**Dissertation**
**submitted to the**
**Combined Faculties for the Natural Sciences**
**and for Mathematics**
**of the Ruperto-Carola University of Heidelberg, Germany**
**for the degree of**
**Doctor of Natural Sciences**

**Put forward by**
**Diplom-physicist: Martina Hub**
**Born in: Schwäbisch-Gmünd**
**Oral examination:12/17/2008**

# The geometric accuracy of medical images and elastic image registration in radiation therapy

**Referees: Prof. Dr. Christian P. Karger**

**Prof. Dr. Wolfgang Schlegel**

# Zusammenfassung

Die Genauigkeit radiologischer Bilder, sowie deren Verarbeitung, sind Voraussetzung für eine erfolgreiche Strahlentherapie. Diese Arbeit befasst sich mit der geometrischen Unsicherheit der Magnetresonanztomographie (MRT) sowie der elastischen Bildregistrierung. Es wurde ein neuartiges Phantom mit zugehöriger Auswertesoftware zur Erkennung von Verzeichnungen in MRT Bildern entwickelt. Die Phantommessungen resultieren in einem dreidimensionalen Verschiebungs-Vektorfeld, das zur Korrektur der Bilder herangezogen werden kann. Tests ergaben eine Reduktion der mittleren geometrischen Fehler von 3.0 ± 3.6 mm auf weniger als 1 mm im Durchschnitt. Zusätzlich wurden Verfahren zur Abschätzung der Unsicherheit von Verschiebungs-Vektorfeldern, resultierend aus B-Spine- sowie Demons-Registrierung, entwickelt. Im Fall der B-Spline-Registrierung wurde dies durch zufällige Variationen der resultierenden B-Spline-Koeffizienten realisiert. Eine Größe zur Erfassung der lokalen Empfindlichkeit der Metrik auf diese Variationen wurde eingeführt. Die signifikante statistische Abhängigkeit zwischen dieser Größe und dem lokalen Registrierungs-Fehler wurde demonstriert. Für den Demons-Algorithmus wurde die lokale Reproduzierbarkeit des Verschiebungs-Vektorfeldes als Unsicherheitsmaß betrachtet. Beide Verfahren wurden anhand künstlich verformter Lungenbilder getestet. Die Verfahren erlauben die Einteilung der Bilder in Sub-Regionen, die sich im Betrag ihres durchschnittlichen Registrierungsfehlers unterscheiden.

# Abstract

Radiotherapy relies on the accuracy of radiological images and image processing procedures. Here, the geometric uncertainty of magnetic resonance imaging (MRI) and elastic image registration are investigated. A new type of phantom to measure MRI distortions and a corresponding evaluation software were developed. As a result of phantom measurement, a tree-dimensional displacement vector field is obtained to correct the images. In tests, the distortions were reduced from 3.0 ± 3.6 mm to less than 1 mm in average. In addition, methods to estimate the uncertainty of the displacement vector field (DVF) were developed for a b-spline and a demons registration algorithm . In case of the b-spline algorithm, this was done by random variations of the coefficients resulting from the registration. A quantity was introduced to characterize the local sensitivity of the similarity measure to these variations. The significant statistical dependence between this quantity and the local image registration error was demonstrated. For the demons algorithm, the reproducibility under multiple registrations was regarded as a measure of uncertainty. The algorithms were tested with artificially deformed lung images. Both methods have the potential to divide an image in sub-regions which differ in the magnitude of their average registration error.

# Table of contents

# Table of abbreviations

**DVF** = displacement vector field

**FID** = free induction decay

**MR =** magnetic resonance

**MRI =** magnetic resonance image

**PMMA** = Polymethylmethacrylat

**PET** = Polyethylenterephthalat

**RF** = radio frequency

**SI** = superior inferior

**SSD** = sum of the squared differences

# 1. Introduction

The aim of radiation oncology is to achieve tumor control while sparing healthy tissue. This requires a conformal dose distribution, which is planned and delivered based on medical images. So radiotherapy essentially relies on the geometric accuracy of these images, which provide the geometrical patient model for treatment planning as well as for verification of the patient setup. Geometric uncertainties are therefore a limitation for the accuracy of dose delivery.

As an example, data from magnetic resonance imaging (MRI), which is used for target delineation may be affected by geometric distortion and therefore may not represent a correct geometrical model of the patient. So there is need to measure and characterize distortion as part of the quality assurance procedure in order to find out whether the images need to be corrected. We propose a method to measure the geometric distortion of MRI images with the help of a newly developed phantom and evaluation software.

Another example is the geometric accuracy of image registration. Due to the incorporation of multiple datasets into the treatment planning process as well as for the verification of the patient setup, registration of medical images plays an important role in the daily routine of radiation oncology, since it is essential to know the geometric correspondence between voxels of different images. This correspondence is obtained from image registration.

Since rigid or affine registration techniques can not describe complex deformations, a number of algorithms have been developed to perform fast elastic image registration [1][2][3][4]. It is essential to characterize, compare and validate these approaches in terms of their geometric accuracy.

This work focuses on elastic mono modal image registration and in particular on b-spline registration as well as on the demons algorithm. For each of these algorithms a method is proposed to automatically detect image regions where elastic image registration is likely to perform well and to distinguish those from regions of the same image where the registration is likely to be less accurate.

# 2. Material and Methods

In this section we first give an introduction to MRI focusing on the geometric accuracy. It is followed by a description of the phantom which was developed to measure geometric distortion of MR images and by a description of the corresponding software, which was developed to detect and correct the distortions.
Secondly an introduction to elastic image registration is given, followed by a description of two commonly used intensity based elastic registration algorithms: the b-spline registration and the demons algorithm. The description focuses on sources of error for each of the two algorithms. Then, new methods to automatically estimate the geometric uncertainty of the resulting displacement vectors are described for both algorithms.

## 2.1. MRI distortion

### 2.1.1 MRI image acquisition

The MR image is a spatial representation of contributions of the nuclear magnetic moments to the amplitude of the magnetic resonance signal. The signal received by the antenna is called free induction decay (FID) and originates from transversal magnetization of the object being imaged. In the following we give a short introduction to MRI [5].

#### 2.1.1.1. The MR signal

The magnetic dipole moment of a proton due to its spin $\vec{I}$ is given by

$$\vec{\mu} = \gamma \cdot \frac{h}{2 \cdot \pi} \cdot \vec{I}$$

where $h$ is the Planck's constant and $\gamma$ the gyromagnetic ratio.

In case an object is exposed to an external magnetic field $\vec{B}$ the torque $\vec{\mu} \times \vec{B}$ changes the orientation of the magnetic moment in time, depending on the angle between the field vector and the magnetization:

$$\frac{d\vec{m}}{dt} = \gamma \cdot \vec{m} \times \vec{B} \qquad \text{(A)}$$

Here $\vec{m}$ represents the magnetization:

$$\vec{m} = \frac{1}{V} \sum_{i=0}^{N} \vec{\mu}_i$$

N is the number of nuclei in volume V.

In equilibrium the resulting magnetization is in alignment with the external magnetic field. In case of the MR experiment this equilibrium is disturbed by excitation with a $90^o$ radio frequency pulse. This pulse is a circular polarized magnetic field $B_1$ with orientation perpendicular to the static magnetic field and with Larmor frequency $\omega_L$. The torque on the magnetic momentum caused by $B_1$ changes the orientation of the magnetization. After the time t, the angle $\alpha$ between the magnetic moment and the static magnetic field is given by

$$\alpha = \gamma \cdot B_1 \cdot t \ .$$

A $180^o$ pulse flips the magnetization by $180^o$ around an axis within the x-y-plane.

Adding relaxation terms $(m_0 - m_z)/T_1$ as well as $m_{x,y}/T_2$ phenomenologically in equation (A) leads to the Bloch equations:

$$\frac{dm_z}{dt} = \gamma \cdot \left(\vec{m} \times \vec{B}\right)_z + (m_0 - m_z)/T_1$$

$$\frac{dm_{x,y}}{dt} = \gamma \cdot \left(\vec{m} \times \vec{B}\right)_{x,y} - m_{x,y}/T_2$$

$T_1$ is the longitudinal relaxation time and accounts for spontaneous transitions of the magnetic moments due to interactions with the surrounding molecular grid. $T_2$ is the transverse relaxation time and accounts for dephasing of the magnetic moments due to local inhomogeneity of the magnetic field.

The solutions of the Bloch equations are:

$$m_{x,y}(t) = m_0 \cdot e^{-i\gamma B_z t - \frac{t}{T_2}}$$

$$m_z(t) = m_0 \cdot (1 - e^{-\frac{t}{T_1}})$$

The FID is induced by the change of the magnetization in time, after excitation.

### 2.1.1.2. Spatial encoding

Since all excited spins contribute to the FID, spatial encoding is necessary to obtain the spatial contributions to the amplitude of the magnetic resonance signal. This can be done by slice selection, phase encoding and frequency encoding. In the following the principle of spatial encoding is described for a spin echo sequence.

*1. Slice selection*

To excite a transversal slice of finite thickness, the 90° RF pulse is modulated with a sinc function $f(t) = \sin(t)/t$ while a gradient $G_z$ in axial direction is present [5]. Since the resonance frequency is a linear function of the magnetic field and the Fourier transformed of the sinc function is rectangular, this results in the excitation of a transversal slice. Selective excitation of transversal slices reduces the dimension of the encoding problem from three to two.

*2. Frequency encoding $G_x$*

In case of frequency encoding, the encoding gradient is present during readout and results in a spatial modulation of the magnetic resonance frequency. Due to the gradient $G_x$ two voxels in distance d along the frequency encoding direction have a relative frequency shift of:

$$\Delta\omega = \gamma \cdot G_x \cdot d$$

The Fourier Transformation of the FID provides the information on the contribution of each frequency, and therefore of each strip perpendicular to the frequency encoding gradient [5].

*3. Phase encoding $G_y$*

In case of phase encoding, the spectra of the image is recorded in k space [5]. This is done by multiple repetition of the sequence while each repetition differs in the strength of the phase encoding gradient and therefore in the wave number k:

$$k = \gamma \cdot G_y \cdot t_0$$

$G_y$ is the phase encoding Gradient, and $t_0$ the duration in which $G_y$ is switched on.

Let us regard a voxel which is located in distance d along the phase encoding direction, relative to some arbitrary reference voxel. The phase encoding Gradient $G_y$ leads to a frequency shift with respect to the reference voxel:

$$\Delta\omega = \gamma \cdot G_y \cdot d$$

after $t_0$ this results in a phase shift of:

$$\Delta\varphi = \gamma \cdot G_y \cdot d \cdot t_0$$

Therefore, the amplitude of the magnetic resonance signal which originates from a voxel in distance d from the reference point becomes:

$$s(d) \sim e^{ik \cdot d}$$

with $k = \gamma \cdot G_y \cdot t_0$.

Variation of k allows to measure the Fourier transformed m(k) of the amplitude of the spatial contribution to the magnetic resonance signal M(d). So M(d), which represents the image intensities can be obtained by Fourier transformation of m(k). Note that M(d) is not equal to any specific s(d).

*4. Acquisition of a 3D MR image*

In case of a typical spin echo sequence, three dimensional encoding is performed by slice selection in z dimension, frequency encoding in x direction and phase encoding in y direction.

For each slice the image $I(\omega(x), y)$ is finally reconstructed by Fourier transformation.

$$I(x(\omega), y) \propto \iint FID(t, k_y) \cdot e^{-i(\omega \cdot t + \gamma \cdot G_y \cdot y \cdot t_y)} dt \cdot dk_y$$

Note that $FID(t, k_y)$ is a set of measured FID signals for many different values of $k_y$.

## 2.1.2 Geometric distortion

The source of distortions in MRI is a deviation of the actual from the specified magnetic field. This can be due to inhomogeneity of the static magnetic field, due to local susceptibility effects or due to non-linearity of the encoding gradients.

In the following we assume that the magnetic field B in position $x_0$ differs from the specified field $B_0$ by $\Delta B$

$$B = B_0 + \Delta B$$

Let $G$ be the frequency or phase encoding gradient. For simplicity we regard just one dimension in space. The frequency shift due to $\Delta B$ in $x_0$ can not be distinguished from the one that appears in case of an ideal field and encoding gradient $G$ in the position

$$x = x_0 + \Delta x.$$

The consequence is the distortion $\Delta x$ of

$$\Delta x = \frac{\Delta B}{G}.$$

Note that the distortion is large in case of small encoding gradients. This issue will be discussed in more detail in the following.

Several ways to correct images for distortions have been proposed in the past:

### 2.1.2.1 Application of spin echos

In case of phase encoding it is possible to remove the effects of field non-uniformities on the FID by a spin echo. This is performed by 180 degrees pulses between excitation and readout of the FID [6] [7]. In that way phase shifts due to inhomogeneity of the static magnetic field or local susceptibility effects are refocused since the perturbation of the magnetic field is present throughout the repetition time - before as well as after the 180 degrees flip. Therefore the flip of the spins reverses the influence of the local field inhomogeneities. Phase shifts induced by the encoding gradient remain present since the encoding gradients are not present throughout the whole repetition time, but are switched on for a limited time $t_0$. However, the spin-echo-technique is time consuming which is a draw back of this method.

### 2.1.2.2 Distortion maps

Distortion maps can be obtained with the help of physical phantoms such as described by [8][9][10][11]. The distortion is measured by determining the location of control points within a phantom of well defined geometry. The MR images can then be corrected based on the distortion map.

### 2.1.2.3 Altered gradients

A method to correct MR images for distortion without involvement of a physical phantom was proposed by [12]. This method requires acquisition of two images with altered gradients and aims to correct images for geometric as well as intensity distortions due to static field inhomogeneity. A drawback of this method is that it does not account for gradient non-linearity which was found to be the major source of geometric distortion in MR images according to [11].

### 2.1.2.4 Characterization of $\Delta B$ and post processing of the images

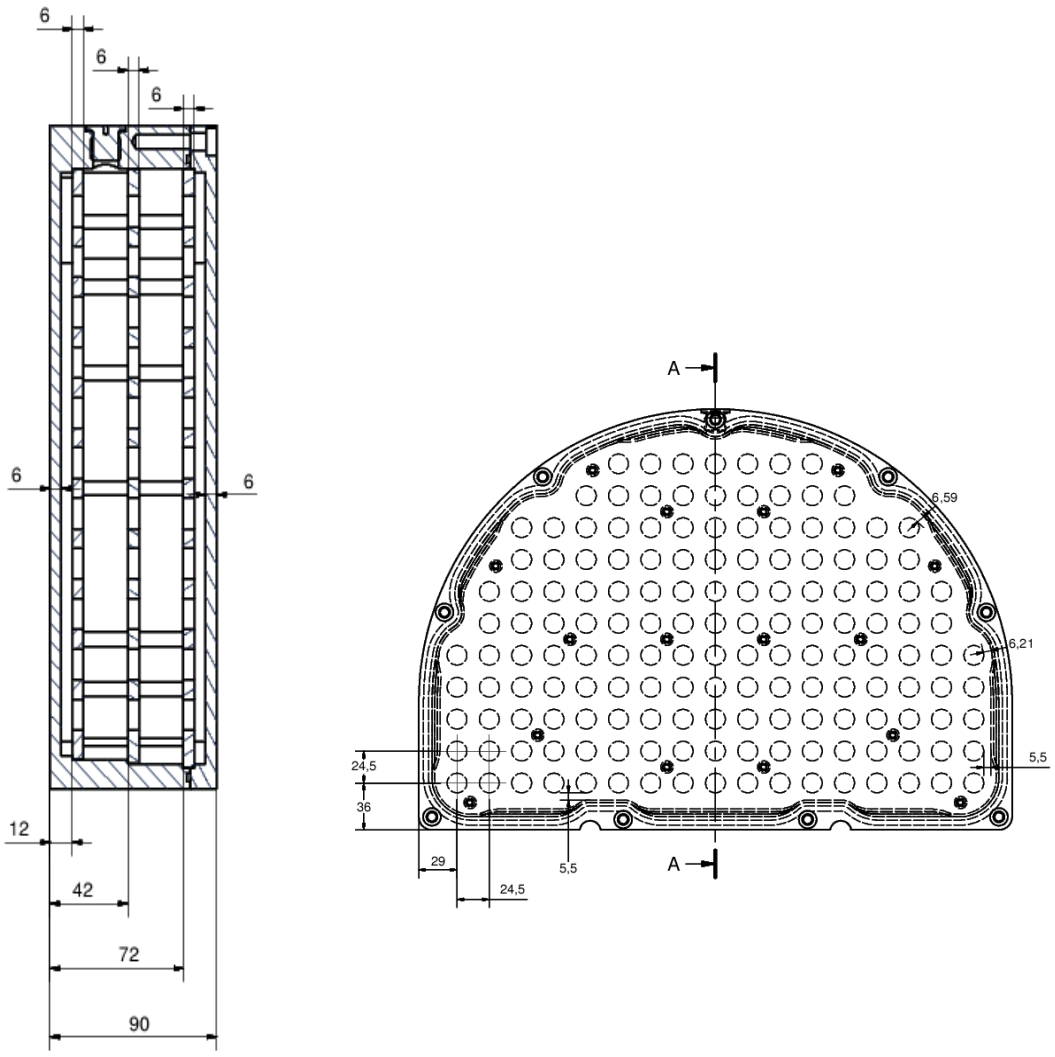
Another way to geometrically correct MR images is to calculate the ground field inhomogeneities as well as the gradient non linearity from the geometry of the scanner coils. The influence of the perturbation on the images is then calculated and the images are corrected in a post processing step after the image acquisition. Nevertheless phantom measurements are necessary to validate the calculated corrections.

## 2.1.3 The phantom design

The phantom was developed in-house as a quality assurance tool for the application of MRI in radiotherapy. It consists of five water-filled modules, which can be handled separately. The modules are made of PMMA (Polymethylmethacrylat). Each module contains three PET (Polyethylenterephthalat) slices with a regular grid of holes, each with a diameter of 15 mm. The distance of the holes is 24.5 mm in plane and 30 mm in axial direction (within each module as well as between slices of neighbouring modules). As the modules are water-filled the holes form a regular grid of control points in the MR image. The software detects these control points from the images and calculates the distortion as deviation between the measured and the mechanically defined positions of the control points.

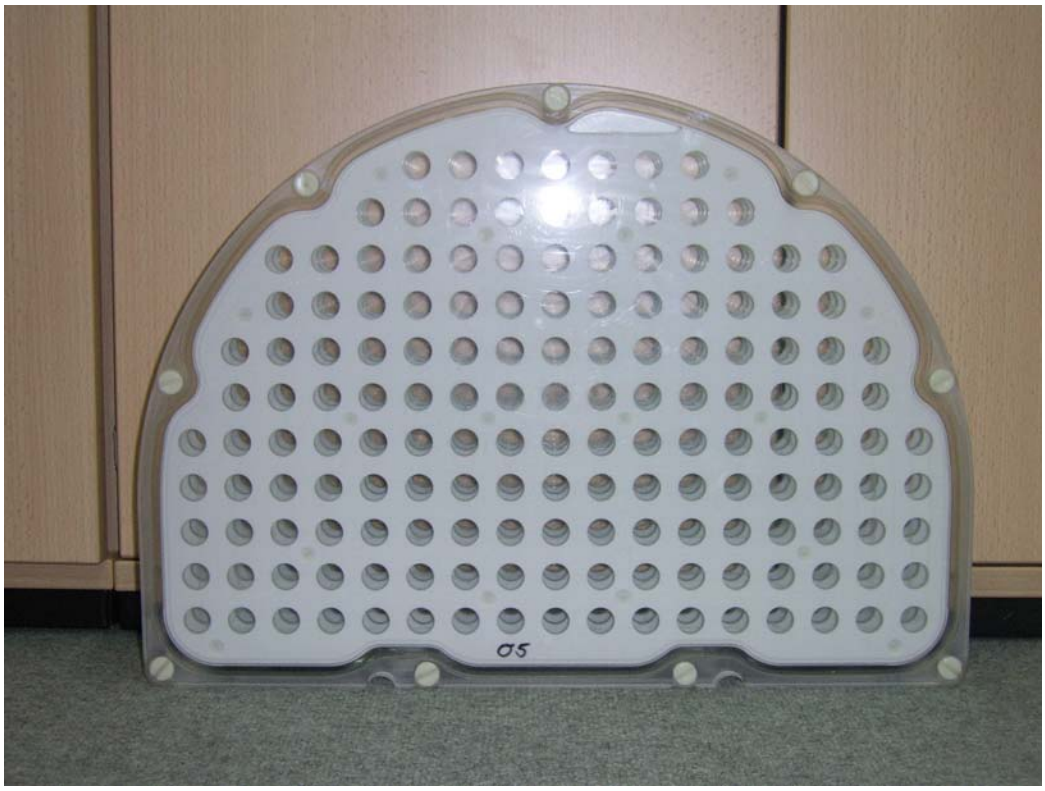
The axial component of the distortion can be expected to be very small in the central slice of the phantom, since in this central region the inhomogeneity of the ground field can be expected to be small and the axial gradient of the magnetic field can be expected to be closer to linear than elsewhere. Deviations between the measured axial positions of central control points and those defined by the geometry of the phantom are therefore likely to be due to imperfect positioning of the phantom. Therefore the axial distortion is assumed to be zero in the central slice. The axial distortion in a control point of another slice is calculated by comparing the measured position of this point with the mechanically defined distance of the respective control point in the central slice. In plane distortions are obtained by comparing the measured control point position with the mechanically defined distance from the control point which is closest to the center of the scanner.

The PET slices are 6 mm thick as shown in Fig. 1 (a) and the control point grid covers 245 mm in lateral and 392 mm in the vertical direction as shown in Fig. 1 (b) and Fig. 2. The field of view in axial direction is 420 mm. Fig. 2 shows how all five modules can be positioned in a frame which is attached to the couch of the MRI scanner during the measurements. The central slice of the central module is positioned to the middle of the MRI-device by aligning the MRI-lasers to marks on the phantom.

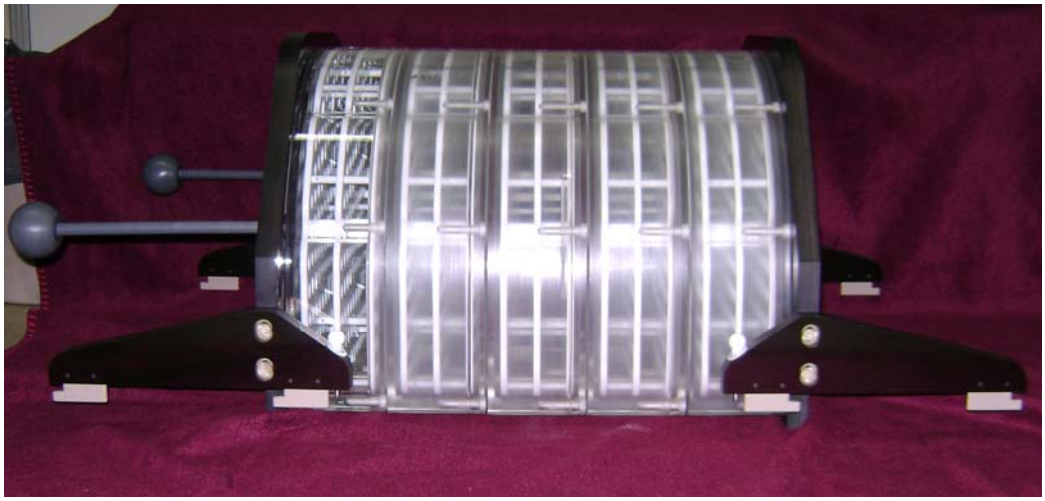



*(a)* *(b)*

*Fig. 1 : Drawing of one module (a) in sagital view and (b) from the front side (distances in mm).*




(a)



(b)

*Fig. 2 : a) one of the modules b) all five modules positioned in the frame*

## 2.1.4 Software to evaluate the distortion

In the following the software to evaluate the image distortions is described. It was developed under Linux using C++ and Qt. It aims to detect the centers of the water filled holes of the PET slices and to determine a distortion map by comparing their locations with the mechanically defined positions. It then corrects the geometry of the MR images based on this map. The description starts with the acquisition technique, followed by the detection of the control points and the correction process.

### 2.1.4.1 Shifted data acquisition

Since the slice distance affects the distortion, the acquisition of data to measure image distortion has to be done with slice distances as used in clinic, which are typically 3 to 5 mm and hence significantly larger than the voxel sizes in the transversal plane. This may lead to a low accuracy in the estimation of control point positions in the axial direction due to low sampling of the intensity profile.
This problem can be solved by acquisition of several images which are shifted relative to one another in the axial direction. The central positions of these images differ by small offsets. These datasets are then combined to one dataset with a smaller virtual slice distance by sorting all slices of all the acquired images according to their SI position.
One data set is acquired without any offset in SI direction. The other images are recorded with offsets in SI and minus SI direction such that all slices of the reconstructed virtual data set have the same slice distance. In this way, the SI position of the slice number i+1 of the virtual data set is differs by an offset $\Delta$ from next slice i in with

$\Delta =$ (original slice distance) / (number of recorded datasets)

Using this technique, the center of the control points in axial direction can be determined with a higher accuracy.

### 2.1.4.2 In plane detection of control points

The location of the control points in plane is obtained by thresholding. The region of the hole is brighter than the PMMA background and can therefore be separated by applying a threshold. Fig. 3 (a) shows the in plane intensity throughout an image of the phantom. Fig. 3 (b) demonstrates the separation by thresholding. The red area represents the detected holes. The center of mass of each red area is considered as the position of the control point.

Since intensity inhomogeneities throughout the slices are present, it is not possible to use one single threshold throughout the slice. This problem can be solved by applying several thresholds. For each threshold that allows the separation of the control point area from the background, a result for the control point position is obtained.

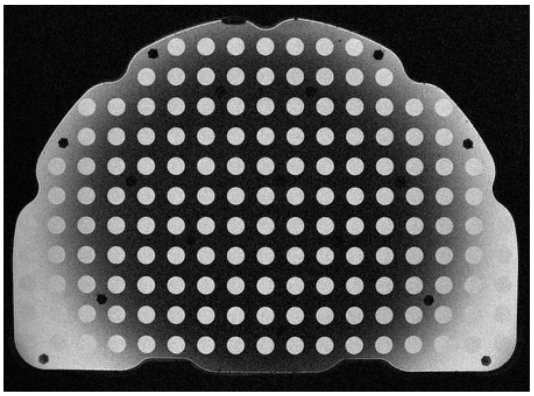
Most of the control points are detected for several thresholds. The average of all the results that were obtained for the same control point and different thresholds is then considered the true position. In case the result obtained for a specific threshold differs from the average by more than 2 mm it is considered an outliner and is therefore disregarded.

As each hole is visible in several slices of the image, the control point detection is done for each slice and the final position of the control point in the transversal plane is the average over the results obtained for all slices in which the control point is visible.

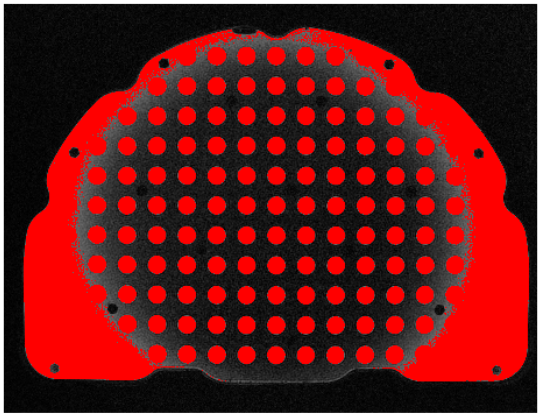
Due to the spatial extend of the phantom some phantom areas are in large distance from the antenna that receives the FID. This causes large variations of the intensity throughout the image. As intensity distortion may affect the control point detection, the intensity distortion is considered to be a two dimensional linear function in the area around each control point and the threshold used for the detection of the control points is adapted accordingly.

It is important to make sure that each of the resulting positions actually represents a control point of the phantom. Therefore, each highlighted area is checked whether it represents a circular shaped object. For this purpose a circle with center in the detected center of mass position is regarded. The distance between the circle and the edge of the highlighted area along each row and column is then evaluated. The average distance between the detected edge and circle is required to be below a threshold of 1.5 voxels. If this is not the case, the highlighted area is not considered to be a control point of the

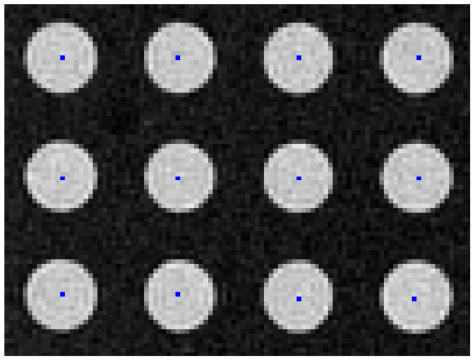
phantom. As geometric distortion may change the size of the holes representing the control points, this procedure is carried out for circles of different radius. If any circle fits the highlighted area the detected point is accepted.



(a)



(b)



(c)

*Fig. 3 : (a) slice of the MRI image, (b) holes detected by threshholding, (c) detected in-plane positions of the control points.*

### 2.1.4.3 Cross plane detection of control points

After the position of a control point in plane is detected, its location in the axial direction is determined. For this purpose image intensity profiles in axial direction are evaluated in the virtual data outside the detected holes.

At the in-plane position of the control point there is only water and hence the axial intensity profile varies solely due to intensity distortion within the images. In areas outside the holes, the axial intensity profile perpendicular to the PET plate has local maxima outside the PET in the region of the water. The difference between both intensity profiles is evaluated and the center of this difference profile is regarded as the axial position of the control point. The center of the intensity difference profile was defined as the point between the two positions with values at half the maximum height and was determined by linear interpolation. The difference profile (Fig. 4) is evaluated at different in plane positions around the hole and the final axial position of the control point is defined as the average over these results. In case a result obtained for a single difference profile differs more than 2 mm from the average it is considered an outliner and is therefore disregarded.

Fig. 5 shows the resulting positions for some control points. Due to the axial distortions, the control points for one phantom slice are in different slices of the MR image.

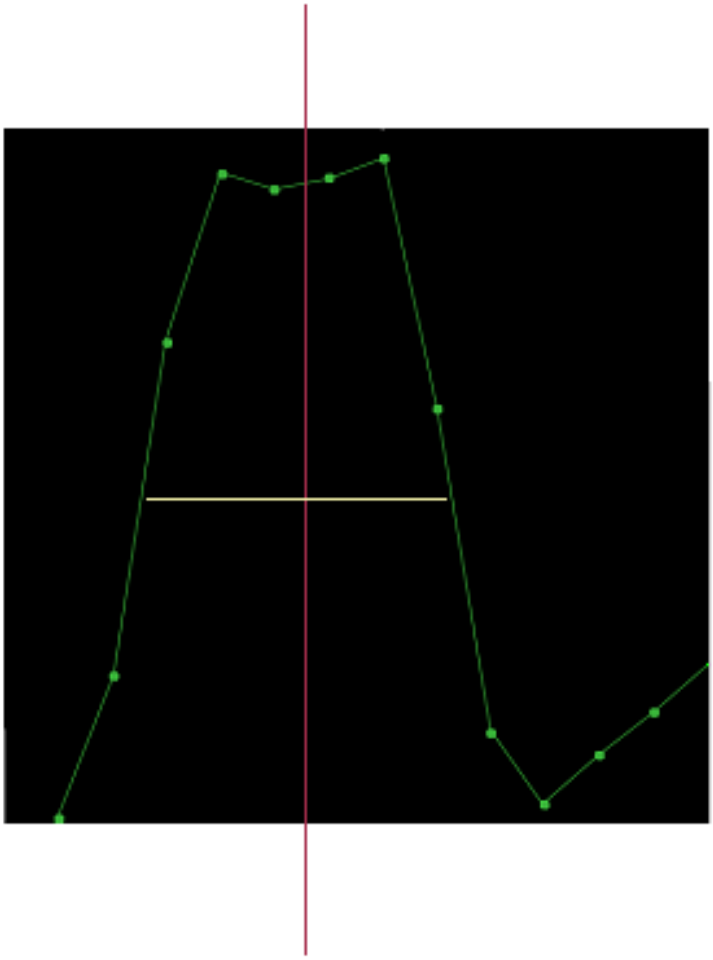


*Fig. 4 : Axial intensity difference profile between center and neighbourhood of a hole (see text).*

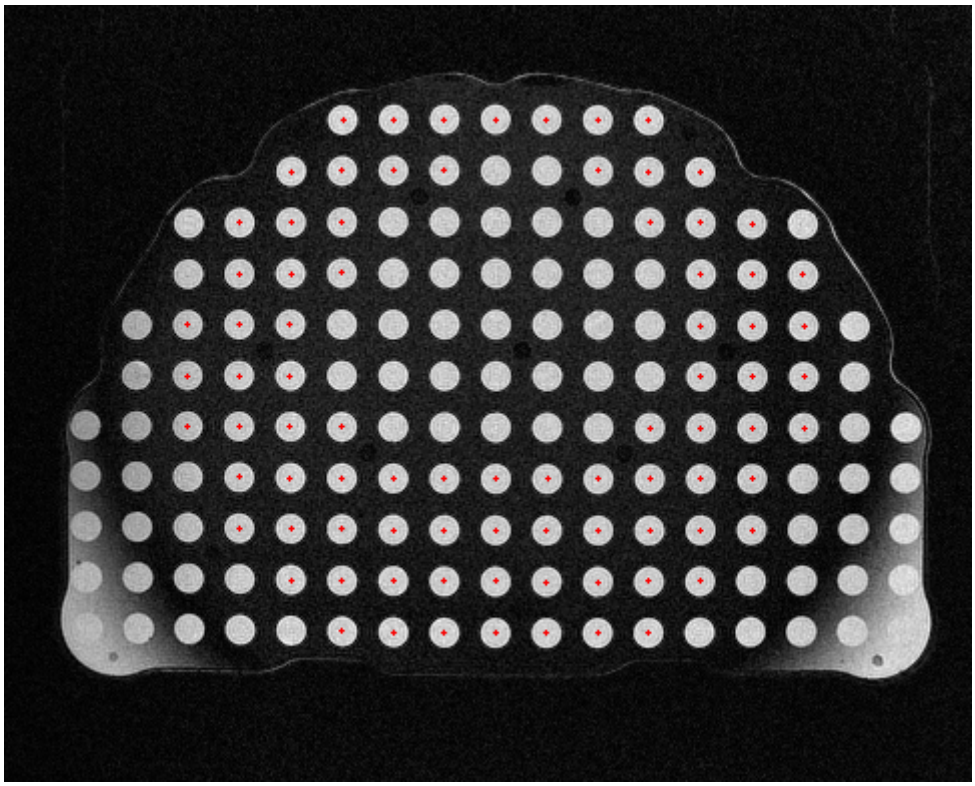


*Fig. 5 : Display of the detected control point positions*

## 2.1.5 Measurements

Mesurements were done with a 1.5 T scanner (Magnetom Symphony, Siemens Medical Solutions, Erlangen, Germany), using two surface coils. The coils integrated in the couch were removed in order to position the phantom closer to the couch and to measure the distortion in this region. The phantom was imaged with a spin echo as well as a flash 3D sequence.
Tab. 1 shows the parameters of these two sequences.

*Tab. 1 : parameters of the MRI sequences.*

| Sequence | Spin echo | Flash 3D |
|---|---|---|
| Slice distance [mm] | 5 | 5 |
| TE [ms] | 97.0 | 4.76 |
| TR [ms] | 1000 | 361 |

### 2.1.6 Correction of the images

To correct the MR images the distortion map is first tri-linearly interpolated to get a displacement vector field which represents the geometric correction in each voxel. This displacement verctor field is then applied to the MR-image. The intensities of the corrected images are obtained by linear interpolation [4].

### 2.1.7 Verification

To verify the over all result of the distortion correction, the algorithm to detect image distortion was again applied to the corrected images of the phantom. The residual geometric distortion is regarded as a measure of the accuracy of the correction.

## 2.2. Uncertainty of elastic image registration

### 2.2.1. Elastic Image registration

Let us regard two images, the test image $f_t$ and the reference image $f_r$:

$$f_t, f_r : Z^3 \to N .$$

The aim of the image registration is to find a displacement vector field (DVF) such that anatomically corresponding voxels of the warped test image (indexed w) and the reference image reach alignment:

$$f_w(x) = f_t(d(x)) \hat{=} f_r(x)$$

Elastic image registration differs from rigid as well as affine registration in the number of degrees of freedom of the transformation. In case of a rigid image registration the

transformation is fully characterized by translation in each direction of space as well as rotation around each of the three axes of the coordinate system. Therefore a rigid registration has 6 degrees of freedom. Affine registration has twelve degrees of freedom since scaling and shearing in all directions are taken into account [13]. The number of degrees of freedom of an elastic image registration is typically much higher than twelve which allows to model complex deformations of the patient anatomy. At the same time a large number of degrees of freedom allows for more possibilities to create wrong deformations.

Some elastic image registration techniques are based on the alignment of contours or on matching landmarks while other approaches are intensity based. The latter group of registration algorithms is driven by intensity differences and intensity gradients. A similarity measure is optimized either directly [4] or in an indirect way [2][3].

## 2.2.2. Two commonly used intensity based algorithms

In this work we focused on b-spline registration as well as the demons algorithm which are both intensity based.

### 2.2.2.1. B-spline registration

#### The basic concept

As proposed in [22] any polynomial spline can be regarded as a superposition of b-spline basis functions. B-spline signal processing has been extensively discussed in [23][24] and its application to multidimensional elastic registration was described in [4].

In elastic b-spline registration the DVF at the voxel position $x \in Z^3$ is represented by

$$d_h^n(x) = \sum_{k \in Z^3} c(k) \cdot \beta^n\left(\frac{x}{h} - k\right),$$

where $\beta^n(x)$ is the basis function of order n, h represents the knot spacing and c(k) represents the coefficients at knot k. In this notation, $d_h^n(x)$ and c(k) represent three-dimensional vectors. The spatial components of these vectors are denoted with an index i ($1 \leq i \leq 3$).

A commonly used metric for mono modality image registration is the SSD (sum of the square differences):

$$SSD = \sum_{x \in \Omega} \left( f_r(x) - f_w(x) \right)^2 ,$$

where $\Omega$ denotes the voxel space of the images.

**Sources of error**

In the following two major sources of errors in b-spline image registration are discussed.

Source 1 : Ambiguity in homogeneous regions

The first partial derivatives of the SSD with respect to the b-spline coefficients were given by [4]:

$$\frac{\partial SSD}{\partial c_{j,m}} = -2 \cdot \sum_{x \in \Omega} \left( f_w(x) - f_r(x) \right) \cdot \left. \frac{\partial f_w(x)}{\partial x_m} \right|_x \cdot \beta\left( \frac{x}{h} - j \right)$$

Note that both, the first as well as second partial derivatives in the coefficients contain components of spatial derivatives of the image intensity. This means that solely those voxels of the test image with nonzero intensity gradient drive the optimization process. Regions without intensity gradient follow passively due to changes of b-spline coefficients in the neighbourhood, but do not guide the deformation process. The metric is not sensitive to misalignments that occur within those homogeneous regions. This limits the registration accuracy as the DVF resulting from the optimization process may not be the only possible result which minimizes the metric and it may therefore deviate from the unknown ground truth.

Source 2 : Misaligned edges

In non-homogenous regions, errors may occur if the optimization does not lead to the global minimum of the SSD-metric. This is the case when the optimization problem is non-convex or if the optimization process ends untimely. In that case corresponding edges may not reach alignment.

Misaligned edges may also occur due to a mismatch of the b-spline model. In this case, the true DVF, $d_{true}(x)$, is not included in the entity of possible b-spline DVFs $\{d_h^n(x)\}$ with degree n and knot spacing h, i.e. $\|d_{true}(x) - d_h^n(x)\| \neq 0$ for the best approximation $d_h^n(x)$.

## 2.2.2.2. The demons algorithm

### The basic concept

The optical flow equation was first given by [25]. It is today often referred to as the demons algorithm due to an interpretation given by [2] where an analogy to Maxwell's Demons is discussed and therefore the optical flow is interpreted as a diffusion process.

Here, the change of the displacement vector $\vec{u}$ in one iteration of the registration process and in some regarded point P of the reference image is given by:

$$\vec{u} = (t-r) \cdot \frac{\vec{\nabla} r}{\left|\vec{\nabla} r\right|^2 + (t-r)^2}$$

where r represents the intensity of the reference image in P and t the intensity of the test image which is warped based on the actual displacement vector field (DVF). $\vec{\nabla} r$ is the intensity gradient in the reference image.

Note that $\vec{u}$ is non-zero only in voxels of the reference image with non-zero intensity gradient. These points drive the registration. To transfer the deformation to

homogenous regions of the image the DVF has to be smoothed. This is done with a Gaussian function after each iteration.

In analogy to the Maxwell's demons a contour can be seen as a membrane. Such a contour is represented by a gradient region which separates a brighter from a darker area of the reference image.

A simple demonstration of the "demons" registration is displayed in Fig. 6. Fig. 6 (a) shows an artificially created reference image. It contains an edge representing a membrane which separates a bright from a dark image area. To demonstrate the registration a shifted version of the same image was created and it is displayed under Fig. 6 (b). It represents the test image. The registration should reverse the shift.

The test image and the reference image are represented in a common coordinate system. So a voxel of the reference image is associated with the point in the warped test image which has identical coordinates. Fig. 6 (a) as well as Fig. 6 (b) represent the same plane in this system. In the region where the intensity gradient of the reference image is non-zero the test image is bright. The direction of $\vec{u}$ is displayed by the grey arrows. Note that an arrow pointing to the left corresponds to a displacement of the test image to the right and vice versa.

The mechanism of the registration is obvious regarding the area where the membrane is located in the reference image. The DVF update $\vec{u}$ is defined such that voxels which belong to the bright area of the test image pass through the membrane solely towards the bright side of the reference image. After alignment is reached, each component of $\vec{u}$ remains zero in further iterations since $(t-r)$ is zero. Fig. 6 (c) shows the deformed test image after "demons registration". To display the deformation, an overlaid regular grid was deformed with the test image according to the DVF.

In case the test image had initially been shifted to the opposite side dark voxels of the test image would have been at the position of the membrane. In this case the sign of $(t-r)$ would have been reversed. Therefore the direction of $\vec{u}$ would have been reversed, such that the dark area in the test image would have passed through the membrane solely towards the dark side of the reference image.

Please note that an initial overlap of corresponding structures is a precondition for a successful registration with the demons algorithm. This issue will be discussed in detail in the next sub-section.

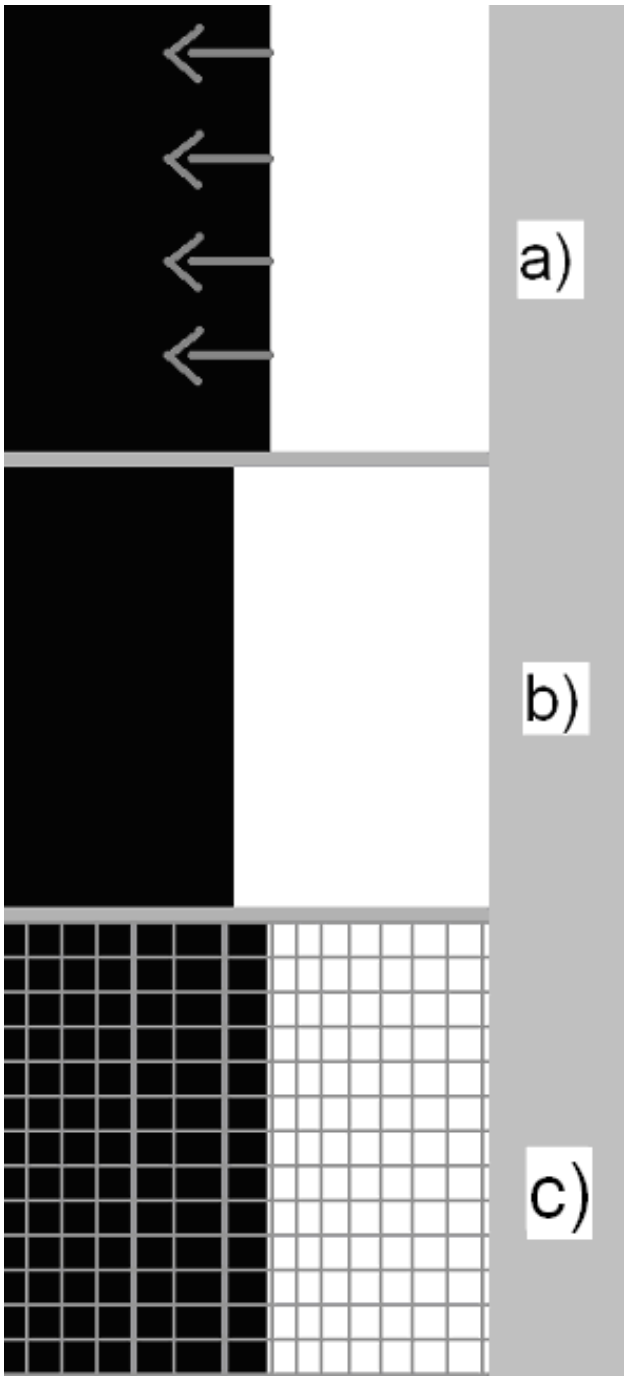



*Fig. 6 : demons registration: those bright voxels of the test image which are located on the left side of the membrane pass through the membrane consisting of a gradient region of the reference image. (a) displays the reference image. The grey arrows represent the direction of* $\vec{u}$ *. (b) shows the test image and (c) represents the warped test image after registration. It was overlaid with a regular grid prior to the deformation.*

**Sources of error**

In the following two mayor sources of error in demons registration are discussed.

Source 1 : Lack of initial overlap

An initial overlap of corresponding structures is essential to ensure the right direction of the displacement $\vec{u}$ described in the previous sub-section.
This is demonstrated in Fig. 7. In Fig. 7 (a) a circular shaped object is displayed. Its difference image with a shifted version of the same object is displayed in Fig. 7 (b). Fig. 7 (c) displays the difference image after 100 iterations of the registration. To display the deformation the test image was overlaid with a regular grid prior to the deformation. In this example the registration reaches alignment. Fig. 7 (d) shows a difference as in (b), but with a larger shift between corresponding structures such that there is no overlap. Fig. 7 (e) displays the difference image after 100 iterations of the "demons" registration. Again the test image was overlaid with a regular grid and then deformed. Alignment could not be reached since initial overlap of corresponding structures was missing.

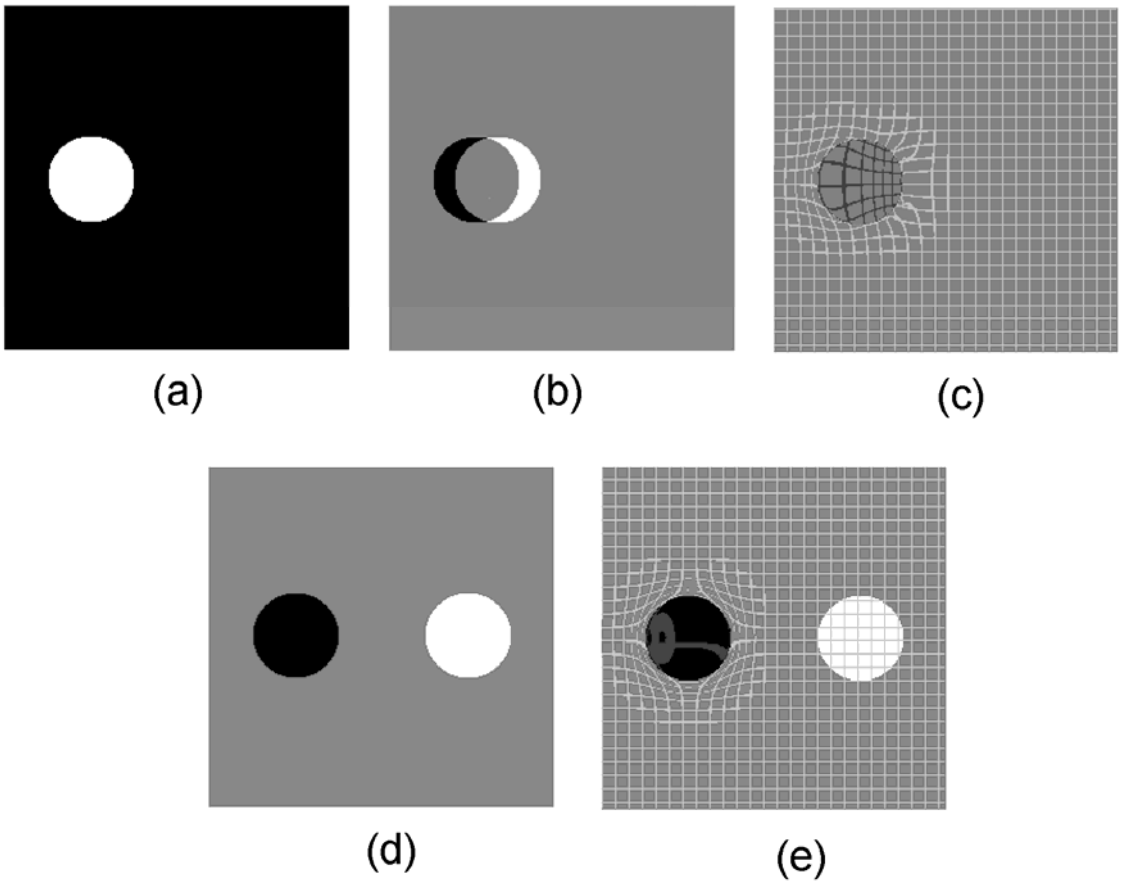



*Fig. 7 : (a) a circular shaped object; (b) difference image of (a) and a shifted version of itself; (c) difference image after 100 iterations; (d) again difference image of (a) and a shifted version of itself, but with larger shift; (e) difference image after 100 iterations. Note: The difference images are calculated by* $I(x) = \left(t(x) - r(x)\right) + \bar{i}$, *where x is a voxel of the reference image., t(x) and r(x) are the intensities of the reference and test image, respectively.* $\bar{i}$ *is an offset of half the magnitude of the difference between the brightest and the darkest voxel in any of the two images.*

It is possible that fine structures such as bronchi initially do not overlap, but reach a state of coarse overlap during the registration process as other larger and initially overlapping structures within the images guide the deformation. The diaphragm may typically be such a larger and initially overlapping structure. Once finer structures have reached coarse overlap the "diffusion process" can take place in the region of their edges and allow further alignment. Structures which do not reach coarse overlap

remain unaligned and are visible as poorly aligned regions in a color overlay or a split screen visualization.

Source 2 : Lack of structure

Voxels in homogenous regions of the reference image can not directly influence the deformation process as $\vec{u}$ is zero in these points due to missing intensity gradients. Still the displacement vectors may be non-zero after the registration due to Gaussian smoothing. The smoothing transfers driving forces that originate from structured regions of the image into the homogenous regions. In these homogenous areas larger registration errors have to be expected than in those voxels that actually guide the deformation process. This uncertainty is a consequence of missing information on the alignment encoded in the image intensities.
This problem can be seen in Fig. 6. Assuming Fig. 6 (a) and Fig. 6 (b) represent the same object with a simple shift, the registration should result in a shift of the whole test image. So in case of a perfect registration the deformed grid in Fig. 6 (c) should be regular after the deformation. However, in further distance from the intensity gradient region the deformation is obviously not correct as the deformed grid is irregular.

## 2.2.3. Estimation of the registration uncertainty

We consider $err_i(x) = |d_i(x) - d_{true,i}(x)|$ to be the local registration error in the spatial dimension i ($1 \leq i \leq 3$), where $d_i(x)$ is the calculated DVF and $d_{true,i}(x)$ the ground truth, which is generally unknown.
In the following two new methods are described to estimate $err_i(x)$. One method refers to the b-spline registration and one method for the demons algorithm.

## 2.2.3.1. B-spline registration

### The basic idea

The uncertainty of the elastic registration is estimated by evaluating the sensitivity of the local metric to moderate and randomly performed variations of the b-spline coefficients which are obtained as result of the b-spline registration.

The local contribution to the global SSD can be calculated from a small region around each voxel. Let $\Delta \subset \Omega$ be a sub-region of the reference image. The contribution $SSD_\Delta$ of $\Delta$ to the global $SSD_\Omega$ is:

$$SSD_\Delta = \sum_{\delta \in \Delta} \left( f_r(\delta) - f_w(\delta) \right)^2 .$$

The aim is to determine the range of geometric deviations which can be performed without increasing the local contribution $SSD_\Delta$ to the overall metric $SSD_\Omega$. In our study $\Delta = x \cdot y \cdot z = 12 \cdot 12 \cdot 18$ mm.

Let $\{c_1,...,c_{3N}\}$ be the set of 3N coefficients resulting from the registration, where N is the number of knots and 3 is the dimension of space. Let $\{r_1,...,r_{3N}\}$ be the set of 3N random variables with the boundaries $b_l$ and $b_u$: $b_l \le r_n \le b_u$ for $1 \le n \le 3N$ . In this study $b_l = -10\ mm$ and $b_u = 10\ mm$ were used. The random value was obtained with the C/C++ function rand().

To generate random test deformations, the coefficients $c_n$ obtained from the registration are replaced by randomly modified coefficients $c_n + r_n$. For each set of modified coefficients, the corresponding DVF is calculated and the test image is deformed accordingly. In the next step the spatial deviation between the randomly modified and the initial deformation is calculated for each dimension of space and each voxel and the local contribution to the global SSD metric is calculated for a region around each voxel. This procedure is repeated K times using different sets of random variations $r_n$.

For each voxel and dimension, the largest deviation between the initial result of the b-spline registration and one of the K test deformations, for which the local SSD is smaller than or equal to the initial local SSD, is stored as a measure of the uncertainty:

$$uncertainty_i(x) = d_{\max,i}(x) = \max_{k=1}^{K} \left\{ \left| d_{k,i}(x) - d_i(x) \right| \right\}$$

where $d_{k,i}(x)$ is one of the K test DVFs and $d_i(x)$ is the result of the b-spline registration. i denotes the dimension of space ($1 \leq i \leq 3$).

The underlying idea of this approach is that an additional random deformation may locally improve the result of the initial registration, which may be affected by the errors described in the following. The maximum spatial deviation of the modified DVFs, which does not increase the local SSD is therefore regarded as a measure of the local registration uncertainty.

**The registration error in image sub-regions**

We do not expect a deterministic dependence between $d_{\max,i}$ and $err_i(x)$ in a specific voxel as the algorithm may have estimated the displacement vector correctly by chance although no image structure is present. In case of statistical dependence, however, the values of $d_{\max,i}$ may allow the estimation of the average of $err_i(x)$ for a larger entity of voxels with similar $d_{\max,i}$ values. Here we explain how to exploit the information that $d_{\max i}$ contains on $err_i(x)$ to divide a dataset in sub-regions that differ in their average local image registration error $err_i(x)$. To do so, the $d_{\max i}$-values are grouped in several intervals and the voxels are classified accordingly. We expect that the average registration error increases with increasing $d_{\max i}$.

**The statistical dependence between $d_{\max,\,i}$ and $err_i(x)$**

The Statistical dependence between $d_{\max,\,i}$ and $err_i(x)$ was investigated with the help of artificially created test data, where the ground truth of the deformation and hence $err_i(x)$ is known. The creation of the test data is described in the next section.

We demonstrate the statistical dependence between $d_{\max,i}$ and $err_i(x)$ by calculating the mutual information (MI) from their marginal and joint distributions. For this, a histogram of equally-sized bins is created for the $d_{\max,i}$- as well as for the $err_i(x)$-values. Let $p_d(n)$ be the probability that a $d_{\max,\,i}$ value belongs to the n-th bin of the histogram of $d_{\max,i}$-values and $p_{err}(m)$ the probability that $err_i(x)$ belongs to the m-th bin of the histogram of $err_i(x)$ values. Let p(m,n) be the probability of the joint event that $d_{\max,i}$ contributes to bin n and $err_i(x)$ to bin m.

The mutual information is then calculated by

$$MI = \sum_{n,m=1}^{N,M} p(m,n) \cdot \log_2 \frac{p(m,n)}{p_d(n) \cdot p_{err}(m)}.$$

N is the number of bins of the $d_{\max,i}$ histogram and M the number of bins of the $err_i(x)$ histogram.

If there is no statistical dependence, the MI should be equal to zero. To demonstrate the statistical dependence, we use the artificially deformed test images for which the ground truth on the deformation is known:

After registering these with the un-deformed images and applying the algorithm described above, a field of $d_{\max,i}$ values is obtained for each dimension of space i. As the deformation of the test data was pre-defined, the ground truth of the deformation $d_{true\,,i}(x)$ and hence the local registration error $err_i(x)$ is known. Subsequently, the initial MI is calculated.

As we are dealing with real world data and therefore with probability distributions that are estimated based on a limited sample, it is not justified to interpret any deviation of

the MI from zero as a statistical dependence. To demonstrate the significance of the increase, the initial $d_{\max,i}$-values are randomly re-distributed over the voxels, conserving the number of entries per bin and hence the shape of the distribution $p_d(n)$. Then the joint probability distribution p(m,n) as well as the MI are recalculated. After this process, no statistical dependence between $d_{\max,i}$ and $err_i(x)$ can be expected and any deviation of the MI from zero is to be regarded as noise. This procedure was repeated 200 times and so the random distribution of the MI-values for the case of statistical independence is estimated. $d_{\max,i}$ and $err_i(x)$ are considered to be statistically dependent, if the initial MI-value is very unlikely to occur according to the distribution of MI-values obtained for the case of statistical independence.

### 2.2.3.1. Demons algorithm

Since the demons algorithm is a non-parameterized method the approach described in the previous section can not be applied. Therefore we propose a second method to capture the uncertainty of the registration which takes both types of errors discussed in the section 2.2.2.2 into account.
The basic idea is to regard the reproducibility of the deformation field as a local figure of merit for the geometric accuracy of the image registration. The approach consists of three steps:

**Step 1: Initial registration**

First an initial registration is done resulting in a DVF which is stored and later on used for any purpose which the registration is needed for.

**Step 2: Repeated registrations from different starting points**

In a second step, after the initial registration an offset is added to each component of each displacement vector which resulted from the initial registration (step 1). The offset is the same in each voxel and for each component of the displacement vector. Therefore, warping the test image with the modified DVF results in a simple shift,

compared to warping with the initial DVF and without modification. The registration is then continued.
Since the demons algorithm warps the test image with the current DVF in each iteration step this process can be regarded as one registration during which the deformation field was perturbed by adding an offset.
The resulting DVF is stored. This process is repeated eight times with offsets that differ in their direction. Each time the resulting DVF is stored. In our study the magnitude of the offset was 3mm for each component of the DVF and in each voxel.

**Step 3: Calculation of the standard deviation of the displacement vector components**

After these eight registrations are finished the standard deviation of each component of the displacement vector is calculated in each voxel. All nine registrations, including the initial one are taken into account.

**Practical application**

We propose to divide the image into sub regions that differ in the magnitude of their average registration error by classifying the voxels according to the standard deviation of the resulting displacement vector components. The larger the standard deviation, the less reproducible the result and the more likely are large registration errors. This can either be due to missing initial overlap of corresponding structures which may lead to misaligned edges or due to missing structure in the images.
The image sub-regions of different average registration error in direction i of the space $0 < i \leq 3$ are obtained by binning the voxels in a histogram of standard deviation values. The voxels of one bin represent one region. This region is not necessarily spatially connected.
In the results section we demonstrate that the average local registration error of all the voxels in one bin increases with increasing standard deviation and hence with the bin number.

## 2.2.4 Test of the algorithms

### 2.2.4.1. Generation of test data

The algorithms were tested on lung datasets. To generate test data with known ground truth of the deformation, 5 clinical lung data sets were artificially deformed. This deformation aims to model the transition of the exhale to the inhale breathing phase. To take the main physiological aspects of breathing into account, the artificial deformations consist of the following components:

1) Extension of the chest in the transversal plane
2) Decompression of the lung in cranio-caudal direction
3) Random deformation
4) Tissue sliding between lung and rip cage

Although these four steps cannot be considered to fully describe all aspects of lung motion, it is regarded as a model, which describes the main physiological components, while providing the ground truth of the deformation.
In the following, the realization of the deformation is described. For each of the steps a DVF is generated and the total DVF is calculated as a superposition of these components. To model tissue sliding, two DVFs are generated. One aims to model the deformation in regions within the chest wall and a second one aims to model the deformation outside the chest wall. An additional step is necessary to combine both deformation fields without folding or tearing in the boundary region.

1) Extension of the chest in the transversal plane

The extension was performed by linear scaling. Let $z_0$ be a position in the region of the diaphragm. Scaling with the factor $g_{xy} = 1 + s_0$ was applied for all $z < z_0$ where $z - z_0$ is the distance from the diaphragm in cranial direction. For $z_0 \leq z < z_0 + \Delta$, the scaling factor was

$$g_{xy}(z) = 1 + s_0 \cdot \left( 1 - \frac{\log\left(1 + (z - z_0) \cdot \alpha\right)}{\log(1 + \Delta \cdot \alpha)} \right),$$

where $\alpha \in R$ is a constant, and $\Delta$ is the range over which $g_{xy}(z)$ decays. For $z \geq z_0 + \Delta$, we use $g_{x,y}(z) = 1$. So the magnitude of $g_{xy}(z)$ is largest in the region of the diaphragm and decreases in cranial direction.

2) Decompression of the lung in cranio-caudal direction

During the transition between inhale and exhale breathing phase, the diaphragm is moved in caudal direction. In the model, this is described by a displacement q(z) in caudal direction:

$q(z) = t_0$ for all $z < z_0$,

$$q(z) = t_0 \cdot \left( 1 - \frac{\log\left(1 + (z - z_0) \cdot \alpha\right)}{\log(1 + \Delta \cdot \alpha)} \right)$$

for $z_0 \leq z < z_0 + \Delta$, and

$q(z) = 0$ for $z \geq z_0 + \Delta$

The closer to the diaphragm, the larger is the displacement. In distance $\Delta$ from the diaphragm, the displacement is zero.

3) Random deformation

In order to obtain an additional deformation which is not regular throughout each slice, an additional three dimensional DVF is randomly created and added to the sum of the deformations which resulted form step one and two. For this, Gaussian functions, were used as base functions and the coefficients were generated randomly. This DVF cannot generally be described as a superposition of b-spline basis functions and hence a model mismatch can be expected.

4) Tissue sliding between lung and rip cage

The deformation inside the chest wall, $DVF_{internal}$, is created as a superposition of the deformation steps one to three. It is considered to describe the *internal* deformation of the lung from exhale to inhale. Due to tissue sliding, however, $DVF_{internal}$ does not describe the deformation outside the chest wall. Therefore, a second deformation field, $DVF_{external}$, is obtained from $DVF_{internal}$ by setting the cranio-caudal component to zero in each voxel. $DVF_{external}$ is considered to describe the *external* deformation in the region of the rips as well as outside the chest wall. $DVF_{external}$ does not describe the deformation of the lung tissue.

To create a combined DVF which approximates the deformation of the anatomy all over the image without folding or tearing, a third step is necessary. In the following we describe an approach to modify $DVF_{internal}$ such that the boundary surface of the lung after warping with $DVF_{internal}$ gets mapped to the same surface as after warping the image based on $DVF_{external}$. This modification allows a simple combination of both DVFs just by using $DVF_{internal}$ for the deformation of the inside and $DVF_{external}$ for the deformation outside the chest wall.
To realize this concept, a mask is created to distinguish the regions inside and outside of the chest wall. The mask is a binary image and allows the calculation of intensity gradients. In a first step, this mask is deformed based on $DVF_{internal}$ and registered with the mask deformed based on $DVF_{external}$. As a result, the deformation field $DVF_{mask}$ is obtained.
This registration was done with the ITK demons implementation. We chose this non-parameterized method, as b-spline deformations should not be involved in creating the test data.
Finally, $DVF_{internal}$ is replaced by $DVF_{internal} + DVF_{mask}$. After this slight modification, a simple combination of $DVF_{internal}$ and $DVF_{external}$ is possible without folding or tearing in the boundary region. Nevertheless, the resulting DVF contains a discontinuity in the region of the pleura which represents tissue sliding.

### 2.2.4.2. Application of the algorithms on test data

The developed algorithms were tested using five lung cases. To generate the test data, the exhale breathing phase from a 4D-CT was used as a starting point and the DVF described above was created using the parameters of Tab. 2. This DVF was then used to simulate the inhale image from the exhale image. For the test of the algorithm, the exhale and simulated inhale images were registered using the inhale image as reference and $d_{\max\, i}$ was calculated for each voxel. As the underlying ground truth of the deformation is given by the artificially created DVF, the relation between $d_{\max\, i}$ and the local registration error $err_i(x)$ can be analyzed.

*Tab. 2 : Parameters used for the generation of the test cases*

| **Parameter** | **Value** |
|---|---|
| $t_0$ | 20 mm |
| $\Delta$ | ≈Extension of lung |
| $\alpha$ | $1.3/\Delta$ |
| $s_0$ | ≈0.12 |
| Spacing of Gaussian Functions | 16 voxels |

# 3. Results

## 3.1. MRI distortion correction

A method to detect and correct distortion in MRI was developed and tested for different sequences.

### 3.1.1 The spin echo sequence

Fig. 8 shows the distribution of the magnitude of the 3D-distortions measured in the control points. In 59% of the control points the distortion was less than 3 mm and in 75 % of the control points the distortion was less than 5 mm however, distortions larger than 20 mm were detected in some control points. The average 3D distortion measured to be 2.9 ± 3.6 mm.

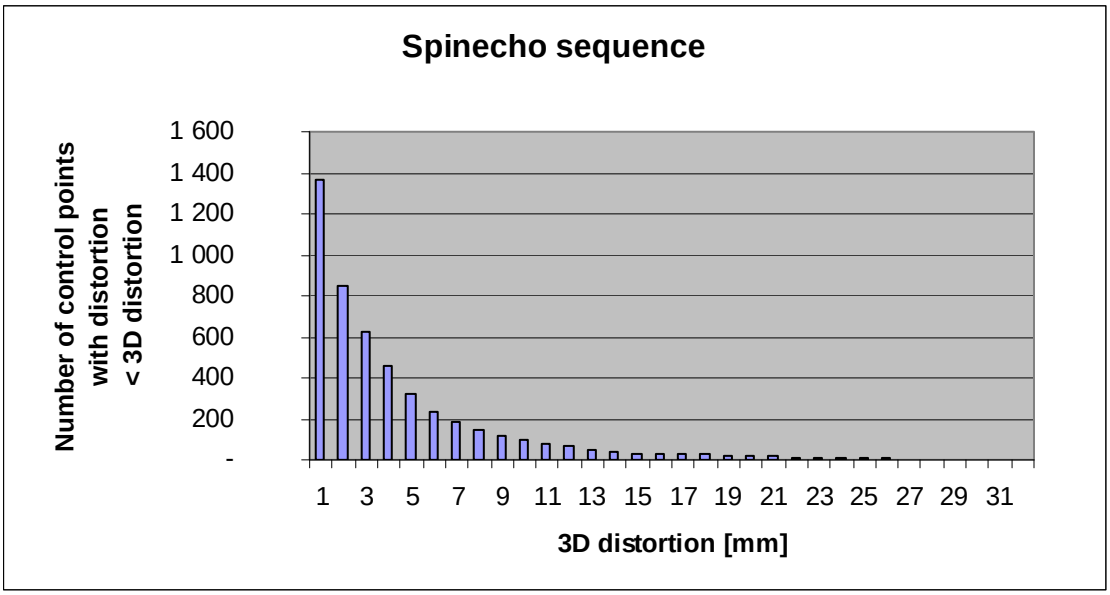



*Fig. 8 : The distribution of the magnitude of the3D distortions for the spin echo sequence.*

Tab. 3 displays the distortions measured for the spin echo sequence before and after correction. For a compact presentation only the values for the control points of the central slice of each module are shown. In module one and five, which are located in the largest distance from the center of the MRI-device, the distortions were largest with values up to 32.6 mm. In the central slice of the central module distortions larger than 2 mm were not detected. The axial component of the distortion is zero by definition in this central slice.

*Tab. 3 : Comparison of the distortions measured for the spin echo sequence before and after geometric correction of the images for the central slice of each module of the phantom. The z-coordinate denotes the axial position relative to the center of the MRI device.*

| | **Distortion[mm] Mean±SD (maximum)** | |
|---|---|---|
| | **Before correction** | **After correction** |
| **Module 1 ( z = -180 mm )** | | |
| 3D | 7.6 ± 5.9 (32.6) | 0.8 ± 1.3 (15.8) |
| Lateral | 4.2 ± 5.1 (25.5) | 0.4 ± 1.0 (12.6) |
| Vertical | 2.6 ± 2.9 (18.3) | 0.4 ± 0.7 (7.8) |
| Axial | 4.8 ± 3.5 (14.2) | 0.3 ± 0.5 (5.4) |
| **Module 2 ( z = -90 mm )** | | |
| 3D | 1.7 ± 1.7 (10.5) | 0.4 ± 0.3 (1.7) |
| Lateral | 0.7 ± 0.5 (4.3) | 0.1 ± 0.1 (0.7) |
| Vertical | 0.6 ± 0.5 (3.9) | 0.3 ± 0.2 (0.9) |
| Axial | 1.1 ± 1.7 (9.4) | 0.2 ± 0.2 (1.3) |
| **Module 3 ( z = 0 mm )** | | |
| 3D | 0.5 ± 0.3 (1.4) | 0.2 ± 0.1 (1.2) |
| Lateral | 0.2 ± 0.2 (0.8) | 0.1 ± 0.1 (0.3) |
| Vertical | 0.4 ± 0.3 (1.4) | 0.2 ± 0.1 (1.2) |
| Axial | 0.0 ± 0.0 (0.0) | 0.0 ± 0.0 (0.0) |
| **Module 4 ( z = 90 mm )** | | |
| 3D | 1.8 ± 1.7 (10.2) | 0.3 ± 0.2 (2.0) |
| Lateral | 0.5 ± 0.6 (4.3) | 0.1 ± 0.1 (0.8) |
| Vertical | 0.9 ± 0.5 (4.2) | 0.2 ± 0.2 (1.9) |
| Axial | 1.2 ± 1.8 (9.0) | 0.2 ± 0.2 (1.2) |
| **Module 5 ( z = 180 mm )** | | |
| 3D | 7.5 ± 5.3 (26.7) | 0.9 ± 1.4 (13.0) |
| Lateral | 3.5 ± 4.3 (20.6) | 0.4 ± 1.0 (9.1) |
| Vertical | 2.7 ± 2.6 (14.5) | 0.5 ± 0.8 (7.5) |
| Axial | 5.2 ± 3.5 (14.6) | 0.5 ± 0.7 (5.5) |

Tab. 4 compares the distortions measured for the spin echo sequence before and after geometric correction as a function of the radial distance from the center of the MRI-device. The magnitude of the distortions increases with increasing distance from the isocenter.

*Tab. 4 : Distortions measured for the spin echo sequence before and after geometric correction as a function of the radial distance from the isocenter.*

| | **Distortion[mm] Mean±SD (maximum)** | |
|---|---|---|
| | **Before correction** | **After correction** |
| **Distance < 100 mm** | | |
| 3D | 0.6 ± 0.3 (1.2) | 0.2 ± 0.1 (0.6) |
| Lateral | 0.2 ± 0.2 (0.7) | 0.0 ± 0.0 (0.2) |
| Vertical | 0.5 ± 0.3 (1.2) | 0.1 ± 0.1 (0.6) |
| Axial | 0.1 ± 0.1 (0.4) | 0.0 ± 0.0 (0.3) |
| **Distance < 200 mm** | | |
| 3D | 1.4 ± 1.0 (5.7) | 0.3 ± 0.2 (2.0) |
| Lateral | 0.6± 0.7(5.5) | 0.1 ± 0.1 (0.7) |
| Vertical | 0.8± 0.6 (4.4) | 0.2 ± 0.2 (1.9) |
| Axial | 0.8 ± 0.8 (3.9) | 0.2 ± 0.2 (1.1) |
| **Distance < 300 mm** | | |
| 3D | 2.9 ± 3.6 (32.6) | 0.4 ± 0.6 (15.8) |
| Lateral | 1.4 ± 2.7 (25.5) | 0.2 ± 0.4 (12.6) |
| Vertical | 1.2 ± 1.5 (18.3) | 0.3 ± 0.3 (7.8) |
| Axial | 1.7 ± 2.3 (14.6) | 0.2 ± 0.3 (5.5) |

Fig. 9 shows examples of overlays of the undistorted control point grid defined by the physical phantom (green) with the measured grid (red). Fig. 9 (a) displays the central slice of the first module of the phantom where the magnitude of the distortions are large compared to the central slice of the central module (Fig. 9 (b) ). The distortions measured for the central slice of the last module ( Fig. 9 (c) ) are similar to those found in the first module ( Fig. 9 (a) ) due to the symmetry of the magnetic fields. The distortions in the transversal plane increase with the axial as well as the lateral distance from the isocenter.

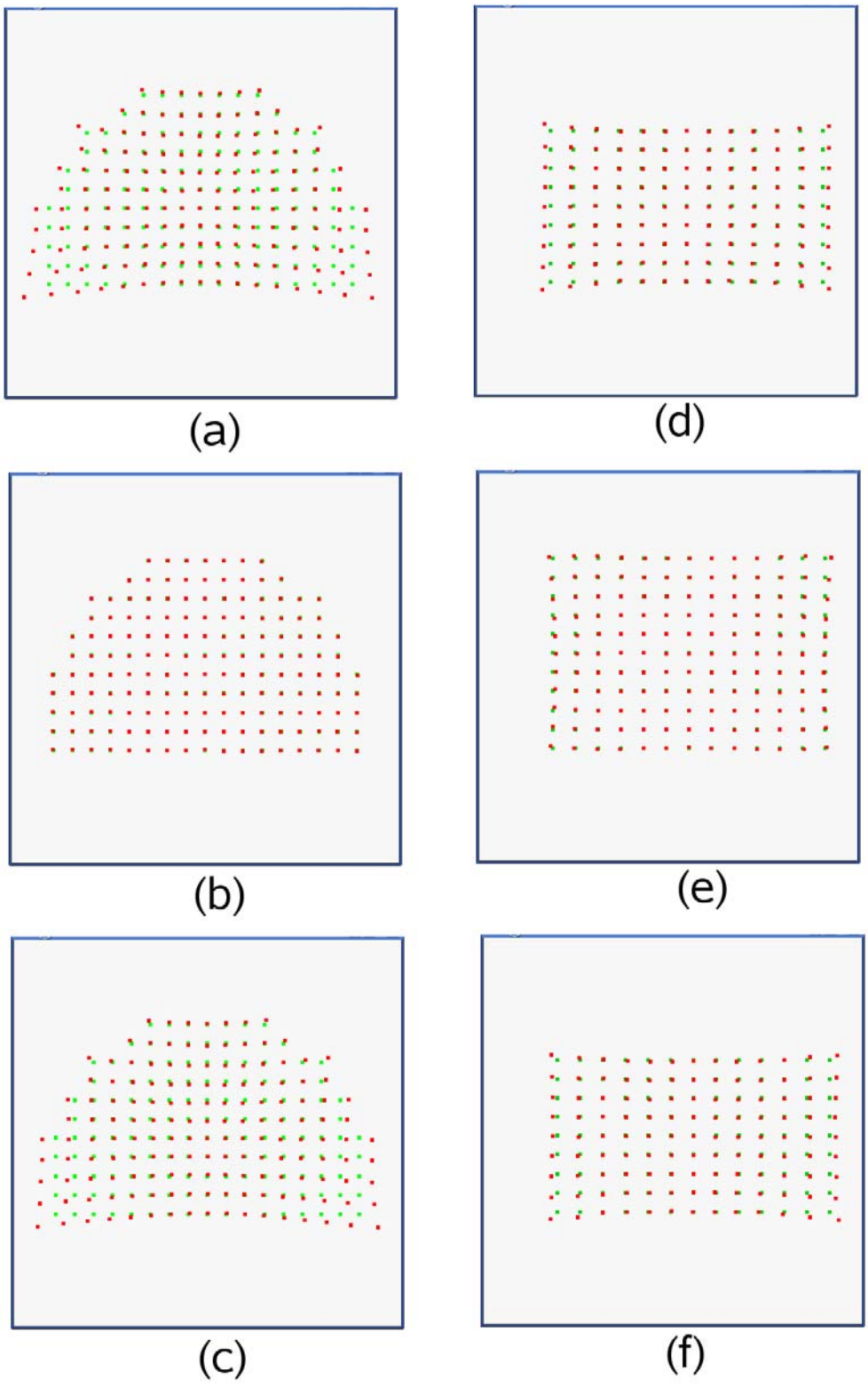



*Fig. 9 : Transversal view of the distortion map for the central slice of the first (a), central (b) and last (c) module of the phantom. In addition, the sagital view for the central (e) plane as well as for planes, shifted laterally by 12.5 cm to the left (d) and to the right (f) are shown.*

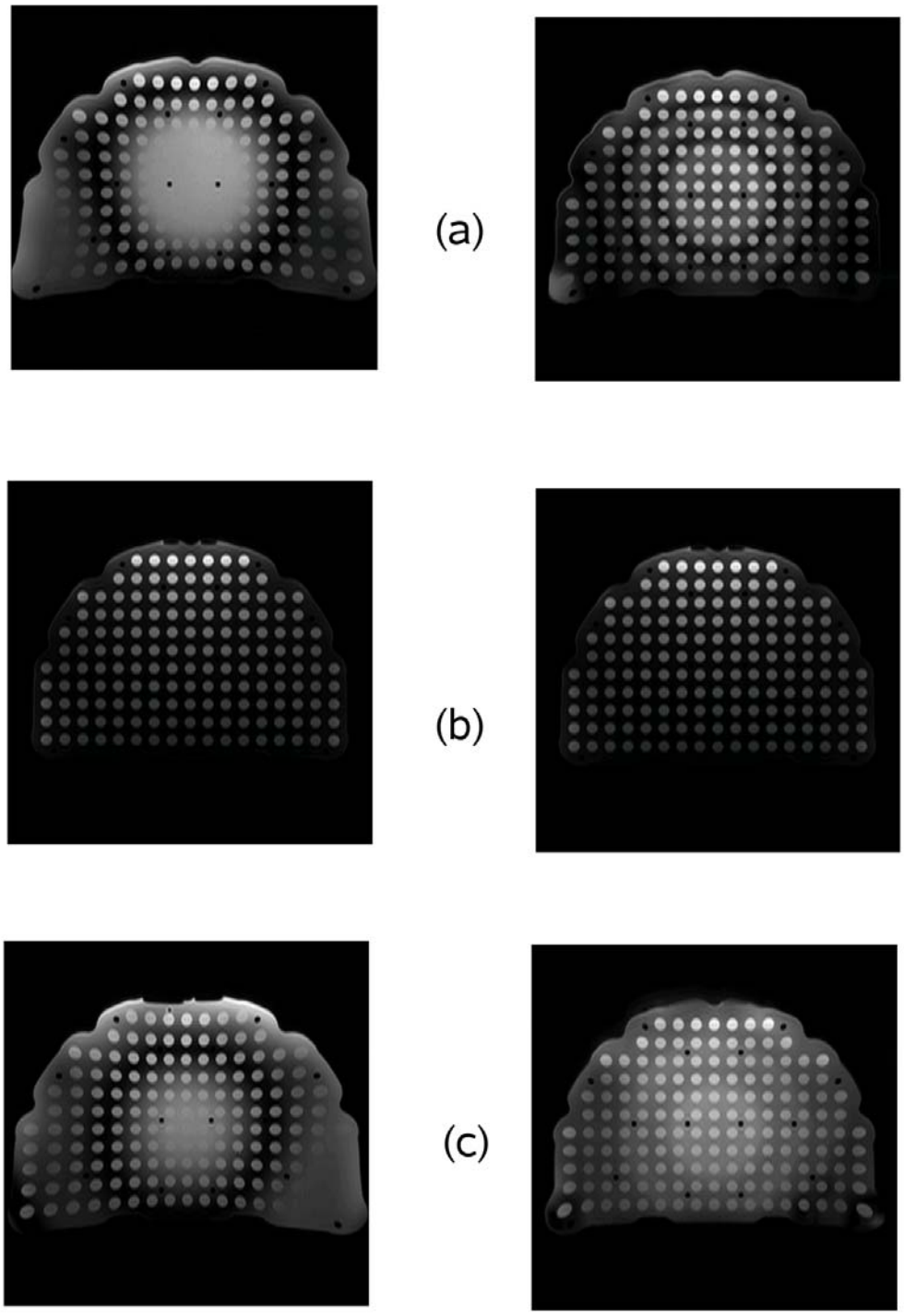



*Fig. 10 : Uncorrected (left) and corrected (right) image of the central slice of the first (a), central (b) and last (c) module of the phantom.*

Fig. 10 shows the uncorrected and corrected CT-images for different slices of the phantom. After geometric correction, the distortions are mainly removed. The corrected images show a regular grid of control points.

## 3.1.2 The flash 3D sequence

Fig. 11 shows the distribution of the magnitude of the 3D distortion which was measured in the control points for the flash 3D image. The distribution looks very similar to the one obtained from the spin echo sequence.

In 57 % of the control points the distortion was less than 3 mm and in 74 % of the control points the distortion was less than 5 mm.
The average 3D distortion was 3.0 ± 3.6.

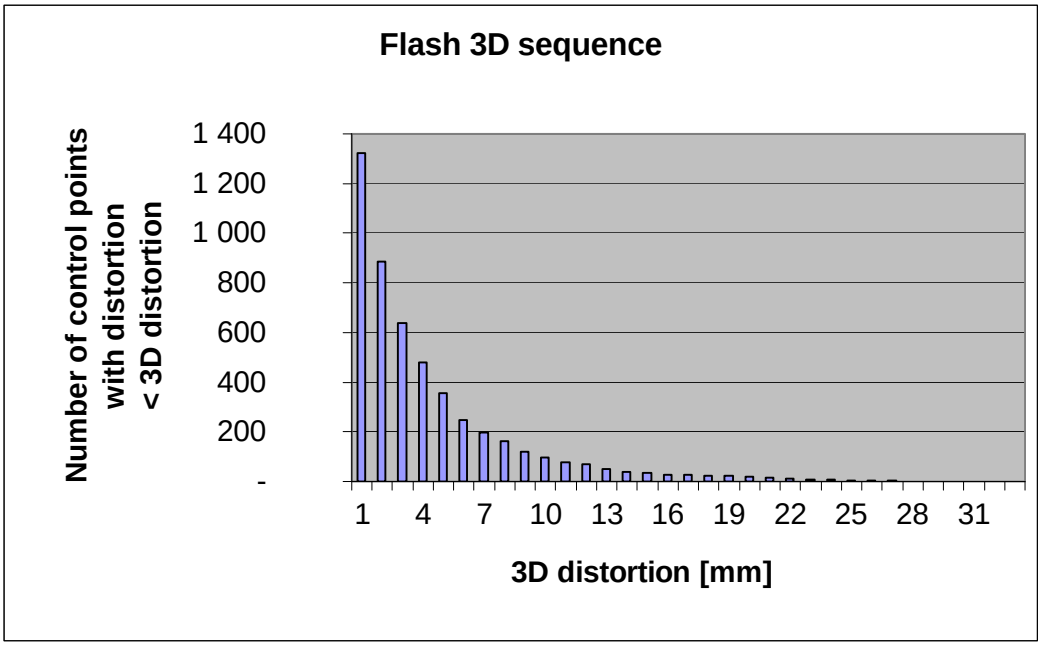



*Fig. 11 : The distribution of the magnitude of the3D distortion over the control points obtained for the flash 3D sequence.*

Tab. 5 displays the distortions measured in the flash 3D sequence before and after geometric correction. For a compact presentation only the values for control points of the central slice of each module are shown. Like in the spin echo sequence the

distortion was largest in modules with the largest distance from the center of the MRI-device.

*Tab. 5 : Comparison of the distortion measured for the flash 3D sequence before and after geometric correction for the central slice of each module of the phantom. The z-coordinate denotes the axial position relative to the center of the MRI device.*

| | **Distortion[mm] Mean±SD (maximum)** | |
|---|---|---|
| | **Before correction** | **After correction** |
| **Module 1 ( z = -180 mm )** | | |
| 3D | 7.7 ± 5.3 (27.5) | 0.9 ± 1.4 (14.7) |
| Lateral | 3.8 ± 4.5 (22.2) | 0.5 ± 1.1 (11.1) |
| Vertical | 2.5 ± 2.4 (13.6) | 0.5 ± 0.7 (7.1) |
| Axial | 5.2 ± 3.6 (13.2) | 0.4 ± 0.6 (6.6) |
| **Module 2 ( z = -90 mm )** | | |
| 3D | 1.8 ± 1.7 (10.0) | 0.5 ± 0.2 (2.3) |
| Lateral | 0.6 ± 0.5 (3.6) | 0.2 ± 0.1 (0.4) |
| Vertical | 0.6 ± 0.6 (4.8) | 0.3 ± 0.1 (0.8) |
| Axial | 1.4 ± 1.7 (8.7) | 0.2 ± 0.3 (2.2) |
| **Module 3 ( z = 0 mm )** | | |
| 3D | 0.6 ± 0.3 (1.4) | 0.4 ± 0.1 (0.7) |
| Lateral | 0.2 ± 0.2 (1.0) | 0.2 ± 0.1 (0.5) |
| Vertical | 0.5 ± 0.3 (1.4) | 0.3 ± 0.1 (0.7) |
| Axial | 0.0 ± 0.0 (0.0) | 0.0 ± 0.0 (0.0) |
| **Module 4 ( z = 90 mm )** | | |
| 3D | 1.7 ± 1.7 (10.7) | 0.5 ± 0.2 (1.7) |
| Lateral | 0.6 ± 0.6 (3.7) | 0.2 ± 0.1 (0.5) |
| Vertical | 0.7 ± 0.6 (4.9) | 0.3 ± 0.1 (0.7) |
| Axial | 1.1 ± 1.7 (8.8) | 0.2 ± 0.2 (1.7) |
| **Module 5 ( z = 180 mm )** | | |
| 3D | 8.0 ± 5.6 (30.1) | 0.7 ± 0.4 (2.9) |
| Lateral | 3.5 ± 4.2 (19.1) | 0.3 ± 0.3 (1.6) |
| Vertical | 2.7 ± 3.0 (18.0) | 0.4 ± 0.3 (1.4) |
| Axial | 5.7 ± 4.0 (15.2) | 0.4 ± 0.3 (2.6) |

Tab. 6 shows the distortion measured in the flash 3D sequence before and after geometric correction as a function of the radial distance from the center of the MRI-device. The distortion increases with increasing distance from the center.

*Tab. 6 : Comparison of the distortions measured in the flash 3D sequence before and after geometric correction as a function of the distance from the isocenter*

| | **Distortion[mm] Mean±SD (maximum)** | |
|---|---|---|
| | **Before correction** | **After correction** |
| **Distance < 100 mm** | | |
| 3D | 0.5 ± 0.2 (0.9) | 0.4 ± 0.1 (0.6) |
| Lateral | 0.1 ± 0.1 (0.5) | 0.1 ± 0.0 (0.3) |
| Vertical | 0.5 ± 0.2 (0.8) | 0.3 ± 0.1 (0.6) |
| Axial | 0.1 ± 0.1 (0.6) | 0.1 ± 0.1 (0.4) |
| **Distance < 200 mm** | | |
| 3D | 1.4 ± 1.0 (5.5) | 0.4 ± 0.1 (2.3) |
| Lateral | 0.5 ± 0.7 (5.0) | 0.1 ± 0.1 (0.5) |
| Vertical | 0.7 ± 0.6 (4.5) | 0.3 ± 0.1 (1.0) |
| Axial | 0.8 ± 0.9 (3.9) | 0.2 ± 0.2 (2.3) |
| **Distance < 300 mm** | | |
| 3D | 3.0 ± 3.6 (30.1) | 0.5 ± 0.5 (14.7) |
| Lateral | 1.3 ± 2.5 (22.2) | 0.2 ± 0.3 (11.1) |
| Vertical | 1.1 ± 1.5 (18.0) | 0.4 ± 0.2 (7.1) |
| Axial | 1.9 ± 2.5 (15.2) | 0.2 ± 0.3 (6.6) |

# 3.2. Elastic image registration

## 3.2.1. The test data

Fig. 12 shows an exhale (a) and a simulated inhale (b) image in comparison to a real inhale image from the 4D-CT (c). The displacement of the diaphragm as well as the tissue sliding between lung and chest wall is similar in the true and the simulated inhale image.

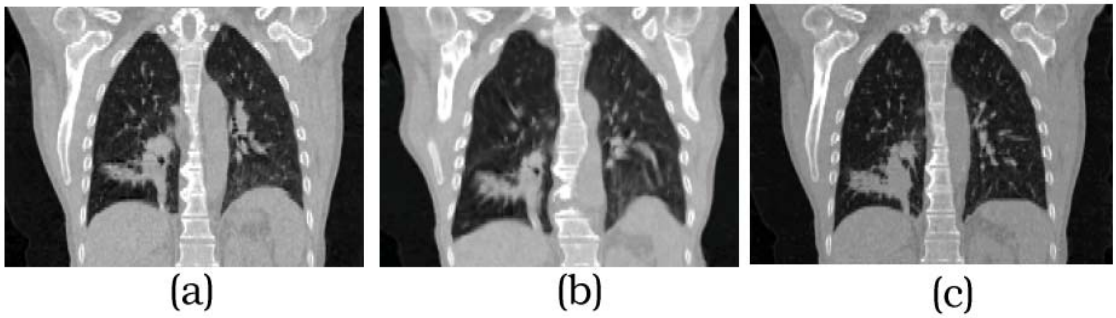



*Fig. 12 : Exhale image from the 4D-CT (a) and inhale image (b) simulated on the basis of the exhale image (a) using the method described in section 2.2.4.1. For comparison an inhale image from the 4D-CT is shown in (c).*

Fig. 13 displays the difference image between the mask deformed based on $DVF_{internal}$ and the mask deformed based on $DVF_{external}$ (a). After applying the final deformation $DVF_{mask}$ (b), both masks coincide in the region of the chest wall. So the chest wall is mapped to the same surface by both deformation fields.

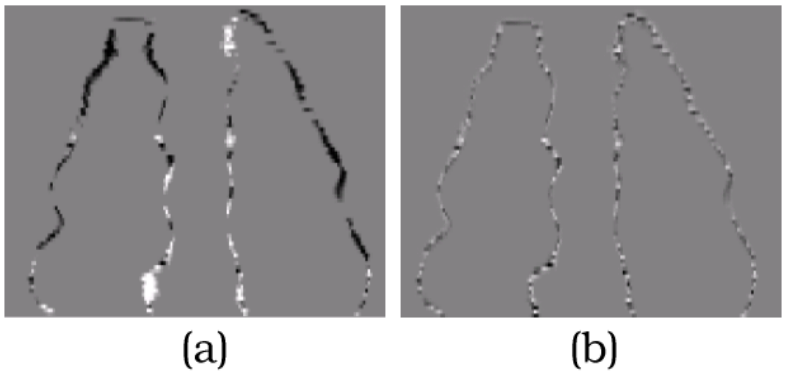



*Fig. 13 : (a) Difference image of the masks for the regions inside and outside the chest wall after deformation with $DVF_{internal}$ and $DVF_{external}$, respectively. (b) The same difference after modifying $DVF_{internal}$ by adding $DVF_{mask}$. (see section 2.2.4.1. ).*

## 3.2.2. Estimation of the registration uncertainty

### 3.2.2.1. The b-spline algorithm

#### Implementation of the algorithm

The b-spline algorithm was implemented using C++ under Linux. It is a dkfz in house implementation of the algorithm described in [4]. This algorithm was enhanced to estimate the registration uncertainty as described above.
The registrations took around 3 minutes for lung data sets with about 256x256x80 voxels that were processed on image resolutions four to two. The final spacing of the b-spline basis functions was 8 voxels in plane and 4 voxels cross plane.
The time for the estimation of the registration uncertainty is proportional to the number of test deformations. In this study we applied 400 test deformations. It took 15 minutes on a PC with a 1.85 GHz processor and without parallel computing.

#### Validation of the b-spline registration with test data – actual registration errors

After registration of the simulated inhale image with the initial exhale image, the resulting DVF was compared with the ground truth and the local registration error was calculated in each point. Tab. 7 shows the average registration error over the voxels of the complete body region for each of the five test cases.

*Tab. 7 : Average registration error and standard deviation for each of the test cases*

| Test case | Mean±SD [mm] | | |
|---|---|---|---|
| | lateral | anterior-posterior | Caudal-cranial |
| 1 | 2.2±2.1 | 2.2±2.1 | 3.8±4.1 |
| 2 | 2.0±2.0 | 2.1±1.9 | 3.3±3.4 |
| 3 | 2.7±2.6 | 2.6±2.5 | 3.6±3.9 |
| 4 | 2.8±2.5 | 2.4±2.3 | 3.8±3.9 |
| 5 | 2.6±2.2 | 2.7±2.4 | 3.7±3.7 |

**Estimation of the registration error for sub regions**

Fig. 14 shows the combined histogram for the $d_{max}$-values and the corresponding true local registration error for case 1. The brightness of the entries represents the number of voxels with the respective combination of $d_{max}$ and err. The histograms demonstrate, how the proposed quantity $d_{max}$ should be interpreted: Although large registration errors may occur for large $d_{max}$-values, small errors are very likely as well, since the algorithm may have estimated the deformation correctly by chance although no image structure is available. Large errors, however, are very unlikely in voxels with small $d_{max}$-values.

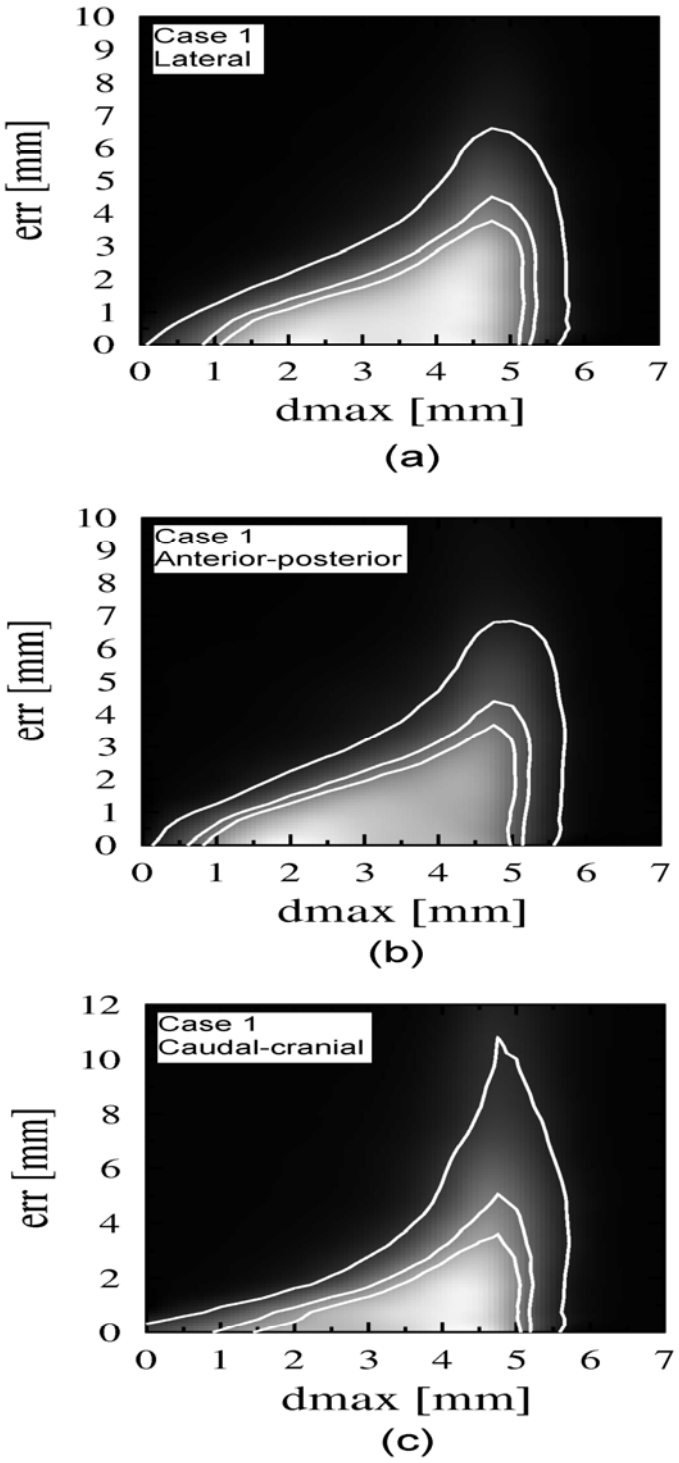



*Fig. 14 : Combined histogram for the $d_{max}$-values and the corresponding true local registration error for case 1.*

As a typical example, the left side of Fig. 15 displays the average local registration errors obtained from the ground truth for case 1 as a function of $d_{max}$ using 15 intervals. It shows that the average registration error increases with increasing values of $d_{max}$. This was found for each dimension of space.

The right side Fig. 15 displays the standard deviation of the local registration error for the same case and the same intervals of $d_{max}$. Large $d_{max}$–values correspond to large standard deviations.

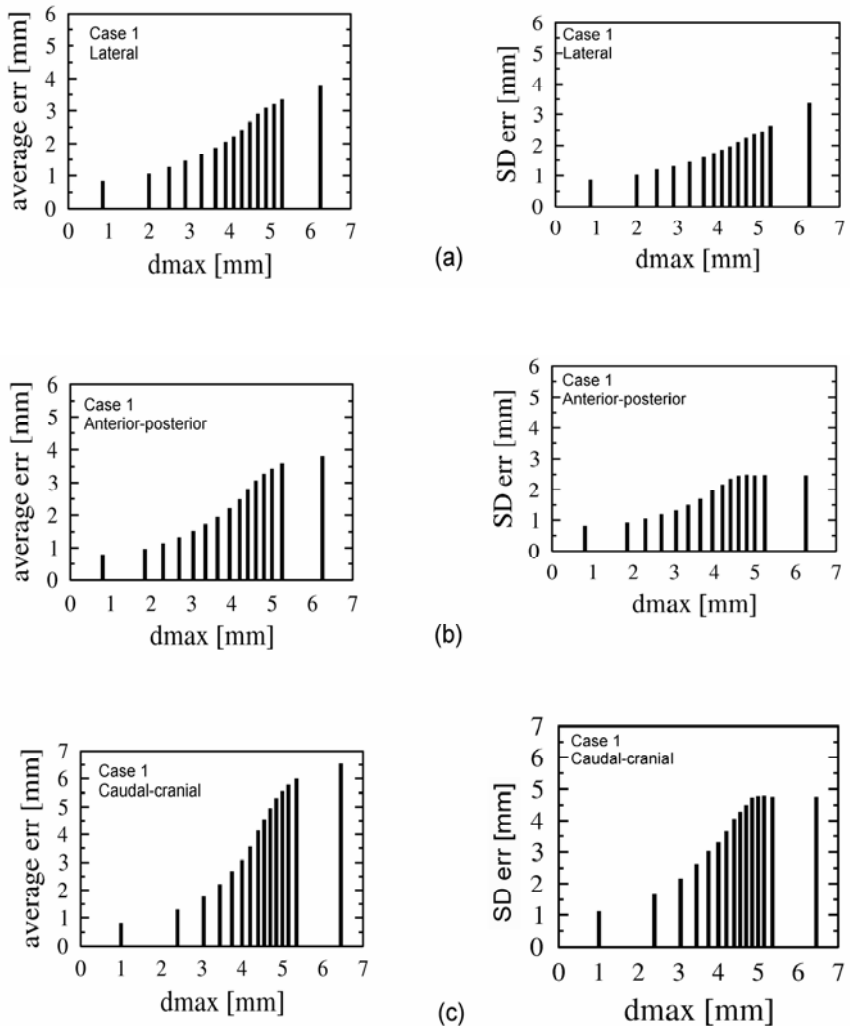



*Fig. 15 :*

*left side: Average registration error (ground truth) as a function of $d_{max}$ for case 1.*

*right side: Standard deviation of the registration error as a function of $d_{max}$ for case 1.*

*Note: To obtain equal number of entries per bin, the intervals were selected non-equidistantly. The bars are centered in the respective interval.*

Tab. 8 summarizes the average registration errors for all test cases investigated in this study. For a compact presentation, $d_{max}$ was binned in 3 intervals only. This data shows that separating $d_{max}$ into three intervals allows separation of voxel entities with different average registration errors. The larger $d_{max}$ (i.e. the bin number), the larger are the average true registration errors as well as the corresponding standard deviations.

*Tab. 8 : Average local registration error for 3 $d_{max}$-bins and al cases. Note: Each bin contains the same number of entries.*

| $d_{max}$-bin | Lateral | Anterior-posterior | caudal-cranial |
|---|---|---|---|
| | Mean±SD [mm] | Mean±SD [mm] | Mean±SD [mm] |
| Case 1 | | | |
| 1 | 1.2±1.2 | 1.1±1.1 | 1.7±2.3 |
| 2 | 2.2±1.9 | 2.2±2.0 | 4.1±4.1 |
| 3 | 3.2±2.6 | 3.4±2.5 | 5.8±4.8 |
| Case 2 | | | |
| 1 | 1.0±1.0 | 1.1±1.2 | 1.8±2.1 |
| 2 | 2.0±1.7 | 2.1±1.7 | 3.5±3.6 |
| 3 | 3.2±2.5 | 3.0±2.1 | 4.7±3.8 |
| Case 3 | | | |
| 1 | 1.4±1.5 | 1.3±1.4 | 1.6±2.1 |
| 2 | 2.9±2.6 | 2.7±2.4 | 3.8±3.7 |
| 3 | 3.8±2.9 | 3.9±2.9 | 5.6±4.4 |
| Case 4 | | | |
| 1 | 1.6±1.7 | 1.3±1.5 | 1.8±2.1 |
| 2 | 2.9±2.5 | 2.6±2.2 | 3.9±3.7 |
| 3 | 3.9±2.7 | 3.5±2.6 | 5.7±4.6 |
| Case 5 | | | |
| 1 | 1.7±1.6 | 1.5±1.6 | 2.0±2.4 |
| 2 | 2.9±2.2 | 2.9±2.4 | 4.0±3.5 |
| 3 | 3.4±2.4 | 3.7±2.6 | 5.4±4.2 |

Fig. 16 shows how a color overlay of $d_{max}$ with the CT image can guide the user of a b-spline algorithm. In the regions displayed in green a user can probably trust in the result of the image registration while in those regions that appear in red, large image registration errors have to be expected.

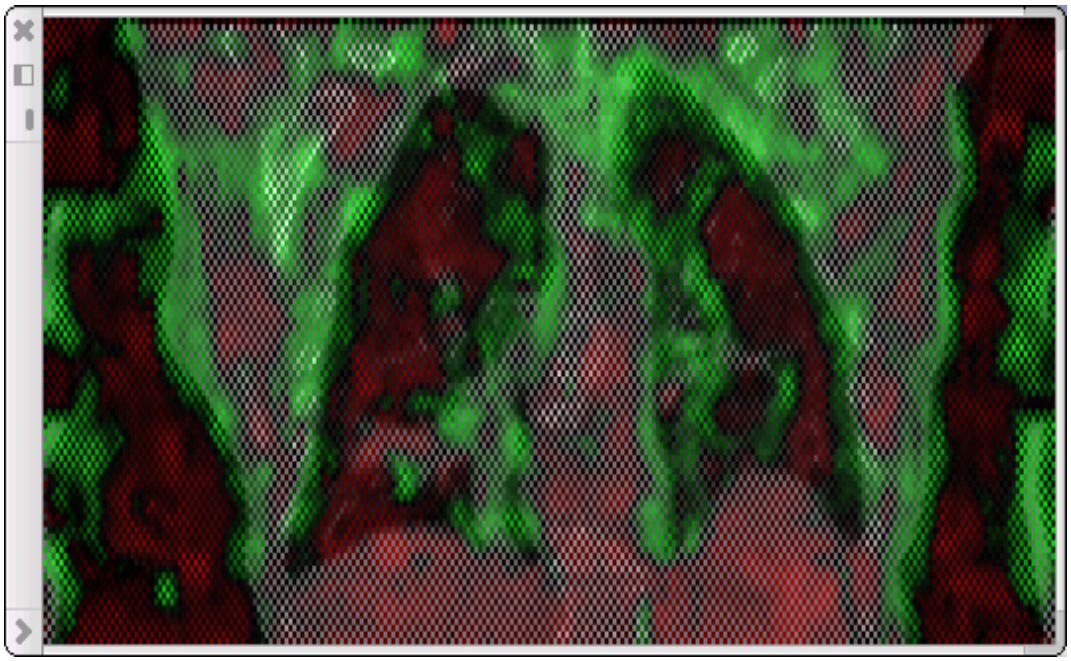



*Fig. 16 : A color overlay between the lateral component of $d_{max}$ and the CT image.*

**The Statistical dependence between $d_{\max,\,i}$ and $err_i(x)$**

Tab. 9 shows the MI-values for the $d_{max}$- and err-values obtained from the analysis of the test cases ($MI_{alg.}$) as well as the largest MI-values obtained for the 200 redistributions ($MI_{indep.}$) described above. This shows that the probability to obtain $MI_{alg.}$ by chance is less than $\frac{1}{200} = 0.005$ and hence the $d_{max}$ can be regarded as statistically dependent on the true local registration error.

*Tab. 9 : MI for the $d_{max}$- and err-values obtained for the test cases ($MI_{alg}$/$MI_{indep}$: prior to/after redistribution, see text)*

| | Lateral | Anterior-posterior | Caudal-cranial |
|---|---|---|---|
| Case 1 | | | |
| $MI_{alg.}$ | 0.20 | 0.19 | 0.18 |
| $MI_{indep.}$ | 0.0003 | 0.0003 | 0.0004 |
| Case 2 | | | |
| $MI_{alg.}$ | 0.21 | 0.23 | 0.25 |
| $MI_{indep.}$ | 0.0002 | 0.0002 | 0.0002 |
| Case 3 | | | |
| $MI_{alg.}$ | 0.17 | 0.18 | 0.22 |
| $MI_{indep.}$ | 0.0003 | 0.0002 | 0.0004 |
| Case 4 | | | |
| $MI_{alg.}$ | 0.17 | 0.22 | 0.24 |
| $MI_{indep.}$ | 0.0003 | 0.0002 | 0.0003 |
| Case 5 | | | |
| $MI_{alg.}$ | 0.14 | 0.18 | 0.20 |
| $MI_{indep.}$ | 0.0003 | 0.0002 | 0.0003 |

## 3.2.2.2. The demons algorithm

### Implementation of the algorithm

For this study the ITK implementation of the demons algorithm was integrated in an in-house developed registration framework. The registration took less than two minutes for lung data sets with about 256x256x80 voxels that were processed on image resolutions four to two.

### The registration results

After registration of the simulated inhale image with the initial exhale image, the resulting DVF was compared with the ground truth and the local registration error was

calculated in each point. Tab. 10 shows the average registration error over all voxels of the complete body region for each of the five test cases.

Tab. 10 : Average registration error and standard deviation for each of the test cases

| **Test case** | **Mean±SD [mm]** | | |
|---|---|---|---|
| | Lateral | anterior-posterior | caudal-cranial |
| 1 | 2.3±2.3 | 2.0±2.0 | 3.1±4.2 |
| 2 | 1.8±1.8 | 1.8±1.9 | 2.7±3.5 |
| 3 | 2.2±2.5 | 1.9± 2.0 | 3.3±4.9 |
| 4 | 2.9±2.7 | 2.5±2.6 | 4.2±5.5 |
| 5 | 2.5±2.4 | 2.7±2.8 | 3.1±4.5 |

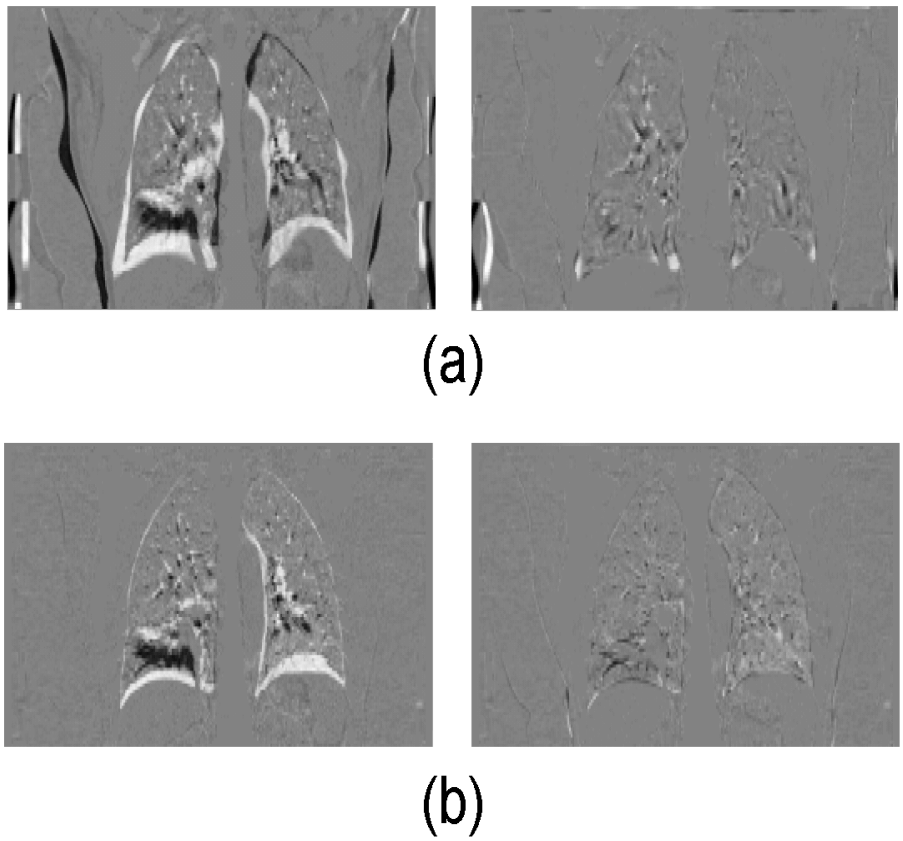



*Fig. 17: (a) difference image between simulated inhale and real exhale before and after the registration for patient 1; (b) displays the same for the real inhale and the real exhale image*

Fig. 17 (a) displays the difference image between the exhale and the simulated inhale dataset before and after the registration for patient 1.
Fig. 17 (b) displays the same for a registration of the exhale with the real inhale from the 4DCT. Probably the registration of the simulated exhale with the inhale image and the registration of the exhale with the real inhale image from the 4DCT are similar challenging for the registration algorithm.

In Fig. 18 (a) the magnitude of the local registration error in caudal-cranial direction is displayed as brightness. Black stands for a small local error and white a large local error.
In Fig. 18 (b) the standard deviation of the caudal-cranial component of the displacement vector is displayed. Fig. 18 (c) displays the difference image between simulated inhale and exhale after registration in the same slice.
Regarding Fig. 18 it is obvious that the standard deviation of the displacement vector component (in Fig. 18 the caudal-cranial component) is not identical with the local registration error. In Fig. 19 some regions where the estimation is rather well are highlighted and distinguished from areas where the estimation performs poorly. However, in regions with large standard deviation large errors are more likely than in those regions where the standard deviation is small. See next sub section for details.

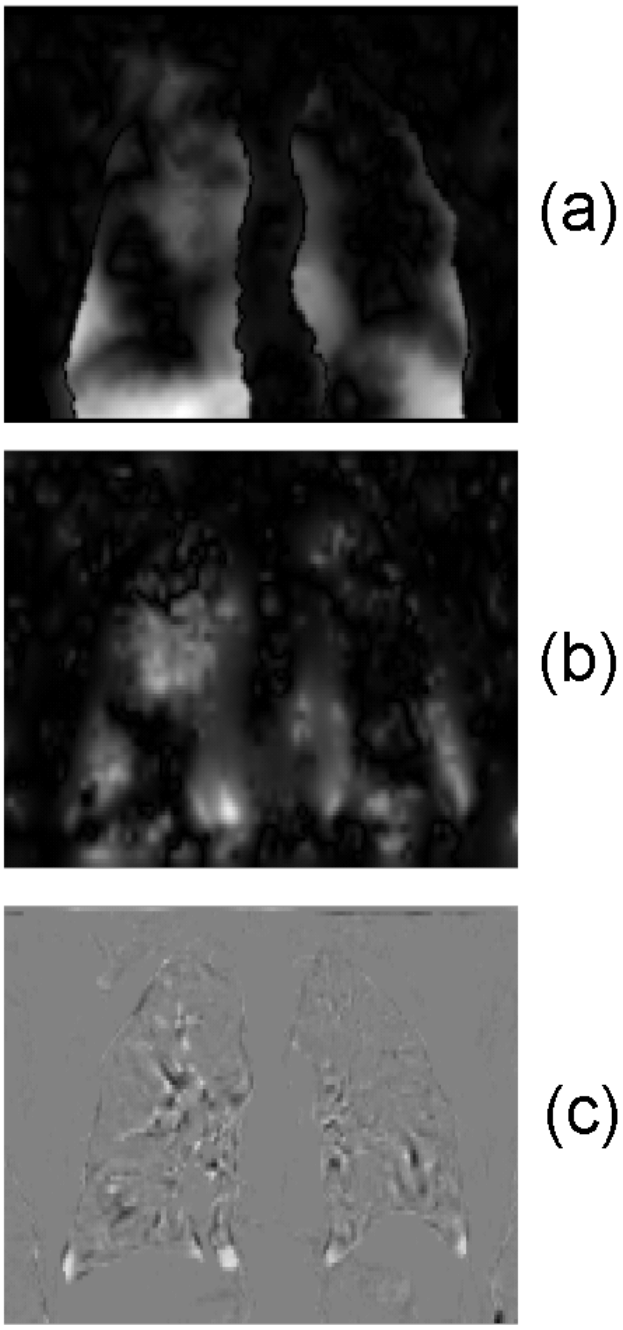



*Fig. 18 : (a) magnitude of the local registration error in caudal-cranial direction, displayed as brightness. Black stands for a small local error and white a large local error. (b) standard deviation of the caudal-cranial component of the displacement vector. (c) difference image between simulated inhale and exhale after registration in the same slice.*

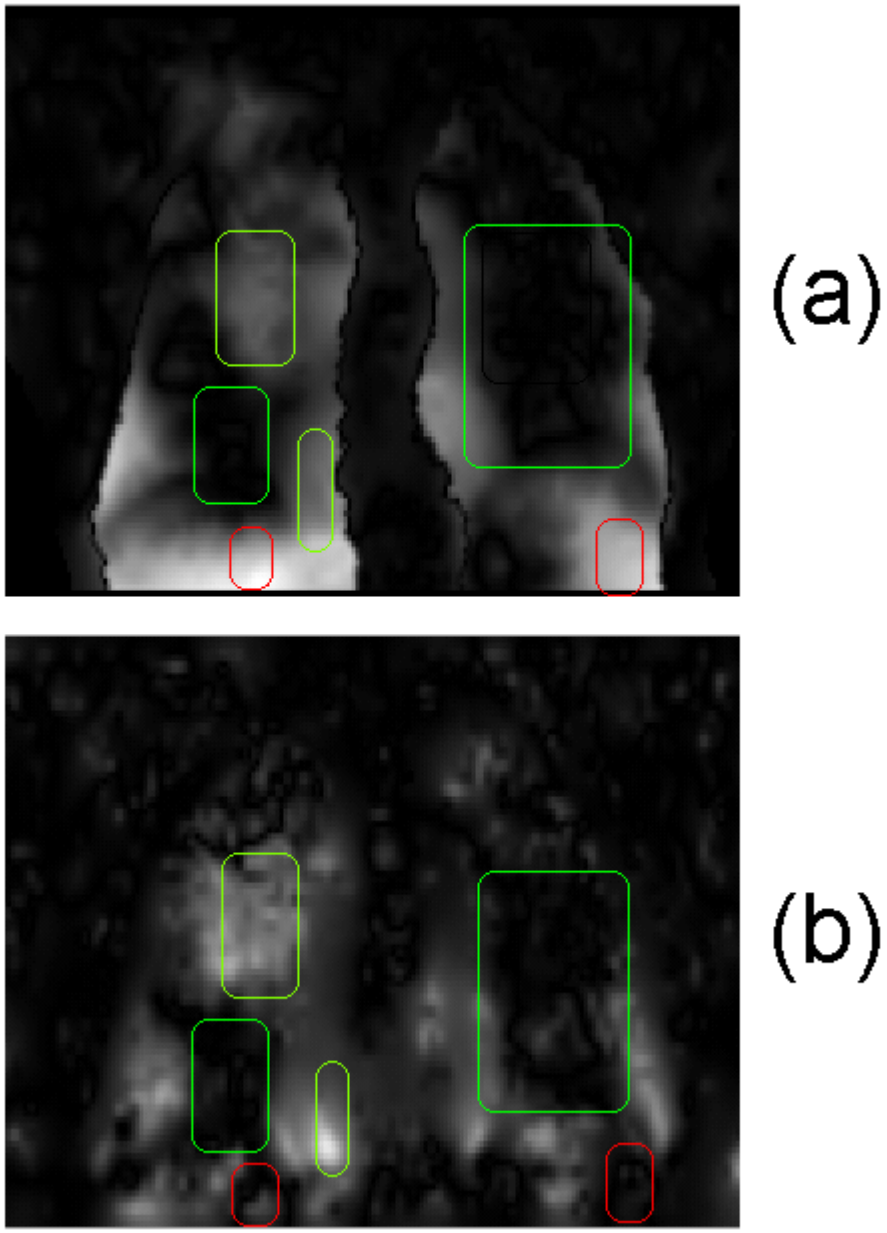



*Fig. 19 : (a) magnitude of the local registration error in caudal-cranial direction, displayed as brightness. Black stands for a small local error and white a large local error. (b) standard deviation of the caudal-cranial component of the displacement vector. In those regions marked in green the estimation of the uncertainty performs rather well. Those regions marked in red show that there may be disagreement between the estimation and the ground truth in other areas.*

**The registration error in sub-regions**

As a typical example, Fig. 20 displays the results obtained for test case 1. The standard deviation of the displacement vector components under multiple registrations, as described in section 2.2.3.1, is binned in 7 intervals. Remember that this standard deviation is regarded as a measure of uncertainty. The bounds of the intervals are chosen non-equidistant, in a way such that the same number of voxels contribute to each bin. In (a), on the left side, the average local registration error in lateral direction is displayed for each bin. Note that the average registration error increases with increasing standard deviation and therefore with increase of the proposed measure of uncertainty. This increase demonstrates that large registration errors are likely to appear in regions where the registration result is not reproducible under multiple registrations.

On the right side the standard deviation of the local registration error within each of the bins is displayed. This standard deviation is rather larger than the average registration error within each bin. This indicates that, regarding a specific voxel the estimation is not reliable while at the same time the estimation of the local registration error is valuable regarding a larger entity of voxels with similar estimation of the registration uncertainty. Fig. 20 (b) displays the same for the anterior-posterior direction. Fig.20 (c) represents the caudal-cranial component.

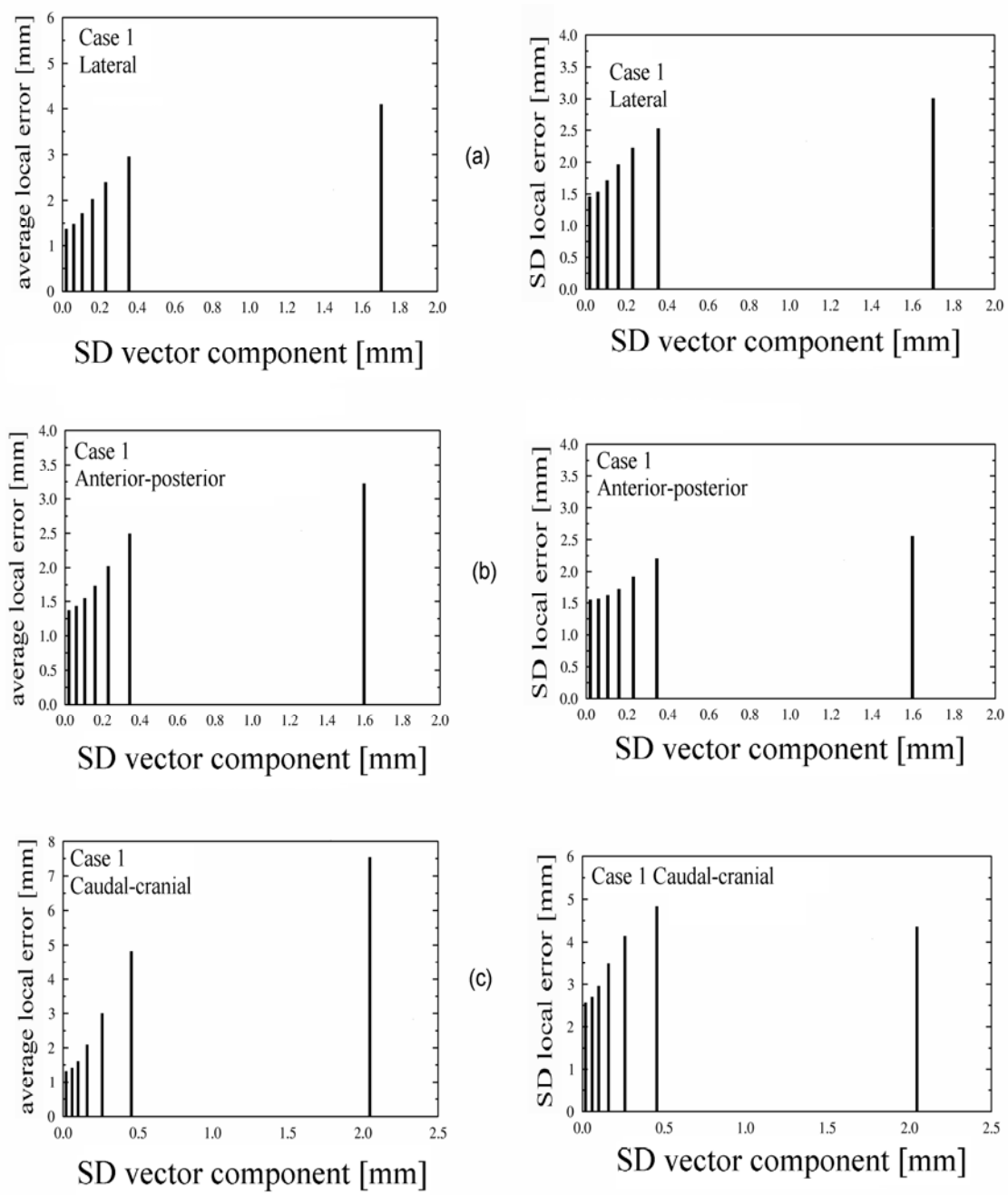



*Fig. 20 :*

*left side: After binning of the standard deviation of the displacement vector components in 7 intervals the average local registration error is displayed for each bin.*

*right side : Standard deviation of the registration error within each bin.*

*Note: To obtain equal number of entries per bin, the intervals were selected non-equidistantly. The bars are centered in the respective interval.*

Tab. 11 summarizes the average registration errors for all test cases investigated in this study. For a compact representation the standard deviation was binned in 3 intervals only. This data shows that separating the standard deviation into three intervals allows separation of voxel entities with different average registration error. The larger the standard deviation (i.e. the bin number), the larger is the average local registration error as well as the corresponding standard deviation.

*Tab. 11 : Average local registration error for 3 bins of the standard deviation of the vector component and all cases. Note: Each bin contains the same number of entries.*

| | Mean±SD [mm] | Mean±SD [mm] | Mean±SD [mm] |
|---|---|---|---|
| Bin | lateral | anterior-posterior | caudal-cranial |
| Case 1 | | | |
| 1 | 1.4±1.5 | 1.4±1.6 | 1.4±2.6 |
| 2 | 2.0±2.0 | 1.7±1.7 | 2.1±3.5 |
| 3 | 3.4±2.8 | 2.8±2.4 | 5.8±4.8 |
| Case 2 | | | |
| 1 | 1.1±1.2 | 1.1±1.3 | 1.1±1.7 |
| 2 | 1.5±1.5 | 1.5±1.5 | 1.8±2.4 |
| 3 | 2.7±2.3 | 2.7±2.4 | 5.3±4.3 |
| Case 3 | | | |
| 1 | 1.4±1.7 | 1.3±1.5 | 1.1±2.3 |
| 2 | 2.0±2.1 | 1.7±1.7 | 1.8±3.3 |
| 3 | 3.4±3.0 | 2.8±2.4 | 7.0±5.9 |
| Case 4 | | | |
| 1 | 2.1±2.2 | 1.8±2.0 | 1.9±3.7 |
| 2 | 2.8±2.7 | 2.3±2.2 | 3.1±4.7 |
| 3 | 3.9±3.0 | 3.6±3.0 | 7.6±6.2 |
| Case 5 | | | |
| 1 | 1.9±2.0 | 2.0±2.2 | 1.5±2.5 |
| 2 | 2.3±2.1 | 2.5±2.6 | 2.0±3.2 |
| 3 | 3.3±2.8 | 3.8±3.2 | 5.9±6.0 |

# 4. Discussion

## 4.1. Detection of MRI distortion

A physical phantom in combination with an evaluation software was developed to determine and correct distortions in MR images used for treatment planning in radiation oncology. Compared to previously described phantoms which are mostly developed for intracranial applications, it detects distortions over a much larger field of view. In addition, the representation of the control points differs from that of other phantoms. The phantom and the evaluation software were successfully tested for two frequently applied MR sequences. After correction of the distortions the magnitude of the residual average distortion measured in the control points was significantly reduced for all slices of the phantom.

The magnitude of the distortion and the principal shape of the distortion map is similar for the flash 3D and the spin echo sequence. Since the distortion depends on the homogeneity of the ground field and the linearity of the gradients, this was expected.

In module one as well as module five of the phantom, which have the largest distance from the center of the MRI-device, there are still large distortions in a few control points, even after the geometric correction. Fig. 21 shows that these control points, are located in the lower left and right corner of the module. The failure of the correction in these points is due to problems with the detection of the control points. Since there is a large change of the distortion throughout these control points, the bright area of the hole actually appears in elliptical rather than circular shape. Since the software expects circular shaped objects, the detection is not accurate for these control points.

Tab. 12, demonstrates, however, that this problem concerns just the two control points in the corners of these slices. Even in slices of module one and five there is no other point with a similar large distortion after the geometric correction. In all other control points the distortion is mainly eliminated after correction. The clear reduction of the average distortion, detected after geometric correction of the images, demonstrates this.

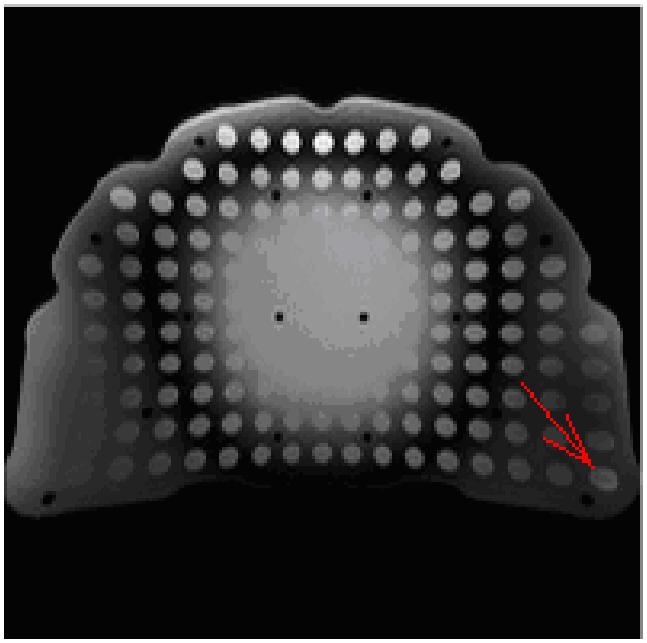


*Fig. 21 : The central slice of module number five. The red arrow points to one of the control points with the largest distance from the center of the MRI-device.*

*Tab. 12 : Residual magnitude of the distortion after correction in those control points with largest failure of the distortion correction.*

| magnitude of the 3D distortion | Position of the **first** module z = -180 mm | Position of the **last** module z = +180 mm |
|---|---|---|
| Largest value [mm] | 15.8 | 13.0 |
| Second largest value [mm] | 1.8 | 11.3 |
| third largest value [mm] | 1.6 | 1.8 |

Since the slice distance affects the distortion, images acquired for quality assurance must have the same slice distances as the images used in the clinical situation. To achieve an acceptable image resolution in the axial dimension, a shifted test data acquisition technique was introduced. For this, the acquisition of several datasets and the construction of a virtual data set with reduced slice distances is necessary, although it is time consuming.

Finally, it should be noted, that a physical phantom can not capture distortions which are due to susceptibility changes induced by the patient. Phantom measurements can solely account for distortion due to ground field inhomogeneity as well as non-linearity of the gradient system. However, since these distortions are present in any image of

any patient, it is necessary to measure and remove this part of the image distortion. In case the manufacturer of the MRI scanner provides post-processing software for this purpose, the geometric accuracy of the resulting images should still be controlled with phantom measurements.

Due to legal requirements with the medical device directive, the application of the distortion correction to clinical MR images appears to be difficult. Nevertheless the developed system can be used to check the geometric accuracy of MR images within routine quality assurance procedures.

## 4.2. Uncertainty of the elastic image registration

Various methods to access the accuracy of image registration algorithms have been proposed such as tracking of landmarks [14][15][16][17], check of the alignment of contours [17][18], evaluation of the overlay of corresponding edges with a color wash technique or a split screen [1] as well an the investigation of the performance on artificially created test cases, where the ground truth on the deformation is known [3][19][4]. Each of these methods has its drawbacks. Visible landmarks consist of voxels that may drive the registration so that the measured accuracy differs from the accuracy achieved in homogeneous regions of the same dataset. The overlay of contours can only be evaluated in case the contours are available in each dataset. This is not the case when the algorithm is used to transfer contours from one dataset to another. Registration of artificially deformed images provides knowledge on the under-laying ground truth of the deformation. In a clinical setting, however, this information is missing and it is the task of the image registration to estimate it.

The color wash or split screen visualization is helpful to display the alignment of corresponding edges, but severe registration errors may be present in homogeneous regions of the image. These errors are not visible in a color wash or split screen image.

So there is need to further investigate registration errors and sources of error in medical image registration. Especially the lack of image structure needs to be taken into account, since this problem is not visible in a color overly and may affect the accuracy of the image registration and thus the accuracy of dose mapping, target delineation as well as anatomy mapping. In this work two different methods for two

different registration techniques are proposed to deal with this problem.

A recently proposed method is based on evaluating the physical fidelity of the deformation field [20][21]. This method was quantitatively tested on a demons algorithm. The authors show that the unbalanced energy is correlated with the geometric registration error. However, the demons algorithm for which this method has been tested is likely to create deformation fields which do not fulfill requirements of physical fidelity, since demons algorithm can perform relatively large changes of the deformation field in regions with small intensity gradients. Due to its large number of degrees of freedom this algorithm is likely to be affected by noise and to create deformation fields which do not fulfill requirements of physical fidelity. Parameterized methods such as b-spline registration do not face the same problems since they are operating on a smaller number of degrees of freedom. For b-spline registration non-fidelity is not likely to be a problem since non-linear as well as non-invertible deformations can be penalized [4]. Deformation of rigid structures has been penalized as well [28] and some authors consider b-spline deformation to be free of folding in case of a multi resolution approach [29]. Nevertheless a deformation field may fulfill requirements of physical fidelity and may still differ significantly from the under-laying ground truth of the deformation due to missing structure or unaligned edges in the images. The method discussed in the next paragraph takes this into account.

## 4.2.1. B-spline registration

### 4.2.1.1. Interpretation of the results

As it is probably impossible to exactly determine the actual registration error for each individual voxel, we focussed on a statistical evaluation. It is important to note that the $d_{max}$ value is not the image registration error itself. In any specific voxel, the registration error is still unknown after running the proposed algorithm as $d_{max}$ is an estimate of the average registration error over a larger entity of voxels. However, large errors are unlikely to appear in regions with small $d_{max}$ values.

The algorithm was tested on artificially deformed images as described in section 2.2.4.1. Here the ground truth on the deformation is known. Such a test dataset is not

identical to a CT dataset reconstructed from a 4DCT. However, the main aspects of breathing motion are taken into account and so we consider the images to be a model for the lung motion, which is suitable to evaluate the proposed method.

It is important to note that $d_{max}$ depends on the choice of the boundaries $b_l$ and $b_u$ of the random variable $r_n$ (see section 2.2.3.1). Therefore an average $d_{max}$ over the complete dataset may be used as a quality estimate to compare different lung registrations only, if the same boundaries were used. The average $d_{max}$ in some sub-region is not identical with the expected average local registration error. The proposed method rather provides the information in which sub-region the mean errors are small or large compared to the range of the expected errors. The magnitude of the absolute value of the average local registration error in each sub-region may be known from clinical experience, as we do have a basic idea about the typical range of b-spline registration errors in the lung from studies such as [16][18][26][27].

Before clinical application of the proposed method it will be necessary to investigate how many test deformations are actually necessary in order to capture the sensitivity of the metric. A smaller number of test deformation could greatly reduce the calculation time. In case 50 test deformations are sufficient, the algorithm could run within less than two minutes on a PC with a 1.85 GHz processor rather than about 15 minutes as for the calculations shown in this study. There is great potential to speed up the calculation by parallel computing.

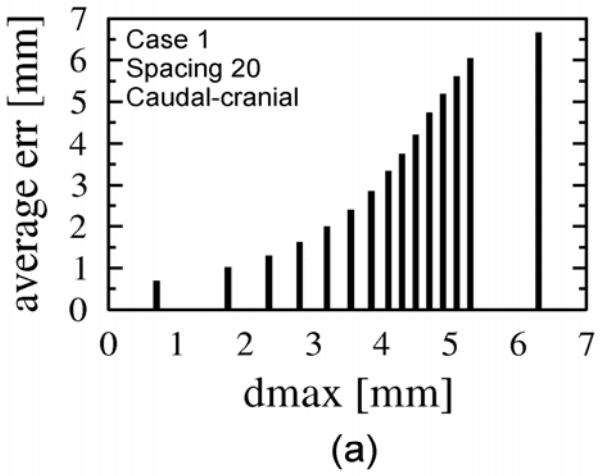
Note that the method was tested solely in presence of deformation and not in presence of growing or shrinking tissue.

#### 4.2.1.2. Robustness

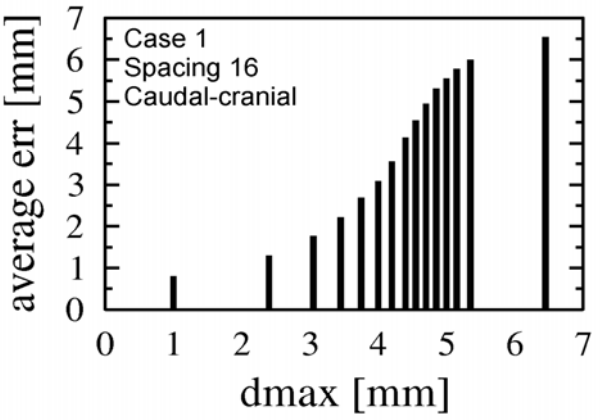
The robustness of the algorithm to estimate the uncertainty of the b-spline registration as well as its potential to classify the quality of the registration result was investigated using three different deformations for case 1. These deformations differ in the magnitude and spatial variability of the DVF. This was achieved by using different spacing of the Gaussian functions that contribute to the random component of the DVF. For this, spacings of 20, 16 and 12 were applied. The smaller the spacing, the stronger is the spatial variability and the larger is the magnitude of the random component of the simulated deformation and hence the more challenging is the

registration problem for the b-spline algorithm. So spacing 20 represents a moderate and spacing 12 a rather challenging test dataset. With these different datasets we show that the proposed method is robust regarding the complexity of the deformations. It can be applied for challenging deformations as well as for rather simple deformations.

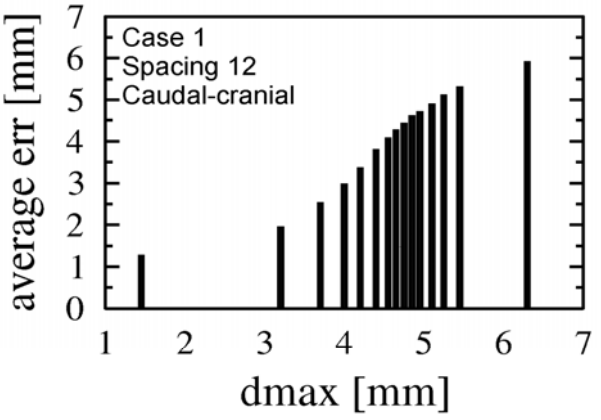
Fig. 22 shows the average local registration error for 15 $d_{max,z}$ bins for spacing 20 (a), 16 (b) and 12 (c). This demonstrates that dividing the image in sub-regions of different average registration error works for varying quality of the registration result.




(a)




(b)




(c)

*Fig. 22 :Aaverage local registration error in cranio-caudal direction for different spacings. Note: To obtain equal number of entries per bin, the intervals were selected non-equidistantly. The bars are centered in the respective interval.*

Fig. 23 shows the difference image before and after each registration for case 1 and for spacings of 20 (a) and 12 (b), respectively, of the Gaussian functions. The figure demonstrates that the registration for spacing 12 results in a less homogeneous difference image after registration than for spacing 20. This indicates that the result of this registration is over all poorer than the result displayed under (a). Tab. 13 displays the average registration errors as obtained from the ground truth and it confirms this impression. The difference images of the real inhale and exhale images obtained from the 4D-CT before and after the registration are shown in (Fig . 23 (c) ). The homogeneity of this difference image is not clearly different from Fig. 23 (a) and Fig.23 (b). Most probably, the quality of the result of the registration between the real inhale and exhale images is comparable to the quality of the result of the test cases.

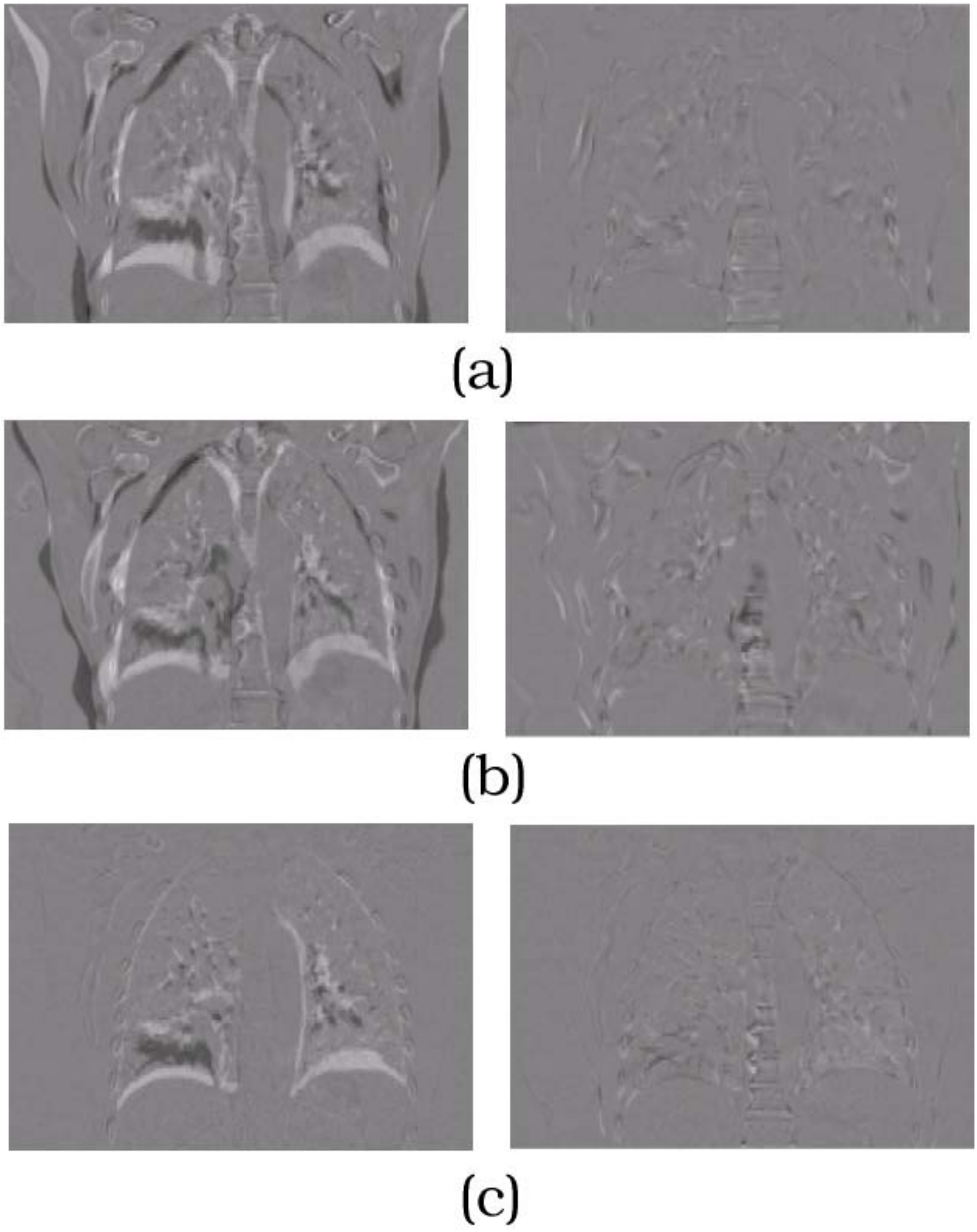



*Fig. 23 : Difference images before (left) and after (right) the registration for case 1 using simulated data ((a) spacing 20, (b) spacing 12) and real data from the 4D-CT (c).*

Tab. 13 shows the average $d_{max}$ and the average local registration error for case 1 for three different spacings. Both, the average $d_{max}$ as well as the average registration error increase with decreasing spacings.

*Tab. 13 : Average registration error and average $d_{max}$ for different spacings*

| | *Lateral* | | *Anterior -posterior* | | *Caudal-Cranial* | |
|---|---|---|---|---|---|---|
| Spacing | $d_{max}$ [mm] | err±SD [mm] | $d_{max}$ [mm] | err±SD [mm] | $d_{max}$ [mm] | err±SD [mm] |
| 20 | 3.5 | 1.9 ±1.9 | 3.3 | 1.8 ±1.8 | 3.8 | 3.3 ±3.7 |
| 16 | 3.8 | 2.2 ±2.1 | 3.7 | 2.2 ±2.1 | 4.1 | 3.8 ±4.1 |
| 12 | 4.2 | 3.1 ±2.7 | 4.2 | 3.1 ±2.6 | 4.4 | 3.9 ±3.8 |

## 4.2.2. Demons algorithm

### 4.2.2.1 Interpretation of the results

Since the approach proposed for b-spline registration can solely be applied for a parameterized method, a different approach is needed for the non-parametric demons algorithm. It is important to note that the calculated standard deviation is not the local image registration error in any specific voxel. Rather, the standard deviation identifies image regions where the algorithm is likely to perform poorly. In any specific voxel the true registration error is still unknown after application of the proposed method. We propose this stochastic approach to estimate the registration uncertainty as it is probably in general impossible to fully determine the actual local registration error.

### 4.2.2.2 Calculation time

The calculation time for nine registrations may be considered as a drawback of the proposed method. However, this method is most suitable for parallel computing. The demons algorithm runs very fast and although the calculation time was a limiting

factor for image registration in the past it may not be a limiting factor in the future as computers are getting faster every year.
As only one of the nine different deformation fields and the standard deviation field need to be stored for further processing, the proposed approach does not extremely increase the effort in data administration.

## 4.2.3. Outlook

Both methods have the potential to distinguish areas where the registration is likely to be accurate from areas where the registration is likely to be less accurate.
In our next study we are planning to investigate an extension of this approach to classify dose mapping errors that are due to registration uncertainties.

### 4.2.3.1. B-splines

In case that a given dose distribution is to be mapped from the test image to the reference image, we propose to evaluate the maximum dose deviation instead of the maximum spatial deviation. In that way the same method may be applied to estimate the error of the dose accumulation instead of the local registration error.
So for each voxel and dimension in space, the largest dose deviation among the K test deformations, for which the local SSD is smaller than or equal to the initial local SSD, can be stored as a measure of the dose-accumulation uncertainty:

$$dose_{d\max}(x) = \max_{k=1}^{K} \left\{ \left| dose_k(x) - dose\ (x) \right| \right\}$$

Where $dose\ (x)$ is the dose mapped to the voxel x based on the result of the b-spline registration and $dose_k(x)$ is the dose mapped to the reference image voxel x based on a deformation with one of the K test DVFs. In our future work, we plan to investigate the statistical dependence between $dose_{d\max}(x)$ and the true dose accumulation error due to registration errors.

#### 4.2.3.2. Demons algorithm

Instead of the standard deviation of the displacement vector component the idea is to warp the dose matrix based on each of the nine resulting deformation fields. The standard deviation of the mapped dose can be calculated in each voxel. It may be interesting to investigate the statistical dependence between the resulting dose standard deviation, obtained from the warped dose distribution and the true dose mapping error.

# Literature

[1] M. L. Kessler “Image registration and data fusion in radiation oncology,” Br J Radiol. 79, 99–108 (2006).

[2] J. P. Thirion, "Image Matching as a Diffusion Process: An Analogy with Maxwell's Demons," Med. Image. Anal. **2**, 243-260 (1998)

[3] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an Accelerated 'Demons' Algorithm for Deformable Image Registration in Radiation Therapy," Phys.Med.Biol. 50, 2887-2905 (2005).

[4] J. Kybic, M. Unser, “Fast parametric elastic image registration,“ IEEE Trans. Image Process. 12, 1427- 1442 (2003)

[5] E. Krestel
Imaging Systems for Medical Diagnostics
1990 by Siemens Aktiengesellschaft, Berlin and Munich

[6] P. Bendel, **“**Echo Projection Imaging-A Method to Obtain NMR Images Undistorted by Magnetic Field Inhomogeneities**”** IEEE Trans Med Imaging. 1985;4(2):114-9.

[7] T. S. Wong, D. Rosenfeld “Spin-Inversion Imaging: A Technique for NMR Imaging under Magnetic Fields with High Field Nonuniformities”
IEEE Trans Med Imaging. 1987;6(2):148-56.

[8] L.R.Schad, H.H. Ehricke, B. Wowra, G. Layer, R. Engenhart, H.U. Kauczor, H.J. Zabel, G. Brix, W.J. Lorenz “Correction of spatial distortion in magnetic resonance

angiography for radiosurgical treatment planning of cerebral arteriovenous malformations.” Magn Reson Imaging.1992;10(4):609-21

[9] S.J. Doran, L. Charles-Edwards, S.A. Reinsberg, M.O. Leach “A complete distortion correction for MR images: I. Gradient warp correction” Phys Med Biol. 2005 Apr 7;50(7):1343-61.

[10] D.Wang, D.M.Doddrell, G.Cowin, “A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging” Magn Reson Imaging. 2004 May;22(4):529-42.

[11] L.N.Baldwin, K.Wachowicz, S.D. Thomas, R. Rivest, B.G. Fallone: “Characterization, prediction, and correction of geometric distortion in 3 T MR images“ Med Phys. 2007 Feb;34(2):388-99.

[12] H. Chang, J.M. Fitzpatrick, “A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities,” IEEE Trans Med Imaging. 1992;11(3):319-29.

[13] J.M. Balter, M.L. Kessler, “Imaging and alignment for image-guided radiation therapy,” J Clin Oncol. 2007 Mar 10;25(8):931-7

[14] K. K. Brock, M.B. Sharp, L.A. Dawson, S.M. Kim, D.A. Jaffray, “Accuracy of finite element model-based multi-organ deformable image registration,” Med. Phys. 32, 1647-1659 (2005).

[15] M. M. Coselmon, J.M. Balter, D.L. McShan, M.L. Kessler, “Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines,” Med. Phys. 31, 2942-2948 (2004).

[16] E. Rietzel, G. T. Y. Chen, “Deformable registration of 4D computed tomography data,” Med. Phys. 33, 4423-4430 (2006).

[17] E. Heath, D. L. Collins, P. J. Keall, L. Dong, J. Seuntjens, “Quantification of accuracy of the automated nonlinear image matching and anatomical labeling (ANIMAL) nonlinear registration algorithm for 4D CT images of lung,” Med. Phys. 2007 Nov; 34(11):4409-21

[18] K. Wijesooriya, E. Weiss, V. Dill, S. Joshi, P. J. Keall, “Quantifying the accuracy of automated structure segmentation in 4D CT images using a deformable image registration algorithm,” Med. Phys. 2008 Apr;35(4):1251-60

[19] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, D. L. G. Hill, “Validation of nonrigid image registration using finite-element methods: Application to breast MR images.” IEEE Trans. Med. Imag. 22, 238-247 (2003).

[20] H. Zhong, T. Peters, and J. V. Siebers, "FEM-Based Evaluation of Deformable Image Registration for Radiation Therapy," Phys.Med.Biol. 52, 4721-4738 (2007).

[21] H. Zhong, E. Weiss, and J. V. Siebers "Assessment of dose reconstruction errors in image-guided radiation therapy" Phys.Med.Biol. 53, 719-736 (2008).

[22] I. J. Schoenberg, ”Cardinal interpolation and spline functions,” J. Approximation Theory, Vol. 2, pp. 167-206, 1969.

[23] M. Unser, A.Aldroubi, and M. Eden, “B-Spline signal processing: Part I - Theory,” IEEE Trans. Signal Process. 41(2), pp. 821-833, (1993)

[24] M. Unser, A.Aldroubi, and M. Eden, “B-Spline signal processing: Part II – Efficient Design and Applications,” IEEE Trans. Signal Process. 41(2), pp. 834-848, (1993)

[25] B. K. P. Horn and B. G. Schunck, "Determining optical flow." *Artificial Intelligence*, vol 17, pp 185-203, 1981

[26] A. Pevsner, B. Davis, S. Joshi, A. Hertanto, J. Mechalakos, E. Yorke, K. Rosenzweig, S. Nehmeh, Y. E. Erdi, J. L. Humm, S. Larson, C. C. Ling, G. S. Mageras, “Evaluation of an automated deformable image matching method for

quantifying lung motion in respiratory-correlated CT images," Med. Phys. 2006 Feb, 33(2) pp.369-76.

[27] R Kashani, J Balter, M Kessler, M Hub, L Dong, L Zhang, L Xing, Y Xie, D Hawkes, J Schnabel, J McClelland, S Joshi, "Objective Assessment of Deformable Image Registration in Radiotherapy - a Multi-Institution Study," 49th AAPM Annual Meeting, Minneapolis, Minneaota, 22-26 July 2007.

[28] D. Ruan, J. A. Fessler, M. Roberson, J. M. Balter and M. Kessler, "Nonrigid Registration with Regularization to Account for Local Tissue Rigidity," SPIE, Feb. 2006.

[29] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, D. J. Hawkes, "A Generic Framework for Non-rigid Registration Based on Non-uniform Multi-level Free-Form Deformations," Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention, p.573-581, October 14-17, 2001

# Acknowledgments

First of all I want to thank Prof. Christian P. Karger for inspiring discussions and all the advice, help and encouragement he gave me during the past years!

It was wonderful to share an office with Christine, Phillip, Ina, Barber, Paola, Sonja, Thorsten, Andres and Sarah. I want to thank all my colleagues for their friendliness.

I would like to thank Prof. Marc Kessler, Rojano Kashani and Prof. James M. Balter for inspiring discussions, encouragement and for their time and patience.

I am most grateful to our God, who is taking care of us, and to my family for being there.