Dense Image Registration through MRFs and Efficient Linear Programming

Ben Glocker [a,b,*,1], Nikos Komodakis [a,c], Georgios Tziritas [c], Nassir Navab [b], Nikos Paragios [a]

[a] *GALEN Group, Mathématiques Appliquées aux Systèmes (MAS)*
Ecole Centrale Paris, France

[b] *Computer Aided Medical Procedures (CAMP)*
Technische Universität München, Germany

[c] *Computer Science Department*
University of Crete, Greece

Abstract

In this paper, we introduce a novel and efficient approach to dense image registration, which does not require a derivative of the employed cost function. In such a context the registration problem is formulated using a discrete Markov Random Field objective function. First, towards dimensionality reduction on the variables we assume that the dense deformation field can be expressed using a small number of control points (registration grid) and an interpolation strategy. Then, the registration cost is expressed using a discrete sum over image costs (using an arbitrary similarity measure) projected on the control points, and a smoothness term that penalizes local deviations on the deformation field according to a neighborhood system on the grid. Towards a discrete approach the search space is quantized resulting in a fully discrete model. In order to account for large deformations and produce results on a high resolution level a multi-scale incremental approach is considered where the optimal solution is iteratively updated. This is done through successive morphings of the source towards the target image. Efficient linear programming using the primal dual principles is considered to recover the lowest potential of the cost function. Very promising results using synthetic data with known deformations and real data demonstrate the potentials of our approach.

Key words: Discrete Optimization, Deformable Registration, Linear Programming

1 Introduction

Medical image analysis [7] is an established domain in computational, mathematical and biological sciences. Recent advances on the acquisition side have made possible the visualization of human tissues as well as physiological and pathological indices related to them either occasionally or periodically. The ability to compare or fuse information across subjects with origins of different modalities is a critical and necessary component of computer aided diagnosis. The term used often to express this need is registration.

The registration problem often involves three aspects, (i) the transformation model, (ii) a similarity criterion and (iii) an optimization strategy.

Registration can be either global or local. Parametric models are often employed to address global registration with a small number of degrees of freedom, such as rigid or similarity. These models refer to a good compromise between performance and computational complexity. Furthermore, the registration problem in such context is well posed since the number of variables to be determined is over-constrained from the number of observations. Dense image registration aims to go further and seeks individual correspondences between observations. The main goal is to determine relationships that locally express the correlation of the observations either for the same subject (acquisitions of different modalities or acquisitions of the same organ in time). Local alignment or dense/deformable registration are the terms often considered to describe this task.

Deformable registration is one of the most challenging problems in medical imaging. The problem consists of recovering a local transformation that aligns two signals that have in general an unknown relationship both in the spatial and intensity domain. Several methods exist in the literature where specific measures are designed to account for this relationship and optimize the transformation that brings together these two signals.

Local image alignment is often performed according to geometric or photometric criteria. Landmark-based methods [16,29] are a classic example of geometric-driven registration. In such a setting, a number of anatomical key points [26]/structures (segmented values) are identified both in the source and the target image and a transformation that aims to minimize the Euclidean distance between these structures is to be recovered. The main limitation of

* Corresponding author.
 Email addresses: `glocker@in.tum.de` (Ben Glocker), `komod@csd.uoc.gr` (Nikos Komodakis), `tziritas@csd.uoc.gr` (Georgios Tziritas), `navab@in.tum.de` (Nassir Navab), `nikos.paragios@ecp.fr` (Nikos Paragios).

these methods is related to the selection and extraction of landmarks, while their main strength is the simplicity of the optimization process.

Iconic registration methods [3] seek for "visual" correspondences between the source and the target image. Such a problem is tractable when one seeks registration for images from the same modality due to an explicit photometric correspondence of the image intensities. Sum of squared differences [15], sum of absolute differences [15], cross correlation [15] or distances on subspaces that involve both appearance and geometry (intensities, curvature, higher order image moments) [6] have been considered. On the other hand it becomes more challenging when seeking transformations between different modalities with a non-linear or only statistical relation of intensities. Such measures have often been used [17] where normalized mutual information [25], Kullback-Liebler divergence [37] and correlation ratio [27] are some of the measures used to define similarity[2] between different modalities.

Once the similarity measure has been defined the next task consists of recovering the parameters that optimize the designed cost function. Parameters can be either searched or estimated. In the first case techniques like exhaustive search can be employed which are time consuming. On the other hand, one can use known optimization techniques, gradient-free or gradient-based to determine the optimal set of parameters starting from an initial guess [19]. These methods require an important customization from one application to another since a correlation exists between the modalities/problem and the selection of the similarity measure. Furthermore, the optimization is often sub-optimal due the non-convexity of the designed cost functions. In particular when considering complex similarity functions defined on the continuous space, then the numerical approximation of the gradient in the discrete domain (image/volume plane) is very challenging leading to erroneous registration results.

The aim of our approach is to overcome both limitations present in all registration methods. Dependency on the similarity measure selection, as well as on the initial conditions in a reasonable computation time.

In this article we propose a novel technique that can either be used for inter or intra modal image registration. Towards satisfying smoothness of the deformation field and reducing the dimensionality of the problem we represent deformation through Free Form Deformations [31]. Our method reformulates registration as an Markov Random Field (MRF) optimization where a set of labels is associated with a set of deformations, and one seeks to attribute a label to each control point such that once the corresponding deformation has been applied, the similarity measure between the source and the target is optimal for all voxels. The optimization procedure is independent from the graph

---

[2] For consistency reasons we always use the term similarity measure, although measures such as sum of squared differences are actually dissimilarity measures.

3

construction, and therefore any similarity measure can be used.

The remainder of this paper is organized as follows: In Section 2 we introduce the proposed registration framework, while in Section 3 we discuss the optimization aspects. Implementation details are given in Section 4 and experimental validation are part of Section 5. Section 6 concludes our paper.

2   Deformable Registration

In order to introduce the concept of our approach [14], we consider (without loss of generality) the 2D image domain. Let us consider a source $f : \Omega = [1, N] \times [1, M] \to \mathcal{R}$ and a target image g. In general, these images are related with a non-linear transformation as well as a non-linear relation between intensities, that is

$$\forall \mathbf{x} \in \Omega \quad g(\mathbf{x}) = h \circ f(\mathcal{T}(\mathbf{x})) \tag{1}$$

where $\mathcal{T}(\mathbf{x})$ is the transformation and h is a non-linear operator explaining the changes of appearance between them. The most common way to formulate the registration problem, is through the definition of a distance between the source and the target image that is to be minimized in the entire domain Ω, or

$$E_{\text{data}}(\mathcal{T}) = \int_{\Omega} |g(\mathbf{x}) - h \circ f(\mathcal{T}(\mathbf{x}))| d\mathbf{x}. \tag{2}$$

Recovering the optimal potential of this objective function is not straightforward. In the case of 2D images, two variables are to be determined while one constraint is available per pixel. The most basic approach to address this limitation is through the use of a regularization function on the space of unknown variables [34], or

$$E_{\text{smooth}}(\mathcal{T}) = \int_{\Omega} \phi(\nabla \mathcal{T}(\mathbf{x})) d\mathbf{x} \tag{3}$$

with ϕ being a convex function imposing smoothness on the deformation field for neighboring pixels. Such a term will make the estimation of the deformation field feasible assuming that the relationship between the signal intensities is known. This hypothesis is not realistic due to the fact that (i) when registering the same modalities this relationship depends on the parameters of the scanner which are not available, (ii) when registering different modalities in most of the cases such an operator does not exist.

In order to overcome this constraint, in the most general case a similarity measure ρ_h is introduced to account for the intensity relation between the two images, or

$$E_{\text{data}}(\mathcal{T}) = \int_{\Omega} \rho_h(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x}))) d\mathbf{x} \tag{4}$$

The definition of the ρ_h depends on the nature of the observed signals as well

4

as the application itself. Once this measure is defined the data term is combined with the smoothness one to determine the objective/cost function under consideration. Gradient-descent is the most common approach to perform the optimization, a method that has some strengths and known limitations. One can claim that this approach is convenient and often it is straightforward to implement. On the other hand, the problem is ill-posed due to the fact that the number of constraints is inferior to the number of variables to be determined. Furthermore, since the cost function is non-convex one cannot guarantee that the obtained solution will be the optimal one. Last, but not least gradient numerical manipulation is not straightforward when projecting from the continuous space to the discrete one.

The above observations lead to a natural conclusion that one should seek (i) dimensionality reduction on the degrees of freedom of the model, (ii) more efficient optimization techniques both in terms of ability to approach the optimal solution with reasonable computational cost, and (iii) techniques that do not require continuous gradient manipulation in discrete spaces.

2.1 Continuous Domain

Since we are interested in local registration, let us introduce a deformation grid $G : [1, K] \times [1, L]$ (usually $K \ll M$ and $L \ll N$) superimposed onto the image (no particular assumption is made on the grid resolution). The central idea of our approach is to deform the grid (with a 2D displacement vector $\mathbf{d_p}$ for each control point) such that the underlying image structures are perfectly aligned. One can assume that the transformation of an image pixel $\mathbf{x}$ can be expressed using a linear or non-linear combination of the grid points, or

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathcal{D}(\mathbf{x}) \quad \text{with} \quad \mathcal{D}(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|) \, \mathbf{d_p} \tag{5}$$

where $\eta(\cdot)$ is the weighting function measuring the contribution of the control point $\mathbf{p}$ to the displacement field $\mathcal{D}$. In such a theoretical setting without loss of generality we consider Free Form Deformations (FFD) based on cubic B-Splines as a transformation model [33]. FFD are successfully applied in non-rigid image registration [24,28,31,32]. Deformation of an object is achieved by manipulating an underlying mesh of uniformly spaced control points. The displacement field for a two-dimensional FFD based on cubic B-Splines at position $\mathbf{x} = (x, y)$ is defined as

$$\mathcal{D}(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} B_l(u) B_m(v) \, \mathbf{d}_{i+l,j+m} \tag{6}$$

5

where $i = \lfloor x/\delta_x \rfloor - 1$, $j = \lfloor y/\delta_y \rfloor - 1$, $u = x/\delta_x - \lfloor x/\delta_x \rfloor$, and $v = y/\delta_y - \lfloor y/\delta_y \rfloor$ where B_l represents the lth basis function of the B-Spline, and $\delta_x = \frac{M}{K-1}$, $\delta_y = \frac{N}{L-1}$ denotes the control point spacing. The three-dimensional version is defined in a straightforward manner.

In order to pose an optimization problem based on such a deformation model, we also have to define a function which allows us projecting information from the image level to the level of control points. This can be seen as a kind of inverse function of $\eta(\cdot)$, which can be defined as

$$\hat{\eta}(|\mathbf{x} - \mathbf{p}|) = \frac{\eta(|\mathbf{x} - \mathbf{p}|)}{\int_\Omega \eta(|\mathbf{y} - \mathbf{p}|)d\mathbf{y}}. \tag{7}$$

where $\hat{\eta}(\cdot)$ computes the influence of an image point $\mathbf{x}$ to a control point $\mathbf{p}$. This is very similar to the term that occurs also in gradient-descent based approaches where the force field based on the derivatives of the similarity measure is also projected to the control points. However, in our case we can simply plug-in this term into the criterion earlier introduced in (4), or

$$E_{\text{data}}(\mathcal{T}) = \frac{1}{|G|} \sum_{\mathbf{p} \in G} \int_\Omega \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \cdot \rho_h(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x})))d\mathbf{x}. \tag{8}$$

Such a term will guarantee photometric correspondence between the two images where the similarity measure is evaluated on the image level but represented on the level of control points. This is an important definition for our framework since it allows us reducing the dimensionality of the dense registration problem. Later, we will see that this exactly the data term used for optimization, so no differentiation of the similarity measure has to be performed.

Obviously, the definition of our data term (8) is only valid for point-wise similarity measures (e.g. sum of squared differences). More complex and statistical measures (e.g cross correlation or mutual information) have to be computed slightly different. First we define another version of $\hat{\eta}$ as

$$\hat{\eta}(|\mathbf{x} - \mathbf{p}|) = \begin{cases} 1, & \text{if } \eta(|\mathbf{x} - \mathbf{p}|) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

Basically, this will mask pixels influenced by a control point $\mathbf{p}$ resulting in a local image patch centered at the control point. From this patch the similarity measure (e.g. cross correlation) is then computed. This imposes currently some limitations on the resolutions of the transformation grid. Since too many grid points will result in too less samples in the local image patches. However, in our experiments we show that in practice this effect does not play a crucial role and statistical measures yield very good registration results.

The transformation due to the interpolation inherits some implicit smoothness properties. However, in order to achieve smooth results also in texture-less or noisy regions, one can consider a smoothness term on the grid domain, or

$$E_{\text{smooth}}(\mathcal{T}) = \frac{1}{|G|} \sum_{\mathbf{p} \in G} \phi(|\nabla_G \, \mathbf{d}_p|) \tag{10}$$

with ϕ being a smoothness penalty function for instance the L_1-Norm. The complete term associated with the registration problem is then defined as the sum of the data and smoothness term, or

$$E_{\text{total}} = E_{\text{data}} + E_{\text{smooth}}. \tag{11}$$

The most common way to obtain the transformation parameters is through the use of a gradient-descent method in an iterative approach [31]. Thus given an initial guess, one updates the estimate according to the following formula $\left[\mathcal{T}^m = \mathcal{T}^{m-1} - \delta t \frac{\delta E_{\text{total}}}{\partial \mathcal{T}} \right]$. Such a process involves the derivative of the energy term with respect to the transformation parameters and therefore it is model and criterion dependent. Slight modifications of the cost function could lead to a different derivative and require novel numerical approximation methods.

2.2 Discrete Domain

Let us now consider a discrete set of labels $L = \{l^1, ..., l^i\}$ corresponding to a quantized version of the deformation space $\Theta = \{\mathbf{d}^1, ..., \mathbf{d}^i\}$. A label assignment $l_{\mathbf{p}}$ to a grid node $\mathbf{p}$ is associated with displacing the node by the corresponding vector $\mathbf{d}^{l_{\mathbf{p}}}$. If a label is assigned to every node we get a discrete labeling $\mathbf{l}$. The displacement field associated with a certain labeling $\mathbf{l}$ becomes

$$\mathcal{D}(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|) \, \mathbf{d}^{l_{\mathbf{p}}}. \tag{12}$$

2.2.1 Data Term Approximation

One can reformulate the registration as a discrete multi-labeling problem, that is assign individual labels $l_{\mathbf{p}}$ to the grid nodes. A common model for representing such problems are Markov Random Fields (MRFs) [23]. The general form of a first-order MRF is

$$E_{\text{MRF}}(\mathbf{l}) = \sum_{\mathbf{p} \in G} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{\mathbf{p} \in G} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \tag{13}$$

where $V_{\mathbf{p}}(\cdot)$ are the unary potentials representing the data term, $V_{\mathbf{pq}}(\cdot, \cdot)$ are the pairwise potentials representing the smoothness term, and $\mathcal{N}$ represents

the neighborhood system of the nodes. We define the unary potentials according to our data term as

$$V_{\mathbf{p}}(l_{\mathbf{p}}) \approx \int_{\Omega} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \cdot \rho_h(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x})))d\mathbf{x} \qquad (14)$$

which can then be seen as local evaluations of the similarity measure. In general the unary potentials are assumed to be independent [23] which is for our application in most of the cases not true according to the influence function $\eta(\cdot)$. Naturally, neighboring control points influence in overlapping areas of the dense displacement field in a linear or non-linear manner. So the potential local similarity caused by displacing a control point can only be approximated. The actual similarity is first known after applying the resulting labeling, i.e. a morphing of the source image. Our approximation scheme for computing the values of $V_{\mathbf{p}}(l_{\mathbf{p}})$ is sketched in pseudo-code in the following:

```
1 for each label α ∈ L
2     for each node p ∈ G
3         V_p(α) = ∫_Ω η̂(|x − p|) · ρ_h(g(x), f(x + d^α))dx
4     end
5 end
```

This scheme can be implemented very efficiently, since the operations that have to be done on the source image f in line 3 are the same for all nodes and basically reduce to a simple translation of f by $\mathbf{d}^{\alpha}$. The approximation error for the potential similarity can be reduced by using only linear weighting functions in $\hat{\eta}(\cdot)$ (while keeping the cubic functions for the smooth transformation). Additionally, the use of a multi-scale incremental approach of successive morphings of the source towards the target image improves the approximation over time.

2.2.2   Incremental Approach and Diffeomorphisms

The incremental approach has another advantage. The number of labels and their capture range play a significant role to the registration process. It is clear that setting the number of labels to infinity will converge to the continuous formulation which though is intractable from computational perspective. On the other side, if the set of labels is too small or misses important displacements the registration process can yield poor results. Therefore we propose to perform several optimization cycles (while resetting the control grid and composing the dense deformation fields on the image level). After each cycle the capture range covered by the label set is refined by a certain scaling factor which enables high accuracy results while boosting the performance of the optimization through small sets of labels. To this end, we can define a series

of cost functions where the data term is computed on

$$V_{\mathbf{p}}^t(\alpha) = \int_\Omega \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \cdot \rho_h(g(\mathbf{x}), f(\mathbf{d}^\alpha + \mathcal{T}^{t-1}(\mathbf{x})))d\mathbf{x} \qquad (15)$$

Recently, one certain type of deformations gained quite a lot of interest. In some applications, e.g. where the deformation field itself is further analyzed or foldings have to avoided, it is desirable to obtain smooth, invertible deformations called diffeomorphisms. Following [30], it is very easy to guarantee diffeomorphic deformations by hard constraints. Since the space of solutions is controlled through the definition of the label set, we can simply restrict the maximum displacement to be 0.4 times the control point spacing [4]. Thus, every morphing will fulfill the diffeomorphic properties, and since we compose the single morphings on the image level and the composition of two diffeomorphisms produces a diffeomorphism, our final solutions are diffeomorphisms as well.

2.2.3 Smoothness Term

The next aspect to be addressed is the definition of the smoothness term $V_{\mathbf{pq}}(\cdot, \cdot)$ in the label domain. A simple smoothness term can be defined as a distance function computing the magnitude of vector differences [14], or

$$V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) = \lambda_{\mathbf{pq}} \left| \mathbf{d}^{l_{\mathbf{p}}} - \mathbf{d}^{l_{\mathbf{q}}} \right| \qquad (16)$$

where $\lambda_{\mathbf{pq}}$ plays the role of a weighting factor which may vary over the spatial domain. The value of the weighting factor is depending on the application and usually it has to be adapted according to the similarity function. Other distance functions can be considered as well, e.g. a quadratic term or truncated terms resulting in piecewise smooth functions. Using the function in (16) as a smoothness term for the registration problem results in a fluid-like registration [10]. This is because only the incremental updates of the deformation field are penalized. If we want to perform a full regularization over time we also have to consider the deformation field from the previous iterations within our distance function, or

$$V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) = \lambda_{\mathbf{pq}} \left| (\mathcal{R}(\mathbf{p}) + \mathbf{d}^{l_{\mathbf{p}}}) - (\mathcal{R}(\mathbf{q}) + \mathbf{d}^{l_{\mathbf{q}}}) \right| \qquad (17)$$

where $\mathcal{R}(\cdot)$ projects the current displacement field on the level of the control points, or

$$\mathcal{R}(\mathbf{p}) = \int_\Omega \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \, \mathcal{D}(\mathbf{x}) \, d\mathbf{x}. \qquad (18)$$

Such smoothness terms together with the data term allows converting the problem of image registration into the form of an MRF formulation as defined

9

in (13). MRFs have been very popular in the area of computer vision [11] in the late eighties and the early nineties. However their main bottleneck at that time was the lack of efficient optimization techniques to recover their lowest potential. Deterministic and non-deterministic algorithms have been considered to address this demand. Iterated conditional modes [1] as well as Highest Confidence First [5] are the most well known deterministic processes which often converge to a local minimum. On the other hand, techniques such as simulated annealing [18] can in theory drive the solution to the optimal one, however in practice the process is rather complicated and increased attention has to be paid to the handling of the temperature decrease. This constrain makes the use of annealing methods almost impractical.

The use of the max-flow/min-cut algorithm [9] and the prove of equivalence with certain MRFs is the main reason of the renaissance for the MRF framework [23], in the late nineties. In particular, the graph-cut algorithm [2] which refers to an efficient implementation of the max-flow/min-cut approach in regular image grids has boosted the attention of the vision community to MRFs. This method can guarantee the global optimum or a good approximation of it (solving a succession of binary problems using the alpha-expansion [35]) under certain conditions [20] which relates the solution with the number of labels and the complexity of the pair/clique-wise potentials. In practice the more complex the interaction terms are, the more challenging the optimization of the objective function is in reasonable computational time. The use of metric or sub-modular functions is the most common constraint related to the definition of the pairwise potential function.

Dense registration is a problem which by default involves a multi-label task while at the same time the regularization terms are often non-linear functions (first and second order derivatives, elastic models, etc.). Therefore assuming that the pairwise potentials are sub-modular functions is unrealistic. Furthermore, one should expect that the level of resolution in the quantized search space will depend on the position of the control point in the image plane. In other words, in areas with strong image content like edges and texture the matching process would be quite precise which will not be the case in smooth areas. Last but not least, given the important number of degrees of freedom, the method should be computationally efficient. Due to the requirements on the pairwise potentials, the use of methods such as alpha-expansion is limited. In the following, a recently proposed MRF optimization strategy based on the primal-dual principle is described. This method provides the needed properties for efficiently solving the problem of image registration within the discrete domain of MRFs.

3 MRF optimization based on Linear Programming

For optimizing the resulting MRF, we seek to assign a label $l_{\mathbf{p}} \in L$ to each node $\mathbf{p} \in G$, so that the MRF energy in (13) is minimized [3]. To this end, a recently proposed method, called Fast-PD, will be used [22]. This is an optimization technique, which builds upon principles drawn from the duality theory of linear programming in order to efficiently derive almost optimal solutions for a very wide class of NP-hard MRFs [21]. When applied to the image registration task, this technique thus offers a series of important advantages compared to the state-of-the-art (see Section 3.2).

For more details about the Fast-PD algorithm, the reader is referred to [21,22]. Here, we will just provide a brief, high level description of the basic driving force behind that algorithm. This driving force will consist of the *primal-dual schema*, which is a well-known technique in the linear programming literature.

3.1 The primal-dual schema for MRF optimization

To understand how the primal-dual schema works in general, we will need to consider the following pair of primal and dual Linear Programs (LPs):

$$\text{PRIMAL: } \min \mathbf{c}^T \mathbf{x} \qquad\qquad \text{DUAL: } \max \mathbf{b}^T \mathbf{y}$$
$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} \qquad\qquad \text{s.t. } \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \qquad (19)$$

Here $\mathbf{A}$ represents a coefficient matrix, while $\mathbf{b}, \mathbf{c}$ are coefficient vectors. Also, $\mathbf{x}, \mathbf{y}$ represent the vectors of primal and dual variables respectively. We seek an optimal solution to the primal program, but with the extra constraint of $\mathbf{x}$ being integral. Due to this integrality requirement, this problem is in general NP-hard and so we need to settle with estimating approximate solutions. A primal-dual f-approximation algorithm achieves that by use of the following principle (illustrated also in Fig. 1(a)):

Primal-Dual Principle 1 *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are integral-primal and dual feasible solutions having a primal-dual gap less than* f, *i.e.:*

$$\mathbf{c}^T \mathbf{x} \leq f \cdot \mathbf{b}^T \mathbf{y}, \qquad (20)$$

---

[3] For similarity measures such as MI we can simply use $-$MI for the data term in order to convert the problem into a minimization problem.

11

(a) The primal-dual principle   (b) The primal-dual schema
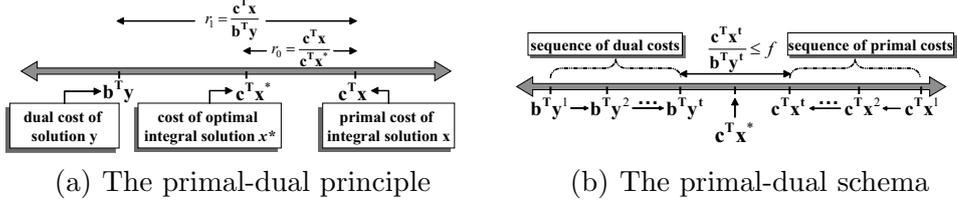
Fig. 1. (a) By weak duality, the optimal cost $\mathbf{c}^T\mathbf{x}^*$ will lie between the costs $\mathbf{b}^T\mathbf{y}$ and $\mathbf{c}^T\mathbf{x}$ of any pair $(\mathbf{x}, \mathbf{y})$ of integral-primal and dual feasible solutions. Therefore, if $\mathbf{b}^T\mathbf{y}$ and $\mathbf{c}^T\mathbf{x}$ are close enough (e.g. their ratio r_1 is $\leq f$), so are $\mathbf{c}^T\mathbf{x}^*$ and $\mathbf{c}^T\mathbf{x}$ (e.g. their ratio r_0 is $\leq f$ as well), thus proving that $\mathbf{x}$ is an f-approximation to $\mathbf{x}^*$.
(b) According to the primal-dual schema, dual and integral-primal feasible solutions make local improvements to each other, until the final costs $\mathbf{b}^T\mathbf{y}^t$, $\mathbf{c}^T\mathbf{x}^t$ are close enough (e.g. their ratio is $\leq f$). We can then apply the primal-dual principle (as in Fig. (a)) and thus conclude that $\mathbf{x}^t$ is an f-approximation to $\mathbf{x}^*$.

then $\mathbf{x}$ is an f-approximation to the optimal integral solution $\mathbf{x}^*$, i.e. $\mathbf{c}^T\mathbf{x}^* \leq \mathbf{c}^T\mathbf{x} \leq f \cdot \mathbf{c}^T\mathbf{x}^*$

Based on the above principle, that lies at the heart of any primal-dual technique, the following iterative schema can be used for deriving an f-approximate solution (this schema is also illustrated graphically in Fig. 1(b)):

Primal-Dual Schema 1 *Keep generating pairs of integral-primal, dual solutions $\{(\mathbf{x}^k, \mathbf{y}^k)\}_{k=1}^t$, until the elements $\mathbf{x}^t$, $\mathbf{y}^t$ of the last pair are both feasible and have a primal-dual gap which is less than f, i.e. condition (20) holds true.*

In order to apply the above schema to MRF optimization, it suffices that we cast the MRF optimization problem as an equivalent integer program. To this end, the following integer programming formulation of (13) has been used as the primal problem:

$$\min \sum_{\mathbf{p} \in G} \sum_{\alpha \in L} V_{\mathbf{p}}(\alpha) x_{\mathbf{p}}(\alpha) + \sum_{\mathbf{p} \in G} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \sum_{\alpha, \beta \in L} V_{\mathbf{pq}}(\alpha, \beta) x_{\mathbf{pq}}(\alpha, \beta) \tag{21}$$

$$\text{s.t.} \sum_{\alpha} x_{\mathbf{p}}(\alpha) = 1 \qquad \forall \, \mathbf{p} \in G \tag{22}$$

$$\sum_{\alpha} x_{\mathbf{pq}}(\alpha, \beta) = x_{\mathbf{q}}(\beta) \qquad \forall \, \beta \in L, \, \forall \, \mathbf{p} \in G \wedge \mathbf{q} \in \mathcal{N}(\mathbf{p}) \tag{23}$$

$$\sum_{\beta} x_{\mathbf{pq}}(\alpha, \beta) = x_{\mathbf{p}}(\alpha) \qquad \forall \, \alpha \in L, \, \forall \, \mathbf{p} \in G \wedge \mathbf{q} \in \mathcal{N}(\mathbf{p}) \tag{24}$$

$$x_{\mathbf{p}}(\cdot), \, x_{\mathbf{pq}}(\cdot, \cdot) \in \{0, 1\}$$

Here, in order to linearize the MRF energy, we have replaced the discrete variables $l_{\mathbf{p}}$ with the binary variables $x_{\mathbf{p}}(\cdot)$ and $x_{\mathbf{pq}}(\cdot, \cdot)$. More specifically, the $\{0, 1\}$-variable $x_{\mathbf{p}}(\alpha)$ indicates that node $\mathbf{p}$ is assigned label α (i.e., $l_{\mathbf{p}} = \alpha$), while the $\{0, 1\}$-variable $x_{\mathbf{pq}}(\alpha, \beta)$ indicates that vertices $\mathbf{p}, \mathbf{q}$ are assigned labels α, β respectively (i.e., $l_{\mathbf{p}} = \alpha$, $l_{\mathbf{q}} = \beta$). Furthermore, the constraints in (22) simply express the fact that each node must receive exactly one label,

while constraints (23), (24) maintain consistency between variables $x_{\mathbf{p}}(\cdot), x_{\mathbf{q}}(\cdot)$ and variables $x_{\mathbf{pq}}(\cdot, \cdot)$, in the sense that if $x_{\mathbf{p}}(\alpha) = 1$ and $x_{\mathbf{q}}(\beta) = 1$ holds true, then these constraints force $x_{\mathbf{pq}}(\alpha, \beta) = 1$ to hold true as well (as desired).

The linear programming relaxation of the above integer program is then taken (by relaxing the binary constraints to $x_{\mathbf{p}}(\cdot) \geq 0, x_{\mathbf{pq}}(\cdot, \cdot) \geq 0$), and the dual of the resulting LP is used as our dual problem. The Fast-PD algorithm is then derived by applying the primal-dual schema to this pair of primal-dual LPs, while using $f = 2\frac{d_{\max}}{d_{\min}}$ [4] as the approximation factor in (20).

3.2 Advantages of the primal-dual approach

Fast-PD has many nice properties, which makes it an excellent candidate for our image registration task. In particular, it offers the following advantages: **1) Generality:** Fast-PD can handle a very wide class of MRFs, since it merely requires $V_{\mathbf{pq}}(\cdot, \cdot) \geq 0$. Hence, by using Fast-PD, our image registration framework can automatically incorporate any similarity measure, as well as a very wide class of smoothness penalty functions. **2) Optimality:** Furthermore, Fast-PD can always guarantee that the generated solution will be an f-approximation to the true optimum (where $f = 2\frac{d_{\max}}{d_{\min}}$). **3) Per-instance approximation factors:** In fact, besides the above worst-case approximation factor, Fast-PD can also continuously update a *per-instance* approximation factor during its execution. In practice, this factor drops to 1 very quickly, thus allowing the global optimum to be found up to a user/application bound. **4) Speed:** Finally, Fast-PD provides great computational efficiency, since it can reach an almost optimal solution very fast and in an efficient manner.

4 Implementation Details & Parameter Setting

In order to prove the concept of our framework, we implemented a deformable registration application in C++. We follow the widely used approach of multi-resolution registration in a course-to-fine manner. The control grid is successively refined by decreasing the grid point spacing by factor of two while at the same time we use a Gaussian pyramid for the image data. In most applications three levels are sufficient. As mentioned before, the deformation grid is reset after each optimization cycle and the resulting displacement fields are incrementally composed on the image level (see also Section 2.2.2) Thus, we can do hierarchical registration without using B-Spline refinement methods. The resolution of the control grid highly depends on the application and is to be

---

[4] $d_{\max} \equiv \max_{\alpha \neq \beta} V_{\mathbf{pq}}(\alpha, \beta), \quad d_{\min} \equiv \min_{\alpha \neq \beta} V_{\mathbf{pq}}(\alpha, \beta)$

specified by the user. If large deformations are expected, one should also start with few control points. In general, we expect that a global pre-registration has been performed ahead of our local registration such that most of the global linear part (translation, rotation) is removed from the images. Therefore, in many settings a control grid resolution of 20 mm grid spacing and refinements to 10 and 5 mm are sufficient as default parameters. These should be changed according to specific problems. Running our framework with a very coarse grid as a potential initialization step efficiently removes global transformation parts such as anisotropic scaling, translation, or shearing.

4.1 Definition of Label Sets

The next aspect and most crucial part for the registration accuracy is the configuration of the discrete set of displacements. Basically, four parameters are controlling the discretization of the solution space. The first one defines the maximum allowed displacement for each level of the multi-scale approach. In scenarios where a diffeomorphic solution is desired, the maximum allowed value of the parameter is bound to the grid resolution (see again Section 2.2.2). Otherwise the value can be freely set by the user. Choices for an appropriate value are problem specific but not very critical since the incremental approach can account for displacements out of the capture range. Additionally, in our software the capture range is visualized at every grid node such that the user can control whether important deformations visible in the images are covered.

The second parameter controls the sampling rate from the zero-displacement up to the maximum displacement. This value is more critical since it directly influences the number of total labels which influences the computational speed of the approximation scheme (see Section 2.2.1). Every additional label causes one extra outer-loop for this scheme. Again, due to the scale-space approach and the incremental morphings we can keep this value quite small. The default value in our application is set to 5.

The third parameter concerns the selection of the type of sampling. We distinguish between dense and sparse sampling of the solution space. A dense sampling results in $(2N + 1)^D$ labels (including the zero-displacement vector) where N is the sampling rate and D the number of dimensions. Using the default sampling rate of $N = 5$ this results in 121 labels for 2D and 1331 labels for 3D. The sparse sampling considers only displacements along the main axis. Therefore, we get $4N + 1$ labels for 2D and $6N + 1$ labels in 3D resulting in 21 respectively 31 labels considering the default sampling rate. The selection of the type of sampling is mainly a compromise between the computational speed of one optimization cycle (including the approximation scheme for the data term) and the number cycles that have to be performed to converge to

satisfactory registration results. In 2D, we usually select the dense sampling since additional outer-loops in the approximation scheme are here not very expensive. In 3D, we normally use the sparse sampling which gives very good results in practice while reducing the computational time immensely, which is shown throughout the experimental validation.

The fourth and last parameter concerning the capture space controls the iterative refinement of the label sets. Since the Fast-PD optimization generates quasi-optimal labelings on the discrete set of labels, usually no further improvement of the registration can be achieved by keeping the same displacement set. A simple scaling factor is multiplied with the initial maximum displacement and the capture range is then resampled. This enables sub-pixel precision on the solution space. By default we set the scaling factor to 0.33 while performing 5 optimization cycles on each pyramid level.

4.2   Similarity Measure Back-Projection

Equation (8) plays a key role in the derivation of our framework. On the one hand, we need this formulation in order to determine the local similarity measures on the control point level. On the other hand, for the unary potential functions of the MRF formulation this implies some problems when using cubic B-Splines in $\hat{\eta}(\cdot)$. We mentioned before, that in general the unary potentials are assumed to be independent, which leads us to our data term approximation scheme. Since the overlapping areas/volumes within the images are rather large for cubic functions, we avoid using them for the data term computation. In practice, linear functions are more appropriate to provide a good balance between speed and accuracy. In all the following experiments, we use linear weighting functions for determining the unary potentials. However, the dense deformation field is computed based on cubic B-Splines in order to obtain smooth results.

5   Experiments on Known and Unknown Deformations

Our framework currently contains a range of well-known similarity measures, namely the sum of absolute differences (SAD) [15], the sum of squared differences (SSD) [15], the normalized cross correlation (NCC) [15], the normalized mutual information (NMI) [25,36], the correlation ratio (CR) [27], and a measure involving an intensity-based and a geometric-based term which combines the sum of absolute differences and image gradient inner product (SADGIP). An additional weighting factor γ is used to control the influence of these two
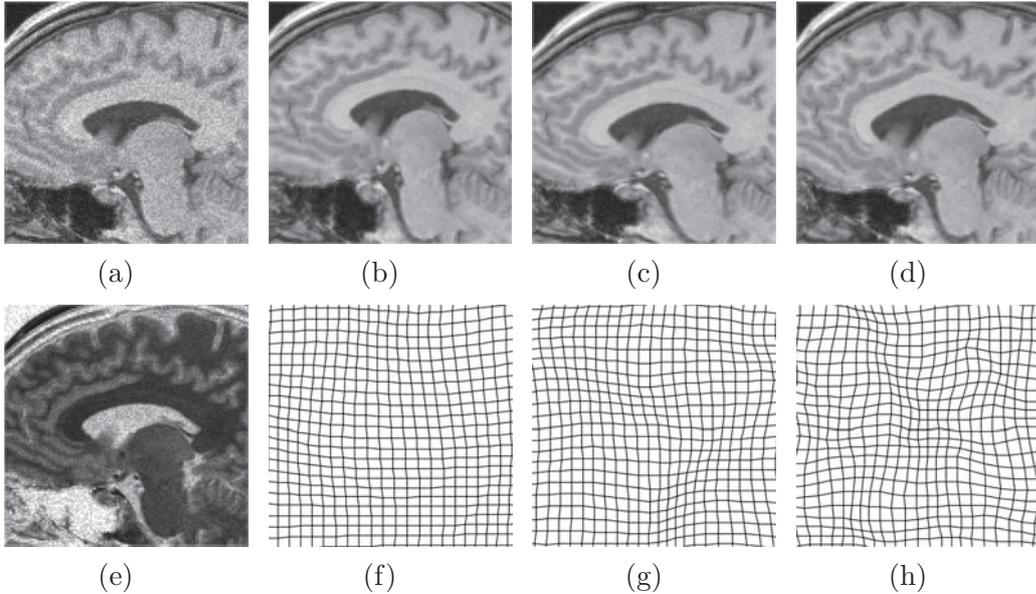
Fig. 2. Data for the first experiment. (a) The source image, and (b)-(d) generated target images with different degree of deformation (denoted as Target A to C). (e) The inverse squared source image used for multi-modal tests. (f)-(h) Deformation fields corresponding to the upper target images.

terms. The SADGIP is defined as

$$
\rho_{\text{SADGIP}}(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x}))) =
$$
$$
(1 - \gamma)|g(\mathbf{x}) - f(\mathcal{T}(\mathbf{x}))| + \gamma \left| \frac{\nabla g(\mathbf{x})}{|\nabla g(\mathbf{x})|} \cdot \frac{\nabla f(\mathcal{T}(\mathbf{x}))}{|\nabla f(\mathcal{T}(\mathbf{x}))|} \right|. \quad (25)
$$

Note that, by setting $\gamma = 1$, this similarity measure might be also used for multi-modal registration. The image gradients are computed using a Sobel filter. We should also note, that the NMI as well as CR measures are based on simple histogramming techniques. This will be changed in future implementations.

We evaluate our framework on several data sets. In general, the evaluation and thus, validation of non-rigid image registration methods is a difficult task. Usually, ground truth data for real deformations - especially in medical applications - is not available. Therefore, we perform several experiments illustrating the potentials of our approach. All our experiments are performed on an Intel 2.16 GHz Mobile CPU.

5.1 Benchmark for Similarity Measures on Synthetic Deformations

The first experiment can be seen as a benchmark for similarity measures. In order to evaluate the efficiency of different measures, we test our method on
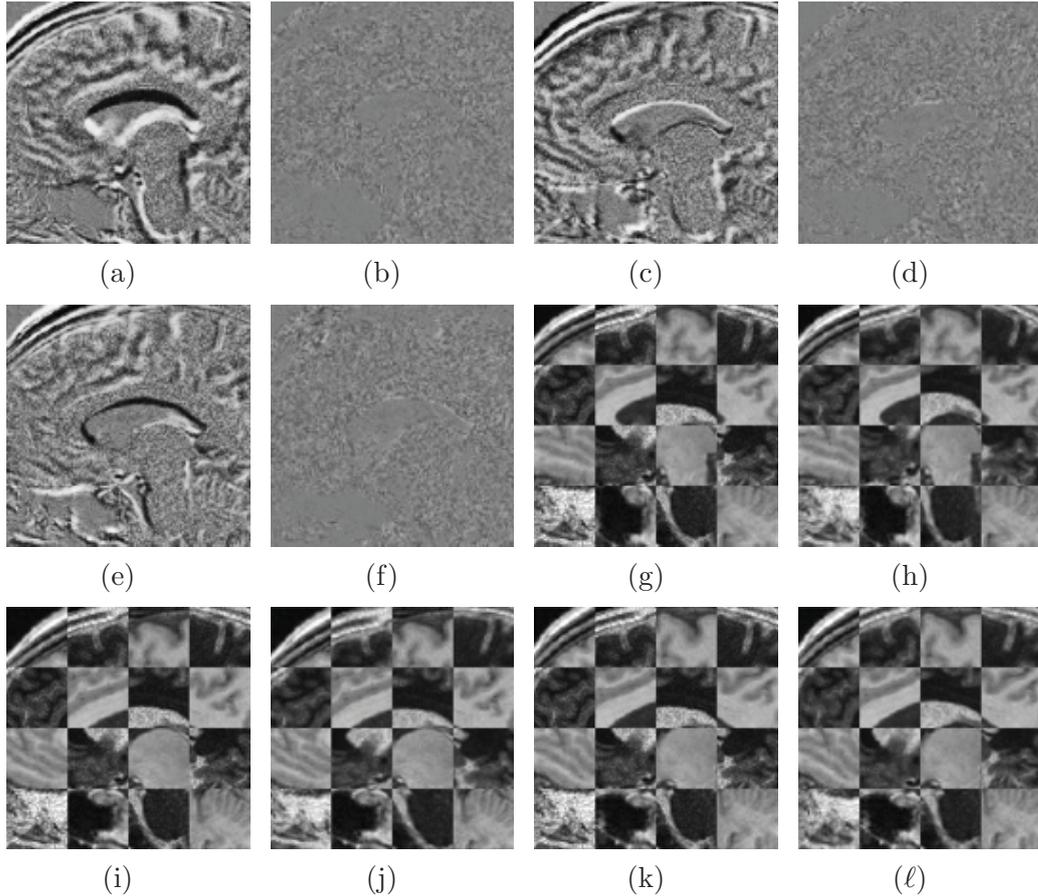
Fig. 3. Exemplary visual results for the first experiments. (a)-(l) Visualization of the image alignment before and after registration using our method with normalized mutual information based on histogramming. For the mono-modal experiment, the difference images are shown (contrast enhanced for print quality) and a checkerboard visualization for the multi-modal registration. Note, that in the latter case, dark areas of the source image should be aligned with bright areas of the target image.

simulated deformations where the ground truth deformation field is known. Three different target images (A-C) are generated from the 2D MRI source image by randomly displacing equally distributed landmarks within a range of -10 to 10 millimeters in both dimensions. The number of landmarks is varied for the three targets in order to obtain different degrees of deformation. The warping is done using thin-plate splines (TPS) with different regularization factors. The resulting targets are shown in Fig. 2. Target A, B, and C (Fig. 2b-d) are generated using 60, 30, and 15 mm spacing between the landmarks, respectively. Before registration we add uniformly distributed random noise with mean 0 and variance 0.01 to the source images (relative to the maximum and minimum intensity). For the multi-modal experiment, we created a second source image with inverted squared intensities. The resolution of the images is 256×256 with an isotropic pixel spacing of 1 mm.

We perform the registration using the default parameters mentioned in Section

17

Target A	Elastix (GD)			Elastix (QN)			Our Method		
Metric	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*
SAD	n/a			n/a			2.25 (1.98)	0.19 (0.14)	10
SSD	7.29 (15.97)	0.61 (1.80)	118	4.17 (4.46)	0.32 (0.32)	160	2.03 (1.73)	0.16 (0.12)	10
NCC	3.35 (3.69)	0.23 (0.20)	233	4.23 (4.17)	0.35 (0.37)	222	2.04 (1.94)	0.16 (0.14)	12
SADGIP (γ=0.1)	n/a			n/a			2.23 (1.98)	0.18 (0.14)	35
SADGIP (γ=1.0)	n/a			n/a			5.26 (9.53)	0.38 (0.61)	27
NMI (Histo)	n/a			n/a			2.73 (2.57)	0.22 (0.20)	45
NMI (Parzen)	3.54 (7.39)	0.27 (0.51)	71	3.96 (7.01)	0.30 (0.42)	77	n/a		
CR (Histo)	n/a			n/a			2.98 (3.17)	0.25 (0.24)	32
SADGIP (γ=1.0)	n/a			n/a			6.75 (14.64)	0.71 (1.79)	26
NMI (Histo)	n/a			n/a			2.10 (3.16)	0.14 (0.14)	46
NMI (Parzen)	1.79 (2.70)	0.11 (0.11)	71	3.02 (7.07)	0.18 (0.32)	124	n/a		
CR (Histo)	n/a			n/a			2.66 (5.31)	0.16 (0.22)	30

Target B	Elastix (GD)			Elastix (QN)			Our Method		
Metric	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*
SAD	n/a			n/a			3.39 (4.75)	0.21 (0.19)	11
SSD	8.97 (19.18)	0.79 (2.27)	117	5.11 (6.75)	0.30 (0.29)	160	3.96 (7.61)	0.23 (0.29)	11
NCC	5.44 (8.16)	0.31 (0.33)	226	5.14 (6.00)	0.34 (0.38)	213	3.39 (4.67)	0.20 (0.18)	12
SADGIP (γ=0.1)	n/a			n/a			3.56 (5.24)	0.22 (0.20)	38
SADGIP (γ=1.0)	n/a			n/a			5.90 (6.54)	0.42 (0.43)	27
NMI (Histo)	n/a			n/a			4.71 (7.47)	0.29 (0.38)	49
NMI (Parzen)	3.68 (5.69)	0.24 (0.30)	70	4.17 (5.55)	0.28 (0.37)	124	n/a		
CR (Histo)	n/a			n/a			4.36 (6.09)	0.27 (0.33)	30
SADGIP (γ=1.0)	n/a			n/a			8.97 (13.37)	0.64 (0.77)	27
NMI (Histo)	n/a			n/a			3.74 (9.21)	0.23 (0.49)	46
NMI (Parzen)	3.54 (7.95)	0.22 (0.44)	71	3.24 (7.23)	0.21 (0.42)	80	n/a		
CR (Histo)	n/a			n/a			2.98 (5.50)	0.18 (0.22)	28

Target C	Elastix (GD)			Elastix (QN)			Our Method		
Metric	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*	*AE*	*MOD*	*TIME*
SAD	n/a			n/a			5.73 (8.46)	0.32 (0.45)	11
SSD	9.55 (14.77)	0.79 (2.33)	115	6.32 (6.06)	0.34 (0.33)	159	5.57 (7.65)	0.30 (0.38)	10
NCC	7.73 (7.47)	0.44 (0.42)	217	6.22 (5.94)	0.33 (0.32)	325	6.67 (7.62)	0.36 (0.43)	12
SADGIP (γ=0.1)	n/a			n/a			4.87 (5.16)	0.26 (0.23)	39
SADGIP (γ=1.0)	n/a			n/a			15.99 (22.63)	0.99 (1.25)	21
NMI (Histo)	n/a			n/a			7.66 (11.32)	0.41 (0.63)	48
NMI (Parzen)	7.57 (13.93)	0.39 (0.72)	70	7.95 (10.79)	0.47 (0.65)	59	n/a		
CR (Histo)	n/a			n/a			6.82 (8.73)	0.35 (0.41)	36
SADGIP (γ=1.0)	n/a			n/a			7.98 (16.07)	0.41 (0.84)	27
NMI (Histo)	n/a			n/a			4.78 (10.11)	0.26 (0.62)	46
NMI (Parzen)	3.32 (5.80)	0.19 (0.37)	70	5.52 (10.80)	0.31 (0.63)	58	n/a		
CR (Histo)	n/a			n/a			4.50 (9.38)	0.25 (0.60)	31

Table 1

Quantitative results for the first experiment. The last three rows in each table are the results for the multi-modal registration. We estimate the mean of the angular error (AE), the mean of the magnitude of differences (MOD), and their respective standard deviations (printed in brackets). The registration time is given in seconds. Elastix (GD) and Elastix (QN) denote the gradient-based methods using gradient-descent and Quasi-Newton optimization.

4. The smoothness factors $\lambda_{\mathbf{pq}}$ are set to the same values for all control points. The values are empirically determined according to the used similarity measure in order to achieve visually good results. For this experiment, we use the full regularization over time defined in Equation (17).

Additionally, we compare our results to an FFD-based registration framework called Elastix [5] [19]. We run the gradient-based registrations with two different

---

[5] Elastix is available for download on http://www.isi.uu.nl/Elastix/.

optimizers, a standard gradient-descent and a Quasi-Newton optimizer using line search strategies [19]. The parameters of both methods are tuned until we achieve visually good results. The number of iterations for the three pyramid levels are set to 500, 250, and 100, respectively. Elastix is using a random sample selection technique in order to increase the computational efficiency. We set the number of samples to 8000, which seems to be a good compromise between accuracy and speed. In some cases where the registration fails, the number of samples is increased to 16000 resulting in an increased computational time. The general configuration for both methods is the same as for our method: same number of pyramid levels, same grid resolution, and the same transformation model based on cubic B-Splines. Furthermore, in case of NMI and CR implementation 64 bins are used for the histograms.

For the quantitative evaluation, two error metrics are considered, namely the angular error (AE) [8] measured in degrees and the magnitude of differences (MOD) measured in millimeters. Thus, we can measure the deviation of the registration results compared to the ground truth. The results for our method and the two gradient-based approaches are presented in Tab. 1. Elastix provides three of our implemented similarity measures, the SSD, NCC, and the NMI, where the latter one is based on Parzen windowing (in contrast to our rather simple implementation of histogramming). For SSD and NCC our method performs best in almost all cases both in accuracy and speed. For NMI the three approaches perform quite similar in accuracy while in some cases the gradient-based approaches are slightly better while our method is always faster. However, since our results for NMI are only based on a simple histogramming approach, the performance is still remarkably good. The visual results for our NMI registration are shown in Fig. 3. The combined measure using intensity and geometrical information from image gradient performs much worse than reported in [14]. This is mainly to the presence of noise in our experiments. Still, in some cases the additional geometrical information can improve the performance of the SAD measure while in other cases, where larger deformations are present it even fails to converge to a satisfactory registration. The use of the image gradients only ($\gamma = 1$) in case of multi-modal registration seems to be not suitable, at least for our experiments.

5.2   Inter Subject Brain Registration

In the next experiment the registration accuracy will be determined using manual segmentations. Eight MRI data sets of the brain are registered where in all of them manual expert segmentations of the gray and white matter are available. The image resolution is $256 \times 256 \times 128$ with a voxel spacing of $0.9375 \times 0.9375 \times 1.5$ mm. The data is part of the Internet Brain Segmentation Repository (IBSR) provided by the Center for Morphometric Analysis at Mas-

Gray Matter	Affine 12-DOF (GD)			Rueckert (GD)			Our Method		
Image	DICE	Sens	Spec	DICE	Sens	Spec	DICE	Sens	Spec
Brain 1	0.7022	0.7679	0.9633	0.8205	0.8547	0.9800	0.8567	0.8936	0.9831
Brain 2	0.7267	0.7236	0.9792	0.8142	0.8125	0.9857	0.8468	0.8489	0.9878
Brain 3	0.6687	0.6047	0.9816	0.8054	0.8059	0.9823	0.8332	0.8194	0.9867
Brain 4	0.7270	0.7924	0.9703	0.8154	0.8524	0.9818	0.8535	0.9065	0.9833
Brain 5	0.6977	0.7341	0.9686	0.8041	0.8449	0.9782	0.8355	0.8787	0.9809
Brain 6	0.7078	0.6328	0.9852	0.8116	0.7615	0.9891	0.8415	0.8112	0.9889
Brain 7	0.7062	0.6793	0.9779	0.8308	0.8303	0.9848	0.8591	0.8725	0.9857
Average	0.7052	0.7050	0.9752	0.8146	0.8232	0.9831	0.8466	0.8615	0.9852
Running Time	4 minutes			3 hours 50 minutes			8 minutes		

White Matter	Affine 12-DOF (GD)			Rueckert (GD)			Our Method		
Image	DICE	Sens	Spec	DICE	Sens	Spec	DICE	Sens	Spec
Brain 1	0.6484	0.6214	0.9842	0.7686	0.7296	0.9910	0.8344	0.7909	0.9944
Brain 2	0.6269	0.6335	0.9863	0.7225	0.6794	0.9929	0.7962	0.8031	0.9924
Brain 3	0.6097	0.5622	0.9887	0.7312	0.7212	0.9899	0.7855	0.7937	0.9909
Brain 4	0.6860	0.6881	0.9866	0.7879	0.8034	0.9900	0.8428	0.8195	0.9947
Brain 5	0.6372	0.6080	0.9853	0.7598	0.7231	0.9912	0.8329	0.8297	0.9921
Brain 6	0.6521	0.6477	0.9882	0.7338	0.8976	0.9808	0.7794	0.8659	0.9876
Brain 7	0.6430	0.5924	0.9884	0.7840	0.8102	0.9881	0.8262	0.8312	0.9916
Average	0.6433	0.6219	0.9868	0.7554	0.7664	0.9891	0.8139	0.8191	0.9920
Running Time	4 minutes			3 hours 50 minutes			8 minutes		

Table 2
Results for the brain registration evaluated on manual segmentations of the gray and white matter. Given is the DICE score, the sensitivity, and the specificity. The running time states the time needed for a single registration. Rueckert (GD) denotes the method in [31] using a gradient-descent optimizer.

sachusetts General Hospital (http://www.cma.mgh.harvard.edu/ibsr/). The T1-weighted images have been positionally normalized into the Talairach orientation (rotation only). We select one data set as the template and register it to the remaining seven data sets. The recovered transformation is then used to warp the template segmentations. In order to compare the warped segmentations to the manual ones, we determine three measures, namely DICE score, the sensitivity, and the specificity. The registration is performed using the NCC similarity measure and an incremental regularization as defined in Equation (16). The weighting factor $\lambda_{\mathbf{pq}}$ is set to 0.005. We use four resolution levels, starting with 40 mm control point spacing which is then refined to 20, 10 and finally 5 mm. The label set scaling factor is set 0.75. The rest of the parameters is set to the default values mentioned in Section 4. A single registration takes about 8 minutes which splits into approximately 7 minutes for the data term computation and 30 seconds for the Fast-PD optimizer plus some seconds for the intermediate warpings.

We compare our results for the MRI brain registration with a 12 degrees-of-freedom (DOF) affine registration (3 rotations, 3 translations, 3 scalings, 3 shears) and the FFD-based registration [6] proposed by Rueckert et al. [31,32] which can be seen as the state-of-the-art in FFD registration. Both methods use a standard gradient-descent optimizer, the NCC similarity measure, and

---

[6] Available on http://wwwhomes.doc.ic.ac.uk/~dr/software/.
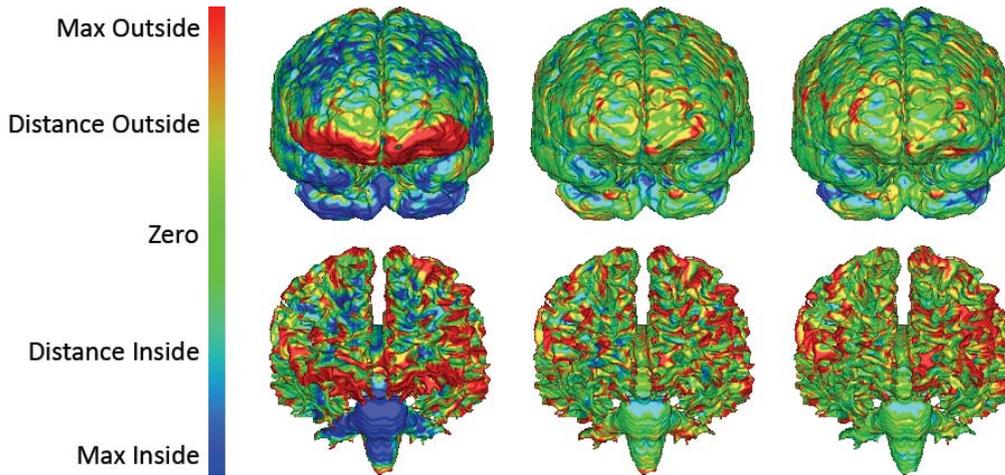
Fig. 4. Color encoded visualization of the surface distance between warped and expert segmentation after affine (left), gradient-descent (middle), and our registration (right) for the Brain 1 data set. The color range is scaled to a maximum and minimum distance of 3 mm. In some regions, the results of the gradient-descent approach seem to be slightly better. However, the actual average surface distance (ASD) after registration for the gray matter is 1.66, 1.14, and 1.00 mm for affine, gradient-descent, and our method, respectively. For the white matter the resulting ASD is 1.92, 1.31, and 1.06 mm.

also a four level resolution approach where the grid resolution for the FFD-based registration is the same as in our method. A single affine registration takes about 4 minutes which is just the half of our deformable registration. The single gradient-descent FFD registration takes more than 3 hours and 50 minutes, which is almost 30 times more than our method.

The quantitative results are presented in Tab. 2. Visual results of the surface distance (SD) for the gray and white matter of one of the data sets are shown in Fig. 4. The SD map is computed using the tool[7] described in [12]. Our method performs best for all three measures while reducing the running time for the FFD registration compared with gradient-descent extremely. We should note that all three methods start from the original images as initialization. Neither our method nor Rueckert's registration is using the affine results as an initialization. Using the very coarse grid of 40 mm ($7 \times 7 \times 6$ grid) in the beginning of the deformable registration, we demonstrate that parts of the affine transformation (e.g. anisotropic scaling) can be successfully recovered by FFD.

---

[7] The tool Valmet is available on http://www.ia.unc.edu/dev/download/valmet/.

6 Discussion

In this paper we have proposed a novel framework for deformable image registration that bridges the gap between continuous deformations and optimal discrete optimization. Our method reformulates registration using an MRF definition, and recovers the optimal solution to the designed objective function through efficient linear programming. Towards capturing important deformations, we propose an incremental estimation of the deformation component. These objectives are met through a discrete labeling problem defined over an MRF graph. Graph edges introduce smoothness on the deformation field, while the unary potentials encode the image support for a given deformation hypothesis versus another. Therefore, the method is gradient-free meaning no computation of the derivative of the employed cost function is needed, it can encode any similarity measure and can recover the optimal solution up to a bound. We have demonstrated the immense computational speedup provided by our framework. In addition, we believe that the intuitive adjustment of the space of solutions which is directly related to the images to be registered is another advantage compared to gradient-descent approaches where the user cannot easily control the search space.

In several applications, building anatomical atlases and models of variations between training examples is feasible. In such a context, one can consider a partial graph where connection hypotheses are determined according to the density of expected deformations. Such a direction will introduce prior knowledge in the registration process and will make the optimization step more efficient. Moreover, the use of shape and appearance models can be considered to perform segmentation through registration. Assuming a prior model that involves both geometry and texture, and given a new volume one can define/recover segmentation through the deformation of the model to the image that is a natural registration problem which can be optimally addressed from the proposed framework. In [13], preliminary but promising results on atlas-based segmentation using our framework are presented.

References

[1] J. Besag. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society*, Series B 48:259–302, 1986.

[2] Y. Boykov and O. Veksler. *Handbook of Mathematical Models in Computer Vision*, chapter Graph Cuts in Vision and Graphics: Theories and Applications, pages 79–96. Springer, 2005.

[3] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache. Iconic feature based nonrigid registration: the pasha algorithm. *Comput. Vis. Image Underst.*,

89(2-3):272–298, 2003.

[4] Y. Choi and S. Lee. Injectivity conditions of 2d and 3d uniform cubic b-spline functions. *Graphical Models*, 62(6):411–427, 2000.

[5] P. B. Chou and C. M. Brown. The theory and practice of bayesian image labeling. *Int. J. Comput. Vision*, 4(3):185–210, 1990.

[6] C. Davatzikos, J. Prince, and R. Bryan. Image registration based on boundary mapping. *IEEE Transactions on Medical Imaging*, 15(1):112–115, 1996.

[7] J. C. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(1):85–106, 2000.

[8] D. J. Fleet. *Measurement of Image Velocity*. Kluwer Academic Publishers, Norwell, MA, USA, 1992.

[9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[10] G.E.Christensen. *Deformable shape models for anatomy.* PhD thesis, Washington University, 1994.

[11] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 6, pages 721–741, 1984.

[12] G. Gerig, M. Jomier, and M. Chakos. Valmet: A new validation tool for assessing and improving 3d object segmentations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, 2001.

[13] B. Glocker, N. Komodakis, N. Paragios, C. Glaser, G. Tziritas, and N. Navab. Primal/dual linear programming and statistical atlases for cartilage segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007)*, Brisbane, Australia, October 2007.

[14] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging*, Kerkrade, Netherlands, July 2007.

[15] J. Hajnal, D. L. G. Hill, and D. J. Hawkes, editors. *Medical Image Registration*. CRC Press, 2001.

[16] P. Hellier and C. Barillot. Coupling dense and landmark-based approaches for nonrigid registration. *Medical Imaging, IEEE Transactions on*, 22(2):217–227, Feb. 2003.

[17] G. Hermosillo, C. Chefd'hotel, and O. Faugeras. Variational methods for multimodal image matching. *Int. J. Comput. Vision*, 50(3):329–343, 2002.

[18] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[19] S. Klein, M. Staring, and J. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing*, 16(12):2879–2890, December 2007.

[20] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159, Feb 2004.

[21] N. Komodakis and G. Tziritas. Approximate labeling via graph-cuts based on linear programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2007.

[22] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic mrfs. In *CVPR*, 2007.

[23] S. Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., 2001.

[24] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using free-form deformations with a local rigidity constraint. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 639–646, 2004.

[25] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, April 1997.

[26] X. Pennec, N. Ayache, and J.-P. Thirion. *Handbook of medical imaging*, chapter Landmark-based registration using features identified through differential geometry, pages 499–513. 2000.

[27] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124. Springer-Verlag, 1998.

[28] T. Rohlfing, J. Maurer, C.R., D. Bluemke, and M. Jacobs. Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *Medical Imaging, IEEE Transactions on*, 22(6):730–741, June 2003.

[29] K. Rohr, M. Fornefett, and H. S. Stiehl. Spline-based elastic image registration: integration of landmark errors and orientation attributes. *Comput. Vis. Image Underst.*, 90(2):153–168, 2003.

[30] D. Rueckert, P. Aljabar, R. A. Heckemann, J. Hajnal, and A. Hammers. Diffeomorphic registration using b-splines. In *Proceedings of 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006)*, October 2006.

[31] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999.

[32] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2001.

[33] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *SIGGRAPH '86: Proceedings of the 13th annual conference on*

Computer graphics and interactive techniques, pages 151–160, New York, NY, USA, 1986. ACM Press.

[34] A. Tikhonov. Ill-posed problems in natural sciences, 1992. Coronet.

[35] O. Veksler. *Efficient graph-based energy minimization methods in computer vision.* PhD thesis, Graduate School of Cornell University, 1999. Adviser-Ramin Zabih.

[36] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[37] L. Zollei, J. Fisher, and W. Wells. *Handbook of Mathematical Models in Computer Vision*, chapter An Introduction to Statistical Methods of Medical Image Registration. Springer, 2005.