Incorporating Prior Knowledge on Class Probabilities into Local Similarity Measures for Intermodality Image Registration

Matthias Hofmann[1,2,3], Bernhard Schölkopf[2],
Ilja Bezrukov[2,3], and Nathan D. Cahill[1,4]

[1] Department of Engineering Science, University of Oxford, Oxford, UK
[2] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[3] Department of Radiology, University Hospital Tübingen, Germany
[4] School of Mathematical Sciences, Rochester Institute of Technology, NY, USA

Abstract. We present a methodology for incorporating prior knowledge on class probabilities into the registration process. By using knowledge from the imaging modality, pre-segmentations, and/or probabilistic atlases, we construct vectors of class probabilities for each image voxel. By defining new image similarity measures for distribution-valued images, we show how the class probability images can be nonrigidly registered in a variational framework. An experiment on nonrigid registration of MR and CT full-body scans illustrates that the proposed technique outperforms standard mutual information (MI) and normalized mutual information (NMI) based registration techniques when measured in terms of target registration error (TRE) of manually labeled fiducials.

1 Introduction

Intermodality image registration describes the task of aligning two images that have been acquired using different modalities and thus have values with different meanings. Typical registration methods are designed in terms of three components: the allowable space of geometric transformations, the similarity measure relating two images, and the optimization technique for finding the optimal value of the similarity measure over the space of allowable transformations.

Mutual information (MI) [1, 2] is one of the most commonly used similarity measures in intermodality image registration. The underlying assumption is that if images values are treated as realizations of some underlying probability distribution function, then the mutual information between both images is maximized when they are aligned. The reason for the success of MI and related measures such as normalized mutual information (NMI) [3] lies in the fact that these measures do not make any underlying assumptions about the modalities. This allows them to be used for a whole range of inter-modality registration problems. In practice, however, most volume registration problems in medical imaging applications use images acquired from a limited number of modalities, including MR, CT, PET, and Ultrasound. When prior knowledge about the images, modalities, or optimal transformations is available, this knowledge can be

exploited to enable better registration results than would be possible with MI or NMI.

Prior knowledge has been previously exploited in a number of ways for image registration [4–9]. Ashburner et al. [4] have shown how incorporating prior knowledge about an object's shape and scale can significantly improve accuracy and reliability for affine registration. Leventon and Grimson [5], Guetter et al. [6], Chung et al. [7], and Sabuncu and Ramadge [8] have all proposed variations on the idea of learning the optimal joint distribution from exemplar aligned images and then designing a similarity measure that reflects the difference between the learned joint distribution and an observed joint distribution. Lee et al. [9] have focused directly on learning an optimal similarity measure for a specific pair of image modalities, assuming exemplar aligned images from those modalities are provided.

While these techniques are useful for certain intermodal registration problems, they seem to rely directly on the image intensities; hence, they may not be directly generalizable to all types of intermodal registration. Whereas CT scanners are calibrated to yield images with an accuracy of within a few Hounsfield units, this is not possible with MR scanners and MR typical field inhomogeneities lead to severe intensity bias. Furthermore, it is not necessarily true that exemplar aligned images will always be available; this is especially true in situations when nonrigid registration is required.

Other previous work has attempted to use prior information in the form of tissue class probabilities to aid in registration. D'Agostino et al. [10] describe how to align images by minimizing the Kullback-Leibler divergence between the actual joint class distribution and an ideal joint class distribution, although their formulation assumes that individual class probabilities in pairs of corresponding voxel positions in two images are independent. Lorenzen et al. [11] avoid making this assumption and propose an approach based on local KL divergence. Although Lorenzen's approach is applicable for the more general problem of *multiple* image registration, it requires the additional step of atlas construction even when only two images must be registered.

In this paper, we show how to design a nonrigid registration algorithm based on vector valued images. Such a nonrigid registration algorithm exploits prior knowledge about the imaging modalities, but unlike other techniques, it does not require exemplar aligned images, independence assumptions, or atlas construction. We compare the performance of the proposed nonrigid registration technique versus standard MI-based registration on full-body MR/CT images, measuring the results in terms of target registration error (TRE) on manually labeled fiducial points.

2 Prior Knowledge on Class Probabilities

In medical images, the intensity at a given voxel depends on the underlying tissue class at the voxel's position. If it is possible to infer the underlying tissue class at each position in an image, this knowledge could potentially be exploited by

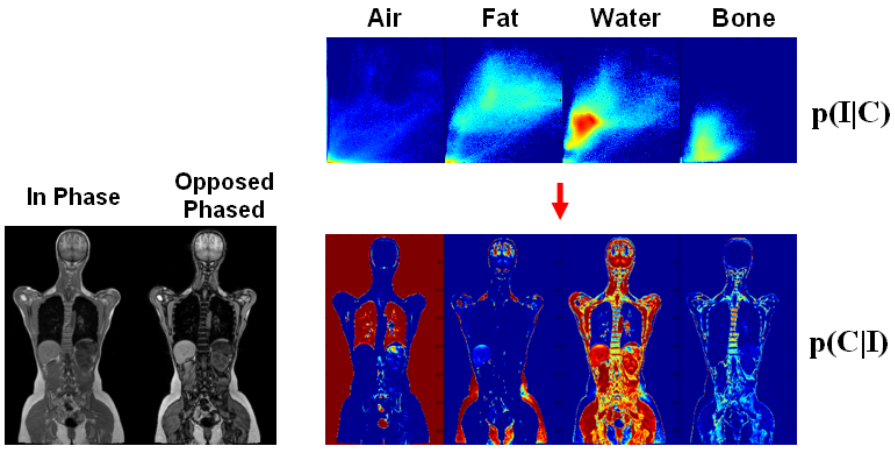

Fig. 1. Estimating class probabilities using Bayes' rule and pre-segmented images. Red/blue colors indicate high/low probabilities, respectfully.

a registration algorithm. In many situations, however, it is difficult to uniquely identify the underlying tissue class at a voxel solely from the intensity value at that voxel. This can be remedied by using additional information that may be available directly from the imaging device, from a pre-segmentation, or from a probabilistic atlas.

One example of additional information that may be provided directly from an imaging device is the phase information in MR imaging that is available after MR reconstruction [12]. This information is not usually used for diagnosis, but may be useful in inferring the underlying tissue class. If no additional direct source of information is available, class probabilities can be estimated based on the intensity values by using pre-segmented images. Assuming that pre-segmented images and the image to be registered have similar intensity characteristics, the probability that a particular voxel has tissue class C, given the voxel intensity I, is given by $p(C|I)$. According to Bayes' rule, $p(C|I) \propto p(I|C) \cdot p(C)$, where $p(C)$ is a prior probability on how often class C is expected to occur. Figure 1 provides an illustration of the process of estimating class probabilities at each voxel in a full-body MR scan.

Alternatively, if there is no direct source of additional information available, and pre-segmentation is not desirable, probabilistic atlases can be used to provide prior information about which tissue class is expected at different positions in the patient [13, 14]. Since the anatomy of different subjects shows structural similarities, the images of different subjects can be brought into alignment through intersubject registration. For example, a template MR image for which tissue labels are available can be registered to a patient MR image. Applying the same transformation on the label image yields a tissue label prediction for the patient MR image. Performing this procedure for several template MR image and label pairs can be used to derive a position-depedendent prior probability $p(C|X)$.

This prior probability can be combined with the intensity-derived class estimate using Bayes' rule:

$$p(C|I,X) \propto p(C|X) \cdot p(I|C) \tag{1}$$

Whether estimated directly from the imaging modality, or indirectly from pre-segmented images or probabilistic atlases, once class probabilities have been derived for each voxel, they can be used to form distribution-valued images. These distribution-valued images can be subsequently used for registration, hopefully yielding better results than registration based solely on the original intensity-valued images.

3 Nonrigid Registration Preliminaries

In this section, we describe a variational framework for registering two images, using MI or NMI as a similarity measure. In the next section, we will extend this framework and introduce new similarity measures to allow for vector-valued and distribution-valued images.

Consider two images, a reference image R and a floating image F, both as functions mapping $\Omega \subseteq \mathbb{R}^d \mapsto \mathbb{R}^N$. Define a *deformation* $\mathbf{\Phi} : \mathbb{R}^d \mapsto \mathbb{R}^d$ by $\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$, and call $\mathbf{u}$ the *displacement*. The general form of the variational registration problem is given by:

$$\min_{\mathbf{u}} \qquad \mathcal{E}(R,F,\mathbf{u}) := \mathcal{S}(\mathbf{u}) + \alpha \mathcal{J}(R,F^{\mathbf{u}}) \quad , \tag{2}$$

where $\mathcal{J}$ is a similarity measure that quantifies the similarity between the reference image R and the deformed floating image $F^{\mathbf{u}} := F(\mathbf{\Phi})$, $\mathcal{S}$ is a regularizer that ensures that the minimization problem is well-posed and that the solution is smooth in some sense, and α is a weighting parameter.

In this paper, we will assume the use of the curvature regularizer [15]:

$$\mathcal{S}(\mathbf{u}) := \int_{\Omega} \sum_{j=1}^{d} (\Delta \mathbf{u}_j(\mathbf{x}))^2 d\mathbf{x} \ . \tag{3}$$

Normally, variational registration proceeds by identifying the Euler-Lagrange equations associated with (2) and embedding them in an artificial time. When the curvature regularizer is used, this process yields the evolution equation:

$$\partial_t \mathbf{u}(\mathbf{x},t) + \Delta^2 \mathbf{u}(\mathbf{x},t) = -\alpha \mathbf{f}(\mathbf{x},R,F^{\mathbf{u}}) \,, \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{4}$$

$$\mathbf{u}(\mathbf{x},0) = \mathbf{u}^{(0)}(\mathbf{x}) \quad , \tag{5}$$

where $\mathbf{u}^{(0)}(\mathbf{x})$ is typically chosen to be 0, and where Dirichlet or Neumann boundary conditions are typically imposed on the boundary of Ω. The equilibrium solution of (4)–(5) yields a stationary point of (2).

To handle large deformations, we follow a strategy similar to fluid registration [16], and solve the evolution equation in terms of the velocity field $\mathbf{v}$ instead of

the displacement field $\mathbf{u}$; i.e.,

$$\partial_t \mathbf{v}(\mathbf{x},t) + \Delta^2 \mathbf{v}(\mathbf{x},t) = -\alpha \mathbf{f}(\mathbf{x}, R, F^{\mathbf{u}}), \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{6}$$

$$\mathbf{v}(\mathbf{x},0) = \mathbf{0} \ , \tag{7}$$

where the velocity and displacement are related by the material derivative:

$$\mathbf{v}(\mathbf{x},t) = \partial_t \mathbf{u}(\mathbf{x},t) + \left(\nabla \mathbf{u}(\mathbf{x},t)\right)^{\mathrm{T}} \mathbf{v}(\mathbf{x},t). \tag{8}$$

The vector field $\mathbf{f}(\mathbf{x}, R, F^{\mathbf{u}})$ that forms the nonhomogeneous term in (4) is called the *force vector*, and it arises from the Gâteaux derivative of $\mathcal{J}$. Force vectors can typically be determined analytically from the similarity measure at hand. For the MI and NMI, we first present the similarity measures and then their corresponding force vectors. The MI and NMI are defined by:

$$MI(R, F^{\mathbf{u}}) := H(R) + H(F^{\mathbf{u}}) - H(R, F^{\mathbf{u}}) \ , \tag{9}$$

$$NMI(R, F^{\mathbf{u}}) := \frac{H(R) + H(F^{\mathbf{u}})}{H(R, F^{\mathbf{u}})} \ , \tag{10}$$

where $H(A)$ and $H(A,B)$ are the marginal and joint entropies, defined by:

$$H(A) = -\int_{-\infty}^{\infty} p_A(a) \log p_A(a)\, da \ , \tag{11}$$

$$H(A,B) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{A,B}(a,b) \log p_{A,B}(a,b)\, da db \ , \tag{12}$$

and where $p_A(a)$ and $p_{A,B}(a,b)$ are the values of the probability density function of A at a, and the joint probability density function of (A,B) at (a,b), respectively.

The corresponding force vectors for MI and NMI are given by:

$$\mathbf{f}_{MI}(\mathbf{x}, R, F^{\mathbf{u}}) = \mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}) \ , \tag{13}$$

$$\mathbf{f}_{NMI}(\mathbf{x}, R, F^{\mathbf{u}}) = \frac{\mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) - NMI(R, F^{\mathbf{u}})\, \mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}})}{H(R, F^{\mathbf{u}})} \ , \tag{14}$$

where

$$\mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|} \left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \ , \tag{15}$$

$$\mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{1}{|\Omega|} \left(\frac{\frac{\partial}{\partial f}\left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f)\right]_{f=F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \ . \tag{16}$$

The force vector for MI is derived in [17], and the force vector for NMI follows a similar derivation.

Once the force vector has been chosen, the evolution equations (4) or (6) can be approximately solved by introducing a discretization that is backward in time and centered in space, and then by using Fourier methods [18] to resolve the linear system that arises at each time step. If (6) is being solved, the displacement field $\mathbf{u}$ can be computed from the velocity field $\mathbf{v}$ at each time step by Euler integration of (8).

4 Nonrigid Registration with Vector-Valued Images

If the images being registered are vector-valued or distribution-valued, the same nonrigid registration framework can be utilized as long as appropriate similarity measures and force vectors are defined.

We now denote the reference and floating images as functions mapping $\Omega \subseteq \mathbb{R}^d \mapsto \mathbb{R}^N$; namely, $\mathbf{R} = (R_1, \ldots, R_N)^{\mathrm{T}}$ and $\mathbf{F} = (F_1, \ldots, F_N)^{\mathrm{T}}$. The bold font indicates a vector quantity, whereas the regular font indicates a scalar quantity.

The variational registration problem (2) can still be used, but with R and F replaced by $\mathbf{R}$ and $\mathbf{F}$, respectively. We must now define similarity measures that operate on these vector-valued images. One such measure can be envisioned as the integral of the point-wise sum of squared differences between vector components. We call this similarity measure the local SSD (LSSD), and define it by:

$$LSSD(\mathbf{R}, \mathbf{F}^{\mathbf{u}}) := \int_{\Omega} \sum_{j=1}^{N} \left(R_j(\mathbf{x}) - F_j^{\mathbf{u}}(\mathbf{x})\right)^2 d\mathbf{x} \ . \tag{17}$$

The corresponding force vector is derived in the Appendix and is given by:

$$\mathbf{f}_{LSSD}(\mathbf{x}, \mathbf{R}, \mathbf{F}^{\mathbf{u}}) = 2\left[\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right]\left(\mathbf{R}(\mathbf{x}) - \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right) \ . \tag{18}$$

Note that the matrix $\nabla \mathbf{F}^{\mathbf{u}}$ is the $d \times N$ Jacobian matrix of $\mathbf{F}^{\mathbf{u}}$ defined componentwise by $\left[\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right]_{jk} = \frac{\partial}{\partial x_j} F_k^{\mathbf{u}}(\mathbf{x})$. The force vector (18) can be inserted into the evolution equations (4) or (6), and nonrigid regisration can proceed as described in the previous section.

If our vector-valued images contain vectors of class probabilities at each voxel, we can make the more restrictive assumption that not only are $\mathbf{R}$ and $\mathbf{F}$ vector-valued images, they are distribution-valued; i.e., $\mathbf{R}, \mathbf{F} : \Omega \subseteq \mathbb{R}^d \mapsto \mathbb{P}_N$, where $\mathbb{P}_N$ is the open probability simplex defined in [19]:

$$\mathbb{P}_N := \left\{ \mathbf{p} \,|\, \mathbf{p} = (p_1, \ldots, p_N) \in [0,1] \wedge \sum_{i=1}^{N} p_i = 1 \right\} \ . \tag{19}$$

Since the images are distribution-valued, other choices for the dissimilarity measure can be made based on probability metrics or premetrics [20]. In this article, we choose to design a similarity measure based on the Kullback-Leibler divergence.

The Kullback-Leibler divergence [21] is a premetric that describes the relative entropy between two distributions. For each position in the reference and floating images, it is defined by:

$$KL(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) := \sum_{j=1}^{N} F_j^{\mathbf{u}}(\mathbf{x}) \log\left(\frac{F_j^{\mathbf{u}}(\mathbf{x})}{R_j(\mathbf{x})}\right) \ , \tag{20}$$

An image similarity measure can be constructed by integrating (20) over the image domain. We call this similarity measure the local KL divergence (LKL),

and define it according to:

$$LKL(\mathbf{R}, \mathbf{F}^{\mathbf{u}}) := \int_{\Omega} KL(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x}))\, d\mathbf{x} \ . \tag{21}$$

The corresponding force vector is derived in the Appendix and is defined componentwise by:

$$[\mathbf{f}_{LKL}(\mathbf{x}, \mathbf{R}, \mathbf{F}^{\mathbf{u}})]_j := -\sum_{k=1}^{N} [\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})]_{j,k} \left(\log\left(\frac{F_k^{\mathbf{u}}(\mathbf{x})}{R_k(\mathbf{x})} \right) + 1 \right) \ . \tag{22}$$

As with the LSSD similarity measure, the LKL force vector can be inserted into the evolution equations (4) or (6), and nonrigid registration can proceed as described in the previous section.

Note that the LSSD and LKL are not the only similarity measures that can be designed for vector-valued or distribution-valued data. Theoretically, any distance measure that can be defined in a vector space or over probability distributions can be developed into a similarity measure for image registration. This shows a potential advantage over the registration approach of Lorenzen et al. [11]. Even though our LKL similarity measure is similar to the one used by Lorenzen, it does not require construction of an atlas; Lorenzen's atlas construction step appears to be possible only when KL divergence is used, and it does not appear to be easily generalizable to other similarity measures.

5 Registration Experiment

To validate the use of distribution-valued images in a practical registration, we developed an experiment on a dataset of 4 human whole body MR and CT images, where patients were scanned with their arms above the head for both the MR and the CT scan. Class probabilities for the classes air, fat, soft tissue and bone were computed for the MR and the CT image. For the MR images they were constructed from a normalized product of $p(I|C)$ and $p(C|I, X)$. Here, $p(I|C)$ was determined by histogram binning the intensities of all voxels that belong to each of the four classes, where the class assignment was determined by using a previously obtained segmentation. The quantity $p(C|I, X)$ was determined by averaging the result of several atlas registrations. For the acquisition of the MR images we used a protocol that acquires opposed phased and inphase images in one scan. Resolution was $2.6 \times 2.6 \times 2.6\text{mm}^3$ and TA = 18s, which allowed for breath hold acquisition. In the CT images, the intensity alone was informative enough and no atlas registration was necessary as an additional source of information. We assumed that for the CT images, the intensity distribution $p(I|C)$ could be approximated by a mixture of Gaussians.

We performed nonrigid registration, both with MI and NMI on the original MR and CT images, and with LSSD and LKL on class probability images. For the MI and NMI computations, histograms (and joint histograms) were estimated using 32 (and 32×32) bins. Linear (or bilinear) interpolation was used to

accumulate partial weights in neighboring bins. Registration in all cases was done over three levels in a multiresolution pyramid, with approximately isotropic sampling at 2.6mm in each dimension at the finest level, and 10.4mm in each dimension at the coarsest level. We varied the α parameter that trades off the contribution of the similarity measure and regularizer, and we experimented with various amounts of smoothing at each resolution level.

In order to assess errors in a physically meaningful sense, we manually placed approximately 30 landmarks of corresponding points in each MR and CT image pair, examples of which are shown in Fig. 2. These landmarks were unknown to the registration algorithms; they were only used retrospectively to compute target registration error (TRE).

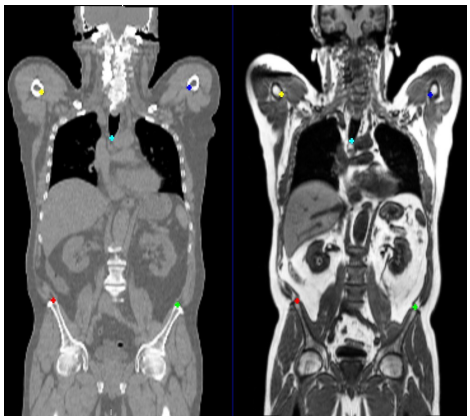

Fig. 2. Manually placed landmarks in CT and MR image.

6 Results

For each similarity measure, we selected the best performing set of values for α and smoothing amount at each resolution level. The resulting target registration errors (TRE) are reported for each patient in Table 1. Results indicate that there is an appreciable improvement in TRE for three out of the four patients when using the vector-valued similarity measures over the MI/NMI.

To illustrate an example registration, Fig. 3 shows fused images before and after a nonrigid registration algorithm using LKL was applied to one of the patient datasets. The left images are slices from an MR dataset that are treated as the reference image and remain unchanged; the right images are slices from a CT dataset that is in its original position (upper) and warped aligned position (lower). In the middle show the fused MR/CT slices, both before (upper) and after (lower) nonrigid registration.

	MI	NMI	LSSD	LKL
Patient 1 Median TRE	17.8	17.6	13.4	13.4
Patient 1 St. Dev. TRE	10.7	10.3	8.0	8.0
Patient 2 Median TRE	13.2	14.4	9.4	8.8
Patient 2 St. Dev. TRE	7.0	6.5	6.0	5.9
Patient 3 Median TRE	16.1	16.9	10.8	10.7
Patient 3 St. Dev. TRE	10.4	9.8	10.6	10.4
Patient 4 Median TRE	13.5	13.9	13.0	13.2
Patient 4 St. Dev. TRE	14.7	14.7	15.3	15.0

Table 1. TRE statistics (in mm) for each patient for each similarity measure.

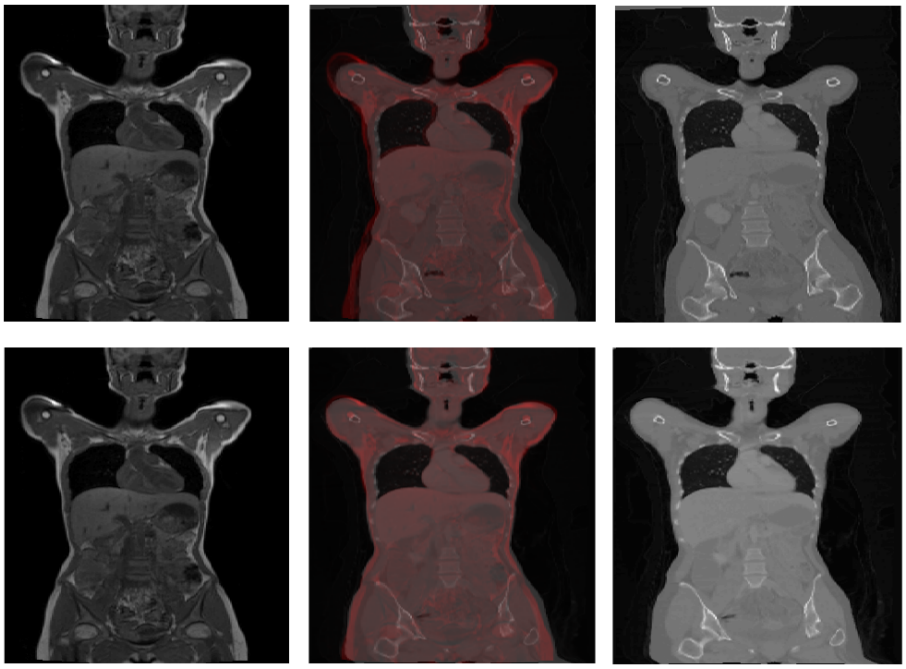

Fig. 3. Example visualization of nonrigid registration. Left images are identical MR slices; right images are CT slices before (upper) and after (lower) nonrigid registration. Middle images are fused versions of the MR/CT slices before (upper) and after (lower) registration.

7 Conclusion

In this article, we have shown that prior knowledge on class probabilities can be incorporated into intermodality image registration. Class probability vectors at each voxel can be constructed from extra information provided by an imaging device, from pre-segmentations, and/or from probabilistic atlases. Incorporation of vector-valued or distribution-valued images into registration requires defining new image similarity measures. In this paper, we presented two such measures, the local SSD, and the local KL divergence, which do not rely on independence assumptions or intermediate atlas formation. When tested on the problem of nonrigid registration of MR and CT full-body images, registration using the proposed measures outperformed standard MI/NMI-based registration a majority of the time when measured in terms of target registration error on manually identified fiducial markers.

Appendix

In this appendix, we derive the force vectors for the LSSD and LKL similarity measures defined in (17) and (21), respectively. We note that for a general similarity measure $\mathcal{J}(\mathbf{R}, \mathbf{F^u})$, the force vector $\mathbf{f}_{\mathcal{J}}(\mathbf{x}, \mathbf{R}, \mathbf{F^u})$ is related to the Gâteaux derivative of $\mathcal{J}$ in the following way:

$$d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) = \int_\Omega \left\langle \mathbf{f}_{\mathcal{J}}(\mathbf{x}, \mathbf{R}, \mathbf{F^u}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x} \ . \tag{23}$$

To aid in the derivation of the force vectors for the LSSD and LKL similarity measures, we present and prove the following Lemma.

Lemma 1 *If the dissimilarity measure $\mathcal{J}(\mathbf{R}, \mathbf{F^u})$ can be expressed as a functional of the form:*

$$\mathcal{J}(\mathbf{R}, \mathbf{F^u}) = \int_\Omega \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x}))\, d\mathbf{x} \ , \tag{24}$$

then the Gâteaux derivative $d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w})$ can be expressed as:

$$d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) = \int_\Omega \left\langle -\nabla \mathbf{F^u}(\mathbf{x}) \left[\nabla_{\mathbf{F^u}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x})) \right], \mathbf{w} \right\rangle d\mathbf{x} \ . \tag{25}$$

Proof. By the definition of the Gâteaux derivative, we have:

$$\begin{aligned} d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) &= \lim_{h \to 0} \frac{1}{h} \left[\int_\Omega \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hw}}(\mathbf{x})\big)\, d\mathbf{x} - \int_\Omega \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x}))\, d\mathbf{x} \right] \\ &= \lim_{h \to 0} \frac{1}{h} \int_\Omega \int_0^1 \frac{d}{ds} \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hsw}}(\mathbf{x})\big)\, ds\, d\mathbf{x} \\ &= \lim_{h \to 0} \frac{1}{h} \int_\Omega \int_0^1 \left\langle \frac{d}{ds} \mathbf{F^{u+hsw}}(\mathbf{x}), \nabla_{\mathbf{F^u}} \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hsw}}(\mathbf{x})\big) \right\rangle ds\, d\mathbf{x} \end{aligned}$$

$$= \lim_{h\to 0} \frac{1}{h} \int_\Omega \int_0^1 \left\langle \left\langle \nabla \mathbf{F}^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x}), -h\mathbf{w} \right\rangle, \nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\big) \right\rangle ds\, d\mathbf{x}$$
$$= \int_\Omega \int_0^1 \left\langle \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}), \mathbf{w} \right\rangle, \nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) \right\rangle ds\, d\mathbf{x}$$
$$= \int_\Omega \int_0^1 \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \left[\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x}))\right], \mathbf{w} \right\rangle ds\, d\mathbf{x}$$
$$= \int_\Omega \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \left[\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x}))\right], \mathbf{w} \right\rangle d\mathbf{x} \qquad \square$$

Now, the LSSD similarity measure (17) can be defined in terms of (24) with:

$$\mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \sum_{j=1}^{N} \left(R_j(\mathbf{x}) - F_j^{\mathbf{u}}(\mathbf{x})\right)^2 \ . \tag{26}$$

For this case, we have:

$$\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = -2\left(\mathbf{R}(\mathbf{x}) - \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right) \ , \tag{27}$$

and (18) is a direct result of Lemma 1 and (23).

The LKL similarity measure (17) can also be defined in terms of (24) with:

$$\mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = KL(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \sum_{j=1}^{N} F_j^{\mathbf{u}}(\mathbf{x}) \log\left(\frac{F_j^{\mathbf{u}}(\mathbf{x})}{R_j(\mathbf{x})}\right) \ . \tag{28}$$

Now for this case, we have:

$$\nabla_{F_k^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \log\left(\frac{F_k^{\mathbf{u}}(\mathbf{x})}{R_k(\mathbf{x})}\right) + 1 \ , \tag{29}$$

and (22) is a direct result of Lemma 1 and (23).

References

1. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical image analysis **1**(1) (1996) 35–51
2. Collignon, A.: Multi-modality Medical Image Registration by Maximization of Mutual Information. PhD thesis, Catholic University of Leuven, Leuven, Belgium (1998)
3. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32** (1999) 71–86
4. Ashburner, J., Neelin, P., Collins, D., Evans, A., Friston, K.: Incorporating prior knowledge into image registration. Neuroimage **6**(4) (1997) 344–352
5. Leventon, M., Grimson, W.: Multi-modal volume registration using joint intensity distribution. In: Proc. Medical Image Computing and Computer Assisted Intervention. (1998) 1057–1066

6. Guetter, C., Xu, C., Sauer, F., Hornegger, J.: Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. In: Proc. Medical Image Computing and Computer Assisted Intervention. (2005) 255–262
7. Chung, A.C.S., Gan, R., Wells III, W.M.: Robust multi-modal image registration based on prior joint intensity distributions and minimization of kullback-leibler distance. HKUST CSE Technical Report, HKUST-CS07-01 (2007)
8. Sabuncu, M., Ramadge, P.: Using spanning graphs for efficient image processing. IEEE Transactions on Image Processing (2008) 788–797
9. Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N.D., Schlkopf, B.: Learning the similarity measure for multi-modal 3d image registration. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2009) to appear.
10. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: An information theoretic approach for non-rigid image registration using voxel class probabilities. Medical Image Analysis **10** (2006) 413–431
11. Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S.: Multi-modal image set registration and atlas formation. Medical Image Analysis **10** (2006) 440–451
12. Bourgeat, P., Fripp, J., Stanwell, P., Ramadan, S., Ourselin, S.: MR image segmentation of the knee bone using phase information. Medical Image Analysis **11**(4) (2007) 325–335
13. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: theory and rationale for its development the international consortium for brain mapping (ICBM). Neuroimage **2**(2PA) (1995) 89–101
14. Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schlkopf, B., Pichler, B.J.: MR-based Attenuation Correction for PET/MR: A Novel Approach Combining Atlas Registration and Recognition of Local Patterns. Journal of Nuclear Medicine **49**(11) (2008) 1875–1883
15. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press (2004)
16. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing **5**(10) (October 1996) 1435–1447
17. Hermosillo Valadez, G.: Variational Methods for Multimodal Image Matching. PhD thesis, Université de Nice - Sophia Antipolis (2002)
18. Cahill, N.D., Noble, J.A., Hawkes, D.J.: Fourier methods for nonparametric image registration. In: Proceedings of the CVPR Workshop on Image Registration and Fusion. (June 2007) 1–8
19. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis **11** (2007) 465–477
20. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International Statistical Review **70** (2002) 419–435
21. Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics **22**(1) (1951) 79–86