A Discrete Search Method for Multi-modal Non-Rigid Image Registration

Alexander Shekhovtsov Juan D. García-Arteaga Tomáš Werner
Center for Machine Perception, Dept. of Cybernetics
Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic
shekhovt@cmp.felk.cvut.cz garcia@fel.cvut.cz werner@cmp.felk.cvut.cz

Abstract

We consider the problem of image matching under the unknown statistical dependence of the signals, i.e. *a signal in one image may correspond to one or more signals in the other image with different probabilities. This problem is widely known as multimodal image registration and is commonly solved by the maximization of the empirical mutual information between the images. The deformation is typically represented in a parametric form and optimization w.r.t. it is performed using gradient-based methods. In contrast, we represent the deformation as a field of discretized displacements and optimize w.r.t. it using pairwise Gibbs energy minimization technique. This has potential advantage of finding good solutions even for problems having many local minima. In experiments we demonstrate that the proposed method working on a single scale achieves comparable performance to a state-of-the-art multi-scale method.*

1. Introduction

Image registration is defined as the process of finding the correspondence between the points in two or more images of the same scene. Its application to medical images has been well studied [9, 13, 12, 11] and extensively used in the fusion of complimentary data acquired from multiple scanning devices, the comparison of images from different individuals, and the follow-up of the changes within the same individual at different times. When imaging biological structures, rigid deformations cannot, in general, compensate for morphological variations between individuals, the gradual changes over time or the natural elasticity of imaged tissues. In these cases, despite its inherent difficulty, non-rigid matching is the preferred image registration option.

The task of estimating the non-rigid deformation that best aligns two images is usually formulated as an optimization problem that minimizes an image dissimilarity criterion between the deformed (registered) images. The deformation, usually taken from a parametric family (*e.g.* B-Splines), is expected to realistically model the physical variations without introducing artifacts. As the deformation is typically described by a high number of parameters, the optimal solution is normally found by a differential or local search method [12]. Because the objective function comparing the images under the deformation has many local minima, it can only be rather poorly optimized by local methods. The problem becomes even more significant when additional parameters need to be estimated; specifically, we consider the case where the statistical dependence between intensities is unknown. Maximizing the likelihood of the observed data w.r.t. both the deformation and the parameters of color dependence model naturally yields an alternating optimization algorithm. Taking the non-parametric form of the unknown statistical color dependence it is known that this maximum likelihood approach is equivalent to registration by maximization of empirical mutual information [14].

In this article we propose to search for the best deformation aligning multi-modal medical images using, instead of local search methods, a global approximate method. Our approach is to consider the deformation as a combination of the local translations of small blocks with the constraint that deviation of translations of neighboring blocks is restricted.

Related Work. Kim *et al*. [7] applied mutual information criterion to solve the stereo correspondence problem. They searched for the best next estimate of the disparity field (deformation) by solving a discrete search problem in the form of pairwise Gibbs energy minimization. They showed that taking the first order Taylor approximation of the variant of empirical mutual information with Parzen window density estimator yields the same objective as alternating the estimation of the deformation and color statistical models in the maximum likelihood approach.

Glocker *et al*. [4] and Shekhovtsov *et al*. [18] applied linear programming relaxation to cope with difficult discrete search problem of the optimal 2D/3D deformation. While discrete optimization methods were proposed earlier for recovering nonrigid deformation [15, 1, 20], methods [4, 18] proposed several approximation tricks which made it realistically applicable for large deformations. While [4] considered normalized mutual information as a possible measure of image dissimilarity they discuss it just very briefly. In particular, it is not clear if it was applied to local windows independently, like normalized cross-correlation, or otherwise. Both [4] and [18] apply similar piecewise translational approximations of the deformation (moving blocks): [4] applies soft regularization on the deformation field and proposes a gradual multi-scale approach while [18] considers hard constraints and focuses on the efficient implementation of single-scale optimization.

Hufnagel *et al.* [5] applied a block matching scheme with independent blocks, with spline post-interpolation. This approach is similar to that of [4, 18], but the blocks are matched independently, meaning that they must be rather large to make a good match and this strongly restricts the class of deformations for which the approach is applicable.

Roche *et al.* [14] presented a systematic study of the image registration problem from the point of view of the maximum likelihood approach, in particular they showed that the non-parametric estimate of statistical color dependence (joint histogram) in maximum likelihood formulation leads to a maximization of the simplest variant of the empirical mutual information.

2. Probabilistic Model

We follow the standard computer vision approach and formulate the registration problem as a maximum a posteriori inference in a constructed probabilistic model. A detailed treatment of image registration within maximum likelihood approach may be found in [14], whereas we focus on a special case.

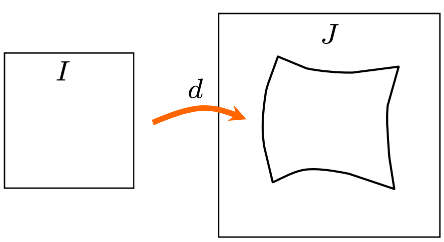

Figure 1. A non-symmetric formulation of registration of image I and image J.

Let I and J be two images to be registered. Let $I : T_1 \to \mathcal{C}_1$ and $J : T_2 \to \mathcal{C}_2$, where T_1 and T_2 are sets of locations and $\mathcal{C}_1$ and $\mathcal{C}_2$ are sets of signals. The deformation is represented by a mapping $d\colon T_1 \to T_2$. We assume that d is an injective mapping, meaning that not all pixels from T_2 will have a preimage. This is illustrated in Fig. 1. This non-symmetric formulation is convenient since it allows image I to be matched to a subregion of J.

Assuming that I and J are geometrically distorted observations of an unknown but equal underlying physical description, one derives (as in *e.g.* [14]) that there is a statistical dependence of how a particular hidden element is imaged in I and how it is imaged in J. In other words, there is a statistical dependence between signals in the two images, which is modeled as $p(c_1, c_2; \theta)$, $c_1 \in \mathcal{C}_1$, $c_2 \in \mathcal{C}_2$, where θ is a parameter defining this distribution.

Treating images I, J and the mapping d as random variables, we model their statistical dependence as

$$p(I, d \mid J) = p(I \mid d, J)p(d). \tag{1}$$

Assuming that pixel signals in I are conditionally independent given d and J, we write

$$p(I \mid d, J) = \prod_{t \in T_1} p(I_t \mid d, J). \tag{2}$$

When pixel t is mapped to pixel $d(t)$ we consider that its signal I_t depends only on the corresponding signal $J_{d(t)}$ and does not depend on the rest of signals in J. This assumption is expressed probabilistically as

$$p(I_t \mid d, J) = p(I_t \mid J_{d(t)}; \theta). \tag{3}$$

Prior distribution $p(d)$, detailed later, expresses how likely is deformation d. In particular, it is often assumed that highly curved deformations are unlikely.

The maximum likelihood estimate of deformation d and parameters θ is formulated now as

$$(d^*, \theta^*) = \operatorname*{argmax}_{d,\theta} \prod_{t \in T_1} p(I_t \mid J_{d(t)}; \theta)p(d). \tag{4}$$

We consider that sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are finite. This could mean that signals in I (resp. J) would be quantized into $|\mathcal{C}_1|$ (resp. $|\mathcal{C}_2|$) bins, if the continuous signal spaces were one dimensional. Alternatively, we can deal with multi-dimensional signals by representing images I (resp. J) by $|\mathcal{C}_1|$ (resp. $|\mathcal{C}_2|$) representative examples (much like a color palette). Our model of the dependence $p(c_1, c_2; \theta)$ is then a Gaussian mixture model with $|\mathcal{C}_1||\mathcal{C}_2|$ Gaussians placed in each possible location (c'_1, c'_2) having diagonal covariance matrix. Formally,

$$p(c_1, c_2; \theta) = \sum_{c'_1, c'_2} \theta_{c'_1, c'_2} g_1(c_1, c'_1) g_2(c_2, c'_2), \tag{5}$$

where

$$g_k(c, c') = \frac{1}{Z_k} \exp\Big[-\frac{\rho_k(c, c')^2}{2}\Big], \quad k = 1, 2; \tag{6}$$

and $\rho_k(c, c')$ is a metric in the space $\mathcal{C}_k$, $k = 1, 2$. Metrics ρ_1 (resp. ρ_2) captures the important information of how close are two different elements of $\mathcal{C}_1$ (resp. C_2). A Gaussian distribution with a different variance can be obtained by changing the metric and the normalization constant accordingly. The metrics are assumed to be predefined and fixed. For (5) to define a distribution the constraint $\sum_{c_1',c_2'} \theta_{c_1',c_2'} = 1$ must hold. Finding the maximum likelihood estimate of parameters θ and evaluating the distribution $p(c_1, c_2; \theta)$ may be seen as very similar to Parzen window kernel density estimation with Gaussian kernel.

2.1. Relation to Mutual Information

Mutual information was simultaneously proposed as an image similarity criterion by Viola and Wells [19] and Collignon *et al.* [3]. Since then, it has rapidly become the preferred choice when registering multi-modal medical images (see the surveys by Pluim *et al.* [13] and Maes *et al.* [11]). Before we make any further assumptions and discuss our approach to the optimization problem (4) we will describe its relation to the registration by maximization of mutual information. The derivation is similar to the one presented in [14] and in [7].

Taking the logarithm of (4) we can write that:

$$d^* = \operatorname*{argmax}_d \Big[\log p(d) + \max_\theta \sum_{t \in T_1} \log p(I_t | J_{d(t)}; \theta) \Big]. \tag{7}$$

Introducing $n_{c_1,c_2} = \sum_{t \in T_1} \delta_{\{I_t = c_1\}} \delta_{\{J_{d(t)} = c_2\}}$ the maximum over θ with additional factor $\frac{1}{N}$ may be expressed as

$$\max_\theta \frac{1}{N} \sum_{\substack{c_1 \in \mathcal{C}_1 \\ c_2 \in \mathcal{C}_2}} n_{c_1,c_2} \log p(c_1 | c_2; \theta), \tag{8}$$

where $N = |T_1| = \sum_{c_1,c_2} n_{c_1,c_2}$. Denoting $\hat{\theta}$ a maximizer of (8), the objective of (8) expresses as

$$\frac{1}{N} \sum_{\substack{c_1 \in \mathcal{C}_1 \\ c_2 \in \mathcal{C}_2}} n_{c_1,c_2} \log \frac{p(c_1, c_2; \hat{\theta})}{p(c_2; \hat{\theta})} =$$
$$\frac{1}{N} \sum_{\substack{c_1 \in \mathcal{C}_1 \\ c_2 \in \mathcal{C}_2}} n_{c_1,c_2} \log p(c_1, c_2; \hat{\theta}) - \frac{1}{N} \sum_{c_2 \in \mathcal{C}_2} n_{c_2} \log p(c_2; \hat{\theta}) =$$
$$-\hat{H}(n_{\mathcal{C}_1,\mathcal{C}_2}) + \hat{H}(n_{\mathcal{C}_2}), \tag{9}$$

where $n_{c_2} = \sum_{c_1} n_{c_1,c_2}$ and $n_{\mathcal{C}_1,C_2}$ stands for the collection of numbers $\{n_{c_1,c_2} \mid c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2\}$ and n_{C_2} for the collection of numbers $\{\sum_{c1 \in \mathcal{C}_1} n_{c_1,c_2} \mid c_2 \in \mathcal{C}_2\}$. The term $\hat{H}(n_{C_1,C_2})$ then denotes the empirical estimate of the entropy from samples represented by counts n_{C_1,C_2}. The estimate is obtained by fitting parameters of the distribution and using the sample mean instead of the expectation. It should be noted that this is only one particular choice of how to estimate the entropy from samples and that other choices are possible.

Now problem (7) can be expressed as:

$$\begin{aligned} d^* &= \operatorname*{argmax}_d \Big[\log p(d) - \hat{H}(n_{\mathcal{C}_1,\mathcal{C}_2}(d)) + \hat{H}(n_{\mathcal{C}_2}(d)) \Big] \\ &= \operatorname*{argmax}_d \Big[\log p(d) + \hat{H}(n_{\mathcal{C}_1}) - \hat{H}(n_{\mathcal{C}_1,\mathcal{C}_2}(d)) \\ &\qquad + \hat{H}(n_{\mathcal{C}_2}(d)) \Big] \\ &= \operatorname*{argmax}_d \Big[\log p(d) + \hat{\mathrm{I}}(n_{\mathcal{C}_1,\mathcal{C}_2}(d)) \Big], \end{aligned} \tag{10}$$

where $\hat{H}(n_{\mathcal{C}_1})$ is the empirical entropy of n_{C_1} which does not depend on d, and $\hat{\mathrm{I}}(n_{\mathcal{C}_1,\mathcal{C}_2}(d))$ is the empirical mutual information. Allowing $p(d)$ to be uniform over a subset of deformations D (and zero outside this subset), it is

$$d^* = \operatorname*{argmax}_{d \in D} \hat{\mathrm{I}}(n_{\mathcal{C}_1,\mathcal{C}_2}(d)), \tag{11}$$

Thus our objective may be stated as the maximization of mutual information. We prefer to work with maximum likelihood formulation because it explicitly clarifies the assumptions about the model and allows a wider range of choices, *e.g.* to specify a nontrivial prior on d.

2.2. Optimization

We follow the standard approach to optimize (4): the objective is alternatively optimized w.r.t. d and θ. In this section we discuss the details of these two steps.

Under fixed d, maximization of the likelihood over θ is obtained by solving (8). For this problem we apply a simple iterative algorithm [6], which in our notation reads:

$$\theta^{\text{new}}_{c_1,c_2} = \theta^{\text{old}}_{c_1,c_2} \frac{1}{N} \sum_{c_1',c_2'} \frac{n_{c_1',c_2'}}{p(c_1', c_2'; \theta^{\text{old}})} g_1(c_1, c_1') g_2(c_2, c_2'). \tag{12}$$

The estimate is initialized to $\theta_{c_1,c_2} = \frac{n_{c_1,c_2}}{N}$, which is easily seen to be the Parzen window estimate, and then updated by several iterations of (12). It should be noticed that the improvement over the initial estimate is not very significant; (12) is used instead of the simple Parzen window estimate to be consistent with maximum likelihood formulation.

Under fixed θ, maximizing the log-likelihood over d reads

$$\operatorname*{argmax}_d \Big[\log p(d) + \sum_{t \in T_1} \log p(I_t | J_{d(t)}; \theta) \Big]. \tag{13}$$

We apply the method [18] to search for the best deformation. This method assumes that the deformation is a mapping $B \to K$, where B is the set of small blocks into which the image I is subdivided and K is the set of possible displacements each block is allowed to move to (see Fig. 2). The prior on d is modeled as a product of pairwise potentials over the set E of pairs of neighboring blocks:

$$p(d) = \frac{1}{Z} \prod_{bb' \in\, E} \phi_{bb'}(d_b, d_{b'}), \tag{14}$$

where $\phi_{bb'}(d_b, d_{b'})$ is 1 or 0 depending on whether the pair of displacements $(d_b, d_{b'})$ is allowed ($d \in D$) or not ($d \notin D$). This is a so-called *hard* model, in which normalization constant Z does not influence the results since all feasible deformations from D will have the same constant probability $\frac{1}{Z}$, and thus the optimal d will not depend on Z.

When model (14) is substituted into (13), the problem can be expressed as

$$d^* = \operatorname*{argmin}_d \Big[\sum_{b \in B} q_b(d_b) - \sum_{bb' \in E} \log \phi_{bb'}(d_b, d_{b'}) \Big], \tag{15}$$

where we denote $q_b(d_b) = -\sum_{t \in b} \log p(I_t \mid J_{d_b(t)}; \theta)$ and switch to minimization, following the convention of energy interpretation. Problem (15) is known as a minimization of pairwise Gibbs energy which is known to be hard. A linear programming relaxation approach is applied [18]. For the scope of this work it is important to notice that [18] takes explicitly into account that the problem has many local minima and a certain convex relaxation of it is solved (in a certain sense suboptimally, for the sake of speed). Nevertheless, we refer to this method as a global one. Note that unlike the approach of *e.g.* [5], there are pairwise constraints on the neighboring locations of blocks, implying that finding the best translations can only be done jointly and does not decompose into finding the best position for each block independently. After the best piece-wise translational deformation is found, we interpolate it to obtain a smooth deformation.

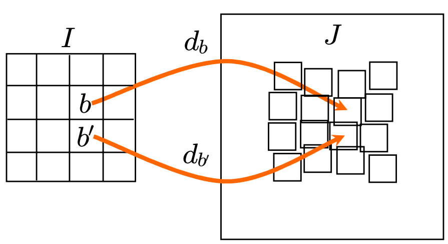

Figure 2. An approximation of deformation d as a piece-wise translational deformation. Image I is regularry subdivided into small blocks, each block $b \in B$ may be moved solidly by displacement d_b, a pair of neighboring blocks is constrained to have small relative displacement, thus preventing fold-overs and discontinuities in the deformation field.

3. Experiments

For a quantitative test we use two synthetic images shown in Fig. 3(b), which correspond to two axial cuts taken

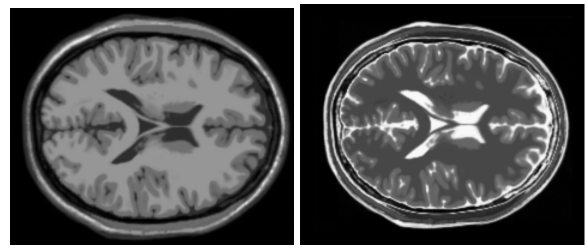

Figure 3. A synthetically generated [2] pair of multi-modal MRI images.

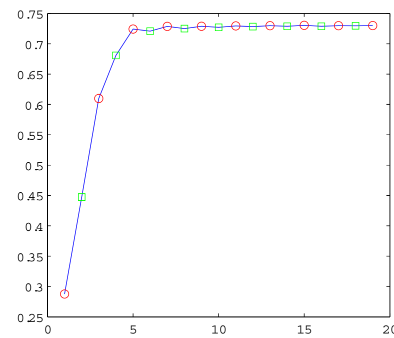

Figure 4. Convergence plot of the log likelihood objective function (empirical mutual information). Circles indicate optimization steps w.r.t. parameters θ of the statistical dependence of signals and boxes indicate optimization steps w.r.t. deformation d.

from a simulated MRI brain scan [2] with different relaxation times. Figure 5(last) shows that there is a non-trivial statistical dependence between colors of the two input images when they are perfectly aligned.

A typical run with second input image deformed like it is exemplified in Fig. 6 and both images corrupted by $\mathcal{N}(0, 0.1^2)$ noise (the signal space is $[0, 1]$) is as follows. The progress in log-likelihood objective during alternating optimization steps is shown in Fig. 4. Note that the plot starts by estimating the signal dependence parameters θ assuming that the initial deformation is the identity. It is seen that the method converges in a few alternating steps. Figure 5 shows the evolution of the conditional density estimate $p(c_1|c_c; \theta)$.

We do not show a visual comparison of registered images, since the registration is quite accurate and it is very difficult to see the differences directly. What we will be interested in instead is to measure quantitatively how the recovered deformation differs from the true deformation r. We tested two types of noise: Gaussian i.i.d. added to each pixel and a clutter-like noise, $\mathcal{B}(\sigma^2)$, which makes bright spots appear in random positions. Here, σ is the variance (in pixels) of the spatial Gaussian mask defining the size of the spots.

We have compared our method against an elastic method based on free-form deformations using B-splines and local

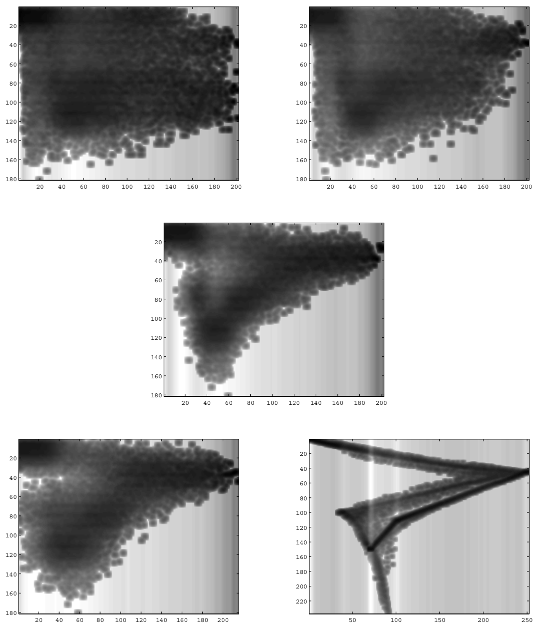

Figure 5. Estimates of the log conditional density $\log p(c_1 \mid c_2; \theta)$ during first 2 and the last iterations. Bottom: the estimate under the ground truth deformation and the estimate for noiseless images under ground truth deformation.

gradient descent optimization of mutual information implemented by the Image Registration Toolkit Software (ITK) [16, 17].

For each recovered deformation the maximum error was measured as

$$\mathrm{err} = \max_{t \in A} \|d_t - r_t\|, \tag{16}$$

which corresponds to the maximum deviation from the ground-truth deformation. The maximum is computed over the subset $A \subset T_1$, which excludes the dark background of the template original image. The deformation of the black background is, indeed, unrecoverable.

By measuring this error over 100 recovered deformations, we estimated the probability $P\{\mathrm{err} \geq \alpha\}$ that on a random sample the error will be α or greater. This probability equals $1 - P\{\mathrm{err} < \alpha\}$ and $P\{\mathrm{err} < \alpha\}$ is computed as empirical cumulative distribution function estimate. Results of our measurements are shown in Figs. 6 and 7. Note that although the errors may seem large (over 10 pixels for some examples), these are maximal errors, whereas average errors over pixels are shown in the Table 1. The registration by both methods align the images' intensities very well and it is left to the regularization prior to solve any ambiguity. Actually, if the ground truth error was not known for every pixel, the error in the estimated deformation would not be discovered. Our current implementation takes around 30 seconds for 317x281 pixel images.

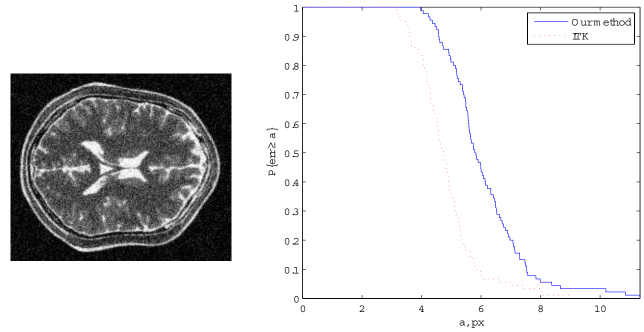

Figure 6. Left: an example of randomly deformed and corrupted by Gaussian noise of variance 0.1^2. The other input image is not deformed but perturbed with the noise as well. Right: the error statistics. For each value of the error the plot shows the estimated probability of the event that an error larger than that would be encountered on a randomly taken sample. Solid line shows our results and dashed like shows results of ITK [16, 17].

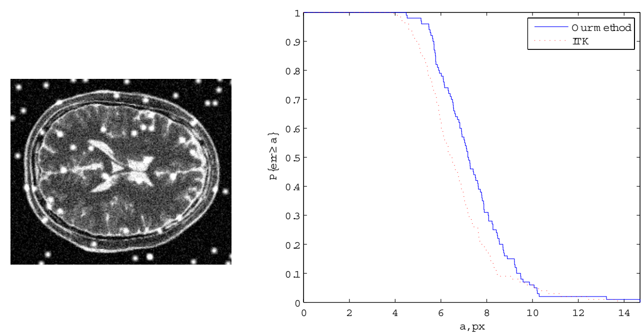

Figure 7. Similar to figure (6). Gaussian noise with variance 0.1^2 and clutter-type noise added to both input images. The deformed and corrupted by noise second image is shown on the left.

	$\mathcal{N}(0,\ 0.1^2)$		$\mathcal{B}(2^2) + \mathcal{N}(0,\ 0.1^2)$	
	our	ITK	our	ITK
AE Mean	0.193	0.14	0.201	0.198
AE Median	0.135	0.0908	0.147	0.135
AE Std	0.199	0.172	0.199	0.216
MOD Mean	0.562	0.419	0.591	0.577
MOD Median	0.443	0.307	0.484	0.453
MOD Std	0.441	0.382	0.433	0.458

Table 1. Statistics of Angular Error (AE, degrees) and Magnitude of Difference (MOD, pixels) errors for different amount of noise. Errors are computed over the area excluding the dark background of the template original image.

4. Discussion and Conclusions

We have presented a method for image registration based on discrete search techniques for mutual information and have given an empirical evidence of its usefulness as an alternative for local optimization of the mutual information criterion in multi-modal images. Tests on simulated 2D medical images show that the method is able to consistently recover large deformations in a robust manner and feasible time. When comparing the precision of the method to a specialized medical image registration system using local optimization (Image Registration Toolkit) the results show that the two systems have similar errors with ITK having a slightly better performance.

One of the problems with the method is that the approximation technique we used in the d-optimization step is of course not guaranteed to find a global optimum. Furthermore, reestimating the deformation sometimes decreases the objective. This happens because [18] finds new approximate optimal solution without taking into account the estimate from the previous step and thus is not guaranteed to improve it.

It should be noted that cubic B-spline based methods have a strong advantage by using physically-inspired regularization on the deformation field. An open research direction is to add to the current hard constraints on the deformation field a regularization term, *e.g.* a bending energy type of regularization, which may be defined using second derivatives of the deformation field. It is clear that with pairwise interactions only, our discrete energy function cannot model second derivatives. It is hence natural to consider higher order Gibbs energy models which could impose this prior.

5. Acknowledgements

The authors have been supported by European Commission grant 215078 (DIPLECS), the Czech government grant MSM6840770038, and Czech Science Foundation Project 102/07/1317. The Image Registration Toolkit used to register the images by free-form deformations was used under Licence from Ixico Ltd.

References

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

[2] C. CA, K. Kwan, and R. Evans. Brainweb: online interface to a 3-D MRI simulated brain database. *NeuroImage*, 5:425, 1997.

[3] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*, pages 263–274, 1995.

[4] 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*, pages 408–420, 2007.

[5] H. Hufnagel, X. Pennec, G. Malandain, H. Handels, and N. Ayache. Non-linear 2D and 3D registration using block-matching and b-splines. In *Bildverarbeitung fuer die Medizin*, pages 325–329, March 2005.

[6] M. Jones and D. Henderson. Maximum likelihood kernel density estimation. Research Report 05/01, The Open University, UK, and University of Newcastle, UK, January 2005.

[7] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *International Conference on Computer Vision*, pages 1033–1040, 2003.

[8] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *International Conference on Computer Vision*, pages 1018–1025, 2005.

[9] H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1):129–149, January 1999.

[10] H. Luan, F. Qi, Z. Xue, L. Chen, and D. Shen. Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. *Pattern Recognition*, 41(1):285–298, January 2008.

[11] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.

[12] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004, 2003.

[14] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology*, 11(1):71–80, 2000.

[15] S. Roy and M.-A. Drouin. Mrf solutions for probabilistic optical flow formulations. In *International Conference on Pattern Recognition*, volume 3, Septembre 2000.

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

[17] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–581, 2001.

[18] A. Shekhovtsov, I. Kovtun, and V. Hlavac. Efficient MRF deformation model for non-rigid image matching. In *IEEE Transactions on International Conference on Pattern Recognition*, 2007.

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

[20] J. Winn and N. Jojic. LOCUS: Learning object classes with unsupervised segmentation. In *International Conference on Computer Vision*, volume 1, pages 756–763, 2005.