B-Spline Registration of 3D Images with Levenberg-Marquardt Optimization

Sven Kabus[a,b], Thomas Netsch[b], Bernd Fischer[a], Jan Modersitzki[a]

[a]Institute of Mathematics, University of Lübeck, Wallstraße 40, 23560 Lübeck, Germany
[b]Philips Research Laboratories, Röntgenstraße 24, 22335 Hamburg, Germany

ABSTRACT

B-splines are a well-known approach for non-rigid image registration. Though successfully applied to various medical applications they exhibit a high computational complexity mainly because of the lack of dedicated optimization methods. In this work we focus on a Levenberg-Marquardt type optimization routine. As a similarity measure we use least-squares functionals such as the sum of squared differences, the cross-correlation and the local correlation measure, respectively. The latter is used for multi-modality registration tasks. The proposed registration algorithm consists of three main parts. In each iteration step one has to (a) build a linear system of equations, (b) solve this system and compute an update, (c) determine the step length for the following iteration step. Appropriate stopping criteria ensure the termination of the registration task. A standard approach for (c) and several modifications are investigated. Using a quadratic model we are able to avoid additional execution of (b) during the step length adaption. Several solvers (Cholesky, CG, pre-conditioning) for (b) have been evaluated. Also, modifications on the most time consuming task (a) are investigated, leading to a speed-up by a factor up to 30. Finally, the algorithm is embedded in a multi-scale framework (both on image and on parameter level) providing additional regularization, an increased capture range and speed-up. Convergence tests have been successfully applied for a priori known transformations. Feasibility of the proposed approach is also shown for clinical applications including PET-CT registrations (19 data sets) and MR mammography.

Keywords: 3D image registration, B-splines, Levenberg-Marquardt optimization, multi-scale decomposition

1. INTRODUCTION

The task of nonrigid image registration of three-dimensional (3D) medical image volumes is to find a vector field of 3D displacements such that each point in the reference image can be mapped onto a corresponding (meaningful) point in the template image. Typical applications include atlas construction, atlas-based segmentation or motion estimation. In the medical area there is an increasing need for comparing images. This can be done by registering the images and aligning them according to the identified vector field. In the case of multi-modality registration tasks the use of the local correlation as similarity measure has been successfully applied. For a detailed description on its use and on applications including CT–MR matching and distortion correction in diffusion tensor imaging we refer to Netsch *et al.*[1]

Registration algorithms can be separated into two basic categories. The first one is known as nonparametric. The methods of this category can be formulated in a variational form giving a necessary condition for a minimum solution which completely determines the final solution.[2] The B-spline registration used in this work belongs to the second category, the parametric methods. Here, not the final vector field but the parameters, representing this field, of an underlying model are determined.

Among registration techniques using parametric free-form deformations,[3] B-splines have shown potential for medical applications such as breast MRI,[4,5] brain[6,7] and cardiac[8] imaging. Compared to related spline representations such as thin-plate splines or elastic body splines, B-splines are computationally very attractive since a change of a control point only affects the transformation within a local neighborhood of the point. Typically, straightforward optimization methods such as steepest descent are employed at reasonable computational cost

Further author information: (Send correspondence to Sven Kabus)
[a] E-mail: {kabus,fischer,modersitzki}@math.uni-luebeck.de, Telephone/Fax: +49 451 7030-426/-436
[b] E-mail: thomas.netsch@philips.com, Telephone/Fax: +49 40 5078-2055/-2510

even for a large number of control points. However, such methods only exploit the gradient direction to determine an update step and the step length has to be adapted during the registration. Note, that a suboptimal update rule may result in a large number of iterations significantly increasing computation time or leading to less accurate registration results.
In this contribution we investigate the optimization of B-spline registration by a Levenberg-Marquardt method in the case of similarity measures that allow for a least-squares formulation such as the sum of squared differences (SSD), the cross correlation (CC), and the local correlation (LC), respectively.

The paper is organized as follows. In Section 2 we describe the problem setting in the context of B-splines. Similarity measures and the proposed optimization method are introduced followed by an outline of the underlying algorithm. Real life problems are presented in Section 3, while Section 4 provides a discussion and draws some conclusions.

2. THEORY AND METHODOLOGY

Given a reference and a template image, the aim is to find a spatial transformation such that the deformed template matches the reference image subject to a suitable similarity measure. In this work the transformation will be expressed by B-splines. Therefore we need to define a mesh of control points. Due to the local support of B-splines a displacement of a control point results in a displacement of the surrounding voxels only. Based on the transformation and on the choice of the similarity measure a functional to be minimized can be defined. The optimization routine then selects from a set of admissible displacements a transformation which minimizes this functional.

2.1. Problem formulation

Denoting by $\mathcal{W} := [0,1] \subset \mathbb{R}$ the normalized range of grey values and by $\Omega^c \subset \mathbb{R}^3$ a compact domain, we can define (discrete) images R, T as mappings $R, T : \Omega \to \mathcal{W}$, where $\Omega := \Omega^c \cap \mathbb{Z}^3$. Let the points $\boldsymbol{x} \in \Omega$ (in the following we will use the term 'voxel') be lexicographically ordered and let $\boldsymbol{x} = (x_1, x_2, x_3)^\mathsf{T} \in \mathbb{Z}^3$ with $0 \leq x_i < n_i$ for all $i = 1, 2, 3$, where n_1, n_2, n_3 denote the number of voxels in x-, y- and z-coordinate direction, respectively, and $N := n_1 n_2 n_3$ the total number of voxels. For the transformation we employ an Eulerian approach (since this results in a one to one mapping). Therefore we describe the transformation of a voxel with respect to its final position. Choosing coefficients $k_1, k_2, k_3 \in \mathbb{R}_{\geq 1}$, describing the desired width (in units of voxels) between any two mesh points (or control points) in each coordinate, a mesh G can be defined by setting

$$G := k_1\mathbb{Z}_{z_1} \times k_2\mathbb{Z}_{z_2} \times k_3\mathbb{Z}_{z_3}$$

$$\text{with} \quad z_i := \lceil \frac{n_i}{k_i} \rceil + 1 \in \mathbb{N} \quad \text{and} \quad \mathbb{Z}_e := \{x \in \mathbb{Z} \mid -1 \leq x \leq e\}.$$

By this setting, a control point is placed in every corner of Ω and an extra layer of control points around Ω is added. This guarantees that every voxel is surrounded by a local mesh cube of size $4 \times 4 \times 4$. Note that due to this extra layer, each voxel on the boundary $\partial\Omega$ is embedded in such a cube and consequently, no boundary conditions have to be selected. Let $g_i := z_i + 2$ denote the number of control points in the i-th coordinate. Thus the total number of control points of a $z_1 \times z_2 \times z_3$ mesh is given by $l_G := g_1 g_2 g_3$. Every control point can be moved into three directions. The displacement of a single control point is encoded by $\boldsymbol{\beta}_{ijk} \in \mathbb{R}^3$ and the collection of all these vectors is denoted by $\boldsymbol{\beta} \in \mathbb{R}^{3l_G}$. Following Rueckert *et al.*[4] and using the Eulerian approach we define a B-spline transformation $\boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}; \boldsymbol{\beta})$ for G as a subtraction of a deformation $\boldsymbol{u}(\boldsymbol{x}; \boldsymbol{\beta})$ from its original position $\boldsymbol{x}$. The deformation $\boldsymbol{u}$ at position $\boldsymbol{x}$ can be described by a tensor product

$$\boldsymbol{u}(\boldsymbol{x}; \boldsymbol{\beta}) = \sum_{i=1}^{g_1} \sum_{j=1}^{g_2} \sum_{k=1}^{g_3} \boldsymbol{\beta}_{ijk} b_{i,3}(x_1) b_{j,3}(x_2) b_{k,3}(x_3) \tag{1}$$

where $b_{\cdot,3}$ refers to a cubic B-spline. As a result of the local support of cubic B-splines (only four B-splines are relevant for a point $x \in \mathbb{R}$, thus there are 4^3 relevant B-splines in 3D) in (1), one has to sum up 64 summands only.

The definition of a functional to be minimized requires the choice of the similarity measure. In general, under a similarity measure we understand a mapping $\mathcal{D} : \Theta_\Omega \times \Theta_\Omega \to \mathbb{R}$, where Θ_Ω is defined as the class of all images defined on Ω with grey values in $\mathcal{W}$. For mono-modal images we employ the normalized sum of squared differences, $\mathcal{D}_{SSD}(R,T) := \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}\left[R(\boldsymbol{x})-T(\boldsymbol{x})\right]^2 = \frac{1}{N}\|R-T\|_2^2$, or cross-correlation while for multi-modal images the local correlation measure is applied. Following Netsch *et al.*[1] for both the squared cross correlation and the squared local correlation a least-squares form can be derived.
Recalling the aim of the registration task we conclude this paragraph with the formulation of the functional to be minimized and employ the sum of squared differences as similarity measure. To this end, let $F_v : \mathbb{R}^{3l_G} \to \mathbb{R}$ be a function* measuring the error for each voxel $\boldsymbol{x}_v \in \Omega$,

$$F_v(\boldsymbol{\beta}) := T(\boldsymbol{x}_v - \boldsymbol{u}(\boldsymbol{x}_v)) - R(\boldsymbol{x}_v), \qquad \boldsymbol{F}(\boldsymbol{\beta}) = (F_1(\boldsymbol{\beta}),\dots,F_N(\boldsymbol{\beta}))^\mathsf{T}. \tag{2}$$

Since the system $\boldsymbol{F}(\boldsymbol{\beta}) = \boldsymbol{0}$ has no solution in general, the optimization problem

$$\text{Solve} \min_{\boldsymbol{\beta}\in\mathbb{R}^{dl_G}} f(\boldsymbol{\beta}) \qquad \text{with} \qquad f(\boldsymbol{\beta}) = \frac{1}{2}\sum_{v=1}^{N}[F_v(\boldsymbol{\beta})]^2 = \frac{1}{2}\left\|\boldsymbol{F}(\boldsymbol{\beta})\right\|_2^2, \tag{3}$$

is considered instead.

2.2. Optimization

For the optimization we choose an iterative method producing a sequence $(\boldsymbol{\beta}^m)_{m\in\mathbb{N}_0}$ with starting value $\boldsymbol{\beta}^0 \equiv \boldsymbol{0}$ and update rule $\boldsymbol{\beta}^{m+1} := \boldsymbol{\beta}^m + \alpha_m \boldsymbol{s}^m$, where $\alpha_m \in \mathbb{R}_+$ denotes a step size and $\boldsymbol{s}^m$ a direction of descent. From the non-negativity and the continuity of the least-squares type similarity measures the existence of a minimum of (3) is guaranteed and, by employing a step size rule (Armijo rule for instance), the sequence $(\boldsymbol{\beta}^m)_{m\in\mathbb{N}}$ has a stationary point $\boldsymbol{\beta}^*$ for suitable search direction $\boldsymbol{s}^m$. Note, that we cannot guarantee $\boldsymbol{\beta}^*$ being the global minimum of (3), furthermore the solution does not have to be unique. Typically, such an optimization routine is accompanied by a multi-scale approach, in order to arrive at a sound minimum.
Newton-type optimization methods are based on the first and second order partial derivatives of the similarity function with respect to the transformation parameters $\boldsymbol{\beta}^m$. However, the calculation of second order derivatives is prohibitively costly for a large number of parameters. Therefore, using a first-order Taylor approach for $\boldsymbol{F}$ and defining

$$q^m : \mathbb{R}^{3l_G} \to \mathbb{R}, \qquad q^m(\boldsymbol{s}^m) = \frac{1}{2}\left\|\boldsymbol{F}(\boldsymbol{\beta}^m) + J_F(\boldsymbol{\beta}^m)\boldsymbol{s}^m\right\|_2^2,$$

as a quadratic model† to f we arrive at the necessary condition for a minimum of q^m,

$$J_F(\boldsymbol{\beta}^m)^\mathsf{T} J_F(\boldsymbol{\beta}^m)\boldsymbol{s}^m \stackrel{!}{=} -J_F(\boldsymbol{\beta}^m)^\mathsf{T}\boldsymbol{F}(\boldsymbol{\beta}^m),$$

which is known as Gauß-Newton approach. Here $J_m := J_F(\boldsymbol{\beta}^m)$ denotes the Jacobian of $\boldsymbol{F}_m := \boldsymbol{F}(\boldsymbol{\beta}^m)$ with entries $\partial F_v(\boldsymbol{\beta}^m)/\partial\beta_i^m$. The matrix $J_m^\mathsf{T}J_m$ is of course positive semi-definite. However, all the experiments in this work show that J_m is rank deficient. This is mainly due to the fact that for voxels in areas with constant or almost constant grey values all partial derivatives with respect to these voxels are 0. This causes zero columns in J_m. To bypass the reduced rank problem a Levenberg-Marquardt type optimization routine[6] may be applied, which adds a matrix $\lambda_m I$, where λ_m is a positive scalar parameter and I the identity matrix, to $J_m^\mathsf{T}J_m$, yielding a regular system to be solved in each iteration step,

$$\left[J_m^\mathsf{T}J_m + \lambda_m I\right]\boldsymbol{s}^m = -J_m^\mathsf{T}\boldsymbol{F}_m. \tag{4}$$

The free parameter λ_m can be seen as a 'trust region radius', thus it works as an adaption of the step length during the iteration. The choice of this radius has a great impact on the convergence of the scheme. If λ_m is chosen too small ($\lambda = 0$ coincides with a Gauß-Newton approach) instabilities of the scheme may arise, leading to a divergence of the iteration. A choice for λ_m which is too large (coinciding with a steepest descent approach) may result in an unnecessary slow convergence rate as illustrated in Section 3.2.

*To evaluate $T(\boldsymbol{x}_v - \boldsymbol{u}(\boldsymbol{x}_v))$ a trilinear interpolation is employed.

†$J_F(\boldsymbol{\beta}^m)^\mathsf{T}J_F(\boldsymbol{\beta}^m)$ gives an approximation to the Hessian $H_F(\boldsymbol{\beta}^m)$.

TOL	$\overline{\lvert I_{indiv}\rvert}$	$\mathcal{D}_{SSD}$	$\lVert\boldsymbol{\beta}-\boldsymbol{\beta}_{ref}\rVert_2$	$\lVert\boldsymbol{\beta}-\boldsymbol{\beta}_{ref}\rVert_\infty$	speed up
0	191	$< 1.000 \cdot 10^{-14}$	$7.9 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	1
10^{-8}	154	$< 1.000 \cdot 10^{-14}$	$7.6 \cdot 10^{-4}$	$2.6 \cdot 10^{-4}$	1.56
10^{-6}	121	$< 1.000 \cdot 10^{-14}$	$7.8 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	2.02
10^{-4}	77	$< 1.000 \cdot 10^{-14}$	$8.3 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	3.44
10^{-2}	18	$5.140 \cdot 10^{-14}$	$5.8 \cdot 10^{-3}$	$3.1 \cdot 10^{-4}$	23.85
10^{-1}	1	$2.700 \cdot 10^{-06}$	$1.5 \cdot 10^{+1}$	$5.3 \cdot 10^{+0}$	33.67

Table 1. Registration of a scaled copy of Figure 1 using a mesh of 2187 parameters, finite difference approximation of the partial derivatives and Cholesky decomposition for different values of TOL. The similarity measure $\mathcal{D}_{SSD}$ before registration was $1.963 \cdot 10^{-3}$.

2.3. Algorithm

The proposed registration algorithm consists of three main parts. In each iteration step one has to

(a) build a symmetric positive semi-definite matrix $J_m^\mathsf{T} J_m$ and the right hand side of (4), $-J_m^\mathsf{T} \boldsymbol{F}_m$,

(b) solve equation (4), determine the step size $\alpha_m \in [0,1]$ and compute the update $\boldsymbol{\beta}^{m+1} = \boldsymbol{\beta}^m + \alpha_m \boldsymbol{s}^m$,

(c) determine the trust region radius λ_{m+1} for the following iteration step.

Establishing the system (4) requires an approximation of the partial derivatives. In this work both finite difference method and chain rule have been employed without major differences in accuracy or computational time. Since (a) does depend on the mesh size and in particular on the image size approximations of the entries of the system matrix have been considered: Taking into account that $J_m^\mathsf{T} J_m$ is of the form

$$\left\{ \sum_{v=1}^{N} \frac{\partial F_v}{\partial \beta_i^m} \frac{\partial F_v}{\partial \beta_j^m} \right\}_{1 \le i,j \le 3 l_G}, \tag{5}$$

it is more convenient to compute the partial derivatives for the same voxel and subsequently to sum up the specific matrix entries instead of calculating each entry of $J_m^\mathsf{T} J_m$ after each other. Due to the local support of the B-splines the temporarily storage can be minimized to a vector of length $3 * 4^3 = 192$, but addressing via an index set is required. Let I_{local} be an index set which does contain the addresses of all the 192 relevant parameters. Note, that I_{local} does not change for voxels whose active control points are the same, thus we recommend instead of a globally lexicographic ordering a blockwise ordering of the voxels.
Within such an index set we observe that the majority of partial derivatives is included in $[-10^{-4}, 10^{-4}]$. Furthermore, 8/27 is an analytical upper bound for $|\partial F_v / \partial \beta_i^m|$. Omitting partial derivatives smaller than a given threshold $TOL \in [10^{-8}, 10^{-1}]$ (thus the absolute value of each summand in (5) is bounded by $0.3 * TOL$) and collecting remaining parameters in an individually index set $I_{indiv} \subset I_{local}$ reduces the number of matrix entry updates significantly from 192^2 to $|I_{indiv}|^2$ per voxel.[‡]
To illustrate this findings we comment on some numerical test runs. Table 1 shows the registration result after 80 iterations for a MR thorax image ($256 \times 256 \times 75$, FOV $450 \times 450 \times 300$ mm^3) deformed by randomly displacing the control points of a $7 \times 7 \times 7$ mesh by at most 20mm. Note that the layer of control points outside the image have not been displaced and moreover have been clamped during registration. The second column displays the dependence of the averaged $|I_{indiv}|$ on the chosen threshold. Moreover, the error in the 2– and ∞–norm in units of voxels (compared to a reference solution $\boldsymbol{\beta}_{ref}$ using $TOL = 0$) are listed. The last column shows the relative speed up factor for (a). So, the outlined techniques clearly pay off.
In contrast to step (a) the computation time for solving the system (λ_m has been determined in the previous iteration step) is independent of the image size. Several solvers are evaluated. A direct Cholesky decomposition allows fast computation for small systems (up to a $4 \times 4 \times 4$ mesh). For larger systems, however, a conjugate

[‡]In fact due to the symmetry of the matrix only 18528 resp. $|I_{indiv}|(|I_{indiv}|+1)/2$ entries have to be updated.

gradient method with a relative decrease in the residuum of 10^6 as stopping criterion turns out to be faster. Note that the convergence of the conjugate gradient method depends on the condition number of the system matrix, thus on the value of λ_m. For small λ_m, i.e. for high condition numbers, the use of a powerful preconditioner may be unavoidable. Once the system has been solved, the algorithm checks if an update $\alpha_m \boldsymbol{s}^m$ with $\alpha_m = 1$ leads to a reduction in the similarity measure. If this is not the case this update is not accepted and α_m will be reduced stepwise until $f(\boldsymbol{\beta}^m + \alpha_m \boldsymbol{s}^m) < 0.995 f(\boldsymbol{\beta}^m)$. To control the trust region radius and to determine the radius for the next iteration step we compare the predicted with the actual reduction of the similarity measure,

$$r_m = \frac{f(\boldsymbol{\beta}^m) - f(\boldsymbol{\beta}^{m+1})}{q^m(\mathbf{0}) - q^m(\boldsymbol{s}^m)}.$$

Depending on the value of r_m and on the relative decrease in f, the trust region radius for the next iteration step, λ_{m+1}, can be determined. This technique avoids additional evaluation of (b) during the trust region radius adaption. However the choice of λ_0 is a difficult task. On the one hand a smaller value for λ_0 could increase the convergence rate but on the other hand it could lead to a failure of convergence and thus to a misregistration. Especially in cases where $\boldsymbol{\beta}^m$ is far away from $\boldsymbol{\beta}^*$, i.e. during the first steps of iteration, we recommend a sufficient large choice for λ_0. In all our experiments, we simply took $\lambda_0 = 16$. Appropriate stopping criteria based on the relative progress of the similarity measure, the relative change of the mesh, and the 2–norm of ∇f relative to f ensure correct termination of the registration.
Experiments show that the capture range of a control point is approximately half the size of a mesh interval, which might be too small for medium or large deformations in the two given images. Therefore the registration algorithm is embedded into a multi-scale approach employing both an image pyramid and a parameter pyramid. Using an image pyramid the small number of voxels on a coarse image level speeds up (a) considerably. The pyramid includes scaled copies of the given image such that the i-th level contains a copy of resolution $2^{-i}n_1 \times 2^{-i}n_2 \times 2^{-i}n_3$. The down-sampling is done by an isotropic Gaussian filter. The maximum level $i_{\max}$, i.e. the start level for the optimization, is chosen such that $16 < 2^{-i_{\max}} \max_i n_i < 64$. The use of a parameter pyramid contributes to both an additional regularization of the optimization and to an increase in the capture range of the optimization. This pyramid includes the proposed mesh resolution on its finest level and coarser resolutions – up to a minimal resolution of a $2 \times 2 \times 2$ mesh – on the higher levels. The algorithm starts on the coarsest level and refines – after fulfilling the stopping criteria – the mesh by factor 2 or 3. If the finest parameter level is reached the algorithm switches to the next level of the image pyramid and proceeds with the same mesh until the stopping criteria are fulfilled again. We recommend to loosen the criteria the higher the actual level is. An example with large deformations is shown in Figure 4.

3. RESULTS

The validation of the registration results is done by visual inspection and by quantitative evaluation of the similarity measure. The visual inspection is restricted to the question of the mesh being plausible and containing no unexpected deformations. For instance, large deformations in image regions with small contrast variations do not affect the similarity measure but might hamper the convergence of the optimization. To ensure the correctness of the algorithm convergence tests are presented in the following subsection. Also, the Levenberg-Marquardt approach is compared to the steepest descent method. Finally, clinical examples are presented.

3.1. Convergence tests

For the quantitative validation of the algorithm we use two methods. Both of them evaluate the reconstruction of a known transformation. The first reference image is obtained by applying a B-spline transformation to a given image, whereas the second reference image is obtained by applying a transformation which does not belong to the class of B-spline transformations. Both are registered to the original image. For the second case rotations up to 16° and translations up to 50mm have been applied to a MR thorax image (cf. Figure 1 (left)) and could be successfully recovered. Figure 1 (right) shows the convergence plot of the registration after displacing the control points randomly. The random transformation with a maximum displacement of a control point of 20mm could be fully recovered.

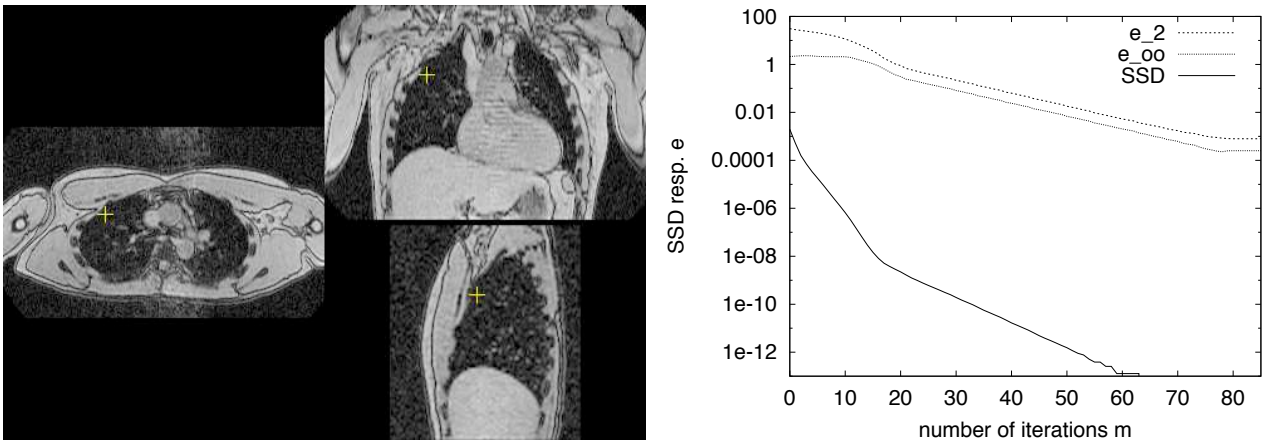

Figure 1. Convergence plot (right) for the registration of a MR thorax image (left). The solid curve displays the similarity measure, the dotted and the dashed curve display the errors $e_2 := \|\beta - \beta_{ref}\|_2$ and $e_\infty := \|\beta - \beta_{ref}\|_\infty$, respectively.

3.2. Steepest descent versus Levenberg-Marquardt

It is interesting to note that in general the Levenberg-Marquardt approach is superior to the common steepest descent approach. Here, the update $\boldsymbol{s}^m$ is determined by taking the negative gradient of f at $\boldsymbol{\beta}^m$,

$$\boldsymbol{s}^m_{sd} := -\nabla f(\boldsymbol{\beta}^m) = -J_m^{\mathsf{T}} \boldsymbol{F}_m.$$

For the step size α_m a value for $\mu \in [0, 1]$ has been chosen (we start with $\mu = 1$ and decrease μ stepwise) such that $\alpha_m \boldsymbol{s}^m_{sd} := \mu / \|\nabla f(\boldsymbol{\beta}^m)\| \boldsymbol{s}^m_{sd}$ leads to a decrease in f. Additionally we implemented the step size control scheme described by Kybic and Unser,[7] where after a successful step, μ is multiplied by 10, otherwise it is divided by 15. We observe (cf. Figure 2) that the Levenberg-Marquardt approach tends to a quadratic convergence while the steepest decent approach (independent of the step size rule) shows a linear convergence rate only. Consequently, the higher computation time per iteration while using the Levenberg-Marquardt approach pays off.

3.3. Examples

Figure 4 (top) shows corresponding ortho-views of MR images ($256 \times 256 \times 125$, FOV $250 \times 250 \times 125$ mm^3) of the knee acquired at two different flexion angles: straight (left) and bent (right). The subtraction images before (left) and after (right) B-spline registration (SSD, mesh size $5 \times 5 \times 5$, three multi-scale levels of the parameter pyramid) are depicted below. The similarity measure is reduced to 5.25%. Figure 3 illustrates the benefits from the multi-scale framework with respect to the parameter pyramid. The solid curve corresponds to a registration of the same data set using only one level of parameters: the algorithm stops after 12 iterations. A further progress in the reduction of the similarity measure can only be achieved at two or three mesh levels (dotted and dashed curves). Note that the leaps at $m = 7$ for the dotted curve and at $m = 9$ and $m = 20$ for the dashed curve result from a level switch of the parameter pyramid.

Feasibility of the proposed approach is also shown for clinical applications (cf. Figure 5). For the PET–CT registration of thorax images (19 studies from three different clinical sites, typically $128 \times 128 \times 116$, FOV $384 \times 384 \times 348$ mm^3) the transmission image is matched to the CT image by means of a $7 \times 7 \times 7$ B-spline mesh and the cross-correlation as similarity measure. To allow the use of subtraction images for visual inspection, a pre-processing is employed which converts the CT image scale to a pseudo-transmission image scale resulting in similar grey-value distributions. In all cases the visual assessment shows that the position of the lungs is well matched after registration. The computation time[§] is on the order of 50 min per registration ($TOL = 0$; cf.

[§]The computation times are taken on a Sun UltraSPARC-II, 400 MHz, 512 MByte RAM machine using C as programming language.

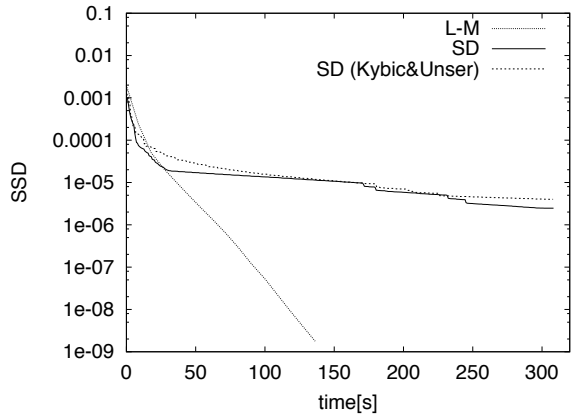

Figure 2. Convergence plot for the registration of a MR thorax image (cf. Figure 1). The optimization has been done by both steepest descent and Levenberg-Marquardt approach.

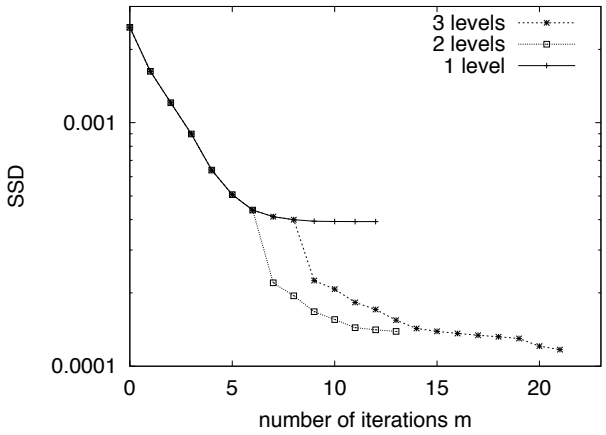

Figure 3. Comparison of a single mono- and multi-scale registration with respect to the number of decomposition levels.

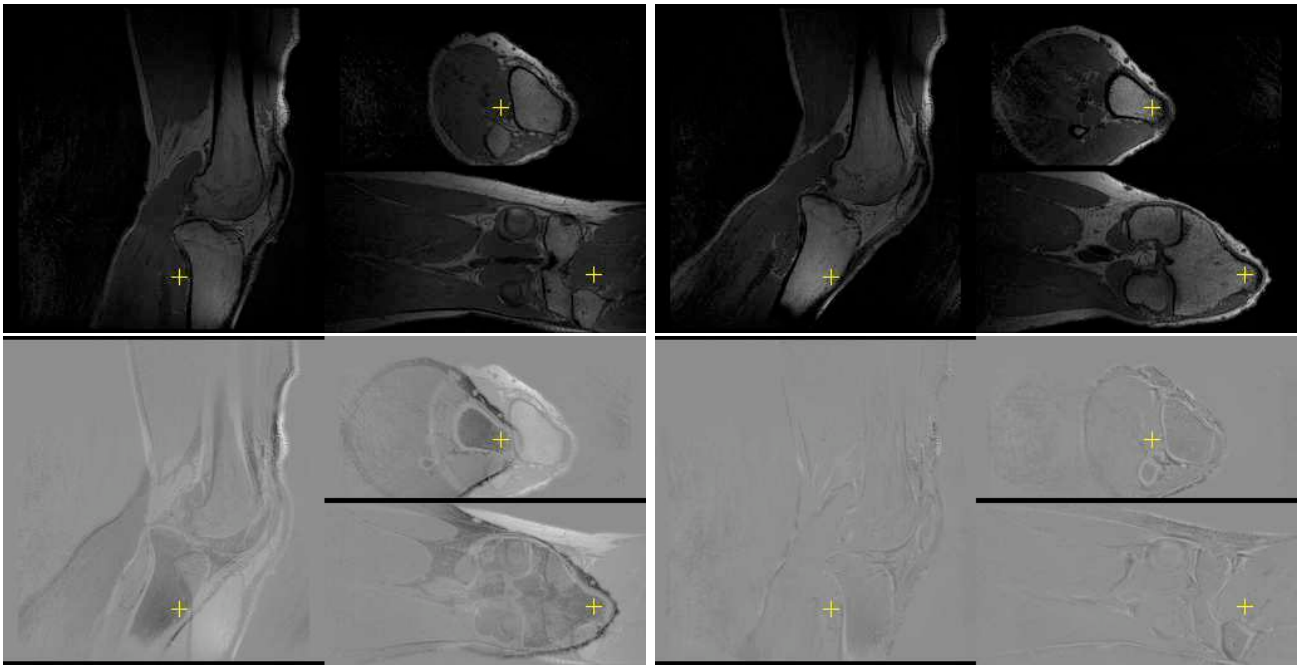

Figure 4. Registration of MR knee images at different flexion angles (top). Subtraction images before and after multi-scale registration are shown below.

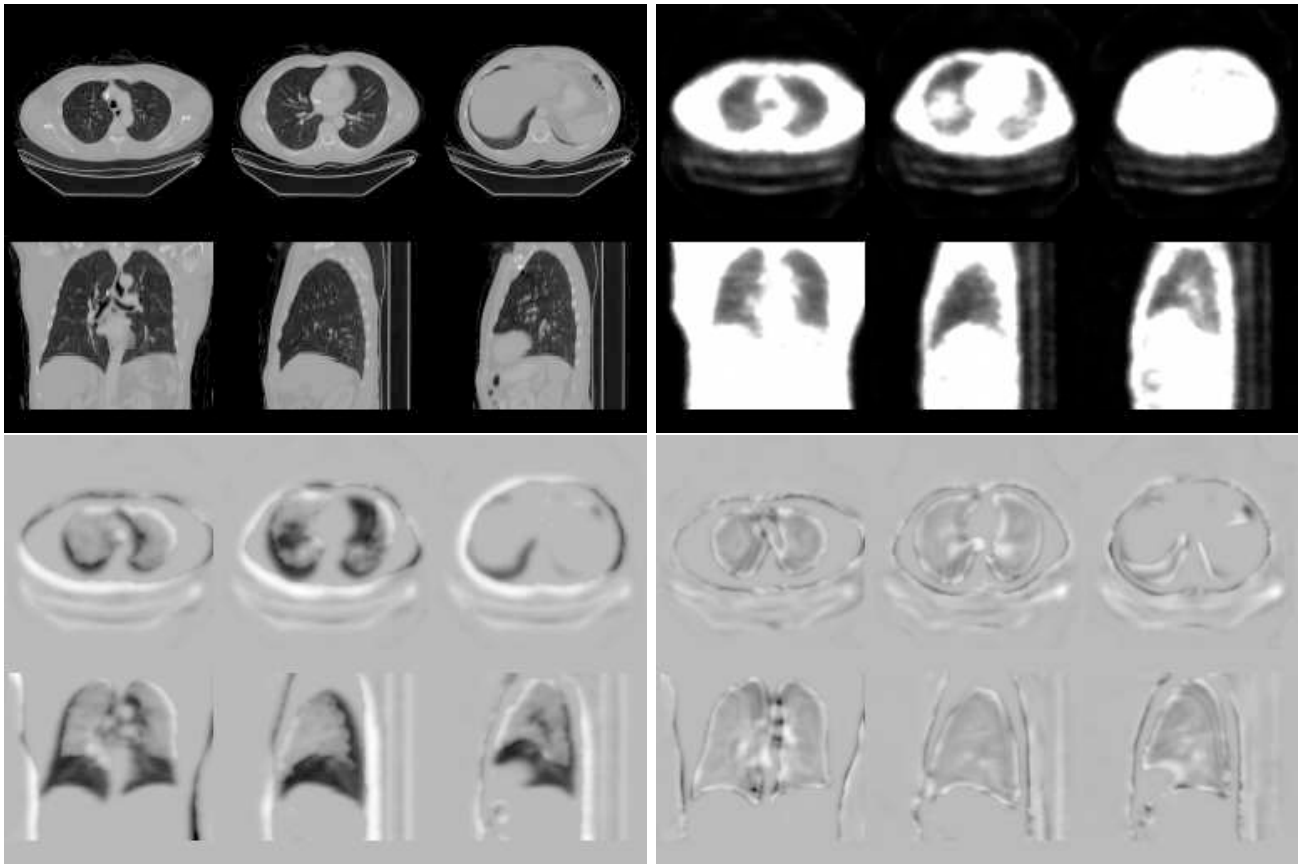

Figure 5. Registration of PET–CT thorax images at different respiration states (top) using the CC similarity measure. Subtraction images before and after registration are shown below.

Section 2.3). By setting $TOL = 0.05$ there is no visible change in the resulting image but the computation time is reduced to 3 min per registration.

As a further clinical application, motion correction of contrast-enhanced MR mammography images has been considered. Note that, due to the uptake of the contrast agent, image contrast and intensity change. Therefore mutual information is suggested as similarity measure.[5] Alternatively, in this work the local correlation is used as similarity measure. Figure 6 shows in each row for two different slices the original images ($128 \times 128 \times 28$, FOV $340 \times 340 \times 140$ mm^3) before (left) and after (second from left) injection of the contrast agent, the subtraction image before registration (centre) and after registration using the SSD measure (second from right) and using the LC measure (right). The normalized sum of squared differences for the registered images using the SSD measure resp. the LC measure does not differ significantly, both are reduced to $\approx 5\%$ indicating a removal of most of the motion-induced differences. A visual inspection indicates, however, a better registration result by using the local correlation as similarity measure. This is confirmed by the calculation of the maximum intensity projections for both measures, as shown in Figure 7.

4. CONCLUSION AND DISCUSSION

The proposed Levenberg-Marquardt optimization technique proves to be a suitable framework for B-spline image registration with least-squares similarity measures such as the sum of squared differences, the cross-correlation and the local correlation measure, respectively. Compared to a steepest descent approach it shows a higher order of convergence. Several variants to calculate the Levenberg-Marquardt parameter λ and the step length

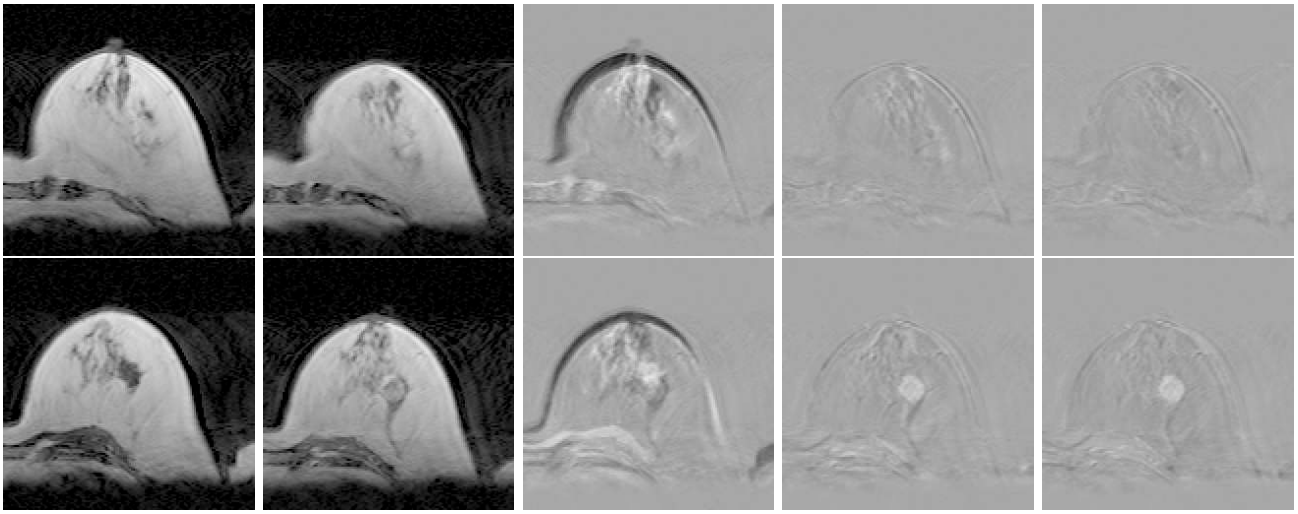

Figure 6. Registration of MR mammography images (original images are left, subtraction image before registration is centred) using the SSD similarity measure (second from right) and the LC similarity measure (right). The top row shows slice 8, the bottom row slice 13.

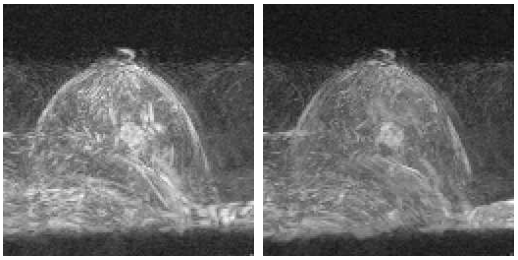

Figure 7. Maximum intensity projections of the subtraction images (cf. Figure 6) after registration using the SSD similarity measure (left) and the LC similarity measure (right).

of the parameter update have been investigated. Moreover, various termination criteria have been discussed. The method robustly handles control points in background areas, does not need boundary conditions for the control points and allows for trading between registration accuracy and computational performance: By only considering relevant image voxel derivatives as summands for the system matrix entries a significant speed-up of the registration by a factor up to 30 can be achieved. A further speed-up and additionally an increase of stability is obtained by applying a multi-scale approach which enables larger deformations as well. Capture range tests in terms of recovering a random deformation reveal the limits of our method. Typically deformations up to half of a mesh width can be fully recovered. Larger deformations could lead to a violation of the mesh topology, for instance if two control points swap places with each other. However, by employing a multi-scale approach and starting with the coarsest mesh resolution of $2 \times 2 \times 2$ the algorithm can cope even with large deformations. Furthermore we applied the algorithm to several medical applications to demonstrate its accuracy, robustness and low computational effort.

ACKNOWLEDGMENTS

PET–CT images in Figure 5 were provided by University of Pennsylvania.

REFERENCES

1. T. Netsch, P. Rösch, A. van Muiswinkel, and J. Weese, "Towards real-time multi-modality 3-D medical image registration," in *International Conference on Computer Vision*, **1**, pp. 718–725, (Vancouver, BC), 2001.
2. B. Fischer and J. Modersitzki, "A unified approach to fast image registration and a new curvature based registration technique," *to appear in Linear Algebra and its Applications*, 2004.
3. J. V. Hajnal, D. L. Hill, and D. J. Hawkes, eds., *Medical Image Registration*, CRC Press, Boca Raton, FL, 2001.
4. D. Rueckert, L. I. Sonoda, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging* **18**(8), pp. 712–721, 1999.
5. E. Denton, L. Sonoda, D. Rueckert, S. Rankin, C. Hayes, M. Leach, D. Hill, and D. Hawkes, "Comparison and evaluation of rigid and non-rigid registration of breast MR images," *Journal of Computer Assisted Technology* **23**, pp. 800–805, 1999.
6. P. Thévenaz and M. Unser, "Spline pyramids for inter-modal image registration using mutual information," in *Image Processing*, *SPIE Proceedings* **3169**, pp. 236–247, 1997.
7. J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Transactions on Image Processing* **12**(11), pp. 1427–1442, 2003.
8. D. Perperidis, A. Rao, R. Mohiaddin, and D. Rueckert, "Non-rigid spatio-temporal alignment of 4D cardiac MR images," in *2nd International Workshop on Biomedical Image Registration*, (Philadelphia, PA), 2003.