Intrasubject Multimodal Groupwise Registration with the Conditional Template Entropy

Mathias Polfliet, Stefan Klein, Wyke Huizinga,
Margarethus M. Paulides, Wiro J. Niessen, Jef Vandemeulebroucke

PII: S1361-8415(18)30028-8
DOI: 10.1016/j.media.2018.02.003
Reference: MEDIMA 1339

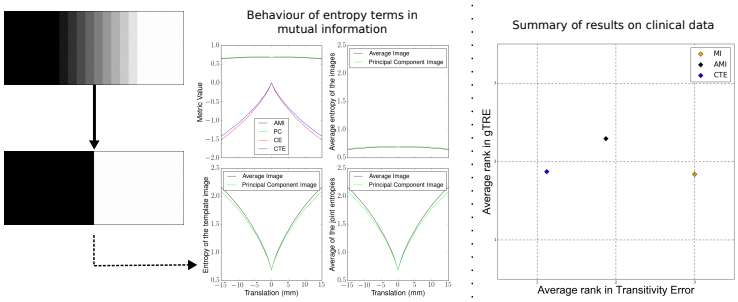
To appear in: *Medical Image Analysis*

Received date: 13 April 2017
Revised date: 6 February 2018
Accepted date: 14 February 2018

Please cite this article as: Mathias Polfliet, Stefan Klein, Wyke Huizinga, Margarethus M. Paulides, Wiro J. Niessen, Jef Vandemeulebroucke, Intrasubject Multimodal Groupwise Registration with the Conditional Template Entropy, *Medical Image Analysis* (2018), doi: 10.1016/j.media.2018.02.003

Highlights

- A novel similarity metric for multimodal groupwise registration is proposed.
- The proposed method showed equivalent or improved registration accuracy.
- Pairwise mutual information is outperformed in terms of transitivity.



Intrasubject Multimodal Groupwise Registration with the Conditional Template Entropy

Mathias Polfliet[a,b,c,*], Stefan Klein[c], Wyke Huizinga[c], Margarethus M. Paulides[d], Wiro J. Niessen[c,e], Jef Vandemeulebroucke[a,b]

[a] *Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics (ETRO), Pleinlaan 2, 1050 Brussels, Belgium*
[b] *imec, Kapeldreef 75, B-3001 Leuven, Belgium*
[c] *Biomedical Imaging Group Rotterdam, Depts. of Radiology & Medical Informatics, Erasmus MC, Rotterdam, the Netherlands*
[d] *Hyperthermia Unit, Department of Radiation Oncology, Erasmus MC Cancer Institute - University Medical Centre Rotterdam, the Netherlands*
[e] *Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, the Netherlands*

Abstract

Image registration is an important task in medical image analysis. Whereas most methods are designed for the registration of two images (pairwise registration), there is an increasing interest in simultaneously aligning more than two images using *groupwise* registration. Multimodal registration in a groupwise setting remains difficult, due to the lack of generally applicable similarity metrics. In this work, a novel similarity metric for such groupwise registration problems is proposed. The metric calculates the sum of the conditional entropy between each image in the group and a representative template image constructed iteratively using principal component analysis. The proposed metric is validated in extensive experiments on synthetic and intrasubject clinical image data. These experiments showed equivalent or improved registration accuracy compared to other state-of-the-art (dis)similarity metrics and improved transformation consistency compared to pairwise mutual information.

Keywords: groupwise image registration, multimodal, conditional entropy, principal component analysis, mutual information

1. Introduction

Biomedical image registration is the process of spatially aligning medical images, allowing for an accurate and quantitative comparison. An increasing number of image analysis tasks calls for the alignment of multiple (more than two) images. Examples include the joint analysis of tissue properties using multi-parametric MRI (Huizinga et al., 2016; Wells et al., 2015), spatio-temporal motion estimation from dynamic sequences (Metz et al., 2011; Vandemeulebroucke et al., 2011), atlas construction (Fletcher et al., 2009; Joshi et al., 2004; Wu et al., 2011) and population analyses (Geng et al., 2009).

One approach to perform such a registration task would be to take one image in the group as a reference and register all other images to this reference in a pairwise manner. However, such an approach has two distinct shortcomings. First, the choice of the reference image inherently biases the resulting transformations and subsequent data analysis towards the chosen reference. Secondly, only a fraction of the total information available within the group of images

*Corresponding author
Email address: mpolflie@etrovub.be (Mathias Polfliet)

is used in each pairwise registration, possibly leading to sub-optimal results.

An alternative is to perform a *groupwise* registration in which all transformations are optimized simultaneously. Transformations are expressed with respect to a common reference space, thereby removing the need for choosing a particular reference image, and the bias associated with that choice. Additionally, a global cost function simultaneously takes into account all information in the group of images. In this work we will address such groupwise similarity metrics for multimodal registration problems.

Multimodal intensity-based pairwise registration is commonly solved using mutual information (MI) (Collignon et al., 1995; Viola and Wells III, 1995; Wells et al., 1996), since it assumes a stochastic relationship between the two images to be registered. Extending MI to groupwise registration leads to a high-dimensional joint probability density function with an exponentially increasing number of histogram bins. Sparsity becomes a major concern as the number of images grows larger and limits the application to small groups of images (Wachinger and Navab, 2013).

A number of alternatives have been proposed to perform multimodal groupwise registration. Orchard and Mann (2010) proposed to use a Gaussian mixture model instead of histograms to approximate the joint probability density functions and Spiclin et al. (2012) approximated the joint probability density functions with a nonparametric approach based on a hierarchical intensity-space subdivision scheme. However, both approaches remain limited by the sparsity in the joint intensity space and perform poorly for large groups of images.

Alternatively, one could represent the intensities as a graph and relate the length of such a graph to the entropy of the images (Hero et al., 2002). Such an approach requires a computationally expensive optimization for the construction of the graph and is not continuously differentiable, making gradient-based optimization difficult.

Zöller et al. (2005) proposed the use of a voxelwise stack entropy. Herein, the intensities of all separate images in the group at a given sampled coordinate are grouped into a one-dimensional probability density distribution. For each sampled coordinate, the entropy is calculated and summed. However, for a low number of images in the group, the probability density functions are sparse which limits its use to larger groups of images.

Wachinger et al. (2007) proposed to accumulate all pairwise estimates of mutual information for all possible pairs of images in the group under consideration. Such an approach leads to a computation time which is proportional to the square of the number of images, making its application to larger groups of images increasingly difficult.

Joshi et al. (2004) developed an interesting metric where the mean squared differences is used as a pairwise metric to compare every image in the group to the average image. Herein the average image is updated in each iteration. They applied the method to monomodal brain atlas construction and it has also been applied to thoracic 4D CT data (Metz et al., 2011) and 4D ultrasound of the liver (Vijayan et al., 2014). The approach carries a number of advantages, such as the linear scaling of the computational complexity with respect to the number of images in the group and the possibility to parallelize the algorithm, making it feasible for both small and large groups of images. Bhatia et al. (2007) proposed to use the normalized mutual information (Studholme et al., 1999) as a pairwise similarity metric and the average image as a template image on monomodal intersubject data. The metric was termed the average normalized mutual information and has been used (together with the average mutual information) in subsequent literature as a metric for multimodal groupwise registrations (Ceranka et al., 2017; Hallack et al., 2014; Huizinga et al., 2016; Polfliet et al., 2016, 2017). However, the use of the average image as the template image might not be appropriate in multimodal data with intensities of varying scales, ranges and contrast.

In this work a novel similarity metric, the conditional template entropy (CTE), is introduced for multimodal groupwise registration based on this principle of pairwise similarity with respect to a template image. Following the original formulation by Joshi et al. (2004), we first design a suitable pairwise metric to be used in the comparison of the template image and

every image in the group. Afterwards we investigate the use of a template image based on principal component analysis.

Given the linear scaling of the computational complexity, the metric can be applied to a wide range of intrasubject multimodal groupwise registration problems, for both small and large groups of images, and can be used as a general purpose metric. The proposed metric is validated in extensive experiments on synthetic and intrasubject clinical data, demonstrating equivalent or improved registration accuracy compared to other state-of-the-art methods and improved transformation consistency compared to pairwise MI.

2. Materials and Methods

2.1. Pairwise Registration

In pairwise registration, a target (moving, floating) image I_T is registered to a reference (fixed, source) image I_R. The transformation $\mathcal{T}_{\boldsymbol{\theta}}$, parameterized by $\boldsymbol{\theta}$, needs to be determined that maps coordinates from the reference image domain to the target image domain (Fig. 1(a)). The registration can be defined as an optimization problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{C}\left(I_R, I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) \ . \tag{1}$$

Here, $\mathcal{C}$ is the cost function or objective value of the registration problem, which is often represented as a weighted sum of a dissimilarity metric, $\mathcal{D}$, and a regularization term, $\mathcal{R}$, such that

$$\mathcal{C} = \mathcal{D} + \lambda\mathcal{R} \ , \tag{2}$$

in which λ is the weight for the regularization.

2.2. Mutual information

In the pairwise approach, mutual information (MI) (Collignon et al., 1995; Viola and Wells III, 1995; Wells et al., 1996) is defined as a similarity metric ($\mathcal{S} = -\mathcal{D}$)

$$\begin{aligned}&\mathcal{S}_{MI}\left(I_R, I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) = \\ &H\left(I_R\right) + H\left(I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) - H\left(I_R, I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) \ .\end{aligned} \tag{3}$$

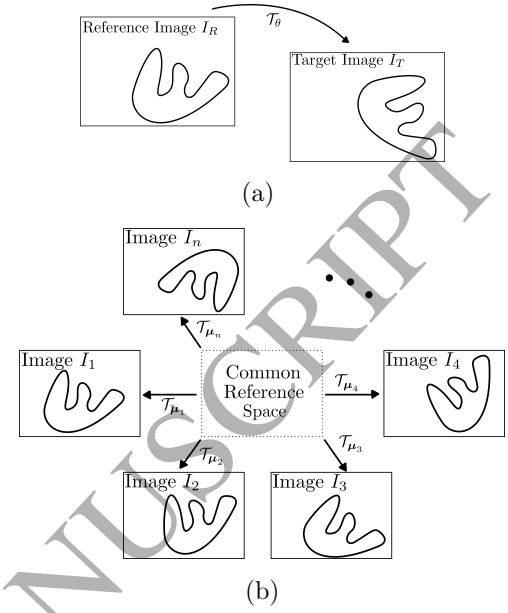


Figure 1: Graphical illustration for (a) pairwise registration (b) groupwise registration

Here, $H(\cdot)$ and $H(\cdot,\cdot)$ refer to, respectively, the marginal and joint entropy of the marginal and joint intensity distributions, often calculated via normalized histograms. In Eq. (3), the first term expresses the complexity of the reference image and the second term is the entropy of the target image mapped onto the reference, which favors transformations that map onto complex parts of the target image. The final term expresses the complexity of the shared or common relationship between the reference and target image. It is maximized when the (statistical or stochastic) relationship is stronger and thus less complex (Wells et al., 1996).

Following Maes et al. (1997), MI can be rewritten in terms of the conditional entropy (CE)

$$\mathcal{S}_{MI}\left(I_R, I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) = H\left(I_R\right) - H\left(I_R | I_T \circ \mathcal{T}_{\boldsymbol{\theta}}\right) \ . \tag{4}$$

The conditional entropy $H\left(A|B\right)$ describes the amount of information that remains in a random variable A once the random variable B is known. With the entropy of the reference image being independent of the transformation parameters, maximization of

the negated conditional entropy and maximization of the mutual information lead to equivalent solutions of the registration problem.

2.3. Groupwise registration

In groupwise registration we consider a group of n images for which the transformations to a common reference frame are unknown. We can consider the following optimization problem to determine these transformations:

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}\left(I_1 \circ \mathcal{T}_{\boldsymbol{\mu}_1}, \ldots, I_n \circ \mathcal{T}_{\boldsymbol{\mu}_n}\right) \quad , \tag{5}$$

where $\mathcal{T}_{\boldsymbol{\mu}_i}$ is the transformation, parameterized by $\boldsymbol{\mu}_i$, that maps the coordinates from the common reference domain to the domain of the i^{th} image (Fig. 1(b)). $\boldsymbol{\mu}$ is the vector formed by the concatenation of all separate transformation parameters $\boldsymbol{\mu}_i$, and I_i is the continuous intensity function of the i^{th} image.

2.4. Template construction

Joshi et al. (2004) proposed the following formulation for monomodal groupwise registration, in which both the transformation parameters and a template image are optimized

$$\begin{aligned}\hat{\boldsymbol{\mu}}, \hat{J} = &\arg\min_{\boldsymbol{\mu}, J} \frac{1}{n|S|} \\ &\sum_{i=1}^{n} \sum_{\boldsymbol{x} \in S} \left(I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x}) - J(\boldsymbol{x})\right)^2 \quad ,\end{aligned} \tag{6}$$

with J the continuous intensity function of a template image, $\boldsymbol{x}$ the coordinate samples drawn from the image and S the set of these samples. The template image can be interpreted as being the image that is most similar to the other images in the group in terms of the mean squared differences. For a given value of the transform parameters, the optimization with respect to the template image J was solved analytically to be the average image

$$J(\boldsymbol{x}) = \overline{I}_{\boldsymbol{\mu}}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x}) \quad . \tag{7}$$

As such, the registration problem in Joshi et al. (2004) is reduced to

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \frac{1}{n|S|} \sum_{i=1}^{n} \sum_{\boldsymbol{x} \in S} \left(I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x}) - \overline{I}_{\boldsymbol{\mu}}(\boldsymbol{x})\right)^2 \quad . \tag{8}$$

2.5. The conditional template entropy

In this work, a novel similarity metric for multimodal groupwise registration is proposed, based on this paradigm in which similarity of the group of images is measured with respect to an iteratively updated template image. Considering the interpretation of the entropy terms given in Section 2.2, we propose to measure similarity using the negated joint entropy of each image in the group with the template image, favoring transformations for which the template explains the group of images well; and the marginal entropies of each image in the group, encouraging transformations that map onto complex parts of the images in the group. Note that this is equivalent to a formulation based on the conditional entropy:

$$\begin{aligned}\hat{\boldsymbol{\mu}}, \hat{J} &= \arg\max_{\boldsymbol{\mu}, J} \frac{1}{n} \sum_{i=1}^{n} H\left(I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}\right) - H\left(J, I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}\right) \\ &= \arg\max_{\boldsymbol{\mu}, J} -\frac{1}{n} \sum_{i=1}^{n} H\left(J | I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}\right) \quad .\end{aligned} \tag{9}$$

Observing the resulting metric, one can notice the resemblance with a formulation based on mutual information. The difference lies in the absence of the marginal entropy of the template image, $H(J)$. As we will demonstrate, this term counteracts the alignment of the group of images. A representative template image is likely to grow sharper when converging towards the optimal registration solution, leading to a reduced complexity of its intensity distribution and a decrease in the marginal entropy, which is opposite of the desired optimization behavior. The proposed method based on conditional entropy as shown in Eq. (9) eliminates this problem.

To find the appropriate template image, we revisit Eq. (6) where the template image could be obtained analytically as the average image. Unfortunately, Eq. (9) cannot be solved analytically with respect to the template image, J, for a given set of transformations if the trivial solution of a constant template image with a single intensity is excluded. Hypothetically, one could set up an optimization scheme where the template image is predefined by a functional relationship and weights corresponding to the images in the group. Herein, the optimization of the transformation parameters could be alternated with the optimization of the weights for the template image. Such nested optimization is error-prone and costly, and undesirable in this context.

Alternatively, instead of maximizing Eq. (9), we propose a more pragmatic approach which maximizes the variance in the template image. By defining J as the linear combination of the images in the group, principal component analysis (PCA) can be used to find the weights associated to the images. This has previously been shown to reduce the noise due to motion in the template image (Melbourne et al., 2007). Additionally, negatively correlated intensities can be accounted for to increase the contrast in the template image, instead of decreasing the contrast as might be the case for simple intensity averaging.

PCA defines a linear transformation from a given high-dimensional space to a low-dimensional subspace whilst retaining as much variance as possible. In this work, PCA is performed with each sampled coordinate as a separate observation and the different images in the group corresponding to different features. The transformation to the 1-dimensional subspace along which the most variance is observed, is given by the eigenvector associated with the largest eigenvalue. As such, the elements of this eigenvector can serve as the weights for the construction of the template image.

$$J\left(\boldsymbol{x}\right) = I_{\boldsymbol{\mu}}^{PCA}\left(\boldsymbol{x}\right) = \sum_{i=1}^{n} v_{i,\boldsymbol{\mu}}\; I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}\left(\boldsymbol{x}\right) \quad . \tag{10}$$

Here, $\boldsymbol{v}_{\boldsymbol{\mu}}$ is the eigenvector associated with the largest eigenvalue and the subscript $\boldsymbol{\mu}$ is added to show its dependence on the transformation parameters. This template image, based on the principal component of the PCA, will hereafter be referred to as the principal component image.

Combining(9) and (10) leads to a novel similarity metric, the conditional template entropy (CTE), where similarity is expressed as the sum of the conditional entropy between every image in the group and the principal component image:

$$\begin{aligned}&\mathcal{S}_{CTE}\left(I_1 \circ \mathcal{T}_{\boldsymbol{\mu}_1}, \ldots, I_n \circ \mathcal{T}_{\boldsymbol{\mu}_n}\right)\\ &= -\frac{1}{n}\sum_{i=1}^{n} H\left(I_{\boldsymbol{\mu}}^{PCA} | I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}\right) \quad .\end{aligned} \tag{11}$$

2.6. Optimization

The proposed metric was implemented as part of the software package `elastix` (Klein et al., 2010) and is publicly available. An adaptive stochastic gradient descent was employed to minimize the cost function (Klein et al., 2009). As such, the negated form of Eq. (11) is used, to allow a minimization to take place. The derivative of the proposed metric with respect to $\boldsymbol{\mu}$ was determined following the approach of Thévenaz and Unser (2000) in which B-splines were used as a Parzen windowing function such that the joint probability density functions p_i between the template image and the i^{th} image in the group become

$$\begin{aligned}p_i\left(\iota, \kappa; \boldsymbol{\mu}\right) = \alpha \sum_{\boldsymbol{x}} \Bigg[&\beta^m\left(\frac{\iota}{\epsilon_{PCA}} - \frac{I_{\boldsymbol{\mu}}^{PCA}\left(\boldsymbol{x}\right)}{\epsilon_{PCA}}\right)\\ &\beta^m\left(\frac{\kappa}{\epsilon_i} - \frac{I_i\left(\mathcal{T}_{\boldsymbol{\mu}_i}\left(\boldsymbol{x}\right)\right)}{\epsilon_i}\right)\Bigg] \quad .\end{aligned} \tag{12}$$

Here, α is a normalization factor to obtain a density function, ϵ is related to the width of the histogram bin and β^m is a B-spline function of the order of m. ι and κ are the discretized intensities corresponding to the template image and images in the group, respectively. With B-splines fulfilling the partition of unity constraint (Thévenaz and Unser, 2000), we have

$$\sum_{\iota \in L_{PCA}} \sum_{\kappa \in L_i} \frac{\partial p_i\left(\iota, \kappa; \boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}} = 0 \quad \forall i \;, \tag{13}$$

where L_{PCA} and L_i are the discrete sets of intensities associated with the principal component and the i^{th} image. This leads to

$$\frac{\partial \mathcal{S}_{CTE}}{\partial \boldsymbol{\mu}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{\iota\in L_{PCA}}\sum_{\kappa\in L_i}\frac{\partial p_i(\iota,\kappa;\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\log\frac{p_i(\iota,\kappa;\boldsymbol{\mu})}{p_{I_i}(\kappa;\boldsymbol{\mu})} \tag{14}$$

With $p_{I_i}(\kappa;\boldsymbol{\mu}_i)$ the probability density function of the i^{th} image. In Appendix A the derivative of the principal component image with respect to the transformation parameters is given.

2.7. Transformation degeneracy

Given the degeneracy of estimating n transformations for n images with an arbitrary global transformation, we chose to constrain our transformation following Bhatia et al. (2004) with

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x}) = \boldsymbol{x}, \quad \forall \boldsymbol{x} \ , \tag{15}$$

i.e the sum of all transformations is the identity, effectively registering the group of images to the mean space. With Rosen's Gradient Projection Method (Luenberger, 1973) this is solved by setting

$$\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}_i}' = \frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}_i} - \frac{1}{n}\sum_{j=1}^{n}\frac{\partial \mathcal{C}}{\partial \boldsymbol{\mu}_j} \ . \tag{16}$$

and using this projected gradient in the stochastic gradient descent optimization.

2.8. Regularization

Following Geng et al. (2009) we used a groupwise regularization term, the groupwise bending energy (GBE)

$$\mathcal{R}_{GBE}(\mathcal{T}_{\boldsymbol{\mu}_1},\ldots,\mathcal{T}_{\boldsymbol{\mu}_n}) = \frac{1}{|S|}\sum_{\boldsymbol{x}\in S}\frac{1}{n}\sum_{i=1}^{n}\sum_{l,m=1}^{d}\left|\left|\frac{\partial^2 \mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x})}{\partial x_l \partial x_m}\right|\right|^2 . \tag{17}$$

Herein, d is the spatial dimension of the images. Regularization was performed in all clinical experiments with a deformable transformation model.

3. Data and Experiments

A total of six experiments were conducted with two on synthetic data and four on clinical intrasubject data. Herein, the proposed conditional template entropy ($\mathcal{S}_{CTE}$) was compared to the average mutual information ($\mathcal{S}_{AMI}$)

$$\mathcal{S}_{AMI}(I_1\circ\mathcal{T}_{\boldsymbol{\mu}_1},\ldots,I_n\circ\mathcal{T}_{\boldsymbol{\mu}_n}) = \frac{1}{n}\sum_{i=1}^{n}\Big[H(\bar{I}_{\boldsymbol{\mu}}) + H(I_i\circ\mathcal{T}_{\boldsymbol{\mu}_i}) - H(\bar{I}_{\boldsymbol{\mu}}, I_i\circ\mathcal{T}_{\boldsymbol{\mu}_i})\Big] \ . \tag{18}$$

Furthermore, two auxiliary similarity metrics were implemented to investigate complementary advantages of the proposed methodology, respectively the advantage of using the conditional entropy (S_{CE}) and the advantage of using the principal component image (S_{PC}).

$$\mathcal{S}_{CE}(I_1\circ\mathcal{T}_{\boldsymbol{\mu}_1},\ldots,I_n\circ\mathcal{T}_{\boldsymbol{\mu}_n}) = -\frac{1}{n}\sum_{i=1}^{n}H(\bar{I}_{\boldsymbol{\mu}}|I_i\circ\mathcal{T}_{\boldsymbol{\mu}_i}) , \tag{19}$$

$$\mathcal{S}_{PC}(I_1\circ\mathcal{T}_{\boldsymbol{\mu}_1},\ldots,I_n\circ\mathcal{T}_{\boldsymbol{\mu}_n}) = \frac{1}{n}\sum_{i=1}^{n}\Big[H(I_{\boldsymbol{\mu}}^{PCA}) + H(I_i\circ\mathcal{T}_{\boldsymbol{\mu}_i}) - H(I_{\boldsymbol{\mu}}^{PCA}, I_i\circ\mathcal{T}_{\boldsymbol{\mu}_i})\Big] \ . \tag{20}$$

For the clinical data, the four previously discussed groupwise similarity metrics were used in addition to the PCA2 metric proposed in Huizinga et al. (2016) and pairwise MI (Eq. 3) as a baseline for comparison. PCA2 was proposed for the registration of images for which the intensity distribution could be represented into a low-dimensional subspace and is given as

$$\mathcal{D}_{PCA2}(I_1\circ\mathcal{T}_{\boldsymbol{\mu}_1},\ldots,I_n\circ\mathcal{T}_{\boldsymbol{\mu}_n}) = \sum_{i=1}^{n} i\lambda_i \ . \tag{21}$$

Herein, λ_i refers to the i^{th} eigenvalue of the correlation matrix of the images in the group. In Huizinga

et al. (2016) it was subsequently validated on monomodal and quantitative MRI image data for which such a low-dimensional subspace exists. PCA2 can be thus considered as a specialist metric specifically designed to register such images. To demonstrate the more generic nature of the proposed methodology, CTE was compared to PCA2 for both quantitative MRI and multimodal image data.

All registrations were performed in an intrasubject manner and the images were normalized by z-scoring to allow for a fair comparison to the similarity metrics employing the average image. In the pairwise registration of a group of images, one image (the first in the sequence) was chosen as a reference to which all others were mapped. Note that other strategies for choosing the reference image in pairwise registrations for a group exist, such as the pre-contrast image in dynamic contrast enhanced sequences (Kim et al., 2011), the end-expiration in 4D CT (Saito et al., 2009) or the mid-way image in computational anatomy (Reuter et al., 2010).

As the optimization strategy, interpolation algorithm, random sampler and transformation model is equivalent for all (dis)similarity metrics, any difference in results can be solely attributed to the use of a different dissimilarity metric.

The proposed methods were validated with two validation criteria. First, the groupwise target registration error (gTRE)

$$\mathrm{gTRE}\left(\boldsymbol{\mu}\right) = \frac{1}{n}\sum_{i\neq r}^{n}\frac{1}{|P_i|}\sum_{j}^{|P_i|}||\mathcal{T}_{i,r}\left(\boldsymbol{p}_{i,j}\right) - \boldsymbol{p}_{r,j}|| \quad (22)$$

was used as a measure for the accuracy of the registration with ground truth annotations of certain anatomical landmarks in the images. In Eq. (22) r is the index of the reference image, P_i the collection of landmarks in the i^{th} image, $\mathcal{T}_{i,r}$ the transformation that maps the coordinates from the i^{th} image to the reference image and $\boldsymbol{p}_{i,j}$ the j^{th} landmark from the i^{th} image. In a groupwise setting $\mathcal{T}_{i,r}$ was determined through the composition of the forward transformation, that maps the coordinates from the common reference space to the reference image, with the inverse transformation, that maps the coordinates from the i^{th} image to the common reference space: $\mathcal{T}_{i,r} = \mathcal{T}_{\boldsymbol{\mu}_r} \circ \mathcal{T}_{\boldsymbol{\mu}_i}^{-1}$(Fig. 2) (Metz et al., 2011). To allow for a fair comparison between pairwise and groupwise registrations, all validation measurements were performed in the same reference space, i.e. the same image which was chosen as a reference in the pairwise registrations.

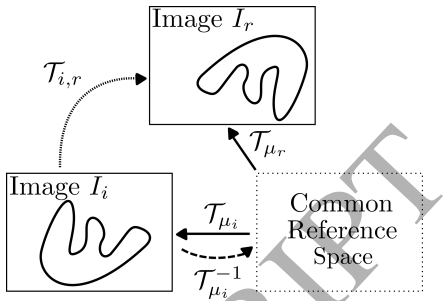


Figure 2: Composition of $\mathcal{T}_{\boldsymbol{\mu}_r}$ and $\mathcal{T}_{\boldsymbol{\mu}_i}^{-1}$ to obtain $\mathcal{T}_{i,r}$

Secondly, we computed the transitivity error (Christensen et al., 2006; Metz et al., 2011) to assess the quality of the transformation

$$\mathrm{Tra}\left(\boldsymbol{\mu}\right) = \frac{1}{|S|}\sum_{\boldsymbol{x}\in S}\sum_{i}^{n}\sum_{l\neq i}^{n}||\mathcal{T}_{i,r}\left(\boldsymbol{x}\right) - \mathcal{T}_{i,l}\left(\mathcal{T}_{l,r}\left(\boldsymbol{x}\right)\right)|| \ . \quad (23)$$

The transitivity error measures the transitive property of the transformations in a group of images and can be interpreted as a measure for the consistency of the transformations in a groupwise setting. For pairwise registration the use of different reference images is required to measure the transitivity and the bias associated with the choice will influence the results, whereas in groupwise registration, all transformations are estimated simultaneously and are inherently transitive (when the inverse transformation is available). As the inverse is approximated iteratively and the source for the transitivity error in the groupwise methods, no comparisons are made among the groupwise metrics based on the transitivity error. The maximum transitivity error of the groupwise methods is reported and compared to the transitivity error of the pairwise method.

The cost function hyperparameters (the number of histogram bins and regularization weight) were chosen such that they optimized the mean gTRE per

dataset. The different regularization weights are reported in Table 1. Due to the arbitrary sign of the projection vector for the principal component image, the number of histogram bins (used to calulate the entropy) are at least doubled compared to the number of histogram bins in registrations using the average image. Other optimization hyperparameters such as the spatial samples in the stochastic optimizer and the number of iterations were set to their default value. All registration hyperparameters in pairwise registrations were kept equal to those in the groupwise approach.

Results for the gTRE were compared in a pairwise manner among all similarity metrics (totaling 64 comparisons). The Wilcoxon signed-rank test was used for significance testing at a significance level of 0.05 adjusted by the Bonferroni correction for multiple comparisons.

3.1. Black&White

To illustrate the effect the entropy term of the template image has on the optimization, an experiment was performed on synthetic data. Eleven identical black-and-white images were progressively and simultaneously translated along the horizontal axis and the similarity metric values were computed. A mask was used to keep the sampling domain constant. Fig. 3 shows a single black-and-white image and the average image of the group of images when they are at maximal displacement (15 mm).

3.2. Multimodal Cubes

To further investigate registration accuracy, 100 registrations were performed on a group of six images (256 × 256 × 256 voxels) each containing two cubes, one surrounding the other. The intensities of the cubes and the backgrounds were set at random intensities to simulate a multimodal setting (Fig. 4). For each group of images a random set of deformable transformations was generated with a grid spacing of 8×8×8 voxels. The gTRE of the corners of the cubes was used to quantify the registration accuracy.

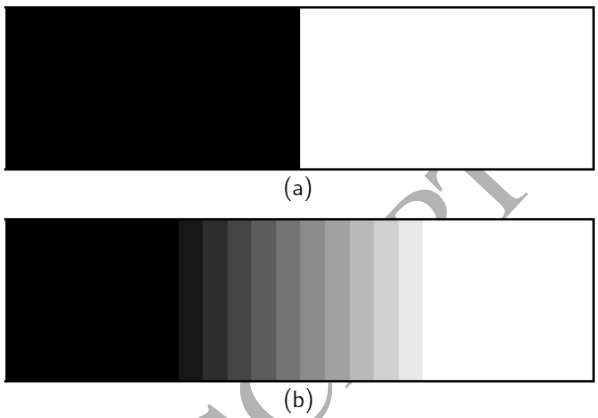


Figure 3: (a) A single black-and-white image (b) Average image of the group at their maximal misalignment.

3.3. Thoracic 4D CT

Thoracic 4D CT data (Fig. 5) was taken from the publicly available POPI and DIR-LAB datasets which include, respectively, 6 and 10 sequences of 10 respiratory phases each (Castillo et al., 2009; Vandemeulebroucke et al., 2011). Thoracic 4D CT data is often considered as monomodal data. However, minor intensity changes can occur due to changes in the voxel density in the lungs associated with the inhalation and exhalation of air (Sarrut et al., 2006) leading several authors to employ adapted or multimodal metrics for lung registration (Murphy et al., 2011).

The POPI dataset contains three patients with 100 manually identified landmarks in the lungs for every breathing phase and three patients with 100 landmarks in end-inspiration and end-expiration phases with an inter-rater error of 0.5±0.9 mm. In the DIR-LAB dataset, all patients have 300 landmarks in the lungs for the inspiration and expiration phases and 75 in the four phases in between and an intra-rater error between 0.70 and 1.13 mm. Accuracy of the registration was determined using the gTRE with respect to the inspiration phase, the first image in the dynamic series.

A deformable registration was performed using cubic B-splines with a final grid spacing of 12.0 mm. Lung masks were used and obtained following Van-

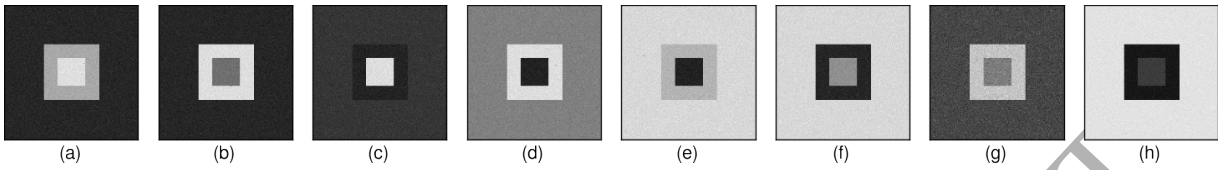


Figure 4: (a-f) A single slice of the six cubes used in the Multimodal Cubes experiment. (g) The average image and (h) the principal component image at alignment.

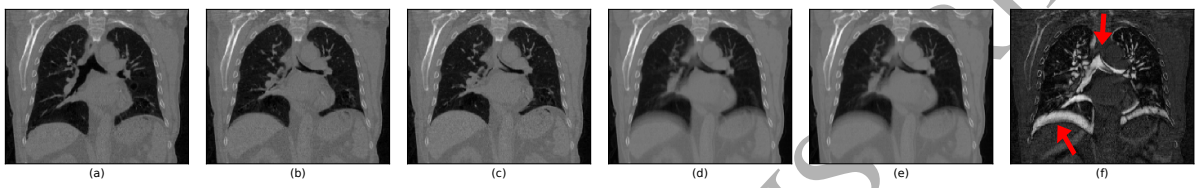


Figure 5: (a - c) Three of the ten phases used in the Thoracic 4D CT experiment. The images differ mainly in the position of the diaphragm and structures in the lungs due to breathing. (d) The average image at misalignment. (e) The principal component at misalignment. (f) Absolute difference image of the average and principal component image. Note that the largest differences occur in regions where motion is present (i.e. the diaphragm), indicated by red arrows. The image contrast is optimized for the range of intensities present in each individual image.

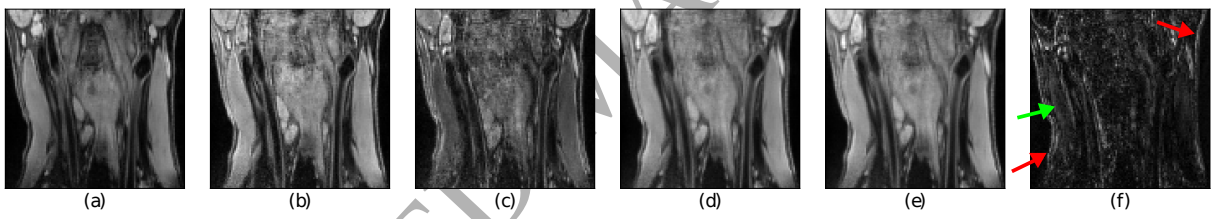


Figure 6: (a-c)Three of the five images used in the Carotid MR experiment. (d) The average image at misalignment. (e) The principal component at misalignment. (f) Absolute difference image of the average and principal component image. Note that the largest differences occur either at borders of structures due to motion, indicated by red arrows, or in homogeneous regions due to the multimodal nature of the data, indicated by a green arrow. The image contrast is optimized for the range of intensities present in each individual image.

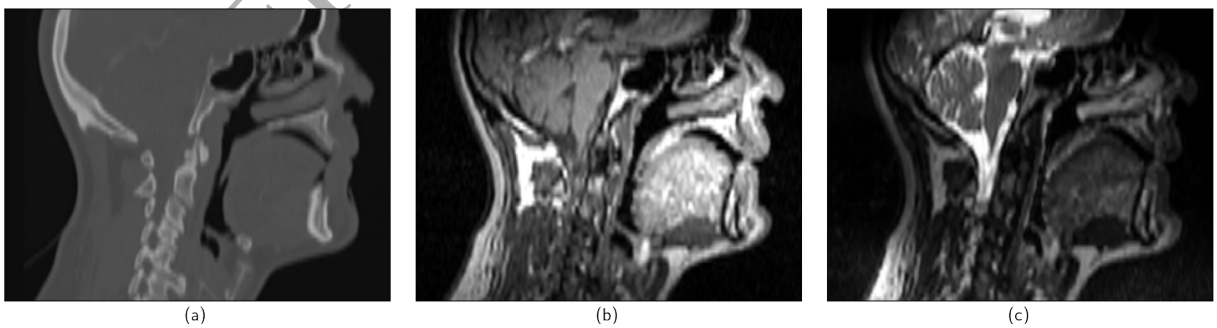


Figure 7: (a) CT image, (b) MR-T1 image and (c) MR-T2 image used in the Head&Neck experiment.

Table 1: The regularization weights used for each metric and clinical dataset.

	Thoracic 4D CT	Carotid MR	Head&Neck	RIRE
PCA2	500	100	2×10^6	-
MI	0.02	50	100	-
AMI	0.05	100	2000	-
PC	0.2	100	2000	-
CE	0.01	100	5000	-
CTE	0.2	100	5000	-

Table 2: Summary of the registration parameters used in the experiments. Two values are reported for the number of histogram bins, separated by a forward slash. The first value reflects the number of bins used in pairwise registration and groupwise registrations based on the average image. The second value gives the number of bins used in groupwise registrations based on the principal component image. Values separated with a backward slash indicate multiple settings within the applied optimization strategy.

Dataset	Histogram bins	Resolutions	Grid spacing	Spatial samples	Iterations
Multimodal Cubes	32/96	2	6.0	2048	2000
Thoracic 4D CT	48/96	4	12.0	2048	2000\4000
Carotid MR	48/128	2	8.0	2048	2000
Head&Neck	64/144	2	64.0	2048	2000
RIRE	48/128	5\2	-	2048	2000

demeulebroucke et al. (2012). For each resolution level 2000 iterations were performed, except for the last resolution where 4000 iterations were allowed.

3.4. Carotid MR

MR image sequences were acquired of the carotid artery by Coolen et al. (2015). The acquisitions were performed with a gradient echo MRI sequence for different flip angles and TE preparation times (Fig. 6). Each sequence consisted of five images and was performed for eight patients. The bifurcation of both carotid arteries was identified for each patient and consequently used as a landmark in the validation of the registration.

For this data we performed a deformable registration with cubic B-splines and a final grid spacing of 8.0 mm. van 't Klooster et al. (2013) has shown that a deformable registration is needed in such acquisitions of the carotid arteries. Masks around the carotid arteries were used as region of interest for registration.

3.5. Head&Neck

As part of radiotherapy planning, 22 patients underwent a CT, MR-T1 and MR-T2 imaging protocol of the head and neck region (Fortunati et al., 2014, 2015; Verhaart et al., 2014)(Fig. 7). In each acquisition between 15 to 21 landmarks were used to quantify the registration accuracy in terms of gTRE. The intra-rater variability of the landmarks was approximately 1mm.

Prior to registration, all images were resampled to the smallest voxel spacing present in the group of images. A deformable transformation was used in two resolution levels using cubic B-splines with a final grid spacing of 64.0 mm, as suggested by Fortunati et al. (2014).

3.6. RIRE

The RIRE database (West et al., 1997) includes 18 patients with up to five different imaging modalities of the brain (Fig. 8). All 18 patients had at least three

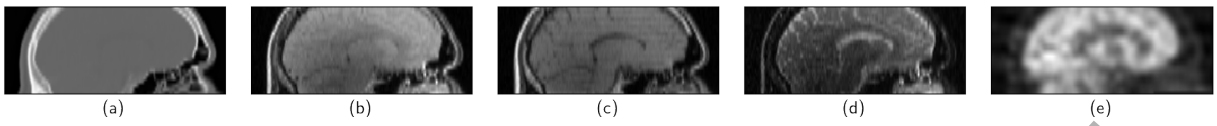


Figure 8: (a) CT image, (b) MR-PD image, (c) MR-T1 image, (d) Mr-T2 image and (e) PET image used in the RIRE experiment.

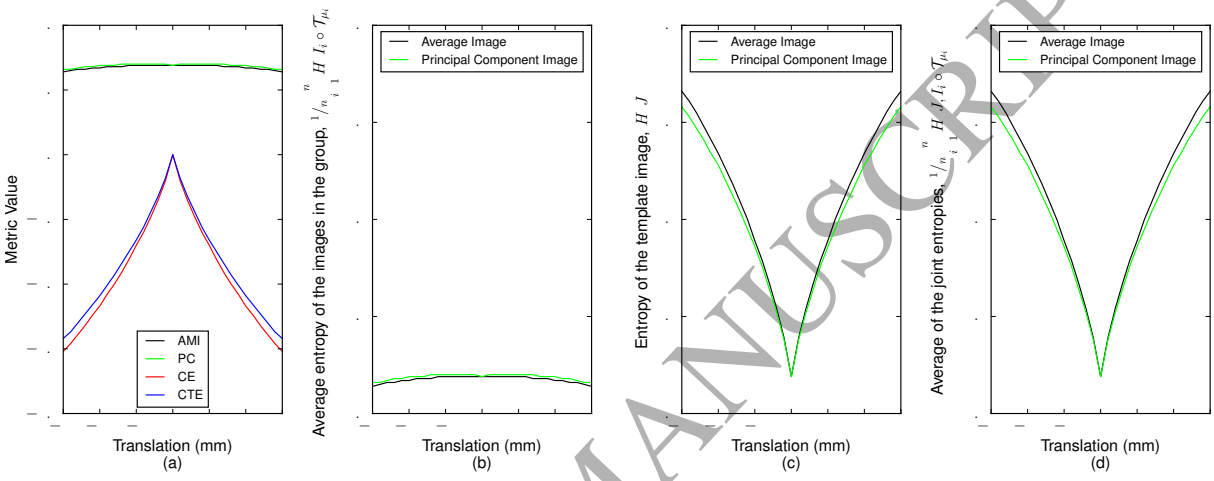


Figure 9: Results for the Black&White experiment where 11 black-and-white images were progressively and simultaneously translated. (a) The metric values. (b) The average of the entropies of the images in the group. (c) The entropy of the template image. (d) The average of the joint entropies.

of the following modalities available: CT, PET, MR-T1, MR-T2, MR-PD. Fiducial markers and a stereotactic frame were used to determine the ground truth transformations for CT to MR and PET to MR. Four to ten landmarks were available for each patient as a ground truth for the registrations and their target registration error was computed through the webform of the RIRE project, where rigid displacements between acquisitions were assumed.

To increase the robustness of the optimization, a two-step approach is used. First, a translation is optimized and used as an initialization for a second full rigid transformation with three translational and three rotational degrees of freedom. The registration was performed with five and two resolution levels, respectively. Similar to the Head&Neck dataset, preprocessing was performed by resampling the images in the group to the smallest voxel spacing.

The registration hyperparameters for the different experiments are summarized in Table 2.

4. Results

4.1. Synthetic Data

The behavior of the metric value and its separate components in the Black&White experiment are shown in Fig. 9 as a function of the translation. The Black&White experiment shows that the metric behavior of $\mathcal{S}_{AMI}$ and $\mathcal{S}_{PC}$ is equal to the behavior of the entropy of the images in the group. The contribution of the entropy of the template image completely cancels out the contribution of the joint entropy in $\mathcal{S}_{AMI}$ and $\mathcal{S}_{PC}$ as can be seen in Fig. 9(c-d). The resulting optimization is only driven by the complexity of the images in the group and not by their shared relationship.

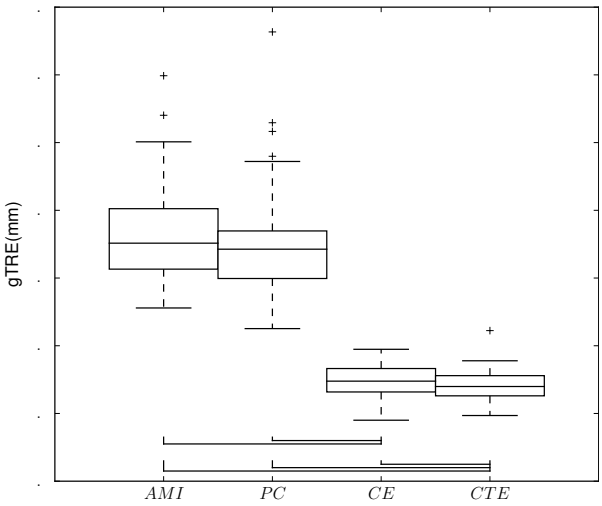


Figure 10: Boxplots for the results of the Multimodal Cubes experiment. Significant differences between two methods are indicated with black bars below the boxplots.

The results for the Multimodal Cubes experiment are shown in Fig. 10. When comparing the similarity metrics, $\mathcal{S}_{CTE}$ (1.71±0.11 mm) significantly outperformed all other entropy-based groupwise metrics (2.80±0.32 mm, 2.73±0.34 mm and 1.74±0.11 for $\mathcal{S}_{AMI}$, $\mathcal{S}_{PC}$ and $\mathcal{S}_{CE}$ respectively).

4.2. Clinical Data

Results for the gTRE in experiments on clinical data are visualized with boxplots in Fig. 11&12.

For the experiments on the Thoracic 4D CT and Carotid MR datasets (Fig. 11), no statistically significant differences were observed in terms of gTRE for the investigated information-based metrics.

In the Head&Neck experiment (Fig. 12) the best results are achieved by $\mathcal{S}_{CTE}$ with a gTRE of 2.74 ± 1.17mm performing significantly better compared to $\mathcal{S}_{AMI}$, $\mathcal{S}_{PC}$ and $\mathcal{D}_{PCA2}$.

Pairwise $\mathcal{S}_{MI}$ performed best in the RIRE experiment (Fig. 12) with a gTRE of 2.29±0.72mm ($\mathcal{S}_{CTE}$, 2.33 ± 0.57mm), but no significant differences were found compared to the other entropy-based metrics. $\mathcal{D}_{PCA2}$ performs worst, with the differences being statistically significant. A group of images was found to be misregistered following Tomaževič et al. (2012) when the gTRE is larger than the largest voxel spacing in the images. No misregistrations were obtained for $\mathcal{S}_{CTE}$, $\mathcal{S}_{CE}$ and $\mathcal{S}_{MI}$ whereas $\mathcal{S}_{AMI}$ and $\mathcal{S}_{PC}$ misregistered two patients and $\mathcal{D}_{PCA2}$ misregistered 14 patients.

In all four experiments on clinical data, pairwise MI performed worst in terms of transitivity, whereas the transitivity error for groupwise metrics reduced to (close to) zero (Table 3).

In Table 4, the values are given for the average runtime of the experiments performed in this work. The use of the conditional entropy does not induce an extra computational burden, whereas the use of the principal component images does. This discrepancy originates from an additional loop over the sampled coordinates, needed to perform the PCA and determine the weights of the eigenvector. Note that for more complex registrations with a regularizer, the additional computation time is relatively small compared to the total cost.

5. Discussion

Results on the Thoracic 4D-CT and Carotid MR dataset showed equivalent performance of the proposed methodology compared to other state-of-the-art methods in terms of registration accuracy.

The results for the Multimodal Cubes, Head&Neck and RIRE results were consistent. In all three datasets the accuracy improved for the proposed formulation compared to $\mathcal{S}_{AMI}$, and the improvement was found to be statistically significant in the former two experiments. Throughout these experiments the behavior of the auxiliary metrics $\mathcal{S}_{CE}$ and $\mathcal{S}_{PC}$ was also consistent. Using the conditional entropy instead of mutual information led to a large improvement, while using the principal component image improved the accuracy modestly. The combination of both contributions led to the best results in all three experiments compared to other groupwise metrics. As expected, the PCA2 metric performed poorly in multimodal registrations where a quantitative model or low-dimensional subspace is not available.

In all experiments based on clinical data, the transitivity of the resulting transformations was com-

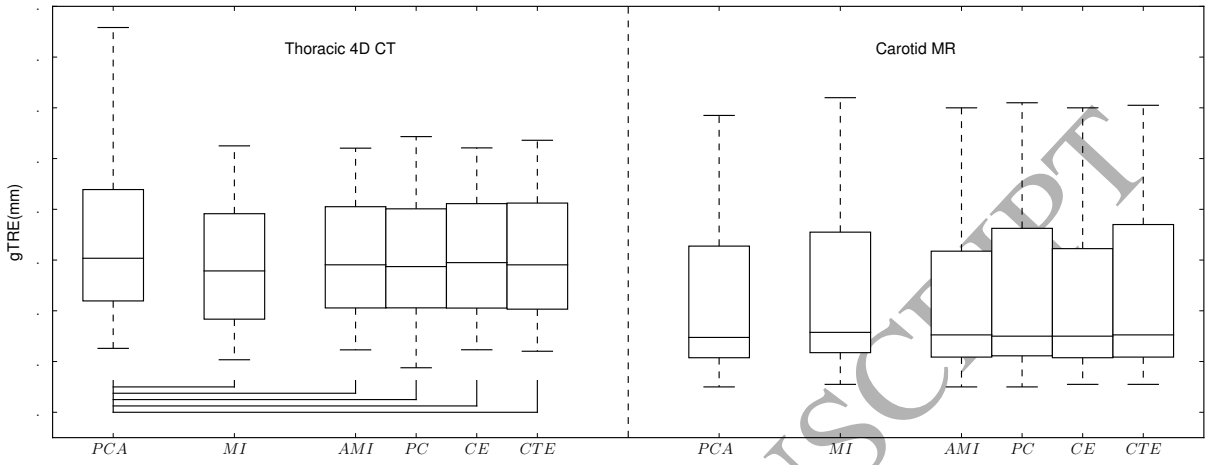


Figure 11: Boxplots for the results of the Thoracic 4DCT and Carotid MR experiment. Significant differences between two methods are indicated with black bars below the boxplots.

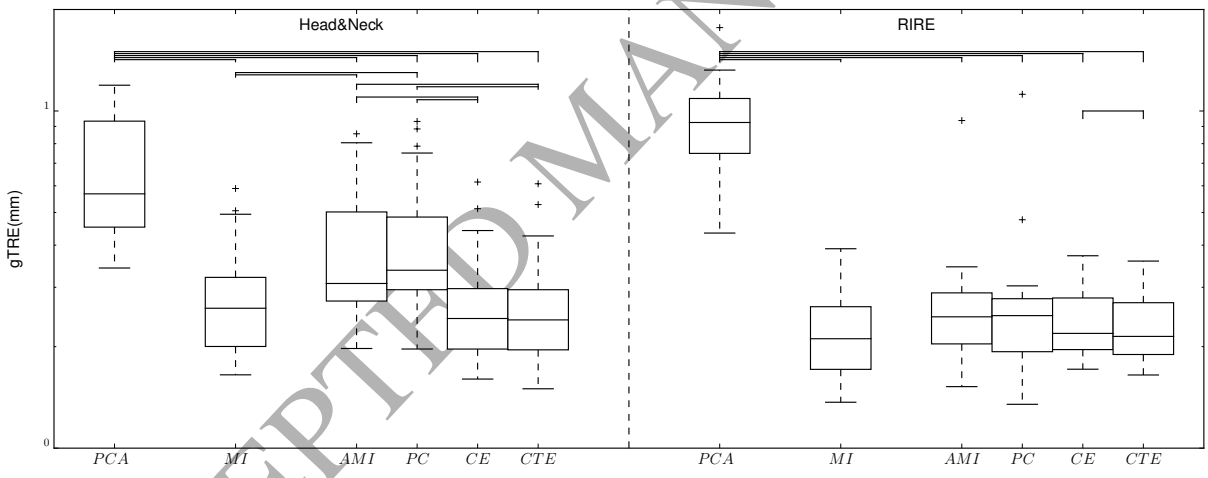


Figure 12: Boxplots for the results of the Head&Neck and RIRE experiment. Significant differences between two methods are indicated with black bars above the boxplots. Note the logarithmic scale on the y-axis.

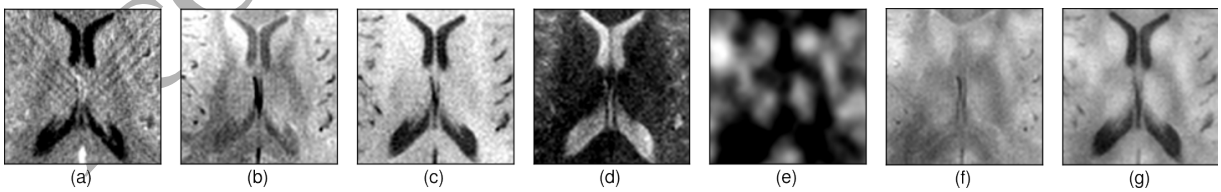


Figure 13: (a) CT image, (b) MR-PD image, (c) MR-T1 image, (d) MR-T2 image, (e) PET image, (f) average image and (g) principal component image when only the subregion of the ventricles is sampled for the RIRE experiment.

Table 3: Average transitivity errors for the clinical datasets. For the groupwise approaches, the maximum average transitivity error among all groupwise methods is reported. The values are given in mm.

	Thoracic 4D CT	Carotid MR	Head&Neck	RIRE
MI	5.65×10^{-1}	2.68×10^{-1}	2.14	1.47
Groupwise approaches	$< 3.39 \times 10^{-2}$	$< 7.66 \times 10^{-3}$	$< 1.85 \times 10^{-2}$	0

Table 4: Average runtime for the registrations in the different experiments. The values are given in minutes.

	Multimodal Cubes	Thoracic 4D CT	Carotid MR	Head&Neck	RIRE
PCA2	-	212	28	20	4
AMI	22	238	31	23	7
CE	22	252	31	23	7
PC	26	248	36	36	54
CTE	26	276	36	36	55

pared to $\mathcal{S}_{MI}$ for groupwise approaches. These results emphasize the added value of the implicit reference space in multimodal groupwise registration. Whereas a pairwise approach has to perform two separate registrations with different reference images to obtain a concatenated transformation, in a groupwise approach all transformations are evaluated simultaneously and with a substantially lower transitivity error. These results are consistent with previous findings in monomodal data (Geng et al., 2009; Metz et al., 2011).

In summary, for experiments based on images where no or modest changes in intensity distributions are present ('Thoracic 4D-CT' and 'Carotid MR'), CTE showed comparable performance to previously proposed groupwise methods and pairwise MI. In experiments with strongly varying intensity distributions ('Multimodal Cubes', 'Head&Neck' and 'RIRE'), CTE showed superior performance to previously proposed groupwise methods and performed on par to pairwise MI, with little to no transitivity error.

Fig. 5(f) and 6(f) highlight the differences in the average and principal component images. Herein, the absolute difference image between the average and principal component image is given in the 'Thoracic 4D CT' and 'Carotid MR' dataset, respectively, for a single patient. Herein, the largest differences occur in regions where the motion is greatest near moving structures or edges. This is consistent with previous work, where the principal component image was used to separate motion present in the images (Feng et al., 2016; Hamy et al., 2014; Melbourne et al., 2007). For multimodal registrations, the benefit of PCA over averaging can be seen by considering cases in which images with an inverted intensity profile are merged into the template image, as shown in Fig. 4(g-h) and Fig. 13. For the 'Multimodal Cubes' experiment, PCA lead to an increase of the contrast-to-noise ratio from 7.4 to 32.5 compared to simple averaging. Fig. 13 shows the average and principal component image when applied to the ventricles for an arbitrary patient in the RIRE dataset. With the T2 modality having an inverted intensity profile, the principal component image is able to retain the contrast in the template image. In the average image the intensities cancel out and the ventricles are poorly visible.

Two limitations should be stated with respect to current work. Firstly, only intrasubject data has been employed. Intersubject data is characterized by greater variability of intensity profiles and morphology, and has been reported to considerably increase

the complexity of groupwise registration (Hamm et al., 2009; Tang et al., 2009). It remains to be verified how CTE would perform when confronted with such data.

Secondly, in this work a methodology was used where the images are deformed and compared to the template image in the implicit reference system. However, previous work has shown that deforming the template image to the images in the group suits a generative model better (Allassonnière et al., 2007; Ma et al., 2008). In methodologies where the template is deformed to the images in the group, no need exists to constrain the transformations to the average deformation space (Eq. 16). This was shown to be advantageous, as such constraints could exclude some legitimate results (Aganj et al., 2017). We expect the proposed metric to perform equally well in such frameworks as it is independent of the transformations that were used.

6. Conclusion

In this work we proposed a novel similarity metric for intrasubject multimodal groupwise registration, the conditional template entropy. The proposed metric was evaluated in experiments based on synthetic and clinical intrasubject data and showed equivalent or improved registration accuracy compared to other state-of-the-art (dis)similarity metrics and improved transformation consistency compared to pairwise mutual information. These improvements were achieved mainly by the use of the conditional entropy, whereas the use of the principal component image contributed modestly in our experiments.

Acknowledgment

The research of S. Klein, W.J. Niessen and W. Huizinga is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 601055, VPH-DARE@IT. Additionally, the authors would like to thank Dr. B.F. Coolen and Dr. A.J. Nederveen for providing the 'Carotid MR' data. The authors would also like to thank the anonymous reviewers for their feedback and remarks on this work, which substantially improved its quality.

Appendix A. Derivative of principal component image

We determined the derivative of the principal component image with respect to the transformation parameters. The principal component image is given by Eq. (10) and repeated here

$$I_{\boldsymbol{\mu}}^{PCA}(\boldsymbol{x}) = \sum_{i=1}^{n} v_{i,\boldsymbol{\mu}} I_i \circ \mathcal{T}_{\boldsymbol{\mu}_i}(\boldsymbol{x}) = \boldsymbol{v}_{\boldsymbol{\mu}}^T \boldsymbol{I}(\boldsymbol{x}) \quad . \tag{A.1}$$

Herein, $\boldsymbol{I}(\boldsymbol{x})$ is the column vector representing all image intensities across the group for a given sampled coordinate. The derivative becomes

$$\frac{\partial I_{\boldsymbol{\mu}}^{PCA}(\boldsymbol{x})}{\partial \boldsymbol{\mu}} = \frac{\partial \boldsymbol{v}_{\boldsymbol{\mu}}^T}{\partial \boldsymbol{\mu}} \boldsymbol{I}(\boldsymbol{x}) + \boldsymbol{v}_{\boldsymbol{\mu}}^T \frac{\partial \boldsymbol{I}(\boldsymbol{x})}{\partial \boldsymbol{\mu}} \quad , \tag{A.2}$$

Following de Leeuw (2007) for the derivative of an eigenvector:

$$\frac{\partial \boldsymbol{v}_{\boldsymbol{\mu}}}{\partial \boldsymbol{\mu}} = -\left(C - eI\right)^{+} \frac{\partial C}{\partial \boldsymbol{\mu}} v_{i,\boldsymbol{\mu}} \quad , \tag{A.3}$$

with C the correlation matrix of the intensities, similar to Huizinga et al. (2016), I the identity matrix, e the eigenvalue associated with $\boldsymbol{v}_{\boldsymbol{\mu}}$ and $^+$ the notation for the Moore-Penrose inverse (de Leeuw, 2007). The derivative of the correlation matrix is given as

$$\begin{aligned}\frac{\partial C}{\partial \boldsymbol{\mu}} =& \frac{1}{|S|-1}\left(\frac{\partial \Sigma^{-1}}{\partial \boldsymbol{\mu}}\left(M-\overline{M}\right)^T\left(M-\overline{M}\right)\Sigma^{-1}\right.\\ &+\Sigma^{-1}\frac{\partial M^T}{\partial \boldsymbol{\mu}}\left(M-\overline{M}\right)\Sigma^{-1}\\ &+\Sigma^{-1}\left(M-\overline{M}\right)^T\frac{\partial M}{\partial \boldsymbol{\mu}}\Sigma^{-1}\\ &\left.+\Sigma^{-1}\left(M-\overline{M}\right)^T\left(M-\overline{M}\right)\frac{\partial \Sigma^{-1}}{\partial \boldsymbol{\mu}}\right)\end{aligned} \tag{A.4}$$

Herein, M refers to the data matrix with the intensities of the images, $\overline{M}$ is the matrix with the average

image intensity repeated along its columns, Σ is the diagonal matrix with the standard deviations of the images intensities as its diagonal elements. All notations correspond to those found in Huizinga et al. (2016) and we have ignored the derivative of the average image intensities likewise.

References

Aganj, I., Iglesias, J.E., Reuter, M., Sabuncu, M.R., Fischl, B., 2017. Mid-space-independent deformable image registration. NeuroImage 152, 158–170.

Allassonnière, S., Amit, Y., Trouvé, A., 2007. Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69, 3–29.

Bhatia, K.K., Hajnal, J., Hammers, A., Rueckert, D., 2007. Similarity metrics for groupwise non-rigid registration, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007. Springer, pp. 544–552. doi:10.1007/978-3-540-75759-7_66.

Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D., 2004. Consistent groupwise non-rigid registration for atlas construction, in: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, IEEE. pp. 908–911. doi:10.1109/isbi.2004.1398686.

Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T., 2009. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. Physics in medicine and biology 54, 1849. doi:10.1002/mrm.25634.

Ceranka, J., Polfliet, M., Lecouvet, F., Michoux, N., de Mey, J., Vandemeulebroucke, J., 2017. Registration strategies for multi-modal whole-body MRI mosaicing. Magnetic Resonance in Medicine .

Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H., 2006. Introduction to the non-rigid image registration evaluation project (NIREP), in: International Workshop on Biomedical Image Registration, Springer. pp. 128–135. doi:10.1007/11784012_16.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory, in: Information processing in medical imaging, pp. 263–274.

Coolen, B.F., Poot, D.H., Liem, M.I., Smits, L.P., Gao, S., Kotek, G., Klein, S., Nederveen, A.J., 2015. Three-dimensional quantitative T1 and T2 mapping of the carotid artery: Sequence design and in vivo feasibility. Magnetic Resonance in Medicine doi:10.1002/mrm.25634.

Feng, Q., Zhou, Y., Li, X., Mei, Y., Lu, Z., Zhang, Y., Feng, Y., Liu, Y., Yang, W., Chen, W., 2016. Liver DCE-MRI registration in manifold space based on robust principal component analysis. Scientific Reports 6.

Fletcher, P.T., Venkatasubramanian, S., Joshi, S., 2009. The geometric median on riemannian manifolds with application to robust atlas estimation. NeuroImage 45, S143–S152. doi:10.1016/j.neuroimage.2008.10.052.

Fortunati, V., Verhaart, R.F., Angeloni, F., van der Lugt, A., Niessen, W.J., Veenland, J.F., Paulides, M.M., van Walsum, T., 2014. Feasibility of multimodal deformable registration for head and neck tumor treatment planning. International Journal of Radiation Oncology* Biology* Physics 90, 85–93. doi:10.1016/j.ijrobp.2014.05.027.

Fortunati, V., Verhaart, R.F., Verduijn, G.M., van der Lugt, A., Angeloni, F., Niessen, W.J., Veenland, J.F., Paulides, M.M., van Walsum, T., 2015. MRI integration into treatment planning of head and neck tumors: Can patient immobilization be avoided? Radiotherapy and Oncology 115, 191–194. doi:10.1016/j.radonc.2015.03.021.

Geng, X., Christensen, G.E., Gu, H., Ross, T.J., Yang, Y., 2009. Implicit reference-based groupwise image registration and its application to structural and functional MRI. Neuroimage 47, 1341–1351. doi:10.1016/j.neuroimage.2009.04.024.

Hallack, A., Chappell, M.A., Gooding, M.J., Schnabel, J.A., 2014. A new similarity metric for groupwise registration of variable flip angle sequences for improved T 10 estimation in DCE-MRI, in: Biomedical Image Registration. Springer, pp. 154–163. doi:10.1007/978-3-319-08554-8_16.

Hamm, J., Davatzikos, C., Verma, R., 2009. Efficient large deformation registration via geodesics on a learned manifold of images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 680–687.

Hamy, V., Dikaios, N., Punwani, S., Melbourne, A., Latifoltojar, A., Makanyanga, J., Chouhan, M., Helbren, E., Menys, A., Taylor, S., et al., 2014. Respiratory motion correction in dynamic MRI using robust data decomposition registration–application to DCE-MRI. Medical image analysis 18, 301–313.

Hero, A., Ma, B., Michel, O.J., Gorman, J., 2002. Applications of entropic spanning graphs. Signal Processing Magazine, IEEE 19, 85–95. doi:10.1109/msp.2002.1028355.

Huizinga, W., Poot, D., Guyader, J.M., Klaassen, R., Coolen, B., van Kranenburg, M., van Geuns, R., Uitterdijk, A., Polfliet, M., Vandemeulebroucke, J., Leemans, A., Niessen, W., Klein, S., 2016. PCA-based groupwise image registration for quantitative MRI. Medical Image Analysis 29, 65–78. doi:10.1016/j.media.2015.12.004.

Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, S151–S160. doi:10.1016/j.neuroimage.2004.07.068.

Kim, M., Wu, G., Shen, D., 2011. Groupwise registration of breast DCE-MR images for accurate tumor measurement, in: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, IEEE. pp. 598–601.

Klein, S., Pluim, J.P., Staring, M., Viergever, M.A., 2009. Adaptive stochastic gradient descent optimisation for image registration. International journal of computer vision 81, 227–239. doi:10.1007/s11263-008-0168-y.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. Elastix: a toolbox for intensity-based medical image registration. Medical Imaging, IEEE Transactions on 29, 196–205. doi:10.1109/tmi.2009.2035616.

van 't Klooster, R., Staring, M., Klein, S., Kwee, R.M., Kooi, M.E., Reiber, J.H., Lelieveldt, B.P., van der Geest, R.J., 2013. Automated registration of multispectral mr vessel wall images of the carotid artery. Medical physics 40.

de Leeuw, J., 2007. Derivatives of generalized eigen systems with applications. Department of Statistics, UCLA .

Luenberger, D.G., 1973. Introduction to linear and nonlinear programming. volume 28. Addison-Wesley Reading, MA.

Ma, J., Miller, M.I., Trouvé, A., Younes, L., 2008. Bayesian template estimation in computational anatomy. NeuroImage 42, 252–261.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. Medical Imaging, IEEE Transactions on 16, 187–198. doi:10.1109/42.563664.

Melbourne, A., Atkinson, D., White, M., Collins, D., Leach, M., Hawkes, D., 2007. Registration of dynamic contrast-enhanced MRI using a progressive principal component registration (PPCR). Physics in medicine and biology 52, 5147.

Metz, C., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J., 2011. Nonrigid registration of dynamic medical imaging data using n D+

t B-splines and a groupwise optimization approach. Medical image analysis 15, 238–249. doi:10.1016/j.media.2010.10.003.

Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al., 2011. Evaluation of registration methods on thoracic CT: the EMPIRE10 challenge. IEEE transactions on medical imaging 30, 1901–1920.

Orchard, J., Mann, R., 2010. Registering a multisensor ensemble of images. Image Processing, IEEE Transactions on 19, 1236–1247. doi:10.1109/tip.2009.2039371.

Polfliet, M., Klein, S., Huizinga, W., de Mey, J., Vandemeulebroucke, J., 2016. The pythagorean averages as group images in efficient groupwise registration, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1261–1264. doi:10.1109/isbi.2016.7493496.

Polfliet, M., Klein, S., Niessen, W.J., Vandemeulebroucke, J., 2017. Laplacian eigenmaps for multimodal groupwise image registration, in: SPIE Medical Imaging, International Society for Optics and Photonics. pp. 101331N–101331N.

Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. Neuroimage 53, 1181–1196.

Saito, T., Sakamoto, T., Oya, N., 2009. Comparison of gating around end-expiration and end-inspiration in radiotherapy for lung cancer. Radiotherapy and oncology 93, 430–435.

Sarrut, D., Boldea, V., Miguet, S., Ginestet, C., 2006. Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. Medical physics 33, 605–617. doi:10.1118/1.2161409.

Spiclin, Z., Likar, B., Pernus, F., 2012. Groupwise registration of multimodal images by an efficient joint entropy minimization scheme. Image Processing, IEEE Transactions on 21, 2546–2558. doi:10.1109/tip.2012.2186145.

Studholme, C., Hill, D.L., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Pattern recognition 32, 71–86. doi:10.1016/s0031-3203(98)00091-0.

Tang, S., Fan, Y., Wu, G., Kim, M., Shen, D., 2009. RABBIT: rapid alignment of brains by building intermediate templates. NeuroImage 47, 1277–1287.

Thévenaz, P., Unser, M., 2000. Optimization of mutual information for multiresolution image registration. Image Processing, IEEE Transactions on 9, 2083–2099. doi:10.1109/83.887976.

Tomaževič, D., Likar, B., Pernuš, F., 2012. Multi-feature mutual information image registration. Image Analysis & Stereology 31, 43–53. doi:10.5566/ias.v31.p43-53.

Vandemeulebroucke, J., Bernard, O., Rit, S., Kybic, J., Clarysse, P., Sarrut, D., 2012. Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. Medical physics 39, 1006–1015. doi:10.1118/1.3679009.

Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D., 2011. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. Medical physics 38, 166–178. doi:10.1118/1.3523619.

Verhaart, R.F., Fortunati, V., Verduijn, G.M., van der Lugt, A., van Walsum, T., Veenland, J.F., Paulides, M.M., 2014. The relevance of MRI for patient modeling in head and neck hyperthermia treatment planning: A comparison of CT and CT-MRI based tissue segmentation on simulated temperature. Medical physics 41, 123302. doi:10.1118/1.4901270.

Vijayan, S., Klein, S., Hofstad, E.F., Lindseth, F., Ystgaard, B., Langø, T., 2014. Motion tracking in the liver: Validation of a method based on 4D ultrasound using a nonrigid registration technique. Medical physics 41. doi:10.1118/1.4890091.

Viola, P., Wells III, W., 1995. Alignment by maximization of mutual information, in: Proceedings

of the Fifth International Conference on Computer Vision, IEEE Computer Society. p. 16.

Wachinger, C., Navab, N., 2013. Simultaneous registration of multiple images: Similarity metrics and efficient optimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, 1221–1233. doi:10.1109/tpami.2012.196.

Wachinger, C., Wein, W., Navab, N., 2007. Three-dimensional ultrasound mosaicing, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007. Springer, pp. 327–335. doi:10.1007/978-3-540-75759-7_40.

Wells, J.A., O'Callaghan, J., Holmes, H., Powell, N.M., Johnson, R., Siow, B., Torrealdea, F., Ismail, O., Walker-Samuel, S., Golay, X., et al., 2015. In vivo imaging of tau pathology using multi-parametric quantitative MRI. Neuroimage 111, 369–378. doi:10.1016/j.neuroimage.2015.02.023.

Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. Medical image analysis 1, 35–51. doi:10.1016/s1361-8415(96)80004-1.

West, J., Fitzpatrick, J.M., Wang, M.Y., Dawant, B.M., Maurer Jr, C.R., Kessler, R.M., Maciunas, R.J., Barillot, C., Lemoine, D., Collignon, A., et al., 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. Journal of computer assisted tomography 21, 554–568. doi:10.1097/00004728-199707000-00007.

Wu, G., Jia, H., Wang, Q., Shen, D., 2011. Sharpmean: Groupwise registration guided by sharp mean image and tree-based registration. NeuroImage 56, 1968–1981.

Zöllei, L., Learned-Miller, E., Grimson, E., Wells, W., 2005. Efficient population registration of 3D data, in: Computer Vision for Biomedical Image Applications. Springer, pp. 291–301. doi:10.1007/11569541_30.