ENHANCING ACCURACY OF SYMMETRIC RANDOM WALKER IMAGE REGISTRATION VIA A NOVEL DATA-CONSISTENCY MEASURE

Lisa Y. W. Tang[1,2] *and Roger Tam*[2] *and Ghassan Hamarneh*[1]

[1]Medical Image Analysis Lab., School of Computing Science, Simon Fraser University, Canada
[2]MS/MRI Lab., Dept. of Radiology, University of British Columbia, Canada

ABSTRACT

Random walker image registration has recently been extended in [1] to enforce inverse consistency in the solution by registering the input images towards a common space that resides midway between the two images. In this paper, we propose a novel extension to [1] to further improve its accuracy. We do so by proposing a voxel selection criterion that examines consistency of the data-likelihood estimates computed between the forward and backward directions. In particular, poor agreement occurs at locations where the top candidate displacement labels preferred by the forward direction conflict with those preferred by the backward direction. Once *data-consistency* is measured at every voxel location, the data-likelihood estimates of locations with low data-consistency are adjusted so that these nodes will contribute minimally to the similarity calculation. Experiments using different image modalities and image similarity measures show that this scheme can improve registration accuracy significantly per statistical analyses.

1. INTRODUCTION

A growing trend [2–4] to solving deformable image registration (DIR) is to adopt a graph-based optimization approach due to its various benefits, including its ability to avoid calculation of gradients of the energy to be minimized, high efficiency (e.g. via primal-dual linear programming) and the ability to obtain unique and globally optimal solutions (e.g., via random walker [2] or dynamic programming [4]). Of the various graph-based optimization strategies proposed in the literature, the formulation based on the random walker optimization that has led to the so-called random walker image registration (RWIR) [2] paradigm has various additional advantages: 1) it assumes no particular graph structure (unlike the formulation of [4] where pixel connectivities are assumed to be tree-structured); 2) it better avoids metrication artifacts [5], i.e. "blockiness" observed in the obtained label field, problems which formulations with graph-cuts are known to be prone to; and 3) it readily affords estimation of the uncertainty distribution of the registration solution for subsequent analyses [6] (as opposed to graph-cuts framework that would need additional expensive computations of min-marginals [7]).

Since its initial adoption for medical image registration, various extensions of RWIR have been proposed that aim to improve the efficiency and accuracy [1, 6, 8] of its predecessors. Recently, Tang et al. [1] extended this frameweork further by introducing symmetry in RWIR by casting a graph-labeling problem in a common space onto which both images are mapped. They do so by employing a label set with *coupled* displacement vectors to represent the registration solution such that the first and second element in each coupled label pair represents the *forward* and *backward* mappings between each input image and the common space. Through the use of a common space wherein the data similarity is measured and spatial regularization is enforced, their method effectively avoids introducing asymmetry in the solution. For brevity, we will denote this framework as SDRW, which stands for Symmetric Diffeomorphic Random Walk registration.

In this work, we propose a novel extension to SRWIR to ensure that registration be steered *only* by voxels with high measured *data-consistency*. In particular, our method contextually examines the data-consistency of the data-likelihood estimates of the coupled labels by examining the agreement between the label preferences driven by the forward and backward directions. Intuitively, poor agreement occurs in locations where the top candidate displacement labels preferred by the forward direction conflicts with those preferred by the backward direction. Once data-consistency is measured at every voxel location, the data-likelihood estimates of locations with low data-consistency are adjusted, thereby ensuring that the image similarity term that drives registration would only consider voxel locations with consistent likelihood estimates. We emphasize that this idea is different from the voxel selection strategy of [9], which are determined based on image gradients (locations of high gradients are deemed more important in their work), while our scheme selects regions based on a fundamentally different criterion (i.e. local trust associated to the consistency of the data dissimilarity estimates). In addition, our selection scheme operates in the common space while [9, 10] performs the selection (and the optimization of the update fields) within each image domain inde-

pendently, and thus inevitably introduces asymmetry. Lastly, the selection scheme of [10] was only tested on linear registration. Note that while we have only applied our proposed method to the RWIR paradigm [2], the idea can similarly be extended to other graph-based symmetric registration algorithms where data likelihood estimates are computed *a priori*, e.g. [4]. However, we advocate the SRWIR framework due to the reasons highlighted above.

2. METHODS

2.1. Brief review of SDRW

Given two images I_1 and I_2 of dimensionality n, DIR aims to find a spatial transformation $\tau : \Omega^1 \subset \mathbb{R}^n \mapsto \Omega^2 \subset \mathbb{R}^n$ that best aligns them. The problem is generally cast as a search for a displacement vector field $\mathbf{U}$, $\tau(\mathbf{x}) = \mathbf{x} + \mathbf{U}(\mathbf{x})$, that best minimizes an energy function E consisting of an image dissimilarity measure $\mathcal{D}$ and a spatial regularization $\mathcal{R}$ that discourages irregularities in $\mathbf{U}$. The image is typically represented with a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ where a pixel coordinate $\mathbf{x}_i$ is represented by a graph node $i \in \mathcal{V}$ (V=$|\mathcal{V}|$) and connectivity between $\mathbf{x}_i$ and $\mathbf{x}_j$ is represented by a graph edge $(i, j) \in \mathcal{E}$. Furthermore, the feasible solution space for $\mathbf{U}(\mathbf{x})$ is typically sampled uniformly, forming a set of K displacement labels $\mathcal{L} = \{\mathbf{v}\}_k^K$, $\mathbf{v}_k \in \mathbb{R}^n$ so that one of these labels will be assigned to each graph node in a manner that minimizes E.

In RWIR, a Laplacian matrix $\mathbf{L}$ encoding connectivities in $\mathcal{E}$ plays the role of enforcing diffusion-based spatial regularization on $\mathbf{U}$ (which under some conditions, is adequate to ensure topology-preservation, as recently shown in [11]). For a given $\mathcal{L}$, a set of image similarity costs to each node can be encoded by a $V \times K$ matrix $\mathbf{D} = (d_{ik})$ so that low image dissimilarity between $I_2(\mathbf{x}_i + \mathbf{v}_k)$ and $I_1(\mathbf{x}_i)$ yield high likelihood values. Usually, d_{ik} is assumed to be spatially independent of d_{jk} and thus may be defined as:

$$d_{ik} = \exp(-\mathcal{D}(I_1(\mathbf{x}_i)), I_2(\mathbf{x}_i + \mathbf{v}_k)), \tag{1}$$

where $\mathcal{D}$ is an image dissimilarity measure $\mathcal{D}$ (e.g. sum of squared differences). Once computed for the entire graph, one then seeks to find a $V \times K$ matrix $\mathbf{P} = (p_{ik})$ (which would give a probabilistic representation of $\mathbf{U}$) that minimizes:

$$E(\mathbf{P}_k) = \sum_{j=1, j\neq k}^{K} \mathbf{P}_k^T \Lambda_j \mathbf{P}_k + (1-\mathbf{P}_k)^T \Lambda_k (1-\mathbf{P}_k) + \alpha \mathbf{P}_k^T \mathbf{L} \mathbf{P}_k, \tag{2}$$

where $\mathbf{P}_k$ represents the k-th component of $\mathbf{P}$ and Λ_k is a diagonal matrix with entries $[d_{1k}, \cdots, d_{Vk}]$. As explained in [2, 8], the first two terms in (2) encourage $\mathbf{P}$ to be proportionally similar to the data-likelihood matrix $\mathbf{D}$, while the last term enforces Laplacian prior on $\mathbf{P}$ that effectively induces diffusion-based regularization on $\mathbf{U}$.

To enforce symmetry in the solution of RWIR, Tang et al. [1] recently proposed to perform labeling on a common space Ω^c with a coupled label set $\mathcal{L}^c = \{(\mathbf{v}, -\mathbf{v})\}_k^K$, where the first (or second) element of each coupled pair represents the spatial mapping between the common domain and Ω^1 (or Ω^2). The algorithm adopts the standard multi-resolution approach where the solution is computed incrementally from a set of (update) displacement fields V_m^1 and V_m^2, where m denotes the m-th resolution. More specifically, the data-likelihood of assigning each coupled label pair $(\mathbf{v}_k, -\mathbf{v}_k)$ to node $\mathbf{x}_i \in \Omega^c$ is then defined as:

$$\mathbf{D}^c = \mathbf{D}^{c1} + \mathbf{D}^{c2} \tag{3}$$

where $\mathbf{D}^{c1}$ and $\mathbf{D}^{c2}$ are the matrices representing the dissimilarity measured between the intermediately deformed images, i.e. with entries:

$$d_{ik}^{c1} = \exp(-\mathcal{D}(I_1(\mathbf{x}_i + V_m^1(\mathbf{x}_i) + \mathbf{v}_k), I_2(\mathbf{x}_i + V_m^2(\mathbf{x}_i) - \mathbf{v}_k))$$
$$d_{ik}^{c2} = \exp(-\mathcal{D}(I_2(\mathbf{x}_i + V_m^2(\mathbf{x}_i) - \mathbf{v}_k), I_1(\mathbf{x}_i + V_m^1(\mathbf{x}_i) + \mathbf{v}_k)),$$

Diffeomorphism is guaranteed by applying the exponential map to the update fields V_m^1 and V_m^2 [1].

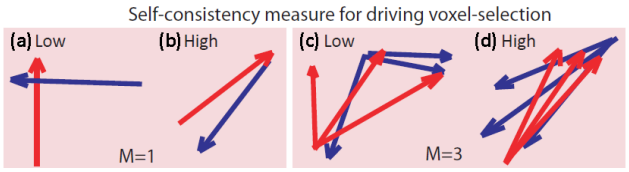

Fig. 1. How the proposed *data-consistency* measure (to be presented in Section 2.1) examines agreement in the data-likelihood estimates computed from both directions. (a)-(b) The red (and blue) arrows represent the top (M=1) candidate point-correspondence as indicated by the likelihood estimated from the forward direction (i.e. $\mathbf{D}_p^{c1}$ in Section 2.2) (and $\mathbf{D}_p^{c2}$ in Section 2.2). (c)-(d) other observed cases with $M = 3$. Our proposed data-consistency measure examines the mutual agreement between the most preferred candidate correspondences and quantifies (a) and (b) to be more trustworthy than (c) and (d).

2.2. Boosting accuracy via measure of consistency

In this work, rather than using $\mathbf{D}^c$ directly, we propose to examine the consistency between $\mathbf{D}^{c1}$ and $\mathbf{D}^{c2}$ prior to solving (2). More specifically, we evaluate the degree of agreement between correspondences implied by these matrices. When the measured agreement at a given node $\mathbf{x}_i \in \Omega^c$ is weak, the data-likelihoods estimated at this node are highly uncertain and thus should be dampened or disregarded.

To measure this uncertainty, we first note that the most probable point-correspondence between Ω^c and Ω^1 at $\mathbf{x}_i$ is given by a=arg max$_k$ d_{ik}^{c1}, and likewise, the most probable correspondence between Ω^c and Ω^2 at the same location is given by b=arg max$_k$ d_{ik}^{c2}. In ideal situations, these candidate correspondences should agree, i.e. distance between $\mathbf{v}_a$ and $-\mathbf{v}_b$ is small, where $\mathbf{v}_a$ is the displacement label indexed

by a ($\mathbf{v}_b$ indexed by b). However, due to noise and ambiguity in the image data, as well as the asymmetry introduced during the discretization process (and inherent asymmetry of some image dissimilarity measures), these correspondences will likely be inconsistent. We thus propose to measure the data-consistency of the data-likelihood estimates computed from both directions and drive registration more by nodes with high data-consistency. To do so, we quantify the agreement between the top M preferred correspondences between the forward and background directions as follows:

$$\mathbf{C}(i) = \exp\left(-\sum_{m=1}^{M}\sum_{n=1}^{M}\frac{1}{2m}\frac{1}{n}||\rho(m,\mathbf{D}^{c1},i) - (-\rho(n,\mathbf{D}^{c2},i))||\right), \tag{4}$$

where ρ returns the displacement label of the m-th top candidate label ranked by $\mathbf{D}^{c1}$ at i; M is a tunable parameter that ensures $\mathbf{C}$ is robust against outlier correspondences by computing a weighted average agreement between the top M candidate correspondences, each of which is weighted by $\frac{1}{2n}\frac{1}{m}$, which was designed based on preliminary experiments. In words, when the displacement vectors implied by the point-correspondences given by the $\mathbf{D}^{c1}$ and $\mathbf{D}^{c2}$ matrices point in different directions (and thus the Euclidean distance between these displacement vectors will be large), the mutual agreement is weak, and thus should give low $\mathbf{C}$ values.

Once $\mathbf{C}$ is computed for every node, we compute its t-th percentile as the threshold for deciding when a node should be regarded as untrusted. We then adjust the data likelihood matrix in (2) so that it would give zero preference to any labels for all nodes deemed untrustworthy by setting:

$$\mathbf{D}^{c}_{ik} = \frac{1}{K}. \tag{5}$$

Note that one may alternatively employ a softer selection scheme where one would instead dampen the effects of nodes with lower data-consistency. However, based on synthetic experiments, we found that employing a hard threshold was more effective than the latter approach. This is largely because the diffusion regularization inherently smooths out the label preferences. Hence, a stricter scheme would be more effective than a soft selection scheme.

Once $\mathbf{D}^c$ is modified as given in (5), we then solve (2) in the same way described in [1]. To circumvent memory issues required by solving the large system of Laplace equations, we perform optimization using preconditioned conjugate gradient, as also suggested in [5].

Fig. 1 demonstrates how our proposed measure reflects uncertainty in the data-likelihood estimates using examples randomly drawn from actual registration trials. More specifically, Fig. 1 portrays the most preferred candidate point-correspondence observed at a point in an actual registration trial, with the blue and red arrows denoting the top candidate displacement labels preferred by the forward and backward directions, respectively. By examining the (weighted average) distance between the preferred correspondences, the proposed measure gives low value when they disagree (displacements labels with high Euclidean distance).

3. EXPERIMENTAL RESULTS

3.1. Materials

We employed 4 public datasets to evaluate the effect of our proposed extension to SDRW of [1]. These include 40 brain magnetic resonance (MR) T1-w images from the UCLA-LPBA dataset [12]; 5 pairs of co-registered Computed Tomography (CT) and Proton-Density (PD) brain images from RIRE[1]; 100 randomly selected pairs of co-registered MR T2-w and PD images from the IXI dataset[2], which consists of images collected at three different hospitals in London; and 12 MR T1-weighted images from the CUMC dataset[3].

3.2. Experiments Conducted

As previous work [1, 8, 13] have provided various comparisons of RWIR [2] with other registration algorithms, we focused in this paper primarily on examining the impact of our proposed extension on the accuracy of SDRW [1]. For brevity, we shall denote our extension as SDRW+C, where 'C' is used to denote the proposed use of data-consistency measure.

Synthetic data. We applied synthetic warps to registered pairs of images to obtain groundtruth solutions. The warps were generated by displacing control points of a free-form deformation (FFD) model at random. Unless otherwise specified, the magnitude of the displacements were sampled from $\mathcal{N}(4,3)$ (in pixels).

In quantifying registration performance, we computed registration error as the mean end-point-error (MEPE) between the recovered and the groundtruth warps. For a fair comparison with SDRW [1], we set $\alpha = 0.025 \times V$, as noted in [1].

[1] http://www.insight-journal.org/rire/
[2] http://biomedic.doc.ic.ac.uk/brain-development/
[3] http://http://www.mindboggle.info/data.html

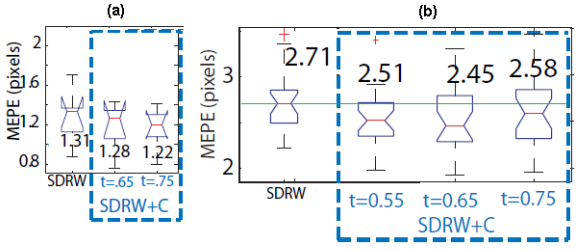

Fig. 2. (a) Results from 50 trials on brain images from IXI. Note that error decreased when data similarity boosting was performed. This was true for various values of t. (b) Repeating the same experiments on PD and CT images from RIRE with 50 random warps. Note that error decreased when data-consistency was employed (i.e. SDRW+C).

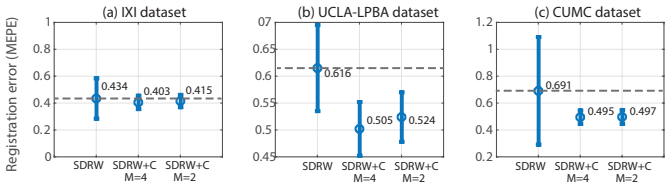

Fig. 3. Comparisons performed on three brain datasets (a) IXI, (b) UCLA-LPBA (c) and CUMC. In general, our method SDRW+C that uses $\mathbf{C}$ for selection of driving voxels led to more stable performance, as reflected by the lower std. dev. of MEPE. Note that improvement in accuracy was marginal for the IXI dataset (per Student's t-test) but significant for the other 2 datasets ($p < 0.05$ per Student's t-test).

Fig. 2a reports evaluation done using 50 synthetic warps applied to 50 pairs of IXI brain images. From Fig. 2a, we can see that the registration error obtained with SDRW+C is lower than the case using just SDRW. Registrations were done using a gradient-based similarity measure as employed in [1]. Repeating the experiment with 50 synthetic warps applied to pairs of PD and CT brain images from the RIRE[4] dataset, using the MIND descriptor [14], yielded similar results as shown in Fig. 2b. We also examined the effect of different values of the threshold value t. As reported in the subfigures, results are relatively stable for the ranges of t tested and that in both datasets, our proposed SDRW+C method improved accuracy of SDRW by as much as 0.26 pixels in MEPE.

To examine the effect of M, we then repeated the experiment with another set of 50 synthetic warps that involved greater magnitude of displacements: $\mathcal{N}(8, 3)$ (in pixels). Fig. 3 reports the comparisons on three brain datasets that we have tested our method on. For these experiments, we use the normalized correlation coefficient as the similarity measure. Similarly, we found our method relatively stable for the values tested for M.

Registration performance evaluated on real data. Lastly, to evaluate our method's performance on registrations of clinical images, we employed the CUMC brain MRI dataset that contains 12 volumes of $256 \times 256 \times 124$ resolution and $0.86 \times 0.86 \times 1.5$mm spacing. This dataset was additionally segmented into 130 anatomical regions and had previously been used as benchmarking of 16 existing deformable registration algorithms, which include the state-of-the-art methods like the symmetric normalization algorithm (SyN) of [15]. For a direct comparison with [8], we followed [8] in using the squared difference (SD) measure for computing image similarity. A total of 12×12 - 12= 132 pairwise registrations were performed for each method. We set $M = 3$ and $t = 0.7$ based on a separate leave-out-one cross-validation experiment performed on a subset (10 randomly selected volumes) of the LPBA dataset that involved a grid-search over $M = \{2, 3, 4, 5\}$ and $t = \{0.6, 0.65, 0.70, 0.75\}$.

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

Results show that our proposed method, SDRW+C, achieved a Jaccard Index of 0.341, which is an improvement over SDRW [1]. We also repeated the registrations using SDRW only (without the adjusted data likelihood matrix as given in (5)). Based on a Student's t-test, the improvement was found to be significant ($p < 0.0024$).

4. CONCLUSIONS

In this paper, we extended the symmetric registration algorithm of [1] by proposing the data-consistency measure $\mathbf{C}$ that measures the quality of the data likelihood estimates and proposed to drive registration only by voxels that have high measured consistency. Based on various experiments involving different image similarity measures, our strategy has improved registration accuracy significantly.

5. REFERENCES

[1] L Tang, S Andrews, and G Hamarneh, "Random walker image registration with inverse consistency," in *ISBI 2015*. IEEE, 2015, pp. 837–840.

[2] D Cobzas and A Sen, "Random walks for deformable image registration," in *MICCAI*, 2011, pp. 557–565.

[3] Ou et al., "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," *Medical Image Analysis*, vol. 15, no. 4, pp. 622–639, 2011.

[4] Heinrich et al., "Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling," in *MICCAI*, 2012, vol. 7512, pp. 115–122.

[5] Couprie et al., "Power watershed: A unifying graph-based optimization framework," *IEEE TPAMI*, vol. 33, no. 7, pp. 1384–1399, 2011.

[6] L Tang and G Hamarneh, "Random walks with efficient search and contextually adapted image similarity for deformable registration," in *MICCAI*, 2013, vol. 8150, pp. 43–50.

[7] D Tarlow and R Adams, "Revisiting uncertainty in graph cut solutions," in *CVPR 2012*. IEEE, 2012, pp. 2440–2447.

[8] K Popuri, D Cobzas, and M Jagersand, "A variational formulation for discrete registration," in *MICCAI*, 2013, pp. 187–194.

[9] G Wu, M Kim, Q Wang, and D Shen, "S-HAMMER: Hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images," *Human Brain Mapping*, vol. 35, no. 3, 2014.

[10] B Oreshkin and T Arbel, "Uncertainty driven probabilistic voxel selection for image registration," *IEEE TMI*, vol. 32, no. 10, pp. 1777–1790, 2013.

[11] S Andrews, L Tang, and G Hamarneh, "Topology preservation and anatomical feasibility in random walker image registration," in *MICCAI*, 2014, pp. 210–217.

[12] Klein et al, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.," *NeuroImage*, vol. 46, no. 3, pp. 786–802, 2009.

[13] L Tang and G Hamarneh, "Random walker image registration with cost aggregation," in *ISBI*, 2014, pp. 1–4.

[14] Heinrich et al., "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical Image Analysis*, vol. 16, no. 7, 2012.

[15] Avants et al., "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26 – 41, 2008.