In the implementations, the control point spacing, the stopping criteria, and the weighting parameters in the cost function need to be determined experimentally.

- Finite element method (FEM)
  FEMs are mesh-based registration methods that model the biomechanical properties of the tissues. These methods aim to include physical behavior in specific registration applications. The main motivation of using these methods is based on the assumption that by creating models of deforming organs that are consistent to their physical properties, the plausibility of the estimated deformation improves and moreover the registration is able to better cope with challenges due to the presence of outliers or large deformations. Therefore, the knowledge regarding the deformation may be achieved exploiting the knowledge about the deformability of the tissues and constructing biomechanical deformation models using the FEM that mimic the properties of the tissues.
  FEM allows more principled control of localized deformations. In fact, FEM does not allow to get an accurate estimation of the displacement of both the surface nodes of a volumetric mesh and the inside nodes according to physical laws (such as linear elasticity). These models divide the image into cells (note that the number of finite elements needs to be found experimentally as trade–off between accuracy and calculation time) and assign to these cells a local physical description of the anatomical structure. For instance, soft tissue can be labelled as elastic, bone as rigid, and so on. Then, external forces such as landmark correspondences or voxel similarity measures are applied to the model, which deforms according to the material behavior in each cell. However, creating data-specific meshes is cumbersome for several reasons: geometry is sometimes highly complex (especially for multi-body 3D images) and placing nodes at good position (typically on high contrast voxels) is not an easy task. As the method aim to faithfully represent anatomical structures with the optimized model, the material properties as well as the necessary geometry and boundary conditions are required and need to be accurately defined. This is a challenging procedure because the understanding of the material properties is often limited. And moreover the definition of the geometry requires an accurate segmentation of anatomical structures as well as appropriately meshing of the image

domain. Uncertainty in the specification of these parameters may lead to undesirable bias.

Non-parametric deformable registration

In non-parametric registration methods, the optimization is not performed in a high-dimensional space of parameters, but directly on the displacement field itself (involving a huge number of variables, equal to the number of voxels times the 3 dimensions of the space). This family of methods is also denoted as optical flow registration, owing to the first non-parametric registration method (Horn et al. 1980). This was originally developed for video sequences, and it is based on the constant intensity assumption. The supportive hypothesis of this method is that the intensity of local time-varying structures is constant under motion in the interval of time between two image acquisitions. Nevertheless, this assumption is no longer valid in serial medical images, hence optical flow methods have not been widely adopted for medical applications.

Most of these schemes proposed for solving the nonparametric image registration problem are typically expressed as an optimization problem of an energy functional, combination of two terms:

$$\mathcal{J}[\mathbf{u}] = \mathcal{D}[R, T'; \mathbf{u}] + \alpha \mathcal{R}[\mathbf{u}] \qquad (3.11)$$

where $\mathcal{D}$, called feature term, is a distance measure that quantifies the image similarity between the reference image R and the template image T under the influence of a displacement field $\mathbf{u}$ (the notation T' denotes that the template is deformed). It represents the energy of the features that accounts for the forces that move each pixel of the template image toward positions that demonstrate similar characteristics in the image reference (i.e. the external forces defined to obtain the desired registration result). The second one, $\mathcal{R}$, is the normalization or regularization term, that is related to the properties of the transformation model and it regularizes the deformation field in order to preserve its smoothness and invertibility (i.e. the internal forces designed to keep the displacement field smooth during deformation and prevent undesirable discontinuities). As displacement vectors are computed independently for every voxel, the field might not be smooth due to image noise or to the specific behavior of the field calculation strategy. It is therefore crucial to include more a priori knowledge into the registration model and constraints. Therefore, the displacement field needs to be regularized in order to get a smoother transformation and to avoid local minima. The parameter α may be used to control the strength of the smoothness of the displacement versus the similarity of the images. The image is deformed until the forces reach an equilibrium.

Figure 3.3 reports an example created to highlight the effect of the regularization term. Figure 3.3(c) shows the result obtained optimizing the functional

using only the feature term. The red channel of the image represents the reference, while the green channel represents the template after the application of the resulting displacement field. It can be noted how the feature term moves the structures of the template towards the similar structures on the reference, but at the same time in some regions the optimization can lead to unrealistic distortions. If the regularization is included, the artifacts disappear and a more realistic solution is achieved (figure 3.3(d)). The regularization may be achieved through the use of either hard constraints or soft constraints. Hard constraints are the constraints that the solution must satisfy in order for the registration to be successful, e.g. regularization based on physics-based models. The most common regularizers presented in the literature are: the elastic, fluid, diffusion, and curvature regularizers. These are inspired by the physical behavior of the tissues under deformation. For instance, the elastic regularizer treats the template image as a linear, elastic body and deform it using forces derived from the image similarity measure. Soft constraints are introduced as additional terms in the energy function that penalize non-regular configurations. Specific goals may be used such as: topology preservation (i.e. equivalent to the invertibility of the deformation field), volume preservation, and rigidity constraints. Volume preservation is of particular interest when it is known that the imaged anatomical structure is not compressible and that all changes are due to motion. A simple example is a bone structure; more complicated cases include deformable structures that preserve their volume such as breast, myocardium and liver. The rigidity constraint, instead, represents a term added to control the local rigidity of the transformation of rigid anatomical structures. Of great importance for biomedical applications are the constraints that may be applied to the transformation in order to exhibit special properties, that include: inverse consistency, symmetry, diffeomorphism. The inverse consistency constraint is aimed to avoid a bias on the registration outcome due to the choice of one image to be the reference and the other to be the template image. This constraint introduces a term that penalize the difference between the forward and backward transformations during their simultaneously estimation (Avants et al. 2008, C. Chen et al. 2007, (Christensen et al. 2001). It also constrains the forward transformation to be the inverse of the backward transformation. A diffeomorphism is an invertible geometrical transformation such that both the function and its inverse are smooth (Arsigny et al. 2006, Vercauteren et al. 2009). Imposing the displacement field to be diffeomorphic means that no physically implausible folding of volume occurs. As organs can be compressed and deformed, but cannot undergo non-invertible spatial transformations, this constraint should be used to ensure that displacement fields represent physical deformations (Arsigny et al. 2006).
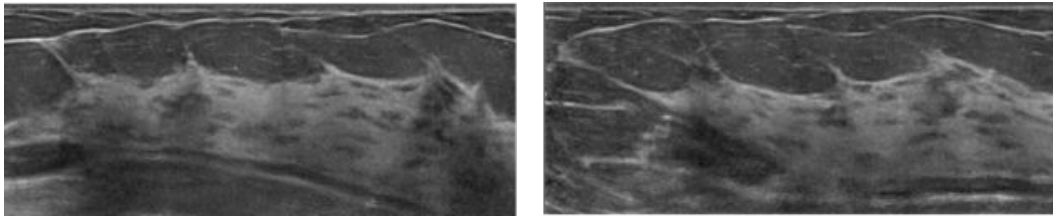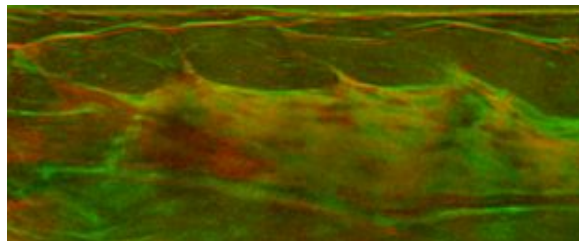
(a)
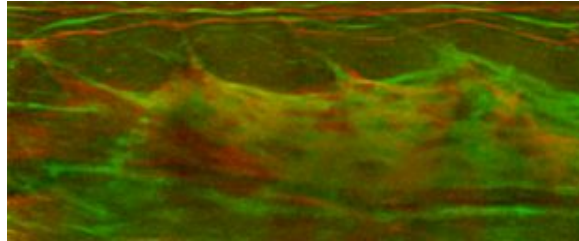
(b)
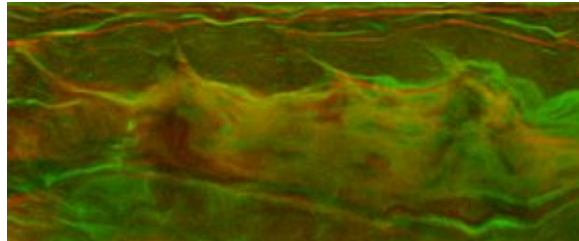
(c)

(d)

Figure 3.3: (a): reference (on the left) and template (on the right) images; (b): overlap image, where the red channel of the image is used for the reference, and the green one for the template. The last two images show the results after the application of the estimated displacement field as moved only by the features (c), or resulting by a balance between feature and normalization terms (d).

3.3.3   Optimization strategy

Optimization refers to the manner in which the transformation is adjusted to improve the image similarity. Hence, the aim of the optimization is to infer the optimal transformation that best matches two images according to the en-

ergy functional comprising the similarity and the regularization terms. As a consequence, the choice of the optimization methods impacts on the quality of the obtained result. Choosing a good optimizer requires a good understanding of the registration problem, and the constraints that can be applied. A good optimizer is one that reliably and quickly finds the best possible transformation. It means that the more the time required by the optimization method to estimate the displacement field (i.e. it is a computationally demanding strategy), the more the chance to converge to local minima, which can result in a good image match that nevertheless is not the best one.

Optimization of the energy functional is achieved by solving the variational (Gâteaux) derivative of $\mathcal{J}[\mathbf{u}]$ (eq. 3.11), i.e. of the regularizer and similarity measure. These results in the corresponding partial differential equations (PDE) known as Euler-Lagrange equations:

$$f(\mathbf{x}, \mathbf{u}) + \alpha A[\mathbf{u}](\mathbf{x}) = 0 \quad \mathbf{x} \in \Omega \tag{3.12}$$

with

$$f(\mathbf{x}, \mathbf{u}) = (R(\mathbf{x}) - T(\mathbf{x} + \mathbf{u}(\mathbf{x}))) \cdot \nabla T(\mathbf{x} + \mathbf{u}(\mathbf{x})) \tag{3.13}$$

where A is the partial differential operator, and it is related to the Gâteaux derivatives of the regularization term $\mathcal{R}$; f is the Gâteaux derivative of the distance measure $\mathcal{D}$, and it may be thought off as the driving force for the associated PDE.

Numerical method for the minimization problem
Most non-parametric registrations are based on an iterative process. The idea is to progressively build a proper displacement field by iteratively improving the matching between the reference image and the template image warped by this displacement field, according to the similarity metric. Hence, over the last decade most of the research effort into nonparametric image registration has gone into the development and use of numerical approaches to solve the Euler-Lagrange equations. There are various ways to define the previous parameters. For example, the step size may be constant, or decrease with each iteration, or such that it minimizes the energy functional along the search direction. The search direction can be specified by exploiting only first-order information or, for example, by also taking into consideration second-order information. It is the choice of these parameters that distinguishes different methods. Most techniques apply a fixed-point iteration scheme directly to the Euler-Lagrange equations, yielding an algorithm that requires successive solutions of linear PDE systems until a stationary solution is found.

Many registration algorithms use existing optimization schemes. Commonly used numerical methods include:

- Gradient descent method, where the energy functional is optimized by following the direction that decreases the energy or its negative gradient.

- Quasi-Newton method, that aims to accumulate information from the previous iterations and take advantage of it in order to achieve better convergence. More specifically, these methods aim to estimate the inverse Hessian matrix and use it to define the search direction.

- Gauss-Newton method. It is of the same family of Newton's methods, but this algorithm does not require the computation of second derivatives. The Hessian is approximated by ignoring derivatives higher than first order.

- Levenberg-Marquardt algorithm. It is a method related to the previous one, where the search direction in this case is given by the inverse Hessian matrix and a weighting factor that regulates the performance of the optimizer with respect to its speed and stability. By decreasing its value, greater speed may be achieved. At the limit, when it equals to zero, the algorithm falls to the previous one. On the contrary, when its value increases, the stability increases as well.

Hierarchical strategy

In order to speed up the convergence of the iterative process or to increase the registration accuracy (i.e. avoid local minima), the optimization can be performed from coarse-to-fine scales (hierarchical strategy) (Lester et al. 1999, Bajcsy et al. 1989). At each scale, the registration method is iteratively solved until convergence. The transformation function that registers the images at a lower scale, will only register the images approximately at a higher scale. Therefore, after registering the images at a scale, the process should be repeated to find more correspondences in the images at a higher scale and compute a more elaborate transformation that can compensate more accurately for the local geometric difference between the images. The idea of the hierarchical approach is to register to coarse (low-resolution) image first and then to use the result as the starting point for finer (high-resolution) image registration, and so on. Hence, the process starts with a coarse estimation of the displacement field, that is improved by using finer resolutions. As the optimization starts, at each scale, from an initialization resulting from the previous scale, this makes the method more robust against local minima. In fact, performing the registration on sub-sampled images not only reduces the computational time (because at the finest image resolutions fewer steps are needed for converging to the optimal solution), but also allows the estimation of large displacements that could not be estimated in the final resolution because of noise and local minima in the optimization process. As the scale of the images is reduced, the geometric difference between the images reduces, making the images more similar and increasing the registration accuracy.

The coarse-to-fine approach can be applied to the image during the registra-

tion method, and even to the patch for the definition of the feature descriptor used for the similarity measure (i.e. coarse-to-fine hierarchical matching criterion), as well as to the grid sampling of the displacement field.

3.4 Validation criteria for a deformable registration

Validation of the registration result is an important part of the image registration procedure. It consists of evaluating both the performance of the registration algorithm and the accuracy of the estimated deformation model. However, the measurement of the registration accuracy is still an ongoing problem for deformable methods. It is a critical task because the ground truth is generally not available (Guo et al. 2006, Sotiras et al. 2013). Registration methods are often validated using external markers, anatomical landmarks, or external fiducial frames as the gold standards (Woods 2000). For landmark approaches an error that expresses the distance between corresponding landmarks post-registration can be computed. However, these landmarks are not always available, and even more, although the error at landmarks can be established, landmarks are not adequate to validate a deformable registration model that varies locally. Therefore the registration error cannot be quantified for the whole image with certainty, as well as the pointwise estimation cannot accurately refer to the whole image. Similarly, as already explained above, a good match obtained with an intensity-based approach cannot guarantee a successful registration result neither give a measure of its accuracy, as a transformation model that gives a good similarity match may be physically not plausible and as many solutions can lead to similar image matching (ill-posedness of the registration problem). Nevertheless, a pointwise accuracy assessment over the whole image, even if possible, is meaningless, or in many applications the true point-to-point correspondence may not even exist (e.g. intersubject registration). These criteria have to be adapted to each specific context.

In the recent years, many efforts have been made on the evaluation of the registration in different contexts (e.g. in (Maintz et al. 1998, Christensen et al. 2003, Urschler et al. 2007, Janssens et al. 2009, Klein et al. 2009, Brock et al. 2010, Rodriguez-Vila et al. 2010). At the moment, no generally applicable method for registration accuracy assessment is available. Anyway, a general method suitable for all applications is not reachable. These validation criteria have to be defined and adapted to each specific context. For instance, the requirements for the registration computing time are different for applications as intraoperative registration tools than in the context of off-line image processing. As well as an estimate of the registration error evaluated satisfactory for a specific task, on the other hand could be considered unacceptable for

another application that requires a high level of accuracy. Therefore, validation usually means showing that a registration algorithm applied to typical data in a given application consistently succeeds with a maximum (or average) registration error acceptable for the application.

Currently, for specific applications, the increasing availability of annotated data sets (e.g. accessible medical image databases and online medical atlases) has made possible evaluation studies. Moreover, the development of evaluation projects for image registration (e.g. Non-rigid Image Registration Evaluation Project – NIREP (Christensen, Geng, et al. 2006) and the increasing understanding regarding the use of surrogate measures for the measurement of the accuracy of registration (Rohlfing 2012) will further facilitate the comparison between different algorithms.

Performance evaluation of deformable registration algorithms

Common criteria for the evaluation of image registration performance are accuracy, reliability, robustness, and computing time.

- Accuracy

Accuracy refers to the difference between true and estimated values. In image registration, this difference can be evaluated as the mean, median, maximum, or root-mean-squared distance between points in the reference image and corresponding points in the template image after they have been resampled to the space of the reference image (target registration error, TRE (Maurer et al. 1997)).
For a given displacement field $\mathbf{u} = (u, v, w)$ and a pair of controls point $(\mathbf{p}, \mathbf{p'})$, where $\mathbf{p'} = (x', y', z')$ is the transformed location of a voxel $\mathbf{p} = (x, y, z)$, the TRE is defined by:

$$TRE = \sqrt{(x + u(\mathbf{p}) - x')^2 + (y + v(\mathbf{p}) - y')^2 + (z + w(\mathbf{p}) - z')^2} \quad (3.14)$$

where the Euclidean distance serves as localization error. The smaller this difference, the more accurate the estimate will be. Accuracy is usually measured in pixels (or voxels for volumes).

- Robustness and reliability

Reliability refers to the number of times an algorithm succeeds in finding a satisfactory answer compared to the total number of tests performed. Robustness refers to the reliability of an algorithm under variations in one or more of its input parameters. Therefore, a registration method is characterized as robust, when its performance does not drastically degrade for small deviations of the input images from the nominal assumptions. For example, the presence of a small fraction of artifacts or outliers results in small changes in

the result. Robustness is, for example, important when encountering images of pathology, i.e. images characterized by the presence of tumors that can be regarded as outliers.

- Computing time

The computing time required by the registration is one of the main concern of the users, even if it depends on the specific application. Computational complexity determines the speed of an algorithm and shows the practicality of the algorithm in real situations. Also pre-processing computing time has to be considered, i.e. the time required to prepare all data needed for the registration. In some cases, pre-process includes human manipulations (e.g. manual delineations) which are usually much more critical than computer processing time.

As already explained, another indicator of registration consistency is the symmetry of the resulting transformations regarding the order of the inputs (Christensen et al. 2003).

Validation criteria

The validation of the registration accuracy is particularly important. Different methods to measure the accuracy can be used:

- landmark correspondences: one common approach is to identify corresponding landmarks or features (i.e. anatomical landmarks or control points landmarks) independently on the reference and template images and evaluate the TRE. The landmarks can be manually or automatically identified. Alternatively, fiducial markers may be placed in the scene and the locations of the fiducials may be used to evaluate the registration accuracy.

- using synthetic or simulation images where the coordinates of true correspondences are known.

- phantom studies, where virtual and physical phantoms provide the gold standards for validation.

- subtraction of images: it is a straightforward evaluation method, commonly used in intra-modality registration. It consists in performing subtraction pre- and post-registration images (i.e. between reference and transformed template) together with quantitative similarity-measures, such as SSD and correlation coefficient.

- voxel-wise comparison between displacement fields: it can be achieved using different kinds of field metrics (e.g. mean square of the difference between fields, median absolute deviation, maximum deviation, etc. (Urschler et al. 2007)).

- visual inspection (carried out by physicians): in most applications, careful visual inspection remains the first and most important validation check.

Chapter 4

Experimental acquisition system setup

Contents

This section refers to a prototype system built for the acquisition of 3D US test images of phantoms using a conventional 2D US transducer. It basically comprises a support for the probe and an application to drive the motion of the US transducer and to control the image capture.

4.1 System components

A laboratory setup was developed to support the analysis of the implemented tridimensional registration techniques (described in the next chapter). It comprises a prototypal US acquisition system and a number of ad-hoc designed phantoms. Its primary target is to test elementary software components (for example the expressiveness of a given feature descriptor) in a controlled environment. A secondary goal is to suggest a way, at least for what concerns the US modality, to generate the required volumetric data from bidimensional US probes, exploring the possibility to apply the developed registration techniques also when only a bidimensional probe is available.

System description
The mechanical support for the probe was designed in collaboration with the mechanics workshop of the Department of Physics at the University of Trieste; the automation required for the probe movement was produced with

the assistance of the workshop of electronics. The physical support has been tested with the following Ultrasound devices: Ultrasonix SONIC series with linear transducerL14- 5W/60; Philips iU22 model with linear transducers L 17-5 and L 12-5.

The apparatus is composed by the following physical components:

- US two-dimensional linear probe;

- a support used to place the probe over the breast of the patient;

- a track system which constrains the movement of the probe along the horizontal plane and allows the whole organ scan: the displacement along the movement direction has a millimetric precision and guarantees a constant-step image acquisition;

- an acoustically transparent membrane placed over or inside the support; the membrane is elastic enough to fit the shape of the breast; during the acquisition process it is filled with water in order to let the probe be immersed in a water bath; this enables the transmission of the ultrasonic signal without the need of a direct contact between the probe and the breast.

Figure 4.1(a) show how an US probe is fastened to the support by means of the vertical arm; the probe acquires frames in the xy plane and moves along the z-axis; the x-motion is controlled by the stepper motor visible on the bottom-left corner of the picture. In figure 4.1(b) the support is seen from above: the track system constrains the arm and let it move only along the x-direction but can be repositioned along the y-direction. In this configuration the membrane is fixed to the bottom of a container and filled with water; a CIRS 051 Triple Modality Breast Biopsy Training Phantom is protected by the elastic membrane and it is about to be scanned. This system allows to avoid the relevant deformations impressed during the scanning performed with ABVS probe (as explained in the second chapter).

To make the overall system work it was also necessary to implement:

- the automated movement of the physical support through the small unipolar stepper motor driven by a programmable microcontroller board (Arduino Uno model);

- a software module that synchronizes the movement system with the proprietary capture software associated with the different manifacturer's equipments; to make it more flexible a programmable mouse simulation tool, able to mimic the actions required for capturing an image, was also provided.
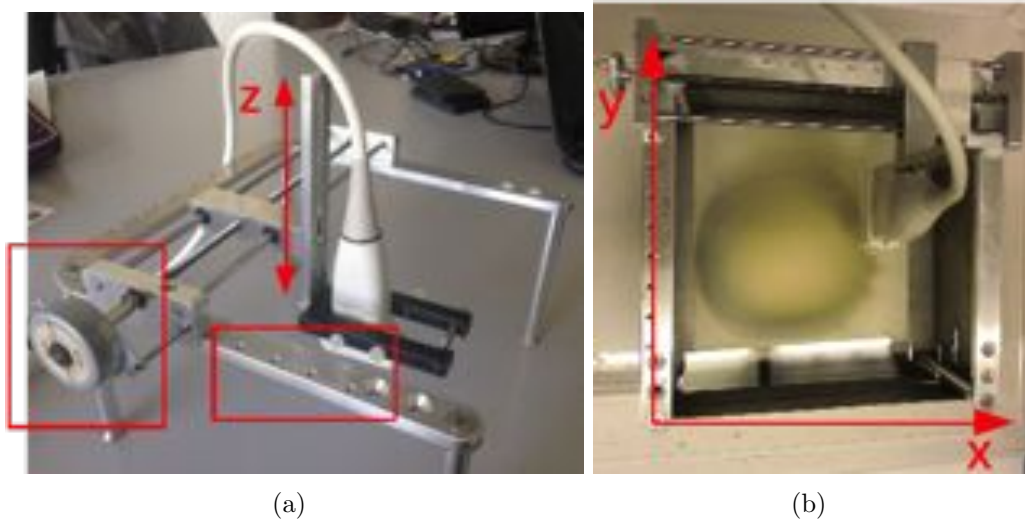
Figure 4.1: pictures of the designed support for US probe. In (a) the red boxes highlight the stepper motor used to move the probe along the x-axis (shown in picture (b)), and the equally spaced holes (at intervals of 2 cm) for the acquisition at different positions along the y-axis (shown in picture (b)). The depth of the probe can be manually set in order to place the probe closer to the object to be scanned, avoiding the contact between probe and object to guarantee the probe's motion.

Figure 4.2(a) shows the connections between a conventional US system with the other components of the experimental prototype for 3D US acquisition. The black box on the bottom contains the electronic part of the prototype (figure 4.2(b)): the programmable microcontroller board controls the stepper motor of the support and handles the communication with the synchronization software module (installed on the conventional US system) via USB cable. The interactions between the different parts of the system are represented in the sequence diagram in figure 4.3: when the synchronization module program starts, it sends a reset command to the microcontroller; the board reacts activating the stepper motor and making the probe move to the x=0 position; then the synch program enters a cycle in which: the micro is told to move the probe to other known positions; once reached, the synch program asks to the acquisition software associated with the manifacturer's US system to acquire one or more frames. The complete path of the probe in the 3D space is depicted in figure 4.4(a). Thanks to the synchronization program each acquired frame is associated with a known position of the probe thus allowing the automated reconstruction of the whole breast volume from the complete series of saved images. It is worth noting that the probe moves backwards and forwards in space covering different sections of the breast and two consecutive sections have in common a large overlap area. In addition to this spatial redundancy the system allows also to obtain a time redundancy, acquiring
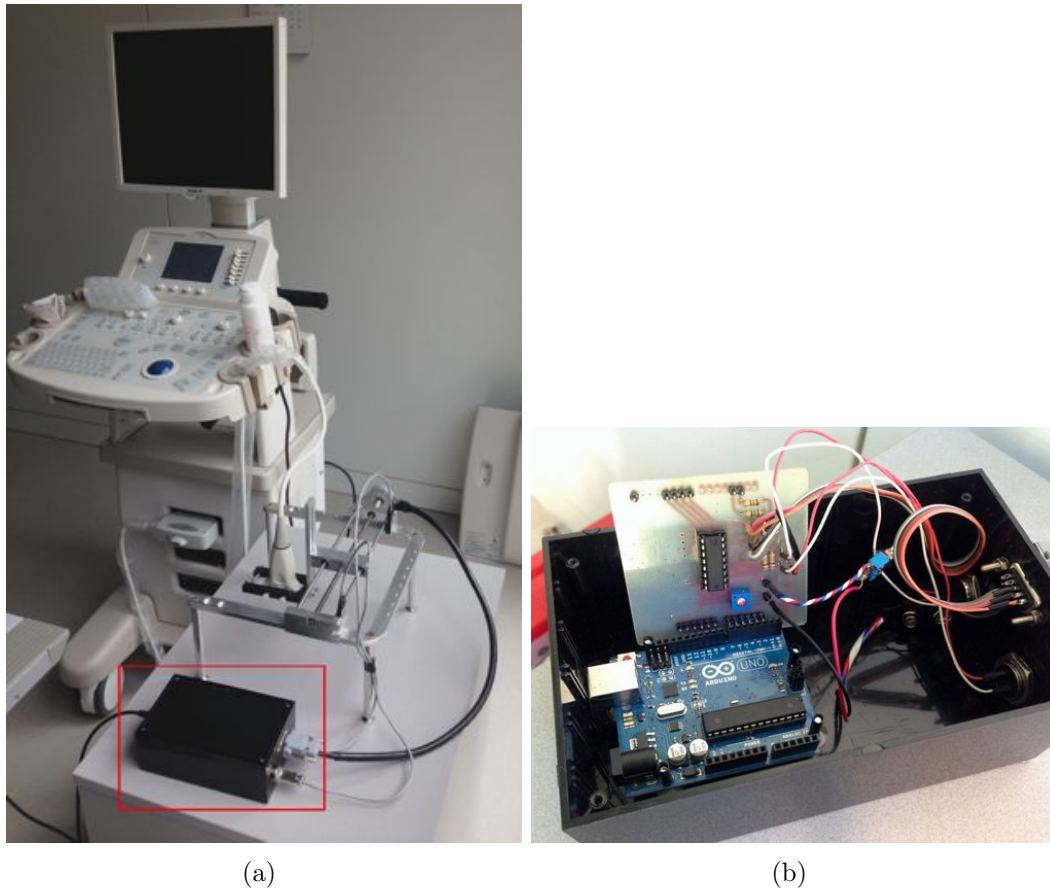
(a)   (b)

Figure 4.2: (a) picture of the whole system comprising the US apparatus and the designed prototype. Figure (b) shows a detail of Arduino board and the printed circuit board made to drive the stepper motor.

more than one frame in a fixed position as shown in figure 4.4(c). Both these redundancies turn out to be useful in the volume reconstruction process as explained in the following sections.

4.2 Stitching and Super-resolution

Lateral stitching

Although the probe is constrained to move along fixed directions during the acquisition process, small displacements with respect to its expected position can take place. Moreover the sensed organ or object can be subjected to a slow, hardly detectable, motion. These kind of events are generally negligible between two consecutive frames but can affect more seriously the space mapping of corresponding frames in two consecutive sections. To avoid the creation of gaps in the reconstructed 3D image data an additional check is required. It is assumed that a simple translation model can describe the
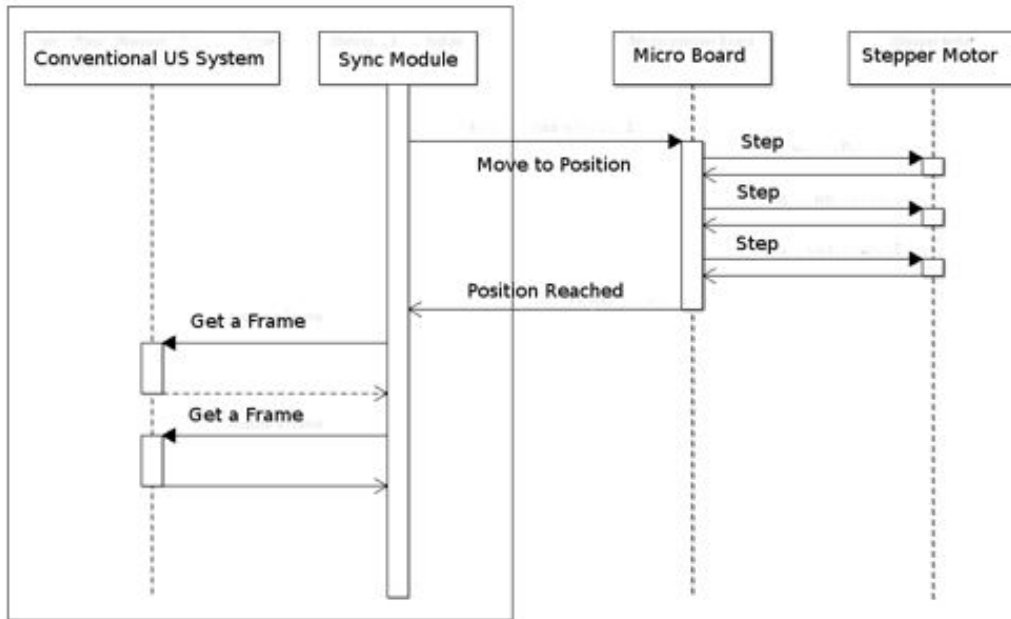
Figure 4.3: Sequence diagram of the basic interaction between the synchro-
nization module with the micro controller board that manages the stepper
motion (on the right) and with the conventional US acquisition system (on
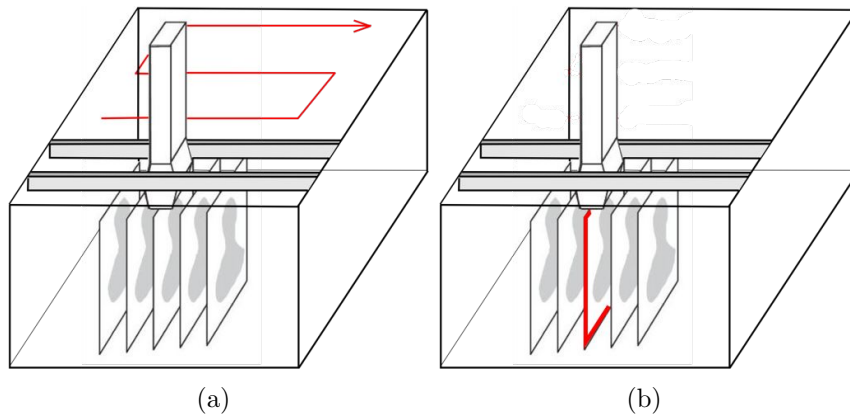the left) that saves a given number of frames in the same position.

Figure 4.4: Acquisition sequence: (a) path of the probe required to retrieve all
the slices that makes up the whole volume.; (b) sequence of frames acquired
in a single position to obtain the time redundancy.

displacement between two consecutive sections (figure 4.5). The expected dis-
placements (given by the synchronization module) are used to initialize these
models. For each couple of consecutive sections a Nelder-Mead (Nelder et al.
1965) optimizer is then run to evaluate the translation that results in the best
Normalized Cross Correlation (NCC) between their overlapping region.

Figure 4.5: Example of a possible small misalignment that will be compensated correlating the information in the overlapped regions.

The simple gradient-blending technique (Rancov et al. 2005) is finally used to create the composite volume without intensity discrepancies near the edges of each section.

Super-resolution
The quality of the acquired frames depends mainly on the characteristic of the available probe. The noise that affects the meaningful signal is a common degradation phenomenon in US images. Moreover the layout of the prototype system and the fact that the probe is not in direct contact with the breast can further reduce the image quality adding unwanted water reflections to the sensed signal. To partially reduce these artifacts and make the resulting image data suitable to be used in deformable registration processes, the available time redundancy is exploited.

To this aim an extension of the Non Local Means algorithm (NLM) has been used. The NLM algorithm was originally proposed for image denoising (Buades et al. 2005). It takes advantage of the high degree of information redundancy commonly embedded in a single image: if we consider an image I and we define for each pixel x of I a neighborhood $N(x, w)$, as the rectangular patch around x of dimensions $[2 \times w + 1, 2 \times w + 1]$, the intensity value in $\mathbf{x}$ is changed by the NLM algorithm in:

$$NLM(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \sum_{y \in I} I(\mathbf{y}) \cdot exp \frac{G_a * \|N(\mathbf{x}, cw) - N(\mathbf{y}, cw)\|}{h^2} \qquad (4.1)$$

where $\mathbf{y}$ is a generic pixel of I, G_a is the Gaussian function with sigma a, $\|N(\mathbf{x}, cw) - N(\mathbf{y}, cw)\|$ is the SSD value between the two patches around $\mathbf{x}$ and $\mathbf{y}$ of aperture cw (comparison window), h is a filtering constant, and $C(\mathbf{x})$ is the normalization factor:

$$C(\mathbf{x}) = \sum_{y \in I} I(\mathbf{y}) exp \frac{G_a * \|N(\mathbf{x}, cw) - N(\mathbf{y}, cw)\|}{h^2}. \qquad (4.2)$$

Defining the similarity measure w between pixel $\mathbf{x}$ and pixel $\mathbf{y}$ as:

$$w(\mathbf{x}, \mathbf{y}) = exp \frac{G_a * \|N(\mathbf{x}, cw) - N(\mathbf{y}, cw)\|}{h^2}. \qquad (4.3)$$

The NLM formula reduces to:

$$NLM(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \sum_{y \in I} I(\mathbf{y}) \cdot w(\mathbf{x}, \mathbf{y}) \qquad (4.4)$$

with

$$C(\mathbf{x}) = \sum_{y \in N(\mathbf{x}, sw)} w(\mathbf{x}, \mathbf{y}). \qquad (4.5)$$

The net result is that the new value in each pixel $\mathbf{x}$ is the weighted average of all the original image pixels $\mathbf{y}$ whose Gaussian neighborhood is similar to $\mathbf{x}$, thus removing noisy pixels from the image. The same approach remains valid if the set of pixels $\mathbf{y}$ is restricted to a search window (sw) around $\mathbf{x}$:

$$NLM(\mathbf{x}) = \frac{1}{C(\mathbf{x})} \sum_{y \in N(\mathbf{x}, sw)} I(\mathbf{y}) \cdot w(\mathbf{x}, \mathbf{y}). \qquad (4.6)$$

This approximation is motivated by the fact that it is more likely to find similar structures near the pixel under analysis. This approach can easily be extended to the case where more than one image representing the same field of view is available: if a series of such images is at hand, the previous formula can simply be changed with a double summation over the whole set of images. In this case also small changes in the position of the sensed underlying structures (or of the sensor) are correctly handled, provided that the spatial misalignment is lower than the search window size. It was demonstrated (Protter et al. 2009) that, besides the deblurring effect, this extension is an optimal solution also for the super resolution problem: in fact if the image pixel spacing is reduced, the information retrieved from a single source image can only fill a part of the target high-resolution matrix (figure 4.6(a)); but the additional information retrieved from the set of quasi-aligned frames can be used to fill the remaining locations (figure 4.6(b)). Each pixel value of the original series of images contributes to the final solution in different locations following the same fuzzy logic of the NLM first formulation. In this case however there's also the possibility to extract a complete patch, called extraction window, around each source pixels y in the multi-frame search window, and let it fill the target matrix near the location x with the relative weight that represents the similarity between y and x.

The NLM algorithm can be applied directly to US data but a further customization has been implemented to improve the homogeneity of the result: the h filtering constant used in the patch similarity evaluation represents the standard deviation of a Gaussian function that normalizes the SSD computed between two patches: if this parameter is high the similarity weight don't change much whether the SSD between the patches is low (they are effectively similar) or not. The net result is a more blurred target image. On

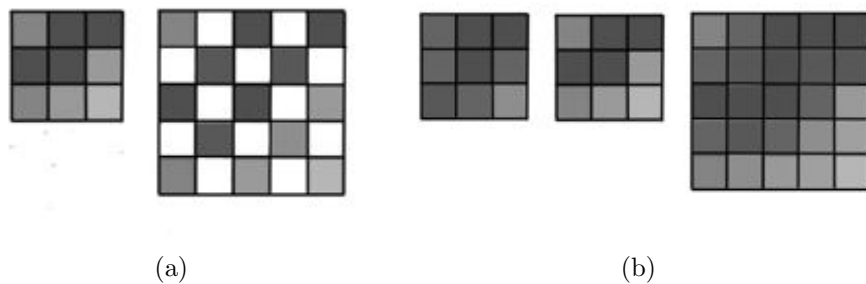<div align="center">(a)   (b)</div>

Figure 4.6: Comparison between a simple upscale operation and super resolution: (a) the information contained in a single source image does not change if the image is upscaled; it is possible to fill the blank positions in the grid but this operation does not provide additional information content. (b) if more source images are used, the high resolution grid can be filled also with the additional information content that derives from the image redundancy.

the contrary if the h parameter is low the result is sharper and more super-resolved, but the algorithm has reduced denoising effect. An average value for h is generally used to obtain a good compromise; but US data, contrary to common-life pictures, exhibit a high degree of variability in the local image contrast and makes the choice for the h parameter more difficult. The customization proposed makes the h parameter dependent on local contrast augmenting the denoising effect for background areas, but preserving the super resolution capability if high-contrast structures are detected.

The integration of this super resolution technique in the prototype system for 3D reconstruction is straightforward: when the probe is moved in a given position a series of frames are acquired sequentially in time and are used to build up a high-resolution frame. Then the stitching process described in the last section uses the high-resolution frames taking into account the associated rescaling factor.

4.3   Example of a breast phantom 3D reconstruction

An example of 3D reconstruction is given to show how the prototype acquisition system together with a conventional bi-dimensional US equipment can produce an US volume suitable to be used in deformable registration processes.

The CIRS Triple Modality Breast Biopsy Training Phantom (model 051, see figure 4.7(a)) is used as the sensed object in an acquisition session. The phantom is squeezable and contains contrasted structures similar to cysts and cancer masses of different dimensions. This is a compressible phantom that

has density and attenuation characteristics simulating an average 50 percent-age glandular breast (BR-12 equivalent) for X-ray, US and MRI. It consists of 6 cystlike and 6 dense masslike objects, where dense mass is 2-8 mm in diameter for core biopsy and cystic mass is 3-10 mm in diameter for needle aspiration. The acquisition process is configured to scan 5 sections of the
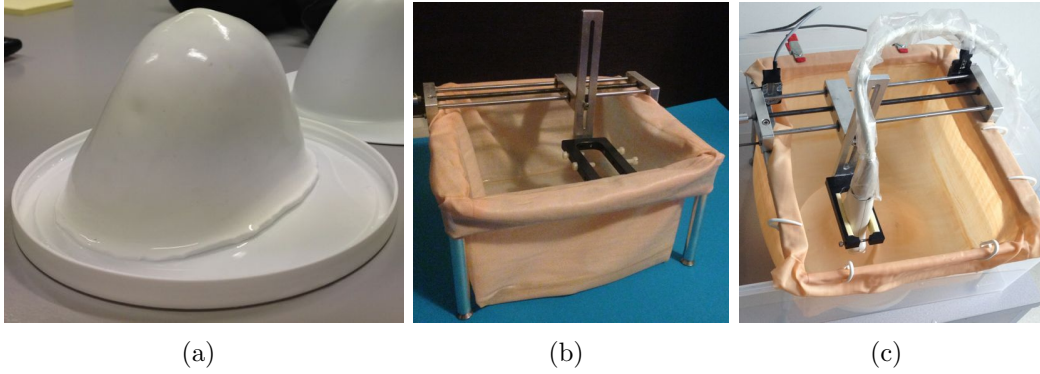

(a)   (b)   (c)

Figure 4.7: (a) picture of the CIRS breast phantom; (b) the home-made latex membrane to isolate the object to be scanned from the water. An example of its use in the US water-bath scanning is reported in (c): the phantom is covered by the membrane, that is then filled with water. During the scan the probe is dipped into the water.

object with an expected overlap of half the image width (figure 4.8). For
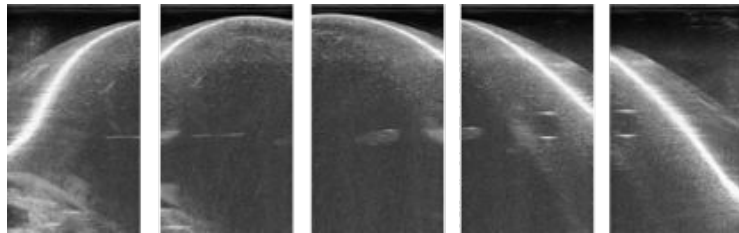
Figure 4.8: Example of a series of images corresponding to a phantom section as acquired at different positions along y-axis, but at the same x-coordinate (refer to the figure 4.1(b) and figure 4.4(a) for a description of the scanning procedure).

each position 9 frames are acquired and the super-resolution is set to produce high-resolution images with a $\times 2$ up-sampling factor.

The lateral stitching technique then blends the corresponding planes and finally creates the 3D volume (figure 4.9).
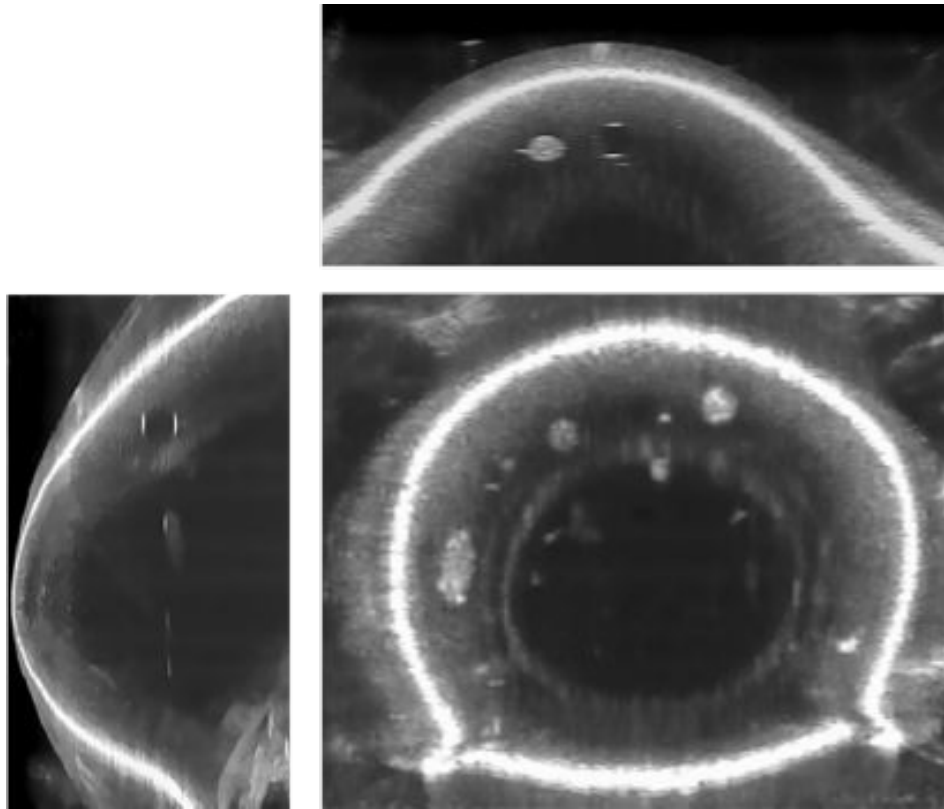
Figure 4.9: three planes of the reconstructed volume: reconstructed plane corresponding to a fixed y position (first row at the bottom), plane resulting from the lateral stitching procedure of the acquired images (secod row, on the left), and the top view corresponding to the coronal plane (second row, on the right).

4.4   Analysis in a controlled environment

With the prototype system a great number of reconstruction parameters can be controlled and changed to test how they can influence the different characterizations of data required in a deformable registration framework to map a structure in a volume to the corresponding structure in another volume. Ad-hoc phantoms were implemented to help this analysis: the underlying idea in their design is to isolate some basic features of the represented objects. In fact in a deformable registration task a lot of components can influence the overall result; reducing the number of such elements (including the artifacts due to the acquisition process, as noise) gives a great support to the study of the behavior of the features analyzed during a registration process. Moreover the prior knowledge about the nature, geometry and sensibility to deformation of these isolated real features can be exploited to evaluate their potential as reliable reference points in the volume mapping. In the next section an example of analysis that takes advantage of ad-hoc designed phantoms and the con-

trolled environment constituted by the experimental setup is given. It briefly describes one of the tests performed before the integration of super-resolution into the prototype acquisition system to evaluate the feasibility and usefulness of such a tool. Although it is not explicitly focused on the evaluation of a deformable model, it indirectly demonstrates how a software tool applied to the raw analyzed data can help in the feature-feature mapping between two representations of the same object in different images enhancing the meaningful information and removing disturbing artifacts.

Super-resolution evaluation
The prototype system was used to collect a sequence of US images of ad-hoc phantoms made for testing the super-resolution algorithm performance. The phantoms have been prepared with agar, and different simple-shaped insertions have been included to image objects with different contrast. Figure 4.10(a) shows one of these phantoms, that consists of an agar cube with a regular pattern of circular holes. When it is located inside the water-filled box for US immersion scanning, all the holes are saturated with water. The coupling between agar and water produces low contrast US image, and these low contrast structures have been used to assess the image quality under realistic and not optimal conditions. Hence, the super-resolved images have been compared with the original acquisitions, in order to qualitatively and quantitatively measure the image quality improvement after the image processing. In figure 4.10 the original scanned image and the corresponding super-resolved image are depicted; the double-resolution image results from the processing of a series of 10 scanned images acquired at a fixed probe position. In figure 4.11 a detail of one of the water-filled circular patterns from the original resolution image series is compared with the same detail of the super-resolved image. A considerable reduction of noise and image definition can be noted in the processed image; this consideration is quantitatively supported by the signal to noise ratio (SNR) estimated both in water-based and agar-based areas. The results obtained in terms of SNR in two regions of interest (ROI), inside the detail (yellow selection) and the surrounding area (red selection), are reported in figure 4.11(c). An average improvement of about 20% has been obtained in the super-resolved image. A higher image quality is obviously a remarkable result, but in this context it is also important to evaluate how much this enhancement can be useful in a registration process. In fact a registration task is required to correctly evaluate (explicitly or implicitly) a great number of correspondences between physical structures imaged in two different acquisitions: even when an object does not change between an acquisition and the other, the imaged data could be very different due to noise, artifacts or deformations introduced by the acquisition process. In the artificial example in figure 4.12(a) the first row shows a circular object imaged at an ideal resolution and without noise, before and after a simple translation. The same object subjected to exactly the same movement is represented in the second row as

(a)   (b)

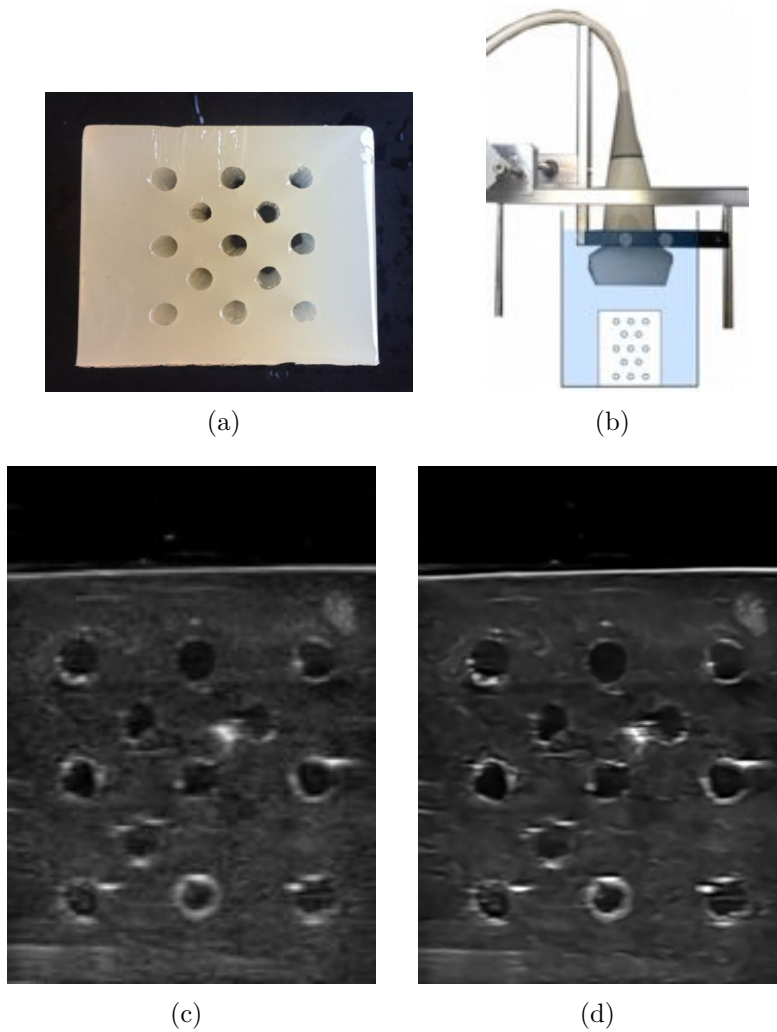(c)   (d)

Figure 4.10: (a) picture of an agar-based phantom with a regular pattern of holes. The acquisition is made using water as coupling medium (b); (c) example of one of the collected images of phantom; (d) super-resolved image resulting from a sequence of 10 different images acquired with the probe at a fixed position. c) and d) are here represented at the same spatial resolution for a more straightforward image quality comparison.

acquired by a low resolution, noisy sensor. It is clear that in the last case even the estimation of such a simple translational model is not trivial and can possibly lead to an inaccurate result. Viewed another way it can be said that the nearer the imaged data is to an ideal representation of the objects, the more likely the registration could lead to an accurate and reliable result. To analyze this subtle but important aspect the holes of the phantom have been modeled by an ideal circle and a simple contouring algorithm has been adopted to outline the borders of each insertion both in the raw low-resolution images and in their super-resolved counterpart. The contouring algorithm is based

(a)   (b)
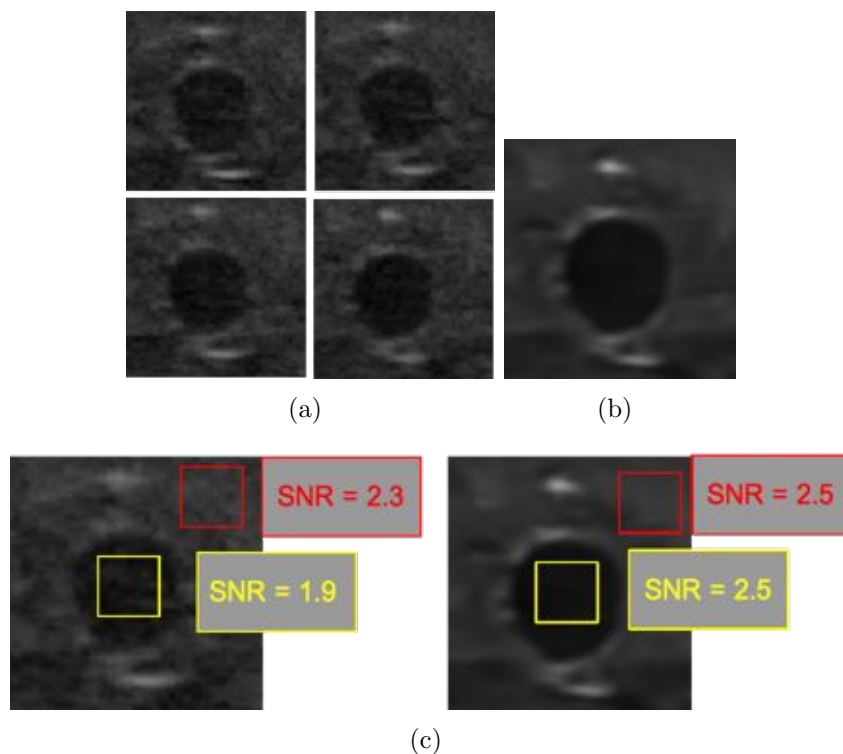

(c)

Figure 4.11: comparison between the original (a) and super-resolved image (b) of one of the inner circular details imaged inside the phantom. In (a) the detail is represented as appear in four images chosen among the collected series. (c): comparison between the original (left) and super-resolved image (right) of one of the inner circular details imaged inside the phantom. In both images the SNR values corresponding to the drawn ROI are reported. The yellow ROI refers to a region inside the detail and red ROI to a uniform region in the surrounding area. The values are estimated at the same scale of analysis (i.e. same image resolution).

on the well-known active-contour method and makes the border expand from the center of each insertion driven by gradients in image intensity. The algorithm is executed until convergence and once obtained a stable solution the points of the resulting contour are fitted to a circle. Hence, the mean distance between each point of the resulting contour and the corresponding reference point on the circle fit has been estimated. This measure reflects the achievable accuracy in the extraction of simple structures boundaries in the acquired US data. Figure 4.12(b) reports the distributions of these distances for both the low-resolution and the super-resolved images. The comparison highlight s the higher accuracy in the contour estimation using the super-resolved image (i.e. a lower mean error in the estimate of the contour points). For a more complete analysis, the same procedure is been repeated for an insert with a rectangular shape. The results have confirmed the same trend also in case of different
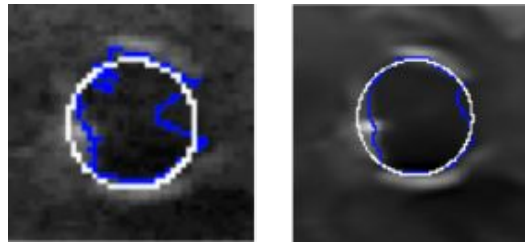
(a)
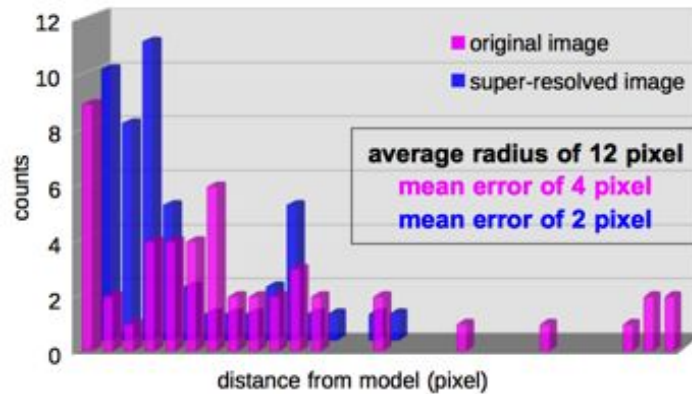
(b)

Figure 4.12: (a): the best fit model is delineated in white, while the blue contour represents the result of the automatic contour algorithm. The result obtains on the original image is shown on the left, while the more accurate contour obtained on the super-resolved image on the right. It can be noted as in the case of the original image the estimated contour is more affected by noise and deviates from the reference model. This result is confirmed by the average distance between the estimated points and those of best fit, as shown in the histogram in figure (b), where in pink the distances relating the original image are reported and in blue those relating the image at higher resolution.

shape of the objects. In summary, a home-made 3D US was developed for the acquisition of test images for the study of the image registration that is the goal of the work. For this development the super-resolution technique has been adapted and applied to images captured with the prototype, proving a powerful and promising technique for this application.

Chapter 5

Automated Multimodal Breast Registration Algorithms (AMBRA)

Contents

5.1 The registration framework

A complete set of tools to support testing, comparing and deriving automated multimodal breast registration algorithms (AMBRA) has been developed. This framework takes into account the requirements that arise from the before mentioned breast related tasks; its architectural design is general enough to handle all these use cases: for example it supports both single and cross modality registration of images acquired with the patient in similar or very different positions, complete or only partial overlapped images. On the other hand it implements a number of specialized techniques explicitly customized for breast; most of them are extensions of more classical and general tools; thus a modular design has been adopted to easy substitute the classical with the novel techniques. In addition it allows a user-independent comparison of different registration methods.

From an architectural point of view the framework is organized in two main

processing blocks (figure 5.1): the initialization and the optimization modules. Both of them define a precise software interface but allow a straightforward customization to address the specific problems and handle the specific images associated to a given clinical task.
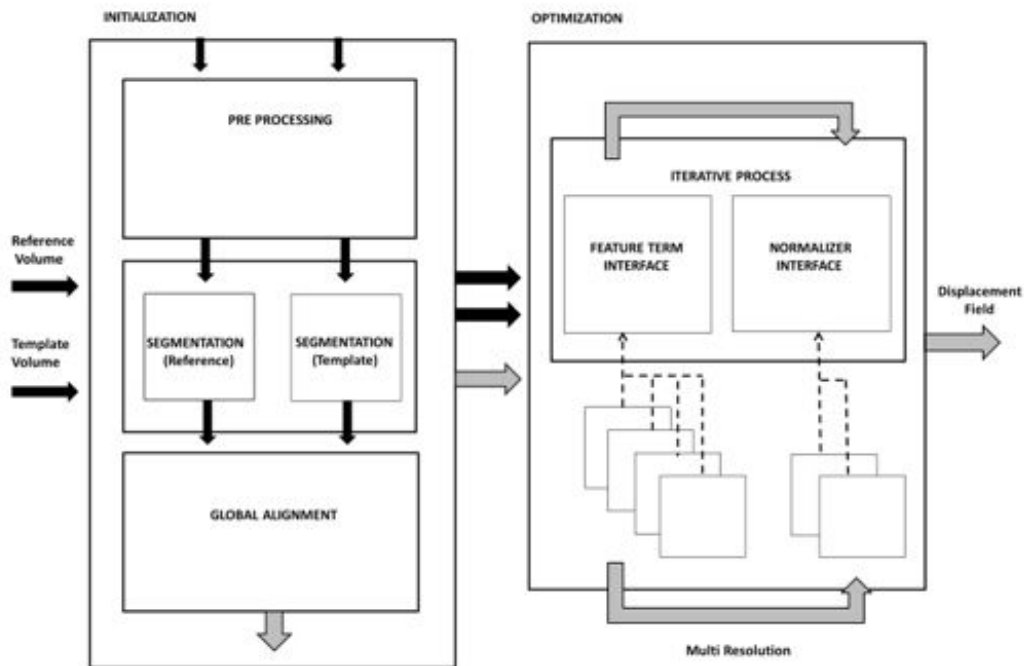

Figure 5.1: The framework's components diagram. The dark arrows represent how image data pass from one component to another. The bright arrows represent the displacement field that is first created by the global alignment module in the initialization block, and then passed to the optimizer that modifies it iteratively. The dotted arrows link different specialized instances of a given interface that can be selectively used during the optimization processp.

5.1.1 Initialization

The purpose of the initialization block is to convert the couple of input images from their original DICOM format (DICOM 1997) to a couple of volume matrices of, if possible, the same size and spatial resolution. A first estimate of the displacement field between the two images should also be provided: this field should represent at least the raw rigid global alignment of the two images and its evaluation can take advantage of prior knowledge about the specific analyzed input data: for example, given a specific clinical task, this knowledge can be represented by the patient expected position or by a reliable segmentation of organs (or other kind of landmarks) that are expected to be seen in the modality-dependent response provided by the acquisition processes asso-

ciated to that task. The initialization block is therefore further subdivided in three conceptually separated procedures: Preprocessing, Modality-dependent Segmentation and Raw Field Initialization or Global Alignment.

Preprocessing

Uniform Sampling

The preprocessing module is necessary to make the available input data conform to a common representation that can be easily handled by the framework. The aim of this block is to convert each volume instance from its original DICOM representation into a 3D matrix, hereafter named Volume, defined as follows:

```
typedef struct Volume
{
    int type;      /* datatype of the volume */
    int depth;     /* z-depth of the volume */
    int width;     /* width of the volume */
    int height;    /* height of the volume */
    void* data;    /* ptr to volume data buffer */
    Rect3D roi;    /* region of interest */
}Volume;
```

Contrary to the DICOM formats, this data structure has a common and uniform spatial sampling along the three orthogonal directions x, y, z in a Cartesian 3D space. The re-sampling operation makes use of trilinear interpolation to evaluate the value of the signal in the out-of-planes locations and takes into account possible rotations of the original slices with respect to the target coordinate system.

Segmentations

As stated before a disadvantage in breast imaging, especially in breast image registration, is the lack of rigid or semi-rigid, easy recognizable organs. Anyway, depending on the modality, the acquisition process and the signal nature, it is sometimes possible to identify one or more morphological structures or regions. For example the nipple location or the skin surface can be detected in quite all the modalities while the chest wall region is visible only in a limited subset of them. The framework segmentation module is a collection of predefined modality-dependent segmentation methods that can be used in the initialization process; in fact the automatically detected position of one or more structures in both the reference and the template volume can obviously help to define an alignment procedure regardless of the single or cross modal nature of the problem.

The outcome of each segmentation method is a 3D mesh defining the segmented object surface. The data structure used to represent the objects is a Vertex System (C. Smith 2006):

```
typedef struct VertexSystem
{
    int vertices_num;
    Vertex** vertices;
}VertexSystem;
```

with Vertex-Vertex connectivity:

```
typedef struct Vertex
{
  void* info;
  int neighboors_num;
  dmVertex** neighboors;
}Vertex;
```

where each Vertex is a 3D point of the surface.

The framework segmentation component allows also a straightforward way to integrate other segmentation procedures (for example for MRI) to the set of algorithms already developed (mainly for US and DBT): the only requirement for such additional methods is to conform to the framework interface and produce a Vertex System representing the surface of the segmented object.

Global Alignment

The global alignment is a user-defined, clinical task-dependent, procedure that produces a first raw displacement field between the reference and template volumes. The goal here is to handle the macroscopic differences in pose and location that affects the two volumes that are mainly due to their possibly different acquisition processes. For this reason it can take advantage of DICOM tags extracted from the original DICOM volumes (accessed by the pre-processing step) and automatically segmented objects (computed in the segmentation step) to estimate, at least, an acceptable global roto-translation from template and reference data.

Even though this rigid transformation can be represented by a limited set of parameters (for example a 3×3 or 4×4 transformation matrix), the framework requires a fully defined Displacement Field DF as the outcome of the initialization block. The field itself is represented by a triplet of Volumes (dX, dY and dZ), representing the displacements along x, y and z in each location of the domain of the reference Volume:

```
typedef struct DisplacementField
{
    Volume* dX;
    Volume* dY;
    Volume* dZ;
}DisplacementField;
```

Given a reference location $Vref = \{x, y, z\}$ the corresponding location in the template Volume can then be found in $Vtem = Vref + DF(Vref) =$

$\{x + dX(Vref), y + dY(Vref), z + dZ(Vref)\}$.

It is worth noting that this kind of output for the initialization block can be easily obtained once a rigid model is available but, at the same time, allows a more complex and localized definition of the deformation (an example can be seen in the use-cases in the last chapter).

5.1.2 Optimization

The output of the first block, the two input volumes data in the form of 3D uniformly-sampled matrices and the initialized field, is the input of the second one, which represents an iterative optimization process that has been designed according to the mentioned general scheme from the registration theory. The two parts of the functional to be minimized give rise to two independent software interfaces – one for the feature term and one for the normalizer. Different implementations for both of them have been developed. For what concerns the feature term each single implementation represents a specific feature metric and the associated distance; some of these metrics are modality independent to fulfill the cross-modality requirement of the whole framework. A similar approach is used for the second term but in this case the selection of the normalizers to be developed and inserted into the framework, followed a different motivation. It was in fact driven by the need of specialization for the breast tissue characteristics (elasticity) and the kind of movements and deformations expected inside the organ (sliding motion of sub structures). The overall optimization process has been arranged in a cyclic procedure that, at each iterative step, makes use of both the feature and the normalizer modules, sequentially updating the displacement field by small differential quantities. In addition to this basic approach other options were embedded in this part of the framework to support more advanced techniques: the adoption of the inverse consistent constraint is supported by an optional module that cooperates with the feature and normalizer terms; a multi-resolution optimization is implemented by re-executing the iterative process at different scales from coarse to fine and taking advantage of support tools to automatically rescale both the field and the volumes data.

In the following sections these concepts are explained in more detail starting from the feature and normalizer interfaces.

Feature Metrics Interface

The aim of this interface is to provide an abstraction, common layer for the feature term of the optimization functional (eq.3.11). Each implementation should provide a way to estimate how much a voxel in the reference Volume is similar to another voxel in the Template Volume taking into account only the

data (modality signal response) in those locations or in their neighborhood. Two main concepts are associated with the idea of a similarity measure:

- **Feature Descriptor**: each location can be characterized by a synthetic description of its underling data. This description is organized in an ordered set of real scalar values named feature vector or feature descriptor:

$$FV = \{f_i\} \ \ with \ \ f_i \in R, \ i \in 0, L \tag{5.1}$$

where L is the descriptor length. In most cases this array is computed retrieving the information from the grayscale values of voxels in a neighborhood of (a patch around) the voxel under analysis. The most important properties associated to a specific kind of feature descriptor are its distinctive power and its degree of invariance. The distinctive power measures its capability to discriminate among similar structures, while its degree of invariance is related to the fact that the provided description remains distinctive even if certain properties of the image data (global or local brightness or contrast, ...) or of the represented structures (scale, rotation, ...) change and thus cannot be assumed constant. For example the most simple, though in most cases useless, feature descriptor is given by the one dimensional feature vector whose unique entry is the original grayscale value of the associated voxel; clearly such a descriptor cannot even distinguish bright noise from a location belonging to a bright region; instead a feature descriptor that is made up by the compound, lexicographically-ordered grayscale values of a squared patch around the target voxel is more able to detect similar structures avoiding the extreme sensitivity to local noise. Anyway the enhanced descriptive power of the second descriptor comes at a cost: voxels that belong to the same structure in the two volumes can be evaluated as very dissimilar if the structure itself is rotated or if the signal in the second acquisition is globally more bright or dark. Other kinds of descriptors can handle these situations thus increasing the degree of invariance. The best compromise between the two properties should be generally found for each specific problem.

- **Distance Metric**: the metric used to evaluate the similarity measure between two locations is bound to the feature descriptor. The type of feature descriptor in fact defines a L-dimensional feature space over which is possible to define different kinds of distances:

$$Dist = R^L \times R^L \to R. \tag{5.2}$$

It is important to note that the distance function is intended as a signed distance and can have also negative values. Generally the distance metric derives directly from the feature descriptor nature but in some cases more than one metric can be associated to the same descriptor influencing the invariance degree or the general behavior of the evaluation.

It is worth noting that the feature descriptors are characterizations of single locations and depend only on the data around those locations. So they can be computed before starting the optimization process for each single voxel of the two input volumes independently.

With these considerations in mind the framework feature vector interface was defined as follows:

```
typedef struct FeatureDetector3DInterface
{
    int (* Detector3DCreate)(Volume* vol,void* static_parameters);
    void (* Detector3DCompute)(Volume* vol,void* dynamic_parameters);
    int (* Detector3DFeatureLenght)();
    char* (* Detector3DGetDescriptorPtr)(Voxel* vox);
    double (* Detector3DGetDescriptorDistance)(char* desc1,char* desc2, int l);
}FeatureDetector3DInterface;
```

where:

- The *Detector3DCreate* method allocates the concrete instance of the specific Feature Detector.

- The *Detector3DCompute* method computes all the feature descriptors associated to every location of the input Volume according to the selected concrete instance logic. It should be called for both the reference and the template volumes before starting the optimization process.

- The *Detector3DGetDescriptorPtr* retrieves at iteration time the description associated with the given Voxel location.

- The *Detecttor3DGetDescriptorDistance* computes at iteration time the similarity distance between two descriptors previously retrieved (one from the reference and the other from the template volume).

It has to be noted that each kind of Feature Descriptor is generally associated to a set of specific parameters: these parameters are classified as static if they can influence the size of the descriptor (and hence its allocation), or as dynamic, if they influence only the values of the computed feature vector entries.

The framework includes also a set of concrete implementations of the Feature Metric Interface that can be conceptually divided in two subsets: modality-dependent and modality-independent. In the following sections the two different sets of Feature Metrics implemented inside the Framework are described in detail focusing the attention on a comparative analysis of their distinctive power in breast images.

Normalizers Interface

A similar polymorphic approach was used for the second term of the optimization functional: also in this case an abstract interface defines the set of

operations that each normalizer is required to implement. In this case each specific instance should provide a way to preserve certain properties of the Displacement Field (for example smoothness), binding it to an acceptable and realistic solution. This means that, contrary to the feature term, the normalizer takes into account the 'shape' of the field and not (only) the two volumes data. In an iterative process this concept can be reformulated in the following terms: at each iteration the normalizer should know how to change the Displacement Vectors towards new values that are more compatible with the desired śhapeór geometrical property of the entire field; this evaluation should be based only on local properties.

Therefore, the formal definition of the normalizer interface is the following:

```
typedef struct FieldNormalizer3DInterface
{
    int (* Normalizer3DCreate)(Volume* ref, void* static_parameters);
    void (* Normalizer3DSetInfluenceArea)(int l);
    void (* Normalizer3DCompute)(Volume* ref,
        DisplacementField* field, void* dynamic_parameters);
    void (* Normalizer3DGetNormalizedField)(Voxel* vox, float* dx, float* dy,
        float* dz);
} FieldNormalizer3DInterface;
```

where:

- The *Normalizer3DCreate* method allocates the concrete instance of the specific Normalizer.

- The *Normalizer3DSetInfluenceArea* method indicates how much local the estimation of the desired field properties should be.

- The *Normalizer3DCompute* method computes the new estimate of the Displacement Field according to the rules that should preserve the expected geometrical field properties.

- The *Normalizer3DGetNormalizedField* retrieves the new local estimate of the field in a given Voxel location.

It has to be noted that also the original Volume data (from the reference volume) is passed as argument to the Normalizer3DCompute method, in this case it could be used only as a ćontextínformation for the regularization logic: in other words the field-reshaping rules of the normalizer can be expressed also in terms of some structural properties of the underling volume data. For example the normalizer can be designed to preserve the smoothness of the field only in sufficiently bright regions, or to apply a piecewise smooth constraint, using the original modality data to detect the regions over which the smoothing effect should be maximized. This idea will be explained in more detail in the next section where the concrete instances of normalizers already implemented in the Framework will be described and where this specific feature of the normalizer interface will be used to handle the breast sliding motion geometric representation problem.

Iterative Process

With these tools in hand the optimization process consists in iteratively updating the Displacement Field starting from its configuration after the Initialization process. The selected Feature Metric and Normalizer instances act in turn to change the field in every location by small differential quantities. The current Displacement Vector $DF = \{dX, dY, dZ\}$ in a reference location $Vref = \{x, y, z\}$ determines its temporary association with the template location $Vtem = \{x + dX(Vref), y + dY(Vref), z + dZ(Vref)\}$. The feature term component searches in a neighborhood of Vtem (within the template Volume) a new location that is more similar to the reference location Vref according to the implemented feature metric distance Dist, and moves the vector towards that local optimum. To do this, the feature descriptor FV of every template location in the searched area is retrieved and compared with the Feature Descriptor FV associated to the reference location. The temporary new configuration of the Displacement Field is then changed by the Normalization module that tries to re-establish, for every location, the geometrical properties desired for the field. The Normalization module updates locally each Displacement Vector towards a new value that is more compatible with the regularization rules. Both the updates due to the Feature Metric and the Normalizer terms are influenced by an associated weight parameter in a range of 0 to 1: 0 means that the Displacement Vectors are left unchanged; 1 that the Displacement Vectors are moved to their optimal value (estimated locally in the current iteration); other values of the weight parameters mean that the Displacement Vectors are moved to intermediate positions. These parameters represent also the strength of their associated term in the overall process. An additional parameter, named search aperture, is shared by the two terms: it controls the search region for the Feature term and the influence area for the Normalizer term.

```
typedef struct DisplacementFieldUpdateParams
{
    int search_aperture;
    float feature_weight;
    float normalization_weight;
    float inverse_consistent_weight;
    float constrains_weight;
}DisplacementFieldUpdateParams;
```

The optimization process repeats these incremental updates until the field is sufficiently stable (no feature vector changes for more than a given small quantity) or a maximum number of iterations is reached.
An Optimizer is then represented by the following structure:

```
typedef struct Optimizer
{
    DisplacementField* field;
    Volume* reference;
    Volume* update;
```

```
    FeatureDetector3DInterface reference_features;
    FeatureDetector3DInterface update_features;
    FieldNormalizer3DInterface normalizer;
}Optimizer;
```

that is composed by the Reference and Template Volumes, the Displacement
Field, two instances (of the same kind) of the FeatureMetric interface (to
maintain the pre-computed feature vectors of both the reference and template
volume) and an instance of the Normalizer interface. The elementary update
operations are performed by the following methods:

```
DisplacementFieldUpdateWithFeaturesTerm(
Optimizer* opt,
DisplacementFieldUpdateParams* pars
);

DisplacementFieldNormalize(
Optimizer* opt,
DisplacementFieldUpdateParams* pars
);
```

Extensions

Besides the basic optimization process the Framework allows to enable some
extended features.

Multi-resolution Optimization. The Feature Descriptors that drive the
optimization process represent a description of the neighborhood of each lo-
cation. Though the size of the neighborhood patch can be customized it
remains fixed during the optimization process. This fact implies that the Fea-
ture term is more sensible to data structures of a given size. On the other
hand also the Normalizer indirectly depends on the resolution of the input
images, in fact the discrete approximations of the field derivatives, required
to correctly apply the regularization rules, are computed within a fixed-size
window expressed in the same coordinate-system of the reference volume. To
make the optimization process able to capture data structures of different
sizes and to conform to normalization rules defined at different wavelengths a
multi-resolution approach can be used: the Displacement Field and the Vol-
umes can be automatically rescaled to different resolutions. It can be noted
that, for what concerns the Field, rescaling affects also the values of its vectors
as they represent displacements in the reference space. The Framework sup-
ports a sequential repetition of the optimization process over different scales
and manages automatically the required transformations of the volume data
and the field. For each scale a different set of iterations parameters and dif-
ferent choices for the Feature and the Normalizer modules can be assigned. A
coarse-to-fine analysis is performed, ending up with a full-resolution process
in the last repetition.

Inverse consistent constraint. The domain of the Displacement Field DF is the same of the reference Volume as it should map each reference location to the corresponding template location. This means that the inverse field IDF mapping the template locations to the reference space is not always derivable: the function is not a one-to-one correspondence , especially after the optimization process. The only way to compute this inverse mapping is to swap the reference and the template volumes and to repeat the whole process for a displacement field defined over the template space. If this approach is followed without constraints the two resulting fields are not one the reverse of the other as both can converge to different local minima. A well known solution is to embed the so-called inverse consistence constraint (Y. Chen et al. 2010, Heinrich, Jenkinson, Papiez, et al. 2013) in an optimization functional that takes into account both the DF and IDF. It represents a force that tends to preserve the one-to-one correspondence of the displacements vectors in such a way that, given a location Vref in the reference space the corresponding location $Vtem = Vref + DF(Vref)$ in the template space obtained applying the direct DF, is mapped back by the IDF in the original location. This term contributes, together with the feature and the normalization terms, in the simultaneous optimization of the two fields leading to a more stable result.

The Framework supports this option making available the method:

```
DisplacementFieldApplyInverseConsistenceContraint(
DisplacementField* field,
DisplacementField* field_inv,
DisplacementFieldUpdateParams* params);
```

where the weight parameter has the same characteristics and meaning as the weight parameters associated to the feature term and the normalizer. If this option is enabled two Optimizers (one for DF and one for IDF) are allocated with shared (but swapped) FeatureDetector3DInterface instances for the reference and the template volumes. Their different DisplacementFields are iteratively and simultaneously updated following, at each iteration, the sequence of actions described for the basic process, but adding the previous method that merges the current direct and inverse fields.

Constrained optimization. When the position of a structure is precisely known both in the reference and in the template image, this knowledge can be used to improve the outcome of the whole optimization process. The displacement vector in the reference position of the structure can be forced to point to the position of the same structure in the template; this prevents possible errors in that location. An indirect benefit of such a constraint is that also the displacement vectors in locations near the forced point are induced to remain in the correct region, reducing the possibility for them to be attracted and captured by wrong local minima.

The Framework allows to define a complete set of such point-to-point correspondences:

```
OptimizerSetConstraints (
Optimizer* opt ,
VoxelSet* reference_points ,
VoxelSet* template_points );
```

Anyway, rather than forcing explicitly the field in these locations, it only pushes them towards their designated values with a weight parameter that controls the strength of this force: the constrains are treated as if they were an additional term in the optimizer functional.

```
DisplacementFieldApplyConstraints (
Optimizer* opt ,
DisplacementFieldUpdateParams* pars );
```

This weighted approach turns out to be very useful when the known correspondences are reliable but not precise. The contoured regions or surfaces resulting from the segmentation block, for example, can be used to establish known and certain correspondences but in most cases they lack in point-to-point accuracy. The weighted approach allows to use these coarse correspondences but, at the same time, gives an opportunity to the other optimization terms to correct them.

5.1.3 Testing and comparison

The architectural design of the Framework allows the derivation of different algorithms and customizations of the optimization process. Some additional software components were added to support testing and comparison procedures:

- **Testing Sets**. For each clinical task a whole set of predefined couples of Volumes are defined. This addition is straightforward but at the same time is very useful to automate the statistical evaluation of a specific method for a given clinical task.

- **Control Points**. A set of manually delineated landmarks can be associated to each couple of Volumes of the Testing Set. It is constituted of a list of point-to-point correspondences used to verify the quality of the resulting Displacement Field.

- **Estimation Metric**. The Target Registration Error (TRE) of these control points is optionally computed after each single optimization process. The average TRE distance computed over a Testing Set can be used as a quantitative measure in the estimation of the method under evaluation.

The modular structure of the Framework allows two different levels of comparison and testing:

Module Comparison. The polymorphism introduced with the two main interfaces of the Optimizer allows to easily substitute a specific implementation of the Feature Metric (or of the Normalizer) with another one. This leads to the fact that, for example, the expressiveness of a Feature Descriptor in a given modality or clinical task can be easily compared with the expressiveness of others: once defined a Testing Set with its control points and enabled the TRE computation, the optimization process can be run over the whole collection of test cases obtaining the average TRE. The same process can be re-run after the substitution of the Feature Descriptor under analysis with other implementations. As the only changed framework component is the Feature Descriptor module, the average TRE can be considered an index of the descriptor distinctive power in the situation characterized by the Testing Set.

It has to be noted that, as each kind of Feature Descriptor (or Normalizer) has a specific set of associated parameters, such a comparison is meaningful only if the best parameter configuration (over the Testing Set) is used for all the metrics or normalizers to be compared. It is crucial in fact to avoid the under-estimation of a metric or of a normalizer due to a wrong parameters selection. To this aim the TRE computation is re-run a lot of times for each Feature Descriptor (or Normalizer), changing its parameters in a range of possible values, and selecting the configuration with the best performance.

Optimization Technique Comparison. The same approach extends to a complete optimization method. The Feature and Normalizer terms, the set of weight parameters, the single- or multi-resolution approach, the application of the different kinds of constraint are included in the definition of the method. The only fixed part of the Framework must be the displacement field initialization strategy. The clear separation of the initialization block from the optimization block allows to compare different optimization methods avoiding the influence of each method's sensitivity to the starting image data dis-alignment.

5.1.4 Algorithm derivation

A registration algorithm for a specific breast clinical task can be derived from the Framework following a number of adaptation steps. The customization required for the initialization block is very different from the customization applied to the optimization block so they are treated separately in the following paragraphs.

Customizing the Initialization

As previously described, the initialization block is the part of the framework that can take advantage of prior knowledge about the data acquisition process. As this part is more affected by customization, depending on the specific task. The global alignment method has been designed as a virtual, user-defined procedure. Thus each single clinical task requires an ad hoc algorithm. As the aim of this step is a coarse initialization of the field, in most cases the problem has a straightforward solution. Anyway for situations in which the global alignment is a challenging task, the availability of the predefined, modality-dependent segmentation tools of the framework can help to reach a reliable method. For example the nipple segmentation procedures, that have already been implemented for US and DBT, can help identifying a coarse translation or rigid transformation between the images: it is always possible in fact to align the detected nipple positions whether the modality of the images is the same or not.

Customizing the Optimization

s Contrary to the initialization block, the optimization process customization does not require any addition to the framework implementation, but only to evaluate the best modules to be used in a specific task. The evaluation includes the estimation of the best parameter configuration for each of these modules. Given a Testing Set with a set of manually delineated landmarks (as described in the Testing and Comparison section) an automatic procedure that compares the TRE average distance of every possible combination of the implemented feature terms and normalizers is available. For each feature descriptor and normalizer, the best specific parameters configurations are retrieve over the set of testing cases; then their TRE average distances are compared to find out the combination of terms that performs better in the case under analysis. It is clearly possible to repeat this procedure over different scales taking advantage of the fact that the scale is a parameter of the whole optimization block. The best configuration of parameters, Feature Metric, Normalizer and the optimization extensions to be applied, are then automatically derived from the data in the Testing Set.

5.2   Feature detectors and normalizers

In the next two sections the series of Feature Metrics and Normalizers already implemented as modules inside the AMBRA Framework will be listed and described. More attention is given to the more advanced instances that were developed to answer to specific breast registration problems, while the most common and well-known among them will only be indicated, briefly describing the way they are integrated in the framework.

The approach used to develop the advanced descriptors and normalizers followed a common work-flow: first their behavior was analyzed on US data acquired with the experimental setup, to take advantage of the control on the acquisition process and of the possibility to isolate some expected data features (using phantoms) in a context where only simplified transformation model are considered; then the same instances are tested and their performance is quantitatively compared using clinical testing sets in which a number of disturbing phenomena are always present. The comparative measures were taken both for US and DBT modality data.

Feature metrics

The feature metrics used as mono-modal reference are the well-known Sum of Absolute Differences (SAD) and Normalized Cross Correlation (NCC). Both are described in the Registration Theory chapter and do not require any additional theoretical consideration. From an implementation point of view both are designed so that a location is described with a feature vector that is essentially formed by the lexicographically ordered series of gray values in a 3D rectangular patch around the location itself. It is convenient to define this patch as a grid of neighbors defined by two parameters: the aperture - that is the number of samples to reach the end of the neighborhood in every direction - and a sampling unit – that is the number of voxels between two consecutive sampled neighbors. The feature vector for each location is computed when the volume data is available. On the other hand they differ in the used distance metric; contrary to the feature vector the distance is computed during the optimization process, as the association between two locations (and hence two feature vectors) is determined by the displacement field and changes dynamically. More in detail:

- **SAD**: to each location is associated a feature vector that is formed by the sequence of gray values extracted from a rectangular 3d patch around it. The original gray values in this case are left unchanged. The distance between two locations is simply the L1 distance between their associated feature vectors.

- **NCC**: a feature vector FV is associated to each location that is essentially the same feature vector computed by SAD but in this case each FV entry is subtracted by the average value computed over the patch itself. So the FV represents the zero-mean patch intensities. The distance between two locations is then computed as the inverse of the cross correlation between the two zero-mean feature vectors divided by the square root of the product of their auto correlations.

$$Dist(FV1, FV2) = -\frac{\sum_{i=0}^{L-1} FV1[i] \cdot FV2[i]}{\sqrt{\sum_{i=0}^{L-1}(FV1[i]^2) \cdot \sum_{i=0}^{L-1}(FV2[i]^2)}} \qquad (5.3)$$

To avoid the re-computation of the auto correlations and the square root at each distance evaluation during the iterative process the feature vector in a location is concatenated with its auto correlation root squared before starting the optimization:

$$\bar{FV}[i] = FV[i] \ \ for \ \ 0 < i < L \tag{5.4}$$

$$\bar{FV}[i] = \sqrt{\sum_{i=0}^{L-1}(FV[j])^2} \ \ for \ \ i = L \tag{5.5}$$

reducing the iteration-time distance computation to:

$$Dist(\bar{FV}1, \bar{FV}2) = -\frac{\sum_{i=0}^{L-1} \bar{FV}1[i] \cdot \bar{FV}2[i]}{\bar{FV}1[L] \cdot \bar{FV}2[L]} \tag{5.6}$$

Depending on the expected variation in local image intensity and contrast (see the third chapter) the SAD or NCC can be selected. These metrics are intrinsically dependent on the modality representation of the breast structures and can be used in all mono-modal tasks.

The reference feature metric used as reference for multimodal tasks is instead the Normalized Mutual Information (NMI) that comes from the information theory.

- **NMI**: also in this case the feature vector FV associated to a given location is the sequence of gray values extracted from a rectangular 3D patch around it. Similarly to NCC an additional entry in the last position of the feature vector is used to maintain a precomputed quantity that in this case is the marginal entropy H of FV.

$$\bar{FV}[i] = FV[i] \ \ for \ \ 0 < i < L \tag{5.7}$$

$$\bar{FV}[i] = H_{j\in 0...L-1}(FV[j]) \ \ for \ \ i = L \tag{5.8}$$

The distance between two locations is then computed as the so-called variation of information normalized in the range of 0 to 1: the joint entropy of the two feature vectors JH is calculated at iteration-time with the first L-values of each feature vectors and the distance is expressed by:

$$Dist(FV1, FV2) = 2 - \frac{H(\bar{FV}1[L] + H(\bar{FV}2[L]))}{JH_{j,k\in 0...L-1}(\bar{FV}2[j], \bar{FV}2[k])} \tag{5.9}$$

As described in the third chapter this metric has been commonly and successfully used in rigid registration tasks but do not scale well to

intensity-based full-field deformable registration models because of its statistical nature: the local metric evaluation required in the context of these registration methods does not allow, in most cases, to collect enough information inside the space-limited neighborhood of a single location to make the description distinctive; so the significant improvement in the degree of invariance comes at the cost of an augmented weakness in the descriptive power. This is a major problem especially for breast applications where also the number of clearly distinguishable structures and their distinctiveness is low.

An advanced feature metric recently introduced (Heinrich, Jenkinson, Bhushan, et al. 2012) to handle multi modal tasks is the Modality Independent Feature Descriptor (MIND). Contrary to MI the topological structure of the intensity pattern inside the patches to be compared is taken into account. Moreover the descriptor does not support unnecessary kinds of invariance apart from the modality independence, thus limiting the loss of descriptive power (for example it is not rotation invariant). A version of this descriptor proposed (Heinrich, Jenkinson, Papiez, et al. 2013) recently to handle noisy data (and hence particularly appropriate when at least one of the scans is an US volume) is the Self Similarity Context (SCC) descriptor. The last part of the section describes these two descriptors focusing the attention on the Self-Similarity concept.

Self-Similarity
The concept has firstly been introduced in the domain of image denoising with the before mentioned NLM algorithm (Buades et al. 2005): recalling the equations (eq.4.1) from the forth chapter on the experimental setup, the NLM algorithm obtains a noise-free pixel in a given location $\mathbf{x}$ computing the weighted average of all the original image pixels y whose Gaussian neighborhood is similar to $\mathbf{x}$; the weights $\mathbf{w}(\mathbf{x}, \mathbf{y})$ used for the averaging are based on the SSD between a patch which surrounds the pixel $\mathbf{x}$ and all other patches in the image. The basic NLM assumption is that every image has a high degree of self-similarity or, in other words, that a high number of pixels $\mathbf{y}$ are similar to $\mathbf{x}$ considering the patch-based similarity measure $\mathbf{w}(\mathbf{x}, \mathbf{y})$. This assumption is generally correct even if the search of such locations $\mathbf{y}$ is limited to a neighborhood of $\mathbf{x}$.

The basic idea of MIND and SSC is to describe a region surrounding a location $\mathbf{x}$ with the topologically ordered map of the similarities $\mathbf{w}(\mathbf{x}, \mathbf{y})$ in each $\mathbf{y}$ instead of the raw intensity values of $\mathbf{y}$. The modality independence of this kind of description assumes that even if an object is sensed by different sensors the similarity map associated to the same physical structure in the two images should be similar, regardless to the specific intensity response of the object to the two different sensors.

This concept can be explained also considering the visual example in figure 5.2. From CT image (figure 5.2(a)) and MRI (figure 5.2(b)) image the same location (green arrow) at the boundary between fat and non-fat tissue are selected. In figure 5.2(c) and 5.2(d) the region around this location is represented by the raw intensity values extracted from the original modality data; obviously the intensity pattern is very different. In figure 5.2(e) and 5.2(f) the self-similarity of the central location with its neighbors computed using the intra-modal $\mathbf{w}(\mathbf{x}, \mathbf{y})$ measure and organized in the small rectangular images is shown: irrespectively of the modality all pixels along the boundary share a similar patch-neighborhood and their similarity measure is high. Out of the boundary the similarity with the central pixel is low but also in this case this is true for both CT and MRI resulting in the same location description. Two parameters are related to and influence the computation of the self-similarity entries: the aperture of the patch (half-width in every direction) used to estimate the similarity between neighbors, and the Gaussian sigma of the weight function $\mathbf{w}(\mathbf{x}, \mathbf{y})$.

- MIND: a feature vector is associated to each location x that is formed by the ordered sequence $\mathbf{w}(\mathbf{x}, \mathbf{y})$ of patch-similarities with a set of locations y in the neighborhood of x. It is possible to define different schemes for the neighbors that contribute to the description. A basic layout used in the framework is represented in figure 5.3(a). AMBRA also allows the definition of other configurations as, for example, a complete grid of neighbors sampled at a constant distance from $\mathbf{x}$, making use of the same convenient definition of the 3D patch defined by the patch aperture and sampling previously considered for the other descriptors. The $\mathbf{w}(\mathbf{x}, \mathbf{y})$ entries for each location are computed before starting the iteration and yielded in the correspondent feature vector. The distance between two locations is then simple the L1 or L2 distance between their associated feature vectors. This is an additional advantage in full-field registration with respect, for example, to MI: it embeds the multi-modal invariance inside the feature descriptor so the metric evaluation at iteration-time is very fast.

- SSC: the underlying concepts and even the implementation details of the SSC descriptor are exactly identical to MIND. The only difference is the layout scheme for the similarity correspondences used in the description of a location $\mathbf{x}$. In this case $\mathbf{x}$ itself is never considered but the feature vector is formed by the similarity distances $\mathbf{w}(\mathbf{y}, \mathbf{z})$ between two different neighbors $\mathbf{y}$ and $\mathbf{z}$ of $\mathbf{x}$ as shown in figure 5.3(b).

The motivation for this layout is to avoid the sensitivity of the MIND description to the central location. This layout prevents in fact the misleading effect of potential noise in $\mathbf{x}$; it makes the feature vector describe the $\mathbf{x}$ location through its context in the underlying image. This characterization makes the
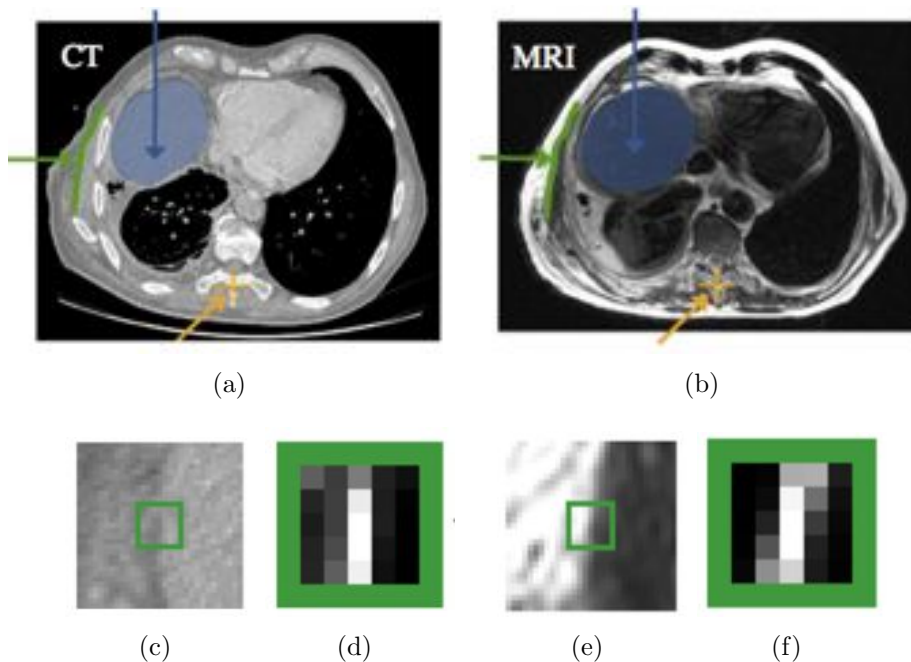
Figure 5.2: proposed concept for the use of MIND for multimodal registration (Heinrich, Jenkinson, Bhushan, et al. 2012). An example using a CT (a) and MRI (b) images is reported in figure. MIND is calculated in a dense manner in both the images. In (c) and (e) an exemplary location with a boundary feature between fat and non-fat tissue is shown. The corresponding descriptors (green colored boxes in (d) and (f), respectively for CT and MRI) are independent of the particular intensity distribution across two images and provide a very good representation of the local shape of an image feature. MIND descriptor is based on the assumption that even though the intensity distribution of an anatomical structure may not correspond across modalities, it is reliable within a local neighbourhood in the same image. In the example even the boundary has different intensity distribution across CT e MRI images, the corresponding descriptors (where high intensities correspond to small patch distances) are independent to the respective modality and can be compared using a simple intensity based metric.

descriptor particularly well-suited for US data. Also in this case the set of neighbors to be compared can be defined through the same grid-patch parameter used for all other descriptors: in this case the layout in figure 5.3(b) is expanded a number of times equal to the patch aperture and the space-multiplier is given by the patch sampling.

Measures

To evaluate the performance of the feature descriptors a comparative test was run on different testing sets. The rationale here was to compare all the de-

(a)   (b)
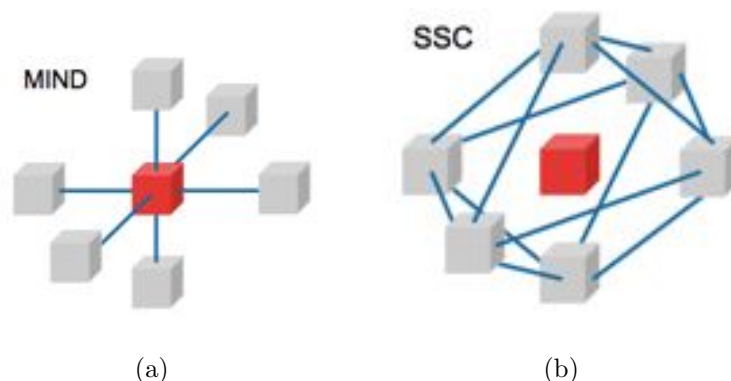
Figure 5.3: concept of self-similarity context (SSC, (b)) compared to MIND (a) with six-neighbourhood (6-NH, Heinrich, Jenkinson, Papiez, et al. 2013). The patch around the voxel of interest is shown in red, all patches within its immediate 6-NH are shown in grey. (a): all patch distances (shown with blue lines) used for MIND within 6-NH take the centre patch into account. (b): geometrical and structural context can be better described by SSC using all patch-to-patch distances, none of which is dependent on the central patch.

scriptors implemented in the framework in a common environment. To this aim the analysis was limited to US-US and DBT-DBT contexts, in which also the mono-modal descriptors can be used.

The first test was run over a testing set of 10 clinical breast US pairs.

Taking advantage of the framework structure a common global alignment procedure was used as the starting point of the deformable registration iteration. The best parameter configuration for each descriptor was first found, running the optimization process multiple times on the whole testing set as described in the previous chapter. The optimal results obtained for all descriptors were then compared, using the global alignment as basic reference (figure 5.4, table 5.1).

Parameter	SAD	NCC	MI	MIND	SSC
Scale	4	2	2	4	4
Search Aperture	6	3	3	3	3
Patch Aperture	3	1	6	2	2
Patch Sampling	2	1	1	1	2
Self Similarity Patch Aperture	-	-	-	1	1
Self Similarity Sigma	-	-	-	30	55

Table 5.1: estimated values corresponding to the optimal configurations for the feature descriptors in case of breast US data.

A similar procedure was adopted for a DBT-DBT testing set (figure 5.5, table 5.2). These results suggest that MIND performs generally better than
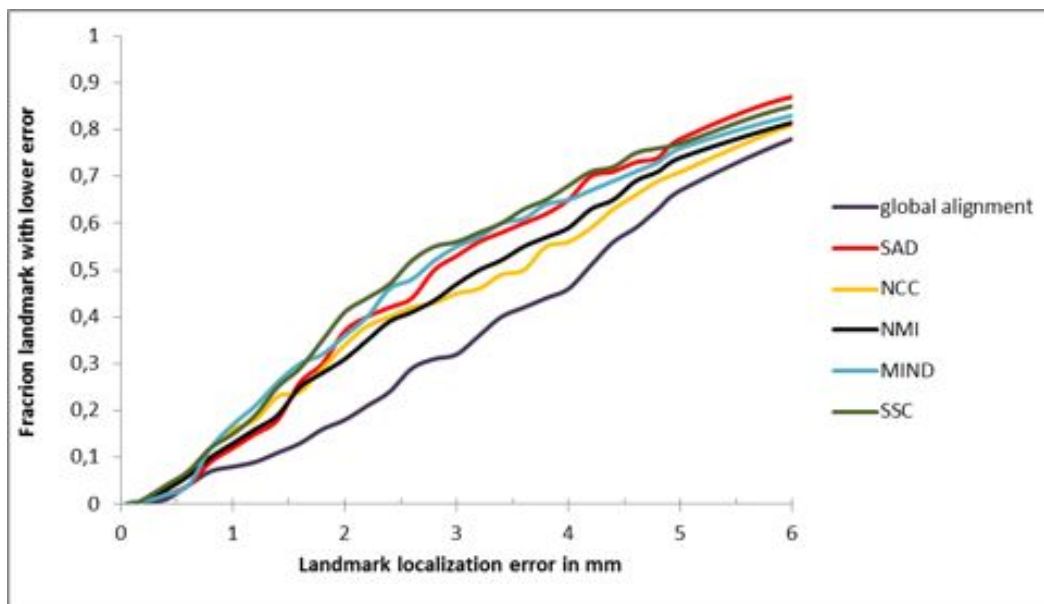
Figure 5.4: Cumulative distribution of TRE, expressed in mm, evaluated for 202 landmarks annotated on 10 US clinical datasets.

Parameter	SAD	NCC	MI	MIND	SSC
Scale	2	2	2	2	2
Search Aperture	3	3	3	3	3
Patch Aperture	1	3	6	1	2
Patch Sampling	1	2	1	2	2
Self Similarity Patch Aperture	-	-	-	2	2
Self Similarity Sigma	-	-	-	20	20

Table 5.2: estimated values corresponding to the optimal configurations for the feature descriptors in case of DBT data.

SAD and NCC also in single modality registration tasks. The additional modality-invariance does not affect, at least for what concerns breast images, its descriptive power. This can be explained by the fact that the self-similarity concept is able to preserve the relevant structural information of underlying data equally or even better than the other tested descriptors. This consideration is supported also by the comparison with NMI that performs poorly precisely because its high degree of invariance at the cost of lower distinctive power. It is also worth noting that SSC performs better than MIND for the US datasets but does not significantly affect the results for what concerns the DBT datasets. This behavior was expected since the main advantage of this descriptor is to reduce the MIND sensibility to noise and hence is more suited for the US noisy data.
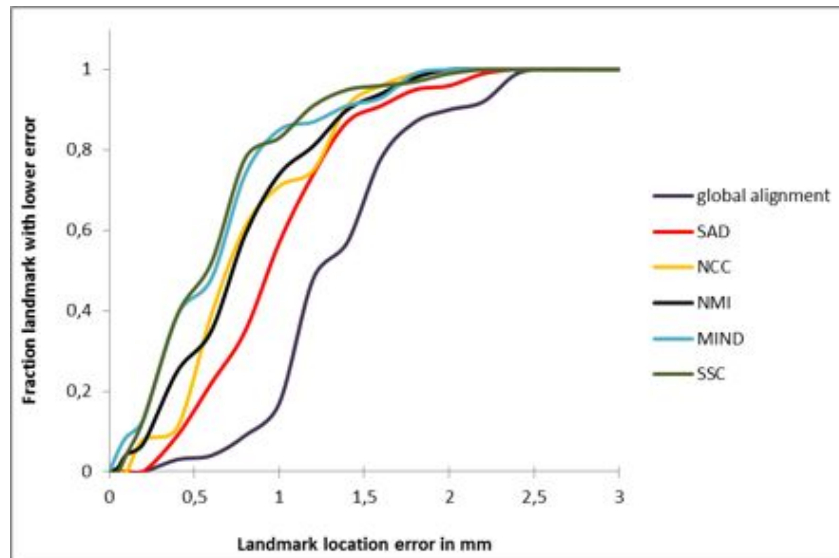
Figure 5.5: Cumulative distribution of TRE, expressed in mm, evaluated for 4 DBT clinical datasets.

Normalizers

In a deformable registration context normalizers represent the way in which the field is expected to change the reference in the template image, supplying the rules the field is subjected to during the optimization; from this point of view they can be seen as relaxed constraints - especially if compared with the fixed constraints implicitly embedded in all rigid transformation models. They act only on the shape of the field and are frequently expressed only in terms of its spatial derivative with the principal aim to preserve the field local smoothness and continuity. Common normalizers have been derived from physical models and the most common among them is the one derived from the diffusion model. For this reason the diffusion regularization term has been taken as reference in the AMBRA framework.

In this section the implementation details of this normalizer and of the more advanced regularization via bilateral filtering are described in more detail.

For the sake of clarity a visual example that demonstrates the effects of these kinds of Normalizer is presented. In figure 5.6(a)-(c) two different breast US scans of the same patient are depicted. In figure 5.6(b)-(d) the same images are represented with overlays that represent the adipose tissue's structures (in yellow) and the glandular tissue's structures (in blue). In the figure 5.6(d) two arrows draw the attention to the motion that these structures are subjected to: the adipose tissue's structures move globally to the left while the glandular tissue's structures to the right.

- **Diffusion**: the diffusion normalizer is based on the heat equation: the
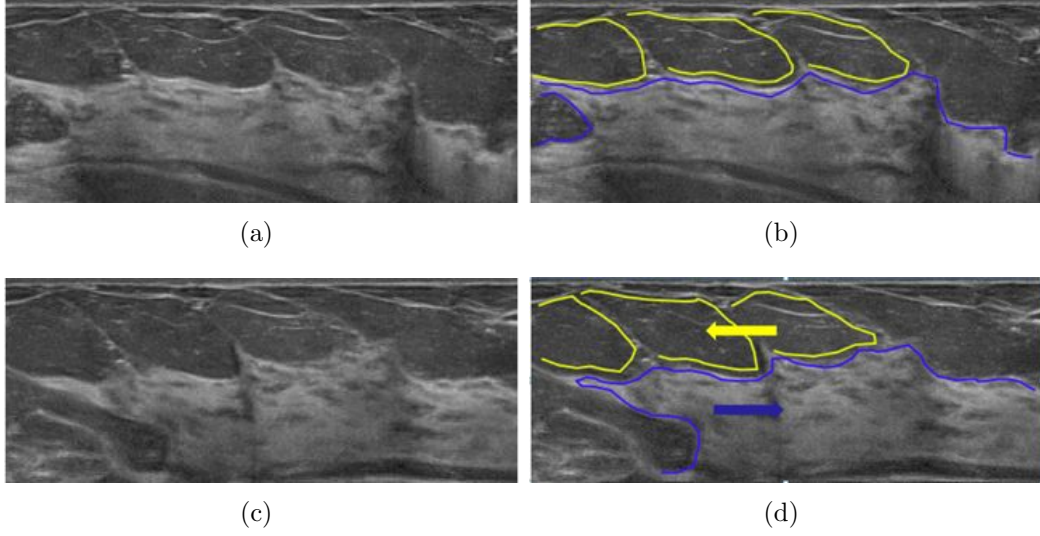
Figure 5.6: Sliding motion example. In (a) a US slice of a breast acquisition is shown. In (b) the adipose structures in the same image are outlined in yellow and the glandular structures in blue. In (c) a different US acquisition of the same breast region is represented, but in this case the breast is subject to a different compression. Figure (d) shows how the adipose structures slid to the left under probe pressure, while the glandular structures moved to the right, compared to the image (b).

Laplace operator is applied with a constant weight alfa to the deformable field as regularization term in the functional to be minimized:

$$\frac{\partial \sum_{x \in Vref} Dist(FV_{ref}(x), FV_{tem}(u(x)))}{\partial t} - a\nabla^2 u = 0 \qquad (5.10)$$

where **u** is the displacement field, Vref is the reference volume, **x** a location in the reference volume, FV_{ref} and FV_{tem} the feature vectors in the reference and template volumes, Dist their distance (as defined in the previous chapter and treated in the previous section) and dt is a small increment in time representing each iteration in the optimization process. Thus this normalizer represents an energy that tries to minimize the first order spatial derivatives of the displacement field. A common implementation of the isotropic diffusion regularization (Thirion 1998) is to apply a Gaussian convolution to the field components at each iteration:

$$u_{t+1}(x)G_\sigma * u_t(x) = W(x) \sum_{y \in N(x)} u_t(y)exp(\frac{\|x - y\|^2}{2\sigma^2}) \qquad (5.11)$$

where N(x) is the neighborhood of x defined by the kernel size, sigma is the standard deviation of the spatial Gaussian function and W(x) is a

normalization factor for the neighborhood of x.

The effect of this regularization method can be seen in figure 5.7. Figure 5.7(a) shows some vectors with different orientation that represent local attractive forces due to the feature term in the yellow region from figure 5.6(d). As can be seen, although most of the vectors point to the (correct) left direction, some of them are captured by local minima and, more in general, the field appears very discontinuous. An iterative convolution of the displacement field with a Gaussian kernel removes the discontinuities giving to the deformation a more acceptable representation (figure 5.7(b)).

The integration in the AMBRA framework was straightforward: it requires only to implement the Normalizer update field method through a Gaussian separable convolution 3D kernel and with a dynamic parameter representing sigma.

(a)   (b)
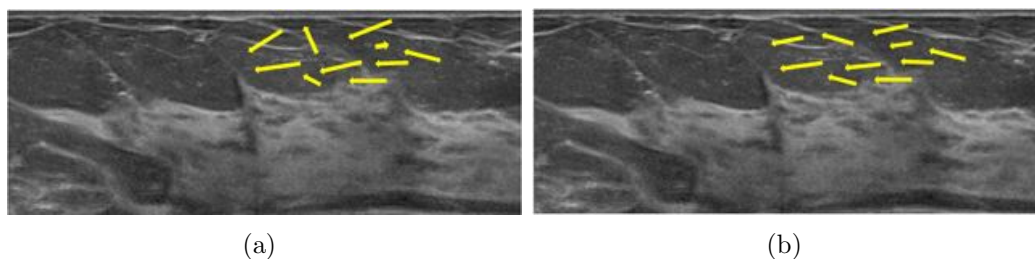
Figure 5.7: Diffusion operator effect: in (a) the raw local forces due the attraction of similar features are depicted; they point on average to the left according with figure 5.6(d). Nevertheless some vectors fell to local minima. Figure (b) shows the resulting field in the same locations when the diffusion operator is applied. In this case the displacement filed exhibits a more smooth and proper deformation.

In complex scenarios however the estimation of plausible deformations requires the application of additional rules to the ill-posed registration problem: when multiple organs move independently towards different directions, a good compromise between the general smoothness of the field and a correct description of local discontinuities could hardly be found. Several approaches have been proposed to overcome this problem but most of them are not practicable in this context: the parametrization of the strength of the normalizer as well as the direction dependent regularization (DDR) schemes based on local structures boundaries (Schmidt-Richberg et al. 2012) are unfeasible due to the fact that both require initial reliable segmentations that are very difficult to obtain for inner breast structures; implicit anysotropic diffusion based on image gradients is a choice but the image contrast depend on modality and cannot be generalized easily (Hermosillo et al. 2002). Other approaches exploit the image local intensities to distinguish tissues properties that can be

used to modulate the regulation term based on prior knowledge (for example Hounsfield units in CT data).

To handle the severe sliding motion of structure inside breast an advanced regularization via adaptive bilateral filtering (Papiez et al. 2014) has been followed and a normalizer based on this approach has been integrated in AM-BRA.

- **Bilateral filtering**: this Normalizer extends the diffusion concepts applying to the displacement field the advanced non-linear filtering technique originally proposed for image denoising (Tomasi et al. 1998) known as bilateral filter. Instead of iteratively convolving the field with a simple spatial Gaussian kernel it uses two combined Gaussian kernels, one defined on the spatial domain and the other defined on the intensity domain:

$$u_{t+1}(x)G_\sigma * G_\rho * u_t(x) = W(x) \sum_{y \in N(x)} u_t(y) exp(\frac{\|x-y\|^2}{2\sigma^2}) exp(\frac{\|I(x)-I(y)\|^2}{2\rho^2})$$

(5.12)

  where I(x) and I(y) are the intensities of the reference volume in the locations x and y, sigma is the standard deviation of the spatial Gaussian function, while rho is the standard deviation of the intensity Gaussian function. The net effect is that the field is averaged in every location only with the neighbors that have a similar intensity. With the assumption that, regardless to the modality used, a breast sub-structure made out of the same tissue is likely to be represented by an almost constant intensity, the bilateral normalizer is expected to allow discontinuities along the borders of different objects preserving at the same time the smoothness of the field in their inner region. In figure 5.8 the behavior of the Bilateral Normalizer is compared with the Diffusion Normalizer: given that the Feature Term pushes the top region correctly to the left and the bottom region to the right (as in figure 5.6(d)), the Diffusion operator (figure 5.8(b)) is likely to make the optimization converge to a displacement field that has a smooth transition between the top and bottom regions. The sliding motion is instead well handled by the Normalizer based on Bilateral Filtering (figure 5.8(a)) as the field is smoothed in top region with only other values from the top region and in the bottom region with only other values from the bottom region.
  The integration of this Normalizer inside AMBRA followes the same logic used in the case of Diffusion instance, with the only difference that in this case at iteration-time the update function makes use of the reference volume (passed as argument in the framework Normalizer Interface) as the context for the local evaluation of the intensity Gaussian weight.

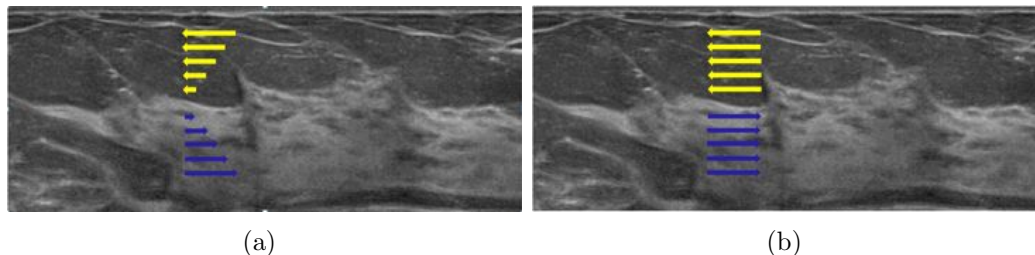(a)   (b)

Figure 5.8: Bilateral operator effect: (a) shows the most likely configuration of the vector field resulting by the application of the diffusion operator to two regions moving in opposite directions under pressure. As a result a smooth transition of the field between the two regions is obtained. Figure (b) shows the result after the application of a bilateral filtering operator; in this case the Gaussian averaging rule is only applied over similar regions (in terms of intensity values). It results in a more acceptable and proper deformation model, allowing a discontinuity of the field at the boundary of the two different regions.

Measures

Comparison tests were run also for the normalizers evaluation. The same testing sets used for the Feature Metric term were chosen.

For what concerns US-US registration the comparison between the optimal results (SSC was chosen as Feature Descriptor) obtained using the Diffusion normalizer and the one based on Bilateral Filtering shows that the seconds performs better and improves the target registration error (figure 5.9). (table 5.3) The DBT results instead does not show any remarkable benefit in

Parameter	Diffusion	Bilateral
Spatial Standard Deviation	4.0	3.5
Intensity Standard Deviation	-	0.2

Table 5.3: optimal values for normalizer's parameters for US data.

the use of the Bilateral Filtering regularization. The TRE rates outlined in the following graph differences only by small quantities. (figure 5.10). (table 5.4) These different behaviors can be better analyzed exploiting the op-

Parameter	Diffusion	Bilateral
Spatial Standard Deviation	4.0	4.0
Intensity Standard Deviation	-	16.0

Table 5.4: optimal values for normalizer's parameters for DBT data.

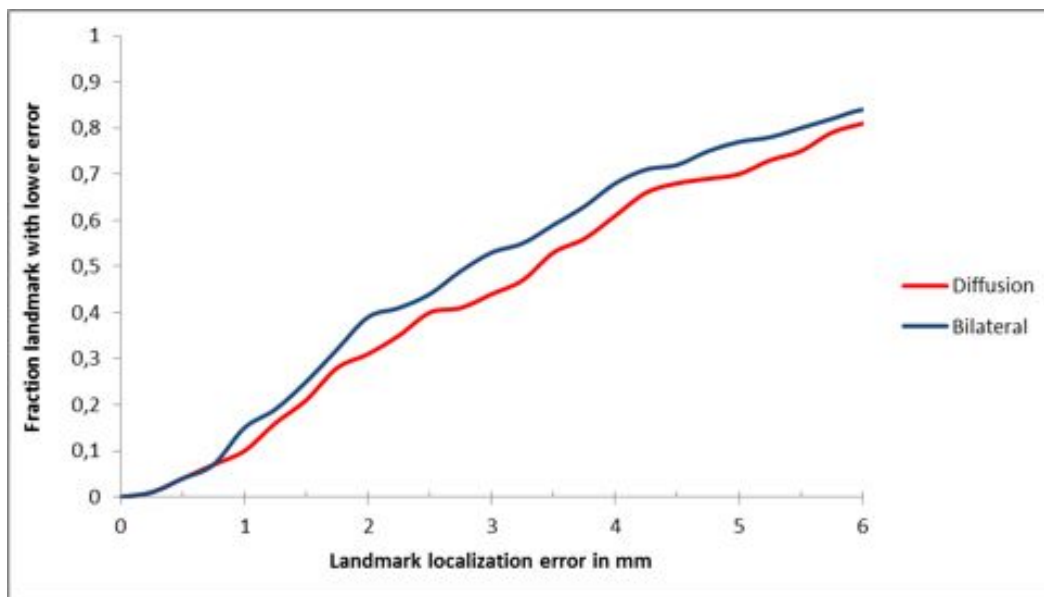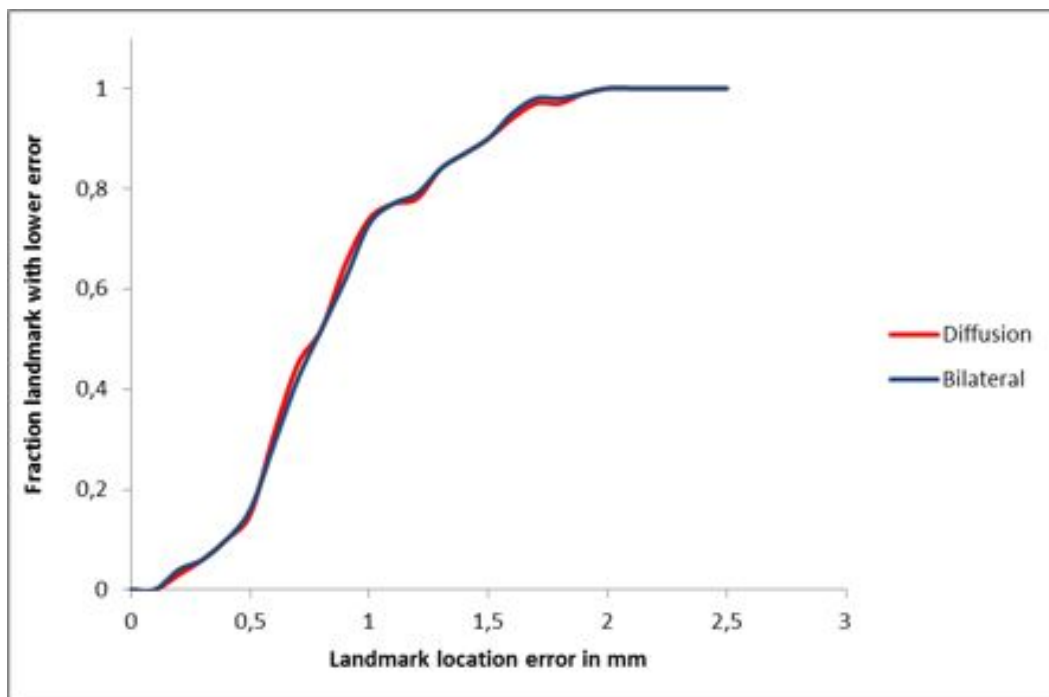timal parameter configuration in the two cases: for the US-US testing set the

Figure 5.9: Comparison of the cumulative distributions of TRE, expressed in mm, obtained using the diffusion and the bilateral filtering regularization for the US clinical datasets.

spatial standard deviation of the Bilateral Normalizer is slightly higher than the standard deviation of the Diffusion Normalizer and the intensity standard deviation is small. Instead in the best configuration for the DBT-DBT case the intensity standard deviation is very high while the spatial standard deviation of both the Diffusion and the Bilateral regularization instances are equal. This means that in the second case the optimal solution of the Bilateral Normalizer approximates a Diffusion Normalizer because the normalized convolution with its intensity-based Gaussian function does not have any influence in the field averaging among neighboring locations. This result can be easily explained by the fact that the sliding motion in the DBT images (acquired in different times) is limited if compared with the US images (acquired in different positions and with different compressions applied to the breast).

Figure 5.10: Comparison of the cumulative distributions of TRE, expressed in mm, obtained using the diffusion and the bilateral filtering regularization for the DBT clinical datasets.

Chapter 6

Case studies

Contents

In this chapter three applications of the registration framework are investigated. Each of them represents a specific derivation of the framework that involves the customization of both the initialization and the optimization part. In each of the following sections, a specific clinical task is addressed. First a brief description of the problem is presented. Then the method derived from AMBRA is explained. Since each of the presented customizations of the framework involves one or more segmentation algorithms (to support the initialization block), the description starts outlining these segmentation procedures. Then the algorithms used for the initialization block is presented. The description of the configuration adopted for the optimizer of the deformable registration algorithm in the case under analysis follows. Finally the registration results are presented. In addition a specific application of the whole algorithm is proposed. As mentioned before, this work is currently focused on

applications regarding the novel 3D imaging technologies for breast. Hence, the proposed case studies refer to DBT and ABUS datasets.

6.1 Intra-modality registration of different Ultrasound breast sections

For the breast examination using the currently available ABUS system, three different scans for each breast need to be collected. As reference point, the area under the nipple has to be imaged in each scan. These three breast views are scanned in the same imaging session and performed by the same technician. Hence neither time-related differences in breast anatomy nor operator-dependent image acquisition differences are expected. The main issue in these datasets are related to the different position of the US probe, that implies a different pressure applied by the probe to the breast during acquisition. As the breast has a non-rigid behavior, the different probe positioning lead to a relevant breast tissues deformations among the scans. Although the ABUS systems allow to save physician's time as the breast scanning is performed by a technician, the images interpretation requires usually a lot of time. Indeed the physician has to read three different volumes for each exam, and the estimated interpretation time is on average 10 minutes for each breast (see the section 2.3 for more details). ABUS workstation allows to display the multiplanar reconstruction of a volume (i.e. transverse, coronal and sagittal planes) and to easily navigate inside a volume. In addition in the same screen two views (i.e. AP, LAT or MED views) referring to two different scans can be displayed. Although the two views present an overlapping area, it appears differently due to the deformations impressed during the scanning. Hence, the same region has to be examined three times. Although the workstation assists the physician in the image reading, for each breast three different volumes have to be scrolled and completely examined. Currently, there is no tool that helps to correlate and compare the highly deformed overlap areas among scans. The purpose of the image registration in this case is to estimate a mapping function between the same breast area imaged in two different US scans. It is aimed to an automatic correlation between the volumes, allowing a more accurate and faster ABUS images interpretation. A dataset for this task consists in two US volumes representing adjacent breast sections and comprising an overlapping area. This area shows relevant deformable displacements of breast tissues. The datasets used for this task were collected using the Siemens ABVS scanner.

Given two scans, for the image registration the nipple is the reference point to perform a global alignment of the input volumes. Then the displacement field in the overlap area between the volumes is estimated. The displacement field can be used to perform an automatic synchronization of the cursor positions

between two different breast views. However, for this task a novel application is proposed. The displacement field is used to deform both volumes in order to perform a fusion of the input volumes in a single one with an extended field-of-view.

For the explanation of the method's steps a test dataset has been selected and the results are presented on bidimensional images for the sake of a clearer visualization. For an easier interpretation, images showing well defined and easily recognizable structures have been chosen (see figure 6.1).

(a)

(b)
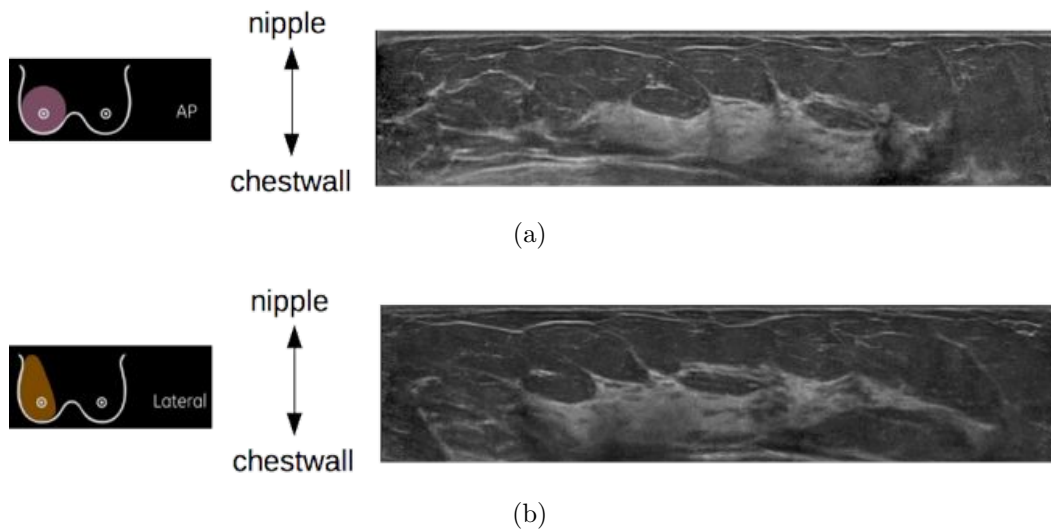
Figure 6.1: an example of two ABVS volumes corresponding to a left breast. Figure (a) shows an axial image extracted from the AP view, while (b) is an axial image of the LAT view.

6.1.1 Nipple segmentation

Irrespectively of the position of the 3D US probe the nipple is always present in the scanned image data as it is used by the physician as a reliable landmark. The nipple solid structure attenuates the US signal in its underlying region (figure 6.2). This characterization of the nipple area has been used for its automated detection. The volumetric data of a breast section is first re-sampled to a matrix with uniform spatial resolution of 0.4 mm in all directions. Then the component tree of all the dark extremal regions of the volume matrix is computed (Matas et al. 2002). A dark extremal region is the dark connected component of an image binarized at a certain threshold level. The component tree is a rooted, connected tree whose nodes are the dark extremal regions for all possible threshold levels. They are linked in such a way that a region R obtained with a threshold T has a parent node (navigating the tree upwards) corresponding to a region that includes R and is obtained with a
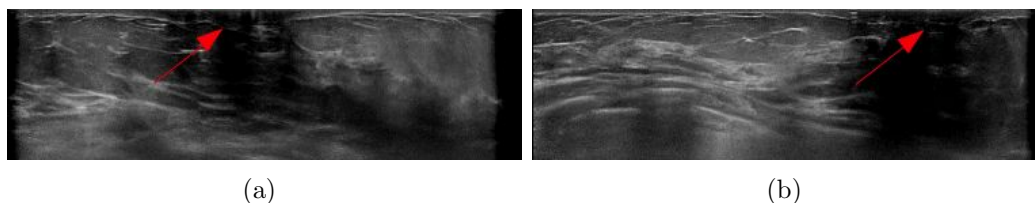
Figure 6.2: Attenuation of the US signal in the region under the nipple in the AP view (a) and the LAT view (b).

threshold $Tp < T$; it has also a set of child nodes (navigating the tree downwards) corresponding to regions included in R and obtained with a threshold $Tc > T$. Finally the maximally stable extremal 3D regions (MSER) are selected (Donoser et al. 2006). The property that characterizes these regions is to slowly change their cardinality (size) moving upwards and downwards the branch of the component tree they belong to. The net effect is to select a set of dark connected components of the volume whose size does not change much if the volume data is binarized with thresholds in a wide range around the threshold associated to the region itself.

The dark MSER set always includes the region underlying the nipple but additional selection steps are required to prevent two major issues:

1. Exclude dark regions whose attenuated signal is caused by other phenomena. An example is the lateral attenuation of the signal due to the lack of coupling between breast and probe as well as a smaller breast size than the probe.

2. Select among the remaining MSERs the one that best fits the nipples area. In fact in some cases the attenuation under the nipple area merges with other attenuated regions, at least for a threshold level that is not sufficiently low.

To this aim first a clustering algorithm is used to connect the MSER that share a common MSER ancestor along the component tree. Each group of regions is represented by this common ancestor (that is the largest MSER of the group). Some characteristics of these representative regions are computed: size, extension (bounding rectangle), location (center of mass) and the three first central spatial moments including variance and skewness. Only regions that are not too small or large and whose bounding rectangle is not too lateral are selected. A special treatment is reserved for the lateral regions: their variance and skewness are used to establish if they are compact (the region is discarded) or not (the region is selected for further analysis); in fact if the nipple region merges partially with a lateral region their union can be classified as lateral but looses the compactness. The group of MSER in the same cluster of the not-discarded lateral region is explored downwards selecting, if any, the

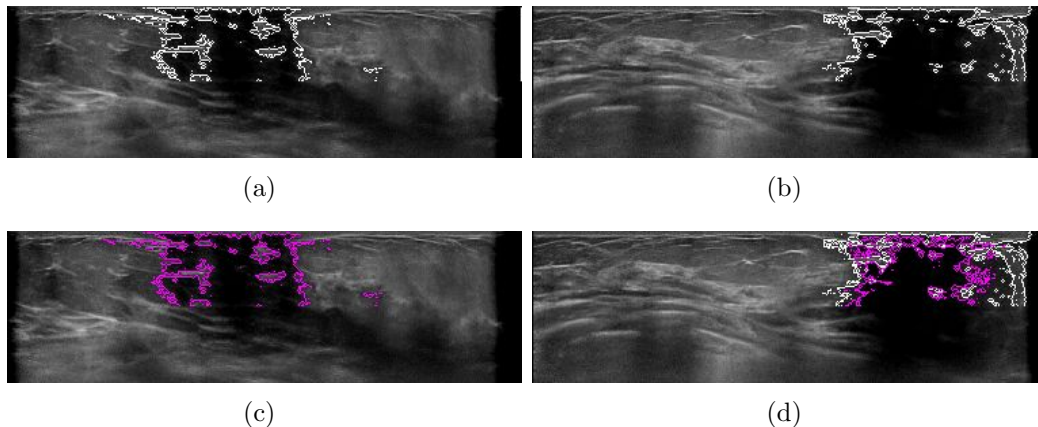first MSER that fulfills all the desired requirements (figure 6.3). Finally a

Figure 6.3: Figure (a) shows the representative region (highlighted in white) of the selected cluster in the AP view. Figure (b) shows that the cluster for the LAT view is correctly selected by the algorithm; however, as the representative region is quite extended and on the side of the image, it includes also the unwanted lateral black region (due to the uncoupling between the probe and the breast). Hence, in the first case, the segmented region (drawn in violet in (c)) corresponds to the parent of the cluster (in the graph model), meanwhile in the second volume the component tree is searched downwards and a sub-region is selected (drawn in violet in (d)). This selection allows to prevent the lateral region.

classifier based on these computed features evaluates the region that is more likely to represent the nipple area. The center of mass of the selected region projected into the skin plane determines the nipple position (figure 6.4). The efficiency and accuracy of the proposed method was tested on a test dataset of 90 US scans. The nipple was detected with a detection rate of 90% inside a range of 15 mm with respect to the ground truth manually annotated by an expert. The procedure succeeded also when the nipple was not completely scanned in the field of view. The average distance between the automated detected nipple position and the reference is $5.1 \pm 3.1mm$. The computation time is 3 seconds on a 2.5 GHz processor.

6.1.2 Method

Initialization

For the current application the initialization block of the framework is implemented with a procedure that begins computing the nipple location in both the reference and template volumes using the nipple segmentation tool described in the previous section. A raw global translation vector:

$$T_{gl} = (dx_{gl}, dy_{gl}, dz_{gl}) \tag{6.1}$$

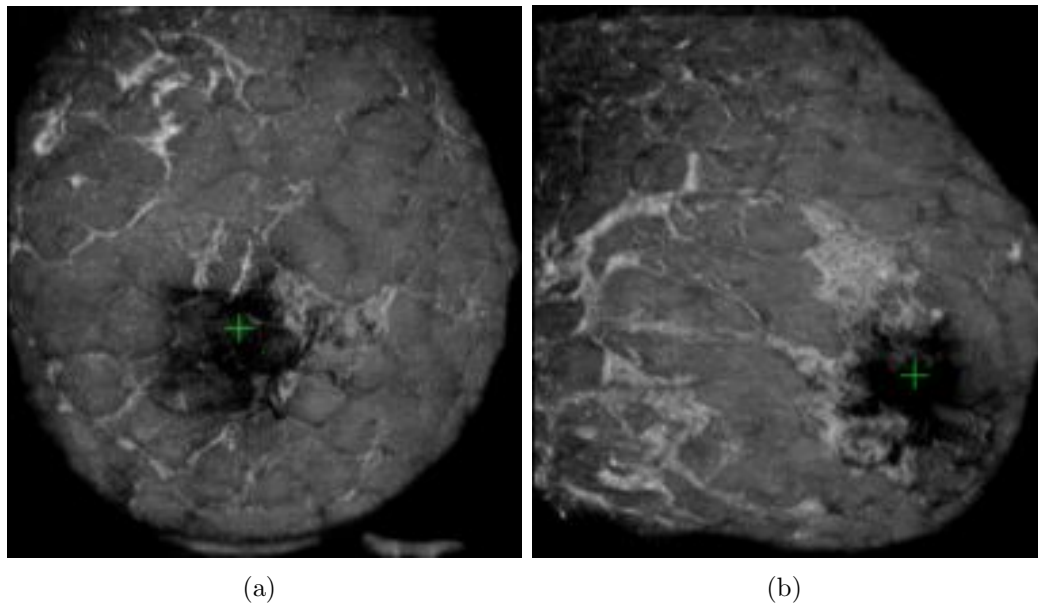(a)   (b)

Figure 6.4: The automatically detected nipple positions in the AP (a) and LAT view (b).

is coarsely evaluated as the displacement between the resulting locations of the nipple in the two volumes. Though the attenuation of the signal in the nipple region could be useful in determining a stable landmark location in both reference and template images, its decreased intensity constitutes a severe artifact. An adaptive intensity enhancement filter based on a sigmoid transfer function (Tao et al. 2006) is then applied to the image data. The filter is modulated by a weighting mask that takes into account both the spatial position of the nipple and the inverse intensity of the original image, in order to restore the original tissue contrast in the region affected by the loss of information (figure 6.5). An additional step regards the estimation of the

(a)   (b)
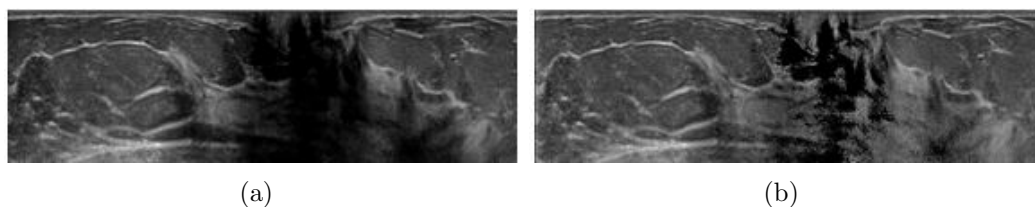
Figure 6.5: (a) example of an image of the breast tissues corresponding to the region under the nipple where a high signal attenuation is shown. The result obtained applying a local contrast enhancement filtering is reported in fIgure (b). It can be noted how the filter selectively normalizes the intensities in the areola region without affecting the surrounding regions, resulting in an equalization of the image intensities distribution.

local movements of the inner structures. In the algorithm workflow, it links the rigid model estimated in (eq. (6.2)) to the deformable model that will be used as the starting configuration for the optimization phase, represented by the displacement field:

$$F(p) = (dx(p), dy(p), dz(p)) \tag{6.2}$$

defined in every spatial position p of the reference geometrical space.

To this aim first the reference space is partitioned in a grid of middle-scale blocks, whose dimensions are comparable to the expected size of distinctive subcutaneous adipose objects or structures embedded in mammary gland that are well characterized in US scans; then a patch window, obtained extending the block dimension to include the neighboring structures, is associated to each block. Finally, taking advantage of the global translation vector (eq. (6.2)), each patch window is searched in the template image only in the neighborhood of its expected location (figure 6.6). The best evaluated local displacement is temporary assigned to the block central position ($\mathbf{p}_{block}$):

$$F_{loc}(\mathbf{p}_{block}) = (dx_{loc}(\mathbf{p}_{block}), dy_{loc}(\mathbf{p}_{block}), dz_{loc}(\mathbf{p}_{block})) \tag{6.3}$$

Unfortunately, due to the floating nature of the anatomical structures represented in breast scans, the size of the search neighborhood for each block needs to be large; this implies a low degree of reliability for the single block local displacement evaluation, particularly where the reference data within the patch window is not enough distinctive. For this reason the local block matching process is associated with a reliable validation procedure: Speeded Up Robust Features (SURF) are extracted from highly characterized points in the scale-space representation of both the reference and the template images (Bay et al. 2006). Next, the two sets of points are cross-matched excluding the couples that do not mutually match (Brown et al. 2007).
For each estimated local translation a confidence index c block is computed as the rate of the SURF points found inside the reference block region, whose matching SURF points lie inside the corresponding template block (figure 6.6 and figure 6.7). The final step in this phase is the initialization of the full resolution deformable registration field in the reference space. First a weighted vector $F(p_{block})$ is assigned at each $\mathbf{p}_{block}$:

$$F(\mathbf{p}_{block}) = w(c_{block})F_{loc}(p_{block}) + (1 - w(c_{block}))T_{gl} \tag{6.4}$$

where $w(c_{block})$ is an increasing function that maps in [0,1] the value of the c block index to the actual weight to be applied to the local translation estimation. Then a trilinear interpolation procedure fills the remaining locations of the field.
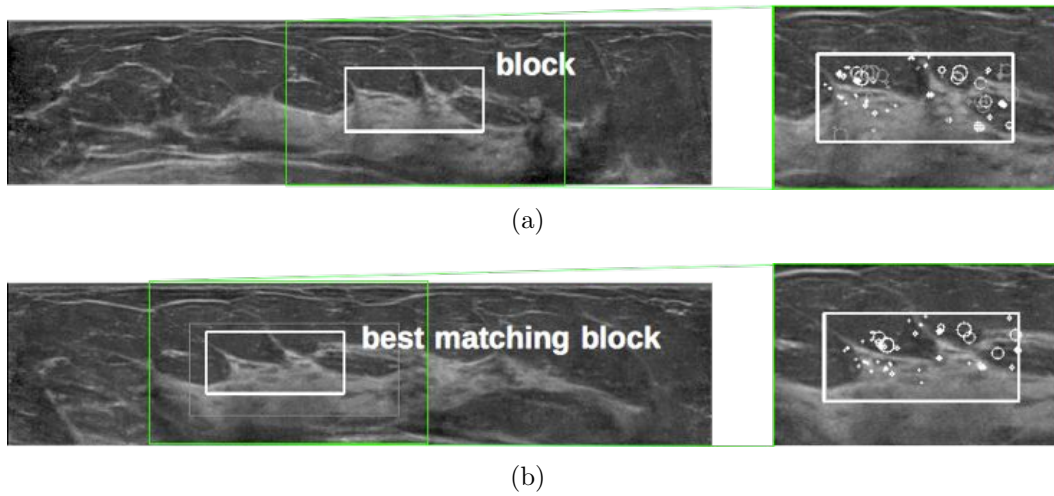
(a)

(b)

Figure 6.6: On the left an example of the local block matching procedure between the reference (a) e template (b). On the right a detail of the SURF validation of block matching: the points lying inside reference patch window are drawn in gray; among these the ones found also in the estimated template block are displayed in white.
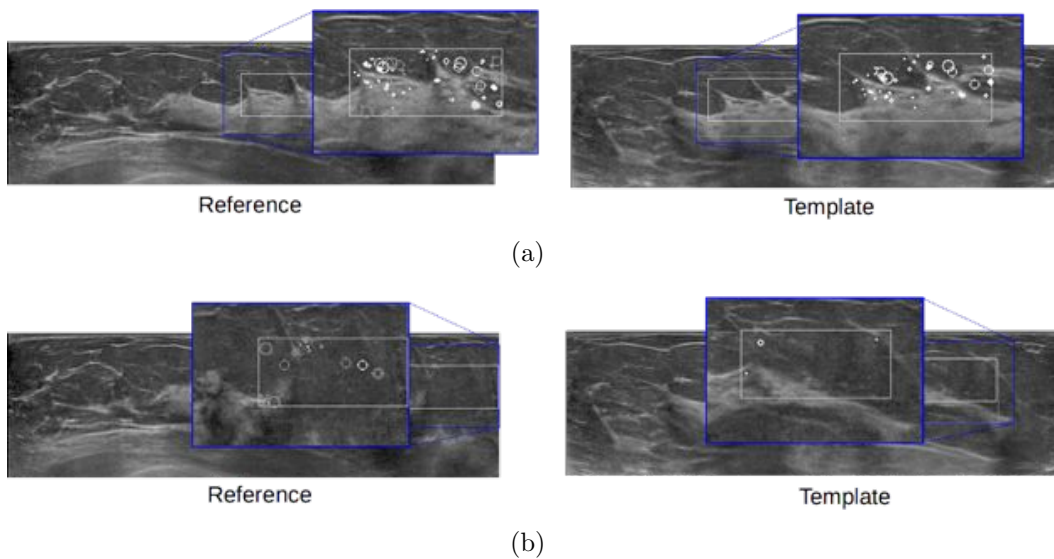

(a)

(b)

Figure 6.7: (a) case of good local matching confirmed by a high confidente index of 0.6, that means that the 60% of SURF points found inside the reference block are correctly mapped in the corresponding block selected in the template; on the contrary, figure (b) shows an example with a low value of the confidence index (equal to 0.2) that suggests a possible error in the matching.

Optimization

The optimization process embedded in AMBRA was customized by selecting the following configuration:

- For what concerns the Feature Metric the Self Similarity Context (SSC) descriptor was selected for its ability to handle US noise and speckle.

- The Bilateral Normalizer was the choice for the regularization term to preserve the severe sliding motion of the most relevant structure inside the breast.

- A Multi Resolution approach in two stages for two different scales was followed: in the first one the field is optimized at low image resolution, with a scale comparable to half size of the blocks defined in the local alignment during initialization; this choice aims to improve the local estimate obtained in that phase with a non rigid transformation model but capturing features of comparable size. The second stage instead makes use of the finest scale to get a full resolution preciseness for the deformable field.

- The inverse consistent constraint option was enabled to grant that the correspondence between the two volumes is invariant to the choice of source and target.

Thus the optimization process ends up with the forward and backward displacement fields $\mathbf{F}$ and $\mathbf{G}$ whose inverse consistency will be used in the application described in the next section.

Results

In figure 6.8(a) a reference (top) and a template (bottom) volume are shown. Figure 6.8(b) shows the registration of the template volume (in green) in the reference space (red). The improvement after the deformable registration (c) with respect to global (a) and local (b) alignment can be noticed. The procedure has been applied on 9 datasets acquired with Siemens ABVS scanner. 141 landmarks have been carefully selected by experts. The evaluation of the results is based on the TRE of anatomical landmarks embedded in the framework. The cumulative distribution of landmark localization error is plotted in figure 6.9. The average error in mm (2 ± 2) achieved with the proposed algorithm is significantly lower than the one (3 ± 3) obtained with the state-of-the-art MIND method (Heinrich, Jenkinson, Bhushan, et al. 2012). The average displacement of landmarks before deformable registration is 4 ± 3 in mm (table 6.1). In table 6.4 the computing times of the algorithm are reported. They refer to two US volumes with the characteristics indicated in tabletable 6.2 processed using a computer with the performance reported in table 6.3.

6.1.3 Image stitching

An interesting application of the presented registration method is a novel merging procedure of two US breast sections (source volumes) in a single ex-

Parameter	avgTRE (mm)	standard deviation (mm)
Global alignment	4	3
MIND	3	3
Proposed method	2	2

Table 6.1: the averageTRE, with the relative standard deviation, estimated after the global alignment, the application of MIND procedure and the proposed method.

US dataset characteristic	Value in mm	Value in pixel
width	$\tilde{1}52$	712
height	$\tilde{4}5$	420
depth	$\tilde{1}67$	318
voxel size	$0.21 \times 0.11 \times 0.52$	
uniform sampling	$0.2 \times 0.2 \times 0.2$	

Table 6.2: technical specifications of the 3D US data.

Computer's property	Value
CPU speed	2.50 GHz
# of processors	12
# of cores	2
RAM	48.0 GB

Table 6.3: overview of the properties of the processor used for the algorithm computing.

Algorithm's step	Time
pre-processing	1'
step #1 (global)	10"
step #2 (local)	5'
step #3 (deformable registration)	1h

Table 6.4: computation time for each step of the registration method.
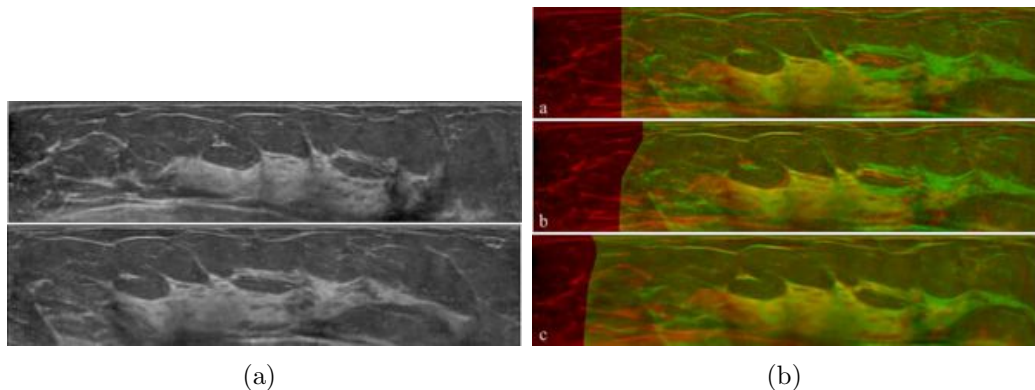
(a)   (b)

Figure 6.8: (a): reference (top) and template (bottom) input images. (b): registration of the template image (green channel) in the reference space (red channel of image) after: a) global alignment (on the top); b) local block matching (in the middle); c) deformable registration with inverse consistency constraints (on the bottom).

tended 3D view (target space). The proposed method is an extension to 3D and to deformable field models of the algorithm based on homographic transfer models used in (Brown et al. 2007) to render a stitched panorama. In that work, once established the rigid transformation model between the two source images, a parameter t with values between 0 and 1 indicates the point of view for the resulting panorama: if $t = 0$ the point of view is the one of the first image and the second image is projected on it; if $t = 1$ the point of view is the same of the second image and the first image is projected on it; for all other values of t, for example 0.5, both images are projected onto a target space that is halfway between the two original points of view. Irrespectively of the value of this parameter the two images are rendered in the target space and the blending operation is applied to the rendered projections. For the sake of simplicity and without loss of generality the novel merging procedure proposed is described in two dimensions (images instead volumes) and assuming that the template image is placed to the bottom-left of the reference image as in figure 6.10. This means that the direct deformable field F is on average lower than 0 for both the x and the y coordinate; and that the inverse deformable field G is on average greater than 0. The average displacement vector Tavg estimated from F and G represents a translation that turns out to be useful to define an appropriate domain for the possible target space of the merged image: this domain is defined as the sum of the original image size and $|T_{avg}|$. Source images are rendered in the obtained target space with a hybrid approach, that combines the deformable and the rigid translation model associated to T_{avg}. Relying on the rigid model three regions can be identified in the target space: the intersection region IR corresponding to the overlapping area, and the two non-overlapping regions, R_{ref} and R_{tem} (fig-
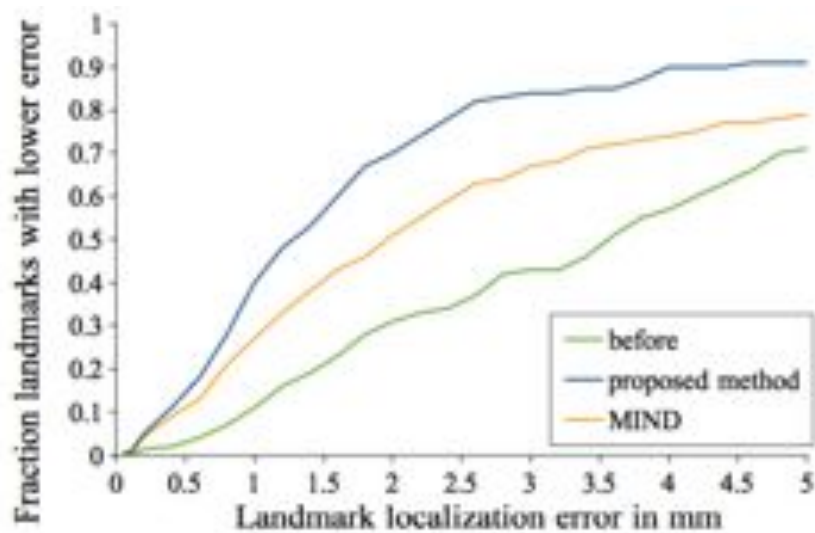
Figure 6.9: Cumulative distribution of target registration error, in mm, evaluated for 141 expert landmarks identified on 9 datasets. Comparison of the results obtained: before registration (only with images alignment, delineated in green line), with the proposed method (blue line) and with the MIND procedure (as proposed in Heinrich, Jenkinson, Bhushan, et al. 2012, orange line).
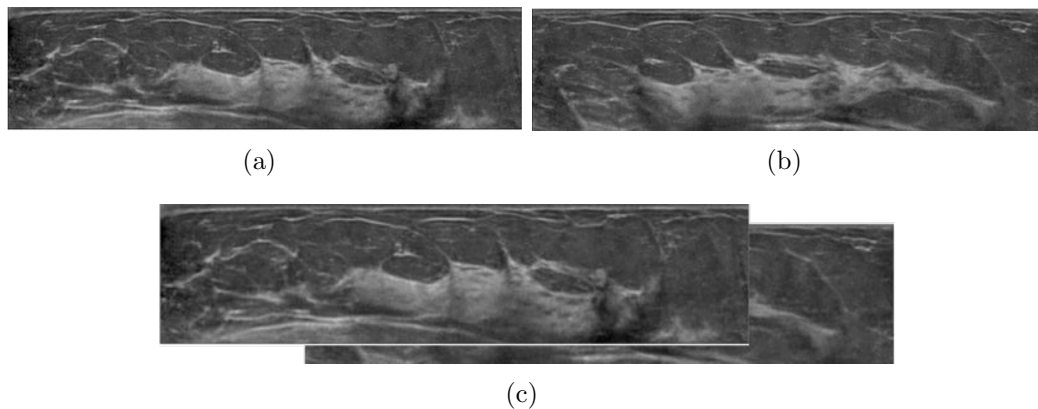

(a) (b)

(c)

Figure 6.10: two axial images of the reference (a) and template (b) volumes and the schematic overlap (c).

ure 6.11). In the analysed configuration the template image is translated in the R_{tem} by T_{avg}. Instead in the IR area a weighted combination of both the rigid and the deformable models is applied: calling q the position inside the IR region, the closer q is to the R_{tem} the more the template projection is affected by the rigid model; the influence of the deformable field F increases as the point q becomes closer to the region R_{ref}. A similar procedure is followed for the reference image: in this case R_{ref} is filled with the reference image

Figure 6.11: $\lambda(\mathbf{q})$ pattern representing the relative weight used to obtain the source images rendering in the extended view. The target space partition into the following three regions are shown: non-overlapping area R_{ref} corresponding to the reference space (bounded by the green line); non- overlapping area R_{tem} corresponding to the translated template space (bounded by the blue line); intersection region IR, within the red line, representing the overlapping space.

left unchanged (or translated by zero); in IR, the closer the position q is to the R_{ref} the more the reference projection is affected by the translation by zero; the influence of the deformable field G (that maps the reference image to the template space) increases as the point q becomes closer to the region R_{tem}.

Thus the displacement fields $(G_{ext}(q), F_{ext}(q))$ used to render the reference and template data in the extended target space are computed by the equations:

$$
\begin{aligned}
G_{ext}(q) &= \lambda(q)G(q - T_{avg}) + (1 - \lambda(q)) \\
&= \lambda(q)G(q - T_{avg})
\end{aligned}
\tag{6.5}
$$

$$
\begin{aligned}
F_{ext}(q) &= (1 - \lambda(q))F(q - 0) + \lambda(q)T_{avg} \\
&= (1 - \lambda(q))F(q) + \lambda(q)T_{avg}
\end{aligned}
\tag{6.6}
$$

where $\lambda(q)$ is a weight function that depends on the $L1$ distance of q from the regions R_{tem} and R_{ref} defined as:

$$
\lambda(q) = \frac{L1(q, R_{tem})}{L1(q, R_{tem}) + L1(q, R_{ref}).}
\tag{6.7}
$$

Once rendered both images in the extended target space, a Multiband Blending strategy (Burt et al. 1983) is used to produce the final image.

The blending technique is completely automatic and produces the extended volume shown in figure 6.12 (axial plane) and in figure 6.13 (coronal plane). It can be noted how structures located in the overlapping area between images are mapped in the extended view with the best representation layout: the
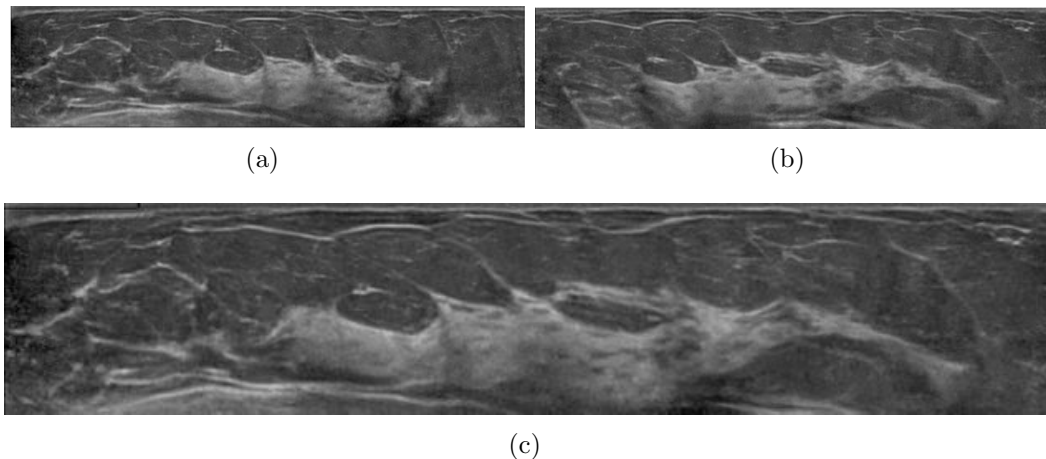
(a)

(b)

(c)

Figure 6.12: (a) reference axial image (related to the left breast's AP view); (b) template axial image (related to the left breast's LAT view); (c) axial section of the volume resulting from the proposed merging procedure of the reference and template images.

central areas of the original images contribute to form the information in the extended view, while lateral regions, that present more severe deformations (right area in the AP view and left area in LAT view, respectively) are excluded in the final representation. The external structures shown in only one acquisition are translated to the target without any geometrical or intensity distortion.

6.2 Intra-modality registration of breast Tomosynthesis images for temporal comparison

The breast cancer screening indications recommend the physicians to define their examination report using both the current and the prior patient's examinations (collected with same imaging modality), in order to feel more confident in the image interpretation and consequently in the cancer detection and prognosis. Given different examinations referring to two different times, local deformations for different breast positioning and compression as well as structural differences owing to physiological anatomy changes are expected. Currently, only a visual comparison can be performed. Typically the physician displays on the same workstation both volumes and scroll the planes one volume at a time until corresponding structures are found. It takes a lot of time and it is not a straightforward task. Moreover, usually a plane-by-plane correspondence is not feasible, as the breast tissues undergo to local deformations and, consequently, corresponding features could be visualized on
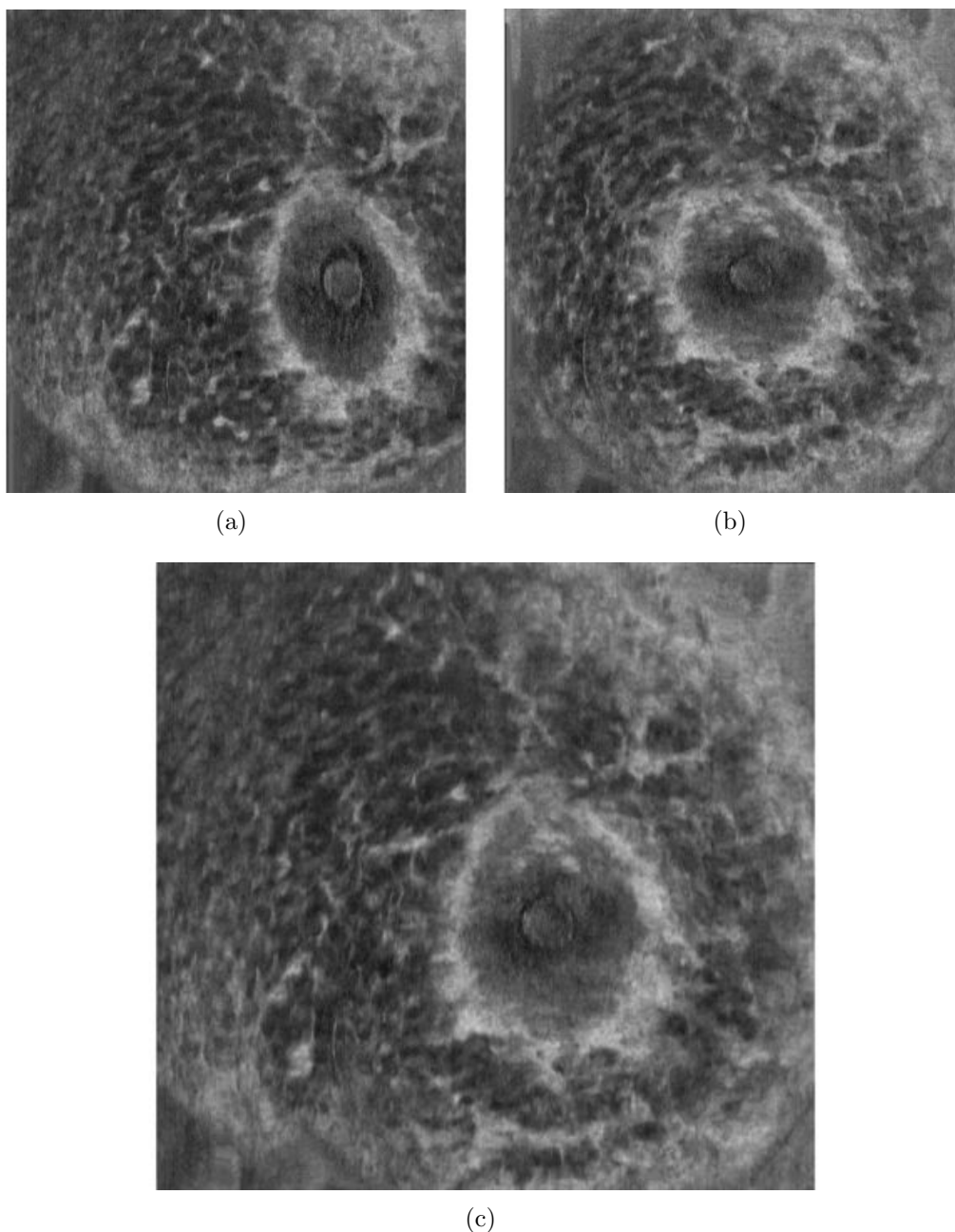
(a)

(b)

(c)

Figure 6.13: a coronal plane of the reference (a), template (b) and the resulting merged volume (c).

different slices. It means that during the images examination the physician has to continuously scroll the planes searching correspondences in order to compare the two volumes. Here a specific application of the image registration to DTB images collected at different times is proposed (an example of two DBT volumes is reported in figure 6.14). In this case the purpose is to estimate a mapping function between the two volumes aimed at supporting

the physician by automatically synchronizing the cursor positions inside the two volumes. Hence, the physician can scroll the planes inside one volume and simultaneously examine the corresponding regions inside the other one. For this task the datasets collected using two different equipments were available: the 3D images of the Siemens and Hologic DBT systems. Negligible differences are noted using these different datasets in terms of the image registration performance.

Given two DBT images, firstly the skin segmentation and the nipple detection are performed for both volumes. These information are used to compute a global alignment, and then the deformable registration field is estimated. Since structural differences including lesion appearance are expected between different examinations, an accurate pointwise correlation is not always feasible, and the displacement field is used to automatically find spatial correlations that allow to simultaneously display corresponding regions inside the two volumes.

6.2.1   Skin and nipple detection

The segmentation of the skin surface in DBT images is straightforward due to the absence of response signal out of the breast; the exact nipple location detection is more difficult but can take advantage of the knowledge derived from the tomosynthesis reconstruction process together with the fact that the density of the nipple is higher than the other breast tissues. The process is divided in two parts.

The first part acts independently on each DBT slice (the planes on the x-y coordinates of the DBT reference system) extracting the skin 2D contour Skin(z) and a candidate nipple location Nipple(z), with z corresponding to the third coordinate in the DBT reference system.

To find the contour Skin(z) each slice (figure 6.15(a)) is binarized with a threshold that corresponds to the 5% of the cumulative distribution function associated with the image histogram (figure 6.15(b)). The edges of the mask are extracted and the border points are ordered from top to bottom (figure 6.15(c)). The resulting contour is uniformly re-sampled along its curvilinear coordinate s with a point distance of 0.2 mm. The candidate Nipple(z) location is found considering the contour Skin(z) curvatures:

- For each point P(s) of Skin(z) the outgoing normal N(s) is first computed as the minor axis of the ellipse that fits the set of its 5 neighbor points (figure 6.16(a)-(b)).

- The field of normal vectors N(s) shown in figure 6.16(c) is further analyzed computing the divergence in every location.
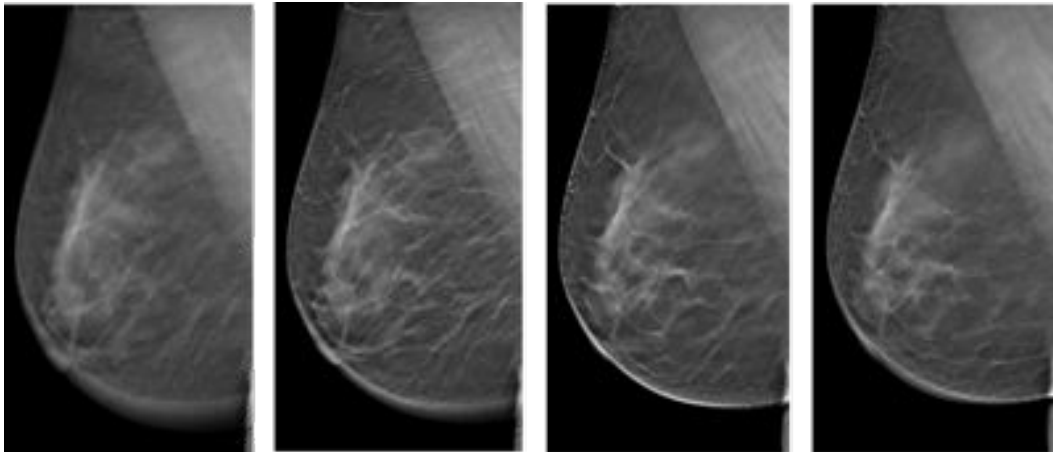
(a)
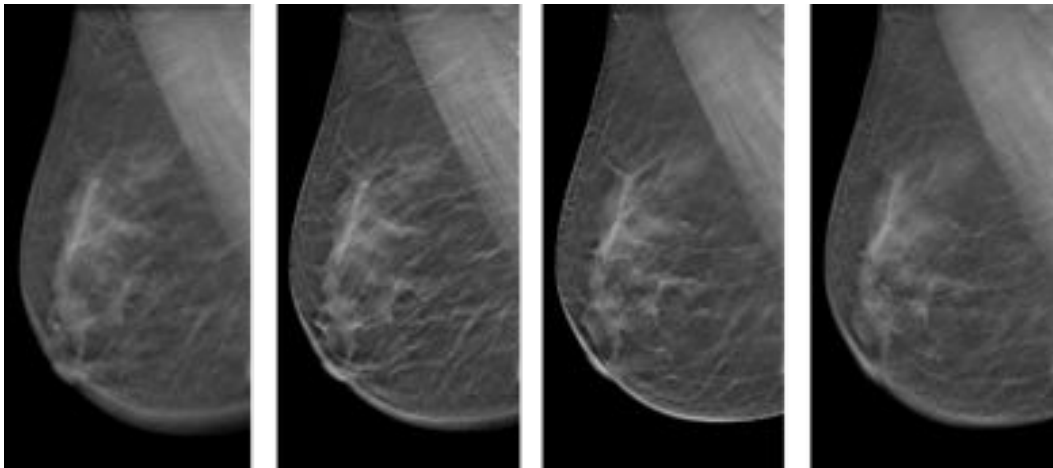
(b)

Figure 6.14: comparison of two DBT volumes of the same patient collected at different times: (a) shows four projections extracted from the volume collected in 2011; figure (b) shows the sequence of the projections (selected at the same depths inside the breast of the volume (a)) corresponding to the volume collected in 2012.

- The candidate Nipple(z) is detected in the point that exhibits a concave-convex-concave pattern (figure 6.16(d)).

The second part of the algorithm moves to a 3D analysis taking advantage of the series of contours Skin(z) and of the candidate nipple locations Nipple(z) for each plane.

The 3D representation of the whole skin surface is simply obtained merging the Skin(z) contours.

The detection of the 3D nipple location requires instead a slightly more com-
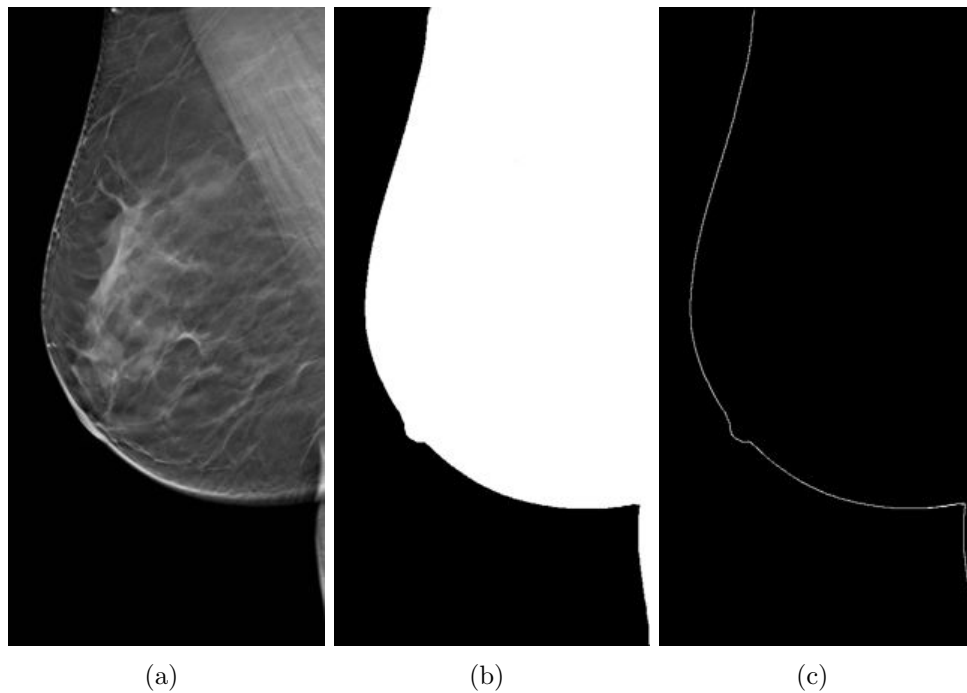
(a)   (b)   (c)

Figure 6.15: Skin detection: the DBT slice in (a) is first binarized to exclude the region outside the breast (b); the a contour is extracted from the edges of the binary image mask (c).

plicated procedure whose goal is to select the right candidate in the set of Nipple(z) locations: the rationale here is to select the z-slice in which the nipple structure is more in focus, taking advantage of the peculiar characteristic of the DBT reconstruction process.

To this aim, first the 3D normal direction to the skin surface is computed in each Nipple(z) location using a method that extends to 3D its bi-dimensional counterpart; then, along each of these directions a uniformly sampled line segment is defined as the set of locations L(z,t) with t representing the displacement from Nipple(z) towards the outgoing direction; the sampling unit of the t-coordinate is 0.2mm while its range, expressed in sampling units, is [-50;+50], resulting in a segment extension of 10 mm both in the outer and in the inner directions.

For implementation convenience it is possible to define a 2D matrix, called the normal projection image (NPI), and its coordinate system {z, t} with z and t as in the preceding description.

The algorithm fills the NPI matrix with values extracted from the original DBT volume in each locations L(z, t) by trilinear interpolation. Figure 6.17(a) shows how the bright signal corresponding to the nipple density (near the im-

(a)   (b)
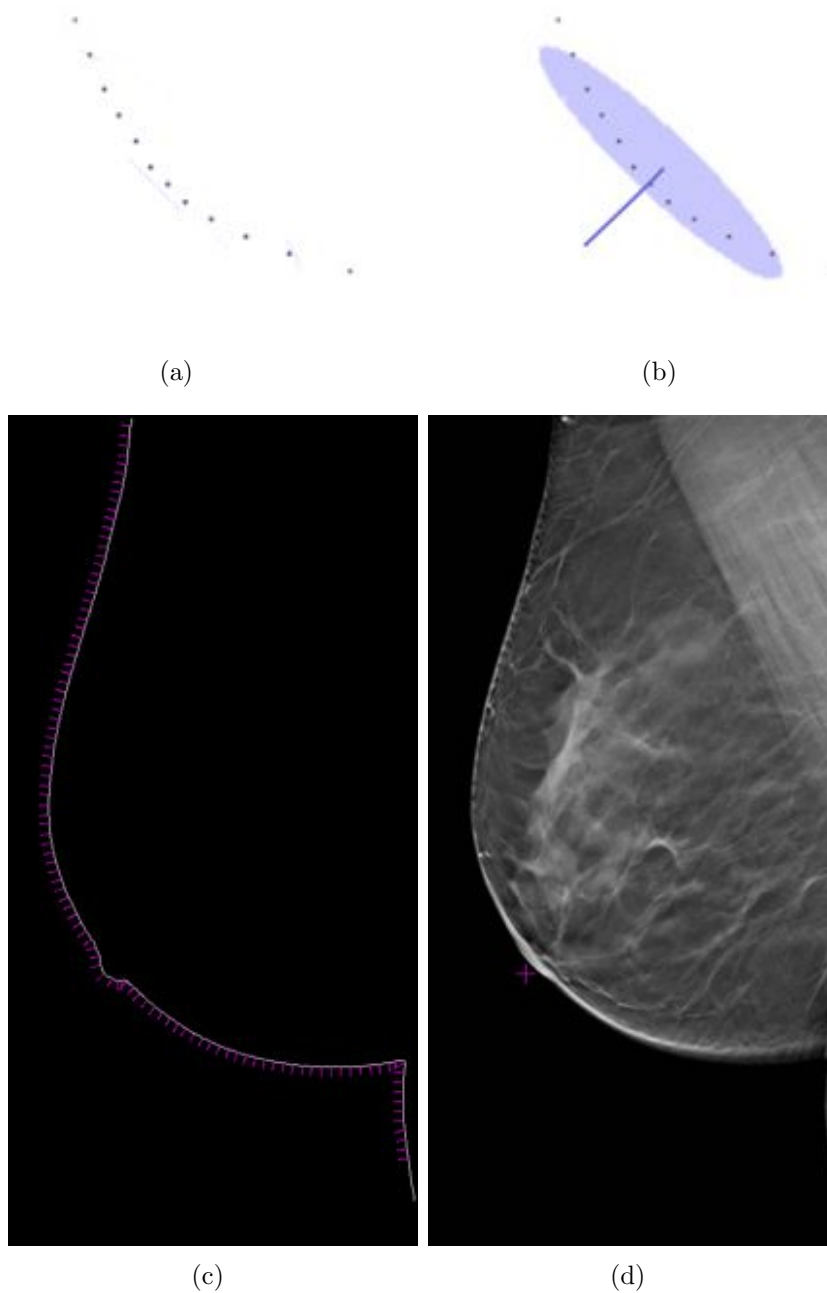

(c)   (d)

Figure 6.16: Nipple detection: the normal orientation in each contour point (a) is computed as the minor axis of the ellipse that fits its neighbors (b); the normals vectors along the whole contour are plotted in (c); the concave-convex-concave pattern of the contour characterizes the nipple location that is delineated with the purple cross in (d).

age center in the vertical direction t) spreads across all the volume depth (horizontal direction z) but it loses focus and intensity in marginal slices; to

enhance this property the NPI matrix is convolved with a fine-scale vertical Sobel filter resulting in the image shown in the second row. The nipple position is found in the point where the bright signal concentrates (figure in the bottom row); this position is finally reverse-mapped in to the original DBT reference system figure 6.17(b)).
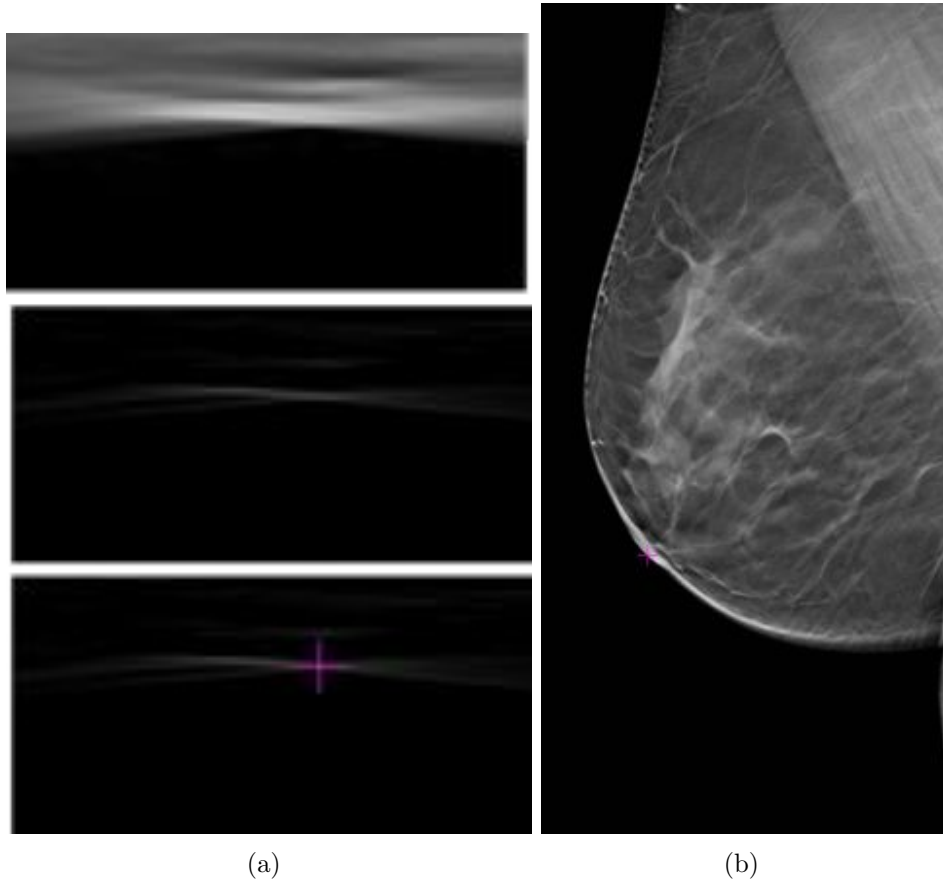

(a)   (b)

Figure 6.17: Nipple localization along the z axis: the top figure in (a) represents the matrix L(z,t) whose each column is extracted from the DBT volume by the trilinear interpolation along the normal segment to the skin surface in the location N(z); t is in the range [-50, 50] with a sampling unit of 0.2 mm both in the inner (top side of the figure) and outer (bottom side of the figure) directions. The matrix L(z,t) is convolved with a fine-scale Sobel edge detection along the vertical direction (central figure in (a)) and the location with the higher response is selected (bottom figure in (a)). In figure (b) the selected location in the L(z,t) space is remapped to the original DBT volume. The purple cross now indicate the estimated nipple position in the volume.

6.2.2 Method

Initialization

Contrary to the previous problem in which the two different sections of the breast overlap only partially, in this case the reference and template volumes are both expected to show the whole breast. Moreover, a lower amount of local distortion is expected. For these reasons the initialization block is implemented evaluating only the global average translation between the source images.

However, in this case not all the available data in a scan is meaningful: the region outside the breast is completely black and, besides the fact that it is useless for a deformable registration algorithm, it can also be misleading for a global alignment procedure based on an intensity-based metric computed over the whole volume.

To this aim the tools for the DBT nipple and skin segmentation are used. Based on the skin segmentation a 3D rectangular region of interest (ROI) is selected on the reference. More in detail the selected ROI(ref) is the rightmost 3D rectangle of a fixed size completely contained inside the breast region; a padding or minimum distance to the skin surface is also required (width, height and depth and padding of the ROI are expressed as a fraction of the whole volume dimensions).

Then a rectangular ROI(tem) of the same dimensions of ROI(ref) is searched in the template volume. Also in this case the region must be completely contained inside the skin surface but no padding is used; the selection criterion is the minimum SSD distance between ROI(tem) and ROI(ref) patches. The method used to search the best correspondence is a Nelder-Mead simplex optimization of the displacement vector T=dx,dy,dz between the ROI(ref) position and the ROI(tem) position.

In figure 6.18(a) the ROI(ref) rectangle is enlightened in yellow while the bright area around it represents the required padding. In figure 6.18(b) the best match ROI(tem) computed by the Nelder-Mead optimizer with an SSD metric is shown in bright in the template volume. Finally the displacement field on the whole reference space is initialized with the value of the estimated global translation T.

Optimization

The optimization process was instead customized by selecting the following configuration:

- For what concerns the Feature Metric the Self Similarity Context (SSC) descriptor was used. The Normalized Cross Correlation (NCC) can also be used with similar performances but SSC was selected due to its higher degree of invariance to modality data. It is worth noting, in fact, that
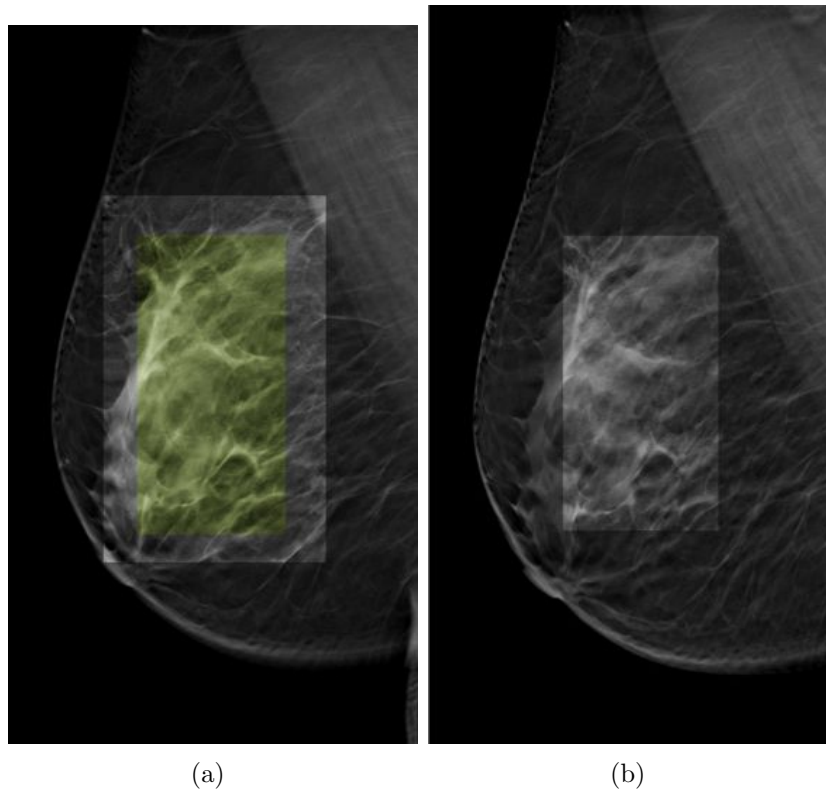
(a)   (b)

Figure 6.18: In figure (a) the rectangular patch ROI(ref) in the reference DBT volume used for the global alignment is highlighted in yellow; while the required padding from the skin surface corresponds to the brighter region around it. The patch ROI(tem) in the DBT template volume that best matches the patch ROI(ref) is shown in figure (b).

using the same descriptor for both US (the previously analysed clinical problem) and DBT (the current task) lays the foundations for the cross-modality applications.

- The Diffusion Normalizer was the choice for the regularization term. In this case there is no need to handle severe sliding motions as the breast is sensed in both the acquisitions with a similar layout. Moreover in the period of time between the first and second scan some internal structures could have changed their dimensions and shape but these kind of deformations are exactly what the physician wants to detect. A diffusion regularization model is more indicated to move such structures in the correct position based on their background context without deforming them locally.

- Constrained optimization: the correspondence between the nipple locations detected in the two source image is used as reliable and precise point-to-point mapping constraint. The strength of the constraints term

limited to this association is relatively high (0.9) with respect to the feature and normalizer terms but has only a local influence. In addition a selection of points on the skin surfaces that depart from the nipple in all directions (angular sampling of 10 degrees) at fixed radial sampling (1.0 mm) is used as a series of weak constraints (weight < 0.2). These additional constraints force the optimization process to maintain the skin-to-skin correspondence during the iterations. The weakness is motivated by the fact that, contrary to the nipple position, these point-to-point correspondence (at least for what concerns their surface curvilinear coordinates) are not precise, and the energy term they should add to the optimization functional is only intended to influence rather than effectively force it.

Results

The algorithm was tested on 8 data sets and the mean TRE error has been computed. The results were compared with the ones reported in Sinha et al. 2009 using a method based on the NMI metric that is theoretically invariant to modality as the SSC metric used here. Exploiting the comparison capabilities of the AMBRA framework the results of an implementation of same algorithm using the NCC metric that is intrinsically monomodal were reported (table 6.5). Figure 6.19(a) shows two different planes of an original DBT cou-

	avgTRE (mm)	std dev (mm)	Number of control points	Number of datasets
Proposed method	1.2	1.0	25	8
AMBRA with NCC	1.3	1.2	25	8
Reference method based on MI	1.8	1.4	15	6

Table 6.5: comparison of the results obtained with two different implementations of AMBRA with a similar application proposed in literature (Sinha et al. 2009).

ple are superimposed with the reference in the green channel and the template in red the channel. The same reference planes are shown in figure 6.19(b) with the template projected on the reference space after the registration process. In figure 6.19(c) the visible accuracy of the registration can be noticed on a detail, extracted from the top section of the first plane placed far from the nipple. The nipple alignment and the neighboring structures superimposition can be seen in figure 6.19(d).
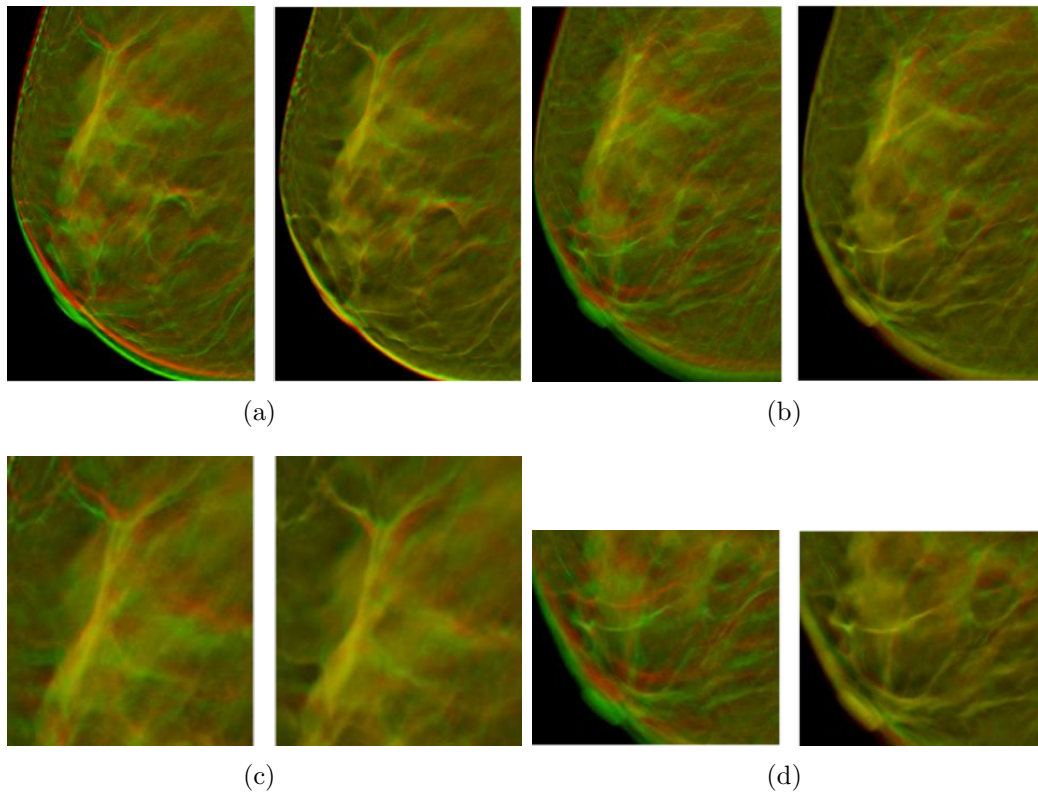
Figure 6.19: some details of overlap image obtained showing the reference volume in the red channel and the template volume in the green channel. In each sub-figure, on the left the overlap before the registration is shown, while on the right the result of the registration is reported.

Dynamic reproiection

A useful application of the proposed algorithm is the derivation of a tool able to remap dynamically a point in a DBT scan acquired in a given date with the corresponding point in another DBT scan acquired before or after the first one.

Applying the proposed algorithm with the first scan as reference and the second as template, the direct displacement field F computed after the algorithm execution is enough to obtain the desired result dynamically; in this case the reverse mapping from target to source is not required: each reference voxel $v = \{x, y, z\}$ is associated by forward mapping to the point $v' = \{x + F(x(v), y + Fy(v), z + Fz(v)\}$ in the template image.

In figure 6.20(a) a selected point in the reference is instantly depicted in the template volume figure 6.20(b).

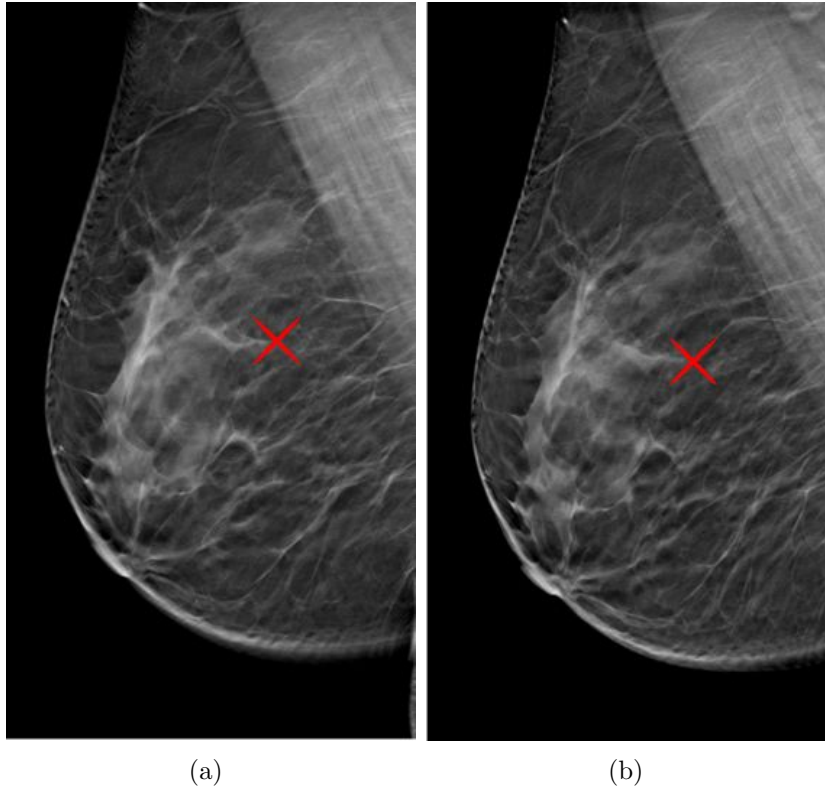(a)   (b)

Figure 6.20: result of the automatic synchronization between a DBT volume collected in 2011 and one collected in 2012. This example shows how the physicians can select a point of interest in one volume, and a corresponding location inside the other volume is automatically displayed using the estimated displacement field.

6.3 Cross-modality registration between breast Tomosynthesis and Ultrasound images

Breast cancer diagnosis can take advantage of complementary information provided by different modalities. 3DUS and DBT are two novel technologies that can be used to scan the breast for diagnostic examinations. The images provided by these modalities are quite different in terms of spatial resolution, image contrast between breast tissues, and moreover are acquired compressing the breast by different positions. Therefore, the breast diagnostic accuracy can be improved combining the data from both these modalities. However, this task involves many issues, firstly the computation of a global mapping function able to co-register the two volumes. After, the relevant differences due to local deformations need to be estimated. Hence, for the multimodal registration task the main purpose is to test if corresponding features between 3D US breast and DBT images can be detected and automatically correlated.

Both volumes were collected with Siemens equipments. As the DBT and ABUS systems are not yet widespread, especially if used in conjunction, the number of datasets available for this study is limited. Currently, only 5 datasets were used. At the current status of the research, the study is restricted to the CC view of DBT images and the AP view of 3D breast US images.

The main issue of this application is that the volumes are represented in different coordinate systems. Hence, given the breast US and DBT images, the first important step refers to compute a reliable alignment of these volumes. As reference point, the automatically detected nipple position is used. Then the skin surface segmented on DBT images and the chest wall segmentation performed on US images are used to define the effective volumes to be matched.

Multimodal breast registration is a very challenging task, that include many issues to overcome. This work represents an initial study towards a cross-modality correlation between 3D breast images.

6.3.1 Chestwall segmentation

The bones of the chest wall attenuate the US signal even more than the nipple but they are less easy to detect in breast images because of their deeper position inside the patient body; only the last few rows in the image generally show this attenuation. Anyway the automated detection of these structures can take advantage of the characteristic pattern of the chest wall with the alternation of bright and dark stripes (where the dark stripes correspond to the bones, see figure 6.21). The algorithm starts re-sampling the US volumetric data exactly with the same resolution adopted for the nipple detection (0.4 mm in all directions). In this case however only half the volume is considered, leaving out the anterior part. For simplicity the y and z axis of the original reference system are swapped in such a way that the x-y planes represent the coronal views while the z-axis is in the anterior-posterior direction (from front to back). The bright-dark pattern characterizing the bones of the chest wall is detected using a multi-scale hessian filter: more in detail the partial second derivatives are computed by convolution of the volume data with 3D box filters that approximate the Laplacian of Gaussians (Bay, Ess, et al. 2008) at each scale s (figure 6.22). The bones detector at scale s is obtained by the response of the vertical second derivative image Lyy subtracted to the horizontal second derivative image Lxx. The range of analyzed scales has been chosen taking into account the possible vertical displacement between two consecutive bones. Finally the scale associated with the maximum average response over the whole volume is selected. A similar approach is adopted to detect the orientation of bones: in most cases in fact (especially in lateral
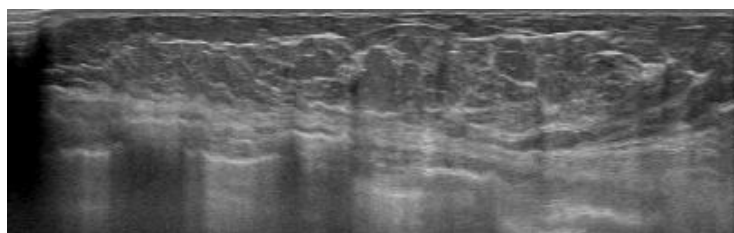
(a)
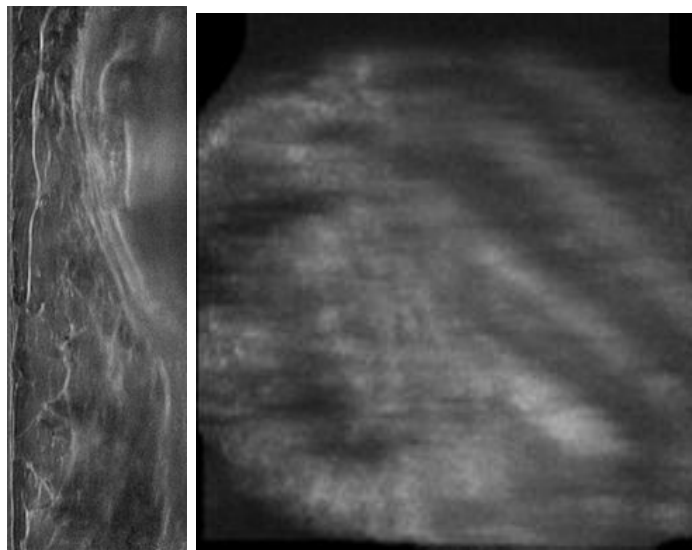
(b)   (c)

Figure 6.21: (a) transverse, (b) sagittal and (c) coronal planes extracted from a US volume showing the chest wall.

breast scans) the bones are not perfectly horizontal. Rather than considering the mixed second derivative Lxy, the volume is first masked by a cylinder with the axis corresponding to the central z axis of the volume, than rotated by different angles; for each angle the average response of the bone detector is computed and the angle with the maximum response is selected.

Once the angle and the scale are found, the corresponding response mask is equalized and the Maximally Stable Volume (MSV, already described in the nipple detection section) is run. This time the MSV algorithm is applied to the response mask and find brights regions with a cardinality compatible with the expected bones size. The borders of these regions are extracted to create a set of bone candidates (figure 6.23). The last check consists in the shape evaluation of these regions: the Random Sample Consensus (RanSaC) algorithm (Fischler et al. 1981) is used to fit the points of each of these contours with an ellipsoid model with the major axis normal to the z-direction. RanSaC is able to find the best fit handling and discovering possible outlier points.
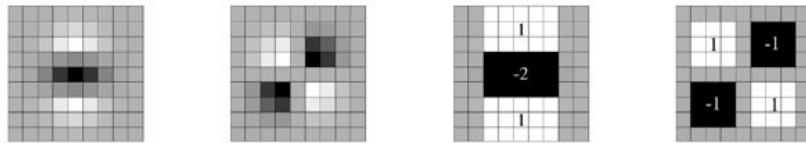
Figure 6.22: From left to right: the (discretised and cropped) Gaussian second order partial derivative in y- (Lyy) and xy-direction (Lxy), respectively; box filter approximations of Lyy and Lxy. The grey regions are equal to zero (Bay, Ess, et al. 2008).

Once found the ellipsoid three cases are possible:

- all points are inliers (their distance to the ellipsoid surface is lower that a given tolerance): the region is considered a valid bone as it is;

- a few points are outliers (caused by small signal anomalies near the bone structure): these points are rejected. If these rejected points cover a specific area the contour is fixed in that part of the surface adding samples derived from the ellipsoid model.

- A lot of points are outliers (this is the case, for example, of a bone underlying the nipple area, in fact the dark region of the nipple can merge with the dark signal inside the bone, creating an unnatural elongation of the response along the z-axis): the most distant outliers are rejected and Ransac is run another time using only the remaining points.

The resulting 3D contours are then converted back in the original reference system and up-sampled to the original volume resolution (figure 6.24). The chest wall surface can be optionally computed approximating the representation of the body trunk with a cylinder of ellipsoidal base whose orientation is fixed and the based dimensions are constrained in a range of values compatible with the human body; the cylinder is fitted to image data in such a way that it must include the detected bones structures but cannot be expanded far beyond them.

6.3.2 Method

Initialization

A multi-modal registration problem like the one faced in this section is more complex if compared with intra-modality registration, not only for what concerns the choice of a suitable feature descriptor able to evaluate correctly the similarity of the same physical structures imaged in a different manner by the two modalities, but also because the images acquisition processes are generally very different, especially for what concerns the patient pose. In the case of
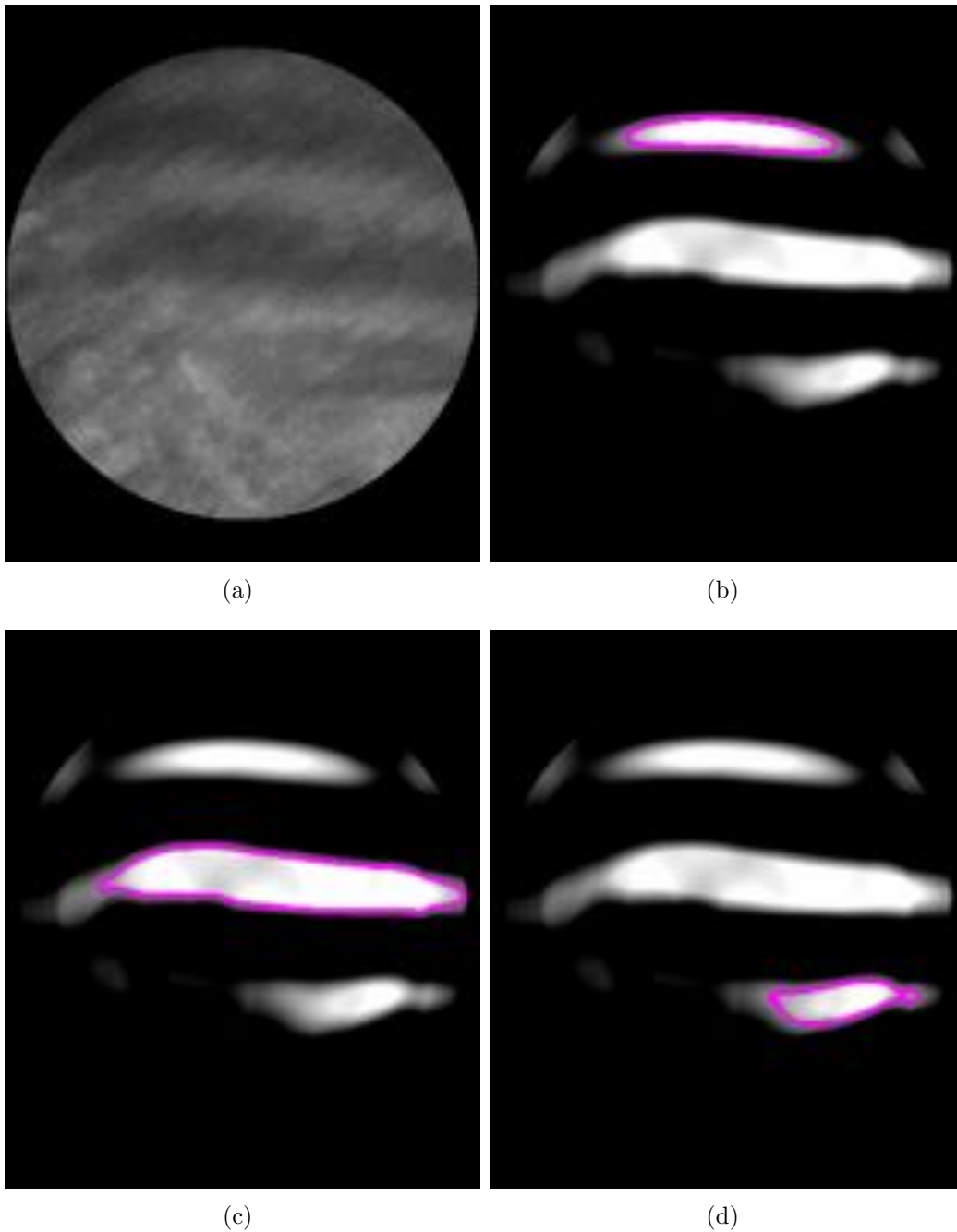
(a)

(b)

(c)

(d)

Figure 6.23: (a): example of the cylindrical mask as applied to the volume
for the segmentation of the chest wall. (b)-(d) figure show the detected bones
structures, delineated with violet contours.

3DUS and DBT scans the problem is further complicated by the fact that the
breast is squeezed in opposite directions and the deformations are very strong.
Moreover the area of overlap do not coincide. To handle all these issues the
initialization block should take advantage of all the available segmentations
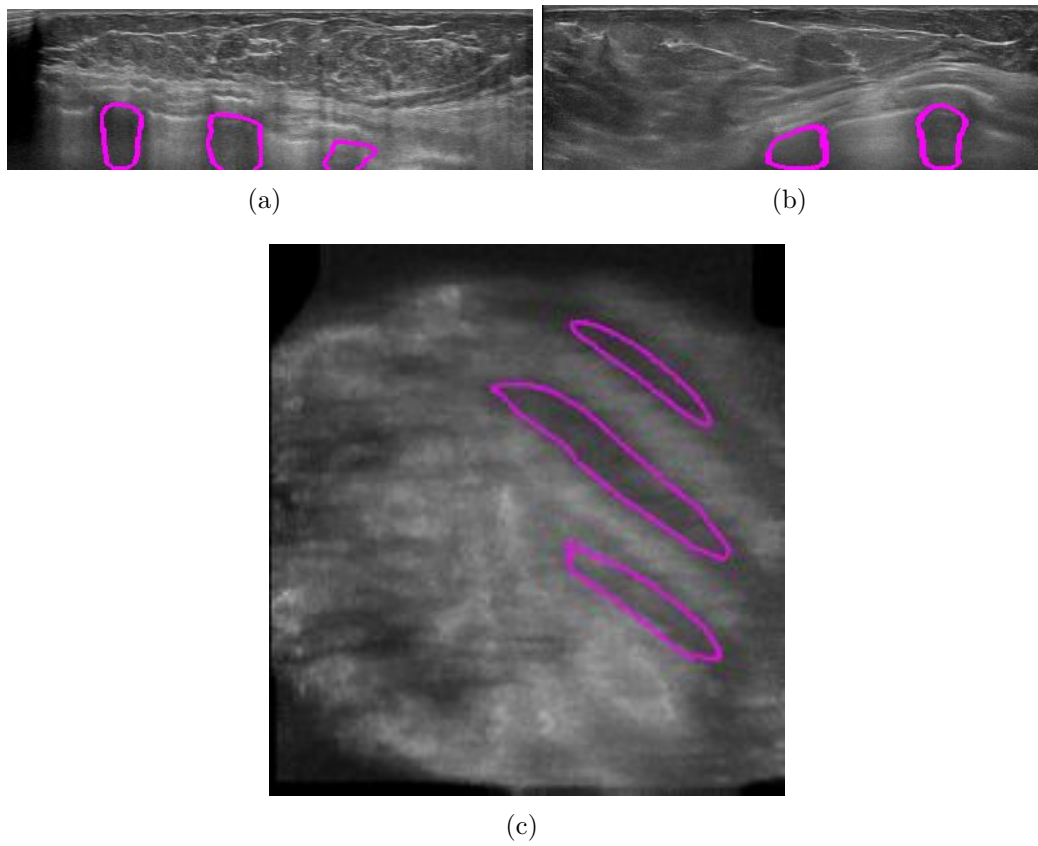
143

(a)

(b)

(c)

Figure 6.24: resulting segmentations in a transverse (a), sagittal (b) and coronal (c) section of the US volume.

and prior information.

Two slices of the DBT and US input volumes of the same breast are shown in figures 6.25(a) and 6.25(c). Figures 6.25(b) and 6.25(d) show the structures detected by the segmentation of nipple (red), skin (green) and chest wall (blue). It has to be noted that the estimated chest wall region in US should be expanded to coarsely compensate the fact that it is not detectable in DBT due to the patient pose during acquisition. Both volumes are then projected into a common geometrical representation. This target representation is obtained by a rotation of the DBT original configuration that places the chest wall-line from the vertical to the bottom horizontal axis (figure 6.26(a)), followed by the inversion of the horizontal direction (figure 6.26(b)). Since the two volumes do not completely overlap on the transverse plane as the US represents only a section of the breast, the DICOM tags of the DBT can be accessed to obtain the actual width in mm of the image. From the nipple segmentation in image coordinates it is straightforward to compute the length in mm of the two red lines in (figure 6.26(c)). A similar process is applied to US and the distances in mm of the nipple from the left and right margins of the
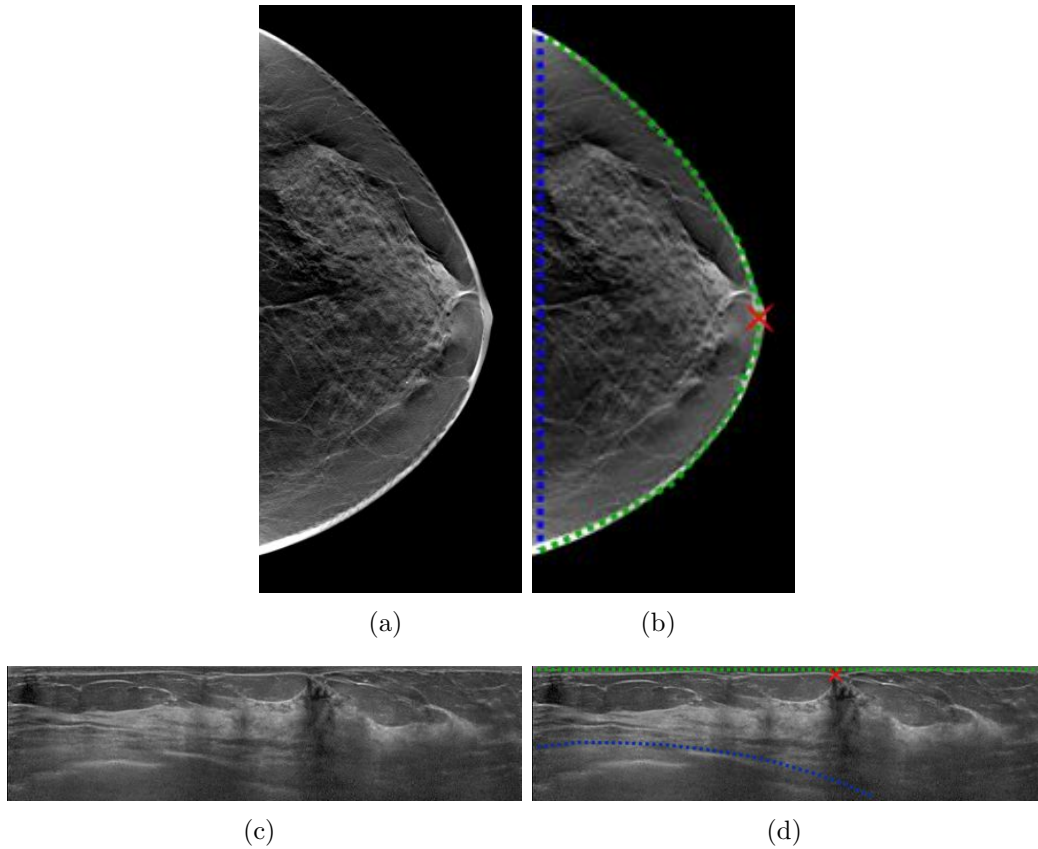
(a) (b)

(c) (d)

Figure 6.25: (a) DBT and US (c) input images. In figures (b) and (d) are displayed the automatically detected nipple position (red cross), and the skin (green line) and chest wall (blue line) segmentations.

imaged US data are reported in figure 6.26(d) as the two yellow lines. The two vertical yellow lines in figure 6.26(e), define approximately the overlap region; and hence, after having expanded these lines by a constant horizontal padding to handle marginal deformations, the target geometrical representation is cut as in figure 6.26(f). A similar procedure is adopted for the selection of the images depth common range. The DBT is projected into the target representation simply applying these transformations to image data. The US data is instead stretched using the chest wall blue line and the skin green line to drive the re-sampling operation. Two possible stretching models are available: the first one simply stretches data along the vertical direction (figure 6.27(a)); the second one tries to model the motion of the lateral region of the breast that, during the acquisition, is squeezed and tend to be pushed even more laterally. With this option the blue and green lines are re-sampled in a fixed number of equally spaced segments. The stretching of the US data is than implemented filling the target representation extracting the source data along these lines from the original US volume (figure 6.27(b)). It is worth noting that this second option is only available when the chest wall is clearly visible;
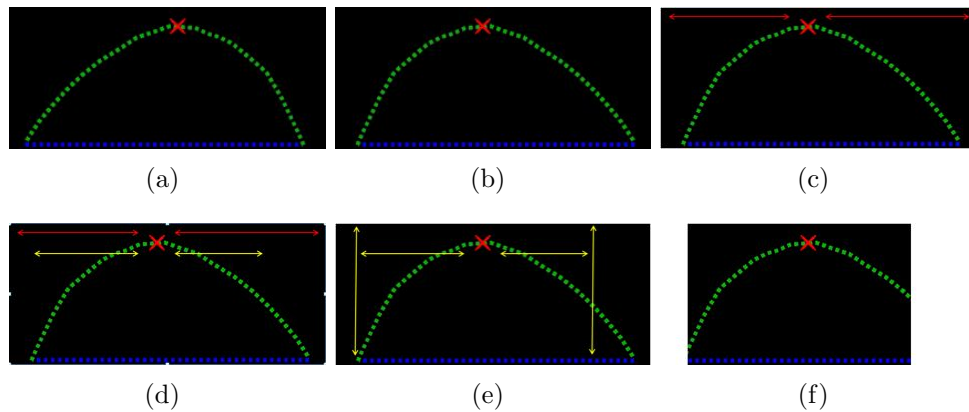
Figure 6.26: sequence of steps aimed to automatically define the effective breast volume imaged in both modalities. The input volume refers to the figure 6.22 (a) and the segmentations delineated in the 6.22 (b). The subfigures (a) report the result after: the rotation (a), the horizontally data inversion (b) of the input volume; (c) shows the distance of the nipple position to the lateral boundaries of the DBT image (measured in mm), while in (d) the yellow arrows delineate the distance (measured in mm) of the nipple to the US image boundaries; the yellow region drawn in (e) represent the breast volume the section of the breast volume estimated to be included in both modalities' field-of-view; (f) is the reduced DBT volume resulting from the matching with the US volume.

otherwise the first option is the only choice. Once found the two mapping

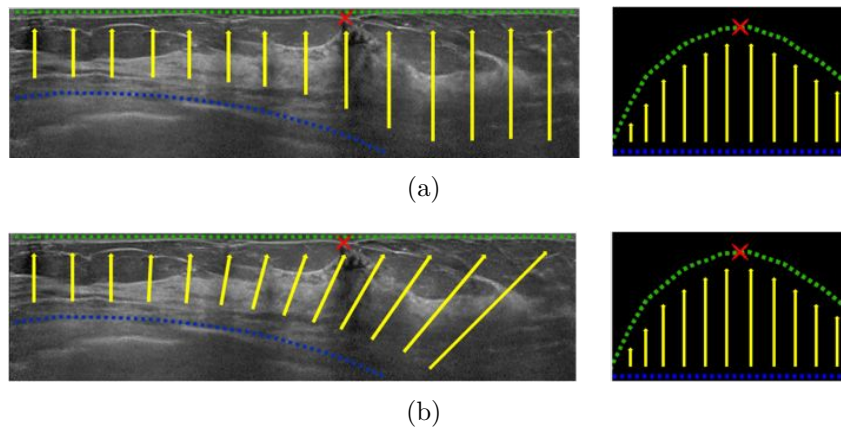

Figure 6.27: schematic representation of the feasible stretching models between the original US space (on the left) and the common target space (on the right).

function between the original volumes and the target space, the volumetric data can be used to fill both the projections; actually the intensity values are not retrieved directly from the original volumes but filtered volumes are used

instead: for what concerns DBT the anisotropic deblurring procedure proposed by Sun et al. 2007 is applied to partially compensate the out-of-focus effect and to obtain a better localization of the underling structures. The US data is instead processed by the adaptive intensity enhancement filter already used in the US-US case, to restore the signal in the dark region under the nipple.

An example of the resulting projections of the DBT and US volumes can be seen in figures 6.28(a) and 6.28(b) respectively. It is worth noting that

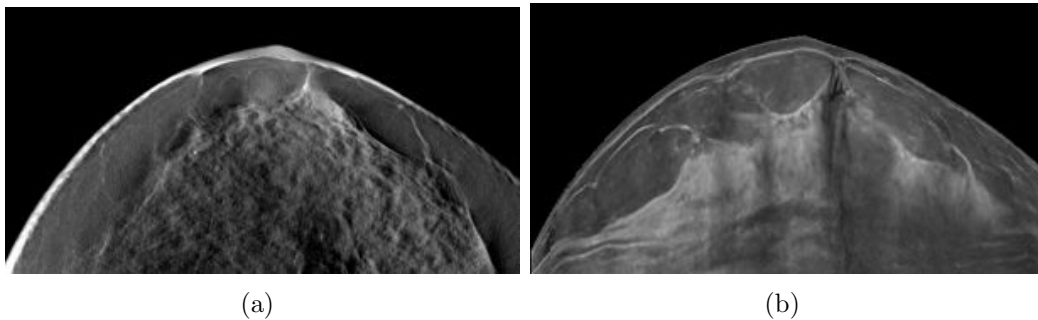(a)   (b)

Figure 6.28: results of the remapping procedure applied to the DBT data (a) and US data (b) in a common and hybrid target space.

the initialization procedure in this case does not compute a displacement field between the two original volumes, but instead transforms input data in the target representation while the initialized field is everywhere zero and is defined in the new common target space. The main motivation for this is to prepare data for the feature descriptors of the optimization phase that does not handle large rotations and reflections. This choice is preferable to the alternative that is represented by the adoption of rotation-invariant feature descriptor: in fact, as was seen in the previous chapters, the addition of more degrees of invariance makes generally the feature metric less descriptive and the complexity of breast multimodal deformable registration requires to avoid any unnecessary complication. Anyway the mappings of both volumes in the target space are well defined and invertible transformations could be stored and used for further processing.

Optimization

The customization of the optimization block in this case faced some additional difficulties, the most relevant was the choice of reliable fiducial marker correspondences between the reference (DBT) and the template (US) volume: as it was explained in the Framework description a set of reliable and precise correspondences is crucial to find automatically the best configuration of the feature descriptors and to compare them over a Testing Set. But in

this case, due to the different geometry and the different signal response, the choice of these correspondences showed a high inter-operator variability (greater than 1.0 mm on average). On the other hand the variation in the average performance of the feature descriptors and normalizers – as computed in the mono-modal tasks or in the quantitative comparison examined in the previous chapter - is generally lower than 1.0 mm. Clearly these considerations imply that the effectiveness of the automatic comparison and selection of the best modules to be used in such a task is subjected to reasonable doubts.

Anyway the configuration selected was the following:

- For what concerns the Feature Metric the Self Similarity Context (SSC) has been chosen. The comparison in this case was limited in a set of descriptors that include only SSC, MIND and NMI for their multi-modal nature.

- The Bilateral Normalizer was the choice for the regularization term as in this case the sliding motion of the most relevant structure is even more severe than in the previous tasks.

- A Multi Resolution approach was followed to make the larger structures - that are easier to detect in both modalities - drive the first part of the optimization process to an acceptable temporary solution; then a full resolution step is performed giving an opportunity to the normalizer and the feature term to adjust the correspondences where it is practicable.

- The inverse consistent constraint option was enabled to take advantage of the correct fits and mediate the weaknesses of the direct and inverse registration processes.

- Different constraints were enabled to exploit the segmentations at hand: the nipple location is set as a strong constraint due to its good localization; the skin surface is only used as a weak constraint to prevent its dis-alignment, but, at the same time, to let the inner structures glide on the surface; the chest-wall was not used at all as it is not detectable in the DBT.

Figure 6.29 depicts an example of the field optimization: figure 6.29(a) shows the result after 100 iterations of the deformable registration and in figure 6.29(b) the more relevant changes in the deformable field are highlighted.

Multimodal correlation

Given the direct field F from the DBT in the target space to US in the target space and the two invertible mapping functions MapDbt and MapUs from original DBT to target DBT and original US to target US respectively, a

(a)   (b)
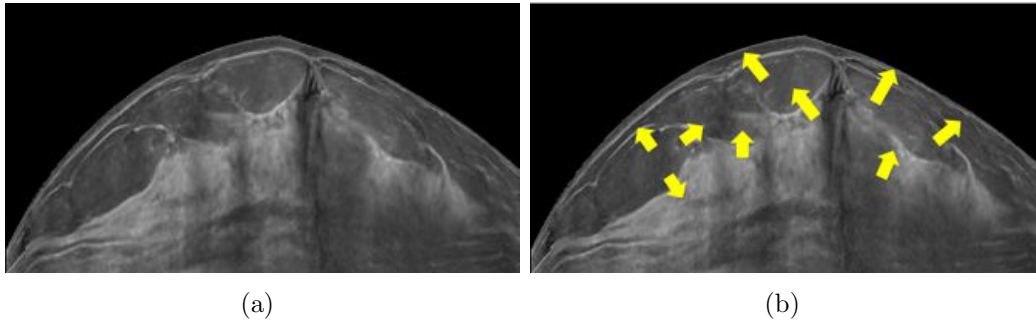
Figure 6.29: the transformed US data after the registration and schematic representation of the estimated displacement field (b).

location in the original DBT can be translated to the corresponding location in the original US volume through the composition:

$$MapDbt \circ F \circ MapUs^{-1} \tag{6.8}$$

while a location in the original US can be projected in the corresponding location of the original DBT through:

$$MapUs \circ G \circ MapDbt^{-1} \tag{6.9}$$

Clearly this possibility is limited to locations in the overlapping region between the two regions. A possible straightforward application is to develop a tool similar to the one proposed for the dynamic re-projection of two DBT scans acquired in different times; but in this case the correlation is obtained between different modalities.

In figure 6.30(a)-6.30(b) some rectangular regions in different planes of a DBT are selected. The corresponding location in the US image are found and shown in figure 6.30(c) and 6.30(d).

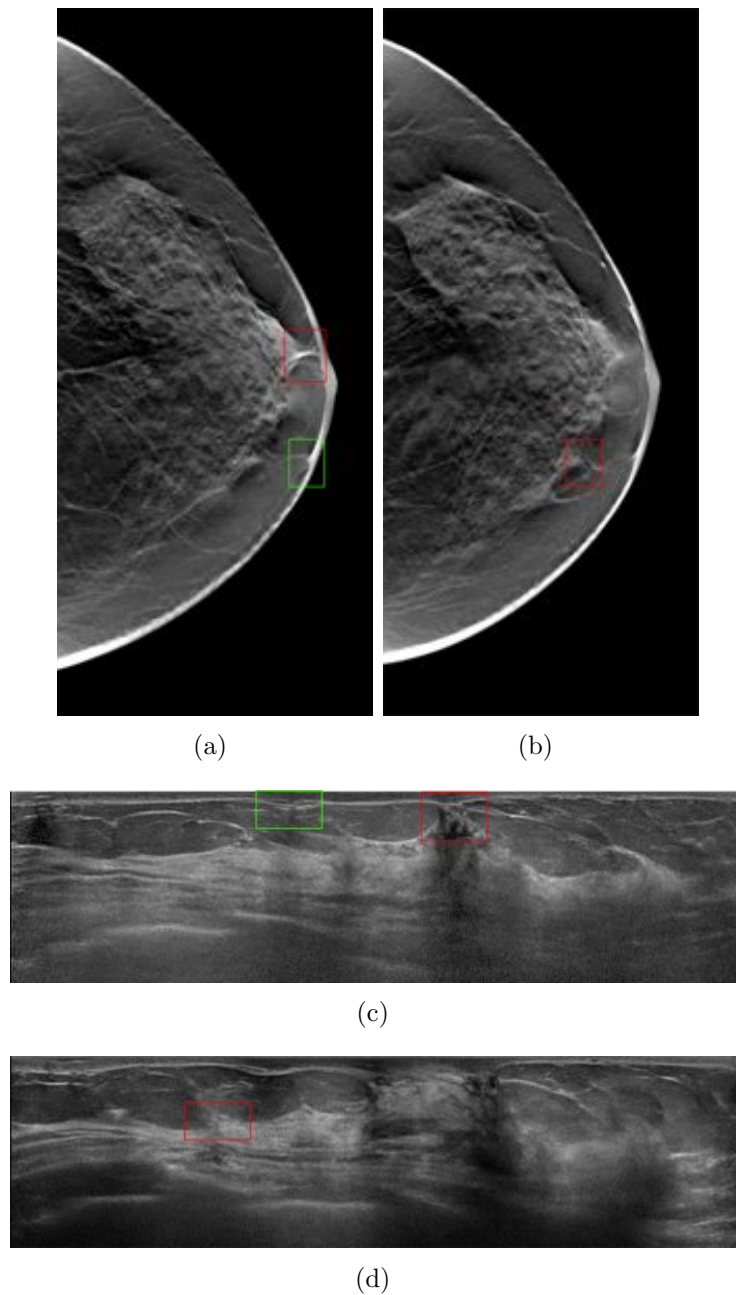Figure 6.30: figure (a) and (b) show two selected frames inside the DBT input volume. Figure (c) and (d) report the corresponding locations found into the US image using the spatial mapping function estimated between the two volumes. The red and green boxes shown how work the automatic synchronization of corresponding regions selected inside the DBT and US volumes.

Chapter 7

Conclusions

Breast screening is widely accepted as an effective tool to significantly reduce the incidence of breast cancer deaths in women population as it allows a detection of tumors in early stages, making the treatment more likely to be successful. Due to the differences in the breast tissue composition in different women and the different kind and nature of cancers that can affect this organ, it is worldwide accepted that there is no imaging technique able to always detect all types of cancer. The imaging technique (or modality) most commonly used is the X-ray mammography that is expected to show the differences in tissue density and discriminate the cancer from the fat tissue; young women, or more specifically women whose breast is made up mostly by glandular or connective tissue, are instead tested with other imaging techniques as these kind of tissues have an appearance similar to cancer in mammograms. US and MRI are two possible alternatives techniques able to detect cancers that are not visible or well characterized in X-ray imaging. It is commonly accepted that each type of exam has its own advantages and drawbacks among which the rate of false positive findings is taken into consideration. For these reasons a comparative analysis of the different kinds of available exams for a single patient is considered an effective solution and it is recommended by the American College of Radiology: it not only allows to detect a larger number of cancer types but also helps to prevent over diagnosis (and consequently over treatment) taking advantage of the compound information that comes from all the modalities together. The same considerations apply to the comparison of exams of the same patient acquired in different times. Until now these indications are commonly implemented by physicians in different ways depending on the modalities and their specific acquisition protocol. The major difficulty in this type of comparison is the different geometrical representation of the imaged data: X-ray mammography represents an integration of the X-signal along the projection axis, while hand-held US probes acquire different bi-dimensional sections of the breast that, without any other additional information, are impossible to be reliably remapped on the mammogram. In recent years the technological advances have brought the opportunity to move the

151

breast imaging modalities from 2D to 3D, with Digital Breast Tomography (the 3D version of X-ray mammograms) and 3DUS scanners (the advanced application of the bi-dimensional US probes). These novel modalities were described in detail in Chapter 2. The volumetric information allows a great improvement in terms of cancer detection and characterization; nevertheless, the novel 3D techniques recently introduced in the clinic context are not yet supported by adequate visualization tools to actually exploit their potentiality. This thesis investigated a fascinating frontier for the technical challenges to be addressed and for new applications with a straightforward impact in the medical context.

The aim of this work is the evaluation of the feasibility of automated registration methods for breast applications. The main difficulty is that the breast is completely made up of soft tissue whose motion under different compressions is hard to predict and to model also for the presence of the sliding motion between breast sub-structures. Another additional complication occurs whenever a cross-modality registration is required as the same structures are represented differently in the two images. Deformable registration algorithms (whose theory has been studied and briefly reported in Chapter 3) are the mathematical tools that can solve these kind of problems. These algorithms, whose aim is to find a dense displacement field between different location in the two images to be registered, had been used in a wide range of contexts and in a great number of medical application. They are generally based on a compromise between two opposite forces: the first one moves the locations of an image towards corresponding locations in the other image, the second one tries to maintain the deformation acceptable and realistic. In most cases they take advantage of some sort of pre-segmentation of expected organs, bones, or rigid structures; in other cases they assume a smooth motion of all the physical structures involved. These partial constraints on the overall displacement estimation are used to help the otherwise ill-posed mathematical problem. However these options are not always available for breast applications. So the study conducted for this research had to find alternative techniques to make the deformable registration approach well suited and feasible in this specific clinical context.
In this thesis two 3D imaging modalities (DBT and 3DUS) have been considered and three specific tasks analyzed. These tasks were chosen as representative of the major clinical needs related to breast images examination: registration of partially overlapped volumes with different compressions in breast tissues (especially for US images), spatial correlation of exams taken at different times (developed for DBT images) and finally the most challenging cross-modality registration (between DBT-US images).
An experimental setup for the controlled acquisition of 3D US data was presented in Chapter 4. A complete set of software tools was organized in a framework (implemented in C language) that allows the automated comparison and

the straightforward derivation of automated multi-modal breast-oriented registration algorithms (AMBRA). The architectural design (see Chapter 5) of this framework is general enough to handle all the mentioned breast-related tasks, supporting both single and cross modality registration, similar of very different patient positions, complete or only partial overlapped images. On the other hand it implements a number of specialized techniques customized for breast; most of them are extensions of more classical and general tools; thus a modular design has been adopted to easy substitute the classical with the novel techniques. It is conceptually divided in two main processing blocks: the first one is the more flexible and customizable part, that should potentially embed in a single tool all the a priori information about the specific task of interest. This part of the framework provides a coarse alignment of the images. The second block represents the deformable registration procedure implementing a generic non-parametric registration scheme. It has a modular structure that enables an easy replacement, testing and comparison of different implementations of the two forces that drive the registration field to the final solution: the metric adopted to describe the similarity between corresponding location in the images, and the rules that should be applied to obtain a deformation compatible with a realistic solution.

Therefore, preliminary evaluations were made to investigate the response of the implemented tools in case of US data. The assessments were carried out isolating simple structural features in US image data (extracted from ad-hoc home-made phantoms) and adopting simple transformation models taking advantage of the US experimental acquisition system, avoiding the relevant deformations impressed by the commercial 3D US system. Afterwards, due to the good results obtained in the preliminary tests, the implemented techniques have been integrated into the framework and compared on clinical datasets. Novel methods recently proposed in literature were compared to the state-of-the-art techniques as reported in the last part of Chapter 5. From this analysis a novel feature metric, named Self Similarity Context (SSC), emerged. It showed a more accurate response than other common descriptors both with US and DBT clinical datasets used for testing. Moreover, this feature descriptor provides an inherently modality-independent representation of data, allowing its application even for cross-modality registration tasks. For what concern the other force involved in the registration method, the so-called normalizer, the recently proposed bilateral filtering field regularization emerged as a suitable solution to handle the sliding motion of inner breast structures. This method showed a higher accuracy in case of US data, but similar behavior to the classical diffusion model for the DBT testing datasets, where the sliding motions are less evident.

Finally the framework with all these novel techniques has been evaluated for the three clinical tasks taken as reference study cases. For each task a specific algorithm has been derived from AMBRA as described in Chapter 6. In this Chapter modality-specific breast segmentation algorithms are presented;

these are an essential part included in the AMBRA initialization block, as reference methods for nipple (one method for US data and on for DBT), skin (for DBT images) and chest wall (for US images) segmentation. The most efforts were indeed dedicated to the customization of the initialization blocks: the alignment procedures combined with the segmentation methods represent the original part of the work. Another innovative contribution is the study of both the best parameters configuration of the feature descriptors and the deformation model applied to breast US and DBT data. The evaluations were performed on a representative testing datasets using a set of pairs of control points, properly annotated for the validation of the results. The most critical issue in this workflow is related to the fact that, in some cases, the physicians were not completely confident in a point-wise localization of a sufficient number of landmarks points. This is due to the lack of reliable and marked features inside the breast that make difficult a pointwise manual correlation, and, in the meantime, because 3DUS and DBT are recent technologies and require a highly specific training. This critical issue is encumbered especially in the case of the cross-modality comparison. The difficulties encountered in the annotation of correspondences between two breast volumes could affect the reliability of the evaluations and suggests to study more robust and specific criteria to validate the application of deformable registration techniques to breast images. On the other hand, this problem highlights the effective impact that an accurate and automatic images correlation could have supporting the breast images interpretation.

The first customized algorithm derived from AMBRA referred to the mapping of corresponding structures in the overlap region between two 3D US breast sections collected with the Siemens ABVS system. In this case the large sliding motion and the different compression of the breast represented the main issue to face. A novel alignment procedure that makes use of a hybrid method (based on both intensity-based patches and robust localized feature points matches) was implemented; this aimed to determine the correspondences between the middle-sized structures in the two images in order to define an initial displacement field. After these structures got sufficiently close, a multi-resolution deformable registration procedure was derived from the framework, to go through subsequent refinements of the displacement field at different scales of analysis. Satisfactory results were obtained using the proposed algorithms, compared with those achieved using other methods in their general-purpose implementation. As expected the normalizer with the best performance was the one based on bilateral filtering. The SSC feature descriptor has been confirmed as the best one for US data. It results probably from its ability to make the description of a feature less affected by noise, compared to other descriptors. This application led to propose a novel technique for blending the two 3D US sections in a target space characterized by an extended field-of-view. The element of innovation is represented by the

use of a deformable transformation model in the stitching procedure. The obtained results suggest the possibility to process the partial input volumes in a single-view output volume to examine. The good quality of the resulting images, that does not present intensity distortions, confirmed at the same time the accuracy in the displacement field estimate.

The second faced task concerns the mapping of two DBT exams of the same patient acquired at different times. In this case a lower local displacement of the breast tissues was expected than the previous application, nevertheless physiological or pathological changes could be occurred. The aim was to compensate the spatial deformations of the unchanged structures due to the different breast positioning, in order to highlight the anatomical differences. Indeed, in this application the deformable registration field was not used to modify and merge the input volumes, increasing the chance to occlude or change some clinically relevant features; on the contrary it was used for an automatic correlation of the volumes in order to provide a simultaneous visualization of corresponding regions. Although extensive evaluations on clinical datasets with evident anatomical differences was not yet performed, in order to validate the method a quantitative assessment of the registration errors of corresponding control points was achieved. For DBT data the diffusion operator was selected than the bilateral regularization; this means that a globally smooth deformation was preferred and anomalies such as a tumor growth are not differently treated, but rather follow their neighborhood's deformation model.

The last tackled task is the most challenging one as it regards the derivation of a multi-modal registration algorithm between DBT and US scans of the same breast. Here the alignment procedure takes advantage of all the segmentations tools embedded in AMBRA. A chest wall segmentation for US images was develop to automatically exclude the regions visible in the US dataset but not scanned in the DBT exam. A complex remapping of the data was then applied to both volumes and the registration was computed on the intermediate space representing the overlapping regions imaged in both the modalities. An application of the estimated displacement field was finally proposed; the automatic synchronization of corresponding locations between the volumes was shown. Up to now the achievable accuracy does not allow an actual pointwise correlation, however the identification of small volumes of interest is feasible. In this case the difficulty to mark reliable pairs of control points, widespread in the volumes for an accurate validation of the results, was more evident: the inter-operator average precision was estimated (>1 mm) greater than the expected variation in the average TRE metric used to evaluate the performances of AMBRA. Nevertheless these initial results are promising enable further investigations. To the best of my knowledge, methods suitable for a comparative study have not yet been proposed in literature for this application.

In summary, the present thesis is a pioneering work in the breast image registration field, especially in 3D breast US registration. The AMBRA framework was developed *ab initio* and the reported results are promising, even in comparison with the limited papers published in this field. The most interesting recent techniques (MIND, SSC, bilateral regularization, etc.) were included and customized for the three proposed breast applications. The derived algorithms demonstrated to be successful and suitable to overcome the challenging issues showing by the breast, related to its soft nature and its inherently non-rigid behavior, and it's intrasubject variability related to the hormone cycle. In this work specific implementations for US and DBT images were developed; however, the flexible and modular architectural structure of the registration framework enable further integrations, including other modalities like MRI. On the basis of this work a research project was developed and submitted to Friuli-Venezia Giulia Region obtaining positive evaluations.

Bibliography

Arsigny, V et al. (2006). "A Log-Euclidean Framework for Statistics on Diffeomorphisms". In: *MICCAI 2006*, pp. 924–931.

Avants, BB et al. (2008). "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain". In: *Med Image Anal* 12.1, pp. 26–41.

Bajcsy, R and S Kovacic (1989). "Multiresolution elastic matching". In: *Comput Vision Graphics Image Processing*. Vol. 46. 1, pp. 1–21.

Bay, H, A Ess, et al. (2008). "Speeded-Up Robust Features (SURF)". In: *Computer Vision and Image Understanding* 110.3, pp. 346–359.

Bay, H, T Tuytelaars, and L Van Gool (2006). "SURF: Speeded Up Robust Features". In: *ECCV 2006*. Ed. by A Leonardis, H Bischof, and A Pinz. Vol. 3951. Springer Berlin Heidelberg, pp. 404–417.

Berg, WA, AI Bandos, et al. (2016). "Ultrasound as the Primary Screening Test for Breast Cancer: Analysis From ACRIN 6666". In: *J Natl Cancer Inst* 108.4.

Berg, WA, Z Zhang, et al. (2012). "Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast risk". In: *JAMA* 307, pp. 1394–1404.

Berg, WA et al. (2006). "Operator dependence of physician-performed whole-breast US: lesion detection and characterization". In: *Radiology* 241, pp. 355–365.

— (2008). "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer: results of the first-year screen in ACRIN 6666". In: *JAMA* 299.18, pp. 2151–2163.

Boyd, NF and et.al. (2007). "Mammographic Density and the Risk and Detection of Breast Cancer". In: *N Engl J Med* 356, pp. 227–36.

Breastcancer.org. http://www.breastcancer.org/.

Brem, RF, L Tabar, and et.al. (2015). "Assessing Improvement in Detection of Breast Cancer with Three-dimensional Automated Breast US in Women with Dense Breast Tissue: The SomoInsight Study". In: *Radiology* 274.3, pp. 663–73.

Brock, KK and Deformable Registration Accuracy Consortium (2010). "Results of a multi-institution deformable registration accuracy study (MIDRAS)". In: *Int J Radiat Oncol Biol Phys* 76.2, pp. 583–596.

Brown, M and DG Lowe (2007). "Automatic Panoramic Image Stitching using Invariant Features". In: *International Journal of Computer Vision* 74.1, pp. 59–73.

Buades, A, B Coll, and JM Morel (2005). "A review of image denoising algorithms, with a new one". In: *SIAM interdisciplinary journal* 4.2, pp. 490–530.

Buchberger, W et al. (1999). "Incidental findings on sonography of the breast: clinical significance and diagnostic workup". In: *AJR Am J Roentgenol* 173, pp. 921–927.

Burt, PJ and EH Adelson (1983). "A Multiresolution Spline with Application to Image Mosaics". In: *ACM Trans. Graph.* 2.4, pp. 217–236.

Bushberg, JT et al. (2002). *The Essential Physics of Medical Imaging.* Lippincott Williams & Wilkins.

Cancer Research UK. http://www.cancerresearchuk.org/.

Chae, EY et al. (2015). "Comparison of Lesion Detection in the Transverse and Coronal Views on Automated Breast Sonography". In: *J Ultrasound Med* 34, pp. 125–135.

Chang, JM et al. (2011). "Breast cancers initially detected by hand-held ultrasound: detection performance of radiologists using automated breast ultrasound data". In: *Acta Radiol* 52, pp. 8–14.

Checka, CM et al. (2012). "The relationship of mammographic density and age: implications for breast cancer screening". In: *AJR Am J Roentgenol* 198.3, pp. 292–5.

Chen, CF, MH Chen, and HT Li (2007). "Fully Automatic and Robust Approach for Remote Sensing Image Registration". In: *CIARP 2007.* Ed. by Springer. L Rueda, D Mery, and J Kittler, pp. 891–900.

Chen, TK et al. (2012). "Importance of transducer position tracking for automated breast ultrasound: Initial assessments". In: *Ultrasonics Symposium (IUS).* IEEE International, pp. 2623–2626.

Chen, Y and X Ye (2010). "Inverse Consistent Deformable Image Registration". In: *The Legacy of Alladi Ramakrishnan in the Mathematical Sciences.* Ed. by K Alladi, JR Klauder, and CR Rao. Springer Berlin Heidelberg.

Cho, N et al. (2006). "Differentiating benign from malignant solid breast masses: comparison of two-dimensional and three-dimensional US". In: *Radiology* 240, pp. 26–32.

Chou, Y et al. (2007). "Automated Full-field Breast Ultrasonography: The Past and The Present". In: *Journal of Medical Ultrasound* 15.1.

Christensen, GE, X Geng, et al. (2006). "Introduction to the non-rigid image registration evaluation project (nirep)". In: *International Workshop on Biomedical Image Registration*, pp. 128–135.

Christensen, GE and HJ Johnson (2001). "Consistent image registration". In: *IEEE Transactions on Medical Imaging* 20.7, pp. 568–582.

— (2003). "Invertibility and transitivity analysis for nonrigid image registration". In: *Electron Imaging*. Vol. 12, pp. 106–17.

Coman, IL, JA Mandel, and K Baum (2004). "Intermodality Nonrigid Breast-Image Registration". In: IEEE.

Crum, WR and et al. (2003). "Zen and the art of medical image registration: correspondence, homology, and quality". In: *NeuroImage*. Vol. 20. 3, pp. 1425–1437.

Crum, WR, T Hartkens, and DLG Hill (2004). "Non-rigid image registration: theory and practice". In: *The British Journal of Radiology* 77, S140–S153.

Dang, PA et al. (2014). "Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations". In: *Radiology* 270.1, pp. 49–56.

Dobbins, JT and DJ Godfrey (2003). "Digital x-ray tomosynthesis: current state of the art and clinical potential". In: *Phys Med Biol* 48.19.

Donoser, M and H Bischof (2006). "3D Segmentation by Maximally Stable Volumes (MSVs)". In: *ICPR 2006*. Vol. 1. 18th International Conference on Pattern Recognition, pp. 63–66.

D'Orsi, CJ et al. (2013). *ACR BI-RADS Atlas, Breast Imaging Reporting and Data System*. 5th. American College of Radiology.

Ducan, W and CR Kerr (1976). "The curability of breast cancer". In: *Br Med J* 2.6039, pp. 781–783.

Feldman, MK, S Katyal, and MS Blackwood (2009). "US artifacts". In: *Radiographics* 29.4, pp. 1179–89.

Ferlay, J et al. (2013). *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Tech. rep. International Agency for Research on Cancer.

Ferrant, M et al. (2000). "Deformable modeling for characterizing biomedical shape changes". In: *Discrete Geometry for Computer Imagery*, pp. 235–248.

Fischler, MA and RC Bolles (1981). "Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Graphics and Image Processing* 24.6.

Fornefett, M et al. (1999). "Radial Basis Functions with Compact Support for Elastic Registration of Medical Images". In: *IVC* 19, pp. 1–2.

Fraunhofer MEVIS. http://www.mevis.fraunhofer.de/.

GE Healthcare. http://www3.gehealthcare.com/.

Georgii, J, F Zohrer, and HK Hahn (2013). "Model-based position correlation between breast images". In: *Medical Imaging*. Vol. Proc SPIE 8670.

Girardi, V et al. (2013). "Breast ultrasound in 22,131 asymptomatic women with negative mammography". In: *Breast* 22.5, pp. 806–9.

Giuliano, V and C Giuliano (2012). "Using automated breast sonography as part of a multimodality approach to dense breast screening". In: *Journal of Diagnostic Medical Sonography* 28.4, pp. 159–165.

Goshtasby, AA (2005). *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley-Interscience.

Guo, Y et al. (2006). "Breast image registration techniques: a survey". In: *Medical and Biological Engineering and Computing* 44.1-2, pp. 15–26.

Gur, D et al. (2009). "Digital breast tomosynthesis: observer performance study". In: *American Journal of Roentgenology* 193, pp. 586–591.

Heinrich, MP, M Jenkinson, M Bhushan, et al. (2012). "MIND: Modality Independent Neighbourhood Descriptor for Multi-modal Deformable Registration". In: *Medical Image Analysis* 16.7, pp. 1423–1435.

Heinrich, MP, M Jenkinson, BW Papiez, et al. (2013). "Towards realtime multimodal fusion for image-guided interventions using self-similarities". In: *MICCAI 2013*. Springer Berlin Heidelberg, pp. 187–194.

Helvie, MA (2010). "Digital Mammography Imaging: Breast Tomosynthesis and Advanced Applications". In: *Radiol Clin North Am* 48.5, pp. 917–929.

Helvie, MA, HP Chan, et al. (2009). "Digital breast tomosynthesis mammography: successful assessment of benign and malignant microcalcifications". In: *Radiologic Society of North America 95th Scientific Assembly and Annual Meeting*.

Helvie, MA, MA Roubidoux, et al. (2007). "Tomosynthesis mammography versus conventional mammography: comparison of breast masses detection and characterization". In: *Radiological Society of North America 93rd Scientific Assembly and Annual Meeting*.

Hermosillo, G, C Chefd'Hotel, and O Faugeras (2002). "Variational methods for multimodal image matching". In: *Int. J. Comput. Vis.* 50, pp. 329–343.

Hill, DLG and et al. (2001). "Medical Image Registration". In: *Phys Med Biol* 46.1, pp. 1–45.

Holden, M (2008). "A review of geometric transformations for nonrigid body registration". In: *IEEE Trans Med Imaging* 27.1, pp. 111–128.

Hong, D et al. (2012). "Multimodality pictorial review of breast diseases affecting skin layer: Imaging features of benign versus malignant diseases". In: *ECR*.

Hooley, RJ, LM Scoutt, and LE Philpotts (2013). "Breast Ultrasonography: State of the Art". In: *Radiology* 268.3.

Hopp, T and et al. (2012). "Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity- based optimization". In: *Med. Image Anal.*

Horn, BKP and BG Schunck (1980). *Determining Optical Flow. Rapport technique*. Tech. rep.

Ibanez, L et al. (2003). *The itk software guide second edition updated for itk version 2.4*.

Inciardi, M (2012). "Breast Ultrasound: What's New in 2013". In: *Imaging Technology News* 10, pp. 39–41.

Isobe, S et al. (2011). "Detectability of breast lesions under the nipple using an automated breast volume scanner: comparison with handheld ultrasonography". In: *Jpn J Radiol* 29.5, pp. 361–365.

Jacob, JMD (2012). "Dense Breast Tissue: Supplemental Screening". In: *Imaging Technology News* 11, pp. 46–49.

Janssens, G et al. (2009). "Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy". In: *Med Phys* 36.9, pp. 4268–4276.

Kaplan, SS (2001). "Clinical utility of bilateral whole-breast us in the evaluation of women with dense breast tissue". In: *Radiology* 221.3, pp. 641–649.

Kelly, KM, J Dean, and WS Comulada (2010). "Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts". In: *Eur Radiol* 20, pp. 734–742.

Kelly, KM, J Dean, SJ Lee, et al. (2010). "Breast cancer detection: radiologists' performance using mammography with and without automated whole-breast ultrasound". In: *Eur Radiol* 20.11, pp. 2557–2564.

Kelly, K and G Richwald (2011). "Automated Whole-Breast Ultrasound: Advancing the Performance of Breast Cancer Screening". In: *Seminars in Ultrasoun, CT and MRI*. Ed. by Elsevier. Vol. 32. 4, pp. 273–280.

Kim, H et al. (2014). "Comparison of conventional and automated breast volume ultrasound in the description and characterization of solid breast masses based on BI-RADS features". In: *Breast Cancer* 21, pp. 423–428.

Klein, A et al. (2009). "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration". In: *Neuroimage* 46.3, pp. 786–802.

Kolb, TM, J Lichy, and JH Newhouse (2002). "Comparison of the performance of screening, mammography, physical examination and breast US and evaluation of factors that influence them: An analysis of 27,825 patient evaluations". In: *Radiology* 225, pp. 165–175.

Kopans, DB (1999). "Breast-cancer screening with ultrasonography". In: *Lancet* 354, pp. 2096–2097.

Kremkau, FW (1998). *Diagnostic ultrasound principles and instruments*. Ed. by Saunders. 5th.

— (2010). *Sonography Principles and Instruments*. Elsevier.

Lander, M and L Tabar (2011). *Automated 3-D Breast Ultrasound as a Promising Adjunctive Screening Tool for Examining Dense Breast Tissue*. Ed. by Elsevier. Vol. 46. Seminars in Roentgenology.

Lester, H and SR Arridge (1999). "A survey of hierarchical nonlinear medical image registration". In: *Pattern Recogn* 32.1, pp. 129–149.

Lin, X, J Wang, and F Hana (2011). "Analysis of eighty-one cases with breast lesions using automated breast volume scanner and comparison with handheld ultrasound". In: *European Journal of Radiology* 81, pp. 873–878.

Loeckx, D et al. (2010). "Nonrigid image registration using conditional mutual information". In: *IEEE Transactions on Medical Imaging* 29, pp. 19–29.

Lo, JY, NC Durham, and JA Baker (2006). "Breast tomosynthesis: assessing patient compression, comfort, and preference". In: *Radiologic Society of North America 92nd Scientific Assembly and Annual Meeting.*

Maintz, JBA and MA Viergever (1998). "A survey of medical image registration". In: *Med. Image Anal.* 2.1, pp. 1–36.

Matas, J et al. (2002). "Robust wide baseline stereo from Maximally Stable Extremal Regions". In: *Proceedings of the British Machine Conference.* Ed. by D Marshall and PL Rosin, pp. 36.1–36.10.

Maurer, CRJ and et al. (1997). "Registration of head volume images using implantable fiducial markers". In: *IEEE Transactions on Medical Imaging* 16.447-462.

Meire, H et al. (2001). *Abdominal and general ultrasound.* Ed. by Churchill Livingstone. 2nd. Vol. 1.

Mendelson, EB and CE Tobin (1995). "Critical pathways in using breast US". In: *Radiographics* 15, pp. 935–945.

Middleton, WD, AB Kurtz, and BS Hertzberg (2004). *Ultrasound.* Ed. by Mosby.

Modersitzki, J (2004). *Numerical methods for image registration.* Ed. by Oxford University Press.

National Cancer Institute. http://www.cancer.gov.

Nelder, JA and R Mead (1965). "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4, pp. 308–313.

Paci, E et al. (2002). "Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990–96". In: *Br J Cancer* 87, pp. 65–69.

Padilla, F and et al. (2013). "Breast mass characterization using 3D automated ultrasound as an adjunct to digital breast tomosynthesis: A pilot study". In: *J Ultrasound Med* 32.1, pp. 93–104.

Papiez, BW et al. (2014). "An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration". In: *Medical Image Analysis* 18.8, pp. 1299–1311.

Pisano, ED, C Gastonis, and E Hendrick et al. (2005). "Diagnostic Performance of Digital versus Film Mammography for Breast Cancer Screening". In: *NEJM* 353, p. 1773.

Pluim, JPW, JBA Maintz, and MA Viergever (2000). "Image registration by maximization of combined mutual information and gradient information". In: *IEEE Transactions on Medical Imaging* 19.8.

Protter, M et al. (2009). "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction". In: *IEEE Transactions on Image Processing* 18.1.

Rafferty, EA et al. (2013). "Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial." In: *Radiology* 266.1, pp. 104–13.

Rancov, V et al. (2005). "An algorithm for image stitching and blending". In: *Proc of SPIE*. Vol. 5701.

Rodriguez-Vila, B et al. (2010). "Three-dimensional quantitative evaluation method of nonrigid registration algo- rithms for adaptive radiotherapy". In: *Med Phys* 37.3, pp. 1137–1145.

Rohlfing, T (2006). "Transformation model and constraints cause bias in statistics on deformation fields. In Medical image computing and computer-assisted intervention". In: *MICCAI 2006* 9, p. 207.

— (2012). "Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable". In: *IEEE Transactions on Medical Imaging* 31.2, pp. 153–163.

Rohr, K et al. (1996). "Point-based elastic registration of medical image data using approximating thin-plate splines". In: *Visualization in Biomedical Computing* 1131, pp. 297–306.

Rotten, D, JM Levaillant, and L Zerat (1999). "Analysis of normal breast tissue and of solid breast masses using three-dimensional ultrasound mammography". In: *Ultrasound Obstet Gynecol* 14, pp. 114–124.

Rueckert, D and et al. (1999). "Nonrigid registration using free-form deformations: application to breast MR images". In: *IEEE Transactions on Medical Imaging* 18, pp. 712–721.

Ruiter, NV et al. (2004). "Model-based registration of X-ray mammograms and MR images of the female breast". In: *Nuclear Science Symposium Conference Record*. Vol. 5. IEEE, pp. 3290–3294.

Schmidt-Richberg, A et al. (2012). "Fast explicit diffusion for registration with direction-dependent regularization". In: *International Workshop on Biomedical Image Registration (WBIR)*. Ed. by Springer-Verlag. 5th. BM Dawant et al.

Sechopoulos, I (2013). "A review of breast tomosynthesis. Part I. The image acquisition process". In: *Med Phys* 40.1.

Shin, HJ et al. (2011). "Automated ultrasound of the breast for diagnosis: interobserver agreement on lesion detection and characterization". In: *AJR Am J Roentgenol* 197, pp. 747–754.

Siemens Healthcare. http://usa.healthcare.siemens.com/ultrasound/radiology/acuson-s2000-abvs-ultrasound-system.

Sinha, SP et al. (2009). "Image Registration for Detection and Quantification of Change on Digital Tomosynthesis Mammographic Volumes". In: *AJ of Roentgenology* 192.2, pp. 384–387.

Sivaramakrishna, R (2005). "3D breast image registration: a review". In: *Technol. Cancer Res. Treatment* 4.1, pp. 1–10.

Sivaramakrishna, R and R Gordon (1997). "Detection of breast cancer at a smaller size can reduce the likelihood of metastatic spread: a quantitative analysis". In: *Acad Radiol* 4.1, pp. 8–12.

Smith, A (2011). *http://hologiced.com/assets/Design Considerations in Optimizing a Breast Tomosynthesis System.pdf*.

Smith, C (2006). "On Vertex-Vertex Systems and their use in geometric and biological modelling". PhD thesis. Department of Computer Science.

Smith, RA et al. (2004). "The randomized trials of breast cancer screening: what have we learned?" In: *Radiol Clin North Am* 42, pp. 793–806.

Sotiras, A, C Davatzikos, and N Paragios (2013). "Deformable Medical Image Registration: A Survey". In: *IEEE Transactions on Medical Imaging* 32.7, pp. 1153–1190.

Sotiras, A and N Paragios (2012). *Deformable Image Registration: A Survey*. Tech. rep. RR-7919. Inria.

Stavros, AT et al. (1995). "Solid breast nodules: use of sonography to distinguish between benign and malignant lesions". In: *Radiology* 196, pp. 123–134.

Stoblen, F et al. (2011). "First Evaluation of the Diagnostic Accuracy of an Automated 3D Ultrasound System in a Breast Screening Setting". In: *Anticancer Research* 31.

Studholme, C and et al. (1999). "An overlap invariant entropy measure of 3d medical image alignment". In: *Pattern Recognition* 32.71-86.

Sun, AX, AW Land, and AR Samala (2007). "Deblurring of tomosynthesis images using 3D anisotropic diffusion filtering". In: *Proc of SPIE*. The International Society of Optical Engineering.

Tabar, L and et al. (1995). "Efficacy of breast cancer screening by age: New results from the swedish two-county trial". In: *Cancer* 75.10, pp. 2507–2517.

Tabar, L et al. (2000). "The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long-term follow-up". In: *Radiol Clin North Am* 38, pp. 625–651.

— (2001). "Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality". In: *Cancer* 91.9, pp. 1724–1731.

Tao, L and VK Asari (2006). *An efficient illuminance-reflectance nonlinear video stream enhancement model*.

Teertstra, H et al. (2010). "Breast tomosynthesis in clinical practice: initial results". In: *European Radiology* 20.1, pp. 16–24.

Thirion, JP (1998). "Image matching as a diffusion process: an analogy with Maxwell's demons". In: *Medical Image Analysis* 2.3, pp. 243–260.

Tomasi, C and R Maduchi (1998). "Bilateral Filtering for Gray and Color Images". In: *Proc of IEEE*. International Conference on Computer Vision.

Unlu, M and et al. (2005). "Deformable model for 3D intramodal nonrigid breast image registration with fiducial skin markers". In: *Proc. of SPIE Medical Imaging*. Vol. 5747.

Urschler, M, S Kluckner, and H Bischof (2007). "A Framework for Comparison and Evaluation of Nonlinear Intra-Subject Image Registration Algorithms". In: *MICCAI*.

Vachon, CM and et al. (2007). "Mammographic density, breast cancer risk and risk prediction". In: *Breast Cancer Res* 9, pp. 217–226.

Vercauteren, T et al. (2009). "Diffeomorphic demons: efficient non-parametric image registration". In: *Neuroimage* 45.1, S61–S72.

Wang, H et al. (2012). "Differentiation of benign and malignant breast lesions: A comparison between automatically generated breast volume scans and handheld ultrasound examinations". In: *European Journal of Radiology* 81, pp. 3190–3200.

Wojcinski, S et al. (2011). "The automated breast volume scanner (abvs): initial experiences in lesion detection compared with conventional handheld b-mode ultrasound: a pilot study of 50 cases". In: *Int J Womens Health* 3, pp. 337–346.

Woods, RP (2000). "Handbook of medical imaging: processing and analysis validation of registration accuracy". In: ed. by Academic, pp. 491–497.

World Cancer Research Fund International. http://www.wcrf.org/.

Yaffe, MJ (2008). "Mammographic density. Measurement of mammographic density". In: *Breast Cancer Res* 10.3, p. 209.

Yaghjyan, L et al. (2011). "Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to tumor characteristics". In: *J Natl Cancer Inst* 103.15, pp. 1179–1189.

Zhang, X et al. (2014). "Improved digital breast tomosynthesis images using automated ultrasound". In: *Med Phys* 41.6.

Zitova, B and J Flusser (2003). "Image registration methods: a survey". In: *Image and Vision Computing* 21.11, pp. 977–1000.