

Pythagorean Mean Images for Efficient Groupwise Registration

Mathias Polfiet^{1,2}, Wyke Huizinga³, Stefan Klein³, Johan de Mey⁴, and Jef Vandemeulebroucke^{1,2}

¹ Vrije Universiteit Brussel, Dept. of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium

² iMinds, Dept. of Medical IT, Gaston Crommenlaan 8 (box 102), B-9050 Ghent, Belgium

³ Biomedical Imaging Group Rotterdam, Depts. of Radiology & Medical Informatics, Erasmus MC, Rotterdam, the Netherlands

⁴ Department of Radiology, Universitair Ziekenhuis Brussel (UZ Brussel), Laarbeeklaan 101, 1090 Brussels, Belgium

Abstract. Groupwise registration is a powerful technique allowing to simultaneously align multiple images using an unbiased approach. Its need arises from population studies or motion estimation across dynamic sequences. An often used class of efficient groupwise metrics measures similarity as the sum of the pairwise similarities between the images and a template image, commonly chosen to be the arithmetic mean image in the current iteration. However, arithmetic averaging in intensity space limits the applications to closely related modalities, and may produce fuzzy images compromising the performance of the metric. Geometric and harmonic averaging is capable of handling range and scale differences without adding computational complexity. Groupwise similarity metrics based on mutual information and the three Pythagorean means were investigated. Experiments performed on monomodal and multimodal data demonstrated superior performance of geometric and harmonic over arithmetic averaging and the corresponding pairwise registration.

1 Introduction

Groupwise registration has shown to be of interest in several areas of research, such as the unbiased construction of atlases [1] and motion estimation across temporal sequences for radiotherapy planning [2]. Scalability of the approach is an important property for groupwise registration, given the fact that such acquisitions can contain over a hundred images to be registered.

The direct application of entropy-based metrics using joint probability density functions (PDFs) such as mutual information [3] (MI), results in an exponentially increasing probability space. Such metrics suffer from the *curse of dimensionality* with the sparsity in the joint PDF limiting their applicability. Several methods have been proposed to work around this problem. Spiclin *et*

al. [4] proposed a method based on hierarchically subdividing the joint intensity space similar to a tree code, and reported good results for up to ten images, after which sparsity limits the robustness of the method.

Instead of focusing on estimating the probability density functions, others have focused on its parent measure, entropy. Hero *et al.* [5] proposed a method based on entropic graphs, where each node in the graph represents an intensity pair in the joint intensity space and the minimal length needed to span the graph is related to the entropy of the system. The main drawback of this method stems from the lengthy optimization required to find this minimal length.

An interesting class of metrics is based on constructing a template image and measuring similarity as the sum of the pairwise similarities between the images and the template [6, 7]. This approach yields an algorithm with linearly increasing computational complexity with respect to the number of images to be aligned, making it suitable for large groupwise registration problems. As a template image, the arithmetic mean image in the current iteration is commonly used, though several authors have demonstrated the importance of the use of a sharp and more representative template to improve robustness and accuracy of the registration [8, 9].

Indeed, arithmetic averaging over the intensity space tends to lead to fuzzy mean images. In addition, it is less suited in the presence of scale and range differences, often present in multimodal data. In this work we propose two novel multimodal metrics for groupwise registration based on internally computing the geometric and harmonic mean images. The performance of these metrics is evaluated for mono- and multimodal groupwise registration, and compared to results when using the arithmetic mean image.

2 Methods and Materials

2.1 Registration Metrics

Following Bhatia *et al.* [7], pairwise mutual information S_{MI} [3] can be extended to arithmetic average mutual information⁵ (AAMI)

$$S_{AAMI}(I_1(\mathbf{x}), \dots, I_n(\mathbf{x})) = \sum_{i=1}^n S_{MI}(I_i(\mathcal{T}_i(\mathbf{x})), \bar{I}_A(\mathbf{x})) \quad (1)$$

where \mathbf{x} is the spatial coordinate, \mathcal{T}_i the spatial transformation and I_i the intensity function associated with the i^{th} image, for which we assumed an interpolation scheme. $\bar{I}_A(\mathbf{x})$ is the voxel-wise arithmetic mean intensity image defined as

$$\bar{I}_A(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I_i(\mathcal{T}_i(\mathbf{x})) \quad (2)$$

⁵ Note that we preferred the use of MI instead of normalised MI, as proposed in [7].

The arithmetic mean tends to lead to fuzzy mean images. In addition, it is not suited for handling differences in intensity ranges and scales as found in multimodal data. We propose the use of the geometric and harmonic mean images, given by

$$\bar{I}_G(\mathbf{x}) = \sqrt[n]{\prod_{i=1}^n I'_i(\mathcal{T}_i(\mathbf{x}))} , \quad (3)$$

$$\bar{I}_H(\mathbf{x}) = \frac{n}{\sum_{i=1}^n \frac{1}{I'_i(\mathcal{T}_i(\mathbf{x}))}} = \frac{n \prod_{i=1}^n I'_i(\mathcal{T}_i(\mathbf{x}))}{\sum_{i=1}^n \prod_{j=1, j \neq i}^n I'_j(\mathcal{T}_j(\mathbf{x}))} . \quad (4)$$

In equations (3) and (4), I'_i is the modified intensity function to ensure non-negative intensities. The three Pythagorean means become equal if and only if, the samples over which are averaged are identical. In all other cases, the arithmetic mean will be higher than the geometric mean, which in turn will be higher than the harmonic mean. As such, intensities tend to fall-off faster at edges, limiting the blurred region, while overlapping structures are retained. We hypothesize that this higher specificity could lead to better registration accuracy.

Two novel groupwise similarity metrics, using MI as a submetric and based on the geometric average image (GAMI) and harmonic average image (HAMI) were implemented as an extension to the software package `elastix` [10] together with AAMI and will be made available in the future. Partial derivatives with respect to the transformation parameters were computed analytically and determined following the approach of Thévenaz *et al.* [11].

To illustrate the behavior of these three metrics, a simulation was developed in which nine squares were simultaneously rotated with a different speed and one was kept stationary. The rotation speed of each square was a multiple of the rotation speed of the first square, such that after a rotation of 90° for the first square, the initial situation is recovered where all squares overlap and another global minimum in the metric space is found. Two simulations were performed, mimicking monomodal and multimodal data. For the monomodal experiment, we used squares with identical intensity. A single image consisted of an outer square, with an intensity of 0.5, an inner square, with an intensity of 1.0, and the background, with an intensity of 0.0. For the multimodal experiment, we set the intensities of the inner and outer squares randomly between 0.5 and 1.0 for all images separately. Figure 1 shows the metric values for a rotation from 0° to 90° for the slowest rotating image.

It can be seen that the proposed metrics have less local minima and those that are common between all metrics are far less pronounced compared to AAMI.

2.2 Monomodal Experiments and Validation

Monomodal 4DCT data was taken from the POPI and DIR-LAB databases [12, 13], consisting of, respectively, 6 and 10 CT images of the thorax for a total

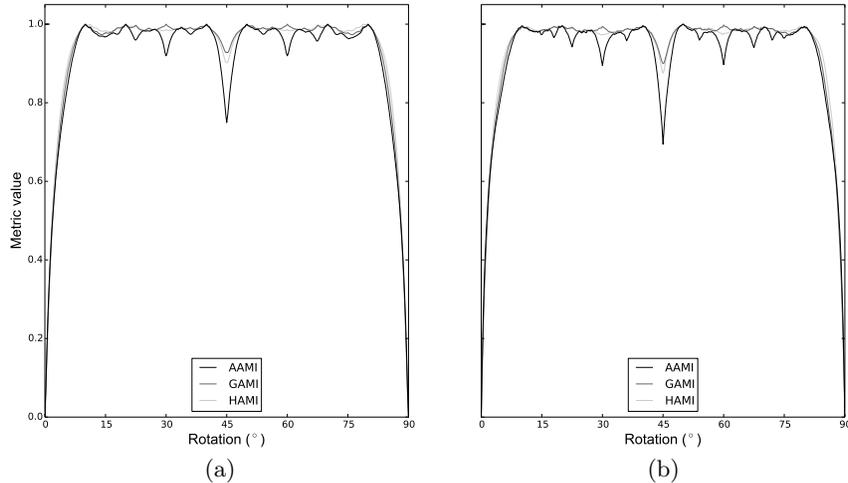


Fig. 1. Behaviour of the three negated similarity metrics under discussion, arithmetic average mutual information (AAMI), geometric average mutual information (GAMI) and harmonic average mutual information (HAMI), when 10 squares are rotated simultaneously in the monomodal case (a) and the multimodal case (b). A lower metric value corresponds to a better match. The angle of the slowest rotating square is given on the x-axis.

of 16 patients. For the POPI database, three patients had 100 manually identified landmarks in the lungs for every breathing phase while the remaining three patients had 100 landmarks in inspiration and expiration phases. The images in the DIR-LAB dataset had 300 landmarks in the lungs for the inspiration and expiration phases and 75 in the four phases in between. Registration accuracy was determined with respect to the inspiration landmarks for all phases in which landmarks were available. We define the groupwise target registration error (gTRE) over a group of images for a single patient as a measure for the accuracy of the registration:

$$gTRE(r) = \frac{1}{n} \sum_{i \neq r} \frac{1}{|P_i|} \sum_{\mathbf{p}_{i,j} \in P_i} \|\mathcal{T}_{i,r}(\mathbf{p}_{i,j}) - \mathbf{p}_{r,j}\|. \quad (5)$$

Herein, r is the reference time point, P_i the collection of landmarks in time point i , $\mathcal{T}_{i,r}$ the transformation that maps the coordinates from the i^{th} timepoint to the reference timepoint and $\mathbf{p}_{i,j}$ the j^{th} landmark from the i^{th} timepoint.

We performed a deformable registration using cubic B-splines and a final control point spacing of 12 mm. For the optimization four resolutions were used together with an adaptive stochastic gradient descent. Registrations were performed with AAMI as a groupwise metric and MI as a pairwise metric and the results were compared to those obtained from the proposed groupwise metrics GAMI and HAMI. All other registration parameters were kept constant. Pair-

wise registrations were only performed for the phases in which landmarks were available with respect to the inspiratory phase whilst groupwise registrations always included all breathing phases.

2.3 Multimodal Experiments and Validation

Multimodal brain images obtained from the RIRE database [14] were used for the multimodal experiments. Data from 18 patients was included for which at least three of the following modalities were available: CT, PET, MR-T1, MR-T2, MR-proton density (PD). Ground truth transformations were determined through the use of fiducial markers and a stereotactic frame for CT to MR and PET to MR. Four to ten landmarks were available for each patient as a ground truth for the registrations, allowing to compute the gTRE.

Given the large range differences present between CT and PET, it is trivial to see that AAMI will fail to obtain a good and unbiased template image. Prior to the registration, the images were normalized between 0 and 1 to eliminate this effect. The registrations were performed with three translational degrees of freedom prior to rigid registrations which had six degrees of freedom. This allowed for a more robust optimization. The performance of groupwise registration using AAMI, GAMI and HAMI, was compared to pairwise registration using MI. Once again all other registration parameters were kept equal.

3 Results and discussion

3.1 Monomodal Data

The results for the deformable registrations of the 4DCT images of the thorax are pooled for all 16 patients and summarized in Table 1. Significance testing was performed using a two-tailed Wilcoxon signed-rank test given the non-normality of the data. Pairwise registrations were outperformed by groupwise registrations using AAMI ($p = 0.11$), GAMI ($p = 0.016$) and HAMI ($p = 0.0019$), illustrating the added value of a groupwise approach. We obtained better accuracies using GAMI with respect to AAMI in all 16 registrations and the difference was significant ($p = 4.4 \times 10^{-4}$). The HAMI metric gave better results than GAMI in all 16 registrations and the difference was significant ($p = 4.4 \times 10^{-4}$).

The obtained accuracies for pairwise registration using MI corresponded well to other results reported for this data and when performing registration without masks. Delmon *et al.* [15] reported 3.82 ± 4.15 mm for a pairwise registration using MI. The improvement in results is most likely due to lower control point spacing used here (32 mm versus 12 mm) which allowed for finer deformations to be modeled.

The use of geometric and harmonic mean images in the groupwise registration framework has little to no computational overhead compared to the AAMI, and constitutes an elegant alternative to approaches which require separate optimization of the template image [8] and minimal spanning tree [9].

Table 1. Results for the registration of the 4DCT of the lungs for the four compared methods: mutual information (MI), arithmetic average MI (AAMI), geometric average MI (GAMI) and harmonic average MI (HAMI). The values, expressed in mm, correspond to the mean, standard deviation, median and maximal groupwise target registration error.

| Method | Mean gTRE | StDev gTRE. | Median gTRE | Max gTRE |
|--------|-----------|-------------|-------------|----------|
| MI | 2.15 | 1.03 | 1.83 | 4.80 |
| AAMI | 2.17 | 1.10 | 1.83 | 5.42 |
| GAMI | 2.12 | 1.06 | 1.79 | 5.22 |
| HAMI | 2.06 | 1.01 | 1.77 | 5.01 |

3.2 Multimodal Data

The results for the rigid registrations on the RIRE dataset are shown in Table 2. In accordance with Tomažević *et al.* [16], we considered the registration successful if the gTRE of a subject was less than the largest voxel spacing of the images under study (8 mm). This lead us to exclude a total of two subjects from the statistical analysis and Table 2 for all metrics to allow for a fair and honest comparison. Two misregistrations were obtained for AAMI (patient 008 and patient 105) and one for GAMI (patient 008).

Overall, groupwise registration using HAMI gave the best results. No misregistrations were recorded and accuracies were significantly better than AAMI ($p = 0.0013$, using a two-tailed Wilcoxon signed-rank test) whilst GAMI also outperformed AAMI significantly ($p = 0.0037$). GAMI and HAMI did not produce significantly different results compared to pairwise MI.

Results per modality pairing are illustrated in Figure 2. Given the limited data available for some modality pairings no significance testing was performed. It can be seen from the figure that a pairwise approach for the PET-MR registrations is better. It is possible that images with low signal-to-noise ratio, such as PET, disturb the groupwise approach by clouding the mean image and wiping away some of the details present therein.

Table 2. Results for the registration of the multimodal brain images for the four compared methods: mutual information (MI), arithmetic average MI (AAMI), geometric average MI (GAMI) and harmonic average MI (HAMI). The values, expressed in mm, correspond to the mean, standard deviation, median and maximal groupwise target registration error.

| Method | Mean gTRE | StDev gTRE. | Median gTRE | Max gTRE |
|--------|-----------|-------------|-------------|----------|
| MI | 2.31 | 0.80 | 2.05 | 4.50 |
| AAMI | 4.19 | 1.91 | 3.46 | 7.62 |
| GAMI | 2.60 | 1.00 | 2.45 | 5.43 |
| HAMI | 2.35 | 0.66 | 2.39 | 3.94 |

Our results for pairwise registration were worse compared to those of Thévenaz *et al.* [11] (reported as the median TRE per modality pair), even though a similar implementation is followed for MI. This discrepancy might be explained by differences in multiresolution strategy and optimization method, or other detailed settings such as the number of histogram bins, interpolation method, etc.

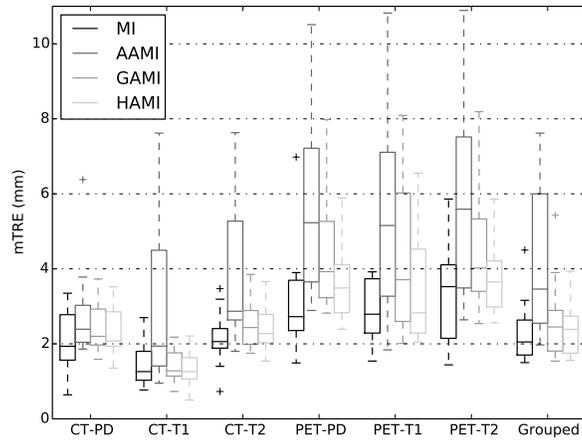


Fig. 2. Boxplots (using John Tukey’s definition) of the gTREs of the 16 registrations per modality pair and grouped over all modalities. The compared metrics were mutual information (MI), arithmetic average mutual information (AAMI), geometric average mutual information (GAMI) and harmonic average mutual information (HAMI).

4 Conclusion

In this work we presented an extension to the average image mutual information through the use of the different Pythagorean mean images. The proposed metrics handle scale and range differences better and have an improved optimization behavior without adding computational complexity.

The geometric and harmonic average mutual information proved to be superior to the commonly used arithmetic average mutual information and the difference was significant for mono- and multimodal experiments.

Acknowledgements

The research of S. Klein and W. Huizinga is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 601055, VPH-DARE@IT.

References

1. Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D.: Consistent groupwise non-rigid registration for atlas construction. In: Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, IEEE (2004) 908–911
2. Schreibmann, E., Thorndyke, B., Li, T., Wang, J., Xing, L.: Four-dimensional image registration for image-guided radiotherapy. *International Journal of Radiation Oncology* Biology* Physics* **71**(2) (2008) 578–586
3. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. In: *Information processing in medical imaging*. Volume 3. (1995) 263–274
4. Spiclin, Z., Likar, B., Pernus, F.: Groupwise registration of multimodal images by an efficient joint entropy minimization scheme. *Image Processing, IEEE Transactions on* **21**(5) (2012) 2546–2558
5. Hero, A., Ma, B., Michel, O.J., Gorman, J.: Applications of entropic spanning graphs. *Signal Processing Magazine, IEEE* **19**(5) (2002) 85–95
6. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* **23** (2004) S151–S160
7. Bhatia, K.K., Hajnal, J., Hammers, A., Rueckert, D.: Similarity metrics for groupwise non-rigid registration. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*. Springer (2007) 544–552
8. Fletcher, P.T., Venkatasubramanian, S., Joshi, S.: The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage* **45**(1) (2009) S143–S152
9. Wu, G., Jia, H., Wang, Q., Shen, D.: Sharpmean: Groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage* **56**(4) (2011) 1968–1981
10. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on* **29**(1) (2010) 196–205
11. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. *Image Processing, IEEE Transactions on* **9**(12) (2000) 2083–2099
12. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D.: Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Medical physics* **38**(1) (2011) 166–178
13. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in medicine and biology* **54**(7) (2009) 1849
14. 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.: Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of computer assisted tomography* **21**(4) (1997) 554–568
15. Delmon, V., Rit, S., Pinho, R., Sarrut, D., Delmon, V.: Direction dependent b-splines decomposition for the registration of sliding objects. In: *Proceedings of the Fourth International Workshop on Pulmonary Image Analysis*. (2011) 45–55
16. Tomaževič, D., Likar, B., Pernuš, F.: Multi-feature mutual information image registration. *Image Analysis & Stereology* **31**(1) (2012) 43–53