A Multilevel Method for Image Registration

Eldad Haber*   Jan Modersitzki*

May 12, 2004

Abstract

In this paper we introduce a new framework for image registration. Our formulation is based on consistent discretization of the optimization problem coupled with a multigrid solution of the linear system which evolve in a Gauss-Newton iteration. We show that our discretization is h-elliptic independent of parameter choice and therefore a simple multigrid implementation can be used. To overcome potential large nonlinearities and to further speed up computation, we use a multilevel continuation technique. We demonstrate the efficiency of our method on a realistic highly nonlinear registration problem.

1   Introduction and problem setup

Image registration is one of today's challenging image processing problems. Given a so-called reference R and a so-called template image T, the basic idea is to find a "*reasonable*" transformation such that a transformed version of the template image becomes "*similar*" to the reference image. Image registration has to be applied whenever images resulting from different times, devices, and/or perspectives need to be compared or integrated; see, e.g. [5, 17] and references therein.

The computation of nonrigid image registration has two main building blocks. The first one is a distance measure $\mathcal{D}$ quantifying distance or similarity of images and the second one is a regularizer $\mathcal{S}$ which outrules non-reasonable solutions. Note that image registration is an ill-posed problem;

---

*Dept of Mathematics and Computer Science, Emory University, Atlanta GA 30322 {haber,modersit}@mathcs.emory.edu.

see, e.g., [17]. Therefore, regularization is inevitable. A common treatment of the registration problem is based on the following variational approach. Find a transformation $\boldsymbol{u}$ minimizing the joint energy

$$\mathcal{J}^\alpha[\boldsymbol{u}] := \mathcal{D}(R, T(\boldsymbol{u})) + \alpha \mathcal{S}(\boldsymbol{u}). \tag{1}$$

Here, $\alpha > 0$ is a regularization parameter and compromises between similarity and regularity.

The formulation (1) is common to many ill-posed inverse problems where the data fitting term $\mathcal{D}$ is balanced with a regularization term using the regularization parameter α. Many authors have dealt with the optimal choice of α using a variety of approaches; see for example [13, 21]. If the difference between the images is associated with random noise then methods which are based on the statistics of the noise such as GCV [8] or χ^2 [19] can be used. However, in image registration the difference between two images (at least in most medical applications) is not associated with random noise. In fact, the difference may be highly structured and only a trained professional (e.g. radiologist) could determine the exact amount of regularization needed. Therefore, we prefer to provide a sequence of registered image pairs for various α's and leave the final choice for an optimal α to a trained professional.

The optimization problem (1) cannot be solved analytically, in general. Thus, numerical schemes and appropriate discretizations are required. Typically, a discretization leads to a finite dimensional non-linear problem, where the number of unknowns can be large. For example, the discrete registration problem for two $128 \times 128 \times 64$ MRI scans results in a non-linear systems of equations in about $3 \cdot 10^6$ unknowns. Thus, highly efficient algorithms are required in order to preform the registration in a reasonable amount of time. This applies in particular, if one is not only interested in a solution for a fixed α but in a sequence of solutions for a variety of different α's. Furthermore, since we want our codes to run in clinical setting, we require only modest computational hardware.

There are two main approaches for the discretization of the registration problem (1). In the first so-called *optimize-discretize* approach one forms the objective function, then differentiates to obtain the continuous Euler-Lagrange equations, which are finally discretized and solved numerically; see, e.g., [14, 6, 17]. The second approach is the so-called *discretize-optimize* approach. Here, one directly discretizes the problem and then solves a finite (but typically high) dimensional optimization problem; see, e.g.,[12]. The

2

advantage of the latter approach is that standard optimization methods can be used. We prefer the discretize-optimize approach, however, in order to take advantage of efficient optimization techniques, all parts of the discrete problem need to be continuously differentiable. We discuss the implications and advantages of this necessity in the paper.

To gain additional computational efficiency and obtain a scalable algorithm, one can either use a Full Approximation multigrid Scheme (FAS) or a multilevel Newton-type scheme with an efficient solver for the linearized systems of equations. In this paper we propose a multilevel inexact Gauss-Newton scheme combined with a multigrid solver for a computation of a numerical solution of (1). The latter approach has an efficiency similar to FAS [20] but it is much more modular. Particularly, the linear solver is decoupled from the optimization strategy. Therefore, the code can be modified easily such as to deal with any differentiable distance measure and/or regularizer. However, for the sake of simplicity, we focus on the so-called *sum of squared differences* (SSD) as a distance measure $\mathcal{D}$ and the elastic potential as a regularizer $\mathcal{S}$:

$$
\mathcal{D}[\boldsymbol{u}] \;=\; \tfrac{1}{2}\int_{\Omega}\Big(T(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x}))-R(\boldsymbol{x})\Big)^2 \; d\boldsymbol{x} \tag{2a}
$$

$$
\mathcal{S}[\boldsymbol{u}] \;=\; \int_{\Omega}\langle \mathcal{B}\boldsymbol{u},\mathcal{B}\boldsymbol{u}\rangle \; d\boldsymbol{x} \tag{2b}
$$

$$
:= \int_{\Omega}\frac{\lambda+\mu}{2}\|\nabla\cdot\boldsymbol{u}\|^2+\frac{\mu}{2}\sum_{i=1}^{d}\|\nabla\boldsymbol{u}_i\|^2 \; d\boldsymbol{x}.
$$

where $\Omega = (0,1)^d$ is the registration domain and λ and μ are the Lam'e constants.

Our discretization is based on the weak form of the problem (1) which naturally leads to staggered grids discretization. To our best knowledge, staggered grids discretization has been introduced to imaging in [12]. In this paper, we proof the h-ellipticity of the discretization for the elastic potential, independent of the problem parameters. As has already been shown in [12], this discretization is also stable for point-wise volume preserving constraints.

Our work relates to the work of Henn and Witsch [14], Clarenz, Droske, and Rumpf [6] and Kalmoun and Ulrich [16]. In [14], a Full Approximation Scheme (FAS) based on the discretized optimality conditions for the continuous problem is presented. In [6], a diffusive rather than an elastic regularizer

is used. In both papers, a d-linear interpolation scheme is used to compute the deformed template. Therefore, the objective function is not continuously differentiable and this may lead to failure of standard optimization techniques. In [16] a multigrid method is developed for the related optical flow problem but with the diffusive regularizer.

This paper is organized as follows. In Section 2, we discuss the discretization of the registration problem (1). Our optimization approach which is based on a Gauss-Newton type scheme (cf., e.g., [18]) is discussed in Section 3. In Section 4, we propose a multigrid method for a solution of the linearized problems. Using a local mode analysis we show that our discretization is particularly amendable to multigrid methods and suggest appropriate multigrid treatment. In Section 5, we combine our nonlinear iteration with a multilevel continuation method. Finally, in Section 6, we present numerical examples that demonstrate the effectiveness of our algorithm.

2 Discretization

Choosing a stable discretization method for a system of partial differential equations (PDE's) with mixed derivatives is a delicate matter. Here, we use staggered grids which are very common for stable discretizations of fluid flow (see, e.g., [7]) and electromagnetics (see, e.g., [23, 10]) where operators such as the gradient, curl, and divergence are discretized. It is also well known that staggered grids are tightly connected to mixed finite elements methods which are commonly used for elasticity [3].

In this section we shortly summarize the discretization we use. Further discussion and details are given in [12].

2.1 Discretizing u

We assume that our discrete images have $m_1 \times \ldots \times m_d$ pixels, where $d = 2, 3$ is the image dimensionality. For ease of presentation, we also assume that each pixel is square with lengths h. In our description we allow for half step indices. As usual in image processing, we identify pixels/voxels with cell centered grid points $x_{j,k,\ell}$, which are therefore labeled with full integers indices. Given a pixel/voxel $x_{j,k,\ell}$, their faces are numbered with a half index, $x_{j\pm\frac{1}{2},k,\ell}$, $x_{j,k\pm\frac{1}{2},\ell}$, and $x_{j,k,\ell\pm\frac{1}{2}}$, and we discretize the ith component u^i of $\boldsymbol{u}$ on the ith face for every pixel/voxel. With some abuse of notation, we denote

the discrete analog of the continuous vector field by $\boldsymbol{u} = (u^1, \ldots, u^d)^\top$, where u^i denotes the grid function which is approximated on the face-staggered grid.

If needed, the derivatives $\partial_j u^k$ are approximated by the short (central) differences,

$$\partial_j u^k \approx \partial_j^h u^k := \tfrac{1}{h}(u^k_{\ldots, i_j + \frac{1}{2}, \ldots} - u^k_{\ldots, i_j - \frac{1}{2}, \ldots}). \tag{3}$$

Note that no boundary conditions are needed to approximate derivatives in the normal directions $(\partial_j u^j)$. For the tangential directions $(\partial_j u^k, \, j \neq k)$ we imposed Neumann boundary conditions.

2.2 Discretizing $\mathcal{S}$

Since many regularizers are phrased in terms of the more complex differential operators gradient ∇ and divergence $\nabla\cdot$, we introduce the notation ∇_j^h and $\nabla^h\cdot$ for the discrete analogs,

$$\nabla_j^h \, u^j = (\partial_1^h u^j, \ldots, \partial_d^h u^j)^\top \quad \text{and} \quad \nabla^h \cdot \boldsymbol{u} = \sum \partial_j^h u^j. \tag{4}$$

Using these discrete analogs the elastic potential $\mathcal{S}$ (2c) is discretized as

$$\mathcal{S}^h(\boldsymbol{u}) = \|\boldsymbol{B}\boldsymbol{u}\|_2^2 := \frac{\lambda + \mu}{2} \| \nabla^h \cdot \boldsymbol{u}\|^2 + \frac{\mu}{2} \sum_{i=1}^{d} \| \nabla_j^h \, \boldsymbol{u}_i \|^2. \tag{5}$$

In the course of the registration process we require derivatives. Upon differentiation of the regularizer we obtain the Navier-Lamé operator

$$\mathcal{S}_{\boldsymbol{u}}^h(\boldsymbol{u}) = (\lambda + \mu)(\nabla^h \cdot)^\top \nabla^h \cdot \boldsymbol{u} - \mu \Delta_h \boldsymbol{u} =: \boldsymbol{A}\boldsymbol{u}, \tag{6}$$

where Δ_h is the usual seven points discrete vector Laplacian. Note that the approach in [6] is diffusion based (i.e., setting $\lambda = -\mu$ in (2c) and (5)), such that coupling terms vanishes. In contrast to [14], where the strong form given by the product $\mathcal{B}^*\mathcal{B}$ is discretized, we discretize $\mathcal{B}$ and build the discrete product $\boldsymbol{A} := \boldsymbol{B}^\top \boldsymbol{B}$. The advantage of the latter approach is that the derivative $\mathcal{S}_{\boldsymbol{u}}^h$ is a true derivative of the discrete function $\mathcal{S}^h$. Moreover, boundary conditions are naturally related to those of the operator $\mathcal{B}$.

2.3 Discretizing T

If we require a continuously differentiable objective function we require to have a continuous image model. Since the images are typically noisy but derivatives are needed we use a smoothing B-spline to approximate the image where the smoothing parameter is chosen using the Generalized Cross Validation method (GCV) [8]. For data interpolation using B-splines see [22]. Since the grid is regular, we can quickly evaluate the spline coefficients using cosine a transform. The continuous smooth approximation is denoted by T^{spline}.

We are heading for fast and efficient optimization scheme and therefore differentiability does play a key role. Thus, although computationally superior, d-linear image approximations can not be used, since they are not continuously differentiable.

Given the staggered grid representation of $\boldsymbol{u}$ we use averaging operators P_j for the transfer to the cell centered positions, we set

$$T(\boldsymbol{u}) := T^{\text{spline}}(x^1 + P_1 u^1, \dots, x^d + P_d u^d),$$

see [12] for details. We denote the Jacobian of T by

$$T_{\boldsymbol{u}} := \frac{\partial T}{\partial \boldsymbol{u}}(\boldsymbol{u}) = \Big(\operatorname{diag}(P_1^\top \partial_1 T), \dots \operatorname{diag}(P_d^\top \partial_d T)\Big), \tag{7}$$

where the partial derivatives $\partial_j T$ are evaluated at the spatial positions $(x^1 + P_1 u^1, \dots, x^d + P_d u^d)$. Note that using a spline approximation for T, $T_{\boldsymbol{u}}$ is a sparse matrix with only four non-zero diagonals.

2.4 Discretizing $\mathcal{D}$

Our discretization of SSD (2a) is straighforward,

$$\boldsymbol{D}(\boldsymbol{u}) := \tfrac{1}{2}\|T(\boldsymbol{u}) - R\|_2^2 \quad \text{and thus} \quad \boldsymbol{D}_{\boldsymbol{u}}(\boldsymbol{u}) = T_{\boldsymbol{u}}(\boldsymbol{u})^\top (T(\boldsymbol{u}) - R).$$

3 Optimization

The discretized analog of the image registration problem (1) for fixed α reads:

$$\text{find} \quad \boldsymbol{u} \quad \text{such that} \quad \boldsymbol{J}^\alpha(\boldsymbol{u}) := \boldsymbol{D}(\boldsymbol{u}) + \alpha \boldsymbol{S}(\boldsymbol{u}) = \min. \tag{8}$$

To solve this problem numerically we use an inexact Gauss-Newton method; cf., e.g., [18].

Starting with the initial guess $\boldsymbol{u}$, at each step we compute the gradient $\boldsymbol{J}_{\boldsymbol{u}}^{\alpha}$ of (8)

$$\boldsymbol{J}_{\boldsymbol{u}}^{\alpha} = T_{\boldsymbol{u}}(\boldsymbol{u})^{\top}(T(\boldsymbol{x}+\boldsymbol{u}) - R(\boldsymbol{x})) + \alpha \boldsymbol{A}\boldsymbol{u} \tag{9}$$

and approximately solve the system of linear equations

$$\boldsymbol{H}\delta_{\boldsymbol{u}} = -\boldsymbol{J}_{\boldsymbol{u}}^{\alpha}, \tag{10}$$

using a multigrid technique which is described in Section 4. Here $\boldsymbol{H}$ is an approximation to the Hessian

$$\boldsymbol{H} = \boldsymbol{M} + \alpha \boldsymbol{A} \tag{11}$$

and $\boldsymbol{M}$ is an approximation to $T_{\boldsymbol{u}}(\boldsymbol{u})^{\top}T_{\boldsymbol{u}}(\boldsymbol{u})$. For example, for $d = 3$, we have

$$T_{\boldsymbol{u}}(\boldsymbol{u})^{\top}T_{\boldsymbol{u}}(\boldsymbol{u}) = \begin{pmatrix} (\partial_1 T)^2 & (\partial_1 T)(\partial_2 T) & (\partial_1 T)(\partial_3 T) \\ (\partial_1 T)(\partial_2 T) & (\partial_2 T)^2 & (\partial_2 T)(\partial_3 T) \\ (\partial_1 T)(\partial_3 T) & (\partial_2 T)(\partial_3 T) & (\partial_3 T)^2 \end{pmatrix}$$

Although from an optimization standpoint it is better to use $\boldsymbol{M} = T_{\boldsymbol{u}}(\boldsymbol{u})^{\top}T_{\boldsymbol{u}}(\boldsymbol{u})$ using this matrix results in a more complicated multigrid solver because simple restriction cannot be used effectively (see also [16]). In order to have a simple multigrid scheme, which uses less storage, we approximated this matrix by

$$\boldsymbol{M} = \begin{pmatrix} (\partial_1 T)^2 & 0 & 0 \\ 0 & (\partial_2 T)^2 & 0 \\ 0 & 0 & (\partial_3 T)^2 \end{pmatrix}.$$

The deformation is updated by $\boldsymbol{u} \leftarrow \boldsymbol{u} + \gamma \delta_{\boldsymbol{u}}$, where γ is chosen according to an Armijo line search, and the process is repeated until convergence, i.e., $\|\boldsymbol{u} - \boldsymbol{u}_{\text{old}}\| \leq \text{tol}$. The algorithm is summarized in Algorithm 1.

3.1   Continuation in α

As discussed in the introduction, the optimal value of the regularization parameter α is in general unknown a-priori. In order to estimate a reasonable α, we follow the strategy suggested in [11]. Starting with a large α_0, we compute a sequence of solutions $\boldsymbol{u}^{\alpha_j}$, where $\alpha_{j+1} < \alpha_j$. This sequence can be viewed as a continuation with respect to the regularization parameter α.

7

Algorithm 1 Gauss-Newton method for image registration:
$\boldsymbol{u} \leftarrow \mathrm{GNIR}(\alpha, \boldsymbol{u})$;
---
   while true **do**
      compute $\boldsymbol{J}_{\boldsymbol{u}}^{\alpha}$ and $\boldsymbol{H}$;
      approximately solve $\boldsymbol{H}\delta_{\boldsymbol{u}} = -\boldsymbol{J}_{\boldsymbol{u}}^{\alpha}$;
      if $\|\delta_{\boldsymbol{u}}\| < \mathrm{tol}$ **then**
         break;
      end if
      using a line search set $\boldsymbol{u} \leftarrow \boldsymbol{u} + \gamma\delta_{\boldsymbol{u}}$
   end while

For a large regularization parameter the registration problem is almost quadratic and therefore its solution requires only a few iterations using a Newton-type optimization technique. Let $j = 0$, starting with α_j, we compute $d_j := \boldsymbol{D}(\boldsymbol{u}^{\alpha_j})$ and expect at least the solution $\boldsymbol{u}^{\alpha_0}$ to under-fit the data. We use $\boldsymbol{u}^{\alpha_j}$ as an excellent starting guess for the computation of the solution of the registration problem for α_{j+1}. We stop, if a solution $\boldsymbol{u}^{\alpha_j}$ becomes to irregular. Irregularity has to be provide by an trained professional and enters into the code by choosing an appropriate value of c for the data-fit. The algorithm is summarized in Algorithm 2.

Algorithm 2 Continuation in α: $[\boldsymbol{u}, \alpha] \leftarrow \mathrm{continuation}(\boldsymbol{u}, \alpha, c)$
---
   set $\boldsymbol{u}_0 \leftarrow \boldsymbol{u}$, choose $\eta < 1$;
   while true **do**
      solve (1) using Alg. 1: $\boldsymbol{u} \leftarrow \mathrm{GNIR}(\alpha, \boldsymbol{u})$;
      if $\mathcal{D}(\boldsymbol{u}) < c$ **then**
         break;
      end if
      relax the regularization parameter: $\alpha \leftarrow \eta\alpha$;
   end while
---

4   A multigrid solution of the linear systems

The challenging part within the nonlinear optimization is the solution of the linear systems. The system can be very large and is strongly coupled but it has a few characteristics which are particularly amendable to multigrid

methods. We now discuss a multigrid method for the solution of the linear systems. General comments on smoothing, prolongation, restriction, and coarse grid solution are given in [20].

4.1 h-ellipticity

To show that the discrete equations (10) are amendable to multigrid methods we first analyze them by local Fourier analysis (LFA). For simplicity, we preform the analysis in 2D but it can be directly generalized into 3D.

Recall that the Hessian for the unconstrained optimization problem reads

$$\boldsymbol{H} = \alpha\boldsymbol{A} + \boldsymbol{M},$$

where $\boldsymbol{M}$ is a non-negative diagonal matrix. A LFA with freezing coefficients takes the worst case scenario into account, i.e. $\boldsymbol{M} = \boldsymbol{0}$. Therefore, we need to study the matrix $\boldsymbol{A}$ alone, which is a discretization of the Navier-Lamé operator

$$
\begin{aligned}
\mathcal{A} &= \mathcal{B}^*\mathcal{B} := -\mu\Delta - (\lambda + \mu)\nabla\,\nabla\,\cdot \\
&= \begin{pmatrix} (2\mu + \lambda)\partial_{x_1,x_1} + \mu\partial_{x_2,x_2} & (\lambda + \mu)\partial_{x_1,x_2} \\ (\lambda + \mu)\partial_{x_1,x_2} & \mu\partial_{x_1,x_1} + (2\mu + \lambda)\partial_{x_2,x_2} \end{pmatrix}
\end{aligned}
$$

where the so-called Lamé constants λ and μ reflect material properties. Staggered grid discretization leads to the following four stencils for the four parts of the operator:

$$
\left.
\begin{aligned}
S^{1,1} &= \begin{pmatrix} 0 & -a & 0 \\ -b & 2(a+b) & -b \\ 0 & -a & 0 \end{pmatrix}, & S^{1,2} &= c\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \\
S^{2,1} &= (S^{1,2})^\top, & S^{2,2} &= (S^{1,1})^\top,
\end{aligned}
\right\} \tag{12}
$$

abbreviating $a = (2\mu + \lambda)$, $b = \mu$, and $c = \lambda + \mu$.

Applying the discretized operators $\boldsymbol{A}_h$ to a grid function $\chi(\boldsymbol{x}, \boldsymbol{\theta}) = e^{\frac{i}{h}\theta^\top \boldsymbol{x}}$, we obtain the symbols of our discretized operators

$$
\widetilde{A}_h(\boldsymbol{\theta}) = \begin{pmatrix} 2aw_1 + 2bw_2 & -4c\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2} \\ -4c\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2} & 2bw_1 + 2aw_2 \end{pmatrix},
$$

where $w_j := 1 - \cos\theta_j = 2\sin^2\frac{\theta_j}{2}$. Its determinant is given by

$$
\begin{aligned}
d(\boldsymbol{\theta}) = \widetilde{d}(\boldsymbol{w}) &= 4abw_1^2 + 4abw_2^2 + 4(a^2 + b^2)w_1w_2 - 4c^2w_1w_2 \\
&= 4ab(w_1 + w_2)^2. \tag{13}
\end{aligned}
$$

Theorem 1 *The h-ellipticity measures of the staggered grid discretizations* (12) *is*

$$E_h(\boldsymbol{A}_h) = \tfrac{1}{4},$$

and in particular independent *of the Lamé constants.*

Proof: Following [20, §8.3.2],

$$E_h(\boldsymbol{A}_h) := \frac{\min\{|\det \widetilde{A}_h| : \ \theta \in T^{\mathrm{high}}\}}{\max\{|\det \widetilde{A}_h| : \ -\pi < \theta_j \leq \pi\}}.$$

From (13), we see that the min and max are independent of λ and μ.

$$\max\{d(\boldsymbol{\theta}) : \ -\pi < \theta_j \leq \pi\} = \max\{\widetilde{d}(\boldsymbol{w}) : \ 0 \leq w_j \leq 1\}$$
$$= \ \widetilde{d}(1,1) = 16ab = 16\mu(2\mu + \lambda),$$
$$\min\{d(\boldsymbol{\theta}) : \ \theta \in T^{\mathrm{high}}\} = \min\{\widetilde{d}(\boldsymbol{w}) : \ \tfrac{1}{2} \leq |w_j| \leq 1\}$$
$$= \ \widetilde{d}(\tfrac{1}{2}, \tfrac{1}{2}) = 4ab = 4\mu(2\mu + \lambda).$$

∎

Remark 1 *The above theorem is remarkable because as the ratio λ/μ becomes larger, the system becomes more ill-conditioned. Nevertheless, the Theorem suggests that an appropriate multigrid implementation works well regardless of that ratio.*

4.2   Smoothing

The h-ellipticity of our discretization (independent of choices of the Lamé constants λ and μ) guaranties that, for example, Kacztmart's relaxation has smoothing properties. Here we have implemented a damped Jacobi smoother with damping parameter of $2/3$. Since we want our code to run on modest computational architecture such as a Laptop, our implementation does not generate the matrix and only matrix vector products are calculated.

4.3   Prolongation and restriction

To fulfill the relation between the order of the differential operator and the sum of the orders of prolongation and restriction, we use linear interpolation for prolongation and its transpose for the restriction.

4.4 Coarse grid operator

Although it is possible to use a Galerkin coarse grid operator for building the coarse grid system $\boldsymbol{A}_H$, a straightforward implementation requires either additional memory or additional computations. For problems, with uniform and structured grids, one can use re-discretization to get a faster implementation. However, since re-discretization can not be applied to the diagonal matrix $\boldsymbol{M}_h$, we choose the following approach leading to a diagonal $\boldsymbol{M}_H$. In the first stage we extract the diagonal entries of $\boldsymbol{M}_h$, this corresponds to a vector with the same dimensions of u_h. Second, we generate the diagonal matrix $\boldsymbol{M}_H$ by restricting this vector and setting the diagonal elements of $\boldsymbol{M}_H$ to the restricted values.

4.5 Multigrid-cycles

The above components are linked together to a V- or a W-cycle. In our implementation we use only a single V-cycle for each non-linear iteration. As we show in the numerical experiments, a $V(3,1)$-cycle is very effective and decrease the residual in one or two orders of magnitude.

5 Multilevel continuation

Image registration can be highly non-linear and therefore, one may require many iterations to achieve a solution. An important method to save computational work and to deal with non-linearities is to use a multilevel continuation. Multilevel continuation is well established for optimization problems and systems of non-linear equations; see, e.g., [1, 2]. However, in image registration it has an additional advantage. Similar to [9, 4], we use the multilevel approach to efficiently identify the relevant range of α's. Thus, our multilevel approach uses continuation in both the grid and regularization parameter. Note that even for approaches where the images may not be smooth, the desired solution $\boldsymbol{u}^\alpha$ is. Therefore, multilevel methods provide an effective tool for the continuation problem.

Given an initial upper bound C for the image distance, using a secant method, we compute a parameter α_C, such that

$$\mathcal{D}(\boldsymbol{u}^{\alpha_C}) \approx C. \tag{14}$$

After α_C is computed, we use Algorithm 2 on the fine grid to obtain a sequence of solution $\boldsymbol{u}^\alpha$, where $c \leq \mathcal{D}(\boldsymbol{u}^\alpha) \leq C$. We summarize the multilevel algorithm in Algorithm 3.

Algorithm 3 Multilevel image registration: $\boldsymbol{u} \leftarrow \mathrm{MLIR}$

---

    choose $\boldsymbol{u}$ on the coarsest grid;
    while true **do**
        solve the registration problem $\boldsymbol{u} = \mathrm{GNIR}(\boldsymbol{u}, \alpha)$; cf. Alg. 1;
        if $|\mathcal{D}(\boldsymbol{u}) - C| \geq \mathrm{tol}$ **then**
            adjust α using one a secant method;
        end if
        if on finest grid **then**
            return;
        end if
        prolongate $\boldsymbol{u}$ to the finer grid: $\boldsymbol{u} \leftarrow \mathrm{prolongate}(\boldsymbol{u})$;
    end while

---

6   Numerical examples

In this section we preform numerical experiments that demonstrate the effectiveness of our algorithm. We present a registration of three-dimensional MRI scans of a human knee (images provided by Thomas Netsch, Philips Medical Solution, Hamburg, Germany). The problem is known to be hard because large nonlinear deformations are needed in order to perform the registration; see also [15].

As apparent from Fig. 1, the reference shows the knee in an almost straight position and the template shows the knee in a bent position. The deformation is thus expected to be highly non-linear. For the registration results shown in Fig. 1, we choose $\alpha = 10^{-3}$, which balances similarity and regularity with respect to our eyeball-norm. Fig. 1 (a,b,c) show a 3D view of the straight (R), bent (T), and registered bent knee ($T(\boldsymbol{u}^\alpha)$). As it is apparent from Fig. 1 (c), the overall registration is reasonable. We also analyze the results in terms of a pairwise comparison of image slices. A 2D view of the generic slice 40 of a total of 64 is shown in Fig. 1 (d,e,f). We add a regular grid to R and T and the deformed grid to $T(\boldsymbol{u}^\alpha)$. Though highly non-linear, the deformation appears to be sufficiently smooth, as can be seen from the
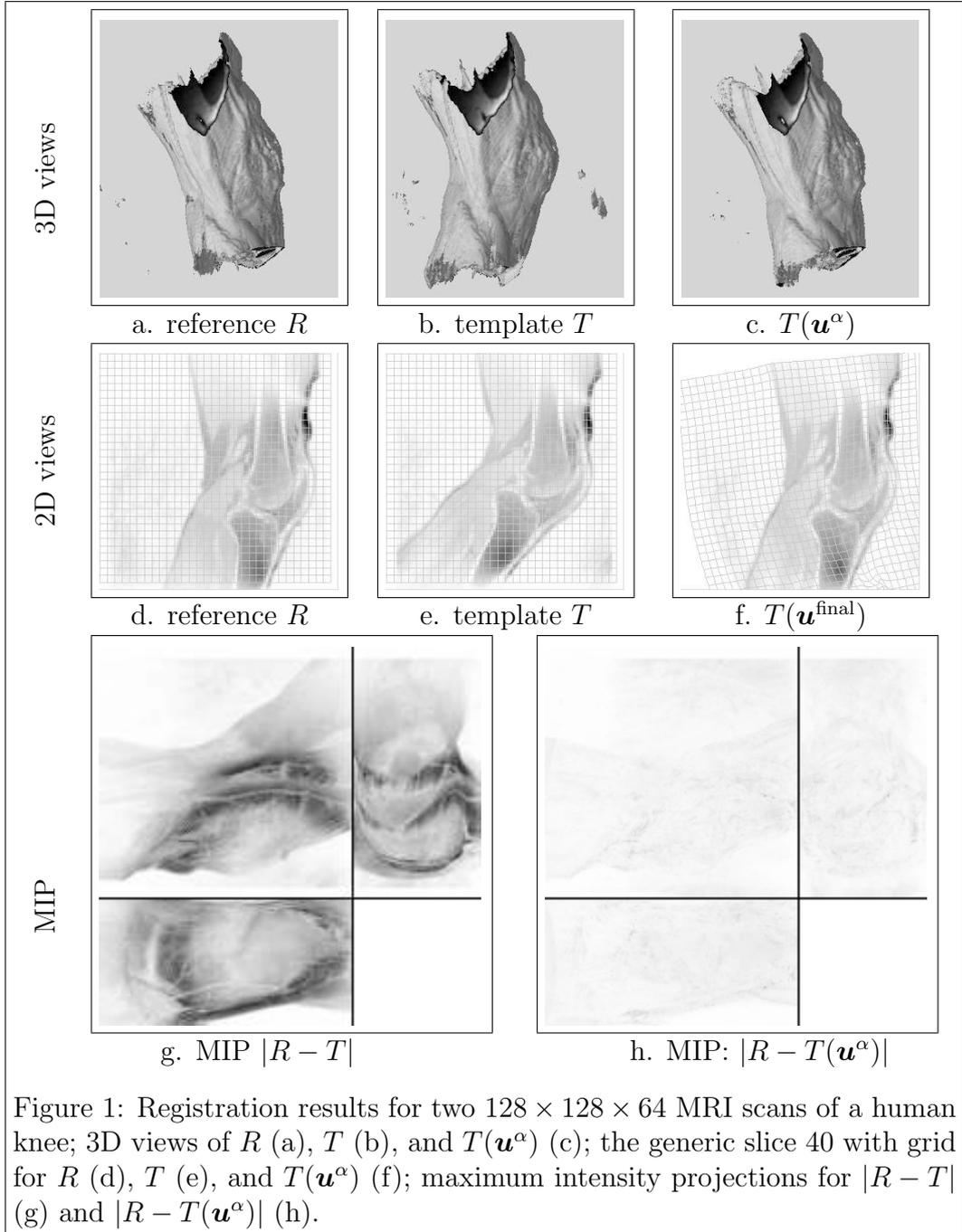
Figure 1: Registration results for two $128 \times 128 \times 64$ MRI scans of a human knee; 3D views of R (a), T (b), and $T(\boldsymbol{u}^\alpha)$ (c); the generic slice 40 with grid for R (d), T (e), and $T(\boldsymbol{u}^\alpha)$ (f); maximum intensity projections for $|R - T|$ (g) and $|R - T(\boldsymbol{u}^\alpha)|$ (h).

13

Table 1: Reductions in the SSD (in %), and iterations (in #) for the multilevel elastic registration, image size is $n^3/2$.

level	n	Iter	SSD
5	8	#13	17.2%
4	16	# 16	17.8%
3	32	# 9	17.9%
2	64	#8	17.2%
1	128	#6	17.2%

deformed grid. Finally, Fig. 1 also shows Maximum Intensity Projections (MIP's) of the distances $|R - T|$ and $|R - T(\boldsymbol{u}^\alpha)|$, respectively. The image distance has been reduced to 17.2%,

$$\frac{\boldsymbol{D}(\boldsymbol{u}^\alpha)}{\boldsymbol{D}(\boldsymbol{0})} \approx 17.2\%, \quad \text{for} \quad \alpha = 10^{-3}.$$

As common to many other registration algorithms we use a two phase registration. In the first phase a multilevel linear preregistration is done followed by a second phase multilevel non-linear registration. The linear registration managed to reduce the image distance to 50.1% and the rest was done using our nonlinear registration.

The computations are performed using MATLAB 6.1 on a 2.2GHz Laptop with 1GB of RAM. The overall time for the registration was about 118 minutes including the computation for the B-spline coefficients ($\approx$ 11 minutes).

In our computations, about one third of the time is needed for the linear part and about two third is needed for the non-linear part. Within the non-linear part, 66% are needed for the computation of $T(\boldsymbol{u})$ and $T_{\boldsymbol{u}}(\boldsymbol{u})$ and 24% are needed to solve the linear system. An overview with respect to levels is presented in Tab. 1.

Running our scheme with a regularization parameter $\alpha = 10^{-2}$ leads to visually indistinguishable results but to a much higher distance (reduction to 25.3% instead of 17.2% for $\alpha = 10^{-3}$). On the other hand, choosing α to small ($\alpha \leq 10^{-4}$ in this example) results in non-physical deformations which are visible by looking at the deformed grid. Therefore, it is easy to derive a lower bound for α. However, choosing an "optimal" regularization parameter α is a delicate matter since it requires a quantification of "physically meaningful".

We do not believe in an automatic detection and therefore leave the final choice to an expert.

To observe the effectiveness of our multigrid algorithm solver we record the reduction of the residual for each of the Gauss-Newton iterations on the finest grid. Note that for each iteration we have a different approximation to the Hessian characterized by the different M's. In our experiment, the multigrid solver was very effective and after only a single $V(3, 1)$-cycle our residual is decreased by approximately two orders of magnitudes. On average, the reduction factor was $2.1 \cdot 10^{-2}$.

7   Summary

In this paper we have developed a multilevel registration scheme for image registration based on a variational formulation. The scheme is based on a multilevel inexact Gauss-Newton scheme with a multigrid solver for the linearized systems. We make extended use of numerical optimization techniques and therefore pay special attention to differentiability issues. We use smoothing B-splines to compute the solution to the forward problem and staggered grids to discretize the transformation. We prove that the discretization is h-elliptic and therefore is amendable to multigrid methods. The numerical experiments support our theoretical prediction and show that we can solve large scale problems using modest computational tools like MATLAB.

References

[1] J. C. Alexander and J. A. Yorke, *The homotopy continuation method: numerically implementable topological procedures*, Trans Am Math Soc **242** (1978), 271–284.

[2] E. Allgower and K. Georg, *Numerical continuation methods*, Springer Verlag, 1990.

[3] Douglas N. Arnold and Ragnar Winther, *Mixed finite elements for elasticity*, Numerische Mathematik **42** (2002), 401–419.

[4] U. Ascher and E. Haber, *Grid refinement and scaling for distributed parameter estimation problems*, Inverse Problems **17** (2001), 571–590.

[5] L.G. Brown, *A survey of image registration techniques*, Surveys **24** (1992), no. 4, 325–376.

[6] U. Clarenz, M. Droske, and M. Rumpf, *Towards fast non–rigid registration*, Inverse Problems, Image Analysis and Medical Imaging, AMS Special Session Interaction of Inverse Problems and Image Analysis, vol. 313, AMS, 2002, pp. 67–84.

[7] C.A.J. Fletcher, *Computational techniques for fluid dynamics*, vol. II, Springer-Verlag, 1988.

[8] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), 215–223.

[9] E. Haber, *A multilevel, level-set method for optimizing eigenvalues in shape design problems*, JCP **to appear** (2004), 1–15.

[10] E. Haber and U. Ascher, *Fast finite volume simulation of 3D electromagnetic problems with highly discontinuous coefficients*, SIAM J. Scient. Comput. **22** (2001), 1943–1961.

[11] E. Haber, U. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse problems **16** (2000), 1263–1280.

[12] E. Haber and J. Modersitzki, *Numerical solutions of volume preserving image registration*, Tech. report, Emory University, 2004.

[13] P. C. Hansen, *Rank deficient and ill-posed problems*, SIAM, Philadelphia, 1998.

[14] Stefan Henn and Kristian Witsch, *Iterative multigrid regularization techniques for image matching*, SIAM J. on Scientific Comp. (2001), 1077–1093.

[15] S. Kabus, T. Netsch, B. Fischer, and J. Modersitzki, *B-spline registration of 3d images with levenberg-marquardt optimization*, Proceedings of the SPIE 2004, Medical Imaging, San Diego 14.–19, 2004.

[16] El Mostafa Kalmoun and Ulrich Rude, *A variational multigrid for computing the optical flow*, Tech. report, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, 2003.

[17] J. Modersitzki, *Numerical methods for image registration*, Oxford, 2004.

[18] J. Nocedal and S. Wright, *Numerical optimization*, New York: Springer, 1999.

[19] R. L. Parker, *Geophysical inverse theory*, Princeton University Press, Princeton NJ, 1994.

[20] U. Trottenberg, C. Oosterlee, and A. Schuller, *Multigrid*, Academic Press, 2001.

[21] C. Vogel, *Computational methods for inverse problem*, SIAM, Philadelphia, 2001.

[22] G. Wahba, *Spline models for observational data*, SIAM, Philadelphia, 1990.

[23] K.S. Yee, *Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media*, IEEE Trans. on antennas and propagation **14** (1966), 302–307.