A Stochastic Approach to Estimate the Uncertainty Involved in B-Spline Image Registration

M. Hub*, M. L. Kessler, and C. P. Karger

***Abstract*—Uncertainties in image registration may be a significant source of errors in anatomy mapping as well as dose accumulation in radiotherapy. It is, therefore, essential to validate the accuracy of image registration. Here, we propose a method to detect areas where mono modal *B*-spline registration performs well and to distinguish those from areas of the same image, where the registration is likely to be less accurate. It is a stochastic approach to automatically estimate the uncertainty of the resulting displacement vector field. The coefficients resulting from the *B*-spline registration are subject to moderate and randomly performed variations. A quantity is proposed to characterize the local sensitivity of the similarity measure to these variations. We demonstrate the statistical dependence between the local image registration error and this quantity by calculating their mutual information. We show the significance of the statistical dependence with an approach based on random redistributions. The proposed method has the potential to divide an image into subregions which differ in the magnitude of their average registration error.**

***Index Terms*—Elastic image registration, local uncertainty.**

I. Introduction

ELASIC registration of medical images may play an important role in the daily routine of radiation oncology in the future as adaptive radiotherapy requires the incorporation of multiple datasets in the treatment planning process as well as for the patient setup [1]. Therefore, it is essential to know the geometric correspondence between voxels of different images. A number of algorithms have been developed to perform fast elastic image registration [2]–[7].

One major group of these algorithms is driven by intensity differences and intensity gradients. A similarity measure is optimized either directly [6] or in an indirect way [4], [5]. The presence of image structure is a basic requirement for the feasibility of these methods. More generally, information on the geometric alignment must be encoded in the image intensities. Local alignment errors are likely to occur in homogenous regions of the datasets where image structure is missing.

Most validation methods do not account for errors due to homogeneity. Exceptions are [8] where the sensitivity of the metric to rigid transformations is evaluated, as well as [7], where the sensitivity of the metric to rigid translations is evaluated in image subregions.

Various methods to validate the elastic registration of medical images have been proposed such as tracking of landmarks [10]–[13], check of the alignment of contours [13], [14], evaluation of the overlay of corresponding edges with a color wash technique as well as split screen visualization [2] or an investigation of the registration result in test cases, where the ground truth on the deformation is known. Such test cases are either obtained by simulating deformations on clinical data [5], [6], [15], [16], [17], or with the help of a physical phantom as in [18] and [19]. In [9], a bootstrap method is proposed to estimate the uncertainty of rigid image registration without ground truth.

Each of these methods has its drawbacks. Visible landmarks may consist of voxels that drive the registration, and may therefore not be representative for errors in homogeneous regions of the same dataset. The overlay of contours can only be evaluated, if the contours are available in both datasets. This is not the case when the algorithm is used to transfer contours from one dataset to another. Registration of artificially deformed images provides the underlaying ground truth of the deformation. In a clinical setting, however, this information is not available and it is the task of the image registration to estimate it, so here this method can not be applied.

The color wash or split screen visualization is helpful to assess 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.

The approach described in [9] is promising, however, this work does not primarily discuss elastic image registration.

A recently proposed method to validate displacement vector fields (DVFs) is based on the investigation of their physical fidelity. This method was quantitatively tested on a "demons algorithm" [20], [21]. Each registration approach, the "demons algorithm" as well as *B*-spline registration has its specific strengths and weaknesses and therefore a validation approach needs to be tailored to the difficulties of the specific registration algorithm. Regarding the "demons algorithm" its large number of degrees-of-freedom allows to describe complex deformations and there are no problems with a potential model mismatch as it can occur in *B*-spline registration. At the same time, due to the large number of degrees-of-freedom the algorithm is likely

Manuscript received November 08, 2008; revised April 02, 2009. First published May 12, 2009; current version published October 28, 2009. *Asterisk indicates corresponding author.*

*M. Hub is with the Department of Medical Physics in Radiation Oncology, German Cancer Research Center, 69120 Heidelberg, Germany (e-mail: m.hub@dkfz.de).

M. L. Kessler is with the Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: mkessler@med.umich.edu).

C. P. Karger is with the Department of Medical Physics in Radiation Oncology, German Cancer Research Center, 69120 Heidelberg, Germany (e-mail: c.karger@dkfz.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2009.2021063

to run into problems with the physical fidelity of the deformation field. Additionally, due to the nature of its optimization, the changes of the deformation field per iteration are relatively large in regions of small intensity gradients. Therefore, the algorithm is rather sensitive to noise.

Parameterized methods, such as *B*-spline registration, face other problems as they are operating with limited degrees-of-freedom and therefore may be affected by model mismatch. The error estimation should take this into account. At the same time the problem of nonfidelity is less likely to be severe in case of *B*-spline transformations as nonlinear as well as noninvertible deformations can be penalized [6]. Deformation of rigid structures has been penalized [22] and some authors consider *B*-spline deformation to be free of folding in case of a multiresolution approach in the knot spacing [23]. For another parameterized registration method noninvertible deformations were avoided by constraints [24]. Nevertheless, a *B*-spline deformation field, which fulfills requirements of physical fidelity may still significantly differ from the underlying ground truth due to missing image structure or unaligned edges as consequence of model mismatch. The method described here is tailored towards a parameterized registration method and takes its specific problems and weaknesses into account.

The basic idea of this approach is to focus on the local sensitivity of the similarity measure to additional deformations after the registration. If the registration has resulted in the correct deformation, and if there is sufficient structure in the image, such that the optimum of the metric is well-defined in the sense of a sharp optimum, the similarity measure should get worse with any additional deformation. In regions where this is locally not the case, either due to missaligned structures or in homogenous regions, we can not distinguish whether the initial or modified displacement vector is the better estimate. This ambiguity can be exploited to evaluate the involved geometrical uncertainty.

The aim of this paper is to propose an algorithm and to demonstrate its potential to divide an image into subregions that differ in the magnitude of their average registration error. A color display of these regions can help a clinician to identify image regions where dose as well as anatomy mapping is likely to be inaccurate.

II. Theory

A. B-spline Registration

Any polynomial spline can be regarded as a superposition of *B*-spline basis functions as proposed in [25]. *B*-spline signal processing has been extensively discussed in [26] and [27] and its application to multidimensional elastic registration was described in [6].

Here, we regard three dimensional images. The result of the registration is a displacement vector field (DVF) which represents the deformation as a displacement vector in each voxel. In *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 $(1 \leq i \leq 3)$. In our study a multiresolution approach in the knot spacing was applied with a finest knot spacing of 32 mm in plane and 24 mm cross plane.

Let us regard two images with integer intensity, the test image f_t and the reference image f_r, both represented in voxel coordinates

$$f_t, f_r : \mathrm{Z}^3 \rightarrow Z.$$

The aim of the elastic registration is to find a set of coefficients $c(k)$ such that anatomically corresponding voxels of the warped test image (indexed w) and the reference image reach alignment

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

A commonly used metric for monomodality image registration is the sum of the squared differences (SSD) of the image intensities

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

where Ω denotes the voxel space of the images.

We consider the DVF with the lowest SSD to be the best solution for the image registration and minimizing the SSD is therefore the optimization problem.

B. Sources of Errors in B-spline Registration

Two major sources of errors in *B*-spline image registration are as follows.

1) Ambiguity in Homogeneous Regions: The first partial derivatives of the SSD with respect to the *B*-spline coefficients were given by [6]

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

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 neighboring knots, but do not guide the deformation process. The metric is not sensitive to misalignment that occurs within these homogeneous regions, so the minimum of the metric may be broad and therefore not well defined. 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 deviate from the unknown ground truth. This general problem is concerning any optimizer, in case the metric is intensity based.

2) Misaligned Edges: In nonhomogenous regions, errors may occur, if the optimization does not lead to the global minimum of the SSD-metric. This is the case, if the optimization problem is not 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 $d_{\text{true}}(x)$ is not included in the entity of possible *B*-spline DVFs $\{d_h^n(x)\}$ with degree n and knot spacing h, i.e., $\|d_{\text{true}}(x) - d_h^n(x)\| \neq 0$ for the best approximation $d_h^n(x)$.

C. Algorithm to Estimate the Local Registration Error

We propose to estimate the uncertainty of the elastic registration 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 subregion of the reference image. The contribution of SSD_Δ of Δ to the global SSD_Ω is

$$\text{SSD}_\Delta = \sum_{\delta_\in \Delta} (f_r(\delta) - f_w(\delta))^2.$$

The aim is to determine the range of geometric deviations, within certain bounds, which can be performed without increasing the local contribution SSD_Δ to the overall metric SSD_Ω.

Let $\{c_1, \ldots, c_{3N}\}$ 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 $\{r_1, \ldots, r_{3N}\}$ be the set of $3N$ random variables which are equally distributed within the intervals defined by the boundaries b_l and b_u: $b_l \leq r_n \leq b_u$ for $1 \leq n \leq 3N$. In this study $b_l = -10$ mm and $b_u = 10$ mm were used as typical registration errors are expected within this range. The size of the region Δ was $12 \times 12 \times 18$ mm.

To generate 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 spatial deviation 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

$$\text{uncertaint}\, y_i(x) = d_{\max,i}(x) = \max_{k=1}^{K'} \{|d_{k,i}(x) - d_i(x)|\}$$

where $d_{k,i}(x)$ is one of the K' test DVFs with not increasing local SSD and $d_i(x)$ is the result of the *B*-spline registration. i denotes the dimension of space $(1 \leq i \leq 3)$. In this work, a value of 400 was selected for K.

The underlying idea of this approach is that the result of the registration is locally ambiguous or may be locally improved by an additional random deformation. This may be due to errors as described in Section II-B. The quantity $d_{\max}$ is therefore regarded as a measure of the local registration uncertainty.

Note that here solely the local SSD is regarded. A test deformation may improve the local SSD, while the global SSD increases due to the influence of the coefficient changes elsewhere, outside the locally regarded area.

D. Registration Error in Image Subregions

We consider $err_i(x) = |d_i(x) - d_{\text{true},i}(x)|$ 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 $d_{\text{true},i}(x)$ the ground truth which is generally unknown. 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 locally present. In case of statistical dependence, however, the values of $d_{\max,i}$ may allow the estimation of the average of $\text{err}_i(x)$ for a larger entity of voxels with similar $d_{\max,i}$ values. In this section we explain how to exploit the information that $d_{\max_i}$ contains on $\text{err}_i(x)$ to divide a dataset in subregions that differ in their average local image registration error $\text{err}_i(x)$. 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 $d_{\max_i}$.

For the demonstration of the statistical dependence of $d_{\max_i}$ and $\text{err}_i(x)$, see Appendix A.

III. Applications

A. Generation of Test Data

The algorithm was tested on lung datasets. To generate test data with known ground truth of the deformation, five clinical lung data sets were artificially deformed. This deformation aims to model the transition of the exhale to the inhale breathing phase. To describe the main physiological aspects of breathing motion a model for lung deformation should take the following components into account:

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 rib cage.

To our knowledge, there is no validated model available, which accounts for all these aspects including especially tissue sliding. In the following we describe how we created artificial inhale from exhale images.

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 $g_{xy}(z) = 1 + s_0$ was applied for all $z < z_0$ where $z - z_0$ is the distance from the diaphragm in cranial direction. For $z < z_0 + \Delta$, the scaling factor was

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

where $\alpha \in R$ is a constant, and Δ 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 below 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

$$q(z) = t_0 \quad \text{for all } z < z_0$$
$$q(z) = t_0 \cdot \left(1 - \frac{\log\left(1 + (z - z_0) \cdot \alpha\right)}{\log(1 + \Delta \cdot \alpha)}\right) \text{for } z_0 \leq z < z_0 + \Delta$$
$$q(z) = 0 \quad \text{for } z \geq z_0 + \Delta.$$

The closer the distance to the diaphragm, the larger is the displacement. In distance Δ 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 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 Rib Cage: The deformation inside the chest wall, $\mathrm{DVF}_{\mathrm{internal}}$, 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, $\mathrm{DVF}_{\mathrm{internal}}$ does not describe the deformation outside the chest wall. Therefore, a second deformation field, $\mathrm{DVF}_{\mathrm{external}}$, is obtained from $\mathrm{DVF}_{\mathrm{internal}}$ by setting the cranio-caudal component to zero in each voxel. $\mathrm{DVF}_{\mathrm{external}}$ is considered to describe the *external* deformation in the region of the ribs as well as outside the chest wall. $\mathrm{DVF}_{\mathrm{external}}$ 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 in the boundary region between lung and rib cage, a third step is necessary. In the following we describe an approach to modify $\mathrm{DVF}_{\mathrm{internal}}$ such that the boundary surface of the lung after warping with $\mathrm{DVF}_{\mathrm{internal}}$ gets mapped to the same surface as after warping the image based on $\mathrm{DVF}_{\mathrm{external}}$. This modification allows a simple combination of both DVFs just by using $\mathrm{DVF}_{\mathrm{internal}}$ to warp the inside and $\mathrm{DVF}_{\mathrm{external}}$ to warp the outside of the chest wall.

To realize this concept, a mask is created to distinguish the region inside from the 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 $\mathrm{DVF}_{\mathrm{internal}}$ and registered with the mask deformed based on $\mathrm{DVF}_{\mathrm{external}}$. As a result, the deformation field $\mathrm{DVF}_{\mathrm{mask}}$ is obtained.

This registration was done with the ITK demons implementation. We chose this nonparameterized method, as *B*-spline deformations should not be involved in creating the test data.

TABLE I
PARAMETERS USED FOR THE GENERATION OF THE TEST CASES

Parameter	Value
t_0	20 mm
Δ	≈Extension of lung
α	$1.3/\Delta$
s_0	≈0.12 mm
Spacing of Gaussian Functions	16 voxels

Finally, $\mathrm{DVF}_{\mathrm{internal}}$ is replaced by $\mathrm{DVF}_{\mathrm{internal}} + \mathrm{DVF}_{\mathrm{mask}}$. After this slight modification, a simple combination of the modified $\mathrm{DVF}_{\mathrm{internal}}$ and $\mathrm{DVF}_{\mathrm{external}}$ 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.

B. Application of the Algorithm on Test Data

The proposed algorithm was tested using five lung cases. To generate the test data, the exhale breathing phase of a 4D-CT was used as a starting point and the DVF described in the previous section was created using the parameters of Table I.

For t_0, 20 mm were chosen since this may be a typical diaphragm displacement in case of normal breathing. Δ was chosen to be about the longitudinal extension of the lung, since the longitudinal decompression of the lung during inhale is largest in the region of the diaphragm and decreases to zero towards the cranial side of the lung. α determines the decay of the longitudinal displacement towards the cranial side of the lung. It was chosen $1.3/\Delta$ to approximate the decay found in a true 4D CT. s_0 was chosen in the same way, to approximate the transversal extension of the chest wall found in a true 4D CT. The spacing was chosen to be 16 voxels in order to create a complex deformation. Other values were tested as well, see next subsection for details. The resulting DVF was then used to simulate the inhale image based on the exhale image. To test 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 can be analyzed.

C. Robustness of the Algorithm

The robustness of the algorithm as well as its potential to classify the quality of the registration result was investigated using three different deformations for case 1, which differ in the magnitude and spatial variability of the DVF. This was achieved by using different spacings of the Gaussian functions that contribute to the random component of the DVF. Spacings 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.

D. Reproducibility

As described in the theory section, the $d_{\max}$ values are obtained from random variations of the coefficients and there-

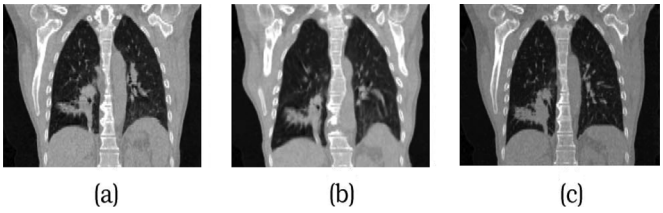

Fig. 1. Exhale image from the 4D-CT (a) and inhale image (b) simulated on the basis of the exhale image (a) using the method described in Section III-A. For comparison an inhale image from the 4D-CT is also shown (c).

TABLE II
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

fore it is important to show that the result for $d_{\max}$ is over all reproducible. To show that the proposed approach yields repeatable results for $d_{\max}$, we applied our method nine times to test case one, using independent sets of coefficient variations. $\{r_1, \ldots, r_{3N}\}$ In each voxel the standard deviation over the $d_{\max}$ values, obtained from the nine applications of the algorithm was calculated. This standard deviation can be regarded as a measure for the reproducibility.

IV. RESULTS

A. Test Data

Although the deformation model described in the previous chapter may not take all aspects of breathing motion into account it does account for the main physiological components. Fig. 1 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.

B. Results of the Registration

After registration of the simulated inhale image with the initial exhale image, the resulting DVF can be compared with the ground truth and the local registration error can be calculated in each voxel. Table II shows the average registration error over the complete body region for each of the five test cases.

C. Registration Error in Subregions

Since $d_{\max}$ is a measure for the uncertainty regarding larger groups of voxels, the average registration error should increase with increasing $d_{\max}$. As a typical example, Fig. 2 displays the average local registration errors obtained from the ground truth

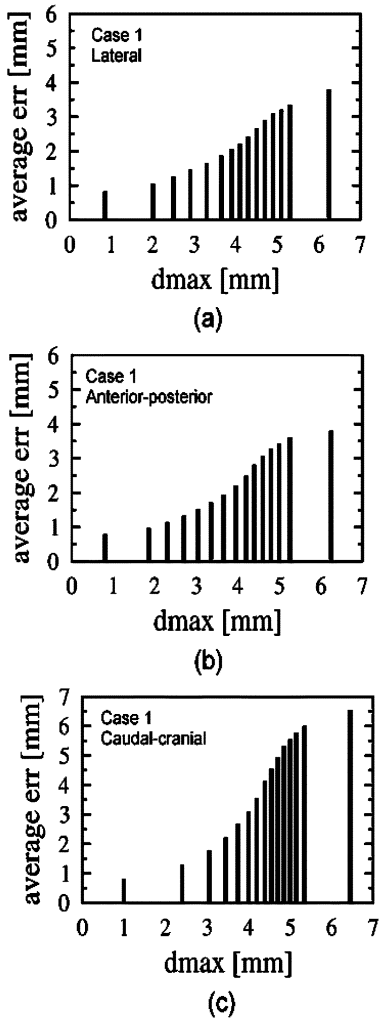

Fig. 2. Average registration error (ground truth) as a function of $d_{\max}$ for case 1. Note: To obtain equal number of entries per bin, the intervals were selected nonequidistantly. The bars are centered in the respective interval.

for case 1 as a function of $d_{\max}$ using 15 intervals. It can be seen that the average registration error indeed increases with increasing values of $d_{\max}$.

It is important to note that not only the average registration error for each bin, but also the standard deviation of the registration error within each bin increases with increasing $d_{\max}$. Fig. 3 displays this 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. To understand this effect, regard Fig. 4, which shows the combined histogram for the $d_{\max}$-values and the corresponding true local registration error for the same case as in Figs. 2 and 3. The brightness of the entries represents the number of voxels showing 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 also very likely as 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. Since bins with large $d_{\max}$ values are likely

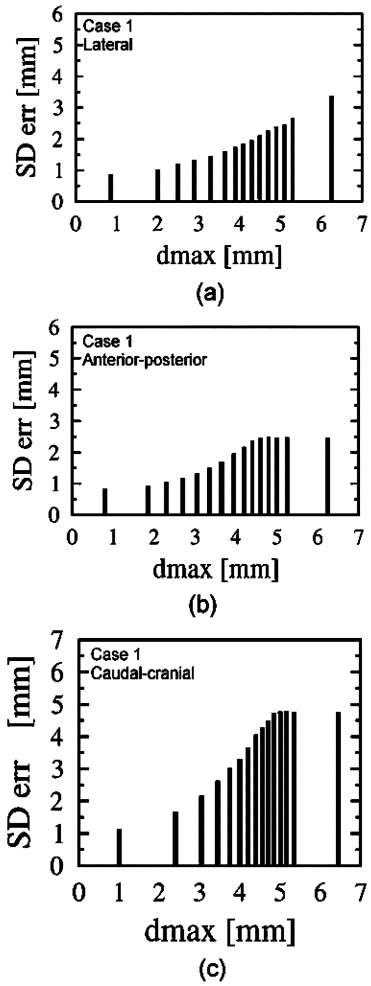

Fig. 3. Standard deviation of the registration error corresponding to Fig. 3 as a function of $d_{\max}$ for case 1. Note: To obtain equal number of entries per bin, the intervals were selected nonequidistantly. The bars are centered in the respective interval.

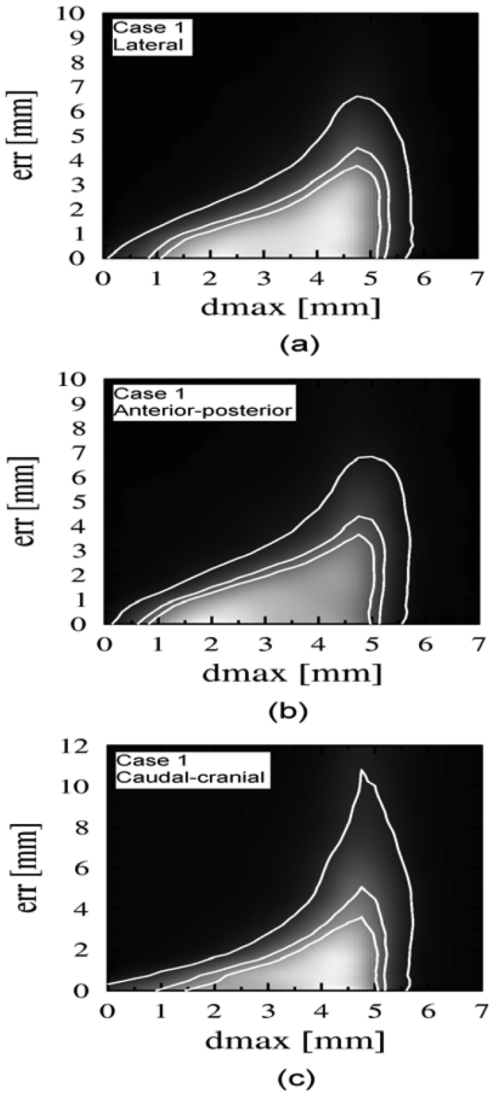

Fig. 4. Combined histogram for the $d_{\max}$-values and the corresponding true local registration error for the same case as in Figs. 2 and 3.

to contain voxels with small as well as voxels with large registration errors the standard deviation of the registration error within each bin increases with increasing $d_{\max}$.

Table III summarizes the average registration errors for all test cases investigated in this work. For a compact presentation, $d_{\max}$ was binned in three intervals only. These data show 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 registration error as well as the corresponding standard deviation.

Fig. 5 displays a color overlay of the lung image with the $d_{\max}$ values. Such a color overlay can help to guide a user of the algorithm and in particular remind him that the *B*-spline registration may not be correct, especially in those areas, displayed in red color.

D. Demonstration of the Robustness

As described in the Section II-C, the algorithm was applied on three artificial deformations simulated for case 1, with spacings 20, 16, and 12 of the Gaussian functions. A smaller spacing of the Gaussian functions causes a more severe model mismatch and therefore poorer results of the registration. The increase of the average local registration error with increasing $d_{\max}$ was found for all the three different deformations. Both, the average $d_{\max}$ as well as the average registration error increase with decreasing spacing of the Gaussian functions.

E. Reproducibility

The average of the standard deviation, over all voxels within the body region, was less than 0.5 mm for each of the three dimensions of space. In addition, the standard deviation of the voxel-based standard deviations, over all voxels within the body region, was less than 0.5 mm for each dimension. We therefore consider the values for $d_{\max}$ to be sufficiently reproducible in case 400 test deformations are carried out.

V. Discussion

As it is probably impossible to determine the registration error for each individual voxel, we focused 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

TABLE III
AVERAGE LOCAL REGISTRATION ERROR FOR $3d_{\max}$-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

algorithm as $d_{\max}$ is an estimate of the average registration error over a larger entity of voxels.

The algorithm was evaluated based on artificially deformed test images as described in Section III-A, for which the ground truth of the DVF is known. These test datasets do not represent true 4D-CTs. However, the main aspects of breathing motion are taken into account and so we consider the data to be a model for the lung motion, which is suitable to evaluate the proposed method. As the results of this evaluation may depend on the spacing of the Gaussian functions, used to generate the test cases, the analysis was repeated for different values of this parameter. The results show that the method performs well, independent from the choice this parameter. Hence, the method is robust against variation of the complexity of the deformation.

Since the standard deviation of $d_{\max}$ under repeated application of the proposed algorithm turned out to be less than 0.5 mm in average, the proposed approach yields sufficiently reproducible values to separate smaller from larger $d_{\max}$-values in the range of several millimeters.

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 II-C). 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 a subregion is not identical with the expected average local registration error. The proposed method rather provides the information in which subregion the 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 subregion 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 [13], [14], [18], [27], [19].

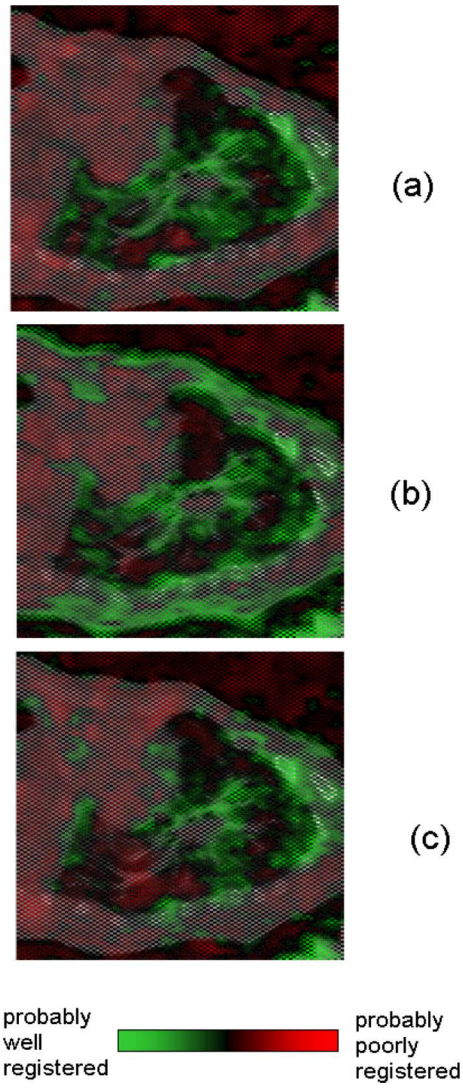

Fig. 5. Overlay of a sagittal lung image with the $d_{\max}$ values obtained for the three dimension in space: (a) lateral $d_{\max}$ component, (b) anterior–posterior $d_{\max}$ component, and (c) caudal–cranial $d_{\max}$ component.

Finally, it should be mentioned that this method was designed and tested for *B*-spline registration based on SSD. However, an estimation of the involved uncertainty by evaluating the local sensitivity of the metric to variations of the coefficients may also be an option to estimate the uncertainty of other parameterized registration algorithms.

VI. CONCLUSION

Based on the method described here it is possible to divide an image into sub regions that differ in the magnitude of their average registration error. A color overlay of $d_{\max}$ with the registered images can help the user to become aware of the registration uncertainty.

VII. FUTURE WORK

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 of 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

$$\mathrm{dose}_{d\max}(x) = \max_{k=1}^{K'} \left\{ \left| \mathrm{dose}_k(x) - \mathrm{dose}_{(x)} \right| \right\}$$

where $\mathrm{dose}\ (x)$ is the dose mapped to the voxel x based on the result of the *B*-spline registration and $\mathrm{dose}_k(x)$ 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 $\mathrm{dose}_{d\max}(x)$ and the true dose accumulation error due to registration errors. A clinician should be able to identify areas of the image where large dose accumulation errors due to imperfect image registration are likely to occur. Registration errors may be one significant, however, not the only error of dose accumulation in the lung, as other error sources such as artefacts of the 4DCT due to imperfect resampling [14], [29] may affect dose accumulation as well.

APPENDIX

Statistical Dependence Between $d_{\max,i}$ and $\mathrm{err}_i(x)$: We demonstrate the statistical dependence between $d_{\max,i}$ and $\mathrm{err}_i(x)$ 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 $\mathrm{err}_i(x)$-values. Let $p_d(n)$ be the probability that a $d_{\max,i}$ value belongs to the nth bin of the histogram of $d_{\max,i}$-values and $p_{\mathrm{err}}(m)$ the probability that $\mathrm{err}_i(x)$ belongs to the mth bin of the histogram of $\mathrm{err}_i(x)$ values. Let $p(m,n)$ be the probability of the joint event that $d_{\max,i}$ contributes to bin n and $\mathrm{err}_i(x)$ to bin m.

The mutual information is then calculated by

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

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

The mutual information is just one member of the class of the f-information measures as described in [30] and it is not the only possible choice for this purpose. However, since it is well known [31] and the most commonly used information measure, we selected the mutual information.

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 described in Section III-A:

After registering these with the undeformed images and applying the algorithm described in the Section II-C, a field of $d_{\max,i}$ values is obtained for each dimension of space i. As the deformation of the test data was predefined, the ground truth of the deformation $d_{\mathrm{true},i}\,(x)$ and hence the local registration error $\mathrm{err}_i(x)$ 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 redistributed 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 $\mathrm{err}_i(x)$ can be expected and any deviation of the MI from zero is to be regarded as noise. This procedure is repeated 200 times and so the random distribution of the MI-values for the case of statistical independence is estimated. $d_{\max,i}$ and $\mathrm{err}_i(x)$ 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.

Since the largest MI-values, obtained for any of the 200 random redistributions were three orders of magnitude smaller than the initial MI-values, the quantities $d_{\max}$ and err can be regarded as statistically dependent.

ACKNOWLEDGMENT

The authors would like to thank R. Kashani and J. M. Balter for inspiring discussions and their help in understanding the research problems and clinical challenges of elastic image registration.

REFERENCES

[1] C. Lee, K. M. Langen, W. Lu, J. Haimerl, E. Schnarr, K. J. Ruchala, G. H. Olivera, S. L. Meeks, P. A. Kupelian, T. D. Shellenberger, and R. R. Manon, "Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 71, no. 5, pp. 1563–71, Aug. 2008.

[2] M. L. Kessler, "Image registration and data fusion in radiation oncology," *Br. J. Radiol.*, vol. 79, pp. 99–108, 2006.

[3] J. M. Balter and M. L. Kessler, "Imaging and alignment for image-guided radiation therapy," *J. Clin. Oncol.*, vol. 25, no. 8, pp. 931–937, 2007.

[4] J. P. Thirion, "Image matching as a diffusion process: An analogy with maxwell's demons," *Med. Image Anal.*, vol. 2, pp. 243–260, 1998.

[5] 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.*, vol. 50, pp. 2887–2905, 2005.

[6] J. K. M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1427–1442, Nov. 2003.

[7] M. Sohn, M. Birkner, Y. Chi, J. Wang, Y. Di, B. Berger, and M. Alber, "Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy," *Med. Phys.*, vol. 35, no. 3, pp. 866–78, Mar. 2008.

[8] J. Wu and S. S. Samant, "Novel image registration quality evaluator (RQE) with an implementation for automated patient positioning in cranial radiation therapy," *Med. Phys.*, vol. 34, no. 6, pp. 2099–112, Jun. 2007.

[9] J. Kybic, "Image registration accuracy evaluation without ground truth," in *Proc. IEEE ISBI 2008*, Paris, France, May 2008, pp. 792–795.

[10] K. K. Brock, M. B. Sharp, L. A. Dawson, S. M. Kim, and D. A. Jaffray, "Accuracy of finite element model-based multi-organ deformable image registration," *Med. Phys.*, vol. 32, pp. 1647–1659, 2005.

[11] M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," *Med. Phys.*, vol. 31, pp. 2942–2948, 2004.

[12] E. Rietzel and G. T. Y. Chen, "Deformable registration of 4D computed tomography data," *Med. Phys.*, vol. 33, pp. 4423–4430, 2006.
[13] E. Heath, D. L. Collins, P. J. Keall, L. Dong, and 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.*, vol. 34, no. 11, pp. 4409–21, Nov. 2007.
[14] K. Wijesooriya, E. Weiss, V. Dill, S. Joshi, and P. J. Keall, "Quantifying the accuracy of automated structure segmentation in 4D CT images using a deformable image registration algorithm," *Med. Phys.*, vol. 35, no. 4, pp. 1251–60, Apr. 2008.
[15] J. A. Schnabel, C. Tanner, A. D. Castellano-Smith, A. Degenhard, M. O. Leach, D. R. Hose, and D. L. G. Hill, "Validation of nonrigid image registration using finite-element methods: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 238–247, Feb. 2003.
[16] P. Rogelj and S. Kovacic, "Symmetric image registration," *Med Image Anal.*, vol. 10, no. 3, pp. 484–93, Jun. 2006.
[17] J. D. Lawson, E. Schreibmann, A. B. Jani, and T. Fox, "Quantitative evaluation of a cone-beam computed tomography-planning computed tomography deformable image registration method for adaptive radiation therapy," *J. Appl. Clin. Med. Phys.*, vol. 8, no. 4, pp. 2432–2432, Nov. 2007.
[18] M. Serban, E. Heath, G. Stroian, D. L. Collins, and J. Seuntjens, "A deformable phantom for 4D radiotherapy verification: design and image registration evaluation," *Med. Phys.*, vol. 35, no. 3, pp. 1094–102, Mar. 2008.
[19] R. Kashani, J. Balter, M. Kessler, M. Hub, L. Dong, L. Zhang, L. Xing, Y. Xie, D. Hawkes, J. Schnabel, J. McClelland, and S. Joshi, "Objective assessment of deformable image registration in radiotherapy-a multi-institution study," presented at the 49th AAPM Annual Meeting, Minneapolis, MN, Jul. 2007.
[20] Zhong, T. Peters, and J. V. Siebers, "FEM-Based evaluation of deformable image registration for radiation therapy," *Phys. Med. Biol.*, vol. 52, pp. 4721–4738, 2007.
[21] H. Zhong, E. Weiss, and J. V. S. Siebers, "Assessment of dose reconstruction errors in image-guided radiation therapy," *Phys. Med. Biol.*, vol. 53, pp. 719–736, 2008.
[22] D. Ruan, J. A. Fessler, M. Roberson, J. M. Balter, and M. Kessler, "Nonrigid registration with regularization to account for local tissue rigidity," *Proc. SPIE*, Feb. 2006.
[23] 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, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Proc. 4th Int. Conf. Med. Image Comput. Computer-Assisted Intervent.*, Oct. 2001, pp. 573–581.
[24] G. K. Rohde, A. Aldroubi, and B. M. Dawant, "The adaptive bases algorithm for intensity-based nonrigid image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1470–9, Nov. 2003.
[25] I. J. Schoenberg, "Cardinal interpolation and spline functions," *J. Approximat. Theory*, vol. 2, pp. 167–206, 1969.
[26] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part I—Theory," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–833, Feb. 1993.
[27] M. Unser, A. Aldroubi, and M. Eden, "B-Spline signal processing: Part II—Efficient design and applications," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 834–848, Feb. 1993.
[28] 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, and G. S. Mageras, "Evaluation of an automated deformable image matching method for quantifying lung motion in respiratory-correlated CT images," *Med. Phys.*, vol. 33, no. 2, pp. 369–76, Feb. 2006.
[29] M. Rosu, J. M. Balter, I. J. Chetty, M. L. Kessler, D. L. McShan, P. Balter, and R. K. Ten Haken, "How extensive of a 4D dataset is needed to estimate cumulative dose distribution plan evaluation metrics in conformal lung therapy," *Med. Phys.*, vol. 34, no. 1, pp. 233–45, Jan. 2007.
[30] J. P. Pluim, J. B. Maintz, and M. A. Viergever, "F-information measures in medical image registration," *IEEE Trans Med. Imag.*, vol. 23, no. 12, pp. 1508–16, Dec. 2004.
[31] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.