Mikko Karjalainen

Time-series registration of thermal breast images

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 21/09/2018

Thesis supervisor:

Prof. Nuutti Hyvönen

Thesis advisor:

Ph.D. Janne Tamminen

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Mikko Karjalainen

Title: Time-series registration of thermal breast images

Date: 21/09/2018 Language: English Number of pages: 6+46

Department of Neuroscience and Biomedical Engineering

Professorship: Biomedical Engineering Code: SCI3059

Supervisor: Prof. Nuutti Hyvönen

Advisor: Ph.D. Janne Tamminen

Dynamic infrared imaging (DIRI) is an emerging technology for the early detection of breast cancer. In this method time-series of thermal breast images are obtained. The patient motion in the time-series can distort the DIRI analysis in such a way that the detection of breast cancer becomes impossible. Image registration can be used to eliminate the patient motion from the time-series data. In this thesis, two different registration algorithms were tested: Thirion's demons algorithm and an algorithm based on an affine transformation. Furthermore, a combined method where the affine method is used as a pre-registration step for the demons method was tested. The algorithms were implemented with Matlab and their performance in the task of registering a time-series of thermal breast images was evaluated using four different performance metrics. The registration algorithms were implemented for time-series data of 20 healthy (no malignant lesions) subjects. The demons method outperformed the affine method and is recommended as a suitable tool for time-series registration of thermal breast images. The combined method achieved slightly improved results compared to the demons method but with significantly increased computation time.

Keywords: Medical imaging, Image registration, Breast cancer, Demons algorithm, Dynamic infrared imaging

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Mikko Karjalainen

Työn nimi: Rinnan lämpökuvien aikasarjojen stabilointi

Päivämäärä: 21/09/2018 Kieli: Englanti Sivumäärä: 6+46

Neurotieteen ja lääketieteellisen tekniikan laitos

Professuuri: Lääketieteellinen tekniikka Koodi: SCI3059

Valvoja: Prof. Nuutti Hyvönen

Ohjaaja: FT Janne Tamminen

Dynaaminen lämpökuvantaminen on lupaava menetelmä rintasyövän aikaiseen havaitsemiseen. Menetelmässä rinnoista otetaan lämpökuvien aikasarja. Kuvantamisen aikana tapahtuva potilaan liike voi vaikeuttaa aikasarjan analysointia niin, että rintasyövän tunnistaminen ei ole mahdollista. Liike voidaan poistaa aikasarjasta kuvastabiloinnin avulla. Tässä työssä tutkittiin kahta kuvastabilointiin kehitettyä algoritmia: Thirionin demons-algoritmia ja algoritmia, joka perustuu affiiniin muunnokseen. Lisäksi tutkittiin yhdistettyä menetelmää, jossa affiinia menetelmää käytetään esiaskeleena demons-menetelmälle. Algoritmien laskenta toteutettiin Matlabilla. Algoritmien tuottaman tuloksen laatua arvioitiin neljällä erillisellä laatumittarilla. Testidatana käytettiin aikasarjoja, jotka oli kuvattu 20:stä terveestä (ei pahanlaatuisia kasvaimia) potilaasta. Demons-menetelmä osoittautui affiinia menetelmää paremmaksi. Demons-menetelmää voidaan suositella rintojen lämpökuvien aikasarjojen stabilointiin. Yhdistetty menetelmä tuotti hiukan parempia tuloksia kuin demons-menetelmä, mutta vaati huomattavasti enemmän laskenta-aikaa.

Avainsanat: Lääketieteellinen kuvantaminen, Kuvastabilointi, Rintasyöpä, Demons-menetelmä, Dynaaminen lämpökuvantaminen

Preface

About six months ago I started working with this thesis. Back then I did not completely know what to expect. I had heard some stories of thesis projects that took forever to finish and were really stressful for their authors. Luckily, my project ended up running quite smoothly from start to finish. This was in large ways thanks to my supervisor Nuutti Hyvönen and advisor Janne Tamminen. Your help during this project was priceless for me.

I would also like to thank Imaqen Ltd for offering me the possibility to work full-time with this thesis. It was a pleasure working with all of you at the office and I hope that you achieve your goals in the future.

The end of my thesis project also marks the end of my studies. Back in 2011 when I arrived to Otaniemi as a freshman, I had no idea how big of an impact the student life would have on me. I learned a lot on lectures and exercise sessions but not nearly as much as I learned from student activities. I ended up organizing parties for hundreds of people, dancing on the stage of student musicals and hanging from a crane at the middle of the night among other things. Some of the most memorable projects I was lucky to be a part of include Tempaus2016 and the 70th anniversary of the Guild of Physics. What really made all these projects so enjoyable for me were all the wonderful people I got to share them with. Special thanks for making my study years the best time of my life so far go to Raati3, TJ'15, Fyysikkospeksi, JTMK'16 and IE'14. I hope I can share a beer or two with you even after many years to come.

Last but not least I would like to thank my family and especially my parents Kirsi and Pekka for their immense support during my studies. During my time at Aalto you may not have been able to help me with my homework anymore, but I could always count on the fact that you would help me with basically any other issue I could encounter. It is also largely thanks to you that I even ended up in Aalto. You taught me to value education and made it easy for me to succeed in the studies before university.

Otaniemi, 21/09/2018

Mikko Karjalainen

Contents

Symbols and abbreviations

Abbreviations

DIRI	Dynamic infrared imaging
SSD	Sum of squared differences
NMI	Normalized mutual information
MR	Magnetic resonance
CT	Computed tomography
PET	Positron emission tomography
LPT	Log-polar transform
FFD	Free-form deformation model
BBO	Breast boundary overlap
FFT	Fast Fourier transform
STH	Spatial thermal homogeneity

1 Introduction

Breast cancer is the most common type of malignancy in women as about 10% of women will be diagnosed with breast cancer before the age of 75 [1]. In 2012 some 1.7 million women were diagnosed with the disease [2]. Survival rates for breast cancer have improved vastly over the decades. In the 1960s only 35% of women diagnosed with breast cancer in the United States were alive after 10 years. In the mid 1990s the 10-year survival rate in the US had already improved to 77%. Despite the major improvement in survival rates, nearly half-a-million women still die of breast cancer each year [3].

Early detection of breast cancer has been proven to be crucial for the survival of the patient [4]. For this reason many countries have established national screening programs for breast cancer. The screening for breast cancer is done by mammography. A mammogram is a special series of X-rays taken from the breasts. According to studies, women who have routine mammograms have a 10-25% lower chance of dying of breast cancer compared to those women who do not have mammograms [4].

Despite the obvious benefits of mammography, it is not a flawless screening method. A major downside of mammography is overdiagnosis. Overdiagnosis refers to the detection of cancers that would have not become apparent during the patients lifetime without the screening [5]. Overdiagnosis causes anxiety to patients and leads to unnecessary medical operations like surgery. It is clear that overdiagnosis occurs but the extent of overdiagnosis is difficult to determine.

Other downsides are also associated with mammography. As mammography uses X-rays, the screening subjects receive a dose of ionizing radiation. Even though the doses are low for a single patient, screenings may cause harm on the population level [5]. Patients experience pain during mammography as the breast is compressed and flattened during the procedure. This even deters some women from attending further screenings [5]. Sometimes a cancer is not found with mammography. This is especially an issue with younger women who have more dense breast tissue as mammography has reduced sensitivity to dense breast tissue [6].

Due to these shortcomings of mammography as a breast cancer screening technique, there is a clear need for an improved screening method. A promising method for breast cancer screening is dynamic infrared imaging (DIRI) [7–9]. In this technique a sequence of thermal images of the subjects breasts are acquired. The spatiotemporal temperature data is then analyzed to detect possible cancerous lesions. The method is painless, does not require the use of ionizing radiation and has better sensitivity than mammography for dense breast tissue.

A major challenge for DIRI is patient motion during imaging [7]. The motion is mainly caused by breathing, but other types of motion are also possible. The patient motion distorts the temporal temperature data making the analysis of the data difficult. The distortions of data can be so severe that detecting cancerous lesions becomes impossible.

One way to prevent patient motion would be to tell patients to hold their breath. This is however not the optimal solution as it can be difficult for the patients to hold their breath for a sufficient time. The preferred solution is a form of digital

image processing known as image registration. Image registration is the alignment of two or more images and it can be used to compensate for the patient motion [10].

Image registration is a common procedure in medical image analysis [10]. Often it is a preprocessing step before further analysis of the images. Besides motion reduction the applications in medical imaging include, for example, the alignment of images acquired using different imaging modalities, registration of images from different subjects and registration of images taken with a long time difference [10].

There is an extremely wide variety of registration algorithms that have been developed for the purpose of medical imaging. Different algorithms have been developed for basically every organ in the human body [10]. A registration algorithm tries to find a geometric transformation that best aligns the images under registration. The algorithm seeks a transformation that optimizes a similarity criterion between the images. In consequence, the basic building blocks of a registration algorithm are a geometric transformation, a similarity measure and an optimization strategy. The algorithms are often classified based on the type of transformation they seek or on the type of similarity measure used. [10]

In this thesis an overview of the algorithms used in medical image registration is presented. The thesis will mainly discuss the different types of algorithms on a general level without going into details of specific algorithms. However, algorithms that are promising for the purpose of registration of thermal breast images will be covered in more detail.

The main algorithm to be tested in this thesis was Thirions demons algorithm [11]. As a comparison method an algorithm based on an affine transformation was used. The affine method was also used as a pre-registration step before the implementation of the demons algorithm to see if the results of the demons method could be improved. All of the registrations were performed with Matlab.

The performance of the algorithms in the task of registering thermal breast images was evaluated using 4 different metrics. Two of these were based on the overall image intensity values: sum of squared differences (SSD) and normalized mutual information (NMI). Two other metrics were based on image features: one was based on the edges detected in the images and the other on corner points detected in the images.

Time-series of thermal breast images from 20 different subjects were used. All of the time-series were registered with each of the three methods and values for all of the evaluation metrics were calculated. The results are presented graphically in chapter 4 of this thesis. Statistical testing was also performed to confirm that the observed differences in the performances of the methods were statistically significant.

This thesis is organized as follows. In section 2, we survey the background of different image registration algorithms. In section 3, the different similarity measures are presented. Furthermore, the numerical experiments carried out in this thesis are described in this section. Section 4 focuses on the results. Finally, in section 5, we will discuss about the significance of the obtained results and about the most important issues for future research.

2 Background

2.1 Medical Image Registration

Image registration can be defined as the process of aligning two or more images [10]. The goal is to find a transformation that best aligns the images or some structures of interest in the images. Medical images acquired using different sensors, at different times or from different viewpoints can contain complementary information that will aid in diagnosis or treatment. Image registration is an important tool in medical image analysis as it helps to uncover that information.

Image registration has many applications in the field of medical image analysis. It can be used to align images acquired using two different imaging modalities. For example the anatomical information from a computed tomography (CT) or magnetic resonance (MR) image can be combined with the functional information obtained from a positron emission tomography (PET) image [10]. The result is a fusion of the two images, for example a PET/CT or PET/MR image (Fig. 1), that can improve the diagnosis and/or treatment of the patient [12].

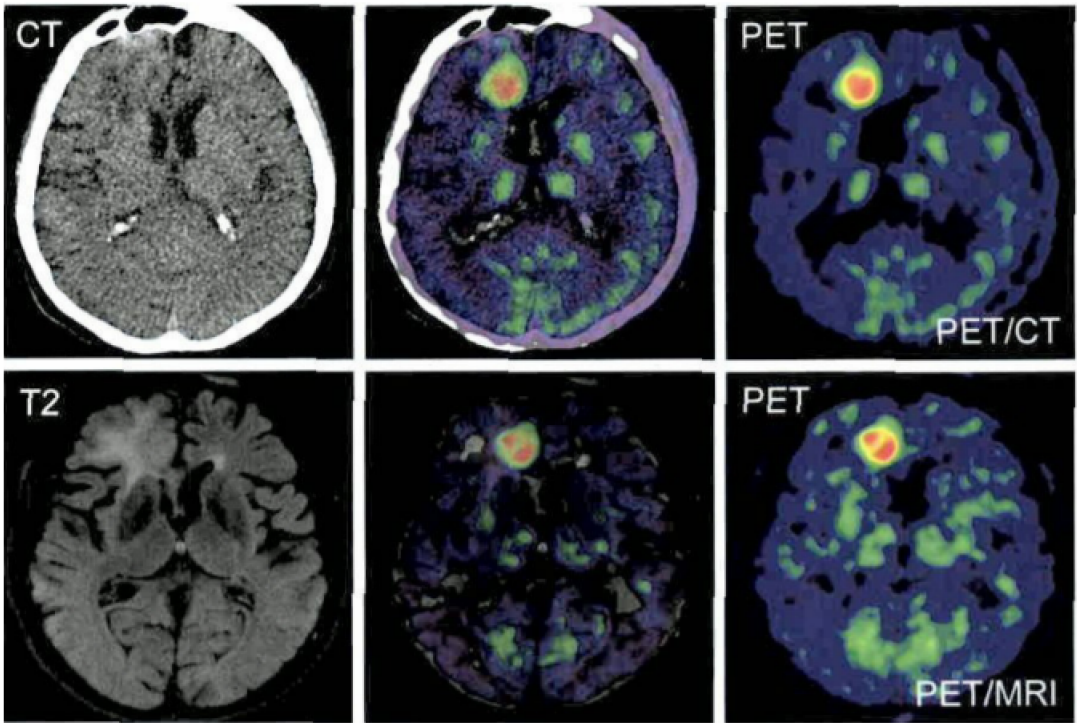

Figure 1: PET/CT and PET/MRI images of a patient with glioblastoma multiforme. Top row presents the CT (left), PET (right) and the PET/CT (center) images. Bottom row presents the T2-weighted MRI (left), PET (right) and the PET/MRI (center) images. The plain CT- and MR-images do not show the glioblastoma visible in the PET images. The combined images show the glioblastoma with increased anatomical accuracy compared to a simple PET-image. Figure is from [13].

In this thesis the point of interest is the registration of images acquired at different times with the same imaging modality. The time points can be close to each other or far apart depending on the application. Time-series of images can for example provide information about organ movement or disease progression. In the context of DIRI we are interested in the temporal behaviour of temperature. During a time sequence of images there is bound to be some unwanted patient motion that can make further analysis of the image sequence more difficult [14]. Image registration can be used to reduce this kind of unwanted motion.

The methods used for image registration can be classified in terms of nine distinctive characteristics [15,16]. These are (1) dimensionality: the registration can be performed between 2D images, between 3D images or 2D images can be registered to 3D images. (2) Nature of the registration basis: it can be extrinsic, intrinsic or non-image based. (3) Nature of the transformation: the transformation can be rigid or non-rigid. In the rigid case, the transformation only consists of translation and rotation. Non-rigid transformations contain more degrees of freedom, possibly millions of them. (4) Domain of transformation: the transformation might be global or local. In a global transformation the image is transformed as a whole. This is the case with rigid transformations. In local transformations different parts of the image can be transformed differently. (5) Degree of interaction: the registration algorithm might require varying levels of user interaction. Some algorithms are completely automatic, some require little interaction, like placing markers on the images, and some require user-guidance during the whole process. (6) Optimization procedure: there exist a variety of different optimization strategies and one has to choose one suitable for the registration problem at hand. (7) Modalities involved: registration can be performed between images acquired with the same imaging modality (mono-modal registration) or with different modalities (multi-modal registration). (8) Subjects involved: registration can be performed on images acquired from the same subject/patient (intra-subject registration) or on images acquired from different subjects/patients (inter-subject registration). Registration can also be performed to a statistical atlas. (9) Objects involved: the images can be of different organs or parts of the human body, like for example breasts, abdomen or brain.

In the context of this thesis many of these characterizations are predetermined. The case we are interested in is mono-modality imaging of breasts. The dimensionality is spatiotemporal 2D/2D and the desired level of interaction is automatic. The main division between the methods considered in this thesis is based on the nature of the transformation. Several rigid and non-rigid methods for image registration will be covered.

Another important classification in image registration is the division between feature based and intensity based methods [17]. This division is based on the chosen similarity measure. In the registration process one tries to find the geometric transformation that optimizes the similarity measure. For example a registration algorithm might try to find a rigid geometric transformation that optimizes a feature based similarity measure [10].

Intensity based methods can use the intensity of the raw voxels, the intensity gradient or statistical information related to the voxel intensities for registration.

Feature based methods extract some structures like sets of points, edges, contours, graphs, surfaces and volumes from the images and base the registration on those features [10].

Intensity and feature based methods use different search strategies for finding the geometric transformation that best aligns the images. Intensity based methods search iteratively for a transformation that optimizes some similarity measure, also known as the cost function, between the images (Fig. 2, left). Depending on the choice of the similarity measure, the goal is either to minimize or maximize it. The similarity measure is based on the voxel intensities. An optimizer is used to define the best search strategy and an interpolator resamples the voxel intensities to the new coordinate system according to the found geometric transformation. Whenever possible, a pre-registration transformation is performed. It pre-aligns the images in terms of the similarity measure. This procedure speeds up the convergence of the optimizer and often prevents convergence to a local optimum [10].

Feature based methods can use a similar approach as intensity based methods but with the distinction that the similarity measure is evaluated for the extracted features rather than for voxel intensities. Also a different strategy is possible: the matching among features is established using some criterion that is based on, for example, geometrical, physical or statistical properties. The geometric transformation is then formed based on the found matching (Fig. 2, right). For example, if the features are sets of points and each point is represented by a descriptor, the 'corresponding costs' are then the 'distances' between the descriptors of the possible point pairs. In this case the similarity measure is the sum of all the 'corresponding costs' [10].

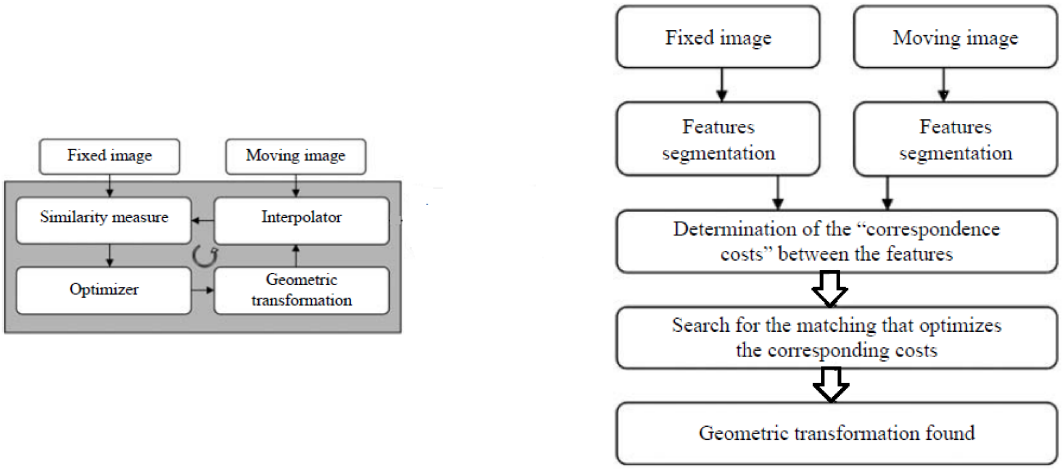

Figure 2: Diagrams depicting the search strategy for optimal geometric transformation between images. The diagram for intensity based methods is on the left and the diagram for feature based methods is on the right. Figures are based on [10].

2.1.1 Rigid image registration methods

The geometric transformations used in image registration can be divided into rigid and non-rigid transformations [10]. Rigid transformations are typically simpler. They can be defined by five parameters or degrees-of-freedom in a 3D space: three translational and two rotational.

In medical image registration rigid transformations are mainly applied in two specific situations. First one is the registration of rigid structures such as bones or organs enclosed in bone; consider e.g, the brain. The other is pre-registration before a more complex geometric transformation [10].

An elegant method to obtain translation and rotation for a pair of images is to use phase correlation together with a log-polar transform (LPT) [18]. The method can also reproduce scale but only to a limited extent. Phase correlation only recovers translation, so LPT is used to recover rotation. The details of this approach are presented in what follows.

The Fourier transform has many beneficial properties for image registration. Translation, rotation, reflection and scaling all have their counterpart in the frequency domain. The frequency domain also provides excellent robustness against correlated and frequency dependent noise [18]. Phase correlation is based on the translation property of the Fourier transform known as the Fourier Shift Theorem [18]. If two images f_1 and f_2 have a displacement $(\mathrm{d}x, \mathrm{d}y)$, that is,

$$f_2(x, y) = f_1(x - \mathrm{d}x, y - \mathrm{d}y), \tag{1}$$

their Fourier transforms F_1 and F_2 are related by

$$F_2(\omega_x, \omega_y) = e^{-i(\omega_x d_x + \omega_y d_y)} F_1(\omega_x, \omega_y). \tag{2}$$

In other words, the images have a phase difference that is directly related to their displacement. It follows from the Fourier Shift Theorem that this phase difference is equivalent to the phase of the cross power spectrum

$$e^{-i(\omega_x d_x + \omega_y d_y)} = \frac{F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)}{|F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)|}, \tag{3}$$

where * denotes the complex conjugate. The inverse Fourier transform of the phase difference is a delta function centered at the displacement [18]. This peak is the point of registration.

The log-polar transform is useful in image registration due to its rotation and scale invariant properties [18]. This is due to the fact that rotation and scaling in the Cartesian coordinate system correspond to pure translation in the log-polar domain (Fig. 3). The polar coordinates are the radial distance from the center ρ and angle with respect to the center θ. Log-polar coordinates are obtained by taking the logarithm of ρ.

LPT is an effective tool for recovering the scaling and rotation. However, it is not suitable for faithful extraction of translation parameters as a slight translation produces a modification of the log-polar image [18]. In order to recover translation

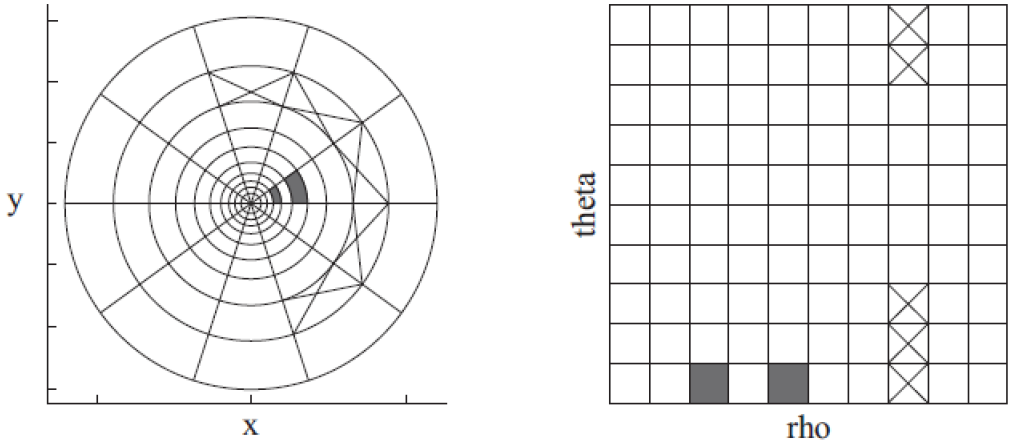

Figure 3: Rotation and scaling in Cartesian coordinates correspond to translation in polar coordinates. Constant angle with increasing radial distance (black boxes) is mapped to ρ-axis and constant radial distance with varying angle (boxes with crosses) is mapped to the θ-axis. Figure is from [18].

and rotation LPT and phase correlation are used together. Fourier transform is first applied over the images. LPT is then performed to the magnitude spectrum to obtain rotation and scaling by using phase correlation in the log-polar domain. This is possible because the magnitude spectra of the images are the same [18]. Translation is obtained with phase correlation as described earlier.

2.1.2 Non-rigid image registration methods

The rigid geometric transformations can be extended to non-rigid transformations by taking into account other parameters besides translation and rotation. A similarity transformation has translation, rotation and uniform scaling. An affine transformation consists of translation, rotation, scaling and shear [10]. An example of an affine transformation is presented in Figure 4. Actually, the rigid and similarity transformations are subsets of affine transformations.

A 3D affine transformation is given by $T(X) = DX + S$, where D is a 3×3 matrix representing the rotation, scaling and shearing and S is a 3×1 vector representing the translation [10]. Affine transformations are sometimes classified as linear. This is not, however, mathematically correct as a function T is linear only if $T(aX + bY) = aT(X) + bT(Y)$. Hence, an affine transformation would be linear only if the translation vector S were null [10]. Affine transformations, like rigid transformations, are rarely used for registering final medical images. However, they are sometimes used for pre-registration [10].

Most organs and structures in the human body are deformable so it is natural that most approaches to medical image registration are based on curved transformations [10]. The curved transformations are often called deformable or elastic transformations. In deformable transformations, the observed signals are associated through a nonlinear dense transformation or a spatially varying deformation

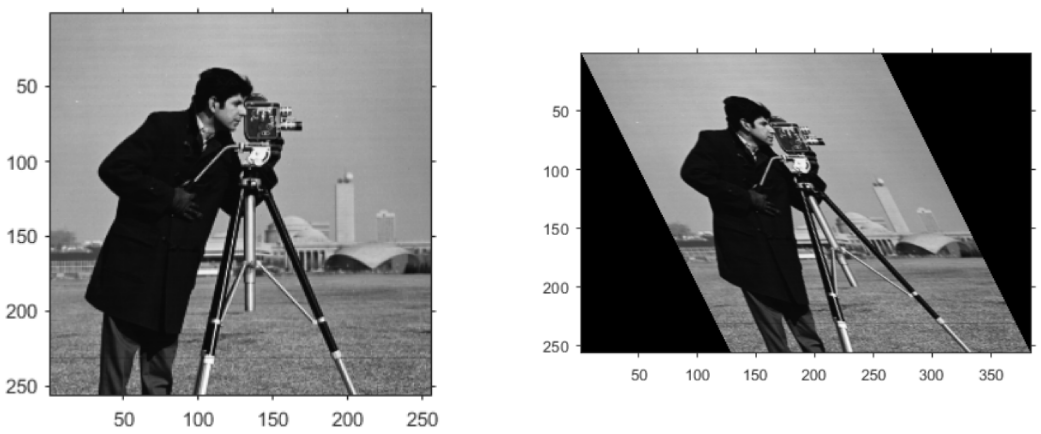

Figure 4: An affine transformation is applied to an image of a cameraman. This transformation has only horizontal shearing. On the left is the original image and on the right the image warped according to the transformation. Images are from [19]

model [20].

Choosing the correct deformation model is crucial for the success of the registration process. The choice of the deformation model includes a trade-off between computational efficiency and richness of description [20]. As the deformation model reflects the classes of transformations that are desirable or acceptable, it limits the solutions of the registration problem. An assumption regarding the nature of the deformation to be recovered is always present in the choice of a deformation model [20].

In medical image registration there are special properties that the transformations should exhibit. These properties include, but are not limited to, inverse consistency, symmetry, topology preservation and diffeomorphism [20]. While searching for the best geometric transformation, constraints can be added to ensure that the transformation has these properties.

Inverse consistent methods estimate simultaneously both the forward and the backward transformation [20]. A data matching term is used to determine how well the images are aligned when one image is deformed by the forward, and the other by the backward transformation. The forward and backward transformations are constrained to be inverse mappings of one another. Inverse consistent algorithms preserve topology but are only asymptotically symmetric.

If an image registration algorithm is asymmetric, the algorithm does not estimate the inverse transformation when the order of the input images is interchanged [20]. Symmetry can be achieved for example by using objective functions that are by construction symmetric to estimate the transformation between the images.

Topology preserving algorithms produce a displacement field that is continuous, onto, and locally one-to-one and has a continuous inverse [20]. For topology preserving displacement fields the value of the Jacobian determinant is greater than zero, wherever it is well-defined.

Diffeomorphic transformations also preserve topology, but they are also required to be smoother. A transformation is diffeomorphic if it is invertible and both the

transformation and its inverse are differentiable. [20]

Deformation models can be classified into four categories. In the first category are models that are inspired by some physical model. The second one includes the models based on interpolation and approximation theory. Third category consists of models that use some specific prior information about the deformation to find the solution. These are called knowledge-based models. Finally, there are models that seek to satisfy some task specific constraint. [20]

The deformation models derived from physical models can be further divided into five classes: elastic body models, viscous fluid flow models, diffusion models, curvature registration and flows of diffeomorphisms [21].

Elastic body models assume that the image under deformation is an elastic body. The body is deformed by a force field and the deformation is described by the Navier-Cauchy Partial Differential Equation (PDE) [20]

$$\mu\nabla^2\mathbf{u} + (\mu + \lambda)\nabla(\nabla \cdot \mathbf{u}) + \mathbf{F} = 0 \tag{4}$$

where $\mathbf{F(x)}$ is the force field that drives the registration based on a matching criterion, $\mathbf{u}$ is the displacement field, μ quantifies the stiffness of the material and λ is Lamé's first coefficient.

The basic idea of an elastic body model is that an external force tries to deform the moving image to match it with the fixed image. An internal force enforces the elastic properties of the material. These forces compete until equilibrium is reached [22]. Elastic body models can be linear or nonlinear. Linear models are simpler but they lack the ability to cope with large deformations [20].

The idea behind viscous fluid flow models is similar as for elastic body models. The difference is that the image under deformation is modelled as a viscous fluid rather than as an elastic body. The Navier-Stokes equation, simplified by assuming a very low Reynold's number, describes the deformation [20]

$$\mu_f\nabla^2\mathbf{v} + (\mu_f + \lambda_f)\nabla(\nabla \cdot \mathbf{v}) + \mathbf{F} = 0. \tag{5}$$

Viscous fluid models do not assume small perturbations so even large deformations can be recovered. The first term in the Navier-Stokes equation (5) guarantees that neighbouring points deform similarly [20]. The velocity field $\mathbf{v}$ is related to the displacement field. The displacement field can be estimated by integrating over the velocity field. The second term allows structures to change in mass and μ_f and λ_f are the viscosity coefficients. Viscous fluid flow models can be computationally inefficient, but several solutions have been proposed to overcome this problem [20].

Instead of a physical model, the deformation model can also be inspired by interpolation or approximation theory. Interpolation theoretic approaches assume that at some locations in the image displacements are known. These known displacements are then interpolated to the whole image domain [20]. In approximation theory, an error is assumed to be present in the estimation of displacements. Hence, the transformation smoothly approximates the known displacements instead of just taking exactly the same values at the control locations [20].

Such models are rich enough for the deformations present in image registration problems and they also have a low number of degrees of freedom. Models based on interpolation theory can be further classified into five categories based on the employed approximation or interpolation method: radial basis functions, elastic body splines, free-form deformations, basis functions from signal processing and piecewise affine models [20].

The free-form deformation models (FFDs) are among the most common types of deformation models used in medical image registration [20]. A rectangular grid $G = K_x \times K_y \times K_z$ is superimposed on the image that gets deformed under the influence of control points. The size of the actual image is $N_x \times N_y \times N_z$ pixels, with $K_x \ll N_x, K_y \ll N_y$ and $K_z \ll N_z$. The displacement field is then presented as

$$\mathbf{u}(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(\mu_x) B_m(\mu_y) B_n(\mu_z) \mathbf{d}_{i+l,j+m,k+n} \tag{6}$$

where $i = \lfloor x/N_x \rfloor - 1, j = \lfloor y/N_y \rfloor - 1, k = \lfloor z/N_z \rfloor - 1, \mu_x = x/N_x - \lfloor x/N_x \rfloor, \mu_y = y/N_y - \lfloor y/N_y \rfloor$ and $\mu_z = z/N_z - \lfloor z/N_z \rfloor$. B_l represents the lth basis function of the B-spline and $\mathbf{d}$ denotes the displacement [20].

This deformation model is simple and is able to efficiently provide smooth deformations [20]. Another advantage is that it only requires a few degrees of freedom to describe local deformations. FFDs gained acceptance in the field of medical image analysis when coupled with cubic-B splines [23]. Many extensions of cubic B-spline FFDs have been developed for several registration problems.

Knowledge-based models can be very useful in the medical field as the registration task is usually well-described. Registration can be performed for images acquired from specific organs or different images are matched to the same target image [20]. For example, while registering breast images, it is possible to introduce prior knowledge about the deformations one tries to recover.

There are two methods for introducing information regarding the deformation in the registration process. If the target image is fixed, a statistical model can be learned by performing pairwise registrations between the target image and some other images that are available. In the actual registration task, deformations that differ from the statistical model can be penalized. The other method is to consider the material properties like tissue elasticity to build a biomechanical model that mimics the properties of the organ/tissue that is imaged [20].

The motivation behind using prior information about the deformation is to increase the robustness and stability of the model [20]. A model is robust if outliers do not drastically affect it. This can be important for example when one tries to search for a tumor in a medical image. The image with the tumor naturally differs from the image of a healthy subject, but the registration should not be affected by such a difference. Stability of the registration method means that small changes in the input data cause only small changes in the result [20]. This is important when subjects are observed over some time and a change in the image in time is associated with an anatomical change, like a lesion or a disease.

Introducing prior knowledge to the registration process can be challenging. In

order to build a reliable statistical model one has to have access to a sufficient amount of learning data [20]. In many situations such an amount of data is not available. Building a reliable biomechanical model can also be a demanding task as the accurate modelling of deformations may be a complex problem.

Marker-based registration methods are often used for the registration of breast images [24]. In [25] it was determined that a 12-marker set is suitable for the registration of a sequence of thermal images of the breasts. Rigid transformations are simple to obtain from marker sets. However, in the context of breast imaging non-rigid transformations are the appropriate choice. In [25] wooden spherical markers were used and they were glued on the skin of the subject before image acquisition. On top of the fiducial markers used to determine the geometric transformation, test markers were used to evaluate the quality of the registration. The number of fiducial markers was changed to find a number of markers after which additional markers did not significantly improve the results of the registration. In all measurements 9 test markers were used. In the registration algorithm the fiducial points had to be detected first. A piecewise linear transformation based on the Delaunay triangulation was used to obtain the geometric transformation. A set with 12 fiducial markers and 9 test markers (Fig. 5) was the best compromise between the quality of the registration and the time it took to prepare the subject.

There are several problems with marker-based registration [26]. Placing the markers is cumbersome and takes time, registration depends on the number of markers, detection algorithms are needed for the accurate localization of the markers and there are limitations on choosing the types of registration parameters. Furthermore, the infrared camera can not detect the temperature of the skin areas beneath the markers.

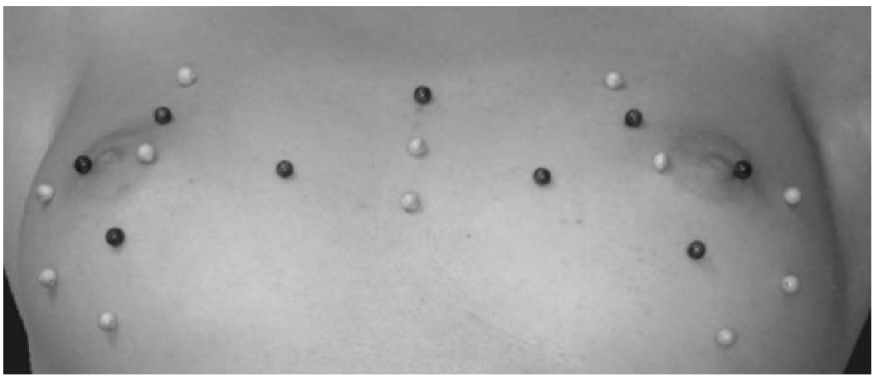

Figure 5: The 12-marker set used in [25] for registration of breast thermal images. White spheres are the fiducial markers and black spheres are used as test markers.

2.1.3 The demons algorithm

In [11] it was proposed that image registration could be performed as a diffusion process. The main idea is to consider the object boundaries in one image as semi-permeable membranes and to let the other image, considered as a deformable grid

model, diffuse through these interfaces [11]. Effectors, called demons, are situated within the membranes and they drive the diffusion process (Fig. 6).

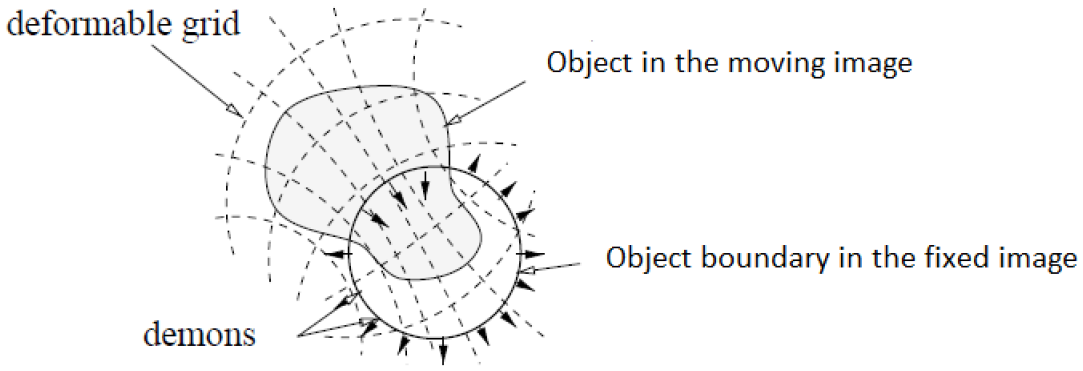

Figure 6: An illustration of the idea behind image registration as a diffusion process. The moving image is modelled as a deformable grid that diffuses through the object boundaries in the fixed image. The diffusion is driven by demons situated at the boundaries. The figure is based on [11].

The diffusion model presented in [11] is inspired by Maxwell's demons. Maxwell used demons to illustrate a paradox of thermodynamics (Fig. 7). In Maxwell's thought experiment we have two containers A and B separated by a semi-permeable membrane. A mixture of gas particles a and b is sealed in the containers. In the beginning the gasses are completely mixed. Now, we place a small demon or demons on the membrane. The demon can distinguish between particles a and b and will only allow particles a to diffuse to the container A and particles b to diffuse to the container B. In the end the container A will only have particles a and the container B will only have particles b. The demon has now decreased the entropy of the system without doing work in a thermodynamical sense. This is a violation of the second law of thermodynamics. The paradox was solved when it was noted that the demons generate a greater amount of entropy to recognize the particles, thus increasing the overall entropy in the world.

An image registration algorithm tries to deform the moving image M so that it matches as well as possible with the fixed image F. The demons are applied to the problem by assuming that the contour of an object O in F is a semi-permeable membrane and by scattering the demons along this contour [11]. It is also assumed that we are able to locally determine, for each contour point in F, a vector perpendicular to this contour and oriented from the inside of the object O to the outside. For example the gradient of the image F is useful for this purpose. M is assumed to be a deformable grid, whose vertices are particles which can be classified as 'inside' or 'outside' [11]. The parallel to Maxwell's demons is the following: in a similar way as in Maxwell's thought experiment, the demons in the registration algorithm push particles labelled 'outside' out of O while particles labelled 'inside' are pushed inside of O. The rigidity of the grid M is determined by the relations between these

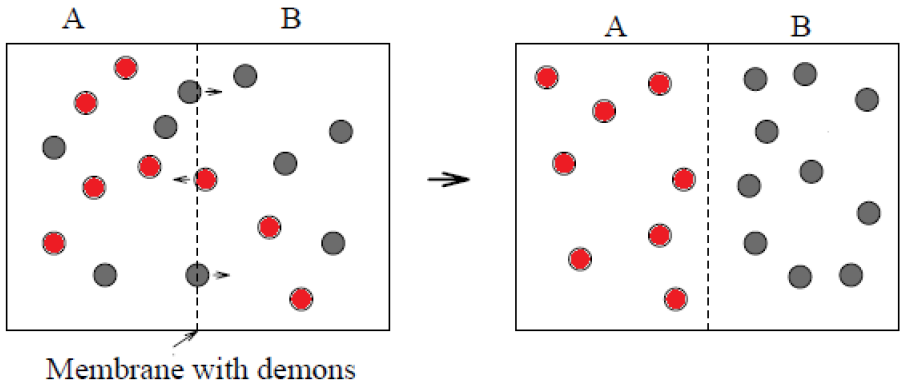

Figure 7: Maxwell's demons are able to separate mixed gases, seemingly violating the second law of thermodynamics. Figure is based on [11].

particles and can range from completely rigid to completely free form [11].

The demons push the pixels with forces inspired by the optical flow equations [11]. Optical flow is used to find small deformations in temporal sequences of images. At a given point P, f is the intensity function in the fixed image F and m is the intensity function in the moving image M. Optical flow assumes that the intensity of a moving object in time is constant. With small displacements this gives the optical flow equation

$$\mathbf{v} \cdot \nabla f = m - f. \tag{7}$$

Equation (7) is not sufficient to define the velocity $\mathbf{v}$ [11]. Ideally $\mathbf{v}$ should be close to zero for small ∇f. With this in mind, we can derive the following expression:

$$\mathbf{v} = \frac{(m - f)\nabla f}{|\nabla f|^2}. \tag{8}$$

In optical flow, $\mathbf{v}$ is a velocity because the images are successive frames in a time-series. However, in a more general sense $\mathbf{v}$ can be simply considered as a displacement [11].

The demon positions D_s can be selected in several ways. The most suitable one for medical image registration is to select all pixels or voxels in the fixed image F, where $\nabla f \neq 0$, as demons [11]. In this case, the interfaces at each point are iso-contours $f = I$, where $I = f(P)$ is constant. This iso-contour is the interface between the inside regions $f < I$ and the outside regions $f > I$ [11]. By comparing the intensities of the moving image M with I, it is possible to automatically label the points of M as 'inside' or 'outside'.

The demons algorithms are iterative. The main steps of the algorithm are 1) computing the demons forces according to the optical flow equations and 2) transforming the image according to these forces [11]. The displacement $\mathbf{v}$ in equation (8) is comparable to the application of an elementary force during one iteration step. The orientation of the force is to the outside (∇f) if P is labelled outside ($m > I$) and to the inside ($-\nabla f$) if P is labelled inside ($m < I$) [11].

For the registration of medical images, the following method is especially suitable [11]. 1) All pixels or voxels are chosen as demons as described earlier. 2) The transformation is completely of a free form, meaning that for every demon (pixel) we get the current elementary displacement $d(P)$. This displacement field is regularized by applying a Gaussian filter with a given σ at every iteration step. 3) The moving image M is transformed according to the displacement field using trilinear interpolation in M. 4) The demon force is given by the optical flow (8). More precisely, the actual force is not calculated but only the result of an application of the force, i.e., a displacement $d = -\mathbf{v}$. [11]

A pyramid (or multigrid) approach is used to make the aforementioned algorithm computationally more efficient [11]. On the finest pyramid level all pixels are used and only few iterations are run. On the next level only $\frac{1}{8}$ of the pixels are used but the number of iterations is multiplied by 4, and so on. The coarsest level is executed first and the finest level last. The user of the algorithm can determine the appropriate number of pyramid levels to be used.

The original demons algorithm had some limitations. For example, it was not diffeomorphic and it lacked a sound theoretical background. In [27] the original demons algorithm was extended to obtain an efficient non-parametric diffeomorphic registration algorithm. Also theoretical justification for the demons algorithm was presented.

The diffeomorphic demons algorithm was used in [26] for the registration of time-series of breast thermal images. The algorithm outperformed two other registration algorithms in terms of several performance measures. The measures used were the normalized mutual information (NMI), the breast boundary overlap (BBO) and the percentage of negative values of the Jacobian determinant of the deformation field.

In conclusion, there is an extremely wide variety of approaches for deformable image registration; one has to carefully decide the model that is best suited for solving the problem at hand. The demons algorithm seems well-suited for the registration of thermal breast images.

2.2 Detection of breast cancer with dynamic infrared imaging (DIRI)

In dynamic infrared imaging one obtains a sequence of thermal images of the subject. This allows for the observation of the spatio-temporal temperature profile. In the context of medical imaging, one can observe the time behaviour of temperature on the skin surface.

It has been demonstrated that DIRI can be used to detect breast cancer [7]. DIRI is an appealing technology for breast cancer detection as it is non-invasive, easy to perform and does not involve any use of ionizing radiation. It is also relatively cheap compared to other imaging modalities.

The idea that the temporal temperature behaviour of cancerous lesions differs from the behaviour of healthy tissue was presented already in 1994 [28]. The abnormalities in the thermal behaviour of cancerous breasts are caused by the nitric oxide (NO) produced by the tumour. Many studies have shown that tumours generate NO

by inducible nitric oxide synthase (iNOS) [29, 30]. This is not only true for breast cancer but for practically every type of cancer. NO has many functions in cancer biology. Each type of cancer may have its own characteristics regarding NO, but the overall biochemical process is not present in the respective benign tissue. NO production seems to be an essential property of cancer [7].

The production of NO by tumours causes regional vasodilation [7]. This vasodilation serves to provide the tumour with increased supply of nutrients and oxygen. NO also enhances angiogenesis: the formation of new blood vessels from pre-existing ones [31]. It is essential for the tumour to grow its own collection of blood vessels via angiogenesis. Besides these mechanisms NO also supports the growth and proliferaration of the tumour in other ways. All of these mechanisms are vital for the survival of the cancer and therefore iNOS can be thought of as a necessary condition for a cancerous disease [7].

The detection and localization of cancerous lesions with DIRI is based on the effect of extravascular NO on the vasculature [7]. Normally vascular tone is modulated by neurohumoral control, which produces small concentrations (significantly smaller than the extravascular concentration in the presence of a tumour) of NO inside blood vessels. Blood flow is also modulated by hydrodynamic cardiogenic pulses. The combination of neurohumoral and cardiogenic modulation results in a highly complex modulation frequency spectrum of blood perfusion. This frequency spectrum ranges from few mHz to more than 10 Hz, so the range spans over several orders of magnitude [7].

The large concentration of extravascular NO in the presence of a tumour causes blood vessels in that region to dilate. They stop responding to neurohumolar modulation and stay constantly dilated [7]. If the affected blood vessels are close to the skin surface, this results in attenuation of the modulation of skin perfusion. If the affected vessels are located deeper, the result is an accentuation of modulation of skin perfusion by cardiogenic pulses. These effects can be observed quantitatively with DIRI [32].

In one of the first studies [7] where DIRI was used to detect breast cancer a series of thermal images of the subjects breast was acquired with a frame rate of 100 Hz. This sampling rate optimally covers the frequency range of measurable perfusion modulations. The duration of the measurement was 11 seconds and 1024 consecutive thermal images were analyzed.

The first step in the data analysis was a fast Fourier transform (FFT) of a time-series of the average temperatures in subareas of interest on the breast surface [7]. As a result of FFT, the power spectra were obtained for each of the 1500-2000 subareas over the studied breasts.

The data analysis yields information on two aspects of cutaneous perfusion: modulation of temperature and modulation of the homogeneity of the cutaneous capillary bed [33]. The temperature modulation that follows changes in the level of perfusion is in the order of tens of millidegrees. From the spatial variance of temperature measured at different pixels in a given small (< 1 cm^2) subarea of the skin, a value called spatial thermal homogeneity (STH) can be derived. The modulation of the homogeneity of the cutaneous capillary bed is observed as the

modulation of STH. STH is of the order of a few millidegrees [34].

In order to separate cancerous breasts from healthy ones several statistical parameters are obtained from the FFT spectra of temperature and the STH FFT spectra. In [7] four significant separators of cancerous and noncancerous breasts were found. Statistical analysis was then performed for these parameters in order to create a single diagnostic parameter with a useful diagnostic sensitivity.

In [8] a wavelet-based multi-scale analysis of temperature fluctuations was used to separate healthy breasts from cancerous breasts. It was shown that besides the differences in the cardiogenic and the neurohumoral rhythms between cancerous and healthy breasts, the complexity of temperature fluctuations about these physiological perfusion oscillations is qualitatively different. The wavelet-based method can be used to observe these differences and thus to separate healthy breasts from cancerous ones.

Thermal stress can be applied to the breasts before the dynamic infrared imaging sequence in order to improve the contrast between cancerous and healthy breasts [9]. The most common type of thermal stress is cooling of the breasts. The reasoning for the increased contrast is that the vessels of the tumor lack the ability to contract in response to neurohumolar modulation. For this reason, the temperature in the tumor region stays almost constant, while the temperature in the other regions is significantly affected by the applied thermal stress. In [9] supervised and unsupervised machine learning methods were applied to time-series temperature data, with thermal stress, to classify breasts as healthy or cancerous. For a sample of 80 subjects, out of which 40 were healthy and 40 had pathologically proven cancer, two classification algorithms showed 100 % classification accuracy.

3 Methods

3.1 Similarity measures for image registration

3.1.1 Intensity based similarity measures

A simple intensity based similarity measure is the sum of squared differences (SSD). The SSD simply calculates the squared differences of pixel intensity values and sums them over the image volume

$$\mathrm{SSD} = \sum_{n=1}^{N} (I_{f,n} - I_{m,n})^2. \tag{9}$$

In equation (9), N is the number of pixels and $I_{f,n}$ and $I_{m,n}$ are the intensities of the nth pixel in the fixed and moving images, respectively. A global minimum of the SSD is searched to find the optimal registration [10]. In order for the SSD to be a suitable similarity measure for an image registration problem, the same objects should have the same intensity values in the moving and fixed images. A simple example of this kind of a setting is the registration of images with a white disc on a black background. In the moving image the disc is in a slightly different place than in the fixed image. The intensity value is 1 for the disc (white) and 0 for the background (black). Now, in every point where the disc is misaligned the squared difference will be 1, contributing to the total SSD value. It is easy to see, that in this simple example optimal registration is achieved when SSD = 0. If, however, some intensity values inside the disc in the moving image were not exactly 1, but say 0.98, this would not be the case anymore.

Hence, the use of the SSD is only appropriate when it can be assumed that the intensity values of objects stay constant between the fixed and moving images. For this reason, the SSD is mainly suitable for inter-subject and monomodality registration [24]. However, SSD is not suitable for the registration of thermal breast images even though the problem is inter-subject and monomodal. This is due to the fact that in thermal images pixel intensities are directly related to temperature [26]. The same object will have different intensity values as the temperature of the object varies. Trying to minimize the SSD value could lead to the loss of the time-temperature signal.

Mutual information is a widely used similarity measure in medical image registration [10]. Unlike the SSD it is suitable for the registration of multimodal images. In [26] its normalized variant was used to evaluate the performance of several algorithms in the task of registering time series of thermal breast images.

Mutual information is an information theoretic measure. More precisely it is based on Shannon's entropy [35]. In communication theory entropy is a measure of information. It is used to measure how much information is conveyed to a receiver of a message via said message. In the simplest cases the message may be for example Morse code or words, but the message can also be an image [35].

In 1928 Hartley formulated a measure for the information included in a message with n symbols s, where s is the number of possibilities for one symbol. For example

in the case of a binary code s would be 2 (possible symbols are 0 and 1). This Hartley's entropy is the basis of many present-day measures [35]. The number of possible messages is s^n. This is, however, not a suitable measure for the amount of information in the message as the amount of information would grow exponentially with increasing message length. This is not realistic, so Hartley wanted a measure that increases linearly with n [35]. Another restriction was that given two messages with lengths n_1 and n_2 and numbers of symbols s_1 and s_2, if the number of possible messages is equal ($s_1^{n_1} = s_2^{n_2}$), then the amount of information should also be equal. These restrictions lead to Hartley's entropy

$$H = n \log s = \log s^n, \tag{10}$$

where log is the base 2 logarithm.

Hartley's entropy depends on the number of possible outcomes. The more possible outcomes there are, the greater is the amount of entropy [35]. If there is only one possible message then entropy is 0, as $\log 1 = 0$. In this sense entropy is also a measure of uncertainty. If the number of possible messages is large, you are uncertain which message you will receive. If there were only one possible message, you would already in advance know the sent message, and therefore no information would be gained.

A major drawback in Hartley's entropy is that it assumes that all symbols are equally likely to occur in a message [35]. This is obviously not true, as for example in written text some letters occur more frequently than others. Shannon's entropy solves this problem by weighting the information per outcome by the probability of that outcome occurring

$$H = \sum_i p_i \log \frac{1}{p_i} = -\sum_i p_i \log p_i. \tag{11}$$

Here p_i is the probability that an event e_i occurs.

The second definition in (11) is more commonly used, but the first one better explains the idea behind Shannon's entropy [35]. The information gained from a message/event is inversely related to the probability of that event occurring. A rare event gives more information than a common one, so more weight is assigned to rare events. On the other hand rare events don't happen very often so the low probability is taken into account.

Shannon's entropy can be explained with an analogy to collecting cards. Let us say that you are collecting some trading cards. The set has 4 different cards and you buy a pack containing 4 cards. The probabilities for obtaining the specific cards might be for example 0.2, 0.3, 0.4 and 0.1. Now according to (11), the Shannon's entropy will be $-0.2 \log 0.2 - 0.3 \log 0.3 - 0.4 \log 0.4 - 0.1 \log 0.1 = 3.14$. For most collectable cards, however, the differences in probabilities will be larger. In some other pack, cards might be classified as 'common', 'rare', 'epic' and 'legendary' with probabilities of occurrence 0.75, 0.20, 0.045 and 0.005, respectively. Now the entropy is only 1.02; the amount of entropy is lower as there is less uncertainty about the card you are going to receive. Most of the time you will receive 'common' cards.

Receiving a 'legendary' card gives you a lot of information but as the probability for receiving such a card is very low, the overall entropy is lower than in the case where the probabilities are closer to each other.

Shannon's entropy can also be calculated for an image [35]. Instead of letters in a message we consider gray values of an image. We obtain the probability distribution of an image by counting how many times each gray value appears in the image. Those numbers are then divided by the total number of pixels in the image. It is also possible to obtain the distribution by dividing the range of gray values (0-1) to bins and by counting how many pixels fall into each bin. Shannon's entropy for an image with only a few gray values is low. An image like that contains little information. On the other hand, an image with many different gray values contains a lot of information and has high entropy. In this sense Shannon's entropy gives us information about the shape of the distribution of gray values in the image. Low entropy corresponds to sharp peaks in the distribution. A high entropy value is obtained when the distribution is flat.

One of the earliest information theoretic measures used for registering images was joint entropy [35]. Joint entropy is calculated from the joint histogram of the two images to be registered. A joint histogram is a two-dimensional plot of gray values in each of the two images for all corresponding points. Thus it is constructed by counting the number of times a combination of gray values occurs. For example if a point (i, j) in an image A has the gray value 0.2 and the point (i, j) in an image B has the value 0.5, then the value of the point (0.2, 0.5) in the joint histogram is increased by one.

The joint histogram changes as a function of the image alignment [35]. When two images are perfectly aligned, the joint histogram will show clear clusters for certain gray values as corresponding anatomical parts overlap each other. When misregistration occurs, anatomical structures overlap with different structures, for example brain with skull. When this happens, new gray value pairs will emerge and the clusters in the histogram will become wider (Fig. 8).

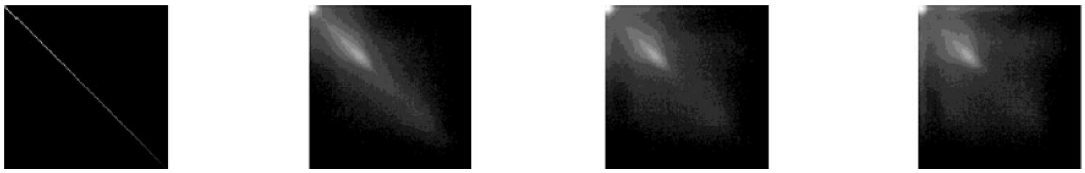

Figure 8: The joint histograms of two identical MR-images with different rotation angles. In (a) the images are correctly aligned and all gray value correspondences lie on the diagonal. In (b), (c) and (d) the other image is rotated by 2, 5 and 10 degrees, respectively. Increasing rotation leads to increased misregistration and increased dispersion of the cluster in the histogram. Figure is from [35]. Copyright © 2003 IEEE.

As discussed earlier, entropy can be seen as a measure of dispersion of gray

values in an image. Low entropy values correspond to clear clusters in images. Therefore minimizing the entropy of a joint histogram leads to optimal registration. To calculate this joint entropy, a joint probability distribution has to be estimated from the joint histogram. This is achieved by dividing each entry in the histogram by the total number of entries. Shannon's entropy for a joint probability distribution is [35]

$$H(A,B) = -\sum_i p(i,j) \log p(i,j). \tag{12}$$

Joint entropy has some drawbacks as a similarity measure. Low entropy values can sometimes be obtained for completely misregistered images [35]. This is the case when the images are misaligned so much that only their backgrounds overlap. The overlap in background results in a peak in the histogram and thus a low entropy value.

To avoid these kinds of situations mutual information was developed by two separate research teams at the same time [36, 37]. Mutual information combines the joint entropy with the individual entropies of the images. Shortly after its introduction, mutual information became the most widely researched similarity measure for medical image registration [35].

Mutual information has several definitions. All of these definitions are interchangeable: one obtains the same value for the mutual information with all of them. However, the different definitions give different views on how mutual information is connected to image registration. [35]

The following definition best explains the term mutual information [35]

$$I(A,B) = H(B) - H(B|A). \tag{13}$$

$H(B)$ is the Shannon's entropy of image B and $H(A|B)$ is the the Shannon's entropy based on the conditional probability $p(b|a)$: the probability of a gray value b in image B given that image A has the gray value a in the corresponding location. If entropy is interpreted as a measure of uncertainty, according to equation (13), mutual information is the amount by which uncertainty about B decreases when A is given. It is the amount of information A contains about B. It is important to note that A and B can be interchanged: $I(A,B)$ is also the information B contains about A, hence it is called mutual information. [35]

The second definition is related to joint entropy $H(A,B)$ [35]

$$I(A,B) = H(A) + H(B) - H(A,B). \tag{14}$$

The mutual information increases when the joint entropy decreases, but the advantage of mutual information over joint entropy is that the marginal entropies $H(A)$ and $H(B)$ are also taken into account. They balance the measure by penalizing transformations that decrease the amount of information in the separate images [35].

The third definition is related to the Kullback-Leibler distance. It is a measure of distance between two distributions p and q and is defined as $\sum_i p(i) \log(p(i)/q(i))$

[35]. To be precise, mutual information can be given as

$$I(A, B) = \sum_{a,b} p(a, b) \log \frac{p(a, b)}{p(a)p(b)}. \tag{15}$$

Equation (15) can be interpreted as the distance between the joint distribution of the images' gray values $p(a, b)$ and the joint distribution in case of independence of images $p(a)p(b)$ [35]. As such, it is a measure of dependence between the images. The assumption is that there is maximal dependence between the images when they are correctly registered. A summary of the different entropies and mutual information is presented in figure 9.

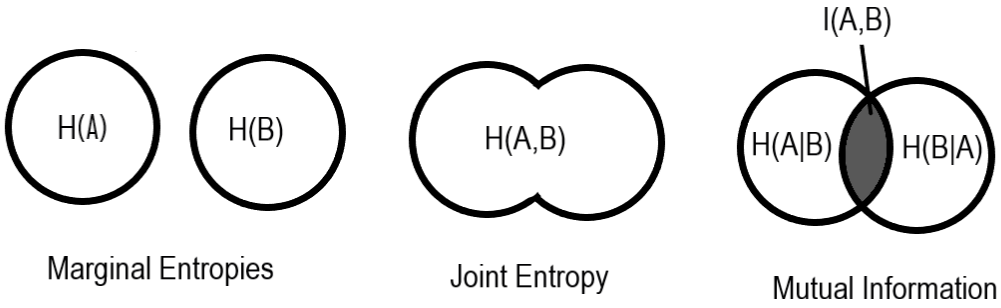

Figure 9: Venn diagrams illustrating the different entropies of two images A and B. The size of the circles represents the amount of information in one image. Figure drawn according to [38].

Mutual information is sensitive to the size of the overlapping part of the images [35]. A decrease in the overlap reduces the statistical power of the probability distribution estimation. In some situations mutual information can increase with increasing misregistration. This is the case when the relative areas of the object and background even out and the sum of marginal entropies increases faster than the joint entropy (see equation (14)). These shortcomings inspired the development of an overlap invariant measure called normalized mutual information (NMI) [38]. NMI is simply the ratio of joint and marginal entropies

$$\mathrm{NMI}(A, B) = \frac{H(A) + H(B)}{H(A, B)}. \tag{16}$$

Maximization of NMI means the minimization of the joint entropy with respect to the marginal entropies [38].

3.1.2 Feature based similarity measures

Feature based similarity measures are an alternative to those based on intensity. Commonly a feature based similarity measure surveys the distance between some

anatomical structures, other regions of interest or marker points between the fixed and warped images. The warped images are the images that are deformed according to the transformation corresponding to the registration algorithm. Naturally one has to first decide which regions or points should be used as a similarity measure. For rigid registration algorithms this is relatively easy: the registration error at any given point is completely determined by errors at three noncollinear landmarks [39].

The selection of features is far more complicated for nonrigid registration algorithms. This is due to the fact that the image is deformed differently at different locations. Therefore, even though some marker points or regions might be well aligned between the fixed and warped images, some other points or regions might be completely misaligned. The chance for this kind of error increases as the distance from the selected points or regions increases [39]. A dense set of control points is required to assure global accuracy of the registration. However, such a set is generally not available. A further problem can arise in inter-subject registration: a subject may lack some landmarks identifiable in another subject [39].

A specific measure for the registration of thermal breast images is the breast boundary overlap (BBO) [26]. BBO measures how well the boundaries of the breasts are aligned. This is one of the most fundamental expectations for the registration of thermal breast images [26].

The first step in calculating BBO is to segment the breast boundaries in all frames. This can be done e.g. using Canny Edge Detection. Furthermore, a strong gradient in every direction is obtained as a scalar value [26]. The SSD of gradient values is calculated between the fixed frame and every warped frame. This value is divided by the number of voxels to obtain the breast boundary difference

$$\text{Boundary}_{\text{difference}} = \frac{1}{n}\sum_{i=1}^{m}(I_{fb} - I_{wb})_i^2, \tag{17}$$

where I_{fb} and I_{wb} are the gradient values of breast boundary for the fixed and the warped images, respectively, m is the total number of frames and n is the number of voxels in one image [26].

The centre of mass (COM) is determined for the whole image and for the left and right breasts separately [26]. Euclidean distances from the breast COMs to the whole image COM are calculated for the fixed and every warped frame (Fig. 10). The distances of the left and right breast from the center are summed. Then SSD of the COM distances is calculated as

$$\text{COM}_{\text{distance}} = \frac{1}{n}\sum_{i=1}^{m}[(I_{fR} + I_{fL}) - (I_{wR} + I_{wL})]_i^2. \tag{18}$$

Here I_{fR} and I_{fL} are distances of the right and left breasts, respectively from the COM of the image in the fixed frame. Hence, these values are not gradient values as in equation (17).

The BBO is determined as the absolute difference between the two aforementioned parameters and its value should be close to zero

$$\text{BBO} = |\text{COM}_{\text{distance}} - \text{Boundary}_{\text{difference}}|. \tag{19}$$

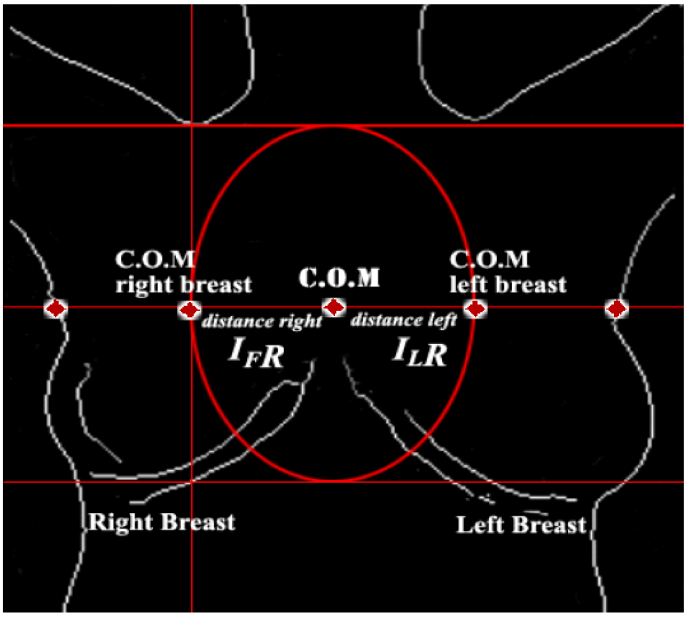

Figure 10: To calculate the BBO, the distances of the COMs of the left and right breasts to the COM of the whole image have to be calculated. Figure is based on [26].

3.2 Description of numerical experiments

In this thesis thermal breast images were registered using two registration algorithms: a registration algorithm that performs an affine transformation and Thirion's demons algorithm. The demons algorithm was chosen as the main algorithm to be tested as it has been shown to be an efficient tool for the registration of thermal breast images [26]. A contributing factor was also that it can be implemented easily with Matlabs `imregdemons` function [40]. An algorithm with an affine transformation was chosen as a comparison algorithm as it can be easily implemented with Matlabs `imregister` function [41]. The demons algorithm was also used to register images that have been pre-registered with the affine method. The results of this combined method were compared to those of the individual methods.

The aforementioned Matlab functions register two images. For this reason, a loop is constructed to perform the registration of the complete time-series. The first image in the sequence is set as the fixed image and all the other images are considered as moving images. In consequence, first the second image is aligned with the first, then the third is aligned with the first and so on, until finally the last image in the sequence is aligned with the first.

An important pre-processing step before using the `imregdemons` function is to perform histogram matching of the images to be registered. This is done using the function `imhistmatch` [42] with 10 000 bins. However, histogram matching poses a problem to the conservation of the temporal temperature signal: matching the histograms of the images will partially destroy the temperature signal. For this reason the `imregdemons` function is only used to calculate the displacement field

from the histogram matched images. The obtained displacement field is then used to warp the moving image, which has not been histogram-matched, onto the fixed image. This is done with the `imwarp` function [19].

3.2.1 Used data

In this thesis time-series of thermal breast images from 20 healthy (no malignant lesions) subjects were registered. The imaging was carried out using a FLIR 655SC infrared camera. Imaging was performed with a frame rate of 50 frames/second. The length of the imaging sequence was 10 seconds, so in total there were 500 frames per subject. In half of the cases, imaging was performed from the right side of the patient and in half of the cases from the left side (Fig. 11). As the 20 subjects, whose data is used in this thesis, have not agreed to have their images published in scientific work, images from a different person will be used as example images (Figs. 11, 14 and 17). This person has given her permission for the use of these images in this thesis.

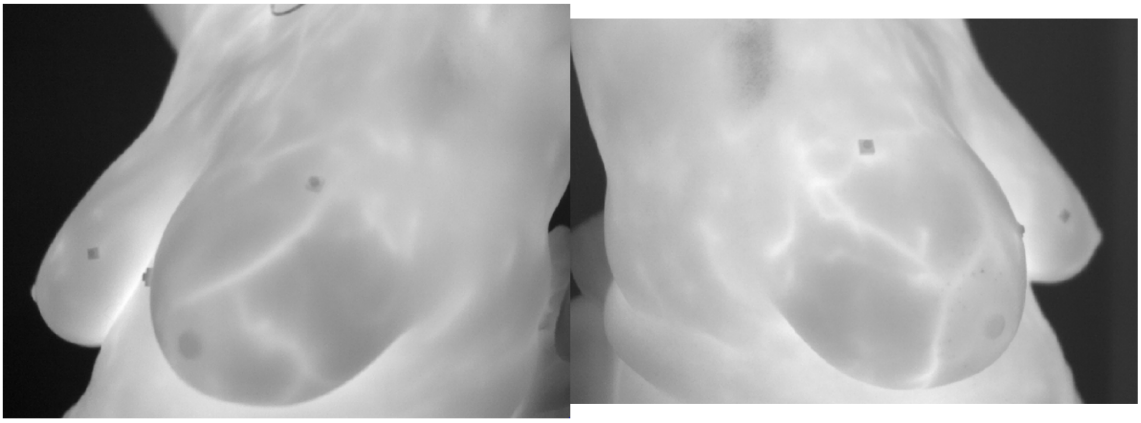

Figure 11: Example thermal breast images. Imaging is performed either from the left side of the subject (left figure) or from the right side (right figure).

3.2.2 Setting the parameters for the imregdemons function

When using the `imregdemons` function in Matlab, the user can define two important parameters: the number of iterations and the amount of smoothing applied to the deformation field [40]. The choice of these parameters affects the registration result and speed.

As discussed earlier, a pyramid approach can be used to perform the iterations of the demons algorithm in order to reduce the computation time of the algorithm. Three pyramid levels were used. In order to determine how many iterations should be run at each level, the demons algorithm was used to register 500 thermal breast images. This task was performed with different numbers of iterations. The time it took to register the images was measured and the average SSD value over the images was calculated to evaluate the performance of the algorithm. At the finest

level of the pyramid the numbers of iterations were 1, 5, 20 and 50. To get the number of iterations at the next level this number was multiplied by four. So the iteration numbers were [16 4 1], [80 20 5], [320 80 20] and [800 200 50], where the first elements correspond to the coarsest level.

Increasing the number of iterations significantly increased the time it took for the algorithm to register the images, while only having a minor impact on the quality of the registration in terms of the average SSD (Fig. 12). Based on these results, [80 20 5] was chosen as the optimal numbers of iterations. Using these numbers improves the registration result compared to the use of lower numbers of iterations in terms of the average SSD, with only minor extra computational cost. Increasing the numbers further results in long computation times with only a minor improvement in the average SSD value.

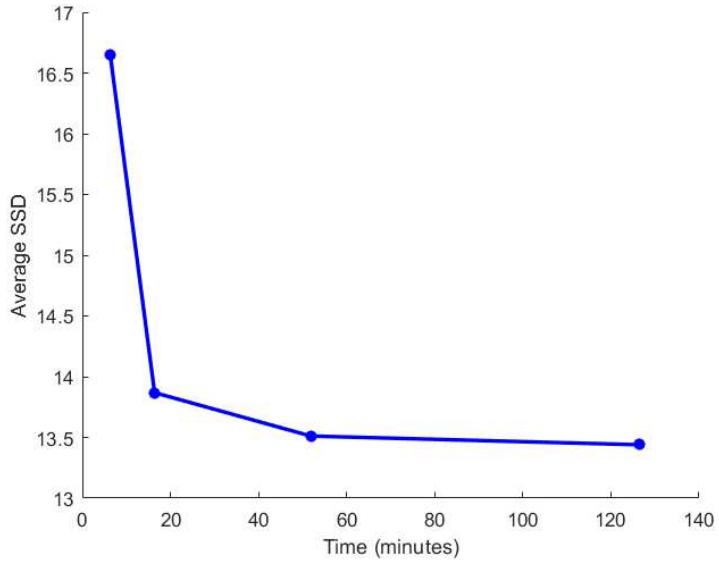

Figure 12: Increasing the number of iterations in the demons algorithm significantly increases the computation time. Increasing the number of iterations at the finest level from 1 to 5 improves the quality of the registration in terms of average SSD, with minor computational costs. Increasing the number of iterations further has little effect on the average SSD, but a major effect on the computation time.

The amount of smoothing is controlled by a value called 'AccumulatedField-Smoothing'. Its value is typically in the range between 1.0 and 3.0 [40]. It controls the amount of diffusion-like regularization applied to the displacement field. The standard deviation of the Gaussian smoothing is employed to regularize the field at each iteration. Large values result in smoother displacement fields and small values in displacement fields with more localized deformations [40].

Different values of this smoothing parameter were tested in a similar manner as with the numbers of iterations. Changing the amount of smoothing between 1.0 and 3.0 did not have a significant effect on the registration time or quality. At the end, a smoothing value of 3.0 was chosen as higher amount of smoothing prevents foldings

and other physically impossible phenomena in the registration result. The effect of changing the amount of smoothing is illustrated in Figure 13.

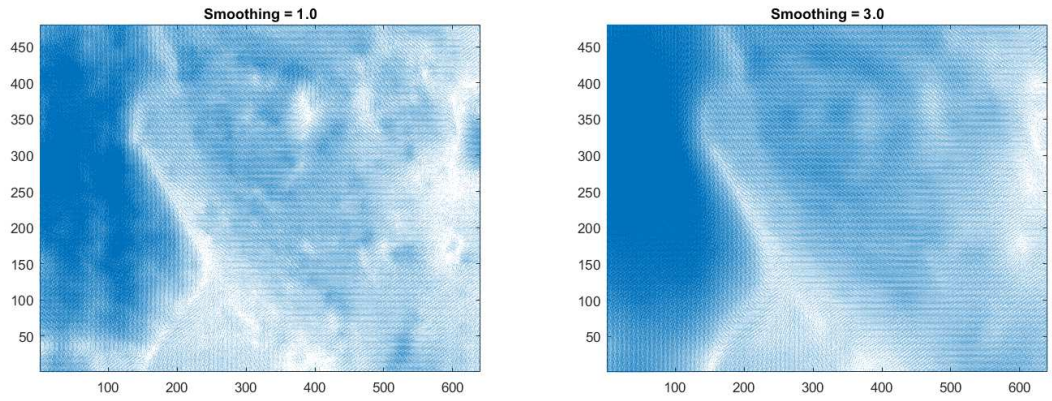

Figure 13: The displacement field calculated by the `imregdemons` function while registering two thermal breast images, with smoothing of 1.0 (left) and 3.0 (right). The displacement field obtained with smoothing of 3.0 is visibly smoother than the one obtained with the value 1.0.

A similarity metric and an optimization method have to be given as parameters to the `imregister` function. The function `imregconfig` [43] is used to obtain these parameters. The default parameters given with the 'Monomodality' option are used.

3.2.3 Evaluating the performance of the registration algorithms

Evaluating the performance of a registration algorithm is an important task as its output would have little value if its accuracy could not be reliably measured [10]. In this thesis, four separate metrics are used to determine which of the chosen methods performs best in the task of registering thermal breast images. Besides these numerical metrics, visual inspection is also used to assure that the algorithms work as intended.

The first two metrics are the intensity based SSD and NMI. Even though it was previously stated that SSD is not a suitable similarity measure for the registration of thermal breast images, it can be used to evaluate how different algorithms perform compared to each other. The problem with the SSD is that the intensities in the thermal images correspond to temperature. For this reason optimal registration does not occur when SSD is 0: the temperature signal always causes some values that are larger than zero. However, it can be argued that the differences in intensities caused by the temperature fluctuations are much smaller than those caused by misregistration. For example, the nipple is always visible in the images as an area colder than the surrounding tissue. If the nipple would be misaligned between images, it would result in a high SSD value.

There is a practical issue related to the use of intensity based metrics to evaluate the performance of the demons algorithm: the `imwarp` function [19] that warps the fixed image according to the displacement field calculated by the `imregdemons`

function creates dark areas by the edges of the image (Fig. 14). These dark areas affect the intensity based metrics. For this reason, the registered images are resized in such a way that these dark areas are excluded from the images that are evaluated with the metrics. This issue does not occur with the affine method but those images are also resized to be of the same size as the ones registered using the demons algorithm. When using the affine method as a pre-registration step before demons registration, the resizing is performed after the demons registration is completed.

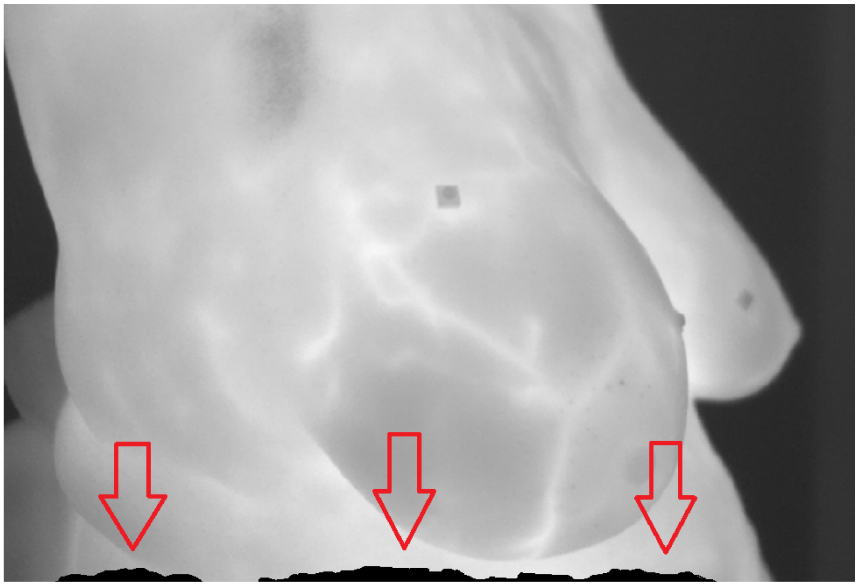

Figure 14: The `imwarp` function creates dark areas at the edges of the registered images. These areas are marked with red arrows and they affect the intensity based metrics. Hence, the images have to be resized to exclude these areas from the final images.

Another issue was observed with the affine registration: in some cases there were a few frames that were completely misregistered (Fig. 15). This happened for 5 subjects and the total number of misregistered frames was 8. These frames were removed from the time-series as they were clear outliers that would affect the values of the metrics. The reason for these misregistrations is unclear.

Even though intensity based metrics can be used to compare different registration methods, they lack any actual physical meaning [10]. It is not possible to know what is the optimal value of these metrics. Furthermore, it has been shown that these metrics can be cheated with a dummy registration algorithm that performs extremely well in terms of these metrics, but does not produce physically meaningful results [39]. This kind of a problem can be circumvented with visual inspection. Because of these problems, however, the use of a feature based metric is necessary. Preferably a metric that gives the registration error in units of pixels, at least in the most important regions, should be used.

The first feature based metric used in this thesis is based on Canny Edge Detection [44]. The edge detection algorithm is implemented with Matlabs `edge` function [45]. The `edge` function allows the user to set a threshold that determines

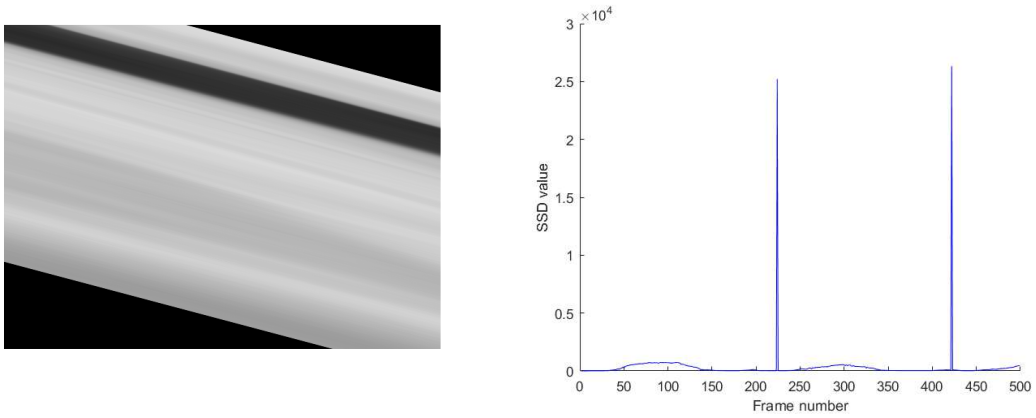

Figure 15: On the left: an example of a misregistered frame. On the right: the SSD value calculated for every frame in a time-series with two clear mistakes. The faulty frames are visible as clear peaks in the SSD plot. These frames are removed before the calculation of the metrics as they would significantly affect the average SSD or NMI value.

the strength of the edges to be detected. All edges that do not meet the threshold will be ignored. The threshold was set in such a way that only one strong edge would be detected in each data set. This was done to ensure that the detected edge is the same in all of the frames. An appropriate threshold was determined to be 0.7. Canny Edge Detection uses two thresholds. The `edge` function sets the lower threshold automatically to be 40 % of the upper threshold [45].

As an output the `edge` function gives an intensity image where the pixel value is 1 in every point where an edge is detected and 0 elsewhere [45] (Fig. 16). Now the average SSD can be calculated for the sets of these edge images in order to determine how much the detected edge has moved on average. More precisely, the metric tells us the average number of pixels where the edge is misaligned compared to the first frame. Low values of this edge SSD indicate good quality of registration.

While detecting the edges, the edge was not completely detected in all of the frames. In some frames the edge was cut off as only a part of it was detected. This cutting off happened at the edges of the frame. This kind of an error has a significant impact on the SSD value. To avoid this, the edge images were reduced to a smaller size to ensure that the areas where the cut-offs happen were not included.

This metric has a few drawbacks. First problem is that the detected edges might be very different between different subjects (Fig. 16). Hence, the results between different subjects are not entirely comparable. For one subject the metric might give more meaningful information about the quality of the registration than for another. Another problem is that even a slight movement of the edge can result in high SSD values. For this reason, the use of this metric as an absolute measure of the quality of the registration is not appropriate.

Another feature based metric used in this thesis is based on the detection of corner points. The Harris-Stephens algorithm [46] is used to detect the points. The detection is performed with Matlabs `detectHarrisFeatures` function [47]. The

Figure 16: The edge images from two different subjects. The value of the edge pixels is 1 and all of the other pixels have the value 0. The detected edges are clearly different between the two subjects.

algorithm does not always detect the same or even the same number of points between the different frames of the data set. In the first frame there might be 5 points detected and in the 10th frame there might be 11. In consequence, the detected points are matched in a way depicted in [48]. The points detected from each of the frames are matched to the points that are detected from the first frame. As an output Matlab gives `cornerPoints` objects that include, for example, the location of the points.

Once we have in hand the locations of the points that were detected from the first frame of the data set as well as the locations of these same points in all of the other frames of the data set, the metric is obtained by first calculating the Euclidean distance between the matching points of the first frame and the nth frame. The average of these distances is then calculated to obtain the average distance of all of the matched points in the nth frame to those of the first frame. This is done for all of the frames. Finally, the mean value of these average distances is calculated. Hence, the final metric is the average distance of certain points to their initial position in the first frame. The metric is calculated in the unit of pixels.

Compared to the other metrics, this average point distance metric has a significant advantage as it is the only one whose result actually has physical significance. The results given by this metric can, at least to some extent, be used to determine if the registration quality is sufficient. A reasonable assumption is that registration quality is good if the average point distance is under 1 pixel.

However, this metric is not by any means perfect. For one, the process of detecting the points is quite random. For some data set the number of matched points might be 5 and for another it might be 2. The points can also be at very different locations. For one subject, they might lie only close to the borders between the subject and the background and for some other subjects, they might be located in central areas. For example, in the data used in this thesis some subjects had markers on their skin. These were often detected by the algorithm. As the locations of the detected points vary, the information given by the metric is not entirely comparable

between different subjects as the importance of the registration quality varies spatially: it is not as important to have good registration in the neck area as it is in the breast area.

Another problem is that the points are not always correctly matched. While calculating the value of this metric for the purpose of this thesis, there were 2 cases out of the 20 total, where a clear mistake in the matching of points occurred (Fig. 17). This kind of an error naturally greatly impacts the value of the metric. The two cases where this error occurred were removed from the data as outliers because in those cases the value of the metric does not give any relevant information about the quality of the registration. The removed subjects were numbers 1 and 19.

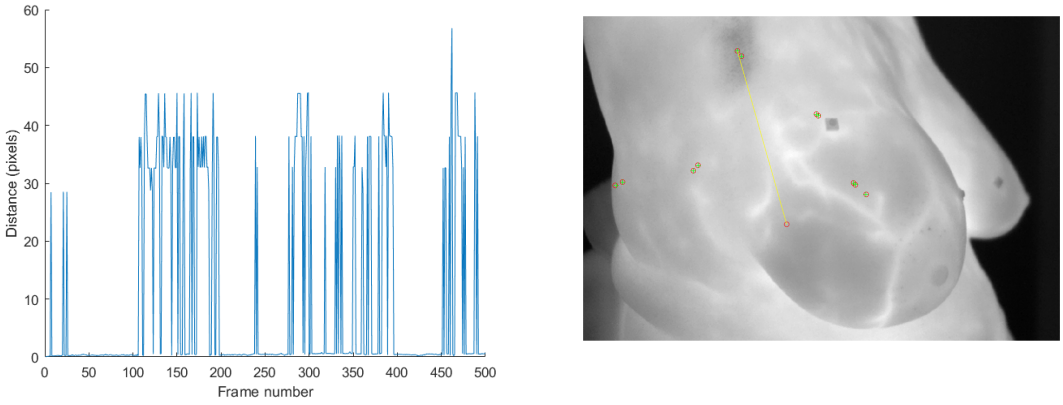

Figure 17: While calculating the value of the average point distance metric, there were two clear outlier cases. On the left is the average distance value for each frame for one of these cases. An example of a point mismatch is depicted on the right. A yellow line is drawn between matched points. The point data is from the subject number 19, but it is overlaid on an example image.

4 Results

4.1 Comparison of the registration methods

In terms of average SSD there is no clear difference between the different registration methods (Fig. 18). For most of the subjects the demons and the combined method give slightly lower values of average SSD than the affine method (low value indicates good quality of registration). However, there are some exceptions, most notably the subject number 8. The 8th subject is a clear outlier for the demons method: its average SSD value is over 120, while the values for other subjects are around 20. Without this outlier the demons method would outperform the affine method, at least slightly. The registration result for the 8th subject was checked visually to spot any clear mistakes, but no such were found.

A boxplot of the SSD results was also drawn in order to better illustrate the differences between the methods (Fig. 19). In this and the following boxplots the black line in the center of the coloured box is the median of the data, the limits of the coloured box are the upper and lower quartiles, the extreme lines show the highest and lowest value excluding outliers, and outliers are marked with circles. The median values of the demons and the combined method are lower than that of the affine method. Even the lower quartile of the affine method has a higher value than the median of the other methods. The combined method has more variability than the other methods. Based on this boxplot, the demons and the combined method slightly outperform the affine method. No clear distinction between the demons and the combined method can be made.

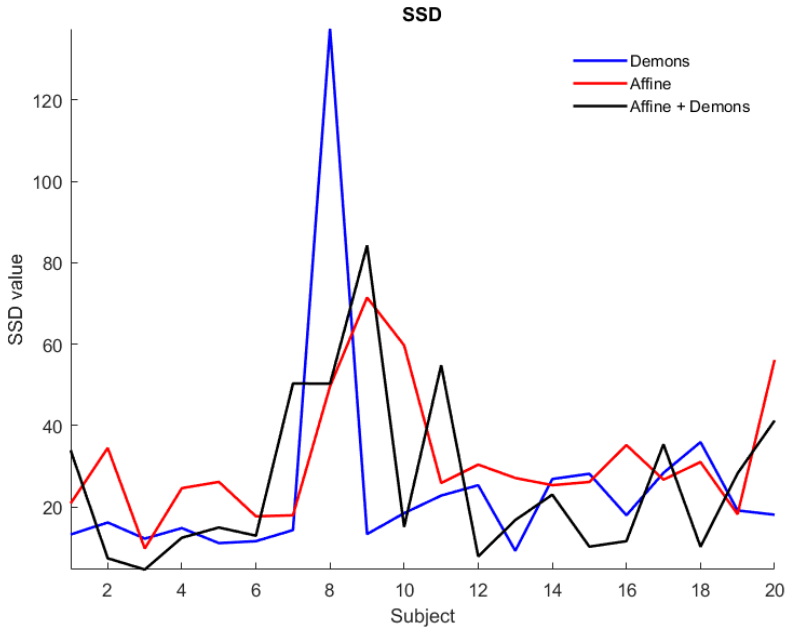

Figure 18: Average SSD values calculated for every subject using each of the three registration methods. Low values indicate good quality of registration. No clear difference between the results is visible in this raw data.

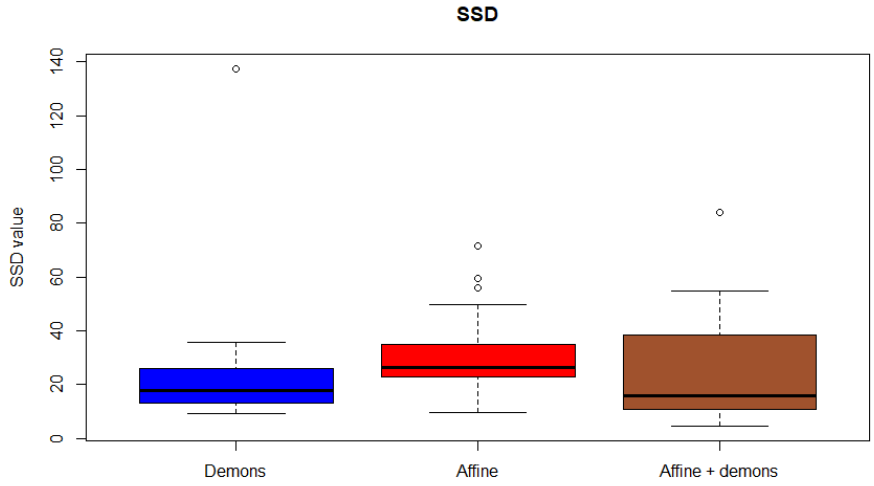

Figure 19: A boxplot of the average SSD results. The combined method has the lowest median and most variability. The affine method has the highest median, and even the lower quartile of the affine method has a higher value than the medians of the other methods.

In terms of the average NMI, the demons and the combined method clearly outperform the affine method (Fig. 20). The average NMI values of the affine method are lower than those for the other methods for every subject (high values indicate good quality of registration). The combined method gives higher values than the demons method for most of the subjects. The difference, however, is not significant.

The boxplot of the NMI results supports these observations (Fig. 21). The upper quartile of the affine method is lower than the lowest values (excluding outliers) of the other methods. The combined method has the highest median but the boxes of the demons and the combined method have a lot of overlap.

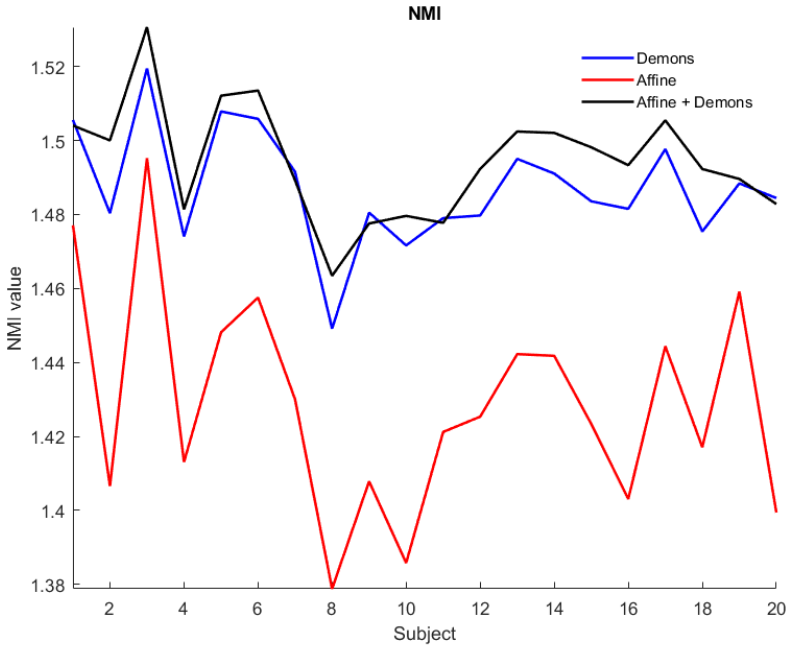

Figure 20: Average NMI values calculated for every subject using each of the three registration methods. High values indicate good quality of registration. The demons method and the combined method consistently outperform the affine method. No evident difference between the demons method and the combined method can be observed.

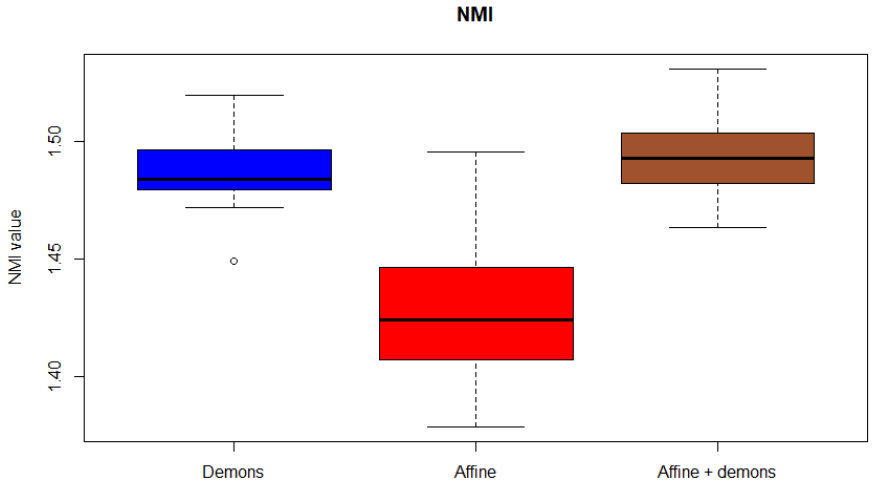

Figure 21: A boxplot of the average NMI values. The values of the affine method are clearly below those of the other methods. The combined method has the highest median, but the box overlaps with the box of the demons method.

In terms of the edge SSD it is not possible to rank the methods (Fig. 22). For one subject the lowest (best) value might be given by the demons method, for another subject it might be given by the combined method and for yet another by the affine method. As with the average SSD, the value given by the demons method for the 8th subject is significantly higher than the values for the other subjects.

The boxplot of the average SSD values further highlights the difficulty of ranking the methods (Fig. 23). There is a lot of overlap between all three boxes. The median of the affine method is slightly higher than those of the other methods. The combined method has a lot of variability as was the case with the average SSD.

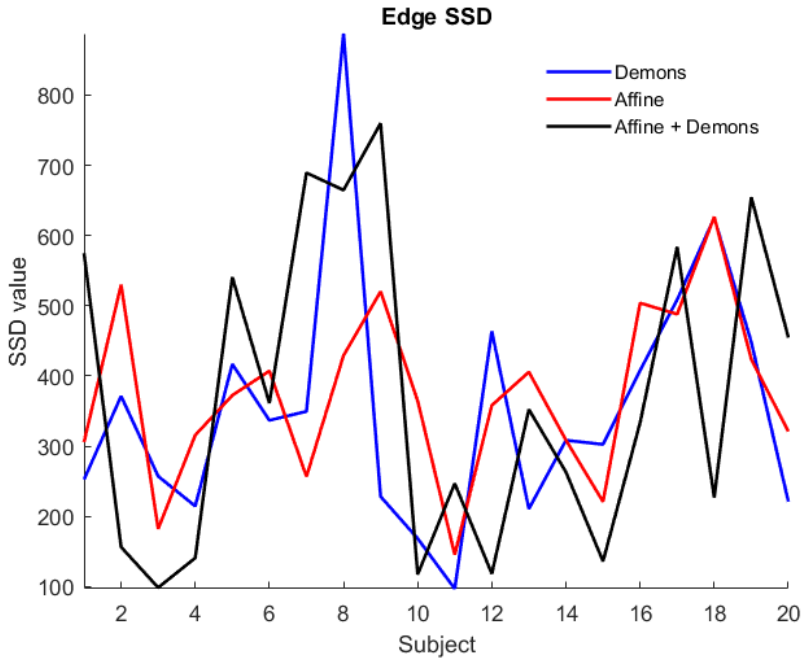

Figure 22: Average edge SSD values calculated for every subject using each of the three registration methods. Low values indicate good quality of registration.

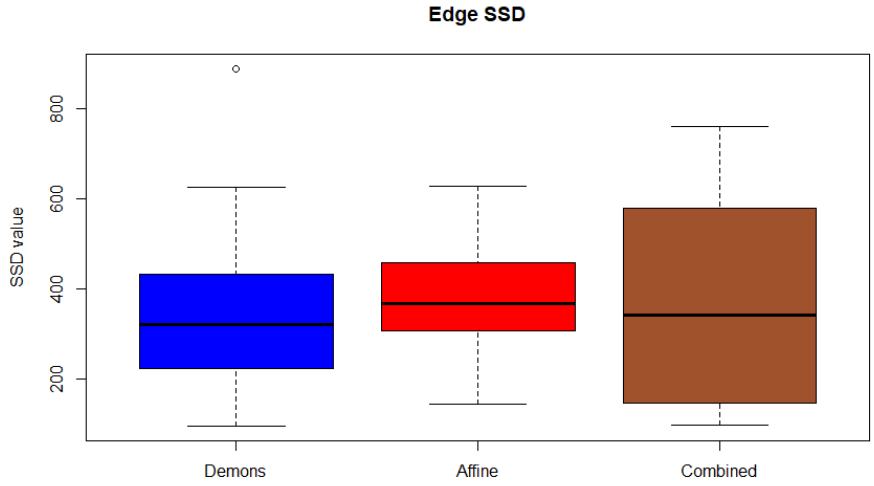

Figure 23: A boxplot of the average edge SSD values. The affine method has the highest median. There is a lot of overlap between all of the boxes.

In terms of the average point distance the demons and the combined method outperform the affine method (Fig. 24). Apart from a few exceptions, the affine method gives higher values than the other two methods (low values are preferred). There is little difference between the results given by the demons and the combined method. As discussed earlier, the results of the subjects 1 and 19 are removed as they were clear outliers resulting from a mistake in the matching of the points.

A boxplot supports these observations (Fig. 25). The values given by the affine method are significantly higher than those given by the other two methods. The lower quartile of the affine method is higher than the median of the other methods. The combined method has the lowest median, but the difference to the demons method is small.

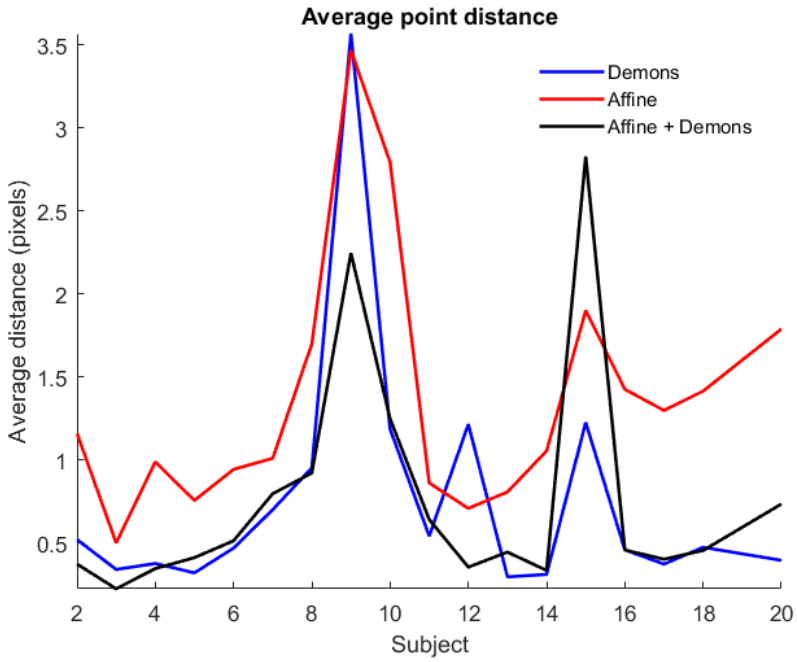

Figure 24: Average point distance values calculated for every subject using each of the three registration methods. Low values indicate good quality of registration. The results from the subjects 1 and 19 are not included in this data. The demons and combined method outperform the affine method apart from a few exceptions.

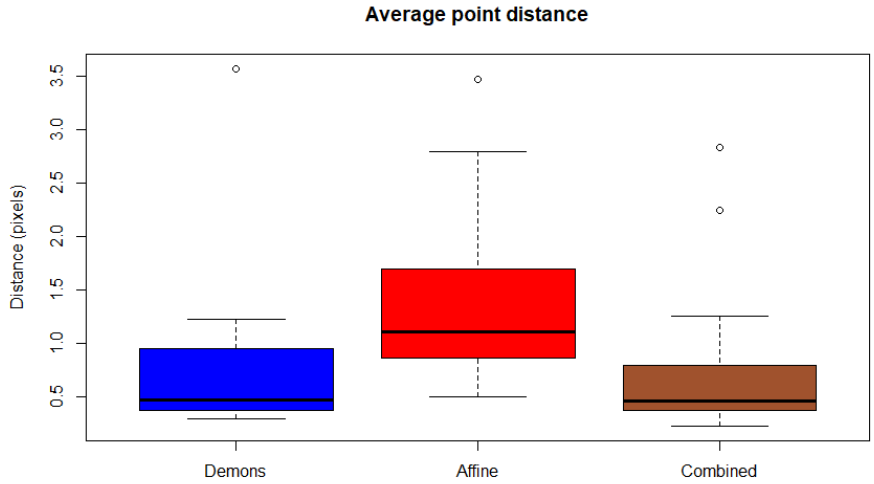

Figure 25: A boxplot of the average point distance values. The results from the subjects 1 and 19 are not included in this data. The values given by the affine method are significantly higher than those given by the other methods. The combined method has the lowest median.

The Wilcoxon signed rank test was used to determine whether these observed

differences between the results are statistically significant. The Wilcoxon signed rank test does not assume normally distributed data and is thus an appropriate statistical test to be used for this data set. The test is performed as a pairwise test. The null hypothesis is that the median difference between the pairs of observations is 0. The alternative hypothesis is that the median difference between the pairs of observations is not 0. Significance level of 0.05 is used: a p-value lower than 0.05 leads to the rejection of the null hypothesis. The p-values are presented in Table 1. For the average SSD the null hypothesis is rejected only for the demons vs affine case. For the average NMI the null hypothesis is rejected in all of the cases. For the edge SSD the null hypothesis is not rejected in any of the cases. For the point distance metric the null hypothesis is rejected in the demons vs affine and the affine vs combined case.

These observations are in line with the ones made based on the graphical representations of the results. Table 2 summarizes the results. A comparison is called a "tie" if the null hypothesis is not rejected. If the null hypothesis is rejected, the method that performs better based on the graphical representation is mentioned as the winner. Both the demons and the affine method get 3 wins, while the affine method does not get any. Out of the used metrics, the average NMI and the average point distance best separate the methods.

	Demons vs Affine	Demons vs Combined	Affine vs Combined
SSD	0.019	0.756	0.202
NMI	$1.9 \cdot 10^{-6}$	0.001	$1.9 \cdot 10^{-6}$
Edge SSD	0.312	0.841	0.985
Point Distance	0.0002	0.580	0.0007

Table 1: p-values obtained using the Wilcoxon signed rank test to compare the different methods. If the p-value is below 0.05 the null hypothesis is rejected.

	Demons vs Affine	Demons vs Combined	Affine vs Combined
SSD	Demons	Tie	Tie
NMI	Demons	Combined	Combined
Edge SSD	Tie	Tie	Tie
Point Distance	Demons	Tie	Combined

Table 2: A summary of the results. According to three of the metrics, the demons method outperformed the affine method. The combined method outperformed the affine method according to two metrics. The combined method outperformed the demons method in terms of the average NMI.

Based on these results, the demons method is recommended to be used to register thermal breast images. The results given by the demons method are significantly better than those given by the affine method. The combined method improves the results of the demons method, but only slightly. Another advantage of the demons method is that it is faster compared to the other methods. Registering one data set

with the demons method took approximately 20 minutes as for the affine method the registration time was around 50 minutes. However, the run times naturally depend on the particular implementations of the algorithms and also on the hardware. The combined method does not improve the results of the demons method in a manner that would justify the resulting increase in the computation time.

4.2 Preservation of the temporal temperature signal

It is crucial for the further analysis of the thermal breast images that the registration preserves the temporal temperature signal as well as possible. To investigate how well the signal is preserved while using demons registration, cosine signals were added to one of the data sets. Signals with frequencies of 0.1, 1 and 2 Hz were used, with amplitudes 0.008, 0.006 and 0.004, respectively. The cosine signal was added to every pixel in the data. Demons registration was then performed on these data sets.

To see how well the signal was preserved, Fast Fourier transform (FFT) was used to calculate the frequency profile for every pixel in the data. This frequency profile was calculated before and after the registration and the results were plotted in the same figure for comparison. The frequency profile was plotted from a single pixel located in the breast area. The added signals with different frequencies were preserved really well (Fig. 26). A clear peak is found at the frequency of the added signal both before and after the registration. Outside of the frequencies of the added signal there are some artefacts. This is an expected outcome as the removal of the motion artifacts due to the registration should somewhat affect the frequency profile. To be precise, the registration reduces the power of lower frequencies as some fraction of the lower frequencies is caused by the motion. It thus seems that a clear frequency trend in the temporal temperature signal is not lost due to demons registration. Especially the higher frequencies (1 and 2 Hz) are preserved well.

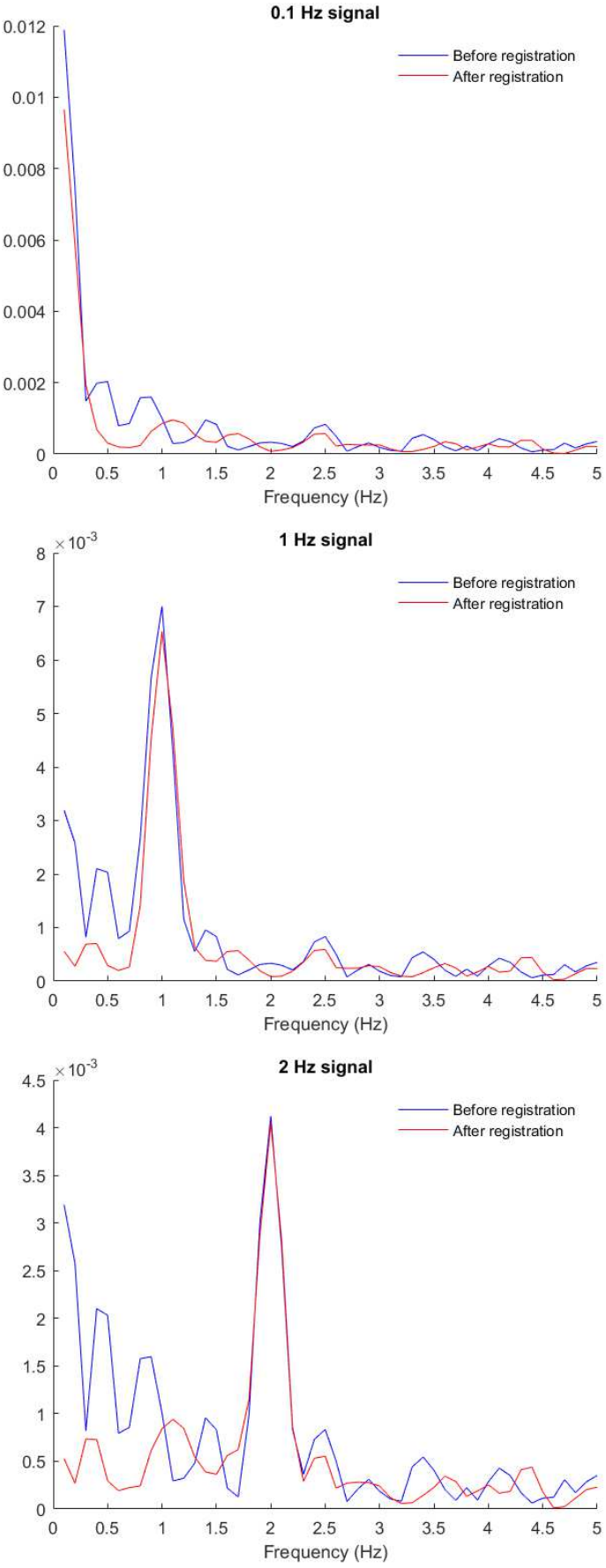

Figure 26: The frequency profiles plotted before and after registration for data sets with an added cosine signal with the frequency of 0.1 Hz (top), 1 Hz (middle) and 2 Hz (bottom).

5 Discussion

An extremely wide variety of image registration algorithms have been developed for the purpose of medical image registration. In this thesis only two of those and their combination was tested. Hence, a question can be raised if the results presented in this thesis actually offer significant proof about the best method for the time-series registration of thermal breast images.

The demons method was chosen as a main algorithm to be tested mainly for two reasons: it had been proven to be an effective method for the registration of thermal breast images in earlier scientific work and it had a built-in Matlab implementation. The fact that there is a Matlab implementation specifically for this algorithm can also be seen as a proof of the usefulness of the method. Furthermore, these first tests with the method were promising, at least based on visual inspection.

The method based on an affine transformation was chosen as a comparison method as it could also be easily implemented with Matlab. A reasonable hypothesis could also be made prior to the numerical experiments that the demons method would outperform the affine method. This is due to the fact that the breathing motion in the time-series data is elastic in its nature, while the affine transformation is not elastic. It would be impossible for the affine method to account for all of the motion.

The combined method was used in order to see if the results of the demons method could be improved this way. The assumption before the experiments was that the results would improve. The reasoning was that the demons method would perform better after all of the non-elastic motion had been removed beforehand: the demons method could only tackle the elastic motion.

The obtained results were quite well in line with the prior assumptions. In particular, this was the case for the comparison between the demons and the affine method. The demons method clearly outperformed the affine method. However, the marginal was not huge. One metric showed a tie between the methods and in terms of the other metrics the demons method was only slightly better.

The comparison between the demons and the combined method was even tighter. Only one out of the four metrics showed a clear improvement in the registration result of the combined method compared to that of the demons method. When deciding whether to use the affine method as a pre-registration step for the demons method, one really has to consider if the slight improvement is worth the significant increase in computation time.

While working on this thesis, the implementation of the registration algorithms was fairly straightforward. Visual inspection also showed that especially the demons algorithm seemed to achieve really good results. The most demanding challenge was to find a numeric metric that would be suitable for verifying (or disproving) the result of visual inspection.

All of the metrics used in this thesis to evaluate the performance of the registration algorithms have their flaws. Based on the scientific work reviewed for this thesis, this seems to be a more general challenge of the entire image registration field. Often the registration algorithms presented in scientific papers are designed

for the purpose of a specific application. After the introduction of the new algorithm the authors proceed to offer proof that their new algorithm outperforms the older solutions. Often they end up calculating a few values that support their algorithms superiority. No 'gold standard' exists for the evaluation of the registration quality.

In this thesis two intensity based measures were used. The core problem of the intensity based metrics is that the optimal value of those metrics is not known in many cases. They do not have any physical meaning. For this reason they can only be used to say if method A is (or seems to be) better than method B. They do not tell by how much a method is better than another. Furthermore, it has been showed that these metrics can be fooled by 'dummy' registration algorithms. The use of the SSD in this thesis was also problematic as the temporal temperature signal affects its value. This might be the reason why the SSD was not really efficient in separating the different methods. The NMI does not suffer from such a problem but has the other drawbacks of intensity based metrics.

The edge SSD metric is basically a combination of an intensity based and a feature based metric. The value calculated is the intensity based SSD but it is based on a physical feature: the edges detected in the image. The value of this metric also has a physical meaning: it is equal to the number of pixels where the edge of the moving image is misaligned compared to the edge of the fixed image. The main problem with the edge SSD is that even a really slight movement of the edge results in a high edge SSD value. This might be the main reason for the fact that the edge SSD was not able to separate any of the registration methods.

The average point distance metric also has a clear physical meaning and it is the only one of the metrics used in this thesis that actually measures the movement of the image in units of pixels. This is really important as such a result can be used to determine if the registration result is sufficient for a specific application: in the context of this thesis for enabling the frequency analysis of the time-series data in such a way that the detection of malignant lesions is possible. The main problem with this metric is that it is prone to clear mistakes. The upside is that these mistakes can be easily detected. Another problem is that the number of matched points between frames can be quite low. If one frame has 640 x 480 = 307 200 pixels and we track 5 of them, we have not really considered the whole story about the movement in the image.

Let us return to the question at the beginning of this chapter: can the demons method really be recommended for the registration of thermal breast images based on the results presented in this thesis? Definitely, not enough proof has been offered to decisively conclude that the demons method is the best option available. However, sufficient proof is presented to conclude that the demons method is a relatively reliable tool for the task. It performs well against the simple affine method and most importantly the average point distance values are consistently smaller than 1 pixel (excluding the clear mistakes). This small movement should not distort the results of DIRI analysis in a way that would make the detection of malignant lesions impossible. Another method might improve the results of the demons method but that improvement probably would not considerably improve the further analysis. The demons method is also fast and easy to implement with Matlab.

References

[1] G. Stamp. Breast cancer. *Encyclopedia of Life Sciences*, 2012.

[2] Chris Woolston. Breast cancer. *Nature*, 527, 2015.

[3] Michelle Grayson. Breast cancer. *Nature*, 485, 2012.

[4] J. Jin. Breast cancer screening: Benefits and harms. *JAMA*, 312, 2014.

[5] M.G Marmot et al. The benefits and harms of breast cancer screening: an independent review. *British Journal of Cancer*, 108, 2013.

[6] Patricia A. Carney et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Annals of Internal Medicine*, 138, 2003.

[7] M. Anbar et al. Detection of cancerous breasts by dynamic area telethermometry. *IEEE Eng. Med. Biol. Mag.*, 20, 2001.

[8] Evgeniya Gerasimova et al. Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. *Frontiers in Physiology*, 5, 2014.

[9] Lincoln F. Silva et al. Hybrid analysis for indicating patients with breast cancer using temperature time series. *Computer Methods and Programs in Biomedicine*, 130, 2016.

[10] Francisco P.M Oliveira and Joao Manuel R.S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17, 2014.

[11] J. P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2, 1998.

[12] Paul G. Gluetz et al. Combined PET/CT imaging in oncology: Impact on patient management. *Clinical Positron Imaging*, 3, 2000.

[13] Andreas Boss et al. Hybrid PET/MRI of intracranial masses: Initial experiences and comparison to PET/CT. *The Journal of Nuclear Medicine*, 51, 2010.

[14] Derek L.G Hill et al. Medical image registration. *Physics in Medicine & Biology*, 46, 2001.

[15] J.B.A Mainz and M.A Viergever. A survey of medical image registration. *Medical Image Analysis*, 2, 1998.

[16] M.A Viergever and J.B.A Mainz et al. A survey of medical image registration - under review. *Medical Image Analysis*, 33, 2016.

[17] D.J Hawkes. Registration methodology: introduction. In J.V Hajnal, D. Hill, and D.J Hawkes, editors, *Medical Image Registration*. New York: CRC Press, 2001.

[18] Jignesh N. Sarvaiya, Suprava Patnaik, and Kajal Kothari. Image registration using log polar transform and phase correlation to recover higher scale. *Journal of Pattern Recognition Research*, 7, 2012.

[19] Mathworks. imwarp. `https://se.mathworks.com/help/images/ref/imwarp.html`, 2018.

[20] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32, 2013.

[21] J. Modersitzki. *Numerical Methods for Image Registration*. New York: Oxford Univ. Press, 2004.

[22] C. Broit. *Optimal Registration of Deformed Images*. PhD thesis, Univ. Pennsylvania, 1981.

[23] J. Kybic and M. Unser. Fast parametric elastic image registration. *IEEE Trans. Image Process.*, 12, 2003.

[24] Yujun Guo, Radhika Sivaramakrishna, Cheng-Chang Lu, Jasjit S. Suri, and Swamy Laxminarayan. Breast image registration techniques: a survey. *Med. Biol. Eng. Comput.*, 44, 2006.

[25] Valentina Agostini, Marco Knaflitz, and Filippo Molinari. Motion artifact reduction in breast dynamic infrared imaging. *IEEE Transactions on Biomedical Engineering*, 56, 2009.

[26] S. Riyahi-Alam, V. Agostini, F. Molinari, and M. Knaflitz. Comparison of time-series registration methods in breast dynamic infrared imaging. *Opto-Electronics Review*, 23, 2014.

[27] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45, 2009.

[28] M. Anbar. Hyperthermia of cancerous breast: analysis of mechanism. *Cancer Lett.*, 84, 1994.

[29] L.L Thomsen et al. Nitric oxide synthase activity in human breast cancer. *Br. J. Cancer*, 72, 1995.

[30] D.C Jenkins et al. Roles of nitric oxide in tumor growth. *Proc. Nat. Acad. Sci. USA*, 92, 1995.

[31] D. Fukumura and R.K Jain. Role of nitric oxide in angiogenesis and microcirculation in tumors. *Cancer Metastasis Rev.*, 17, 1998.

[32] M. Anbar, C.A Brown, and L. Milescu. Objective identification of cancerous breats by dynamic area telethermometry (DAT). *Thermology Int.*, 9, 1999.

[33] M. Anbar and R.F Haverly. Local 'micro' variance in temperature distribution evaluated by digital thermography. *Biomedical Thermology*, 13, 1994.

[34] M. Anbar, M.W Grenn, M.T Marino, L. Milescu, and K. Zamani. Fast dynamic area telethermometry (DAT) of the human forearm with a Ga/As quantum well infrared focal plane array camera. *Eur. J. Thermology*, 7, 1997.

[35] Josien P.W. Pluim, J.B. Antoine Maintz, and Max A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22, 2003.

[36] A. Collignon and F. Maes et al. Automated multi-modality image registration based on information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Information Processing in Medical Imaging*. Dordrecht, The Netherlands: Kluwer, 1995.

[37] W.M. Wells, P. Viola, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. In *Medical Robotics and Computer Assisted Surgery*. New York: Wiley, 1995.

[38] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32, 1999.

[39] Torsten Rohlfing. Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. *IEEE Transactions on Medical Imaging*, 31, 2012.

[40] Mathworks. imregdemons. `https://se.mathworks.com/help/images/ref/imregdemons.html`, 2018.

[41] Mathworks. imregister. `https://se.mathworks.com/help/images/ref/imregister.html`, 2018.

[42] Mathworks. imhistmatch. `https://se.mathworks.com/help/images/ref/imhistmatch.html`, 2018.

[43] Mathworks. imregconfig. `https://se.mathworks.com/help/images/ref/imregconfig.html`, 2018.

[44] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, 1986.

[45] Mathworks. edge. `https://se.mathworks.com/help/images/ref/edge.html`, 2018.

[46] Chris Harris and Mike Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, 1988.

[47] Mathworks. detectHarrisFeatures. `https://se.mathworks.com/help/vision/ref/detectharrisfeatures.html`, 2018.

[48] Mathworks. matchFeatures. `https://se.mathworks.com/help/vision/ref/matchfeatures.html`, 2018.