

# SCIENTIFIC REPORTS



OPEN

## Groupwise image registration based on a total correlation dissimilarity measure for quantitative MRI and dynamic imaging data

Jean-Marie Guyader<sup>1</sup> , Wyke Huizinga<sup>1</sup>, Dirk H. J. Poot<sup>1,2</sup>, Matthijs van Kranenburg<sup>3,4</sup>, André Uitterdijk<sup>4</sup>, Wiro J. Niessen<sup>1,2</sup> & Stefan Klein<sup>1</sup> 

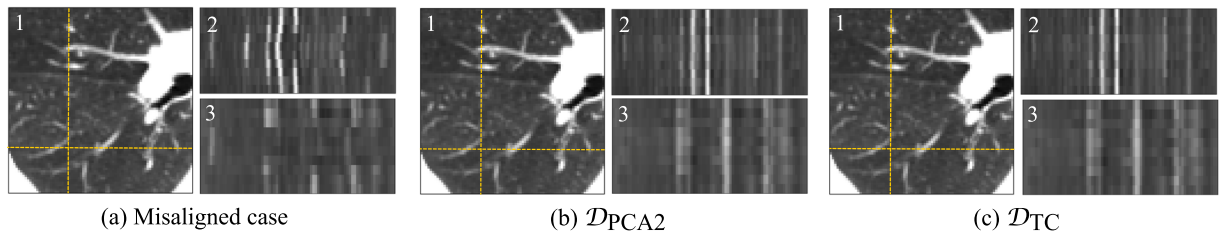
The most widespread technique used to register sets of medical images consists of selecting one image as fixed reference, to which all remaining images are successively registered. This pairwise scheme requires one optimization procedure per pair of images to register. Pairwise mutual information is a common dissimilarity measure applied to a large variety of datasets. Alternative methods, called groupwise registrations, have been presented to register two or more images in a single optimization procedure, without the need of a reference image. Given the success of mutual information in pairwise registration, we adapt one of its multivariate versions, called total correlation, in a groupwise context. We justify the choice of total correlation among other multivariate versions of mutual information, and provide full implementation details. The resulting total correlation measure is remarkably close to measures previously proposed by Huizinga *et al.* based on principal component analysis. Our experiments, performed on five quantitative imaging datasets and on a dynamic CT imaging dataset, show that total correlation yields registration results that are comparable to Huizinga's methods. Total correlation has the advantage of being theoretically justified, while the measures of Huizinga *et al.* were designed empirically. Additionally, total correlation offers an alternative to pairwise mutual information on quantitative imaging datasets.

Intensity-based image registration using the maximization of mutual information is commonly used for aligning pairs of medical images that do not have similar intensity distributions, or are acquired from different modalities<sup>1–3</sup>. Mutual information belongs to the family of pairwise dissimilarity measures. Pairwise methods quantify the alignment of a moving image with a fixed reference image. The optimization process performed in the context of pairwise registration therefore considers only two images simultaneously.

Nowadays, imaging datasets often contain more than two images, acquired from different modalities, different time points or different subjects, for instance. When more than two images have to be registered, the pairwise registration scheme is not always the most adapted. Firstly, the choice of reference image to which the remaining image are registered can be arbitrary, but may also influence the registration results, as shown by Geng *et al.*<sup>4</sup>. Secondly, pairwise registration does not allow the registration of all images in a single optimization procedure, which prevents taking into account all image information simultaneously.

Conversely, groupwise image registration methods are fully symmetric (i.e. all images play the same role in the registration procedure), and they consist of a single optimization procedure. Given the success of mutual information in the context of pairwise image registration, this paper specifically focuses on groupwise registration

<sup>1</sup>Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, The Netherlands. <sup>2</sup>Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands. <sup>3</sup>Departments of Radiology, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, The Netherlands. <sup>4</sup>Department of Cardiology, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, The Netherlands. Correspondence and requests for materials should be addressed to J.-M.G. (email: [jeanmarie.guyader@gmail.com](mailto:jeanmarie.guyader@gmail.com))



**Figure 1.** Registration results for a CT-LUNG dataset. The images denoted ‘2’ and ‘3’ stack the voxel information of  $G = 10$  images at the locations defined by the dotted lines drawn in the image denoted ‘1’ (vertical line: ‘2’; horizontal line: ‘3’).

techniques that are based on the concept of mutual information. Though the formulation of mutual information for two images is unique, several multivariate versions exist for its generalization for more than two images. We provide theory about the main multivariate dissimilarity measures based on mutual information, that could be used for the groupwise registration of medical images. These dissimilarity measures are called interaction information<sup>5</sup>, total correlation<sup>6</sup> and dual total correlation<sup>7</sup>. Total correlation is the groupwise dissimilarity measure that we propose to adapt in the context of groupwise image registration.

A preliminary version of our work was presented at a conference<sup>8</sup>. In the present article, we provide full theoretical developments, extensive implementation details, and additional experimental analyses.

Competing state-of-the-art dissimilarity measures for groupwise registration include the sum of variances developed by Metz *et al.*<sup>9</sup>, the groupwise mutual information method of Bhatia *et al.*<sup>10</sup>, and the groupwise dissimilarity measures based on principal component analysis (PCA) previously developed by Huizinga *et al.*<sup>11</sup>. The expression of the total correlation dissimilarity measure that we propose is remarkably close to Huizinga’s PCA-based groupwise dissimilarity measures, which were shown to outperform pairwise and groupwise state-of-the-art methods on qMRI datasets. The experiments conducted in this article consist of using groupwise total correlation for the registration of a dynamic CT imaging dataset, and of five quantitative magnetic resonance imaging (qMRI) image datasets. Registration results are compared to Huizinga’s methods, but also to pairwise registration based on mutual information.

## Results

Groupwise registration based on the total correlation dissimilarity measure  $\mathcal{D}_{TC}$  that we propose in this study is tested on six different types of image datasets, which overall represents 42 subjects. Dynamic series of CT images were acquired for the first type of image dataset, denoted CT-LUNG. The five other types of datasets, denoted T1MOLLI-HEART, T1VFA-CAROTID, ADC-ABDOMEN, DTI-BRAIN, and DCE-ABDOMEN, are qMRI datasets for which multiple MR images were acquired using different acquisition parameters (or at multiple time points after injection of a contrast agent). For these five qMRI datasets, we fitted a qMRI model to the image intensities at each spatial location, and extracted quantitative images: spin-lattice relaxation time ( $T_1$ ) images for T1MOLLI-HEART and T1VFA-CAROTID, apparent diffusion coefficient (ADC) images for ADC-ABDOMEN, mean diffusivity (MD) images for DTI-BRAIN, and transfer constant ( $K^{trans}$ ) images for DCE-ABDOMEN. More details on the image datasets are provided in the Experiments section of the present article.

**Registration accuracy.** Figure 1 provides a visualization of the image alignment for a CT-LUNG dataset, gathering 10 CT images acquired at different time points from the lung area of a patient. Misalignments due to respiratory motion are visible when no registration is applied between the images (Fig. 1a), while they disappear after applying image registration based on Huizinga’s  $\mathcal{D}_{PCA2}$  (Fig. 1b) or on the total correlation dissimilarity measure  $\mathcal{D}_{TC}$  proposed in this article (Fig. 1c). Visual differences between the results obtained with  $\mathcal{D}_{PCA2}$  and  $\mathcal{D}_{TC}$  are more limited and harder to identify.

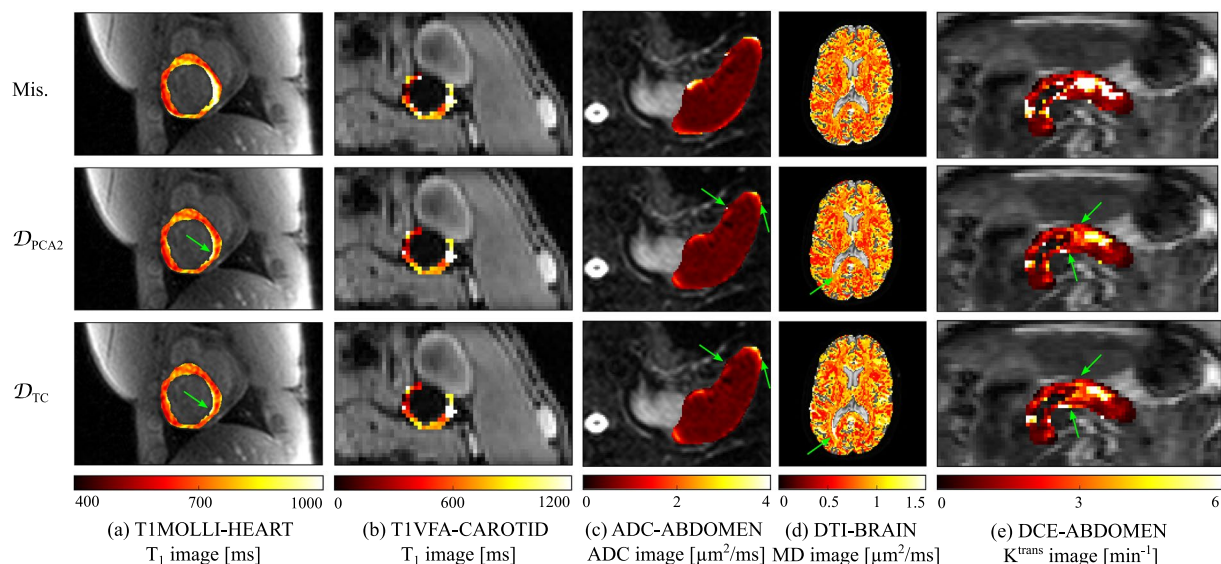
For the five qMRI datasets, Fig. 2 provides quantitative parameter images obtained by applying curve fitting to the images before registration, after registration using Huizinga’s  $\mathcal{D}_{PCA2}$  groupwise dissimilarity measure, and after registration using the total correlation dissimilarity measure  $\mathcal{D}_{TC}$  proposed in this article. The fitting models used to derive the qMRI images assume that spatial correspondence is ensured between the images used for curve fitting. It is therefore expected that quantitative images obtained after image registration will be more reliable than without image registration<sup>11,12</sup>. Based on Fig. 2, visual differences in the estimates tissue maps are easily noticeable between the case before image registration, on the one hand, and the cases with  $\mathcal{D}_{PCA2}$  or  $\mathcal{D}_{TC}$ , on the other hand. Such differences are particularly visible at organ interfaces. Slighter changes, identified by green arrows, can be identified between the tissue maps obtained with  $\mathcal{D}_{PCA2}$  and  $\mathcal{D}_{TC}$ .

Full registration accuracy results in terms of landmark/volume correspondence (mTRE or Dice coefficient), registration transformation smoothness (denoted  $STD_{\det(\partial T_g / \partial x)}$ ), and uncertainty estimation (Cramér-Rao lower bound, denoted CRLB), are provided as supplementary material (Tables S1 to S6) for the following dissimilarity measures: pairwise mutual information  $\mathcal{D}_{MI}$ , Huizinga’s dissimilarity measures based on PCA  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$ , and the total correlation dissimilarity measure proposed in this article  $\mathcal{D}_{TC}$ .

Table 1 presents a partial version of the registration accuracy results, based on the middle value of the control point spacings that were used for the non-rigid B-spline transformation model: 13 mm for CT-LUNG, 64 mm for T1MOLLI-HEART, 16 mm for T1VFA-CAROTID, 64 mm for ADC-ABDOMEN, and 64 mm for

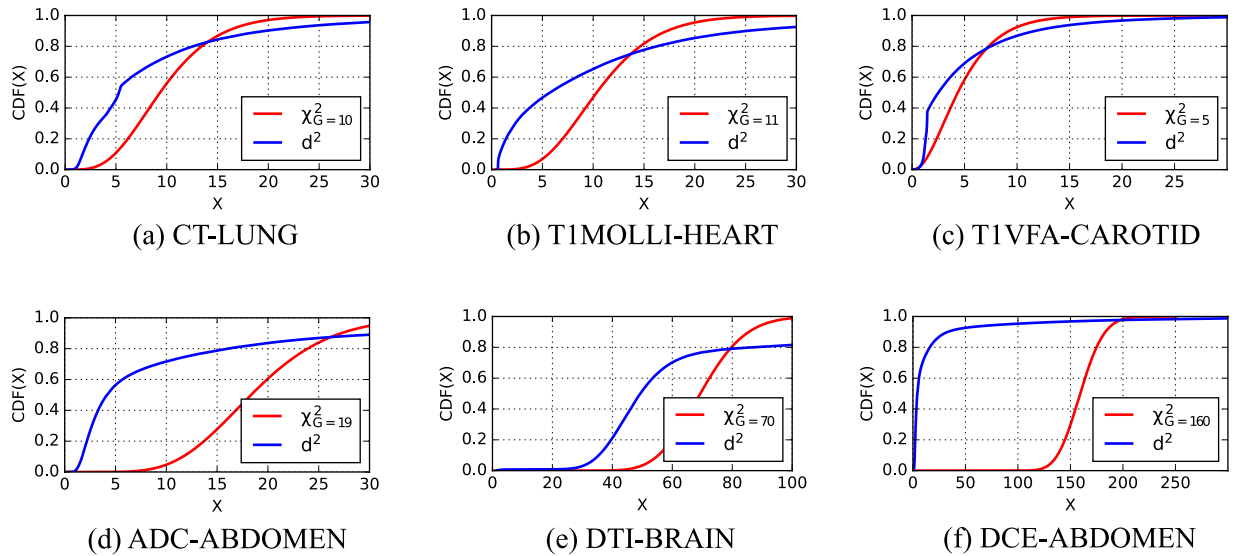
	CT	T1MOLLI	T1VFA	ADC	DTI	DCE
	LUNG	HEART	CAROTID	ABDOMEN	BRAIN	ABDOMEN
<b>(a) Dice coefficients or mTRE values (mean value <math>\pm</math> standard deviation)</b>						
	mTRE [mm]	Dice [%]	mTRE [mm]	Dice [%]	—	mTRE [mm]
Mis.	6.72 $\pm$ 2.51	48 $\pm$ 8	1.47 $\pm$ 0.54	70 $\pm$ 4	—	8.49 $\pm$ 4.54
$\mathcal{D}_{MI}$	1.43 $\pm$ 0.23	37 $\pm$ 11	1.22 $\pm$ 0.43	64 $\pm$ 16	—	6.46 $\pm$ 2.32
$\mathcal{D}_{PCA}$	1.40 $\pm$ 0.37	53 $\pm$ 7	1.11 $\pm$ 0.42	71 $\pm$ 5	—	6.11 $\pm$ 2.33
$\mathcal{D}_{PCA2}$	1.56 $\pm$ 0.55	52 $\pm$ 11	1.08 $\pm$ 0.39	75 $\pm$ 5	—	5.99 $\pm$ 2.18
$\mathcal{D}_{TC}$	1.42 $\pm$ 0.40	53 $\pm$ 11	1.09 $\pm$ 0.40	74 $\pm$ 5	—	6.18 $\pm$ 2.40
<b>(b) Transformation smoothness <math>STD_{\det(\partial T_g/\partial x)}</math> [%] (mean value <math>\pm</math> standard deviation)</b>						
Mis.	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	—	0 $\pm$ 0
$\mathcal{D}_{MI}$	15 $\pm$ 4	7 $\pm$ 2	2 $\pm$ 0	8 $\pm$ 3	—	4 $\pm$ 2
$\mathcal{D}_{PCA}$	8 $\pm$ 2	2 $\pm$ 1	2 $\pm$ 1	3 $\pm$ 2	—	4 $\pm$ 2
$\mathcal{D}_{PCA2}$	7 $\pm$ 2	1 $\pm$ 1	1 $\pm$ 0	3 $\pm$ 1	—	2 $\pm$ 1
$\mathcal{D}_{TC}$	8 $\pm$ 2	2 $\pm$ 1	1 $\pm$ 0	5 $\pm$ 2	—	4 $\pm$ 2
<b>(c) Uncertainty estimation <math>90^{\text{th}} \sqrt{\text{CRLB}}</math> of the fitted parameters (mean value <math>\pm</math> standard deviation)</b>						
	—	$T_1$ [ms]	$T_1$ [ms]	ADC [ $\mu\text{m}^2/\text{ms}$ ]	MD [ $\mu\text{m}^2/\text{ms}$ ]	$K^{\text{trans}}$ [ $\text{min}^{-1}$ ]
Mis.	—	92 $\pm$ 19	>1000	1.37 $\pm$ 0.83	0.096 $\pm$ 0.029	2.84 $\pm$ 2.30
$\mathcal{D}_{MI}$	—	97 $\pm$ 16	501 $\pm$ 83	0.25 $\pm$ 0.05	0.084 $\pm$ 0.028	3.64 $\pm$ 4.13
$\mathcal{D}_{PCA}$	—	87 $\pm$ 16	498 $\pm$ 93	0.23 $\pm$ 0.06	0.085 $\pm$ 0.029	1.52 $\pm$ 1.18
$\mathcal{D}_{PCA2}$	—	83 $\pm$ 12	510 $\pm$ 110	0.27 $\pm$ 0.05	0.084 $\pm$ 0.028	1.27 $\pm$ 0.92
$\mathcal{D}_{TC}$	—	77 $\pm$ 13	500 $\pm$ 96	0.32 $\pm$ 0.05	0.085 $\pm$ 0.028	1.87 $\pm$ 1.79

**Table 1.** Registration results.

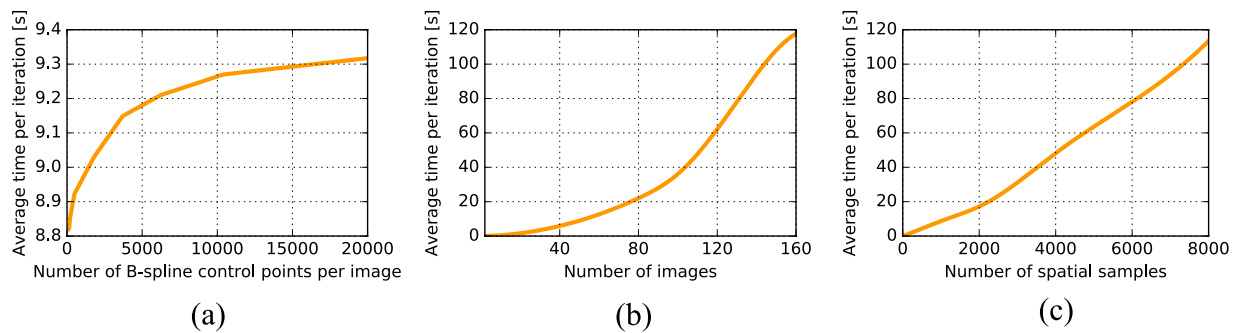


**Figure 2.** Tissue maps generated before image registration (top), after image registration with  $\mathcal{D}_{PCA2}$  (middle), and after image registration with  $\mathcal{D}_{TC}$  (bottom). The fitted values are shown in the myocardium for T1MOLLI-HEART, in the carotid artery wall for T1VFA-ABDOMEN, in the spleen for ADC-ABDOMEN, in the brain parenchyma for DTI-BRAIN, and in the pancreas for DCE-ABDOMEN. Slight visual changes between the tissue maps obtained with  $\mathcal{D}_{PCA2}$  and  $\mathcal{D}_{TC}$  are identified by green arrows.

DCE-ABDOMEN. Registration performances in terms of landmark correspondence (mean target registration error, denoted mTRE) or overlap of volumes of interest (Dice coefficients) are given in Table 1a. For all dataset, better alignments (i.e. lower mTRE) or overlaps (i.e. higher Dice coefficients) were obtained with the groupwise measures  $\mathcal{D}_{TC}$ ,  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  than with pairwise mutual information  $\mathcal{D}_{MI}$ , with one exception: the mTRE obtained with  $\mathcal{D}_{PCA2}$  for the CT-LUNG dataset is higher than the mTRE obtained with  $\mathcal{D}_{MI}$ . The Dice coefficients and mTRE results are very similar for  $\mathcal{D}_{TC}$ ,  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$ . The only case for which  $\mathcal{D}_{TC}$  performs slightly worse than the two other groupwise measures is on the DCE-ABDOMEN dataset. Table 1b provides values for the



**Figure 3.** Cumulative distribution functions for one subject of the six image datasets (aligned case). The observed CDF (blue) is compared with the theoretical CDF of a chi-square distribution with  $G$  degrees of freedom (red).



**Figure 4.** Average time per iteration with respect to the number of B-spline control points per image (a), the number of images  $G$  (b), and the number of spatial samples (c).

transformation smoothness  $\text{STD}_{\det(\partial T_g / \partial \mathbf{x})}$ . In all cases,  $\mathcal{D}_{\text{TC}}$ ,  $\mathcal{D}_{\text{PCA}}$  and  $\mathcal{D}_{\text{PCA2}}$  yield lower (i.e. better) values of  $\text{STD}_{\det(\partial T_g / \partial \mathbf{x})}$  than  $\mathcal{D}_{\text{MI}}$ . The only case for which  $\mathcal{D}_{\text{TC}}$  performs slightly worse than the two other groupwise measures is on the T1VFA-CAROTID dataset. Table 1c provides uncertainty estimations of the qMRI fit ( $90^{\text{th}} \sqrt{\text{CRLB}}$ ). The results indicate that the values of  $90^{\text{th}} \sqrt{\text{CRLB}}$  are lower (i.e. better) with  $\mathcal{D}_{\text{TC}}$  than with  $\mathcal{D}_{\text{MI}}$  for the T1MOLLI-HEART and DCE-ABDOMEN datasets, while they are quite similar for T1VFA-CAROTID and DTI-BRAIN, and higher (i.e. worse) for the ADC-ABDOMEN dataset. The  $90^{\text{th}} \sqrt{\text{CRLB}}$  obtained with  $\mathcal{D}_{\text{TC}}$  is higher than the  $90^{\text{th}} \sqrt{\text{CRLB}}$  obtained with  $\mathcal{D}_{\text{PCA}}$  and  $\mathcal{D}_{\text{PCA2}}$  for two datasets (ADC-ABDOMEN and DCE-ABDOMEN), while it is similar or better for three datasets (T1MOLLI-HEART, T1VFA-CAROTID, and DTI-BRAIN). The full results (Tables S1 to S6) are consistent with the results presented in Table 1a–c.

**Multivariate joint normality.** As detailed in the Method section, the computation of the total correlation dissimilarity measure  $\mathcal{D}_{\text{TC}}$  that we propose is based on the approximation that the intensity distribution of the images to register is multivariate normal. Cumulative distribution functions (CDF) of the squared Mahalanobis distance  $d^2$ , representing the intensity distribution for each of the six dataset types, are plotted in Fig. 3. According to these plots, none of these measure CDF follows the theoretical multivariate normal CDF ( $\chi_G^2$  distribution), which suggests that the image intensities of the images do not follow a multivariate normal distribution.

**Computational efficiency of total correlation  $\mathcal{D}_{\text{TC}}$ .** Figure 4 illustrates the evolution of the average time per iteration obtained with groupwise total correlation  $\mathcal{D}_{\text{TC}}$  for three image registration parameters: the number of B-spline control points per image, the number of images  $G$ , and the number of spatial samples taken to evaluate the dissimilarity measure. The results show that the average registration time per iteration monotonically increases with each of the considered registration parameter. With the present implementation of  $\mathcal{D}_{\text{TC}}$  and of the registration components of the elastix software used to perform the registrations, the results indicate that the

number of B-spline control points has a limited influence on the average time per iteration as it remains close to 9 seconds for the whole span of numbers of B-spline control points that we considered. The experiments suggest that the number of images  $G$  influences the computation time most. For instance, when the number of image is  $G = 40$ , the average iteration time is 5 seconds, while this time reaches about two minutes for  $G = 160$  images. In terms of the number of spatial samples, multiplying the number of spatial samples by 4 ends up in an average time per iteration that is multiplied by 6.

## Discussion

The focus of this paper was to adapt a multivariate version of mutual information in the context of the group-wise registration of medical images, so that it can be used to register two or more images in one optimization procedure.

Among the main multivariate versions of mutual information, namely interaction information  $\mathcal{D}_{II}$ , total correlation  $\mathcal{D}_{TC}$  and dual total correlation  $\mathcal{D}_{TC}$ , total correlation  $\mathcal{D}_{TC}$  theoretically allows to quantify the shared information between any subset of the images to register. Besides, the expression of total correlation is particularly straightforward to apply for the registration of  $G \geq 2$  images, provided that the image intensity distribution is approximated by a multivariate normal distribution.

The expression of the approximated total correlation dissimilarity measure  $\mathcal{D}_{TC}$  that we devise is remarkably analogous to the expressions of two other dissimilarity measures  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  introduced by Huizinga *et al.*<sup>11</sup>, which were developed based on the intuition that an aligned set of images can be described by a small number of high eigenvalues. The expressions of these dissimilarity measures are all sums of functions of the eigenvalues of the correlation matrix  $\mathbf{K}$  (compare Equations (18), (25) and (26)). Huizinga *et al.*<sup>11</sup> had proposed to weigh more the last eigenvalues (the  $\lambda_i$  with the highest  $i$  indexes) than the first ones (the  $\lambda_i$  with the lowest  $i$  indexes) so that as much variance as possible is explained by a few large eigenvectors. The form of  $\mathcal{D}_{TC}$  obtained in this study confirms the intuition of Huizinga *et al.*<sup>11</sup>, since the natural logarithm in Equation (18) also puts more weight on the lower eigenvalues than on the higher ones.

Results obtained on a dynamic imaging dataset and on five multi-parametric datasets show that the total correlation method that we propose yields comparable results as PCA-based methods of Huizinga *et al.*<sup>11</sup>, and better registration results than pairwise mutual information  $\mathcal{D}_{MI}$ . The main advantage of  $\mathcal{D}_{TC}$  with respect to  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  is that it is more theoretically justified: the contribution of each eigenvalue used to compute  $\mathcal{D}_{TC}$  is automatically calibrated and follows naturally from the concepts of multivariate mutual information, whereas for  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$ , the eigenvalue calibration was empirically chosen.

Our study shows that even though the intensity distribution of the datasets to register is not multivariate normal (Fig. 3),  $\mathcal{D}_{TC}$  yields registration results that are better than mutual information and similar to the PCA dissimilarity measures of Huizinga *et al.*<sup>11</sup>. This is the case for a total of six diverse multi-parametric datasets, which suggests that approximating the intensity distributions, as done in this article, yields optimization minima that result in comparable or better registration accuracies than other state-of-the-art pairwise and groupwise techniques. On multi-parametric datasets, the results suggest that the approximation by a multivariate normal distribution is not detrimental to the registration results.

In the current implementation of the total correlation registration technique, increases in the number of images  $G$  have the largest impact on the average time per iteration, which is not surprising as both the amount of image data to register and the number of transformations to estimate scale with a factor  $G$ ; moreover, estimating the correlation matrix  $\mathbf{K}$  and its eigenvalue decomposition become increasingly computationally demanding. Further optimizations could improve the scalability of total correlation with respect to the number of images. The computation time also scales linearly with the number of spatial samples. Thanks to the use of the stochastic gradient descent optimization routine, we were able to use a relatively low number (2048) of spatial samples in our experiments, while still achieving accurate registration.

Other possible applications of the total correlation dissimilarity measure proposed in this article include motion tracking in ultrasound image sequences<sup>13,14</sup>, motion compensation in dynamic PET<sup>15</sup> or dynamic contrast-enhanced CT<sup>16</sup>, and for population template construction<sup>17</sup>. Future research should validate the performance of the method in such contexts.

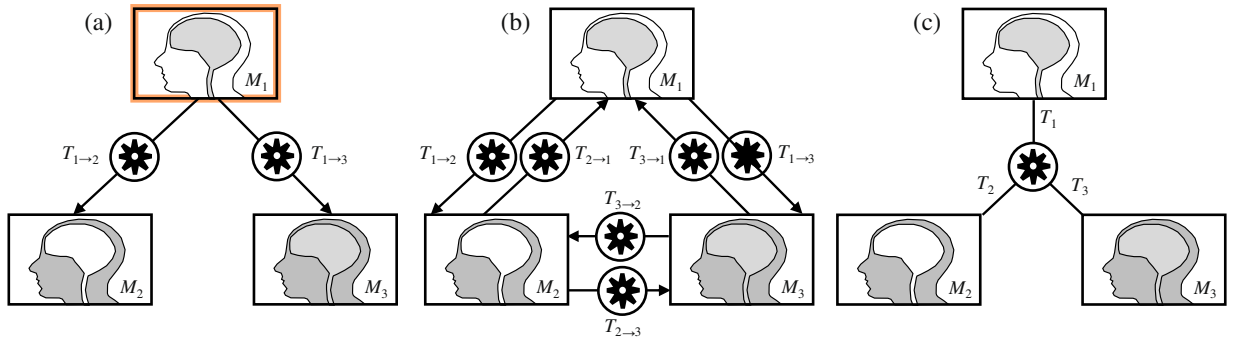
## Conclusion

In conclusion, we proposed an implementation of an approximated version of total correlation  $\mathcal{D}_{TC}$  for which the registration results are comparable to the results obtained with the dissimilarity measures of Huizinga *et al.*<sup>11</sup>, on multi-parametric datasets. Our results indicate that approximating the intensity distributions by a joint normal distribution for the sake of efficient calculation of the entropy, used to derive total correlation  $\mathcal{D}_{TC}$ , does not constitute a limitation in the practical application of  $\mathcal{D}_{TC}$  to quantitative imaging datasets. Total correlation  $\mathcal{D}_{TC}$  has the advantage of being elegant and theoretically justified, while the dissimilarity measures  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  proposed by Huizinga *et al.*<sup>11</sup> were elaborated empirically. Additionally, groupwise total correlation  $\mathcal{D}_{TC}$  offers an alternative to pairwise registration based on mutual information on multi-parametric imaging datasets.

## Method

Let us consider  $\mathcal{M} = \{M_1, \dots, M_G\}$ , a series of  $G$  images that have to be registered. Each image  $M_g$  consists of  $N$  voxels. To quantify how well the  $G$  images are aligned, a dissimilarity measure has to be defined. In this study, we consider dissimilarity measures based on the concepts of mutual information. We choose the convention to formulate the measures as dissimilarity measures instead of similarity measures, so that the registration problem can be written as a cost function minimization problem.





**Figure 5.** (a) Pairwise registration scheme (the orange frame indicates that this method requires the selection of a reference image), (b) semi-groupwise registration scheme proposed by Seghers *et al.*<sup>20</sup>, and (c) groupwise registration scheme.

**Pairwise mutual information.** Mutual information is a robust measure that is commonly used for the pairwise registration of datasets of medical images, including multimodal datasets<sup>3</sup>. For  $G = 2$  images  $M_1$  and  $M_2$ , the negated mutual information  $\mathcal{D}_{MI}$  is computed as follows<sup>1,3</sup>:

$$\mathcal{D}_{MI}(M_1, M_2) = H(M_1, M_2) - H(M_1) - H(M_2) \tag{1}$$

with  $H(M_1)$  the entropy<sup>18</sup> of image  $M_1$ ,  $H(M_2)$  the entropy of image  $M_2$ , and  $H(M_1, M_2)$  the joint entropy of  $M_1$  and  $M_2$ . For two images  $M_1$  and  $M_2$ , the joint entropy can be computed as follows<sup>19</sup>:

$$H(M_1, M_2) = - \sum_{x_1} \sum_{x_2} P(x_1, x_2) \ln [P(x_1, x_2)] \tag{2}$$

where  $x_1$  and  $x_2$  represent the discrete values of images  $M_1$  and  $M_2$ , respectively.  $P(x_1, x_2)$  is the probability of these values occurring together.  $P(x_1, x_2) \ln [P(x_1, x_2)]$  is defined to be 0 if  $P(x_1, x_2)$  equals 0.

When the dataset of images to register contains  $G > 2$  images, it is still possible to use a pairwise method to register the images, but several independent registration procedures have to be performed. A typical method consists of selecting one of the images as fixed reference, and then successively applying pairwise registration with the remaining  $G - 1$  images considered as moving images (Fig. 5a). This technique is not well suited to registration problems for which there is no obvious reference image. Besides, the registration results may be different according to the choice of fixed reference image, as shown by Geng *et al.*<sup>4</sup>. Seghers *et al.*<sup>20</sup> introduced a method that we will refer to as semi-groupwise, which is based on multiple pairwise registrations and does not require the selection of a reference space. For each  $i$ ,  $M_i$  is taken as fixed image and  $G - 1$  independent registration are performed between each remaining image,  $M_j$ , yielding  $G - 1$  transformations  $T_{i \rightarrow j}$  per fixed image  $M_i$ . Each image  $M_i$  is then resampled into an average or mid-point image space using  $\bar{T}_i^{-1}(x)$ , the inverse of the arithmetic mean of the transformations  $T_{i \rightarrow j}$  (Fig. 5b). The method of Seghers *et al.*<sup>20</sup> has the disadvantage of requiring  $G \times (G - 1)$  registration procedures, which becomes computationally complex when  $G$  grows. It also does not allow to register all images in a single optimization procedure.

**Groupwise dissimilarity measures based on multivariate mutual information.** Groupwise registration techniques allow to register  $G \geq 2$  images. In this study, we will focus on groupwise techniques that allow to register all images in one optimization procedure, and that treat the images equally (Fig. 5c). In particular, the order in which the images are supplied should have no influence on the value of the groupwise dissimilarity measure  $\mathcal{D}(M_1, M_2, \dots, M_G)$ , and therefore no influence on the registration results.

This article more precisely focuses on groupwise generalizations of mutual information, given the wide interest and range of applications of that dissimilarity measure in the context of pairwise image registration<sup>3</sup>. There exist multiple multivariate forms of mutual information<sup>5-7</sup>, the concepts of which can be used for groupwise image registration.

The first multivariate generalization of mutual information is known as interaction information<sup>5</sup>, denoted  $\mathcal{D}_{II}$ . It measures the amount of information shared by all the images. For the  $G$  images of  $\mathcal{M}$ , negated interaction information is written:

$$\mathcal{D}_{II}(\mathcal{M}) = \sum_{V \subseteq \mathcal{M}} (-1)^{G-|V|} H(V) \tag{3}$$

with  $V \subseteq \mathcal{M}$  meaning that  $V$  can be any subset of images of  $\mathcal{M}$  (e.g. if  $G = 3$ , then  $V$  successively represents the following subsets of images:  $\{M_1\}$ ,  $\{M_2\}$ ,  $\{M_3\}$ ,  $\{M_1, M_2\}$ ,  $\{M_1, M_3\}$ ,  $\{M_2, M_3\}$ , and  $\{M_1, M_2, M_3\}$ ),  $|V|$  the number of images in the corresponding subset, and  $H(V)$  the joint entropy of the subset  $V$ . For  $G$  images  $M_1 \dots M_G$ , the joint entropy is the generalization of Equation (2):

$$H(M_1, \dots, M_G) = - \sum_{x_1} \dots \sum_{x_G} P(x_1, \dots, x_G) \ln [P(x_1, \dots, x_G)] \tag{4}$$

where the  $x_1, \dots, x_G$  are the values of images  $M_1, \dots, M_G$ , respectively. The same definitions as for  $P(x_1, x_2)$  and  $P(x_1, x_2) \ln [P(x_1, x_2)]$  are directly extended for  $P(x_1, \dots, x_G)$  and  $P(x_1, \dots, x_G) \ln [P(x_1, \dots, x_G)]$ . Interaction information quantifies the information shared together by images  $M_1, \dots, M_G$ <sup>21</sup>. This means that if at least one of the images of  $\mathcal{M}$  shares no information with all other images, the interaction information will be zero<sup>21,22</sup>.

The second form of multivariate mutual information, called total correlation<sup>6</sup>, measures the amount of information shared between any subset of  $\mathcal{M}$ . The negated total correlation is written as:

$$\mathcal{D}_{TC}(\mathcal{M}) = H(\mathcal{M}) - \left[ \sum_{g=1}^G H(M_g) \right] \tag{5}$$

with  $H(\mathcal{M})$  the joint entropy of the images of the set  $\mathcal{M} = \{M_1, \dots, M_G\}$ .

The third form is a refinement of total correlation called dual total correlation<sup>7</sup>, and can be written as:

$$\mathcal{D}_{DTC}(\mathcal{M}) = \left[ \sum_{g=1}^G H(M_g | (\mathcal{M} \setminus M_g)) \right] - H(\mathcal{M}) \tag{6}$$

with  $\mathcal{M} \setminus M_g$  the set of images  $\{M_1, \dots, M_G\}$  without  $M_g$ .  $H(M_g | (\mathcal{M} \setminus M_g))$  is the conditional entropy<sup>19</sup> of  $M_g$  given  $\mathcal{M} \setminus M_g$ . In other terms,  $H(M_g | (\mathcal{M} \setminus M_g))$  is the entropy of the image  $M_g$  given the knowledge of images  $\{M_1, \dots, M_{g-1}, M_{g+1}, \dots, M_G\}$ .

Theoretically, both total correlation and dual total correlation quantify the amount of shared information between all possible combinations of images, while interaction information only quantifies the amount of information shared by all images<sup>23</sup>. Venn diagrams<sup>19,23,24</sup> for  $\mathcal{D}_{II}$ ,  $\mathcal{D}_{TC}$  and  $\mathcal{D}_{DTC}$  are shown in Fig. 6. In the context of image registration,  $\mathcal{D}_{TC}$  and  $\mathcal{D}_{DTC}$  seem more adapted than  $\mathcal{D}_{II}$ , as they are built to quantify shared information not only between all images, but also between any of their subsets<sup>21,22</sup>. In particular, including an image with little dependence towards the others would impair the registration of the remaining images when using  $\mathcal{D}_{II}$ , while this would theoretically not be the case when using  $\mathcal{D}_{TC}$  or  $\mathcal{D}_{DTC}$ . We therefore chose to consider the dissimilarity measures based on total correlation in order to construct a groupwise dissimilarity measure.

**Groupwise total correlation.** In this section, we describe how total correlation, as expressed in Formula (5), can be brought to practical use in the context of image registration. As such, computing total correlation implies computing the joint entropy  $H(\mathcal{M})$ , but this computation is subject to the curse of dimensionality<sup>25</sup>: the evaluation of joint entropy requires to compute a  $G$ -dimensional joint histogram that becomes increasingly sparser as  $G$  increases, and therefore becomes computationally prohibitive.

Let us consider a random variable  $\mathbf{X} \in \mathbb{R}^G$  following a  $G$ -variate normal distribution given by:

$$f(\mathbf{X}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp\left(-\frac{1}{2}(\mathbf{X} - \mu)^T \mathbf{C}^{-1}(\mathbf{X} - \mu)\right) \tag{7}$$

with  $\mu \in \mathbb{R}^G$  an expectation vector,  $\mathbf{C} \in \mathbb{R}^{G \times G}$  a covariance matrix, and with  $\det(\cdot)$  the determinant operator. Ali Ahmed *et al.*<sup>26</sup> have shown that the entropy of the multivariate normal variable  $\mathbf{X}$  may be written as:

$$H(\mathbf{X}) = \frac{G}{2} + \frac{G}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\mathbf{C})) \tag{8}$$

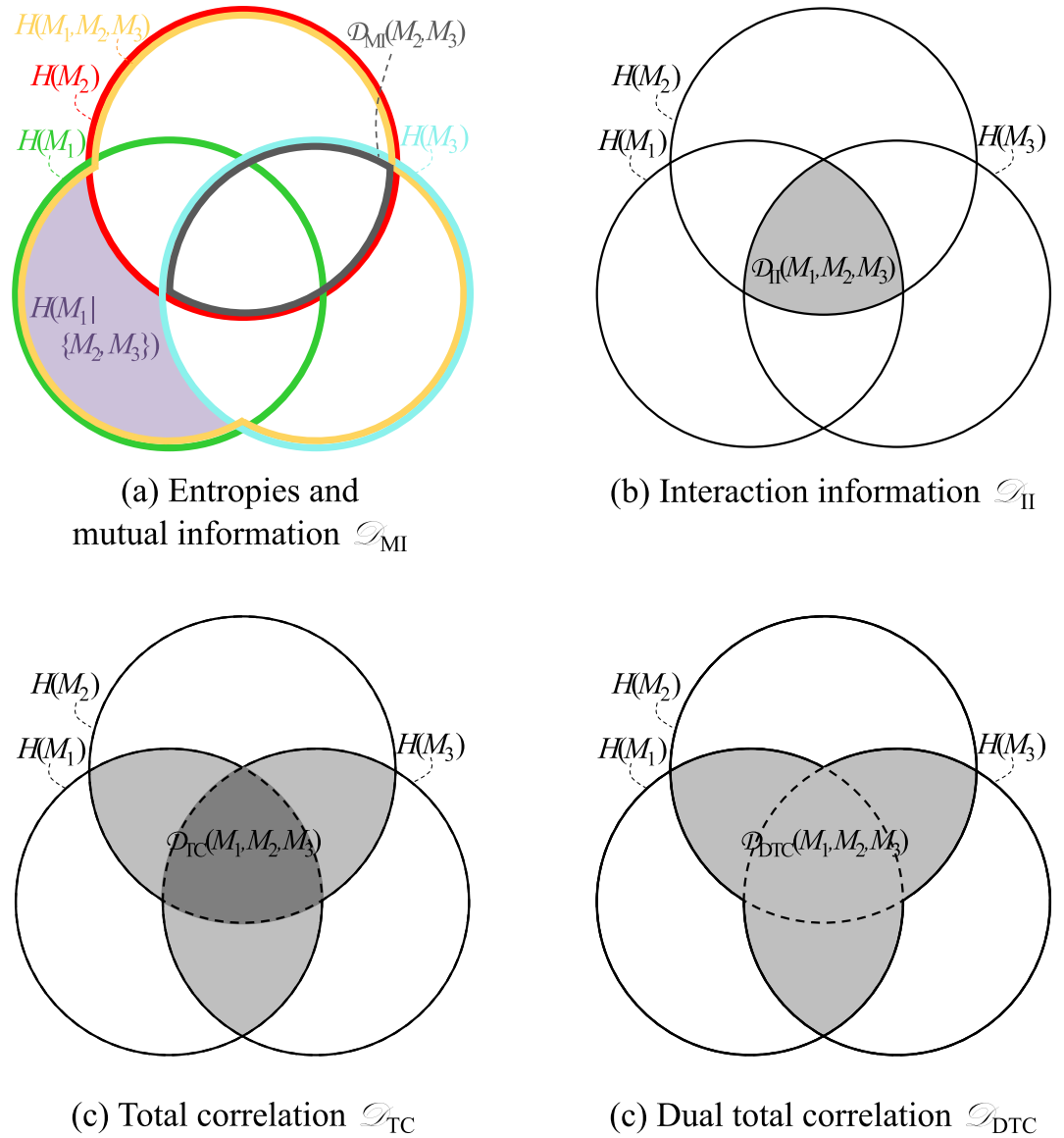
To circumvent the curse of dimensionality, and make it possible to use registration in a groupwise manner on datasets containing any number  $G \geq 2$  images, we propose to use Equation (8) in the context of  $G$  images  $\mathcal{M} = \{M_1, \dots, M_G\}$ . For the sake of efficient calculation of the entropy, we approximate the intensity distribution of the images by a joint normal distribution, and we make the hypothesis that the minimum of the resulting cost function is still a good solution for the underlying registration problem. Let  $\mathbf{M}$  be a  $N \times G$  matrix in which each image  $M_g$  is represented by a column. The matrix  $\mathbf{C}$  of covariances between the images  $M_g$  is obtained as follows:

$$\mathbf{C} = \frac{1}{N-1}(\mathbf{M} - \overline{\mathbf{M}})^T(\mathbf{M} - \overline{\mathbf{M}}) \tag{9}$$

with  $\overline{\mathbf{M}}$ , a matrix that has in each of its column the column-wise average of  $\mathbf{M}$ . To make the method robust to linear intensity scalings and offsets, we incorporate an intensity standardization (i.e. z-score) within the definition of the dissimilarity measure. This is done by computing the entropy  $H(\mathcal{M})$  using the correlation matrix  $\mathbf{K}$  instead of the covariance matrix  $\mathbf{C}$ , with:

$$\mathbf{K} = \Sigma^{-1} \mathbf{C} \Sigma^{-1} \tag{10}$$

with  $\Sigma$  a diagonal matrix with the standard deviations of the columns of  $\mathbf{M}$  as its diagonal elements. A diagonal element  $\Sigma_{gg}$  of  $\Sigma$  verifies:



**Figure 6.** Venn diagram representations for three images  $M_1, M_2$  and  $M_3$ . (a) The green, red and cyan circle represent the entropy of each image. The fact that the images share information is symbolized by the fact that these circles overlap. Subfigures (b), (c) and (d) were constructed based on Equations (3), (5) and (6). In (c), the dark grey area signifies that its contribution to the dissimilarity measure is twice as high as the contribution of each light-grey area.

$$\Sigma_{gg} = \frac{1}{N-1} \sum_{i=1}^N (M_{g,i} - \bar{M}_g)^2 \tag{11}$$

where the  $M_{g,i}$  are the individual voxel values and  $\bar{M}_g$  the average voxel value of image  $\bar{M}_g$ . By construction, each diagonal element of the correlation matrix  $\mathbf{K}$  is equal to 1. The expression of the joint entropy therefore becomes:

$$H(\mathcal{M}) = \frac{G}{2} + \frac{G}{2} \ln(2\pi) + \frac{1}{2} \ln(\det(\mathbf{K})) \tag{12}$$

Equation (12) can also be used to derive the marginal entropies  $H(M_g)$ . When considering only one image  $M_g$ , the correlation matrix  $\mathbf{K}$  is the scalar 1. All  $H(M_g)$  are therefore constant and equal to:

$$H(M_g) = \frac{1}{2} + \frac{1}{2} \ln(2\pi) \tag{13}$$

By combining Equations (5), (12) and (13), we define the dissimilarity measure based on total correlation  $\mathcal{D}_{TC}$  as follows:



$$\mathcal{D}_{\text{TC}}(\mathcal{M}) = \frac{1}{2} \ln(\det(\mathbf{K})) = \frac{1}{2} \sum_{j=1}^G \ln \lambda_j \quad (14)$$

using  $\det(\mathbf{K}) = \prod_{j=1}^G \lambda_j$ , with  $\lambda_j$  the  $j^{\text{th}}$  eigenvalue of  $\mathbf{K}$ , and  $\lambda_j > \lambda_{j+1}$ . Such a simple expression was not found for dual total correlation, which is why we selected total correlation as groupwise dissimilarity measure.

**Gradient-based optimization and implementation.** To implement the approximated version of  $\mathcal{D}_{\text{TC}}$  provided in Equation (14), we define an interpolation scheme based on B-splines. This scheme associates with each original image  $M_g$  a continuous and differentiable function  $M_g(\mathbf{x})$  of the spatial coordinate  $\mathbf{x}$ . The aim is to simultaneously bring the images  $M_g(\mathbf{x})$  to an average space by means of a transformation  $T(\mathbf{x}, \mu)$ , where  $\mu$  is a vector containing the parameters  $\mu_g$  that correspond to the transformation  $T_g(\mathbf{x}, \mu_g)$  related to each image  $M_g$ . Examples of transformation models are the affine model, or the non-rigid model in which deformations are modeled by cubic B-splines<sup>27</sup>.

In the groupwise scheme, the measure  $\mathcal{D}$  quantifies the dissimilarity between all transformed images  $M_g(T_g(\mathbf{x}, \mu_g))$ . We adopted the pull-back definition of a warped image. Groupwise registration can therefore be formulated as the constrained minimization of the dissimilarity measure  $\mathcal{D}$  with respect to  $\mu$ , as previously proposed by Huizinga *et al.*<sup>11</sup>:

$$\hat{\mu} = \arg \min_{\mu} \mathcal{D}(M_1(T_1(\mathbf{x}, \mu_1)), \dots, M_G(T_G(\mathbf{x}, \mu_G))) \quad (15)$$

subject to the following constraint, allowing to define a mid-point space<sup>28</sup>:

$$\sum_{g=1}^G \mu_g = 0 \quad (16)$$

The implementation of the total correlation dissimilarity measure  $\mathcal{D}_{\text{TC}}$  was performed as part of the open source software package elastix<sup>29</sup>. The adaptive stochastic gradient descent (ASGD) developed by Klein *et al.*<sup>30</sup> is used as optimization method for image registration. This method randomly samples positions in the image space at each iteration in order to reduce computation time. Sampling is done off the voxel grid, which was shown to be necessary to reduce interpolation artefacts<sup>29</sup>. A multi-resolution strategy is used: the images are Gaussian-blurred with a certain standard deviation, which is decreased at each resolution level. This means that the large deformations are corrected first, and that finer deformations are corrected in subsequent levels. Linear interpolation is used to interpolate the images during registration, which reduces computation time, but cubic B-spline interpolation was used to produce the final registered images. For the chosen ASGD optimization method, the gradient of the dissimilarity measure is needed. Based on Equation (14) and van der Aa *et al.*<sup>31</sup>, we find:

$$\frac{\partial \mathcal{D}_{\text{TC}}}{\partial \mu} = \frac{1}{2} \sum_{j=1}^G \frac{1}{\lambda_j} \frac{\partial \lambda_j}{\partial \mu} = \frac{1}{2} \sum_{j=1}^G \frac{1}{\lambda_j} \left( v_j^T \frac{\partial \mathbf{K}}{\partial \mu} v_j \right) \quad (17)$$

where  $v_j^T$  is the  $j^{\text{th}}$  eigenvector of  $\mathbf{K}$ . Similarly to van der Aa *et al.*<sup>31</sup>, we assume that the repetition of eigenvalues is unlikely.

When the eigenvalues  $\lambda_j$  tend towards zero, evaluating  $\mathcal{D}_{\text{TC}}$  implies taking the natural logarithm of a near-zero number (as shown in Equation (14)), which might result in a failing optimization. We therefore introduce an adjusting constant  $c \in \mathbb{R}^+$  that is added to the eigenvalue  $\lambda_j$  before taking the natural logarithm:

$$\mathcal{D}_{\text{TC}}(\mathcal{M}) = \frac{1}{2} \ln(\det(\mathbf{K} + c\mathbf{I})) = \frac{1}{2} \sum_{j=1}^G \ln(\lambda_j + c) \quad (18)$$

where  $\mathbf{I}$  is the identity matrix. The gradient of the adjusted total correlation dissimilarity measure therefore becomes:

$$\frac{\partial \mathcal{D}_{\text{TC}}}{\partial \mu} = \frac{1}{2} \sum_{j=1}^G \frac{1}{\lambda_j + c} \frac{\partial \lambda_j}{\partial \mu} = \frac{1}{2} \sum_{j=1}^G \frac{1}{\lambda_j + c} \left( v_j^T \frac{\partial \mathbf{K}}{\partial \mu} v_j \right) \quad (19)$$

To derive an appropriate value for  $c$ , we make the assumption that the first mode, corresponding to  $\lambda_1$ , accounts for half of the total data variation. Given that the trace of  $\mathbf{K}$  is equal to the sum of its eigenvalues, we can write that  $\text{tr}(\mathbf{K}) = \sum_{i=1}^G \lambda_i$ . In addition, the diagonal elements of the correlation matrix  $\mathbf{K}$  are all equal to 1, which induces that  $\text{tr}(\mathbf{K}) = G = \sum_{i=1}^G \lambda_i$ . The assumption that the first mode accounts for half of the total data variation therefore yields  $\lambda_1 = G/2$ . We then constrain the ratio  $(\lambda_1 + c)/(\lambda_G + c)$  to  $G$ , so that the weights  $1/(\lambda_i + c)$  in Equation (19) remain within a known, finite range. We also make the assumptions that  $c \ll G$  and that  $\lambda_G \ll c$ . This leads to the solution  $c = 0.5$ . In addition to solving a computational issue, the constant  $c$  introduces a lower bound on the variance associated with each eigenvector. Initial experiments confirmed that with this choice for  $c$ , occasional numerical instabilities were successfully eliminated, while not visibly affecting the results in other cases.

Based on Equation (10), the expression of  $\partial \mathbf{K} / \partial \mu_p$  in Equation (19) becomes:

$$\begin{aligned} \frac{\partial \mathbf{K}}{\partial \mu_p} &= \frac{\partial}{\partial \mu_p} \left( \frac{1}{N-1} \Sigma^{-1} (\mathbf{M} - \bar{\mathbf{M}})^T (\mathbf{M} - \bar{\mathbf{M}}) \Sigma^{-1} \right) \\ &= \frac{1}{N-1} \left[ \frac{\partial \Sigma^{-1}}{\partial \mu_p} (\mathbf{M} - \bar{\mathbf{M}})^T (\mathbf{M} - \bar{\mathbf{M}}) \Sigma^{-1} + \Sigma^{-1} \left( \frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \bar{\mathbf{M}}}{\partial \mu_p} \right)^T (\mathbf{M} - \bar{\mathbf{M}}) \Sigma^{-1} \right. \\ &\quad \left. + \Sigma^{-1} (\mathbf{M} - \bar{\mathbf{M}})^T \left( \frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \bar{\mathbf{M}}}{\partial \mu_p} \right) \Sigma^{-1} + \Sigma^{-1} (\mathbf{M} - \bar{\mathbf{M}})^T (\mathbf{M} - \bar{\mathbf{M}}) \frac{\partial \Sigma^{-1}}{\partial \mu_p} \right] \end{aligned} \tag{20}$$

The property of commutativity of the dot product yields:

$$v^T \mathbf{A} \mathbf{B} v = v^T \mathbf{B}^T \mathbf{A}^T v \tag{21}$$

with A and B, two matrices and v a vector. Using Equations (19–21), the derivative of  $\mathcal{D}_{TC}$  with respect to an element  $\mu_p$  becomes:

$$\begin{aligned} \frac{\partial \mathcal{D}_{TC}}{\partial \mu_p} &= \frac{1}{N-1} \sum_{j=1}^G \left[ \frac{1}{\lambda_j + c} \times \left\{ \mathbf{v}_j^T \Sigma^{-1} (\mathbf{M} - \bar{\mathbf{M}})^T (\mathbf{M} - \bar{\mathbf{M}}) \right. \right. \\ &\quad \left. \left. \times \frac{\partial \Sigma^{-1}}{\partial \mu_p} \mathbf{v}_j + \mathbf{v}_j^T \Sigma^{-1} (\mathbf{M} - \bar{\mathbf{M}})^T \left( \frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \bar{\mathbf{M}}}{\partial \mu_p} \right) \Sigma^{-1} \mathbf{v}_j \right\} \right] \end{aligned} \tag{22}$$

To obtain  $\partial \Sigma^{-1} / \partial \mu_p$ , the diagonal elements  $\Sigma_{gg}^{-1}$  of the diagonal matrix  $\Sigma^{-1}$  can be derived one by one:

$$\frac{\partial \Sigma_{gg}^{-1}}{\partial \mu_p} = \frac{\partial}{\partial \mu_p} \left( \frac{1}{N-1} \sum_{i=1}^N (M_{g,i} - \bar{M}_g)^2 \right)^{-\frac{1}{2}} = -\frac{\Sigma_{gg}^{-3}}{N-1} \left( \mathbf{M} - \bar{\mathbf{M}} \right)^T \left( \frac{\partial \mathbf{M}}{\partial \mu_p} - \frac{\partial \bar{\mathbf{M}}}{\partial \mu_p} \right) \Bigg|_{gg} \tag{23}$$

The quantity  $\partial \mathbf{M} / \partial \mu_p$  is computed as follows:

$$\frac{\partial M_g(\mathbf{T}_g(\mathbf{x}, \mu_g))}{\partial \mu_p} = \left( \frac{\partial M_g}{\partial \mathbf{x}} \right)_{\mathbf{T}_g(\mathbf{x}, \mu_g)}^T \left( \frac{\partial \mathbf{T}_g}{\partial \mu_p} \right)_{(\mathbf{x}, \mu_g)} \tag{24}$$

It was verified that the derivative  $\partial \bar{\mathbf{M}} / \partial \mu_p$  of the mean intensities was negligibly small and it was therefore ignored in the implementation.

**Related groupwise dissimilarity measures.** Huizinga *et al.*<sup>11</sup> previously presented two dissimilarity measures, the expressions of which are close to the total correlation measure presented in the current article. Huizinga’s dissimilarity measures are based on PCA and rely on the idea that an aligned set of multi-parametric images can be described by a small number of high eigenvalues, since the underlying model  $m_g$  is low-dimensional (i.e. the size  $\Gamma$  of  $\theta$  is lower than  $G$ ). A misaligned set of multi-parametric images would, on the contrary, be characterized by an eigenvalue spectrum that is more flat: more eigenvalues of average intensity are required for describing the data in that case.

The first dissimilarity measure introduced by Huizinga *et al.*<sup>11</sup>, denoted  $\mathcal{D}_{PCA}$ , quantifies the difference between the sum of all eigenvalues and the sum of the first few eigenvalues:

$$\mathcal{D}_{PCA}(\mathcal{M}) = \sum_{j=1}^G \lambda_j - \sum_{j=1}^L \lambda_j = \sum_{j=L+1}^G \lambda_j \tag{25}$$

with  $L$  a user-defined constant with  $1 \leq L \leq G$ , and  $\sum_{j=1}^G \lambda_j = \text{tr}(\mathbf{K}) = G$ . This means that  $\mathcal{D}_{PCA}$  is the sum of the lowest  $G-L$  eigenvalues. Contrary to  $\mathcal{D}_{PCA}$ , the second dissimilarity measure, denoted  $\mathcal{D}_{PCA2}$ , does not require the selection of an arbitrary cut-off  $L$ . It consists of weighting the last eigenvalues more than the first ones:

$$\mathcal{D}_{PCA2}(\mathcal{M}) = \sum_{j=1}^G j \lambda_j \tag{26}$$

The dissimilarity measures of Huizinga *et al.*<sup>11</sup> were developed based on different ideas than total correlation:  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  were developed based on the concepts of PCA, while  $\mathcal{D}_{TC}$  is a multivariate derivation of mutual information. Nevertheless, the expressions of  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$ , on the one hand, and of  $\mathcal{D}_{TC}$ , on the other hand, happen to resemble each other quite closely: all of them consists of a sum of functions of the eigenvalues.

The main disadvantage of Huizinga’s  $\mathcal{D}_{PCA}$  with respect to the other techniques is that it requires to choose the cut-off  $L$ . In  $\mathcal{D}_{PCA2}$ , this user-defined constant is avoided, but the weights  $j$  in Equation (12) are actually still chosen arbitrarily. For the total correlation dissimilarity measure  $\mathcal{D}_{TC}$  that we propose is that the contribution of each

eigenvalue follows naturally from the derivation of mutual information. A key asset of  $\mathcal{D}_{TC}$  is therefore that the influence of each eigenvalue is automatically calibrated, because the expression of the dissimilarity measure is derived from the concept of mutual information.

**Implementation codes.** The implementation of total correlation will be made available within the open source image registration package elastix, downloadable at the following address: <http://elastix.isi.uu.nl>.

## Experiments

The quantitative imaging datasets previously considered by Huizinga *et al.*<sup>11</sup> are covered by the more generic term of multi-parametric datasets, i.e. datasets  $\{M_1, \dots, M_G\}$  for which the images  $M_g$  are characterized by an underlying model  $m_g$  describing their intensity values, such that:

$$M_g(\mathbf{x}) = m_g(\theta(\mathbf{x})) + \varepsilon(\mathbf{x}) \quad (27)$$

with  $\theta$  a vector (dimension  $\Gamma < G$ ) containing the model parameters, and  $\varepsilon$  the noise at coordinate  $\mathbf{x}$ . An example of a model is the monoexponential model  $m_g(\theta) = S_0 \exp(-b_g u_g^T \mathbf{D} u_g)$  used in diffusion tensor imaging, with  $\theta = (S_0, D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33})$ ,  $u_g$  the diffusion gradient direction vector,  $\mathbf{D}$  a  $3 \times 3$  symmetric diffusion tensor, and  $b$  the b-value<sup>32</sup>.

In particular, Huizinga *et al.*<sup>11</sup> applied the groupwise dissimilarity measures  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  to a variety of multi-parametric datasets, and compared the results with other state-of-the-art techniques: pairwise mutual information  $\mathcal{D}_{MP}$ , the accumulated pairwise estimates (APE) introduced by Wachinger and Navab<sup>33</sup>, the groupwise sum of variances designed by Metz *et al.*<sup>9</sup>, and the groupwise mutual information method of Bhatia *et al.*<sup>10</sup>. Huizinga *et al.*<sup>11</sup> concluded that their measures  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  yielded better or equal registration results with respect to the other tested methods.

The present experiment uses total correlation  $\mathcal{D}_{TC}$  as groupwise dissimilarity measure for the registration of the same datasets as in Huizinga *et al.*<sup>11</sup>. On these datasets, the methods of Huizinga *et al.*<sup>11</sup> were shown to be the best ones, which is why we will compare the registration results of  $\mathcal{D}_{TC}$  with  $\mathcal{D}_{PCA}$  and  $\mathcal{D}_{PCA2}$  only. The results reported by Huizinga *et al.*<sup>11</sup> for the other dissimilarity measures are directly comparable with the results reported in the present study.

**Description of the six image datasets.** The first dataset, denoted CT-LUNG<sup>34</sup>, consists of ten patient subsets containing  $G = 10$  three-dimensional CT images of the thorax. The intensity distribution in this dynamic imaging dataset are analogous in all images, which means that the model  $m_g$  can be considered as a constant (see Equation (27)): it is therefore a particular case of multi-parametric dataset. The second study, denoted T1MOLLI-HEART<sup>35</sup>, consists of nine  $T_1$ -weighted MRI datasets of porcine hearts with transmural myocardial infarction of the lateral wall.  $G = 11$  two-dimensional images were acquired for nine subjects. For each registration case, a voxelwise curve fitting was applied to the registered images, producing quantitative  $T_1$  maps. The third study, denoted T1VFA-CAROTID<sup>36</sup>, involves MRIs of the carotid arteries.  $G = 5$  three-dimensional images were acquired for 8 human patients. For each patient, the images were registered and fitted to obtain quantitative  $T_1$  maps. The fourth study consists of DW-MR images of the abdominal region, and is denoted ADC-ABDOMEN<sup>12</sup>. Five datasets, each of them including  $G = 19$  three-dimensional images, were registered and fitted to produce ADC maps. The fifth study is denoted DTI-BRAIN<sup>37-41</sup> and consists, for each of the five considered datasets, of registering diffusion tensor images (DTI) of the brain. The number of images to register varied between  $G = 33$  and  $G = 70$  for each dataset<sup>11</sup>. The fitted parameter is the mean diffusivity ( $MD$ ). The sixth study involves DCE images of the abdomen. Five DCE-ABDOMEN<sup>42</sup> datasets were acquired, each of them containing  $G = 160$  three-dimensional images. The fitted parameter of interest considered in this study is  $K^{trans}$ . The full descriptions of the fitting models are provided by Huizinga *et al.*<sup>11</sup>.

All human data used in this study came from anonymized datasets. Data from the CT-LUNG dataset was obtained from a publicly available dataset<sup>34</sup> available at the following address: <https://www.dir-lab.com>. The ethics committee of the Academisch Medisch Centrum, Amsterdam, the Netherlands, approved the research related to the T1VFA-CAROTID and DCE-ABDOMEN datasets. The Research Ethics Committee of the Royal Marsden Hospital, United Kingdom, approved the research related to the ADC-ABDOMEN dataset. The medical ethics committee for research in humans of the University Medical Center Utrecht, the Netherlands, approved the research performed on the DTI-BRAIN dataset. Informed consent was obtained from all patients in human datasets. Porcine data from the T1MOLLI-HEART dataset were approved by the Animal Ethics Committee of the Erasmus MC Rotterdam, the Netherlands. All studies were carried out in accordance with the relevant guidelines and regulations.

**Registration characteristics.** We selected the same registration settings as Huizinga *et al.*<sup>11</sup>, for comparison purposes. The dissimilarity measures were applied in identical environments. Apart from the dissimilarity measure, all other registration settings such as the choice of optimizer, number of resolutions, number of iterations or number of considered samples, were identical. Two resolutions of 1000 iterations were used for all six image datasets. To account for deformations caused by heart-pulsations and breathing, we used a B-spline transformation model for the CT-LUNG, T1MOLLI-HEART, T1VFA-CAROTID, ADC-ABDOMEN and DCE-ABDOMEN datasets. The registrations were performed for three distinct B-spline grid spacings: 32 mm, 64 mm and 128 mm for the T1MOLLI-HEART, ADC-ABDOMEN, DCE-ABDOMEN datasets, 8 mm, 16 mm and 32 mm for the T1VFA-CAROTID dataset, and 6 mm, 13 mm and 20 mm for the CT-LUNG dataset. All results are reported as supplementary material (Tables S1 to S6). Results for the intermediate values of the spacings (i.e. either 64 mm, 16 mm or 13 mm), are reported in the Results section of this article. To account for deformations

caused by head motion and eddy current distortions, we used an affine transformation model for the DTI-BRAIN dataset. When applying  $\mathcal{D}_{PCA}$ , the value of  $L$  was 1 for CT-LUNG, 3 for T1MOLLI-HEART, 1 for T1VFA-CAROTID, 4 for ADC-ABDOMEN, 7 for DTI-BRAIN, and 4 for DCE-ABDOMEN.

**Evaluation measures.** No ground truth alignment was available for any of the six datasets considered. However, registration performance was evaluated based on four different measures, described in Huizinga *et al.*<sup>11</sup>, and briefly described in this section.

The first two measures are based on landmark correspondence and overlap of volumes of interest. Landmarks were manually defined on images of the T1VFA-CAROTID and DCE-ABDOMEN datasets. The correspondence between the corresponding landmarks was evaluated by computing a mean target registration error (mTRE). In the T1MOLLI-HEART case, segmentations of the myocardium were outlined on between 6 and 9 images per patient. In the ADC-ABDOMEN case, the spleen was manually delineated on 8 images. For these two cases, the overlap between the segmented structures was then evaluated using a Dice coefficient. For the DTI-BRAIN study, neither landmarks nor structures could be reliably identified on the diffusion weighted images, which is why no overlap or point correspondence was calculated<sup>11</sup>.

The second measure quantifies the smoothness of the transformation obtained through registration. Extreme and non-smooth deformations are unexpected. The smoothness of the deformation field can therefore be used to identify such undesirable transformations. A smoothness quantification can be obtained by computing the standard deviation of the determinant of  $\partial T_g / \partial \mathbf{x}$  over all  $\mathbf{x}$  for all images:  $\text{STD}_{\det(\partial T_g / \partial \mathbf{x})}$ . Smoothness was quantified for all datasets except for DTI-BRAIN because an affine transformation was used in that last case. The smoother the transformation, the lower the quantity  $\text{STD}_{\det(\partial T_g / \partial \mathbf{x})}$ .

The last evaluation measure is an uncertainty estimation of the qMRI fit. For the five qMRI datasets, curve fittings were performed to generate quantitative images. The fitted values were evaluated in the myocardium for the T1MOLLI-HEART dataset ( $T_1$  values), in the carotid artery wall for the T1VFA-CAROTID dataset ( $T_1$  values), in the spleen for the ADC-ABDOMEN dataset (ADC values), in the brain parenchyma for the DTI-BRAIN dataset (MD values), and in the pancreas for the DCE-ABDOMEN dataset ( $K^{\text{trans}}$  values). The qMRI models were fitted using a maximum likelihood estimator that takes into account the Rician characteristic of the noise in MR data. We used the fitting same method as Huizinga *et al.*<sup>11</sup>, based on the work of Poot *et al.*<sup>43</sup>. The uncertainty of these fitted qMRI model parameters can be quantified by the 90<sup>th</sup> percentile of the square root of Cramér-Rao lower bound (CRLB), which provides a lower bound for the variance of the maximum likelihood parameters. This uncertainty estimate is denoted  $90^{\text{th}} \sqrt{\text{CRLB}}$ .

**Assessment of multivariate joint normality.** As mentioned in the Method section, the computation of the total correlation dissimilarity measure  $\mathcal{D}_{TC}$  that we propose is based on the approximation that the intensity distribution of the images to register is multivariate normal. For most datasets, however, the intensity distribution is expected not to be multivariate normal. The underlying idea is that the approximated dissimilarity measure will result in the same minimization result as if the approximation had not been done.

A second interest of the experimental setting is therefore to evaluate how multivariate normal the intensity distributions are for the six types of datasets that are registered in this study, and in the light of the registration accuracy results, to assess whether the approximation that we made can be considered as sensible on multi-parametric datasets.

The joint normality of two images can be assessed by computing and visualizing their joint histogram. Assessing joint normality on more images requires other methods. A possible graphical approach to analyze the multivariate joint normality of  $G$  images is to compare the distributions of observed Mahalanobis distances with the distribution of a chi-square distribution with  $G$  degrees of freedom  $\chi_G^2$ . A squared Mahalanobis distance  $d_i^2$  (with  $i = 1 \dots N$ ) can be computed at each voxel location  $M_g(i)$ , by:  $d_i^2 = (y_i - y_m)^T \mathbf{S}^{-1} (y_i - y_m)$ , with  $y_i = [M_1(i), \dots, M_G(i)]^T$ , the sample mean vector  $y_m = \sum_{i=1}^N y_i / N$ , and the sample covariance  $\mathbf{S} = \sum_{i=1}^N (y_i - y_m)(y_i - y_m)^T / (N - 1)$ . It has been shown that the sample squared Mahalanobis distance converges to  $\chi_G^2$  when  $y_i \sim \mathcal{N}_k(y_m, \mathbf{S})$ <sup>44</sup>. To graphically check whether the distribution of intensities of  $\mathbf{M}$  is joint normal, we will plot the cumulative distribution function (CDF) of  $d^2$  and  $\chi_G^2$  in the same graph. If the CDF of the squared Mahalanobis distances  $d^2$  approaches this of  $\chi_G^2$ , then we will consider the data as joint normal.

**Computational efficiency of total correlation  $\mathcal{D}_{TC}$ .** To study the computational efficiency of the proposed total correlation dissimilarity measure  $\mathcal{D}_{TC}$ , the average time per iteration is studied by varying three registration parameters: the number of images  $G$  that are simultaneously registered, the number of spatial samples taken to evaluate the groupwise dissimilarity measure, and the number of B-spline control points of the transformation model used to warp the images. The influence of these three parameters on the average time per iteration is studied by varying each of them while setting the two remaining ones at values in the range of those described in the Registration characteristics section:

- when the number of B-spline control points evolves, the number of images  $G$  is set to 50, and the number of spatial samples to 1024. The numbers of B-spline control points per image vary between 50 and 20000;
- when the number of images  $G$  evolves, the number of B-spline control points is set to 500 per image, and the number of spatial samples to 1024. The numbers of images  $G$  cover the characteristics of the images described in the 'Description of the six image datasets' section (i.e.  $G = 5 \dots 160$ );
- when the number of spatial samples evolves, the number of B-spline control points is set to 500 per image, and the number of images  $G$  is set to 50. We considered numbers of spatial samples between 16 samples and 8192.

## References

- Viola, P. & Wells, W. Alignment by maximization of mutual information. *Int J Comput. Vis.* **24**, 137–154 (1997).
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. & Suetens, P. Image registration by maximization of mutual information. *IEEE T Med Imaging* **16**, 187–198 (1997).
- Pluim, J., Maintz, J. & Viergever, M. Mutual information based registration of medical images: a survey. *IEEE T Med Imaging* **22**, 1–21 (2003).
- Geng, X., Christensen, G. E., Gu, H., Ross, T. J. & Yang, Y. Implicit reference-based group-wise image registration and its application to structural and functional MRI. *NeuroImage* **47**, 1341–1351 (2009).
- McGill, W. J. Multivariate information transmission. *Psychom.* **19**, 317–325 (1954).
- Watanabe, S. Information theoretical analysis of multivariate correlation. *IBM J Res Dev* **4**, 66–82 (1960).
- Han, T. S. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control.* **36**, 133–156 (1978).
- Guyader, J.-M. *et al.* Total correlation-based groupwise image registration for quantitative MRI. *In WBIR*, 186–193 (2016).
- Metz, C., Klein, S., Schaap, M., van Walsum, T. & Niessen, W. Nonrigid registration of dynamic medical imaging data using nD + t B-splines and a groupwise optimization approach. *Med Image Anal* **15**, 238–249 (2011).
- Bhatia, K. K., Hajnal, J., Hammers, A. & Rueckert, D. Similarity metrics for groupwise non-rigid registration. *In MICCAI* **2**, 544–552 (2007).
- Huizinga, W. *et al.* PCA-based groupwise image registration for quantitative MRI. *Med Image Anal* **29**, 65–78 (2016).
- Guyader, J.-M. *et al.* Influence of image registration on apparent diffusion coefficient images computed from free-breathing diffusion MR images of the abdomen. *J Magn Reson. Im* **42**, 315–330 (2015).
- Gao, Z. *et al.* Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. *Med Image Anal* **37**, 1–21 (2017).
- Gao, Z. *et al.* Motion tracking of the carotid artery wall from ultrasound image sequences: a nonlinear state-space approach. *IEEE T Med Imaging* **37**, 273–283 (2017).
- Jiao, J., Searle, G. E., Schnabel, J. A. & Gunn, R. N. Impact of image-based motion correction on dopamine D3/D2 receptor occupancy—comparison of groupwise and frame-by-frame registration approaches. *EJNMMI Phys.* **2**, 1–15, <https://doi.org/10.1186/s40658-015-0117-0> (2015).
- Korporaal, J. G. *et al.* Dynamic contrast-enhanced CT for prostate cancer: relationship between image noise, voxel size, and repeatability. *Radiol.* **256**, 976–984, <https://doi.org/10.1016/j.juro.2011.02.2679> (2010).
- Huizinga, W. *et al.* A spatio-temporal reference model of the aging brain. *NeuroImage* **169**, 11–22, <https://doi.org/10.1016/j.neuroimage.2017.10.040> (2018).
- Shannon, C. E. A mathematical theory of communication. *Bell Syst Tech J* **27**, 379–423 (1948).
- Cover, T. M. & Thomas, J. A. Elements of information theory (Hoboken, New Jersey, 2005).
- Seghers, D., Agostino, E. D., Maes, F. & Vandermeulen, D. Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. *In MICCAI*, 696–703 (Rennes, Brittany, 2004).
- Bell, A. J. Co-information lattice. *In 4th International Symposium on Independent Component Analysis and Blind Source Separation*, 921–926 (2003).
- Galas, D. J., Sakhanenko, N. A., Skupin, A. & Ignac, T. Describing the complexity of systems: multivariable set complexity and the information basis of systems biology. *J Comput. Biol* **21**, 118–140 (2014).
- Timme, N., Alford, W., Flecker, B. & Beggs, J. M. Synergy, redundancy, and multivariate information measures: an experimentalist's perspective. *J Comput. Neurosci* **36**, 119–140 (2014).
- Venn, J. On the diagrammatic and mechanical representation of propositions and reasonings. *Philos Mag Ser.* **5**(9), 1–18 (1880).
- Bellman, R. *Adaptive control processes: a guided tour* (Princeton University Press, New Jersey, 1961).
- Ahmed, N. A. & Gokhale, D. V. Entropy expressions and their estimators for multivariate distributions. *IEEE Trans Inf Theory* **35**, 688–692 (1989).
- Rueckert, D. *et al.* Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Trans Med Imag* **18**, 712–721 (1999).
- Balci, S., Golland, P., Shenton, M. & Wells, M. Free-form B-spline deformation model for groupwise registration. *In MICCAI*, 23–30 (Statistical Registration Workshop) (2007).
- Klein, S., Staring, M., Murphy, K., Viergever, M. & Pluim, J. Elastix: a toolbox for intensity-based medical image registration. *IEEE T Med Imaging* **29**, 196–205 (2010).
- Klein, S., Pluim, J., Staring, M. & Viergever, M. Adaptive stochastic gradient descent optimisation for image registration. *Int J Comput. Vis.* **81**, 227–239 (2009).
- van der Aa, N., Ter Morsche, H. G. & Mattheij, R. R. M. Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem. *Electron J Linear Al* **16**, 300–314 (2007).
- Le Bihan, D. *et al.* MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiol.* **161**, 401–407 (1986).
- Wachinger, C. & Navab, N. Simultaneous registration of multiple images: similarity metrics and efficient optimization. *IEEE Trans Pattern Anal* **35**, 1–14 (2012).
- Castillo, R. *et al.* A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol* **54**, 1849–1870 (2009).
- Uitterdijk, A. *et al.* VEGF165A microsphere therapy for myocardial infarction suppresses acute cytokine release and increases capillary density, but does not improve cardiac function. *Am J Physiol-Heart C* **309**, 396–406 (2015).
- Coolen, B. F. *et al.* Three-dimensional quantitative T1 and T2 mapping of the carotid artery: sequence design and *in vivo* feasibility. *Magn Reson. Med* **75**, 1–10 (2015).
- Leemans, A., Sijbers, J., De Backer, S., Vandervliet, E. & Parizel, P. Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. *Magn Reson. Med* **55**, 1414–1423 (2006).
- De Geeter, N., Crevecoeur, G., Dupré, L., Van Hecke, W. & Leemans, A. A DTI-based model for TMS using the independent impedance method with frequency-dependent tissue parameters. *Phys Med Biol* **57**, 2169–2188 (2012).
- Wang, H.-C., Hsu, J.-L. & Leemans, A. Diffusion tensor imaging of vascular parkinsonism: structural changes in cerebral white matter and the association with clinical severity. *Arch Neurol* **69**, 1340–1348 (2012).
- van der Aa, N. *et al.* Does diffusion tensor imaging-based tractography at 3 months of age contribute to the prediction of motor outcome after perinatal arterial ischemic stroke? *Stroke* **42**, 3410–3414 (2011).
- Reijmer, Y. D. *et al.* Improved sensitivity to cerebral white matter abnormalities in Alzheimer's disease with spherical deconvolution based tractography. *PLoS One* **7**, 1–8 (e44074) (2012).
- Klaassen, R. *et al.* Motion correction of high temporal 3T dynamic contrast enhanced MRI of pancreatic cancer - preliminary results. *In ISMRM*, 3667 (Poster) (2014).
- Poot, D. H. J. & Klein, S. Detecting statistically significant differences in quantitative MRI experiments, applied to diffusion tensor imaging. *IEEE T Med Imaging* **34**, 1164–1176 (2015).
- Timm, N. H. *Applied multivariate analysis* (Springer, New-York, 2002).



## Acknowledgements

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking (<http://www.imi.europa.eu>) under grant agreement nr. 115151 (QuIC-ConCePT project), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) (EU-FP7) and EFPIA companies in kind contribution. Funding was also provided by EU-FP7 under grant agreement no. 601055, VPH-DARE@IT. The research of M. van Kranenburg was generously supported by stichting DIRA - Moerman *et al.* The authors would also like to thank: DIR-lab and Richard Castillo for providing the CT-LUNG data; H.M.M. van Beusekom and R.J.M. van Geuns for providing the T1MOLLI-HEART data. The acquisition of the T1MOLLI-HEART data was financially supported by Agentschap NL (SENER-NOVEM): "A novel approach to myocardial regeneration" to H.M.M. van Beusekom *et al.* under grant nr. ISO43050.; B.F. Coolen and A.J. Nederveen for providing the T1VFA-CAROTID data; N.M. deSouza, L. Bernardin and N. Douglas, Institute of Cancer Research, London, UK, for providing the ADC-ABDOMEN data. The ADC-ABDOMEN data were acquired in the context of the QuIC-ConCePT project. A. Leemans for providing the DTI-BRAIN data; R. Klaassen, A.J. Nederveen and H.W.M. van Laarhoven for providing the DCE-ABDOMEN data. The acquisition of the data was financially supported by the Dutch Cancer Society under grant no. UVA 2013-5932 to H.W.M. van Laarhoven.

## Author Contributions

All authors designed the experiments and reviewed the manuscript. J.-M.G., W.H. and D.H.J.P. performed the experiments. J.-M.G. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-31474-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018