Random Walks with Efficient Search and Contextually Adapted Image Similarity for Deformable Registration

Lisa Y.W. Tang and Ghassan Hamarneh

Medical Image Analysis Lab., Simon Fraser University, Burnaby, Canada
{lisat,hamarneh}@sfu.ca

Abstract. We develop a random walk-based image registration method [1] that incorporates two novelties: 1) a progressive optimization scheme that conducts the solution search efficiently via a novel use of information derived from the obtained probabilistic solution, and 2) a data-likelihood re-weighting step that contextually performs feature selection in a spatially adaptive manner so that the data costs are based primarily on trusted information sources. Synthetic experiments on three public datasets of different anatomical regions and modalities showed that our method performed efficient search without sacrificing registration accuracy. Experiments performed on 60 real brain image pairs from a public dataset also demonstrated our method's better performance over existing non-probabilistic image registration methods.

1 Introduction

Deformable image registration (DIR) has many important applications, some of which include multi-modal fusion, atlas-construction, computer-aided intervention, and quantification of disease progression. A recent trend [1–5] to solving registration problems is to adopt a discrete optimization approach due to its potential benefits, e.g. efficient optimization (via primal/dual linear programming [3]) and the ability to obtain unique and globally optimal solution (via random walks [1], dynamic programming [4]).

The computational complexity of these methods, however, increases with the number of unknown variables and with the number of discrete values each variable can take. As remedies, researchers have employed a coarse-to-fine discretization of the search space [1, 3, 4] and/or reduced the allowed degrees of freedom of the solution via use of parametric transformation models (e.g. free-form deformations, FFD [2, 3]). However, schemes in [1–3] require iterative updates on the transformation, while [4] assumes tree-structured voxel connectivities and so enforces spatial regularizations only along edges of the tree.

In this paper, we develop a novel scheme for a random walk-based image registration (RWIR) that examines the solution search space efficiently. Starting from a coarse sampling of the deformation space (to obtain a label set of candidate displacements), we progressively and selectively add new samples (labels) to refine the initial sampling, using information extracted from the probabilistic registration solution. Our method makes no assumption on pixel connectivities and does not iteratively alternate between transformation update and image interpolation, as required in [2–4]. While standard

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 43–50, 2013.

coarse-to-fine sampling schemes (e.g. [2–4]) only refine displacements of selected (control) points and resample the search space *uniformly*, our method does not impose such restrictions, as we will explain in Sec. 2.

Related to our progressive search strategy are [6, 7]. In [7], a multi-scale energy pyramid (MSEP) scheme was proposed for minimizing Markov random field energies. Nevertheless, their approach only facilitates the optimization of an energy with a predefined label set and and does not examine the search space dynamically like we do. In [6], an iterative algorithm that refines the initial label set based on the measured uncertainty of the current solution was proposed. As reported in [8], however, estimating uncertainty in their obtained solutions requires additional expensive computation. Conversely, our approach employs information *already* available in the probabilistic solution from RWIR and so does not require expensive computation.

As our second contribution in this paper, we incorporate a *contextual* feature weighting strategy into RWIR. In contrast to a *learning* approach [5, 9] that requires training data, we employ information only derived from input data. Unlike [2], which selected a single subset of features over the *entire* image domain, our proposed scheme selects features in a *spatially adaptive* manner.

2 Methods

2.1 Efficient Solution Search for RWIR

Let I_a and I_b be discrete representations of two images in a d-dimensional domain $\Omega \subset \mathbb{R}^d$. DIR aims to find a spatial transformation $T : \Omega \mapsto \mathbb{R}^d$ that best aligns I_a and I_b, and is often done via minimization of an energy function of the form:

$$\arg\min_{T} \sum_{\mathbf{x}\in\Omega} \mathcal{D}(I_a(\mathbf{x}), I_b(T(\mathbf{x}))) + \alpha\mathcal{R}(T) \tag{1}$$

where $\mathcal{D}$ measures the dissimilarity between I_a and the transformed image I_b, $\mathcal{R}$ denotes a regularization term that encourages T to maintain certain smoothness properties and α is a weighting that balances between these two terms.

Under RWIR [1], the image grid is represented as a regular graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ where each pixel coordinate $\mathbf{x}_p$ is represented by a graph node $p \in \mathcal{V}$, $|\mathcal{V}| = V$; $\mathcal{E}$ is encoded by a Laplacian matrix $\mathbf{L}$. For a given sampling of the feasible solution space of $T(\mathbf{x})$ as represented by a set of K displacement labels $\mathcal{L} = \{\mathbf{v}_k\}$, $\mathbf{v}_k \in \mathbb{R}^d$ $(k = 1, \cdots, K)$, the image dissimilarity costs associated with $\mathcal{L}$ are encoded by a $V \times K$ matrix $\mathbf{D} = (\mathbf{D}_{pk})$, whose elements may be computed as:

$$\mathbf{D}_{pk} = e^{-\mathcal{D}(I_a(\mathbf{x}_p), I_b(\mathbf{x}_p+\mathbf{v}_k))}, \tag{2}$$

which maps low dissimilarities to high potential values. One then finds a probabilistic field $\mathbf{U} : \Omega \mapsto \mathbf{u}$ ($\mathbf{u} \in \Delta_K$, the K-dimensional unit simplex) that minimizes:

$$E(\mathbf{U}_k) = \sum_{j=1, j\neq k}^{K} \mathbf{U}_k^T \Lambda_j \mathbf{U}_k + (1 - \mathbf{U}_k)^T \Lambda_k (1 - \mathbf{U}_k) + \alpha \mathbf{U}_k^T \mathbf{L} \mathbf{U}_k, \tag{3}$$

where $\mathbf{U}_k$ denotes the k-th component of $\mathbf{U}$; Λ_j is a diagonal matrix with entries $[\mathbf{D}_{1k}, \cdots, \mathbf{D}_{Vk}]$. Briefly, the first and second term in (3) correspond to the data and regularization term in (1), respectively. When $\mathbf{D}$ is normalized to rows with unity sum, the RHS of Eq. (3) is positive definite and allows for a unique and global solution for $\mathbf{U}_k$ [1].

Critical to registration accuracy is $\mathcal{L}$ that defines the search space of T. In our approach, we start with a coarse initial set $\mathcal{L}$ and progressively examine the search space by extracting information from $\mathbf{U}$ to guide the generation of a new label set $\mathcal{L}_{refine}$ that would complement $\mathcal{L}$. By initializing $\mathcal{L}$ as a coarse sampling of the search space and then progressively adding new labels to $\mathcal{L}$ that are worthy of exploration, we reduce the computation needed for exploring, from the outset, the entire search space in a dense manner. Note that as we only enlarge the label set, each iteration of the progressive search will always result in a globally optimal solution with respect to the enlarged set, as opposed to *compositive schemes*, e.g. [2, 3], in which only the iterative updates are globally (sub-) optimal.

Central to the proposed method is a mechanism that estimates when and which new labels are added. In coarse-to-fine approaches [2–4], new labels are generated from a denser but *uniform* sampling of a smaller search space. This strategy is, however, inefficient as all unexplored regions may be considered. Instead, we employ the probabilistic solution $\mathbf{U}$ to identify new candidate labels that best explore the more probable regions of the search space. As we will show in Sec. 3, our progressive search is highly efficient because elements in $\mathcal{L}_{refine}$ are not restricted to be uniform samples of unexplored search regions. Specifically, for each node p, we examine its set of top J candidates S_p, as ranked by $\mathbf{U}_p$, and define a *confidence measure* $\mathcal{C}$ that quantifies the amount of agreement between the elements in S_p by estimating its probability-weighted inter-distances:

$$\mathcal{C}(\mathbf{U}_p) = \sum_{i=1}^{J} \sum_{j=1, j\neq i}^{J} f(\mathbf{U}_p, i) ||\mathbf{v}_{h(\mathbf{U}_p,i)} - \mathbf{v}_{h(\mathbf{U}_p,j)}||^{-1} \tag{4}$$

where $\mathbf{U}_p$ is a shorthand for $\mathbf{U}(p, \cdot)$; $h(\mathbf{U}_p, j)$ returns the index to the j-th top probable candidate at p and $f(\mathbf{U}_p, j)$ is the corresponding probability. Intuitively, node i has high measured confidence when its S_i forms a tight cluster around a region in the search space that is worthy of further exploration. Hence, we generate a refining label based on S_i by computing the vector mean[1] $\mathbf{m}_i$ of S_i, which corresponds to a uniform sampling of the convex hull of S_i. Visiting all nodes with high measured confidence and collecting corresponding $\mathbf{m}_i$ yields a new candidate label set $\mathcal{L}_{potential}$. Noting that the same label set is examined by all nodes and the complexity of solving (3) increases with K, we leverage the fact that $\mathcal{L}_{potential}$ typically includes many similar labels and bound K by selecting N representatives from $\mathcal{L}_{potential}$. To do so, we remove from $\mathcal{L}_{potential}$ elements that have low Euclidean distances to those already in $\mathcal{L}$ and then perform vector quantization (k-means) on $\mathcal{L}_{potential}$ to produce a compact set $\mathcal{L}_{refine}$ that has minimum expected distortion to $\mathcal{L}_{potential}$. We then rerun RWIR using the enlarged set $\mathcal{L} \leftarrow \mathcal{L}_{refine} \bigcup \mathcal{L}$. The procedure is repeated, generating $\mathcal{L}^n_{refine}$ after iteration n, and

[1] Using the mean is computationally efficient and achieves improved results. Other more computationally expensive strategies can be applied in the future, e.g. employ regression techniques to map $\mathbf{U}$ to target registration error, which is then used to steer the sampling process.

terminates when the average inter-distances between $\mathcal{L}^n_{refine}$ and $\mathcal{L}^{n-1}_{refine}$ is less than a user-defined threshold θ that governs the desired application-dependent precision.

2.2 Contextual Approach to Spatially Adaptive Feature Weighting for RWIR

When multiple sources of information are available, one can improve registration robustness and accuracy via multi-feature registration [2, 5, 9]. Many approaches have been proposed for optimizing a given feature set, some of which include optimal feature subset selection [2] and spatially adaptive feature weighting [5]. Here, we employ a *contextual* approach, as opposed to a *learning* approach [5, 9] that requires training data, which is generally challenging to collect. Specifically, let $\mathbf{D}^c$ denote the data cost matrix computed from component c that has been linearly scaled to unit variance. Then, RWIR employing F components can be performed by replacing $\mathbf{D}$ in (2) with $\mathbf{D}^{multi} = \phi^1 \circ \mathbf{D}^1 + \phi^c \circ \mathbf{D}^c + \ldots \phi^F \circ \mathbf{D}^F$, where ϕ^c is a $V \times K$ matrix of weights that is determined contextually based on the desired properties of $\mathbf{D}$.

As explained in [10], two[2] desirable properties of a similarity measure $\mathcal{D}$ are accuracy and certainty in identifying matches. Since the former relies on unknown groundtruth data, we address the latter. Note that because [10] focuses on evaluating a *single* similarity measure, their approach to estimating certainty of $\mathcal{D}$ can no longer be applied for cases involving multiple features. We thus infer the certainty of $\mathbf{D}^c$ as follows. For each node p, its M best matches correspond to the displacement labels that are associated with the M lowest dissimilarity costs of $\mathbf{D}^c$. If these matches are spatially inconsistent, $\mathbf{D}^c$ has low certainty. Hence, we measure the inter-distances between the top M matches and define *feature certainty* at node p for component c as:

$$FC(p,c) = e^{-\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M} ||\mathbf{v}_{g(\mathbf{D}^c,p,i)} - \mathbf{v}_{g(\mathbf{D}^c,p,j)}||} \tag{5}$$

where $g(D^c, p, j)$ returns the index to the j-th top candidate ranked by $\mathbf{D}^c$ at p. Finally, we note that as $\mathbf{D}$ can be computed efficiently via the use of shift operators [3], the main computational bottleneck is solving (3), which depends on V and K.

3 Results

Materials. Our test data included 40 brain MR-T1 images from the LONI dataset [13]; a set of MR-T1, MR-T2, and PD thigh images from the visible human (VH) dataset[3]; and 5 pairs of MR-T1 and MR-T2 brain images from RIRE[4].

When generating the initial label set in all experiments to be described, the search space was sampled in the spherical coordinate system, bounded by the maximum displacement in G, with sampling rate R (equal spacing on the colatitude, longitude, and radial axes). Unless otherwise stated, $R = 5$.

[2] Other properties examined in [10] relate to the convergence behaviour of $\mathcal{D}$ in the context of local optimization, where $\mathbf{D}$ should generate as few local minima on E as possible. When one adopts graph-based optimization with a guaranteed unique global solution [11], problems due to local minima of E are effectively avoided [12].

[3] `ftp://nlmpubs.nlm.nih.gov/visible/bitmaps/mri/`

[4] `http://www.insight-journal.org/rire/`

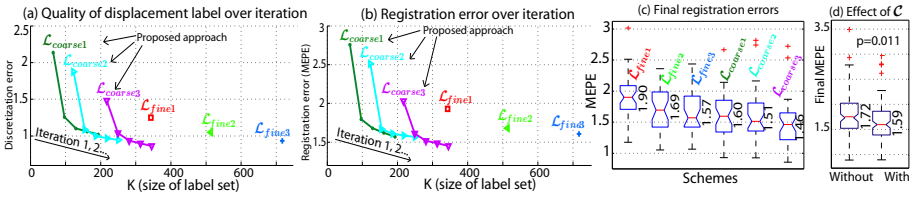

Fig. 1. (a) Discretization and (b) registration error over iteration. Note that our approach with $\mathcal{L}_{coarse2/3}$ achieved final registration errors lower than those achieved with $\mathcal{L}_{fine1/2/3}$. (c) Boxplot of the final registration errors obtained for each scheme. When compared to $\mathcal{L}_{fine3}$ alone, our method with $\mathcal{L}_{coarse3}$ also yielded statistically significantly lower errors (p=0.034). (d) $\mathcal{L}_{refine}$ generated based on $\mathcal{C}$ yielded errors that were significantly lower than the case without $\mathcal{C}$.

Experiment I: Synthetic Misalignment. Experiments in this section involved recovering known warps (denoted as G) that have been applied to I_b. The warps were generated by randomly displacing control points of a B-spline FFD model, where the magnitude of displacements (in voxels) were sampled from $\mathcal{N}(8, 2)$.

a) Examining Efficiency of Proposed Scheme. For a given label set $\mathcal{L} = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ and a warp G, the registration error is lower bounded by the (normally unknown) *discretization error* (DE): $\frac{1}{V}\sum_{x\in\mathcal{V}} \min_{k\in\{1,\cdots,K\}} ||G(x) - \mathbf{v}_k||$. Measuring DE would allow us to examine the quality of the label progressions. Clearly, DE may be lowered by simply adding more labels, but that comes at the price of increasing computational complexity of the method. One strength of our method is in adding new labels that are effective in decreasing DE. To examine this trade-off between K (number of labels) and DE, our first set of experiments proceeded as follows. We employed our method using 2 levels of coarse discretization of the search space for the initial label set ($\mathcal{L}_{coarse1/2/3}$) and compared the results with those obtained with 3 levels of finer discretization ($\mathcal{L}_{fine1/2/3}$), but without our progressive search. For coarse levels, R ={3,4,5} and for fine levels, R ={6,7,8}. The normalized cross-correlation (NCC) similarity measure was used; the size of $\mathcal{L}^n_{refine}$ was $N = 32$. Fig. 1a plots the DE over iterations averaged over 40 image pairs from the LONI dataset. Note how the curves using our progressive search are closer to the origin (i.e. lower DE and smaller K) than those without progressive search (original RWIR [1]). For example, progressive search with $\mathcal{L}_{coarse1}$ needed only about 200 labels to achieve DE =1, whereas $\mathcal{L}_{fine2}$ needed more than 500 labels. It also took $\mathcal{L}_{fine3}$ more than 700 labels to achieve the same DE that our method achieved using $\mathcal{L}_{coarse2}$ with about 250 labels. Further, when we calculated the actual obtained registration error (mean end-point-error[5], MEPE), we observe that our method with $\mathcal{L}_{coarse2/3}$ achieved errors lower than those using $\mathcal{L}_{fine1/2/3}$ as shown in Fig. 1b, while using a smaller label set. Moreover, Fig. 1c shows that our method with $\mathcal{L}_{coarse3}$ gave significantly lower MEPE than $\mathcal{L}_{fine3}$ (p=0.034). In summary, our proposed strategy performed efficient search without sacrificing accuracy.

b) Effect of $\mathcal{C}$. One may alternatively generate $\mathcal{L}_{potential}$ from *all* voxels, regardless of $\mathcal{C}$. However, this adds noise to the quantization of $\mathcal{L}_{potential}$. To examine this effect,

[5] Target registration error calculated over the whole image domain [3].

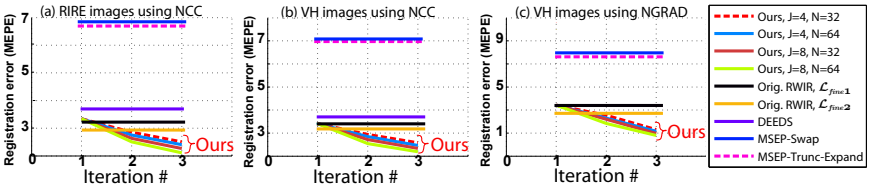

Fig. 2. Registration error of different methods for (a) RIRE and (b-c) VH datasets. In (c), NGRAD was used. To avoid direct comparison of error reduction rate across different methods, we plotted only the final errors for Orig. RWIR, DEEDS and MSEP. Note that MSEP gave results inferior to those of RWIR; one reason may be that graph-cuts performs poorly for large label set [7]. For the purpose of these experiments, θ was set to 2 pixels.

we repeated experiments in (a), using the same image pairs, label sets ($\mathcal{L}_{coarse1/2/3}$) and parameters ($N$,$J$), but generated $\mathcal{L}_{potential}$ from all voxels. As Fig. 1d reports, because the labels generated based on $\mathcal{C}$ were closer to the optimal displacements than the case with all voxels considered, the obtained errors were significantly lower (p=0.034).

c) Multi-model Image Registrations of RIRE and VH Datasets. We next registered 60 image pairs using data from RIRE and compared our method to the following methods: 1) DEEDS [4], which employs coarse-to-fine optimization and compositive scheme; and 2) MSEP [7], which uses either alpha-beta swap or truncated alpha-expansion algorithms (graph-cuts). For fairest possible comparison, RWIR and MSEP used the exact same matrices $\mathbf{D}$ and regularization strength α. Fig. 2a reports the registration performance based on registration error. For our method, we explored 2 choices of N and J (4). As Fig. 2a-c show, our method was not sensitive to these parameters. Our method also attained the lowest error after the second iteration. When compared to non-progressive search that uses $\mathcal{L}_{fine1/2}$, our method also attained lower errors after the second iteration. Repeating the experiment on images from the VH dataset using correlation in intensities (NCC) or normalized intensity gradients (NGRAD) gave similar results, as shown in Fig. 2b-c. Note that MSEP gave the worst results, affirming the observations of [7,11] that graph-cuts do not perform well for large K. Further, MSEP does not implement coarse-fine labeling but only approximates an energy corresponding to a fine labeling with an energy pyramid, and thus could only explore a limited set of labels. Conversely, DEEDS did better, but not the best, possibly because it employs (1) a tree-structure to encode regularization (vs. 6-connected grid as done in our case), and (2) stochastic, and thus less accurate, approximation of $\mathbf{D}$.

d) Multi-feature Registration with and without Adaptive Feature Weighting. We next examined the effect of incorporating ϕ into RWIR. We performed 100 trials on image pairs from VH and RIRE. The explored features included intensity (NCC), normalized intensity gradient (NGRAD), and the MIND descriptor (with radius=0) [14]. Fig. 3 shows the registration error obtained by RWIR under different feature choices and regularization settings; when our proposed feature weighting ϕ was used, the results were more accurate than the case of using all features or any of the features alone.

Experiment II: Real Registrations. Finally, we registered 60 randomly selected image pairs from the LONI dataset. As only segmentation is available for each image,

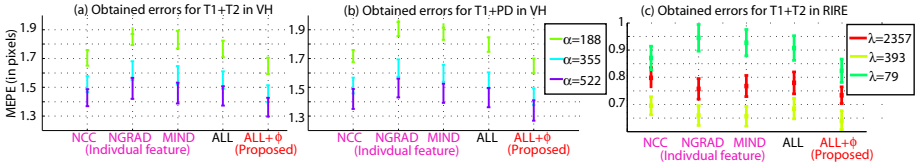

Fig. 3. Effectiveness of proposed adaptive feature weighting scheme as evaluated on VH and RIRE images. Shown are the mean and std. registration errors obtained by our approach under 3 regularization settings (α). Our method, denoted as ALL+ϕ, yielded the lowest MEPE, when compared to each feature alone (columns 1-3) and to the case using all features but no adaptive weighting (ALL).

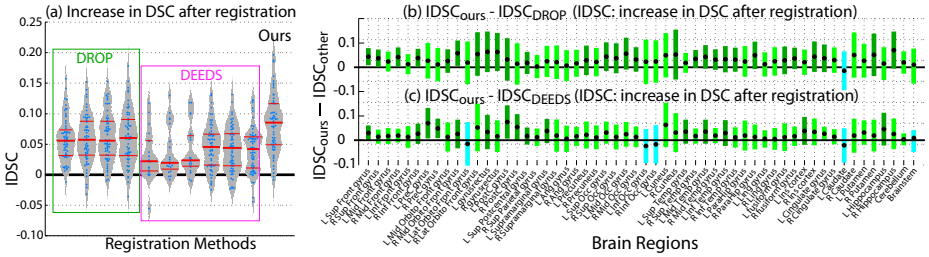

Fig. 4. Evaluation on LONI dataset. (a) Increase in DSCs (IDSC) of different methods. For DROP and DEEDS, 20 and 10 regularization settings were used respectively. Due to space constraints, only best subsets are shown. (b) Boxplot of the differences in IDSCs between our method and DROP (or DEEDS) for each anatomical region [13]. Green indicates positive difference (i.e. ours is better). Dark green indicates statistically significant positive difference. Blue indicates negative difference (ours is worse). Dark blue indicates stat. sign. negative difference (absent here).

registration accuracy is evaluated based on Dice similarity coefficient (DSC) between the corresponding segmentations of the warped and reference images. For details on the region names and segmentation protocols, refer to [13].

We performed registration using our proposed method and compared it to DEEDS and DROP [3], which is another registration method that adopts coarse-to-fine, multi-resolution registration. For both DROP and DEEDS, we repeated registration with different regularization settings. Conversely, we only tested our method with one regularization setting, i.e. α=0.025$\times V$ and report results obtained when J=4, N=64. For a fair comparison, all methods used the NCC similarity measure. Fig. 4a compares these methods in terms of **increase in DSC** (IDSC) after registration. We also show in Fig. 4b-c the difference in IDSC between our method and DROP (or DEEDS) and performed non-parametric one-way analyses of variance on the obtained IDSC to evaluate the statistical significance in the differences. As Fig. 4 shows, our method was only inferior in terms of IDSC for at most 5 out of 56 regions, but achieved statistically significantly higher IDSCs than DROP for 29 regions and higher IDSCs than DEEDS for 14 regions.

4 Conclusions

We proposed an efficient, progressive solution search and a contextual approach to spatially adaptive feature selection to enhance the original RWIR algorithm [1]. We also validated our method against three state-of-art registration frameworks [3,4,7] that employ discrete optimization. Results revealed that our method achieved higher accuracy over these methods and made significant improvement over the original framework in terms of both accuracy and efficiency. Future work consists of extending the evaluation to include other state-of-art methods examined in [13].

References

1. Cobzas, D., Sen, A.: Random walks for deformable image registration. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 557–565. Springer, Heidelberg (2011)
2. Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C.: DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Med. Image Anal. 15(4), 622–639 (2011)
3. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Image Anal. 12(6), 731–741 (2008)
4. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 115–122. Springer, Heidelberg (2012)
5. Tang, L., Hero, A., Hamarneh, G.: Locally-adaptive similarity metric for deformable medical image registration. In: ISBI, pp. 728–731 (2012)
6. Glocker, B., Paragios, N., Komodakis, N., Tziritas, G., Navab, N.: Optical flow estimation with uncertainties through dynamic MRFs. In: CVPR (2008)
7. Bagon, S., Galun, M.: A unified multiscale framework for discrete energy minimization. CoRR, vol. 4867 (2012)
8. Kohli, P., Torr, P.: Measuring uncertainty in graph cut solutions. Comput. Vis. Image Und. 112(1), 30–38 (2008)
9. Li, Y., Verma, R.: Multichannel image registration by feature-based information fusion. IEEE Trans. Med. Imag. 30(3), 707–720 (2011)
10. Skerl, D., Likar, B., Pernus, F.: A protocol for evaluation of similarity measures for non-rigid registration. Med. Image Anal. 12(1), 42–54 (2008)
11. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power watershed: A unifying graph-based optimization framework. IEEE Trans. Pattern Anal. Mach. Intell. 33(7), 1384–1399 (2011)
12. Lempitsky, V., Rother, C., Roth, S., Blake, A.: Fusion moves for markov random field optimization. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1392–1405 (2010)
13. Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G., Collins, D., Gee, J., Hellier, P., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3), 786–802 (2009)
14. Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, M., Schnabel, J.: MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. 16(7), 1423–1435 (2012)