Dense Image Registration through MRFs and Efficient Linear Programming

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

[a] *GALEN Group, Laboratoire de Mathématiques Appliquées aux Systèmes Ecole Centrale de 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, fast, efficient and gradient free approach to dense image registration. 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 dissimilarity measure) projected on the control points, and a smoothness term that penalize local deviations on the deformation field according to a neighborhood system on grid. Towards a fully discrete approach the search space is quantized resulting in a fully discrete model. In order to account for large deformations and produce a finer and finer resolution 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 extreme potentials of our approach.

Key words: Discrete Optimization, Deformable Registration, Linear Programming

* Corresponding author. TUM, Institut für Informatik / I16, Boltzmannstrasse 3, 85748 Garching bei München, Germany

Email addresses: `glocker@in.tum.de` (Ben Glocker), `komod@csd.uoc.gr` (Nikos Komodakis), `nikos.paragios@ecp.fr` (Nikos Paragios),

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 with them either occasionally or periodically. The ability to compare or fuse information across subjects with origins from 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) deformation model, (ii) dissimilarity criterion and (iii) 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 non-linear relationship. Several methods exist in the literature where specific measures are designed to account for this non-linearity 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 [17,28] are a classic example of geometric-driven registration. In such a setting, a number of anatomical key points [25]/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 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 [4] seek for “visual” correspondences between the

tziritas@csd.uoc.gr (Georgios Tziritas), navab@in.tum.de (Nassir Navab).

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 [16], sum of absolute differences [16], cross correlation [16] 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 relation of intensities. Non-linear measures have often been used [18] where normalized mutual information [24], Kullback-Liebler divergence [33] and correlation ratio [26] are some of the measures used to define similarity between different modalities.

Once the dissimilarity 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. These methods require an important customization from one application to another since a correlation exists between the modalities/problem and the selection of the dissimilarity measure. Furthermore, the optimization is often sub-optimal due the non-convexity of the designed cost functions. Furthermore, in particular when considering complex dissimilarity 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 dissimilarity measure selection, as well as to 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. 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 dissimilarity measure between the source and the target is minimal for all voxels. The optimization procedure is independent from the graph construction, and therefore any dissimilarity measure can be used.

The reminder 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 [15], we consider (without loss of generality) the 2D image domain. Let us consider a source $f : \Omega = [1, N] \times [1, M] \rightarrow \mathcal{R}$ 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 [32], 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 linear relationship between the signals 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 dissimilarity measure ρ is introduced to account for the non-linear transformation relating 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 as the application itself. Once this measure is defined the data term is combined with the smoothness one to determine the objective 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 objective 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 $\mathcal{G} : [1, K] \times [1, L]$ (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

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

where $\eta(\cdot)$ is the weighting function measuring the contribution of the control point p to the displacement field $\mathcal{D}$. The position vector of point p is denoted as $\mathbf{p}$. In such a theoretical setting without loss of generality we consider Free Form Deformations (FFD) based on cubic B-Splines as a transformation model. FFD are successfully applied in non-rigid image registration [27,31]. 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 is defined as

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

where $i = \lfloor x/K \rfloor - 1$, $j = \lfloor y/L \rfloor - 1$, $u = x/K - \lfloor x/K \rfloor$, and $v = y/L - \lfloor y/L \rfloor$ and where B_l represents the lth basis function of the B-Spline. The three-dimensional version is defined in a straightforward manner.

By defining the registration problem based on such a deformation model we can now rewrite the criterion earlier introduced,

$$E_{\text{data}}(\mathcal{T}) = \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} \int_{\Omega} \eta^{-1}(|\mathbf{x} - \mathbf{p}|) \; \cdot \rho_h(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x})))d\mathbf{x}. \tag{7}$$

where $\eta^{-1}(\cdot)$ is the inverse projection for the contribution to the objective of the image pixel $\mathbf{x}$ according to the influence of the control point p,

$$\eta^{-1}(|\mathbf{x} - \mathbf{p}|) = \frac{\eta(|\mathbf{x} - \mathbf{p}|)}{\int_{\Omega} \eta(|\mathbf{y} - \mathbf{p}|)d\mathbf{y}}. \tag{8}$$

Such a term will guarantee photometric correspondence between the two images. Hence, this term is also called the data term. The transformation due to the interpolation inherits some implicit smoothness properties. However, in order to avoid folding of the deformation grid, one can consider a smoothness term on the grid domain, or

$$E_{\text{smooth}}(\mathcal{T}) = \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} \phi(|\nabla_{\mathcal{G}}\, \mathbf{d}_p|) \tag{9}$$

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

$$E_{\text{total}} = E_{\text{data}} + E_{\text{smooth}}. \tag{10}$$

The most common way to obtain the transformation parameters is through the use of a gradient-descent method in an iterative approach [30]. Thus given an initial guess, one updates the estimate according to the following formula $\left[\mathcal{T}^m = \mathcal{T}^{m-1} - \delta t \frac{\delta E_{\text{total}}}{\partial T}\right]$. 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 $\mathcal{L} = \{u^1, ..., u^i\}$ corresponding to a quantized version of the deformation space $\Theta = \{\mathbf{d}^1, ..., \mathbf{d}^i\}$. A label assignment u_p to a grid node p is associated with displacing the node by the corresponding vector $\mathbf{d}^{u_p}$. The displacement field associated with a certain discrete labeling u becomes

$$\mathcal{D}(\mathbf{x}) = \sum_{p \in \mathcal{G}} \eta(|\mathbf{x} - \mathbf{p}|)\, \mathbf{d}^{u_p}. \tag{11}$$

One can reformulate the registration as a discrete optimization problem, that is assign individual labels u_p to the grid nodes such that

$$E_{\text{data}}(u) = \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} \int_\Omega \eta^{-1}(|\mathbf{x} - \mathbf{p}|) \rho_h(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x})))d\mathbf{x} \approx \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} V_p(u_p) \tag{12}$$

where $V_p(\cdot)$ represents a local dissimilarity measure. There is a main issue coming along when using such an MRF based formulation. In general, the so-called singleton potential functions $V_p(\cdot)$ are assumed to be independent. This is obviously not the case when using higher order polynomials for the weighting function in our transformation model. Actually it is not possible to pre-compute the exact data term corresponding to a certain discrete labeling since the resulting deformation depends on all grid nodes influencing the local image patch. Therefore, we propose an approximation scheme for an efficient pre-computation of the look-up table containing the $|\mathcal{L}| \times |\mathcal{G}|$ data terms. For a given image pair this computation can be achieved by simple global shift operators. The entries of one table row (corresponding to a certain label u_p) are determined simultaneously by translating the source image with the associated displacement vector $\mathbf{d}^{u_p}$. The entry for node p and label u_p is determined by

$$V_p(u_p) = \int_\Omega \eta^{-1}(|\mathbf{x} - \mathbf{p}|) \rho_h(g(\mathbf{x}), f(\mathbf{x} + \mathbf{d}^{u_p}))d\mathbf{x}. \tag{13}$$

Thanks to the use of the inverse projection and the local grid node support on the deformation field, the error that is made during this approximation is relatively small. Since image points which are close to a grid node and mostly affected by its displacement will also have the most influence when back-projecting the dissimilarity measure.

Unfortunately, the mentioned approximation scheme including the inverse projection can only be used for point-to-point dissimilarity measures. For more complex measures, such as Mutual Information or Cross Correlation, the inverse projection cannot be used. Here, we can simply approximate these measures in the local neighborhood of a grid node. In case of Mutual Information for instance, we generate a local joint histogram from the neighboring image patches next to certain control point.

Since the computation of the data term is only based on global translations it is very fast and straightforward. Additionally, it allows to plug in any dissimilarity measure without modifications on the scheme itself. The measures are only considered on the image domain and no further analytical differentiation is needed.

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 updating the dense deformation field in a compositional fashion) which allow us to keep the set of labels quite small and thus the optimization to be fast. This iterative approach improves the accuracy of the results since the data term approximation gets better and better after updating the deformation after each cycle. Furthermore, in case that an optimal grid node displacement is not covered by the capture range of the first cycle we can still converge to the optimal position. Additionally, we can successively refine the quantized deformation space to enable high sub-pixel accuracy. To this end, we can define a series of cost functions, or

$$E_{\text{data}}^t(u) = \sum_{p \in \mathcal{G}} \iint_{\Omega} \eta^{-1}(|\mathbf{x} - \mathbf{p}|)\rho_h(g(\mathbf{x}), f(\mathbf{d}^{u_p} \circ \mathcal{T}^{t-1}(\mathbf{x})))d\mathbf{x}. \tag{14}$$

We should note, that from an optimization point of view we achieve (quasi) optimal solutions for the discrete labeling in every cycle.

Recently, one certain kind of deformations gained quite a lot of interest. In some applications, e.g. where the deformation field itself is further analyzed, it is desirable to obtain smooth, invertible deformations called diffeomorphisms. Following Rueckert et al. [29], we can guarantee diffeomorph deformations by simply setting the maximum allowed displacement to the bounds derived in [29].

The next aspect to be addressed is the definition of the smoothness term in the label domain. One can express distances between the deformation vectors using differences between labels if a ranking has been considered within the definition of the label set, or

$$E_{\text{smooth}}(u) = \frac{1}{|\mathcal{E}|} \sum_{p,q \in \mathcal{E}} V_{pq}(u_p, u_q) \tag{15}$$

where $\mathcal{E}$ represents the neighborhood system associated with the deformation grid $\mathcal{G}$. For the distance $V_{pq}(\cdot,\cdot)$ (also called pairwise potentials) we consider a simple piecewise smoothness truncated term based on the euclidean geometric distances between the deformations corresponding to the assigned labels:

$$V_{pq}(u_p, u_q) = \lambda_{pq} \min\left(|\mathbf{d}^{u_p} - \mathbf{d}^{u_q}|, T\right) \tag{16}$$

with T being the maximum penalty and λ_{pq} being a (spacially varying) weighting to control the influence of this prior term. Basically, this is a discrete approximation of the smoothness term defined in Equation 9 extended by the piecewise property. We propose two ways of incorporating the smoothness function in the optimization process. Incremental regularization [4] of the transformation in each cycle independently yields a fluid-like registration [11]. This has the advantage that the pairwise potentials always fulfill the metric properties which is needed for MRF optimization algorithms such as the α-expansion algorithm [3]. Full regularization over time can be achieved by back-projecting the dense deformation field to the (reseted) grid nodes in a similar way as it is done for the data term. Here, an optimization method is needed which can handle semi-metrics which is the case for the below mentioned algorithm based on linear programming.

Such a smoothness term together with the data term allows to convert the problem of image registration into the form of an MRF [23] in a discrete domain, or

$$E_{\text{total}}(u) = \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} V_p(u_p) + \frac{1}{|\mathcal{E}|} \sum_{p,q \in \mathcal{E}} V_{pq}(u_p, u_q). \tag{17}$$

MRFs [12] have been very popular in the area of computer vision in late eighties and 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 the local minimum. On the other hand, techniques like simulated annealing [19] can in theory drive the solution to the optimal one, however in practice the process was rather complicated and important attention was to be paid on the handling of the temperature decrease. This constrain have made the use of annealing methods almost impractical.

The use of the max-flow/min-cut algorithm [10] and the prove of equivalence with certain MRFs was the main reason of renaissance for the MRF framework, in 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 communities in MRFs. This method can guarantee the global optimum or a good approximation of it (solving a succession of binary problems using the alpha-expansion [2]) under certain conditions [20] which related 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 is the optimization of the objective function in reasonable computational time. The use of metric or sub-modular functions is the most common constraint related with 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 computational efficient. Due the pairwise potentials requirements, the use of methods like graph-cuts, or other max-fow/min-cut is limited.

3 MRF optimization based on Linear Programming

For optimizing the resulting MRF, we seek to assign a label $u_p \in \mathcal{L}$ to each node $p \in \mathcal{G}$, so that the MRF energy in (17) is minimized. 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. When applied to the image registration task, this technique thus offers a series of important advantages compared to prior art (see Section 3.2).

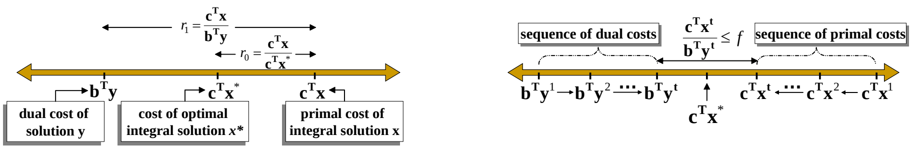
For more details about the Fast-PD algorithm, the reader is referred to [22,21]. Here, we will just provide a brief, high level description of the basic driving force behind that algorithm. This driving force will consist of the *primal-dual schema*, which is a well-known technique in the linear programming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

$$\begin{array}{llll} \textsc{Primal:} & \min \mathbf{c}^T\mathbf{x} & \qquad \textsc{Dual:} & \max \mathbf{b}^T\mathbf{y} \\ & \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} & & \text{s.t.} \quad \mathbf{A}^T\mathbf{y} \leq \mathbf{c} \end{array} \tag{18}$$

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



(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 $\le f$), so are $\mathbf{c}^T\mathbf{x}^*$ and $\mathbf{c}^T\mathbf{x}$ (e.g. their ratio r_0 is $\le f$ 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 $\le f$). We can then apply the primal-dual principle (as in Fig. (a)) and thus conclude that $\mathbf{x}^t$ is an f-approximation to $\mathbf{x}^*$.

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.:*

$$\mathbf{c}^T\mathbf{x} \le f \cdot \mathbf{b}^T\mathbf{y}, \tag{19}$$

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

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 (19) 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 (17) has been used as the primal problem:

$$\min \sum_{p\in\mathcal{G}} \sum_{l\in\mathcal{L}} V_p(l)x_p(l) + \sum_{(p,q)\in\mathcal{E}} \sum_{l,l'\in\mathcal{L}} V_{pq}(l,l')x_{pq}(l,l') \tag{20}$$

$$\text{s.t.} \sum\nolimits_l x_p(l) = 1 \qquad \forall\, p \in \mathcal{G} \tag{21}$$

$$\sum\nolimits_l x_{pq}(l,l') = x_q(l') \qquad \forall\, l' \in \mathcal{L},\ (p,q) \in \mathcal{E} \tag{22}$$

$$\sum\nolimits_{l'} x_{pq}(l,l') = x_p(l) \qquad \forall\, l \in \mathcal{L},\ (p,q) \in \mathcal{E} \tag{23}$$

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

Here, in order to linearize the MRF energy, we have replaced the discrete variables u_p with the binary variables $x_p(\cdot)$ and $x_{pq}(\cdot,\cdot)$. More specifically, the $\{0,1\}$-variable $x_p(l)$ indicates that node p is assigned label l (i.e., $u_p = l$), while the $\{0,1\}$-variable $x_{pq}(l,l')$ indicates that vertices p,q are assigned labels l,l' respectively (i.e., $u_p = l$, $u_q = l'$). Furthermore, the constraints in (21) simply express the fact that each node must receive exactly one label, while constraints (22), (23) maintain consistency between variables $x_p(\cdot), x_q(\cdot)$ and variables $x_{pq}(\cdot,\cdot)$, in the sense that if $x_p(l) = 1$ and $x_q(l') = 1$ holds true, then these constraints force $x_{pq}(l,l') = 1$ 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 $x_p(\cdot) \geq 0, x_{pq}(\cdot,\cdot) \geq 0$), 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 $f{=}2\frac{d_{\max}}{d_{\min}}$[1] as the approximation factor in (19).

3.2 Advantages of the primal-dual approach

Fast-PD has many nice properties, which makes it a perfect 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 $V_{pq}(\cdot,\cdot) \geq 0$. Hence, by using Fast-PD, our image registration framework can automatically incorporate any dissimilarity 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.

[1] $d_{\max} \equiv \max_{a\neq b} d(a,b)$, $\quad d_{\min} \equiv \min_{a\neq b} d(a,b)$

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 while at the same time we use a Gaussian pyramid for the image data. As before mentioned, the deformation grid is reseted after each optimization cycle and the resulting displacement fields are concatenated in a compositional fashion. The resolution of the control grid depends on the image dimensions as well as the present deformations. However, a general setup of starting with 20 mm grid spacing and refinements to 10 and 5 mm is usually sufficient for many applications. Even if we usually assume a rigid or affine pre-registration of the image data, the use of a very coarse grid can efficiently compensate for global anisotropic scaling, translation and shearing in advance to the local registration.

We mentioned the crucial part of defining the set of labels and the corresponding quantized version of the displacement space. Setting these parameters can for instance be done by visually inspecting the images and the present deformations. The user has to define the maximum displacement [2] $\mathbf{d}^{\max}$ and the number of sampling steps N in each dimension. In 2D we discretize the continuous domain within the rectangular range defined by the maximum displacement and thus we form a set of $(2N+1) \times (2N+1)$ labels (including the zero-displacement vector). In 3D we use a sparse sampling along the six main directions and thus we form a set of $6N + 1$ labels. In our software, we visualize the capture range at every grid node such that the user can make sure that the important deformations are covered. Another parameter controls the capture range refinement after each cycle which is defined as a simple scaling of $\mathbf{d}^{\max}$. The refinement allows us to keep the number of sampling steps quite small while achieving sub-pixel accuracy. In practice it turned out that setting $N = 5$ is mostly sufficient (resulting in 121 labels for 2D, and 31 labels for 3D). We think that this intuitive adjustment of the space of solutions turns out to be another advantage compared to gradient-descent approaches where the user can hardly control the search space. Additionally, prior information on the present deformations can be easily incorporated in our framework.

The last important parameter regards the regularization control. In order to make this parameter more independent from the input data (and its absolute intensity values) we can use intensity normalization of an image pair. Still, tuning is sometimes needed depending on the used dissimilarity measure.

We believe that besides the need for fast and efficient deformable registration

[2] $\mathbf{d}^{\max}$ can also be set automatically to the bounds for diffeomorph transformations of 0.42 times the control point spacing [29].

methods in medical applications with hard time constraints such methods are also extremely valuable in the stage of development and for parameter tuning. Our framework can provide direct visual feedback to the user at run time thanks to its efficiency. We will give more details about the running times in the experiments section.

5 Experiments on Known and Unknown Deformations

In order to demonstrate the flexibility of our framework, we implemented a range of well-known dissimilarity measures [3], namely the Sum of Absolute Differences (SAD) [16], the Sum of Squared Differences (SSD) [16], the Normalized Cross Correlation (NCC) [16], the Normalized Mutual Information (NMI) [24], the Correlation Ratio (CR) [26], and a measure involving an intensity-based and a geometric-based term which combines the Sum of Absolute Differences and image gradient information (GRAD). An additional weighting factor γ is used to control the influence of these two terms. The GRAD is defined as

$$\begin{aligned} \rho_{\mathrm{GRAD}}(g(\mathbf{x}), f(\mathcal{T}(\mathbf{x}))) = (1-\gamma)|g(\mathbf{x}) - f(\mathcal{T}(\mathbf{x}))| + \\ + \gamma \left(\frac{\nabla g(\mathbf{x})}{|\nabla g(\mathbf{x})|} \cdot \frac{\nabla f(\mathcal{T}(\mathbf{x}))}{|\nabla f(\mathcal{T}(\mathbf{x}))|} \right). \end{aligned} \tag{24}$$

Note that, by setting $\gamma = 1$, this dissimilarity measure can also be used for multi-modal registration.

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 hopefully illustrating the great potentials of our approach.

5.1 Realistic Synthetic Registration

The first experiments are concerning the nature of the free choice of the dissimilarity measure inherent in our framework. In order to evaluate the efficiency of different measures we test our method on simulated realistic data. The target image is generated from the 2D MRI source image by randomly displacing the control points of an superimposed FFD grid. For the experiments we use three different grid resolutions in order to simulate different degrees of deformations

[3] We use the term dissimilarity measure for consistency reasons since any similarity measure (e.g. $1 - NMI$) can be converted to a dissimilarity.

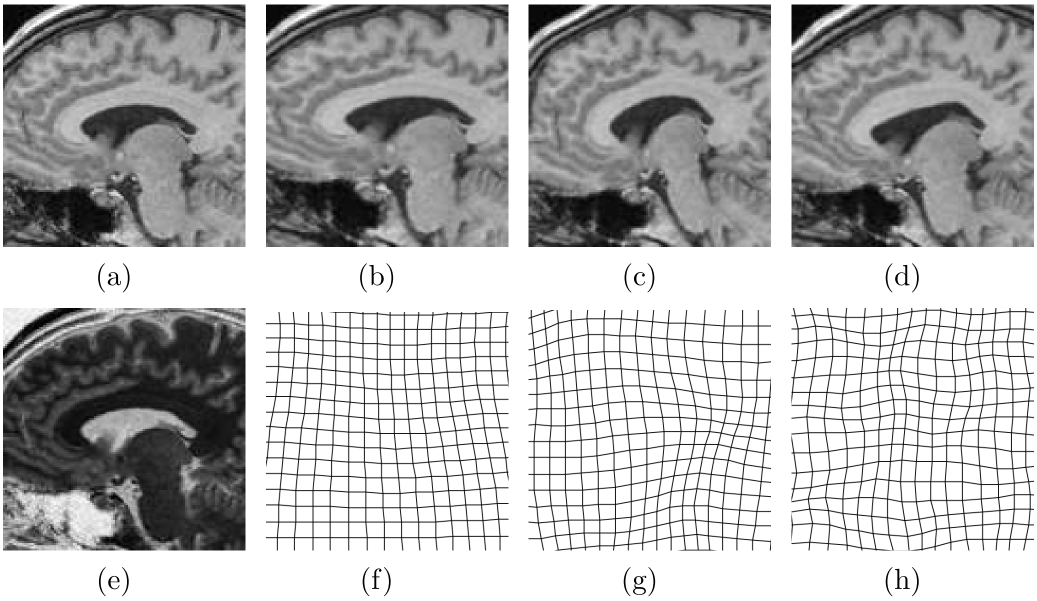


Fig. 2. Realistic synthetic data. (a) The source image, and (b)-(d) generated target images with different degree of deformation. (e) The inverse squared source image used for multi-modal tests. (f)-(h) Deformation fields corresponding to the upper target images.

Measure	AE Mean	AE Median	AE Std	MOD Mean	MOD Median	MOD Std
SSD	1.16	0.38	6.12	0.12	0.06	0.38
SAD	1.16	0.35	6.43	0.10	0.05	0.33
GRAD $\gamma = 0.5$	0.95	0.39	3.70	0.10	0.06	0.21
GRAD $\gamma = 1.0$	0.78	0.49	1.10	0.10	0.08	0.10
CR	0.75	0.38	2.19	0.09	0.06	0.21
NCC	0.73	0.42	1.82	0.08	0.06	0.09
NMI	0.57	0.39	0.67	0.07	0.06	0.05
GRAD $\gamma = 1.0$	2.22	0.58	4.56	0.72	0.09	2.02
CR	1.00	0.48	3.07	0.12	0.07	0.35
NMI	0.61	0.43	0.61	0.08	0.07	0.05

Table 1
60 mm control point spacing and a maximum displacement of 10 mm used for target generation (see Fig. 2b). The upper part shows the results for the mono-modal experiment, the lower part is for the multi-modal experiment.

(see Fig. 2). We should note that none of these resolutions is later used within the registration. We also perform an experiment where the three multi-modal measures are compared namely the NMI, CR, and GRAD ($\gamma = 1.0$). Therefore, we use the inverse squared source image (see Fig. 2e) such that no linear relation between the intensities is given.

The image resolution in the first experiments is 256×256 pixels with an isotropic pixel spacing of 1 mm. The registration is performed using a three-

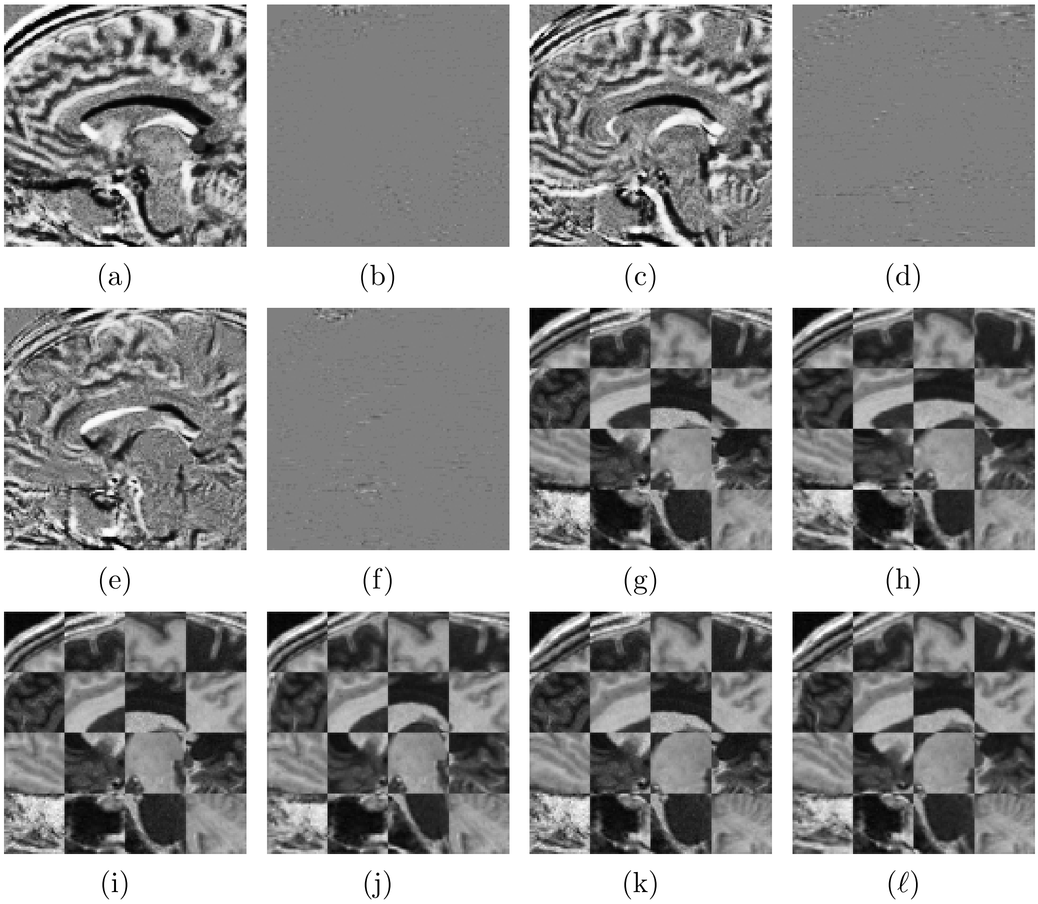


Fig. 3. Exemplary results for the realistic synthetic data experiment. (a)-(f) Difference images before and after mono-modal registration using SAD. (g)-(ℓ) Checkerboard visualization before and after multi-modal registration using NMI.

level image and grid pyramid. For the grid we start with a control point spacing of 20 mm, successively refined to 10 and 5 mm. The maximum displacement $\mathbf{d}^{\max}$ for each level is set to the bounds for diffeomorph transformations [29] while using 5 sampling steps in each direction (121 labels in total). The label refinement parameter is set to 0.5 meaning that the displacement range is halved after each cycle. On each level we perform 5 optimization cycles.

One complete registration takes between 5-60 seconds depending on the used dissimilarity measure (SAD is the fastest, NMI the slowest). The results are shown in Tab. 1-3. For the evaluation, two error metrics are considered, namely the angular error (AE) [8] measured in degrees and the magnitude of difference (MOD) measured in millimeters. We only consider the deformation field within a region of interest containing the brain tissue. In the mono-modal experiment, all of the used measures are able to recover the unknown transformations quite well. For multi-modal the NMI performs best in our experiments. However, it turned out that the GRAD measure purely based on gradient information is also usable in the case of our multi-modal setting. Note that the computation

Measure	AE Mean	AE Median	AE Std	MOD Mean	MOD Median	MOD Std
NCC	1.81	0.91	5.13	0.12	0.08	0.26
SSD	1.72	0.95	3.73	0.12	0.08	0.22
SAD	1.64	0.85	3.89	0.11	0.08	0.22
CR	1.47	0.88	3.03	0.10	0.08	0.12
GRAD $\gamma = 0.5$	1.47	0.84	2.74	0.11	0.08	0.21
GRAD $\gamma = 1.0$	1.36	0.87	2.00	0.10	0.08	0.08
NMI	1.05	0.71	1.11	0.08	0.07	0.05
GRAD $\gamma = 1.0$	3.51	0.98	12.58	0.22	0.09	0.73
NMI	2.33	0.84	9.10	0.15	0.08	0.49
CR	2.20	1.00	7.37	0.15	0.09	0.37

Table 2
30 mm control point spacing and a maximum displacement of 7.5 mm used for target generation (see Fig. 2c). The upper part shows the results for the mono-modal experiment, the lower part is for the multi-modal experiment.

Measure	AE Mean	AE Median	AE Std	MOD Mean	MOD Median	MOD Std
GRAD $\gamma = 1.0$	4.64	1.86	11.97	0.20	0.09	0.55
GRAD $\gamma = 0.5$	4.08	1.47	12.57	0.20	0.07	0.64
NCC	3.29	1.57	9.79	0.16	0.07	0.51
CR	2.83	1.60	6.31	0.13	0.08	0.30
SSD	2.68	1.37	7.54	0.13	0.06	0.45
SAD	2.16	1.18	5.84	0.10	0.06	0.34
NMI	1.66	1.24	1.72	0.07	0.06	0.08
CR	3.20	1.80	6.70	0.17	0.09	0.52
GRAD $\gamma = 1.0$	2.69	1.83	2.69	0.11	0.09	0.09
NMI	2.42	1.45	5.98	0.10	0.07	0.27

Table 3
15 mm control point spacing and a maximum displacement of 5 mm used for target generation (see Fig. 2d). The upper part shows the results for the mono-modal experiment, the lower part is for the multi-modal experiment.

of the GRAD measure is slightly slower than the SAD and much faster than NMI or CR.

5.2 Inter and Intra Subject Registration

The next experiment is aiming at the registration accuracy. An automatic segmentation should be performed by mapping a template segmentation to an image to be segmented. The mapping is done by registration of the original intensity images. The recovered transformation is then used to warp the template segmentation and compared to manual expert segmentation. For the

Image	OR	Sens	Spec
1	0.6959 / 0.8752 (26%)	0.7466 / 0.8752 (17%)	0.9650 / 0.9847 (2%)
2	0.7188 / 0.8365 (16%)	0.7082 / 0.8365 (18%)	0.9796 / 0.9894 (1%)
3	0.6964 / 0.8053 (16%)	0.6402 / 0.8053 (26%)	0.9821 / 0.9878 (1%)
4	0.7261 / 0.8927 (23%)	0.7803 / 0.8927 (14%)	0.9718 / 0.9851 (1%)
5	0.6959 / 0.8672 (25%)	0.7252 / 0.8672 (20%)	0.9695 / 0.9828 (1%)
6	0.7001 / 0.8012 (14%)	0.6214 / 0.8012 (29%)	0.9854 / 0.9901 (0.5%)
7	0.6900 / 0.8601 (25%)	0.6554 / 0.8601 (31%)	0.9779 / 0.9867 (1%)

Table 4
Brain registration results for gray matter tissue. The left values are the results after using an affine registration. The right values are the results after using our method. In brackets is the improvement in percentage.

Image	OR	Sens	Spec
1	0.6426 / 0.8390 (31%)	0.6109 / 0.7969 (30%)	0.9845 / 0.9945 (1%)
2	0.6201 / 0.8006 (29%)	0.6236 / 0.8018 (29%)	0.9863 / 0.9929 (1%)
3	0.6336 / 0.7913 (25%)	0.5997 / 0.8040 (34%)	0.9882 / 0.9908 (0%)
4	0.6839 / 0.8440 (23%)	0.6831 / 0.8174 (20%)	0.9867 / 0.9950 (1%)
5	0.6382 / 0.8373 (31%)	0.6062 / 0.8244 (36%)	0.9857 / 0.9929 (1%)
6	0.6526 / 0.7869 (21%)	0.6486 / 0.8762 (35%)	0.9882 / 0.9877 (0%)
7	0.6411 / 0.8289 (29%)	0.5865 / 0.8293 (41%)	0.9888 / 0.9921 (0%)

Table 5
Brain registration results for white matter tissue. The left values are the results after using an affine registration. The right values are the results after using our method.

following experiments, we set the maximum displacement corresponding to the diffeomorphism bounds while scaling the range after each cycle by 0.75. The deformation space is sub-sampled in 5 steps in each of the six main directions (yielding 31 labels in total). We perform 5 optimization cycles per pyramid level. The warped template segmentation is compared to the manual expert segmentation by computing the overlap ratio (OR), sensitivity, and specificity. The comparison of the segmentations is done using the tool [4] described in [13].

The first set of data targets an inter subject segmentation task of MRI brain images. The data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. MR images often vary a lot in the range of intensities depending on the used protocol and scanner device. Thus, intensity based inter subject registration becomes a difficult task. In this experiment, 8 data sets are available each with gray and white matter segmentations performed by experts. The image resolution is $256 \times 256 \times 128$ with a voxel spacing of $0.9375 \times 0.9375 \times 1.5$ mm. We use the NCC measure and

[4] Available on http://www.ia.unc.edu/dev/download/valmet/

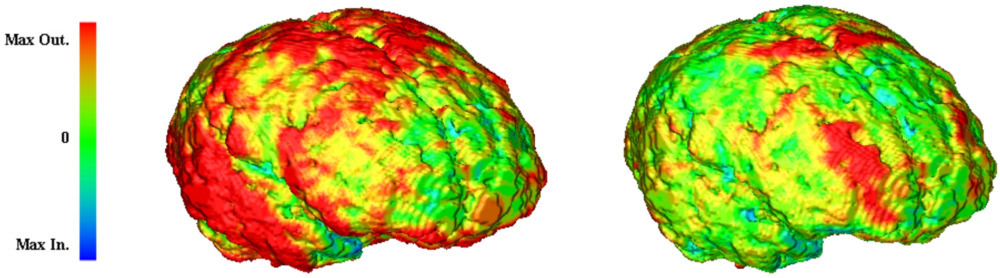


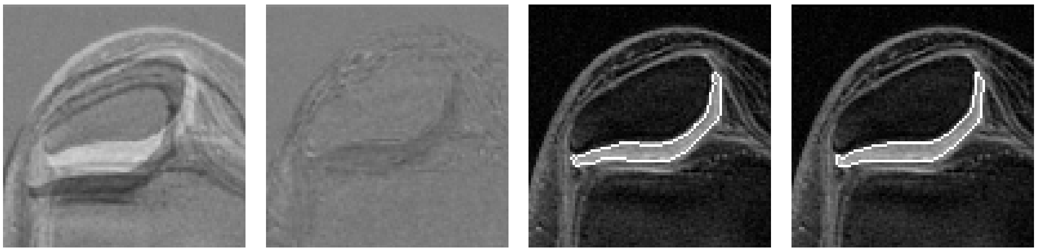
Fig. 4. Color encoded visualization of the surface distance between warped template and expert segmentation after affine (left) and deformable (right) registration using our method.

Image	OR	Sens	Spec	ASD
1	0.913	0.924	0.999	0.229
2	0.904	0.918	0.999	0.248
3	0.915	0.942	0.999	0.210
4	0.908	0.946	0.999	0.237
5	0.886	0.905	0.999	0.282
6	0.893	0.922	0.999	0.280

Table 6
Results for the cartilage segmentation experiment.

a four-level image pyramid with 40, 20, 10, and 5 mm control point spacing. A significant improvement of the segmentation could be achieved compared to affine registration. The results are shown in Tables 4 and 5. Here, one deformable registration using our framework takes about 550 seconds.

The medical application in the second segmentation task is similar to the one described in [9]. An automatic segmentation of the cartilage should be performed. Assuming that manual segmentations are available, one may create statistical models for an atlas-based segmentation procedure [14]. In our experiment, 7 data sets ($256 \times 256 \times 20$ voxels / $0.625 \times 0.625 \times 3.0$ mm spacing), all manually segmented by medical experts, are available. The MRI data was acquired for a follow-up experiment. We are able to achieve a fully automatic segmentation of one data set in less than 30 seconds. We use a three-level pyramid (20, 10, and 5 mm control point spacing) and the SAD measure. The registration is restricted to a narrow band of interest obtained by the template segmentation. After registration we achieve an average OR of 0.903, an average sensitivity and specificity of 0.926 and 0.999, and an average surface distance (ASD) of 0.248 mm (see also Table 6 and Fig. 5). In [14], we also present an approach for creating a statistical cartilage model using our framework for atlas-based segmentation.



(a) (b) (c) (d)

Fig. 5. Exemplary slice for the MRI cartilage registration. (a) Difference image before and (b) after registration. (c) In comparison, the overlay of the the warped template segmentation and (d) the expert segmentation.

5.3 Comparison to State-of-the-Art.

Schnabel et al. [31] propose a non-rigid image registration method [5] based on B-Spline FFD together with a gradient-descent optimization. Their approach can be seen as the state-of-the-art in FFD based registration. In order to obtain meaningful comparable results we try to set the registration parameters as similar as possible. Both algorithms are using the same deformation model and the SSD measure. The test data are two CT volumes showing the heart of a pig. The image resolution is $128 \times 128 \times 88$ with a voxel size of $0.848 \times 0.848 \times 1.25$ mm. Due to the heart beat a deformation of the shape is clearly visible. We run both methods on a deformation grid with 10 mm control point spacing. Within the region of interest enclosing the heart and an average SSD error of 12278 before registration, we achieve an average SSD error of 3180, where the other method converges to a value of 3402. Also, by visual perception of the difference images we can achieve better results (see Fig. 6). Last but not least, the running time of our algorithm is less than 2 minutes in contrast to a running time of more than 2 hours for the other method (AMD Athlon64 2.21 GHz). We should note, that this experiment was not performed to obtain the best registration of the two data sets, but rather to compare the two algorithms. With our standard pyramidal approach we obtain a SSD error of 1233 by same running time of about 2 minutes.

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

[5] Available on http://wwwhomes.doc.ic.ac.uk/∼dr/software/

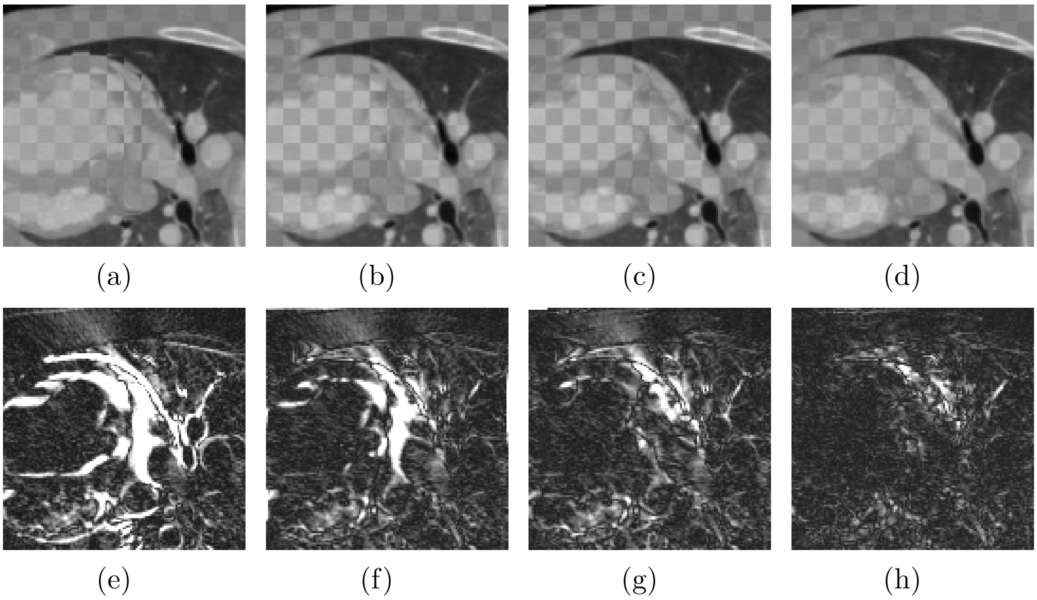


Fig. 6. (a) Checkerboard visualization before registration, (b) after registration using the method in [31], (c) after registration using our method, and (d) after registration using our approach with a pyramidal settings. In the same order, the difference images are shown in (e)-(h).

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 singleton 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 cost function is needed, can encode any dissimilarity measure and can recover the optimal solution up to a bound.

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.

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] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, Nov. 2001.

[4] 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.

[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, 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] J. Folkesson, E. Dam, O. F. Olsen, P. Pettersen, and C. Christiansen. Automatic segmentation of the articular cartilage in knee mri using a hierarchical multi-class classification scheme. In *Medical Image Computing and Computer-Assisted Intervention*, October 2005.

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

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

[12] 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.

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

[14] 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.

[15] 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.

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

[17] P. Hellier and C. Barillot. Coupling dense and landmark-based approaches for nonrigid registration. *IEEE Medical Imaging*, 22(2), Feb. 2003.

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

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

[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. A new framework for approximate labeling via graph cuts. In *IEEE International Conference on Computer Vision*, 2005.

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

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

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

[25] 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.

[26] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention*, 1998.

[27] 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. *IEEE Medical Imaging*, 22(6), 2003.

[28] 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.

[29] 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.

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

[31] 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*, 2001.

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

[33] 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.