LOCALLY-ADAPTIVE SIMILARITY METRIC FOR DEFORMABLE MEDICAL IMAGE REGISTRATION

Lisa Tang[1], *Alfred Hero*[2], *and Ghassan Hamarneh*[1]

[1]Medical Image Analysis Lab., School of Computing Science, Simon Fraser University
[2]Departments of EECS, BME and Statistics, University of Michigan - Ann Arbor

ABSTRACT

More and more researchers are beginning to use multiple dissimilarity metrics or image features for medical image registration. In most of these approaches, however, weights for ranking the relative importance between the selected metrics are empirically tuned and fixed for the entire image domain. Different parts of a medical image, however, may contain significantly different appearance properties such that a metric may only be applicable in certain image regions but less so in other regions. In this paper, we propose to adapt this weighting to generate a locally-adaptive set of dissimilarity metrics such that the overall metric set encourages proper spatial alignment. Using contextual information or via a learning procedure, our approach generates a vector weight map that determines, at each spatial location, the relative importance of each constituent of the overall metric. Our approach was evaluated on 2 datasets of 15 computed tomography (CT) lung images and 40 brain magnetic resonance images (MRI). Experiments show that our approach of using a locally-adaptive set of dissimilarity metrics gives superior results when compared against its non-region specific variant.

1. INTRODUCTION

One essential component in medical image registration (MIR) is the image dissimilarity metric. As no single metric is suitable for all applications, many definitions have been proposed for different applications, e.g. mutual information, cross-correlation, sum of squared differences (SD), etc.

In the past decade, researchers have began to combine multiple dissimilarity metrics or image features to boost registration performance. For instance, Shen and Davatzikos [1] proposed the HAMMER approach that combines segmentation- and feature- based information for brain image registrations. Liao et al. [2] also combined an improved version of mutual information, a feature-based metric, and a local descriptor for brain image registration. In [3], Tang and Hamarneh matched shapes by combining geometric, topological, and intensity-based features. For registration of lung images, Cao et al. [4] combined a measure called vesselness difference (VD) with a conventional intensity-based measure.

In all aforementioned works, an empirically tuned set of weights (or a single scalar weight) is used to linearly combine the involved dissimilarity metrics. These weights, however, are global in the sense that they remain constant across the image domain. In this paper, we argue that different regions in medical images contain different appearance properties (i.e. due to the differences in the underlying tissue appearance (e.g. textures) and thus, registration in a particular image region should be driven by the most relevant dissimilarity metric(s) or image feature(s) in that region. In neuro-images, for instance, there is no strong reason to use the same metric when measuring dissimilarity in regions belonging to the white matter, gray matter, or cerebrospinal fluid. Likewise, in lung image registration, a vesselness-based measure, e.g. [4], operates best within the lung regions (where vessels reside) but becomes inferior when it operates on other anatomical regions that lack vessels, as our experimental results in Section 3.1 will demonstrate.

Accordingly, we propose to employ a learning approach to construct a locally-adaptive metric that fuses multiple dissimilarity metrics or image feature sets. The learned metric encodes prior knowledge about each metric's effectiveness in driving correct registrations and places different amounts of emphasis on each component of the fused set according to local image content. While we may manually design such locally-adaptive combination of dissimilarity metrics, we also advocate a learning approach so that the learned metric is derived from a corpus of exemplars, thereby formulating a general method that applies to different applications and need not require manual intervention.

To the best of our knowledge, our work is most similar to [5] in that it also learns a weighting function to obtain a spatially adaptive combination of a feature set. However, their work does not make use of a heterogeneous feature set as we advocate, and is solely for surface matching. The authors of [6, 7] also learn a novel dissimilarity metric from a set of training images, but both are fundamentally different from our approach of computing and employing a *locally-adaptive* set of dissimilarity metrics.

2. METHODS

Deformable image registration seeks to recover a transformation $\mathcal{T}$ that best aligns two images I_a and I_b. Generally, the problem involves minimizing a weighted sum of two penalty terms, i.e.:

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

where S denotes a dissimilarity function between two images I_a and I_b, $\mathcal{R}$ denotes a regularization term that encourages $\mathcal{T}$ to maintain certain smoothness properties (e.g. being continuous or homeomorphic), and α is a weight that balances these two terms.

We propose to build S using a group of image metrics, which includes dissimilarity between extracted features as a special case (further details on feature-based metrics are presented in Section 3.2). Specifically, let there be a set of dissimilarity metrics, $\{S_1, S_2, ..., S_{|S|}\}$, where each component might be an image metric computed between an image pair (e.g. intensity difference, cross-correlation, local mutual information, etc.), or those that are defined in terms of features (e.g. vesselness difference [4], Gabor responses, etc.), which have the form of

$$S_i(\mathbf{x}, \mathbf{y}, I_a, I_b) = |\mathcal{F}_j(\mathbf{x}, I_a) - \mathcal{F}_j(\mathbf{y}, I_b)|^2 \tag{2}$$

where $\mathcal{F}_j$ denotes the j-th feature extracted from I_a or I_b. For brevity, we will now refer to a dissimilarity metric that is defined on extracted features simply as another metric.

Our approach generates and employs a weight function $W : \Omega \times |S| \mapsto \mathbb{R}$ that maps each spatial location to a vector where component i of the vector is a relative weight assigned to S_i such that the overall dissimilarity between I_b and I_a becomes

$$S(\mathbf{x}, \mathbf{y}, I_a, I_b) = \sum_{i}^{|S|} W(\mathbf{x}, i) S_i(\mathbf{x}, \mathbf{y}, I_a, I_b) \tag{3}$$

where i is the index of the i-th metric component and $W(\mathbf{x}, i)$ denotes the weight at $\mathbf{x}$ for the i-th metric, and $\mathbf{y} = \mathcal{T}(x)$. The weight function W should be designed so that high importance is only given to metrics at regions where they are effective in producing proper image alignment and vice versa. We will illustrate two approaches in generating W: contextual or learned. The contextual approach is adopted if prior knowledge about the appropriateness of certain metrics is available. For the example of CT lung registration using VD and SD where lung masks can be reliably created, one may then design a weighting scheme that employs specific metrics in specific regions. Registration results for this scenario will be presented in Section 3.1. Alternatively, when we do not have such prior knowledge or when the size of the metric set is prohibitively large, a learning approach is used where the weight function is learned from a training set of registered images. We next detail this learning approach.

2.1. Learning the Weight Function

Let there be a set of N linearly registered images $\mathcal{I}$, where each spatial coordinate $\mathbf{x} \in \Omega$ corresponds in each image in the set. The registered images shall provide training data from which our method estimates the effectiveness of a particular metric in aligning images properly. Specifically, at every spatial location, we collect a set of metric values where a metric is computed between pairs of aligned images. For brevity, we denote the samples collected at $\mathbf{x}$ for metric i as $Q_{aligned}(\mathbf{x}, i)$. To learn when a metric fails to align images, we also collect another set of samples of metric values where a metric is computed between pairs of misaligned images (which can be generated by applying random warps or global translations to each of the aligned images in $\mathcal{I}$). We will denote these samples as $Q_{misaligned}$.

Recall our goal of computing a weight function W that favors dissimilarity metrics that produce low values at aligned regions and high values at misaligned areas. Precisely, we should assign high weights to dissimilarity metrics that remain consistently low across images at $\mathbf{x}$ and consistently high outside the periphery of $\mathbf{x}$ (i.e. at non-corresponding locations). Otherwise, the metric should contribute minimally to S. Therefore, we propose the following energy cost:

$$E_{w_1}(W) = \sum_{\mathbf{x}\in\Omega} \sum_{i}^{|S|} \frac{W(x,i) M_{aligned}(\mathbf{x}, i) V_{aligned}(\mathbf{x}, i)}{M_{misaligned}(\mathbf{x}, i)}; \tag{4}$$

$$\text{subject to } \sum_i W(x,i) = 1,\ \forall i, W(x,i) \geq 0 \tag{5}$$

where $M_{aligned}$ and $V_{aligned}$ are the mean and covariance of $Q_{aligned}$ and similarly defined for $Q_{misaligned}$. Intuitively, the optimal W minimizing (5) would favor metrics that generate a low value in $\frac{M_{aligned}}{M_{misaligned}}$ (low dissimilarity values over high). Note that this cost resembles those proposed in [8], in which Brown et al. learned a set of local image descriptors for image classification; due to our relatively smaller sample size, we therefore omit an additional step of dimensionality reduction on the training samples. We also propose an alternative energy cost that examines the difference between the means of the samples (rather than their ratio), subject to the same constraints in (5):

$$E_{w_2}(W) = \sum_{\mathbf{x}\in\Omega} \sum_{i}^{|S|} W(x,i) e^{-\frac{|M_{misaligned}(\mathbf{x},i) - M_{aligned}(\mathbf{x},i)|}{V_{aligned}(\mathbf{x},i)^2}}. \tag{6}$$

Motivated by the questions raised in [9], we also questioned whether different weight vectors of neighbouring locations would affect the performance of the learned metric. Accordingly, we examined the impact of spatial smoothness[1] of W and examined the impact of imposing spatial regularization on W, thus yielding the following optimization problem

[1] In contrast to our work, [9] proposed spatially adapting α, the weight between the data cost and the smoothness regularization.

for W:

$$\arg\min_{W} E_{w_{1|2}}(W) + \lambda \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{E}} \sum_{i}^{|S|} |W(\mathbf{x},i) - W(\mathbf{y},i)|^2 \quad (7)$$

where $\mathcal{E}$ is the set of pixel connectivities of the image grid, the second term enforces spatial regularization on W by penalizing the difference between the weight vectors of two spatial neighbours $\mathbf{x}$ and $\mathbf{y}$, and λ is an empirically tuned weight that adjusts the amount of regularization. In Section 3.2, we will examine the impact of spatial regularization on W on registration accuracy.

2.2. Graph-based registration

We employ a graph-based approach for image registration [10], in which we seek to label each spatial coordinate $\mathbf{x}$ of I_a with a displacement vector $\mathbf{t}$ such that the entire label field forms a displacement vector field $\mathcal{T}$. Mathematically, image registration incorporating the spatially adaptive W is now formulated as the minimization of the following MRF energy:

$$\arg\min_{\mathcal{T}} \sum_{\mathbf{x}\in\Omega} \sum_{i}^{|S|} W(\mathbf{x},i) S_i(\mathbf{x}, \mathbf{x}+\mathbf{t}_{\mathbf{x}}, I_a, I_b) + \alpha \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{E}} \mathcal{R}(\mathbf{t}_{\mathbf{x}}, \mathbf{t}_{\mathbf{y}}) \quad (8)$$

where $\mathbf{t}_{\mathbf{x}}$ is the translation assigned to $\mathbf{x}$ as specified by $\mathcal{T}$, etc.

3. EXPERIMENTAL RESULTS

3.1. Lung image registration

As a proof of concept, we show the use of a contextually generated weight function for the registration of lung CT images from the POPI dataset [11]. We follow the approach of [4] of combining the VD measure with an intensity-based measure, but rather than employing both metrics in a globally constant manner, we will weight the metrics in a spatially varying manner. Using contextual information to generate lung masks[2], we created W that places high emphasis on the VD measure in regions within the lungs and high emphasis on SD for regions outside the lungs. Fig. 1a shows an example weight function for a target image in the dataset. Then, for each lung image pair, we computed W from the template image I_a and registered each source image I_b to I_a.

To assess registration accuracy, we computed the target registration error (TRE) between expert-defined point-correspondences, which were also provided in the POPI dataset. However, all point-correspondences provided were within the lungs. In order to evaluate registration accuracy more thoroughly, we examined the quality of the alignment between lung surfaces, which can be reliably extracted from the images via thresholding. Fig. 1b-c show an example result from one registration trial. From the figure, we can see that the use of VD alone gave the worse performance for the alignment of lung surfaces as reflected by lower DSC (Dice similarity coefficient). On the other hand, the equally combined use of VD and SD improved the alignment of lung surfaces but the TRE between the point-correspondences within the lungs had increased dramatically. With the use of contextually weighted combination of VD and SD, registration achieved a relatively high DSC score while maintaining a relatively low TRE. In Fig. 1f-g, we report the averaged TRE and DSC scores as obtained over 15 registration trials. Evidently, while both the equal and weighted schemes achieved similar DSC scores, the former scheme gave much higher TRE scores than those achieved by the weighted scheme. This quantitatively reflects that proper alignment in regions both inside and outside the lungs can best be achieved by the spatially adaptive scheme.

[2]Lung masks were made available in [11]. However, we created our own lung masks using image intensities (via thresholding based on the Hounsfield Units and performed subsequent hole-filling procedures) to show general applicability of our approach.

3.2. Brain MRI Registration

We next evaluated our learning approach to construct W. For this experiment, we employed a dataset of 40 rigidly aligned brain MRI images provided in [12]. We chose this dataset because it also contains corresponding probabilistic segmentations from which we can measure accuracy of our obtained registration results.

Our validation pipeline was as follows. We performed 40

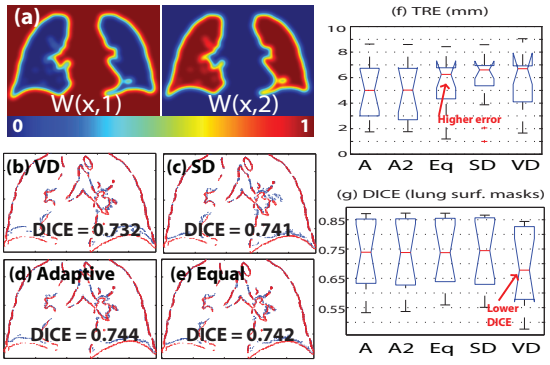


Fig. 1. Evaluating the contextual approach. (a) Components of W created for an image using a Gaussian-smoothed lung mask (kernel width σ). (b)-(e) Comparison between registrations using different metric schemes: result of using (b) VD measure only, (c) SD measure only, (d) a weighted combination of VD and SD that is adaptive to contextual information, and (e) equally combined VD and SD. While VD and the weighted scheme gave comparable TRE measures, the use of VD alone gave lower DSC (alignment of lung surfaces poor). Conversely, while equal weight gave slightly higher DSC than the weighted one, its TRE was *much higher* than the weighted one. (f-g) Boxplots of TRE between point-correspondences, and DSC scores that examine the alignment of lung surface points after registration as performed under different metric schemes. "A", "A2" denotes the use of adaptively weighted metric combination (with $\sigma = 3$ mm and $\sigma = 6$ mm, respectively) and "Eq" denotes equal weighting.

trials where each image in the set acted as a template and the remaining images were randomly separated into a set of $m = 15$ training images and a set of $n = 40 - m$ test images. In each trial ($c = 1 \cdots 40$), the weight function W_c of the template I_c was constructed from a separate set of $Q_{aligned}$ and $Q_{misaligned}$. The former was generated by evaluating metric S_i between each of the $\binom{m}{2}$ pairs of aligned training images at every spatial location $\mathbf{x}$. For $Q_{misaligned}$, we introduced misalignment between each pair prior to metric evaluation, i.e. we evaluated $S(\mathbf{x}, \mathbf{x} + \mathbf{t}_q, I_u, I_v)$ where $\mathbf{t}_q$ is a translation from the set $\{[a\ a], [a\ -a], [-a\ a], [-a\ -a]\}$ with $a = \{3, 6\}$, and $u \neq v$, $u, v \leq m$. Next, we computed $M_{aligned}$, $V_{aligned}$, etc. and optimized W_c for I_c using (5) or (6). Then, for the actual evaluation of our method, we applied random thin-plate-spline warps (pixel displacements in range of [-8,8] pixels) to the n test images (and their segmentations J_n) and subsequently performed registration between the warped test image and I_c by minimizing (8). Quality of registration result was then measured by computing the *reduction* in mean segmentation error (MSE) between the registered probabilistic segmentations (as compared to the MSE evaluated before registration).

Fig. 2 compares the registration results using an equally weighted metric set and the proposed schemes, i.e. performing minimization of (5) and (6), with and without spatial regularization. The metrics employed were sum of intensity difference (SD), gradient magnitude difference (GMD) and normalized gradients difference (NGD). From the figure, we see that the use of W with either (6) achieved more accurate registrations than an equally weighted metric set ($p = 0.008$ at 95% confidence interval). Interestingly, with spatial regularization, (6) did not improve MSE as significantly well as the other schemes, while (5) was relatively insensitive to the effect of spatial regularization on W. Based on the amount of reduction in MSE, we conclude that W as learned from (5) gives the best registration results.

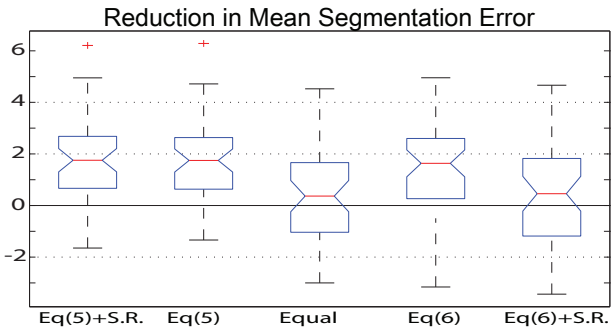


Fig. 2. Comparison of registration results as obtained without and with W as computed by different schemes. S.R. denotes enforcing spatial regularization on W. Note that reduction in MSE is greatest when W, as optimized with (5), was used.

4. CONCLUSION

We have shown two approaches to fusing and combining multiple dissimilarity metrics and image features in a weighted manner for image registration. When we have prior knowledge about the data, our empirical experiments showed that the accuracy of lung image registrations can be improved using a contextually computed weight function. When a set of roughly aligned images are available, we also showed how the weight function can be learned. Again, our experiments demonstrated that using a weighted combination of metrics as optimally determined via a learning procedure gave rise to higher registration accuracies than those achieved without such adaptive weighting. We are currently working towards extending our approach to multi-modal registration. We also foresee that the use of this locally-adaptive learned metric can improve the accuracy of groupwise registrations.

5. ACKNOWLEDGMENTS

LT and GH were partially supported by Natural Sciences and Engineering Research Council (NSERC), and AH was partially supported by the US National Institutes of Health, grant number 2P01CA087634-06A2.

6. REFERENCES

[1] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imaging*, vol. 21, no. 11, pp. 1421–1439, 2002. 1

[2] S. Liao and A. Chung, "A novel multi-layer framework for non-rigid image registration," in *ISBI*, 2010, pp. 344 –347. 1

[3] L. Tang and G. Hamarneh, "SMRFI: Shape Matching via Registration of Vector-Valued Feature Images," in *CVPR*, 2008, pp. 1–8. 1

[4] Cao et al., "Unifying vascular information in intensity-based nonrigid lung ct registration," in *International workshop on Biomedical Image Registration workshop*, 2010, pp. 1–12. 1, 2, 3

[5] F. Steinke, B. Schlkopf, and V. Blanz, "Learning Dense 3D Correspondence," in *Advances in Neural Information Processing Systems*, 2007, pp. 1313–1320. 1

[6] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N.D. Cahill, and B. Scholkopf, "Learning similarity measure for multi-modal image registration," in *CVPR*, 2009, pp. 186 –193. 1

[7] M. Bronstein, A. Bronstein, F. Michel, and N. Paragios, "Data fusion through cross-modality metric learning using similarity-sensitive hashing," in *CVPR*, 2010, pp. 3594–3601. 1

[8] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Trans. Pattern Anal. Mach. Intel*, vol. 33, pp. 43–57, 2011. 2

[9] Yeo et al., "Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex," *IEEE Trans. Med. Imaging*, vol. 29, no. 7, pp. 1424–1441, Jul 2010. 2

[10] Glocker et al., "Dense Image Registration through MRFs and Efficient Linear Programming," *Medical Image Analysis*, vol. 12(6), pp. 731–741, 2008. 3

[11] Vandemeulebroucke et al., "Point-validated pixel-based breathing thorax model," in *Int'l Conf. on the Use of Comput. in Radiation Therapy*, Toronto, Canada, 2007. 3

[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. 3