Towards fast non-rigid registration

U. Clarenz, M. Droske, M. Rumpf

ABSTRACT. A fast multiscale and multigrid method for the matching of images in 2D and 3D is presented. Especially in medical imaging this problem - denoted as the registration problem - is of fundamental importance in the handling of images from multiple image modalities or of image time series. The paper restricts to the simplest matching energy to be minimized, i.e., $E[\phi] = \frac{1}{2}\int_\Omega |f_1 \circ \phi - f_2|^2$, where f_1, f_2 are the intensity maps of the two images to be matched and ϕ is a deformation. The focus is on a robust and efficient solution strategy.

Matching of images, i.e., finding an optimal deformation ϕ which minimizes E is known to be an ill-posed problem. Hence, to regularize this problem a regularization of the descent path is considered in a gradient flow method. Thus the initial value problem $\partial_t \phi = -\mathrm{grad}_g E[\phi]$ with some regular initial deformation $\phi(0) = \phi_0$ is solved on a suitable space of deformations $\Omega \to \Omega$. The gradient grad_g is measured w.r.t. a suitable regularizing metric g. Existence and uniqueness of solutions is demonstrated for different types of regularizations. For the implementation a metric based on multigrid cycles on hierarchical grids is proposed, using their superior smoothing properties. This is combined with an effective time-step control in the descent algorithm. Furthermore, to avoid convergence to local minima, multiple scales of the images to be matched are considered. Again, these image scales can be generated applying multigrid operators and we propose to resolve the pyramid of scales on a properly chosen pyramid of hierarchical grids. Examples on 2D and large 3D image matching problems prove the robustness and efficiency of the proposed approach.

1. Introduction

Image assisted diagnostics and surgery planning requires robust and valid segmentation and classification results and an analysis of the temporal change of anatomic structures. This can only be achieved properly if images recorded with different imaging machinery or at different times can suitably be correlated to each other. Various techniques have been proposed to solve this registration problem. They all ask for an "optimal" deformation which deforms one image such that there is an "optimal" correlation to another image with respect to a suitable coherence measure.

Mainly two different approaches have been discussed in the literature [**5, 6, 7, 9, 15, 17, 21**]. On the one hand, so called elastic registration techniques deal with a regularization of the energy, typically adding a convex energy functional based on gradients to the actual matching energy. The regularization energy is regarded as a penalty for "elastic stresses" resulting from the deformation of the images. This approach is related to the well known classical Tikhonov regularization of the originally ill-posed problem. On the other hand,

1991 *Mathematics Subject Classification.* Primary 65M30, 65M32, 65M55; Secondary 62H35, 68U05, 93A30.

viscous flow techniques are taken into account. They compute smooth paths from some initial deformation towards the set of minimizers of the matching energy. Thereby, a suitable regularization of the velocity, e.g., adding an artificial viscosity, ensures a certain problem dependent smoothness modulus. This class of methods can be interpreted as a gradient flow approach with respect to a metric which penalizes non–regular descent directions. Taking into account a time-step discretization this methodology is closely related to iterative Tikhonov regularization methods [**12, 20**]. Preparing this paper we got aware of recent results by Henn and Witsch [**13**]. They proposed to use multigrid smoothers in iterative Tikhonov regularization and proved its applicability for non-rigid registration in medical imaging. Concerning the impact of multigrid smoothing, this approach is closely related to our approach. In fact using the gradient flow perspective we enlighten the problem from a different point of view. In addition we embed the approach in a scale of matching problems which enable the computation of more global deformations.

Furthermore let us recall the *optical flow* method in image processing. The task is to extract motion fields from image time sequences. We ask for the time discrete motion velocity between two images of a time sequence, i.e., a short time deformation which is again a matching problem. If the motion is only piecewise smooth a simple regularization adding a Dirichlet-integral would not be able to retain the often discontinuous deformations on image edges. Thus Nagel and Enkelmann proposed an anisotropic quadratic form for the gradient of the deformation which regularizes edges of the image only in the tangential direction [**8, 18**].

Due to the non–convexity of the minimization problem in image registration it might be difficult to find the absolute minimum in case of larger deformations. Alternatively, one can consider a convolution of the images with a large corresponding filter width which destroys much of the detailed structure, match those images, and then successively reduce the filter-width and iterate the process [**3, 19, 24**]. This procedure is comparable to an annealing algorithm, where the filter width plays the role of the temperature.

In this paper we will consider one of the simplest image intensity based matching energies and apply a gradient descent approach for its minimization. The focus is on the robustness and efficiency of the proposed method and not on the generality of the approach with respect to its range of applications. In Section 8 we will give an outline on further research directions. The building blocks of the presented method are:

- a suitable choice of the regularizing metric (especially based on multigrid cycles),
- standard effective time-step control methods in the gradient descent method but now taking into account the selected new metric,
- a multiscale approach considering a series of successively smoothed images
- scale dependent grid resolution, i.e., solving coarse scale problems for sufficiently smooth images on coarse grids,
- and finally, scale dependent stopping criteria, which prevents us from resolving fine deformation details already on much too coarse scales.

Altogether these ingredients ensure a superior performance of the resulting algorithm. It allows the efficient computation of large scale matching problems with large solution deformations. Specifically, medical images of a resolution 129^3 can be matched based on deformations in the space of piecewise trilinear, continuous functions in a few minutes on a LINUX PC with a Pentium IV, 1.7 GHz processor and the resulting deformations are reasonably smooth. In what follows we will introduce the continuous model in Section 2,

show existence and uniqueness of solutions in Section 3. Section 4 contains the description of the chosen scale space method. We explain the spatial and temporal discretization in Section 5. In Sections 6 and 7 we collect the algorithmic ingredients and applications respectively and in Section 8 further extensions towards nonlinear metrics and different matching energies are sketched and we draw conclusions.

2. Problem and approach

Given two images $f_1, f_2 : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$ and $d = 2, 3$, we would like to determine a deformation $\phi : \Omega \to \mathbb{R}^d$ which maps Ω onto Ω and maps grey values in the first image f_1 via a deformation ϕ to grey values at the deformed position in the second image f_2 such that

$$f_1 \circ \phi \approx f_2 \,.$$

For the ease of presentation we assume $\Omega = [0,1]^d$ throughout this paper. We consider u as the perturbation of ϕ from the identity $\mathbb{1}$ which means $\mathbb{1} + u = \phi$. To optimize the deformation with respect to a proper match of the two images we define the most basic energy E depending on the displacement u (resp. the deformation ϕ):

$$E[u] = \frac{1}{2} \int_\Omega |f_1 \circ (\mathbb{1} + u) - f_2|^2 \,. \tag{E}$$

In what follows we use either ϕ or u as the argument of the energy E. If u is an ideal deformation the above energy vanishes. Thus we ask for minimizers of the problem $E[u] \to \min$ in some Banach space $\mathcal{V}$. Obviously, this problem is ill-posed. Consider a deformation ϕ and for $c \in \mathbb{R}$ the level sets $\mathcal{M}_c^1 = \{x \in \Omega \,|\, f_1(x) = c\}$. Then for any displacement Λ which keeps $\mathcal{M}_c^1$ fixed for all c, the energy does not change, i.e.,

$$E[\phi] = E[\Lambda \circ \phi]$$

This especially holds true for a possible minimizer. Hence, a minimizer – if it exists – is non-unique and the set of minimizers is expected to be non-regular and not closed in a usual set of admissible displacements.

A minimizer u of (E) is characterized by the necessary condition $E'[u] = 0$, where $E'[u] \in \mathcal{V}'$ for the dual space $\mathcal{V}'$ of $\mathcal{V}$. This condition can be expressed in weak form:

$$\int_\Omega (f_1 \circ (\mathbb{1} + u) - f_2) \nabla f_1 \circ (\mathbb{1} + u) \cdot \theta = 0 \,,$$

for all $\theta \in [C_0^\infty(\Omega)]^d$. Suppose $[L^2(\Omega)]^d$ is embedded in the space $\mathcal{V}'$. Under obvious integrability conditions we obtain the L^2-representation of E'

$$\operatorname{grad}_{L^2} E[u] = (f_1 \circ (\mathbb{1} + u) - f_2) \nabla f_1 \circ (\mathbb{1} + u) \,. \tag{2.1}$$

We investigate a gradient flow approach to solve this matching problem. A gradient of a functional $E : \mathcal{V} \to \mathbb{R}$ is defined as the representation of the Fréchet derivative $E' \in \mathcal{V}'$ in a metric $g(\cdot,\cdot)$ on $\mathcal{V}$, i.e.,

$$g(\operatorname{grad}_g E[u], \theta) = \langle E'[u], \theta \rangle \,.$$

One frequently identifies $E'[u]$ with the gradient of E with respect to the L^2-product. Here we introduce a different length measurement on the space of deformations and consider a general gradient flow

$$\begin{aligned} \partial_t u &= -\operatorname{grad}_g E[u] \,, \\ u(0) &= u_0 \,. \end{aligned}$$

for a suitable metric $g(\cdot\,,\cdot)$ on $\mathcal{V}$. Thus we ask for a solution $u : \mathbb{R}_0^+ \to \mathcal{V}$, such that

$$g(\partial_t u, \theta) + \int_\Omega (f_1 \circ (\mathbb{1} + u) - f_2) \nabla f_1 \circ (\mathbb{1} + u) \cdot \theta = 0\,, \tag{2.2}$$

for all $\theta \in [C_0^\infty(\Omega)]^d$. The choice of $\mathcal{V}$ and the metric g on $\mathcal{V}$ is related to a regularization of the matching problem (cf. Section 3). At least for finite time we ensure $\mathcal{V}$-regularity of the deformation. The representation of the metric g in the duality pairing $(\mathcal{V}',\mathcal{V})$ will be denoted by $A : \mathcal{V} \to \mathcal{V}'$, i.e.,

$$g(\varphi, \psi) = \langle A\,\varphi, \psi \rangle$$

for all $\psi \in \mathcal{V}$. Hence, if the inverse of A is regular in a suitable sense (cf. Section 3) the gradient flow can be rewritten as an ODE in the Banach space $\mathcal{V}$:

$$\partial_t u = -A^{-1} E'[u]\,.$$

In the following section we give examples for A and corresponding metrics g and show existence and uniqueness of solutions.

3. Existence and uniqueness

In what follows let us assume $\mathcal{V}$ to be a Banach space. Furthermore suppose that there is a second Banach space $\mathcal{W} \supset \mathcal{V}$ which is embedded in the dual space $\mathcal{V}'$ of $\mathcal{V}$. Hence, we can state the following

THEOREM 3.1. *Let A be a linear isomorphism $A : \mathcal{V} \to \mathcal{W}$. If $E'[\mathcal{V}] \subset \mathcal{W}$ and $E'[\cdot] : \mathcal{V} \to \mathcal{W}$ is Lipschitz continuous, then there exists a unique solution of the problem:*

For given initial data $u_0 \in \mathcal{V}$, find a solution $u : \mathbb{R}_0^+ \to \mathcal{V}$, such that

$$\begin{aligned} \partial_t u &= -A^{-1} E'[u]\,, \\ u(0) &= u_0\,. \end{aligned}$$

Remark: Theorem 3.1 especially ensures that solutions of the gradient flow are $\mathcal{V}$-regular for finite times. Let us emphasize that in general we can neither expect the $\mathcal{V}$-norm to be uniformly bounded in time nor that there exists a steady state.

The proof is a straightforward application of the Picard-Lindelöf Theorem in Banach spaces. We have shown that there is a L^2-representation $\mathrm{grad}_{L^2} E$ of E' (cf. Section 2.1), if f_1 and f_2 are of suitable regularity. Therefore in case $\mathcal{W} = [L^2(\Omega)]^d$ the inclusion $E'(\mathcal{V}) \subset \mathcal{W}$ is valid. Let us prove Lipschitz continuity of $\mathrm{grad}_{\mathcal{W}} E = \mathrm{grad}_{L^2} E$.

LEMMA 3.2. *Let $\mathcal{V} = \mathcal{V}' = \mathcal{W} = [L^2(\Omega)]^d$; then the derivative of the energy E w.r.t. $\mathcal{W}$ is Lipschitz continuous, i.e.,*

$$||\mathrm{grad}_{L^2} E[u_1] - \mathrm{grad}_{L^2} E[u_2]||_{L^2} \leq C\,||u_1 - u_2||_{L^2}$$

if $f_1 \in C^{1,1}(\mathbb{R}^d)$ and $f_2 \in L^\infty(\Omega)$.

Proof: Let $u_1, u_2 \in \mathcal{V}$. Then we have

$$\begin{aligned} & \mathrm{grad}_{L^2} E[u_1] - \mathrm{grad}_{L^2} E[u_2] \\ = \;& (f_1 \circ (\mathbb{1} + u_1) - f_2) \nabla f_1 \circ (\mathbb{1} + u_1) - \\ & (f_1 \circ (\mathbb{1} + u_2) - f_2) \nabla f_1 \circ (\mathbb{1} + u_2) \\ = \;& [(f_1 \circ (\mathbb{1} + u_1) - f_2) - (f_1 \circ (\mathbb{1} + u_2) - f_2)] \nabla f_1 \circ (\mathbb{1} + u_1) - \\ & (f_1 \circ (\mathbb{1} + u_2) - f_2) [\nabla f_1 \circ (\mathbb{1} + u_1) - \nabla f_1 \circ (\mathbb{1} + u_2)]\,. \end{aligned}$$

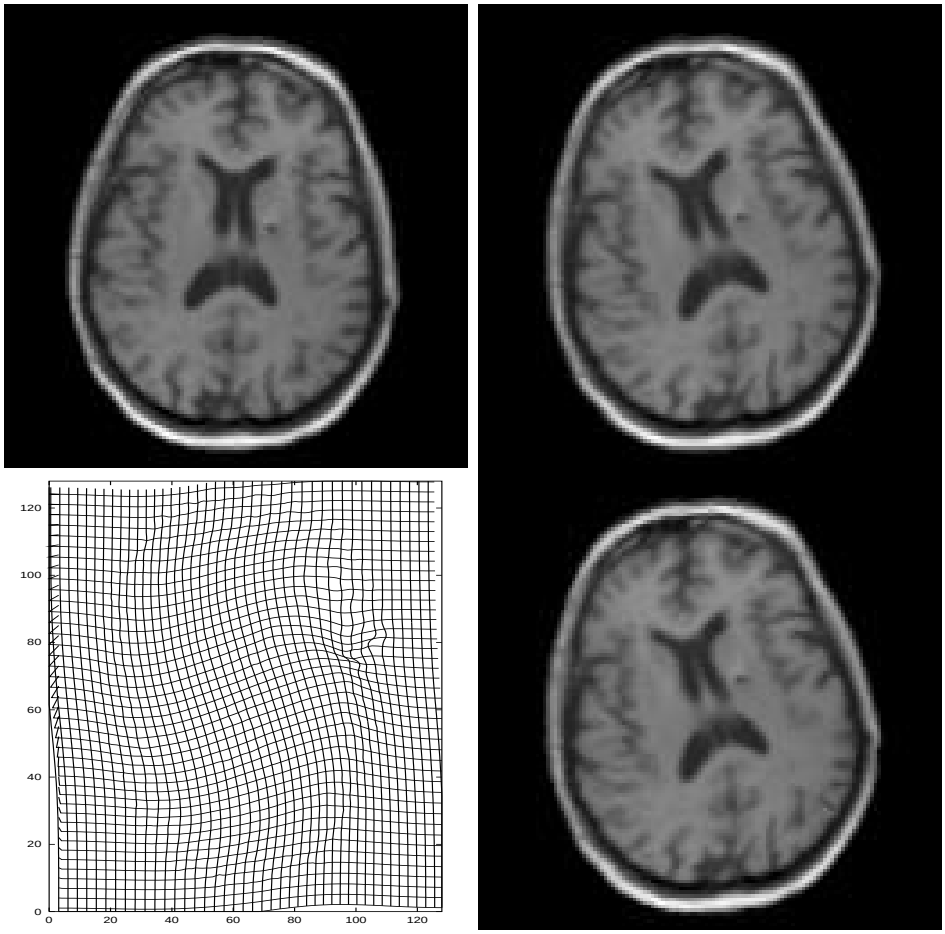

FIGURE 1. *The images show the 2D–matching result of an artificially deformed MRI-slice with* 129^2 *pixels. From top left to bottom right: original image, distorted image by rotational twist, application of the deformation to a uniform grid , matching result.*

On account of our regularity assumptions we can finish our proof of Lipschitz continuity:

$$\begin{aligned} & |\text{grad}_{L^2} E[u_1] - \text{grad}_{L^2} E[u_2]| \\ \leq \quad & ||f_1||_{C^1} \, ||f_1||_{C^{0,1}} \, |u_1 - u_2| + (||f_1||_{C^0} + |f_2|)||f_1||_{C^{1,1}} |u_1 - u_2| \end{aligned}$$

which leads to

$$\begin{aligned} & ||\text{grad}_{L^2} E[u_1] - \text{grad}_{L^2} E[u_2]||_{L^2(\Omega)} \\ \leq \quad & ||f_1||_{C^1} \, ||f_1||_{C^{0,1}} \, ||u_1 - u_2||_{L^2(\Omega)} + \\ & ||f_1||_{C^0} ||f_1||_{C^{1,1}} \, ||u_1 - u_2||_{L^2(\Omega)} + \\ & ||f_1||_{C^{1,1}} ||f_2||_{L^\infty} \, ||u_1 - u_2||_{L^2(\Omega)} \, . \end{aligned}$$

□

Let us now consider several examples:

(i) In case $A = \mathbb{1}$ existence and uniqueness are shown by the above Lemma and Theorem 3.1.

(ii) The proof of Lemma 3.2 clearly extends to $\mathcal{V} = [H^{s,2}(\Omega)]^d$, where $s \geq 0$ and $\mathcal{V}' = [H^{s,2}(\Omega)']^d$ because in this case $[H^{s,2}(\Omega)]^d \hookrightarrow [L^2(\Omega)]^d \hookrightarrow [H^{s,2}(\Omega)']^d$.

For our purpose of image matching the regularity induced by the L^2-metric will not be sufficient to obtain proper approximations of energy minimizers for our ill-posed problem w.r.t. actual applications. Thus we cannot expect to obtain smooth deformations in case $A = \mathbb{1}$ and $\mathcal{V} = \mathcal{V}' = \mathcal{W} = [L^2(\Omega)]^d$, even if we start with smooth initial deformations. Therefore we deal with spaces $\mathcal{V}$ of higher regularity and suitable operators A representing a metric:

(iii) We might choose the Helmholtz operator $A = \mathbb{1} - \frac{\sigma^2}{2}\Delta$ for $\sigma \in \mathbb{R}^+$. The metric representing A is

$$g(v, w) = (v, w)_{L^2} + \frac{\sigma^2}{2}(\nabla v, \nabla w)_{L^2} .$$

This choice corresponds to an implicit time discretization of the heat equation with time-step $\tau = \frac{\sigma^2}{2}$ and is thus related to Gaussian filtering with a filter width σ. As corresponding spaces we take into account $\mathcal{V} = [H^{1,2}(\Omega)]^d$, $\mathcal{V}' = [H^{1,2}(\Omega)']^d$ and $\mathcal{W} = [L^2(\Omega)]^d$. The isomorphism property of A and thereby the Lipschitz continuity of A^{-1} is well known in this case. Thus we have an existence and uniqueness result at hand but now with improved solution regularity.

(iv) We can further improve the regularity of the deformations. Choosing $\mathcal{V} = [H^{2,2}(\Omega)]^d$, $\mathcal{W} = [L^2(\Omega)]^d$ and $\mathcal{V}' = [H^{2,2}(\Omega)']^d$ together with the operator $A = (\mathbb{1} - \frac{\sigma^2}{2}\Delta)^2$. The corresponding metric is given by

$$g(v, w) = (v, w)_{L^2} + 2\frac{\sigma^2}{2}(\nabla v, \nabla w) + \frac{\sigma^4}{4}(\Delta v, \Delta w) ,$$

and A^{-1} is again well defined and Lipschitz continuous.

So far we have shown that using suitable metrics g one can improve the regularity of resulting deformations. Finally let us add a remark on the use of a true Gaussian filtering instead of $A = \mathbb{1} - \frac{\sigma^2}{2}\Delta$ (cf. (iii)). Consider the ODE

$$\partial_t u = -B \operatorname{grad}_{L^2} E[u] ,$$

where $B = HESG\left(\frac{\sigma^2}{2}\right)$ and $HESG$ indicates the heat equation semi-group. We look at $\mathcal{V} = C_0^m(\Omega)$ for $m \geq 0$. In this case we don't have an interpretation of this ODE as a gradient flow with respect to a norm induced by a metric. Nevertheless, we obtain C^m regular deformations for any $m \geq 0$ and finite time.

4. A scale space approach

As already stated in the introduction for typical image intensity functions f_1, f_2 the energy $E[\cdot]$ is non-convex and we expect an energy landscape with many local minima. This implies that gradient descent paths mostly tend to asymptotic states which only locally minimize the energy. Following Alvarez et al. [**2**] we consider a continuous annealing method based on a scale of image pairs $f_{1,\epsilon}$, $f_{2,\epsilon}$, where $\epsilon \geq 0$ is the scale parameter. Here we consider scale spaces of images generated by a scale space operator $S(\cdot)$ which maps an initial image f onto some coarser image, i.e.,

$$f_\epsilon = S(\epsilon) f .$$

The scale parameter ϵ allows to select fine grain representations corresponding to small values of ϵ and coarse grain representations with most of the image details skipped for larger values of ϵ. A frequently used scale space is the linear one based on Gaussian filtering. In fact we can take into account the heat equation semi group $\{HESG(\frac{\epsilon^2}{2})\}_\epsilon$ on the bounded domain Ω with imposed Neumann boundary conditions, where ϵ is the filter width parameter, i.e., $S(\epsilon) = HESG(\frac{\epsilon^2}{2})$. Here, we confine with this basic filter. Alternatively, other scale space operators such as morphological ones may be incorporated. Let us emphasize, that with respect to the final implementation we actually consider an efficient and effective approximation of this operator. Finally, we formulate the arising scale of problems: For given $\epsilon \geq 0$ we consider an energy

$$E_\epsilon[u] = \frac{1}{2}\int_\Omega |f_{1,\epsilon} \circ (\mathrm{Id} + u) - f_{2,\epsilon}|^2 \, .$$

and the corresponding gradient flow

$$\begin{aligned} g(\partial_t u_\epsilon \, , \, \theta) &= -\langle E'_\epsilon[u_\epsilon], \theta\rangle \\ u_\epsilon(0) &= u_{0,\epsilon} \, . \end{aligned}$$

We are left to choose the initial data $u_{0,\epsilon}$ for the evolution on scale ϵ. Here we expect the minimizer or a sufficiently good approximation of the same problem on a coarse scale to be a suitable starting point to approach the global minimum on the finer scale. Algorithmically, we select a sequence of scales

$$\epsilon_k = \beta_1 \, 2^{-\beta_2 k} \, , \; \beta_1, \, \beta_2 > 0 \, , \tag{4.1}$$

where $k < k_{max}$ and $\epsilon_{k_{max}} = 0$. Thus we compute discrete counterparts of the continuous solutions $u_{\epsilon_k}(T_k)$ for end times T_k sufficiently large and set

$$u_{\epsilon_k}(0) = u_{\epsilon_{k-1}}(T_{k-1}) \, .$$

For fixed β_2 the parameter β_1 is chosen such that $\epsilon_{k_{max}-1} = h$. For details we refer to Section 6.

5. Discretization

Concerning the time discretization our approach can be interpreted as a gradient flow in a Banach space. The energy functional E is Fréchet differentiable if we assume certain regularity of f_1 and f_2 (see Section 3). Therefore taking into account the energy, its Gâteaux derivative and the metric g on L^2, we are able to recall time-step controlled descent algorithms well–known in continuous optimization problems [**16**].

We will consider a time discretization as well as a spatial discretization in the following section. The spatial discretization is a standard finite element method. In addition we make use of multigrid techniques.

Time discretization. Aiming at an efficient implementation of a discrete gradient flow we apply a suitable time-step control. Thus, it pays off to consider the gradient flow perspective not only as a conceptually intuitive setting but also in the application of classical numerical tools. A time-step control strategy for the minimization of energy functionals on $\mathbb{R}^m$ turns into a time-step control for our discrete generalized gradient descent algorithm. We only have to replace the Euclidian distance in $\mathbb{R}^m$ by the norm induced by $g(\cdot,\cdot)$ on $\mathcal{V}$. We consider the explicit scheme:

$$\frac{u^{n+1} - u^n}{\tau_n} = -A^{-1}E'[u^n] \, .$$

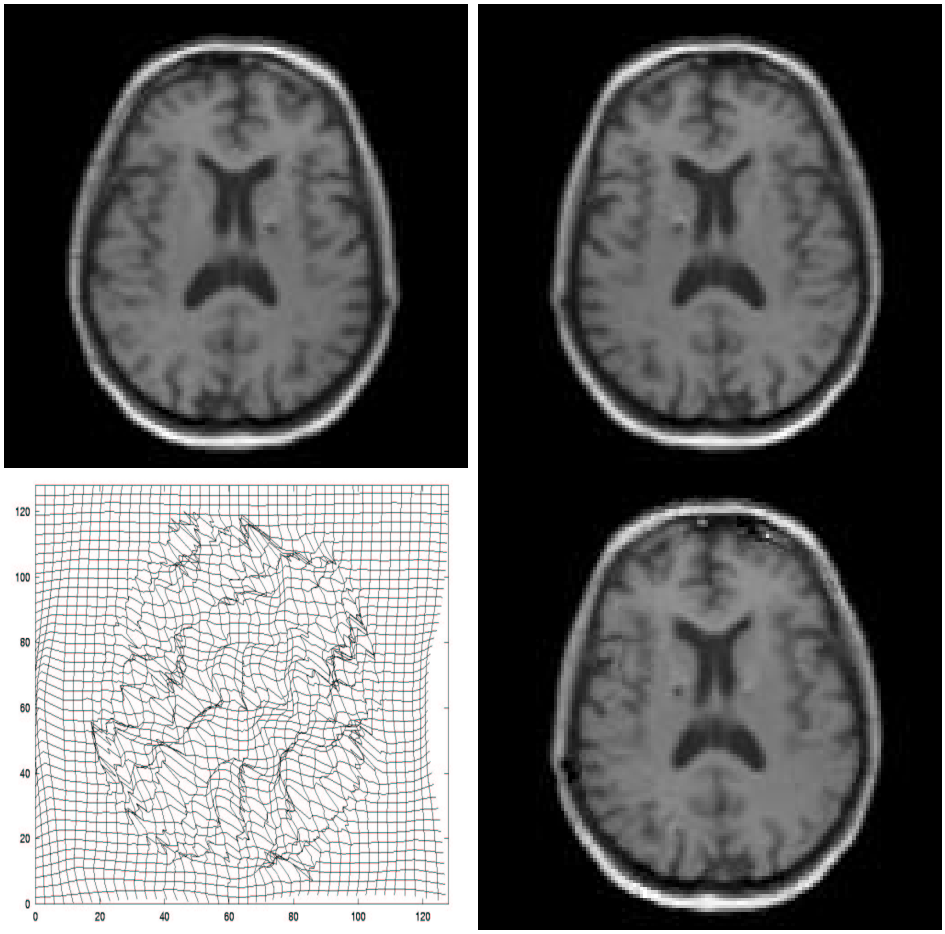

FIGURE 2. *Here the 2D–matching result of a mirrored MRI-slice with* 129^2 *pixels is shown. From top left to bottom right: original image, mirrored image, application of the deformation to a uniform grid, matching result.*

Thus we construct a sequence $(u^n)_{n=0,\dots}$, such that u^n approximates $u(t_n)$ with $t_n = \sum_{i=1}^{n} \tau_i$. The actual focus is not on the quality of the approximation but on a fast and robust descent. In our implementation we determine τ_n using Armijo's rule. As will become clear in our considerations, various other time-step control strategies can also be considered.

Armijo's rule determines each time-step τ_n by choosing τ_n such that for $\sigma \in (0, \frac{1}{2})$

$$-\frac{E[u^{n+1}] - E[u^n]}{\tau_n \langle E'[u^n], A^{-1}E'[u^n]\rangle} \geq \sigma\,.$$

Using the metric g this inequality can be expressed as

$$-\frac{E[u^{n+1}] - E[u^n]}{\tau_n g(A^{-1}E'[u^n], A^{-1}E'[u^n])} \geq \sigma\,. \tag{5.1}$$

We provide in each time-step a solution τ_n of the above inequality computing $k_n \in \mathbb{Z}$ as the smallest integer such that for some fixed $\beta \in (0,1)$

$$E[u^n] - E[u^n - \beta^{k_n} A^{-1} E'[u^n]] \geq \sigma \beta^{k_n} \|A^{-1} E'[u^n]\|_g^2 \,. \tag{5.2}$$

Here $\|\cdot\|_g$ denotes the norm induced by the metric g. It is always possible to find τ_n by that algorithm as long as u^n is not a local minimum of E. Indeed the function

$$h(t) := \frac{E[u^n] - E[u^n - tA^{-1}E'[u^n]]}{t\|A^{-1}E'[u^n]\|_g^2}$$

clearly converges to 1 if $t \to 0$ and tends (on account of the boundedness of E) to 0 if $t \to \infty$.

Spatial discretization. Now we describe the spatial discretization of equation (2.2). The set $\Omega = [0,1]^d$ is given as the union of squares or cubes E_i for i in an index set J_h. The set of elements $\{E_i\}_{i \in J_h}$ forms the mesh $\mathcal{M}_h$. Here the subscript h indicates the grid size. We confine to grids which are generated by iterated subdivision into 4 squares or 8 cubes respectively.

Thus the resulting grids form a pyramid with grid sizes $h_l = 2^{-l}$ for $l = 0, \cdots, l_{\max}$. The set of vertices of the mesh $\mathcal{M}_h$ is denoted by $\mathcal{N}_h$. Interpreting pixel or voxel values of a 2D or 3D image as nodal values we consider discrete images (F_1, F_2) as piecewise multilinear continuous functions on $\mathcal{M}_h$. The corresponding multilinear finite element space is denoted by $\mathcal{V}^h$.

We suppose $\{\Psi^i\}_{i \in I_h}$ to be the canonical nodal basis of $\mathcal{V}^h$, where I_h is the index set corresponding to $\mathcal{N}_h$. Hence we obtain $F_i = \sum_{j \in I_h} F_i^j \Psi_j$ as the representation of the image F_i in this basis, where $F_i^j = F_i(x_j)$ for the node $x_j \in \mathcal{N}_h$ corresponding to the basis function Ψ_j. Analogously, we take into account $[\mathcal{V}^h]^d$ as the set of discrete deformations. Hence the fully discrete algorithm reads as follows:

For given initial displacement U^0 find a sequence of displacements $(U^n)_n$ in $[\mathcal{V}^h]^d$ which solve

$$g_h\left(\frac{U^{n+1} - U^n}{\tau_n}, \Theta\right) = -\langle E'[U^n], \Theta \rangle \,,$$

for all test functions $\Theta \in [\mathcal{V}^h]^d$. Here the metric g_h is supposed to be a suitable approximation of the original metric g. The computation of E' induces the evaluation of $f_1 \circ \phi$. The spatial discretization of ϕ is defined on all nodes x_i and we define $f_1 \circ \phi$ as the bi- or trilinear interpolation of $(f_1 \circ \phi)(x)$ for all $x \in \mathcal{N}_h$.

Let us throughout this presentation denote the nodal vector or nodal vector dependent functional by a bar on top of the corresponding function or functional respectively. Then we can rewrite our scheme and obtain

$$A_h(\bar{U}^{n+1} - \bar{U}^n) = -\tau_n \bar{E}'[\bar{U}^n] \,,$$

or alternatively

$$\bar{U}^{n+1} = \bar{U}^n - \tau_n A_h^{-1} \bar{E}'[\bar{U}^n] \,. \tag{5.3}$$

Here $A_h = M_h G_h$, where M_h is a mass matrix – in our case the lumped mass matrix [**22**] – and G_h the matrix representation of the discrete metric with respect to the product induced by M_h, i.e.,

$$g_h(X, Y) = M_h G_h \bar{X} \cdot \bar{Y} \,.$$

This defines G_h uniquely. In the nodal basis $\{\Psi^i\}_{i\in I_h}$ we have

$$(G_h)_{ij} = (M_h)_{ii}^{-1}\, g_h(\Psi^i, \Psi^j)\,,$$

where $((M_h)_{ii})_{i\in I_h}$ is the diagonal lumped mass matrix.

Furthermore, $\bar{E}'[\bar{U}]$ is the vector of partial derivatives of $\bar{E}[\bar{U}]$ with respect to the nodal values $\bar{U}$ of U on the grid nodes.

Hence, starting with U^0 we compute a sequence of discrete displacement functions $(U^n)_{n=0,\cdots}$ approximating $u(t_n)$ for $t_n = \sum_{i=1}^{n} \tau_i$. Let us once more emphasize that our main intention is not a proper consistency with the time continuous problem. Indeed we intend a rapid energy descent along the piecewise linear path in the space $\mathcal{V}^h$. Simultaneously, we expect the timestep control to have a significant impact on the $\mathcal{V}$-norm of the solution. Hence an a priori upper bound of the time step seems feasible. An analysis of the problem has to be investigated in the future.

6. Improving the method's efficiency

In this section we will outline how the registration algorithm can be further improved concerning efficiency. The ingredients are a multigrid approximation of the smoothing operator A_h^{-1}, the numerical treatment of problems corresponding to coarse image scales on coarse grids and effective stopping criteria for the minimization procedure on coarse scales. In what follows these efficiency aspects are discussed in detail.

Multigrid. In Section 3 we showed existence as well as uniqueness for the flow $\partial_t u = -A^{-1}E'[u]$ on the space $\mathcal{V} = [H^{1,2}(\Omega)]^d$ and for $A = \mathbb{1} - \frac{\sigma^2}{2}\Delta$. The approach known to be the most efficient to solve such a linear system of equations is a multigrid method. It leads to an already optimal complexity of $O(n_h)$ if n_h is the cardinality of $\mathcal{N}_h$. But even better, already a single multigrid V–cycle is characterized by nice smoothing properties [**4, 10**] which we suppose to be the essential and sufficient property of A_h^{-1}. In fact we are not interested in any convergence but only in the smoothing properties of the multigrid cycle.

Henn and Witsch [**13**] already used this improvement in their iterative Tikhonov regularization approach.

We replace the operator A_h^{-1} in our discrete flow by the operator MGM_h representing a single multigrid V–cycle for the solution of a linear problem with the discrete operator $\mathbb{1} - \frac{\sigma^2}{2}\Delta_h$. Hence, we consider a sequence of grids $(\mathcal{M}_{h_l})_{l=0,\cdots,l_{\max}}$ with successively finer grid size h_l (e.g. $h_l = 2^{-l}$). Then the building blocks of our multigrid operator are

- on each grid $\mathcal{M}_{h_l}$ with discrete function space $\mathcal{V}^l := \mathcal{V}^{h_l}$ Jacobi iterations as smoother and
- standard prolongation and restriction operators defined on $\mathcal{V}^l$.

Finally, we are left to choose the number of pre-smoothing and post-smoothing steps in our V–cycle (cf. Fig. 3 for comparison of the resulting filter kernels). In our applications we confine with a single pre– and post-smoothing step. Table 1 lists the resulting computing times for the components of a single time-step in the discrete descent algorithm. Finally, let us underline that the discrete metric g_h is now induced by the multigrid operator MGM_h as follows: We consider MGM_h as an approximation of $\left(\mathbb{1} - \frac{\sigma^2}{2}\Delta_h\right)^{-1}$. For our discrete metric g_h this means $g_h(U,V) = MGM_h^{-1}\bar{U}\cdot\bar{V}$. Still this can be regarded as some approximation of the original metric $g(u,v) = (u,v)_{L^2} + \frac{\sigma^2}{2}\,(\nabla u, \nabla v)_{L^2}$.

Coarse scale problems on coarse grids. As the scale space operator S (cf. Section 4) actually applied in our approach we initially use an implicit step of the heat equation semigroup, i.e., we consider

$$\left(1\!\!1 + \frac{\epsilon^2}{2} M_h^{-1} L_h\right) \bar{F}_{i,\epsilon} = \bar{F}_i \,,$$

where M_h is the lumped mass matrix and L_h the usual stiffness matrix. Here ϵ plays the role of a filter width parameter. Again, with respect to an improved efficiency, we replace the exact linear solver for this system by the corresponding multigrid V–cycle, now acting on images and not on displacement descent directions (cf. Section 3 example (iii)). Let us recall that the scale parameter is chosen as in (4.1). Now we couple the sequence of meshes $\mathcal{M}_{h_l}$ and corresponding function spaces $\mathcal{V}^l$ on the one hand and the sequence of scales on the other hand. Thus we restrict on scale k the whole problem to grid level $l(k)$. In the case where S corresponds to the heat equation, a suitable choice for $l(k)$ would be the smallest integer such that

$$h_{l(k)} \leq \alpha \epsilon_k$$

for some constant $\alpha > 0$. In the concrete applications we have chosen the parameter $\alpha \in [\frac{1}{4}, 2]$.

Effectively coupling scales. Furthermore it is not required to reach a local minimum of the corresponding energy E_{ϵ_k} by the discrete gradient descent on level k. Expressed in formal geometric terms it is sufficient if for some integer n_k a discrete displacement $U_k^{n_k}$ of the discrete gradient descent sequence $(U_k^n)_{n=0,\dots}$ on scale k enters the attractive region of the global minimum of the energy $E_{\epsilon_{k+1}}$ on the next finer scale $k+1$. By construction in the $k_{\max}$-th step we would then end up in the contraction region of the actual energy E. Unfortunately, the above condition is rather implicit.

Hence, we confine with a heuristic stopping criterion for the discrete gradient descent on scale k, i.e., we turn to the next finer scale, if for some fixed constant γ we observe that for some norm $||\cdot||$

$$\|U_k^{n+1} - U_k^n\| \leq \gamma\, \epsilon_k$$

holds. Actually, in the application we consider $\gamma = \frac{1}{2}$ and evaluate the L^2-norm of the displacement update.

If the iteration is finished on a certain scale k a regularization is applied to the deformation obtained on this scale before starting the iteration on the next scale. (We use our multigrid regularization approximating $HESG(\frac{\epsilon_k^2}{2})$.)

Based on this multilevel approach, the overall cost of the multiscale method is optimal. Here by optimal we mean that the number of considered scales does not contribute significantly to the overall cost. Indeed the overall cost reduces to the cost for the gradient descent on the finest grid and the finest scale times a constant C. For $\alpha = 1$ and an equal number of gradient descent iterations on all scales the overall cost geometrically decays with decreasing scale and grid level respectively, i.e., $C = 4/3$ in 2D and $C = 8/7$ in 3D. Due to our adaptive stopping criteria the actual factor for the offset cost for solving a multiscale problem is even smaller (cf. Table 2 which lists the required number of iterations and the computing times on every scale for an application problem in 3D).

7. Applications

Let us now collect numerical results in 2D and 3D for some test cases. In two dimensions it is still feasable to solve the resulting linear systems for $A = \mathbb{1} - \frac{\sigma^2}{2}\Delta$ by CG iterations. In three dimensions we define A^{-1} by one multigrid cycle related to the operator $A = \mathbb{1} - \frac{\sigma^2}{2}\Delta$ (cf. Section 6). We applied the presented algorithm on synthetic problems with large deformations as well as on medical MR-images which also lead to a non-local matching problem. In Figures 1, 2, and 4 2D-results for the matching of an artificially twisted resp. mirrored brain with the original image are depicted.

Figures 7 and 5 show the 3D-matching of a strongly rotated synthetic image and a reflected MR-image versus the corresponding original. All Figures corresponding to results in 3D show planar slices through the 3D volume.

The synthetically generated matching problem should demonstrate an important advantage of the multiscale approach, namely the capability to handle the registration of heavily distorted images, where the distortion itself is additionally highly non–rigid and non–local.

Naturally, the applicability to medical images is of fundamental importance for the evaluation of the method. Although we have confined to the most simple matching energy for a starting point, we wanted to get some insight on the fundamental behaviour of the gradient descent on realistic MR-images. Due to the fact, that both hemispheres of a healthy brain have – apart from minor geometrical differences – the same fundamental structure, a reflection provides a useful and solvable test example and is comparable to a matching problem from a patient to a reference image from an atlas. Thus, our aim is to find the displacement which describes both hemispheres given the corresponding other hemispheres and not to find the global minimum, which would be the reflection itself. This is naturally ruled out by the regularization and the gradient descent approach. The method is capable to compute a rather regular approximation of a local minimum, but with convincing coherence of the deformed first image and the initial second one, which can be seen by comparing the sulci of the cortex on the reference image with those of the matching result. Due to its high contrast to the surrounding tissue, the ventricles are perfectly matched as well. In contrast to the synthetic example, the resulting matching deformation varies locally in magnitude because some regions match initially quite well while others have to be deformed quite drastically. It turns out in the experiments, that the choice of σ in the metric operator A should not be too large in order not to destroy local variations in the deformation.

Process	**Duration**
V-cycle (single component)	3.3s
computation of $E[u]$ and grad $E[u]$	5.25s
computation of $\langle E'[u], \phi \rangle$	5.38s
computation of $E[u]$	1.23s
time–step control	1-3s

TABLE 1. *Approximate computing times for the key ingredients of each gradient descent step in our algorithm on a reference PC (Pentium IV, 1.7 Ghz, 1Gb RAM) applied on 3D images with* 129^3 *voxels.*

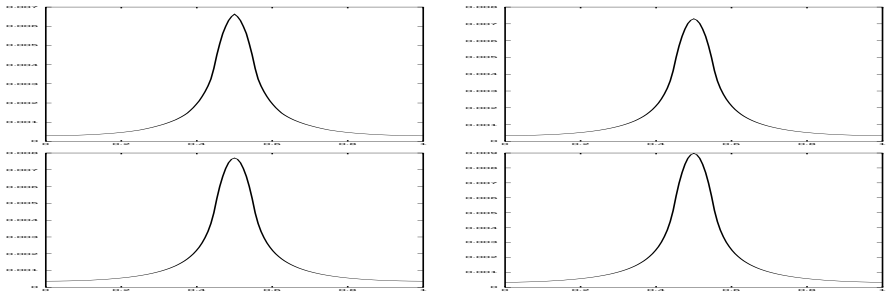

FIGURE 3. *Profiles through a three dimensional data set after the application of one multigrid cycle for the solution of the discrete heat equation with corresponding filter width $\sigma = 0.1$ and a discrete centered Dirac distribution as initial image. We applied 1 through 3 pre- and post-smoothing steps. These correspond to the actually applied kernels in a multigrid smoothing cycle. The bottom right image shows the profile corresponding to the exact discrete solution. One can clearly observe, that the overall shape varies only very slightly when the number of smoothing steps is changed.*

scale k	**filter width** ϵ_k	**steps** n_k	**grid level** $l(k)$	**time**
0	.250	5	5	<1s
1	.177	3	5	<1s
2	.125	3	5	<1s
3	.088	4	6	9s
4	.062	3	6	7s
5	.044	4	7	67s
6	.031	6	7	95s
7	.022	6	7	96s
8	.016	5	7	82s
9	.0	5	7	83s

TABLE 2. *Iteration counts on different scales due to the adaptive stopping criteria and the corresponding absolute timings for the computation on the correspondingly chosen grid levels. Here we assume $\alpha = \frac{1}{4}$,$\gamma = \frac{1}{2}$.*

8. Conclusion and Outlook

We have presented an efficient and robust registration method which is capable to compute large deformations on fine two and three dimensional grids. Here the focus is not on the generality of the presented method but on the acceleration potential based on the gradient flow perspective and the multigrid and multiscale approach for this class of optimization problems. Computations on medical images have pointed out the method's applicability and the quality of the achievable results. Nevertheless let us emphasize that the method is restricted to intensity matching applications. It is applicable to images of the same modalitiy for example to register a medical MR-image of a patient to a equally weighted reference image in a medical atlas library or to describe temporal deformations

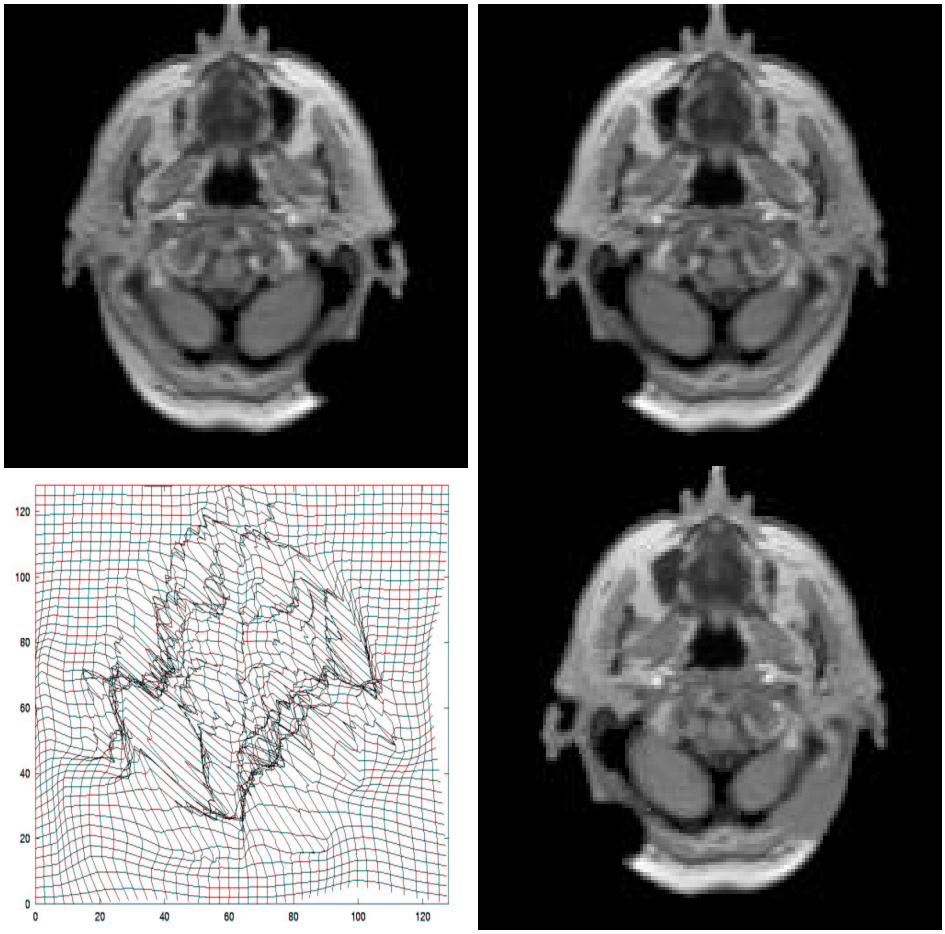

FIGURE 4. *Due to the multiscale approach, the method is capable to generate relatively large displacement fields, here depicted for a 2D example. From top left to bottom right: original image, plane–mirrored image, application of the deformation to a uniform grid, matching result.*

of subsequently aquired images of the same patient. It is however not capable to correlate image morphologies, i. e. it is not grey scale invariant [**1**], which is important for the registration of CT, MRI, PET and ultrasound datasets for example. Instead of matching image intensity one may consider image morphologies only and try to match them between images of different modality or different time steps from a sequence of images. The morphologies are characterized uniquely by the entity of level sets and their Gauss maps respectively. Hence, we will investigate a cost functional to be minimized which measures the effect of the deformation on the image Gauss maps instead of the image intensities. Furthermore in certain applications it turns out to be useful (cf. Section 1) to consider a non–homogeneous metric on the space of deformations depending on the images and image features. Furthermore constraints on the velocity fields such as a vanishing divergence – which ensures volume preservation of the resulting deformation at least in the time continuous case – can be included in the approach via a projection method.

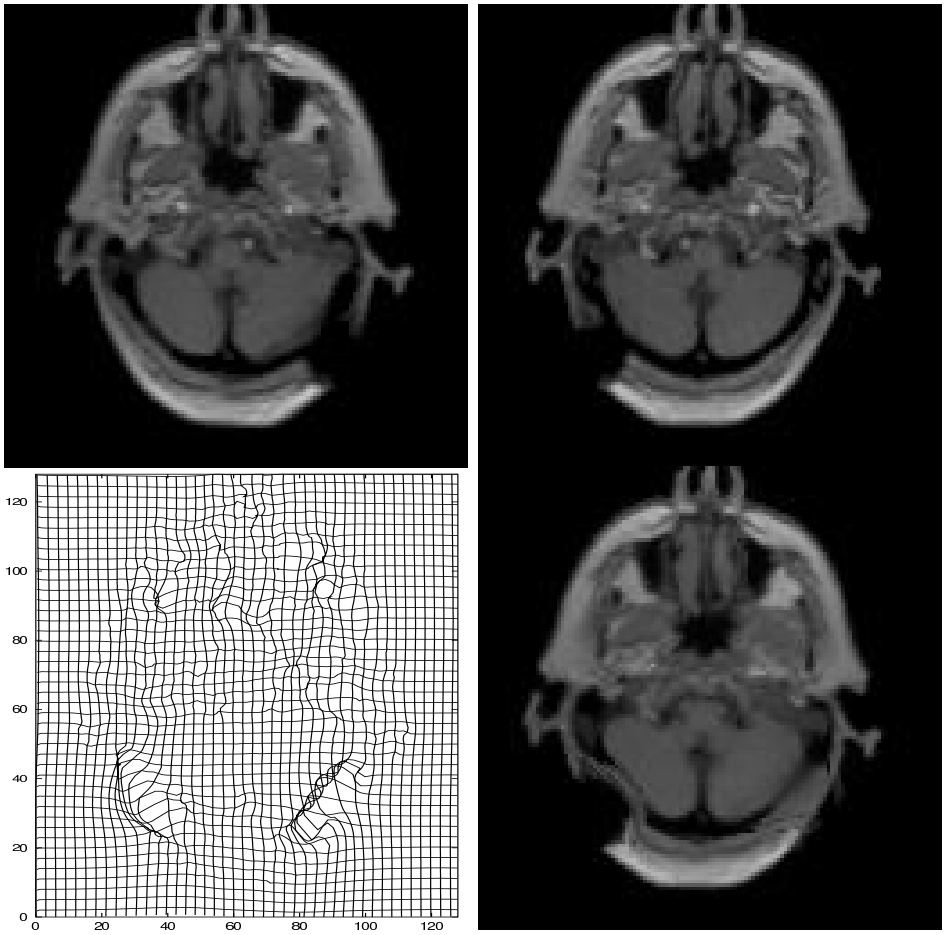

FIGURE 5. *In this 3D–matching example the second image is generated from the first by reflecting the original at a central mirror plane. Thus the matching process has to cope with locally large deformations. From top left to bottom right: axial slice through the original 3D image, second image generated by reflection, deformation applied to a uniform grid, matching result.*

Acknowledgements

We thank Folkmar Bornemann for giving us insight on a different topic of image smoothing and image segmentation which nevertheless mainly inspired the presented approach to the coupling of scales and the corresponding stopping criteria, Joachim Weickert for interesting discussions on optical flow problems and Rainer Lachner from the BrainLAB AG, Munich for providing some of the test cases.

This work is partially funded by the Deutsche Forschungsgemeinschaft (DFG) – Programme SPP 1114, Mathematical methods for time series analysis and digital image processing.

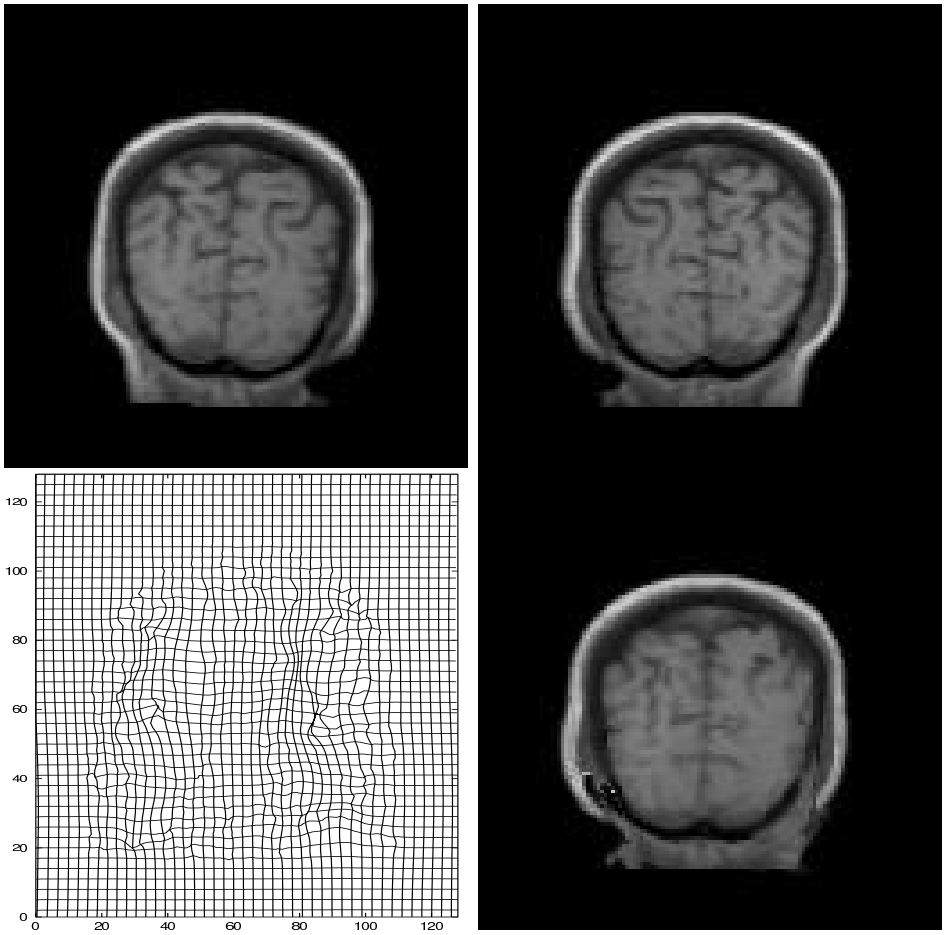

FIGURE 6. *From top left to bottom right: saggital slice through the original 3D image, second image generated by reflection, deformation applied to a uniform grid, matching result.*

References

[1] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. Axioms and fundamental equations of image processing. *Arch. Ration. Mech. Anal.*, 123:199–257, 1993.

[2] L. Alvarez, J. Weickert, and J. Sánchez. A scale–space approach to nonlocal optical flow calculations. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision. Second International Conference, Scale-Space '99, Corfu, Greece, September 1999*, Lecture Notes in Computer Science; 1682, pages 235–246. Springer, 1999.

[3] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39:41–56, 2000.

[4] F. Bornemann and P. Deuflhard. The cascadic multigrid method for elliptic problems. *Num. Math.*, 75(2):135–152, 1996.

[5] G. E. Christensen, S. C. Joshi, and M. I. Miller. Volumetric transformations of brain anatomy. *IEEE Trans. Medical Imaging*, 16, no. 6:864–877, 1997.

[6] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Trans. Medical Imaging*, 5, no. 10:1435–1447, 1996.

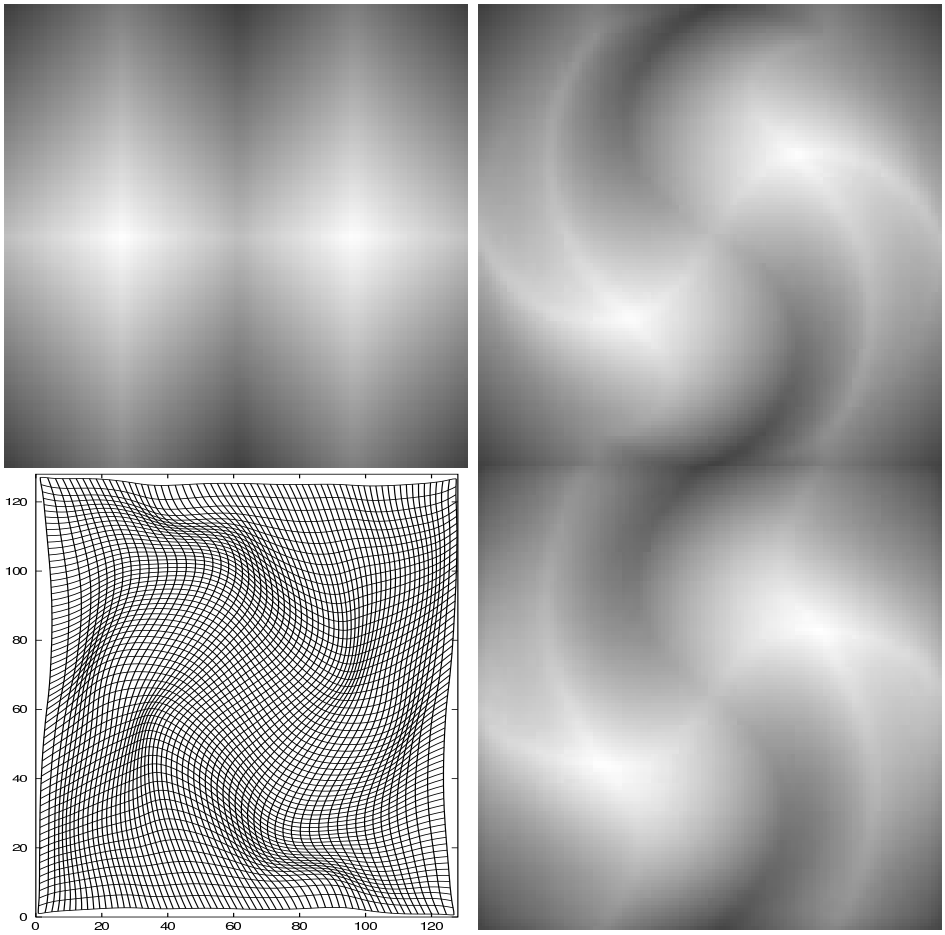

FIGURE 7. *The images show the 3D–matching results of a synthetically generated problem (rotational twist by $\frac{\pi}{3}$) with resulting rather large deformations. From top left to bottom right: slice through the original volume image, second image generated via artificial deformation of original image, deformation applied to a uniform grid, matching result.*

[7] C. A. Davatzikos, R. N. Bryan, and J. L. Prince. Image registration based on boundary mapping. *IEEE Trans. Medical Imaging*, 15, no. 1:112–115, 1996.

[8] R. Deriche, P. Kornobst, and G. Aubert. Optical–flow estimation while preserving its discontinuities: A variational approach. In *Proc. Second Asian Conf. Computer Vision (ACCV '95, Singapore, December 5–8, 1995)*, volume 2, pages 290–295, 1995.

[9] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly Appl. Math.*, LVI, no. 4:617–694, 1998.

[10] W. Hackbusch. *Multigrid Methods and Applications*. Springer, Berlin/Heidelberg, 1985.

[11] W. Hackbusch. *Iterative solution of large sparse systems*. Springer, Berlin, 1994.

[12] M. Hanke and C. Groetsch. Nonstationary iterated tikhonov regularization. *J. Optim. Theory and Applications*, 98:37–53, 1998.

[13] S. Henn and K. Witsch. Iterative multigrid regularization techniques for image matching. *SIAM J. Sci. Comput. (SISC)*, Vol. 23 no. 4:pp. 1077–1093, 2001.

[14] W. Hinterberger, O. Scherzer, C. Schnörr, and J. Weickert. Analysis of optical flow models in the framework of calculus of variations. Technical Report 8, Computer Science Series, University Mannheim, 2001.
[15] S. C. Joshi and M. I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Medical Imaging*, 9, no. 8:1357–1370, 2000.
[16] P. Kosmol. *Optimierung und Approximation*. de Gruyter Lehrbuch, 1991.
[17] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi–modal volume registration by maximization of mutual information. *IEEE Trans. Medical Imaging*, 16, no. 7:187–198, 1997.
[18] H. H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:565–593, 1986.
[19] E. Radmoser, O. Scherzer, and J. Weickert. Scale–space properties of regularization methods. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision. Second International Conference, Scale-Space '99, Corfu, Greece, September 1999*, Lecture Notes in Computer Science; 1682, pages 211–220. Springer, 1999.
[20] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering, 1998.
[21] J. P. Thirion. Image matching as a diffusion process: An analogy with maxwell's demon. *Medical Imag. Analysis*, 2:243–260, 1998.
[22] V. Thomée. *Galerkin - Finite Element Methods for Parabolic Problems*. Springer, 1984.
[23] P. A. Viola. Alignment by maximization of mutual information. Technical Report AITR-1548, 1995.
[24] J. Weickert. *Anisotropic diffusion in image processing*. Teubner, 1998.
[25] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information, 1996.

GERHARD-MERCATOR-UNIVERSITÄT DUISBURG, LOTHARSTRASSE 65, 47048 DUISBURG
E-mail address: [clarenz, droske, rumpf]@math.uni-duisburg.de