



HAL
open science

3D nonrigid medical image registration using a new information theoretic measure.

Bicao Li, Guanyu Yang, Jean Louis Coatrieux, Baosheng Li, Huazhong Shu

► **To cite this version:**

Bicao Li, Guanyu Yang, Jean Louis Coatrieux, Baosheng Li, Huazhong Shu. 3D nonrigid medical image registration using a new information theoretic measure.. *Physics in Medicine and Biology*, 2015, 60 (22), pp.8767-90. 10.1088/0031-9155/60/22/8767 . hal-01253712

HAL Id: hal-01253712

<https://hal.science/hal-01253712>

Submitted on 11 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Three-dimensional Nonrigid Medical Image Registration Using a New Information Theoretic Measure

Bicao Li^{1, 2, 3}, Guanyu Yang^{1, 2, 3}, Jean Louis Coatrieux^{1, 4, 5, 6}, Baosheng Li^{1, 7} and Huazhong Shu^{1, 2, 3}

¹Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, 210096 Nanjing, China

²Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, 210096 Nanjing, China

³Centre de Recherche en Information Médicale Sino-français (CRIBs), Nanjing, 210096, China

⁴INSERM, U1099, Rennes, F-35000, France

⁵LTSI, Université de Rennes 1, Rennes, F-35042, France

⁶Centre de Recherche en Information Médicale Sino-français (CRIBs), Rennes, F-35000, France

⁷Department of Radiation Oncology, Shandong Cancer Hospital, 250117 Jinan, China

E-mail: shu.list@seu.edu.cn

Abstract

This work presents a novel method for nonrigid registration of medical images based on the Arimoto entropy, a generalization of the Shannon entropy. The proposed method employed the Jensen-Arimoto (JA) divergence measure as a similarity metric to measure the statistical dependence between the medical images. Free-form deformations were adopted as the transformation model and the Parzen window estimation was applied to compute the probability distributions. A penalty term is incorporated into the objective function to smooth the nonrigid transformation. The goal of registration is to optimize an objective function consisting of a dissimilarity term and a penalty term, which would be minimal when two deformed images are perfectly aligned using the limited memory BFGS optimization method, and thus to get the optimal geometric transformation. To validate the performance of the proposed method, experiments on both simulated 3D brain MR images and real 3D thoracic CT data sets were designed and performed on the open source *elastix* package. For the simulated experiments, the registration errors of 3D brain MR images with various magnitudes of known deformations and different levels of noise were measured. For the real data tests, four data sets of 4D thoracic CT from four patients were selected to assess the registration performance of the method, including ten 3D CT images for each 4D CT data covering an entire respiration cycle. These results were compared with the normalized cross correlation and the mutual information methods and show a slight but true improvement in registration accuracy.

Keywords: Arimoto entropy, divergence measure, free-form deformations, nonrigid registration.

1. Introduction

Image registration is a very common and useful technique in numerous fields including medical imaging, computer vision, remote sensing, etc. It plays a major role in medicine for multimodal diagnosis and computer-aided surgery (Maintz *et al* 1998, Hill *et al* 2001). Due to the motion of tissue and organ of the patients, nonrigid deformations must be taken into account. This problem can be addressed by using nonrigid geometric registration methods aimed at finding the optimal elastic transformation between the two images to be aligned. Feature-based and intensity-based alignment methods (Khader *et al* 2011) are classically considered for such purpose.

For the feature-based registration methods, their accuracy largely depends on the choice and the extraction of features, which usually include salient image elements such as corner points, edges, ridges, surfaces (Maintz *et al* 1996), and others. Let us mention, among many other schemes, SUSAN (Smith *et al* 1997) based on the minimization of a local image region and SIFT (Lowe 2004) that are invariant to scale and orientation. Szeliski and Coughlan introduced a fast registration algorithm for 3D anatomical surfaces (Szeliski and Coughlan 1997). In their approach, a distance criterion was minimized to search the optimal mapping between two surfaces. Another feature-based approach consists of using the moment invariants. Flusser and Suk described a blurring geometric moment invariant (Flusser and Suk 1998) and deduced a combined blurring and rotation invariant of geometric moment which were applied to feature extraction in image registration. On the basis of Flusser's work, Bentoutou *et al* (2002) utilized the combined invariants for image registration in digital subtraction angiography (Bentoutou *et al* 2002). A robust discriminative clustering method based on mutual information of supervoxels is presented to extract the features of brain MRI (Kong *et al* 2015). However, the registration accuracy of feature-based approaches is directly determined by the feature extraction quality and completeness. This preprocessing step is not trivial in particular in medical imaging where low contrast and noise are observed.

Another kind of registration methods accounts for intensity-based approaches, in which the spatial mapping of the images to be registered relies on the image intensity values, without any feature extraction or prior segmentation. Hoh *et al* (1993) used the sum of absolute differences (SAD) and the stochastic sign change (SSC) similarity measures for rigid registration of cardiac PET images (Hoh *et al* 1993). Similarly, the sum of squared differences (SSD) was also applied to intra-modality image registration. They all assume that the corresponding intensity values of the two images to be aligned are similar. To improve the robustness of image registration, the correlation coefficient was proposed by Lemieux *et al* (1994). However, it was pointed out that the correlation coefficient is sensitive to outliers (Huber 1981). Even a few outliers may greatly degrade the registration quality as it was shown later (Kim and Fessler 2002). To overcome this problem, a robust correlation coefficient (Kim and Fessler 2004), serving as registration metrics, was described. However, the correlation methods, if of interest for monomodal registration, are not suitable for the registration of multisensor (i.e. multimodal) images due to their totally different physics used (Maurer *et al* 1993).

Another popular framework for intensity-based registration methods is based on the concept of information theory. The maximization of mutual information (MI) for multimodal image registration was independently introduced by Collignon *et al* (1995), Wells *et al* (1996), Maes *et al* (1997), Viola and Wells (1997). Collignon *et al* (1995) applied the MI registration algorithm to human head images and achieved a sub-voxel registration accuracy. In the paper by Maes *et al* (1997), a partial volume interpolation was proposed to estimate the marginal and joint image intensity distributions, and so, smoothing the objective function and

reducing the number of local maxima. Sub-voxel accuracy, high robustness and good immunity to noise and geometric distortion were shown. A new way to compute the image entropy and mutual information was reported by Viola and Wells (1997). The MI based methods do not require any preprocessing and they have found a large variety of applications. However, in some applications, mutual information may fail in registering images, especially when the image misalignment is due to a change in the field of view (FOV). To address this issue, a method called normalized mutual information (NMI) (Studholme *et al* 1999), robust to a partial overlap of the registered images, was proposed. NMI was defined as a ratio of the sum of the marginal entropies and the joint entropy. This approach was validated on MRI-CT images with different FOVs. Pluim *et al* (2000) presented a new information measure by combining the MI with gradient information. The matching results of MR-T1 and PET images indicated that this new method yielded a better registration than MI and NMI. A generalized method using the so-called f-information measure was later reported by the same group (Pluim *et al* 2001) and further extended to a series of other similarity measures, including V-information, I_α -information, M_α -information, χ_α -information and Rényi measure (Pluim *et al* 2004). Another registration criterion, based on cumulative probability distributions, called cumulative residual entropy (CRE) (Wang *et al* 2003) was proposed whose theoretical relations (Rao *et al* 2004) with Shannon entropy were further analyzed and compared with MI-based methods (Wang *et al* 2007). The CRE measure, being based on the cumulative distribution functions (CDF) rather than the probability distribution functions (PDF), was shown more robust to noise.

An alternative information-based approach was introduced by Studholme *et al* (2006). They defined a nonrigid viscous fluid registration scheme, which uses a new similarity measure calculated over a set of overlapping sub-regions of the image. A third channel, representing a spatial label to extend the intensity joint histogram, was estimated by B-spline Parzen windows (Thévenaz *et al* 2000). A similarity measure called conditional mutual information (cMI) (Loeckx *et al* 2010) was presented. Similarly, cMI was calculated by extending the joint histogram to a third dimension, representing the spatial distribution of the joint intensities. The Shannon entropy possesses the additivity property, which means that the joint entropy of a pair of independent random variables is the sum of the individual entropies. However, as correctly indicated by Antolin *et al* (2010), this property does not take the correlation between the variables into account. To overcome this problem, they introduced the pseudo-additive Tsallis entropy.

Inspired by Antolin's work, we have presented a novel divergence measure called the Jensen-Arimoto (JA) divergence as the registration metric for 2D-2D rigid medical registration (Li *et al* 2014). It was shown that the classic mutual information is a particular case of JA divergence as $\alpha = 1$, where α is a parameter appeared in JA divergence measuring the departure degree from pseudo-additivity. In this paper, we further investigate the concavity of Arimoto entropy and boundedness property of JA divergence and apply it for 3D nonrigid image registration. We introduce a 3D nonrigid registration technique with JA as the similar measure term and bending energy selected as the regularization term. Free-form deformations (FFDs) are employed as the transformation model and the limited memory Broyden-Fletcher-Goldfarb

-Shanno (L-BFGS) optimization method is used to optimize the JA divergence. In our optimization scheme, we estimate the continuous probability distributions using the B-splines Parzen window, so that the analytical gradient of JA divergence can be obtained.

The rest of this paper is organized as follows. In section 2, we introduce the new divergence measure based on Arimoto entropy. The subsequent nonrigid registration method

is detailed in section 3. Section 4 provides our experimental results on simulated 3D MR brain images and real 3D thoracic CT data with a comparison to the standard nonrigid methods based on the mutual information. Concluding remarks and perspectives are sketched in section 5.

2. Theory

In this section, the Arimoto entropy is introduced and its properties are examined. Then, a new divergence called the Jensen-Arimoto divergence is presented and some basic properties are derived.

2.1. Arimoto Entropy

Given a random variable X , the Shannon entropy of X is a measure of the amount of uncertainty included in the random variable (Cover *et al* 2006), $H(X) = -\sum_{x \in X} p(x) \log p(x)$. Consider another random variable Y , the conditional entropy of X given Y is defined as the amount of uncertainty left in X when knowing Y (Maes *et al* 1997). The mutual information is related to the reduction of the entropy of X by the conditional entropy of X given Y , and accounts for the degree of disparity between the two random variables. That is, when two random variables are independent from each other, the entropy of X is equal to the conditional entropy of X given Y so that the value of MI equals to zero, while MI achieves its maximum if Y is completely dependent on X . MI has been widely used in statistical applications.

The Arimoto entropy (Arimoto 1971), a generalization of Shannon entropy, was introduced by Arimoto and further developed by other researchers (Boekee *et al* 1980, Liese *et al* 2006). Its definition is given by

$$A_\alpha(X) = \frac{\alpha}{\alpha-1} \left[1 - \left(\sum_{i=1}^M p_i^\alpha \right)^{\frac{1}{\alpha}} \right] \quad \alpha > 0, \alpha \neq 1 \quad (1)$$

where $P = (p_1, p_2, \dots, p_M)$ denotes the M probability distributions on X .

Some significant properties of the Arimoto entropy were deduced (Boekee *et al* 1980). Here, we only consider three of them.

Non-negativity:

$$A_\alpha(X) \geq 0 \quad \alpha > 0, \alpha \neq 1 \quad (2)$$

Pseudo-additivity:

$$A_\alpha(X, Y) = A_\alpha(X) + A_\alpha(Y) - \frac{\alpha-1}{\alpha} A_\alpha(X) A_\alpha(Y) \quad (3)$$

Concavity:

$$A_\alpha(tX_1 + (1-t)X_2) \geq tA_\alpha(X_1) + (1-t)A_\alpha(X_2) \quad t \in (0, 1) \quad (4)$$

The detailed proof of these properties can be found in Liese *et al* (2006). In (3), X and Y are two independent random variables, $A_\alpha(X, Y)$ is the joint Arimoto entropy between X and Y , for $\alpha \in (0, 1) \cup (1, +\infty)$. Unlike the pseudo-additivity of the Arimoto entropy, the joint Shannon entropy has the property of additivity when two random variables are independent, i.e., $H(X, Y) = H(X) + H(Y)$. Pseudo-additivity means that the Arimoto entropy is a nonextensive entropy and the nonextensive property signifies that the correlations between X and Y are considered. To demonstrate the utility of this property in image registration problem, an illustration will be given in section 4.2.

As mentioned in section 1, the parameter α in (3) weights the degree of nonextensivity. The nonextensivity gradually becomes weak as the parameter is tending to 1, and the Arimoto

entropy also approximates to Shannon entropy that accounts for the complete extensive entropy. Figure 1 illustrates the Arimoto entropy of a Bernoulli distribution, $P = (p, 1-p)$, for various α values. It was shown that Arimoto entropy is concave if $\alpha \in (0, 1)$ or $\alpha \in (1, +\infty)$. Moreover, the limit of the Arimoto entropy is equal to the Shannon entropy when $\alpha \rightarrow 1$. The curve for $\alpha = 1$ in figure 1 represents the Shannon entropy, computed using the natural logarithm. As shown in figure 1, the measure of uncertainty by Arimoto entropy increases as α decreases and is not less than Shannon entropy for $0 < \alpha < 1$, and decreases as α increases and is equal or less than Shannon entropy for $\alpha > 1$.

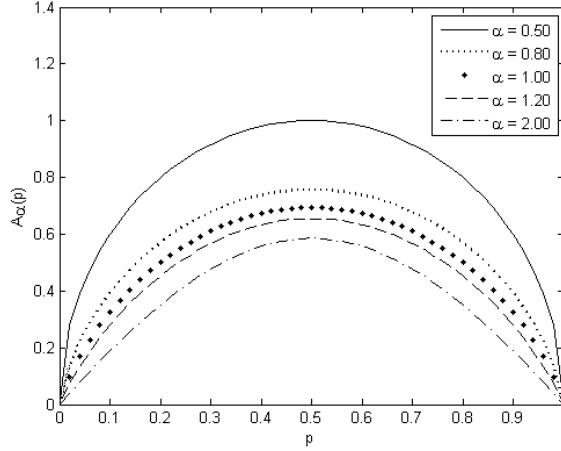


Figure 1. Arimoto entropy $A_\alpha(\mathbf{p})$ of a Bernoulli distribution $\mathbf{p}=(p, 1-p)$ for various α values.

2.2. The Jensen-Arimoto Divergence

Following the derivation reported in Lin *et al* (1991), we define a new divergence measure based on the Arimoto entropy called the Jensen-Arimoto (JA) divergence and study its properties in the sequel. The Jensen-Shannon (JS) divergence was used to measure the distance of two random variables and, as we will see, the JA divergence has the same nature as JS divergence.

Definition 1: Let $X(x_1, x_2, \dots, x_M)$ be a random variable, and $P(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M)$ be M probability distributions on X . The Jensen-Arimoto divergence is defined as

$$JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) = A_\alpha\left(\sum_{i=1}^M \omega_i \mathbf{p}_i\right) - \sum_{i=1}^M \omega_i A_\alpha(\mathbf{p}_i) \quad (5)$$

where $A_\alpha(\cdot)$ is the Arimoto entropy for $\alpha > 0$, $\alpha \neq 1$ and ω_i are weight factors such that $\sum_{i=1}^M \omega_i = 1$ and $\omega_i \geq 0$. When the nonextensive parameter α tends to 1, Arimoto entropy converges to the standard Shannon entropy and the limit of JA divergence in (5) is identical to the traditional mutual information using L'Hopital's rule. In the following, we study the properties of the JA divergence.

Proposition 1: The JA divergence is nonnegative, symmetric and equal to zero if and only if all probability distributions are identical to each other for $\alpha \in (0, 1) \cup (1, +\infty)$.

Proof: For a real concave function φ , with numbers x_1, x_2, \dots, x_M in its domain, and positive weights ω_i , the Jensen inequality can be stated as

$$\varphi\left(\sum \omega_i x_i\right) \geq \sum \omega_i \varphi(x_i). \quad (6)$$

Due to the concavity of Arimoto entropy, we replace the function φ in (6) by the Arimoto entropy function and we can easily verify the non-negativity of JA divergence over the interval $\alpha \in (0, 1) \cup (1, +\infty)$. In addition, the JA value will not be changed when we exchange any two probabilities of equation (5), which leads to the symmetry of JA divergence. And it vanishes (with the identity axiom as a distance) if and only if all probability distributions are equal to each other. Here, we can set every \mathbf{p}_i equal to $1/M$, and substitute them into (5), we obtain $JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) = 0$.

Four requirements must be fulfilled for a distance (Li *et al* 2004): non-negativity, identity axiom, symmetry axiom and triangle inequality. The JA divergence is not a true distance since it does not satisfy the triangle inequality. Nevertheless, this does not degrade the performance of JA as a measure of the disparity among probability distributions. JA has similar properties as the mutual information, which measures the amount of information that one random variable includes in another random variable. Thus, we can use the JA divergence to evaluate the similarity between the two images to be registered.

Proposition 2: For $\alpha \in (1, 2]$, the JA divergence is a convex function of $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$.

A detailed proof can be found in Li *et al* (2014). In the optimization scheme of JA divergence, its derivative can be analytically calculated. To ensure the correct image alignment during the optimization and to guarantee the feasibility of the optimization method, α is restricted to the interval $(1, 2]$ in the rest of this paper.

Proposition 3: The maximum of JA divergence is obtained when $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$ are degenerated distributions, that is, $\mathbf{p}_i = \delta_{ij}$, δ_{ij} denoting the Kronecker symbol, and this maximum equals to $A_\alpha(\omega)$.

Since JA divergence is a convex function of $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M$, The proof can be easily verified (He *et al* 2003). Combining propositions 1 and 3, we can see that the JA divergence is bounded, $0 \leq JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) \leq A_\alpha(\omega)$.

3. Description of the proposed method

We detail here our nonrigid registration method. Firstly, we need to choose an appropriate transformation model. Then, in this transformation space, the JA divergence measure used as the registration criterion is derived together with the L-BFGS optimization method.

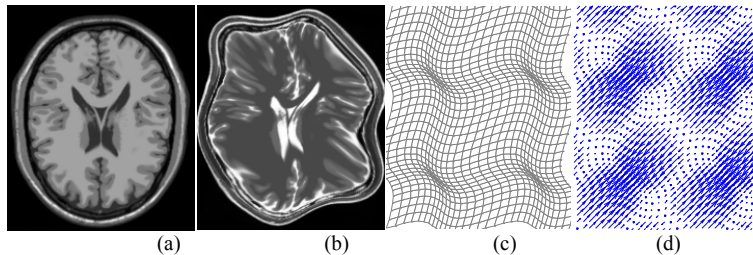


Figure 2. Two corresponding slices in MR T1 and MR T2 volumes and the deformation between them. (a) The reference image; (b) the deformed float image; (c) a known smooth deformation field to generate (b); (d) the corresponding deformation vector.

3.1. Registration Framework

Given two misaligned images to be registered, one is selected as the float image F , and the other considered as the reference image R . As an example, figure 2 depicts the corresponding slices in two volumes to be registered: there are, from left to right, the T1-weighted MR and

T2-weighted MR images, the true deformation field and the vector field between the two images. The image registration process consists to search the optimal transformation between F and R . Let x and x' be an arbitrary point in R and its corresponding point in F , respectively. The spatial mapping function between the corresponding points in R and F can be formulated by

$$x' = T_\mu(x), \quad (7)$$

where T_μ is the transformation function (the global function for rigid or affine registration and the local function in case of deformation) with parameters μ .

For the 3D case, $x = (x, y, z)^T$ and $x' = (x', y', z')^T$, the registration of F to R can be simply described as an optimization problem

$$\begin{aligned} T^* &= \arg \min_{\mu} D(F \circ T_\mu, R) \\ &= \arg \min_{\mu} D(F(T_\mu(x)), R(x)), \end{aligned} \quad (8)$$

where D denotes a suitable dissimilarity measure that achieves its minimum when $R(x)$ and $F(T_\mu(x))$ are completely aligned. The symbol “ \circ ” stands for an operation such that the float image is transformed by the space mapping T_μ .

However, the nonrigid deformation should be generally characterized by a smooth transformation. For that, a penalty term is introduced for regularizing the local deformation. Taking the regularization term into account, the objective function E is written as

$$E = D(F(T_\mu(x)), R(x)) + \lambda S(T_\mu(x)), \quad (9)$$

where λ is the weight parameter balancing the tradeoff between D and the smoothness term S . To align the deformed float image $F(T_\mu(x))$ with the reference image $R(x)$, the optimal parameter set μ minimizing the objective function E must be searched. So, in our registration framework, three crucial issues are concerned: the transformation model, the registration metrics (i.e. objective function) and the optimization scheme. In the next sections, these factors are examined in detail.

3.2. Transformation Model

Several popular transformation models including rigid, affine, perspective and curve (elastic) mapping models have been reported (Pluim *et al* 2003). In this paper, we address a nonrigid transformation dealing with organ or tissue motion. The local deformations though parameterized transformations can be preferably described by the free form deformation (FFD) model. In addition, a number of desirable properties of B-splines for modeling the deformation, including the inherent control of smoothness and separability in the multidimensional case, have been introduced (Wang *et al* 2007). Therefore, to simulate this elastic deformation, the FFD (Mattes *et al* 2003) based on cubic B-splines as the transformation model, is employed. For 3D images, let Φ be a $n_x \times n_y \times n_z$ mesh along the x , y and z directions of control points $[\omega_i, \omega_j, \omega_k]^T$ with a uniform spacing δ . Note that to improve the registration efficiency and save calculation time, a rigid or affine registration is often applied as a rough initial step before implementing nonrigid registration (Klein *et al* 2010), especially in the presence of a large global displacement (translation or rotation) between two images to be registered. In our work, we have chosen two phases of one respiration cycle as two registered images, and the large global displacement between them does not exist. So a rigid registration is not necessary prior to the nonrigid registration.

The mesh points serve as parameters of the free form transformation based on B-splines and the choice of mesh points determine the amount of nonrigid deformation. Usually, a large

spacing of mesh points allows for global nonrigid deformation, while a small spacing accounts for local nonrigid deformation. Therefore, we need to choose the mesh point set according to the degree of nonrigid deformation between the two test images.

Then, the 3D transformation at point $\mathbf{x} = [x, y, z]^T$ can be stated using a linear combination of cubic B-spline kernels as

$$T_\mu(\mathbf{x}) = \sum_{ijk} \mu_{ijk} \beta^{(3)}\left(\frac{x-\omega_i}{\delta}\right) \beta^{(3)}\left(\frac{y-\omega_j}{\delta}\right) \beta^{(3)}\left(\frac{z-\omega_k}{\delta}\right), \quad (10)$$

where μ_{ijk} denotes the vector of deformation coefficients related to the mesh and the control points (for its computation, the reader can refer to (Mattes *et al* 2003)). In the following, we adopt the transformation model defined by (10).

3.3. Registration Criteria

In the image registration process, the value of the similarity measure for two images is expected to be maximum when they are perfectly aligned. However, as stated in section 3.1, our purpose is searching for the minimum of a dissimilarity measure. Therefore, in accordance with section 3.1, a negative sign is assigned to the JA divergence measure in (5) and a new dissimilarity measure D is produced.

$$D(F(T_\mu(\mathbf{x})), R(\mathbf{x})) = -JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) \quad (11)$$

Combining (11) and (9), the optimal spatial mapping between $F(T_\mu(\mathbf{x}))$ and $R(\mathbf{x})$ can be formulated as

$$\mu^* = \arg \min_{\mu} E(F(T_\mu(\mathbf{x})), R(\mathbf{x})) = \arg \min_{\mu} \{-JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) + \lambda S(T_\mu(\mathbf{x}))\} \quad (12)$$

where μ^* is the estimated parameters set, $\mathbf{p}_i = p_i(F(T_\mu(\mathbf{x}))|R(\mathbf{x}))$, $1 \leq i \leq M$, represents the conditional probability distribution of the transformed float image given the reference image.

Let $f = (f_1, f_2, \dots, f_M)$ and $r = (r_1, r_2, \dots, r_M)$ be two sets of the intensity values in $F(T_\mu(\mathbf{x}))$ and $R(\mathbf{x})$ respectively and M the number of the chosen intensity bins to estimate the histogram. Then, the conditional probability can be rewritten as $\mathbf{p}_i = p(F=f_j|R=r_i) = p(f_j|r_i)$. The dissimilarity measure D shown in (11) represents the opposite number of the JA divergence with different weights and α values. In (5), we use the marginal probability distribution of the reference image as the weight (a prior), that is, $\omega_i = p(R(\mathbf{x})) = p(R=r_i) = p(r_i)$. Substituting ω_i and \mathbf{p}_i into (11), the new dissimilarity measure D used in our method is given by

$$\begin{aligned} D(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) &= -JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M) \\ &= \frac{\alpha}{\alpha-1} \left\{ \left[\sum_{j=1}^M \left[\sum_{i=1}^M p(r_i) p(f_j | r_i) \right]^\alpha \right]^{\frac{1}{\alpha}} - \sum_{i=1}^M p(r_i) \left[\sum_{j=1}^M p(f_j | r_i)^\alpha \right]^{\frac{1}{\alpha}} \right\} \quad (13) \\ &= \frac{\alpha}{\alpha-1} \left\{ \left[\sum_{j=1}^M p(f_j)^\alpha \right]^{\frac{1}{\alpha}} - \sum_{i=1}^M \left[\sum_{j=1}^M p(r_i, f_j)^\alpha \right]^{\frac{1}{\alpha}} \right\} \end{aligned}$$

$S(T_\mu(\mathbf{x}))$ in (12) is the penalty term aimed at smoothing the transformation and its expression is

$$\begin{aligned}
S(T_\mu(x)) = & \frac{1}{V} \int_0^X \int_0^Y \int_0^Z \left[\left(\frac{\partial^2 T}{\partial x^2} \right)^2 + \left(\frac{\partial^2 T}{\partial y^2} \right)^2 + \left(\frac{\partial^2 T}{\partial z^2} \right)^2 \right. \\
& \left. + 2 \left(\frac{\partial^2 T}{\partial x \partial y} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial x \partial z} \right)^2 + 2 \left(\frac{\partial^2 T}{\partial y \partial z} \right)^2 \right] dx dy dz,
\end{aligned} \tag{14}$$

where V denotes the volume of image domain in which the deformation T defined in (10) is estimated. X , Y and Z denote the dimensions of the image domain along the three directions, respectively. Generally, the histogram method is used to calculate the marginal probability $p(r_i)$ and the conditional probability $p(f_j|r_i)$ involved in (13). Then, the new registration criterion defined in (12) is minimized using an optimization method.

3.4. Optimization

The search algorithm mainly influences the accuracy and calculation time of a registration method. Hence, selecting an appropriate optimization method plays a crucial role in a registration framework. Compared to the gradient descent methods, the use of second-order information in Newton Raphson algorithm can provide better theoretical convergence properties (Klein *et al* 2007). However, the computation of the Hessian matrix and its inverse is time-consuming. As for the quasi-Newton methods, the inverse of Hessian matrix is approximated and the second order derivatives of the objective function are not calculated. In this paper, we adopt the L-BFGS (Nocedal *et al* 1980), a popular variant of the quasi-Newton approach, to minimize the objective function shown in (12). A second-order Taylor approximation (Press *et al* 2007) of the objective function E at μ is derived as follows

$$E(\mu + \Delta\mu) \approx E(\mu) + \Delta\mu^T \cdot \nabla E(\mu) + \frac{1}{2} \Delta\mu^T \cdot \nabla^2 E(\mu) \cdot \Delta\mu, \tag{15}$$

where $E(\mu)$ denotes the objective function in (12) for simplicity, $\Delta\mu$ represents the increment of the parameter vector μ and superscript T is the transpose operator. The second term on the right hand side of (15) is the directional derivative of E at μ in the direction $\Delta\mu$. For the L-BFGS optimization algorithm, the parameter vector μ is updated at iteration number $k+1$ using the previous iteration number k as

$$\mu^{(k+1)} = \mu^{(k)} - (H^{(k)})^{-1} \cdot \nabla E(\mu^{(k)}), \tag{16}$$

where the superscript k stands for the iteration number, $(H^{(k)})^{-1}$ is the inverse of Hessian matrix for every iteration, and ∇ denotes the gradient operator. Here, note that the elements of $(H^{(k)})^{-1}$ are computed using an approximation instead of the actual Hessian matrix at each iteration. In the sequel, we need to compute the derivative of the objective function E with respect to μ . An important prerequisite of this method is the calculation of gradients of the objective function. The derivative of E is given by

$$\frac{\partial E}{\partial \mu} = \left[\frac{\partial E}{\partial \mu_1}, \frac{\partial E}{\partial \mu_2}, \dots, \frac{\partial E}{\partial \mu_n} \right], \tag{17}$$

where n denotes the number of parameters (degree of freedom, DoF) in the transformation model. In our scheme, for the non-rigid registration, the amount of DoFs depends on the size of the chosen mesh in FFD model. For example, if the mesh is selected with the size $10 \times 10 \times 10$, the number of DoFs can reach about 3000 for 3D images as reported by Rueckert *et al* (1999). The derivative of the penalty term S in the objective function has been already reported in Rohlfing *et al* (2003). Therefore, we only provide in this work the derivative calculation of the dissimilarity measure.

Note that the L-BFGS optimization scheme is stopped if either the difference of objective

function value between two consecutive iterations is less than 0.01 or the number of iterations reaches a prefixed value NMAX. In our work, NMAX was set to 100.

3.4.1. Probability estimation

To avoid the discretization errors and to carry out a robust optimization scheme, the analytical gradient of the new dissimilarity measure D needs to be calculated. As the form of the probability distributions in (13) is not continuous and differentiable, we adopted a Parzen-window to estimate the continuous probability distribution function from the image data. Many kernel functions can be used as far as they meet the requirements to be smooth, symmetric, to have zero mean and integrate to one (Khader *et al* 2011). Some popular kernel functions involve Gaussian (Tustison *et al* 2011) and B-spline functions (Thévenaz *et al* 2000). In this study, the B-spline functions are used to estimate the marginal and conditional probabilities. Let $\beta^{(0)}$ be the zero-order B-Spline function and $\beta^{(3)}$ the cubic B-Spline function. Due to the independence of the reference image to the transformation parameters μ , a zero-order B-spline was used and the marginal probability distributions of $R(x)$ is expressed by

$$\beta(r_i) = \frac{1}{V} \sum_{x \in \Omega} \beta^{(0)} \left(r_i - \frac{R(x) - R^0}{\Delta b_R} \right) \quad (18)$$

For the transformed float image, the cubic B-spline function was used to estimate the probability distribution as previously proposed (Wang *et al* 2005, Khader *et al* 2011). In this case, the estimated probability of $F(T_\mu(x))$ is given by

$$\beta(f_j) = \frac{1}{V} \sum_{x \in \Omega} \beta^{(3)} \left(f_j - \frac{F(T_\mu(x)) - F^0}{\Delta b_F} \right) \quad (19)$$

and the joint probability of $F(T_\mu(x))$ and $R(x)$ by

$$\beta(r_i, f_j; \mu) = \frac{1}{V} \sum_{x \in \Omega} \beta^{(0)} \left(r_i - \frac{R(x) - R^0}{\Delta b_R} \right) \beta^{(3)} \left(f_j - \frac{F(T_\mu(x)) - F^0}{\Delta b_F} \right) \quad (20)$$

where Ω denotes the image domain in which the marginal and joint probability density functions were estimated. V is the total number of voxels in Ω . In (18), (19) and (20), intensity values of the two images are normalized by the minimal intensity values, R^0 , F^0 and the range of the chosen bins width Δb_R , Δb_F of the reference and float image, respectively.

3.4.2. Derivative of the proposed dissimilarity measure

The derivatives of the marginal probability $\beta(f_j)$ and of the joint one $\beta(r_i, f_j; \mu)$ with respect to μ are

$$\frac{\partial \beta(f_j)}{\partial \mu} = -\frac{1}{V \cdot \Delta b_F} \sum_{x \in \Omega} \beta^{(3)} \left(f_j - \frac{F(T_\mu(x)) - F^0}{\Delta b_F} \right) \left(\frac{\partial F(t)}{\partial t} \Big|_{t=T_\mu(x)} \right) \cdot \frac{\partial (T_\mu(x))}{\partial \mu} \quad (21)$$

and

$$\begin{aligned} \frac{\partial \beta(r_i, f_j; \mu)}{\partial \mu} = & -\frac{1}{V \cdot \Delta b_F} \sum_{x \in \Omega} \beta^{(0)} \left(r_i - \frac{R(x) - R^0}{\Delta b_R} \right) \\ & \cdot \beta^{(3)} \left(f_j - \frac{F(T_\mu(x)) - F^0}{\Delta b_F} \right) \left(\frac{\partial F(t)}{\partial t} \Big|_{t=T_\mu(x)} \right) \cdot \frac{\partial (T_\mu(x))}{\partial \mu} \end{aligned} \quad (22)$$

where $\beta^{(3)}$ is the derivative of the cubic B-Spline function, $\partial F(t)/\partial t$ denotes the simple gradient of the transformed float image $F(T_\mu(x))$, $\partial (T_\mu(x))/\partial \mu$ is calculated by the selected

3D FFD transformation model. Referring to the expression of dissimilarity measure in (13) and combining it with (21) and (22), we can easily show that the derivative of D is

$$\frac{\partial D}{\partial \mu} = \frac{\alpha}{\alpha-1} \left\{ \left(\sum_j \theta(f_j) \right)^\alpha \sum_j \theta(f_j)^{\alpha-1} \frac{\partial \theta(f_j)}{\partial \mu} - \sum_i \left(\sum_j \theta(r_i, f_j) \right)^\alpha \sum_j \theta(r_i, f_j)^{\alpha-1} \frac{\partial \theta(r_i, f_j)}{\partial \mu} \right\} \quad (23)$$

where $\partial \theta(f_j)/\partial \mu$ and $\partial \theta(r_i, f_j)/\partial \mu$ represent respectively the derivatives of the estimated marginal probability of the transformed float image in (21) and the joint probability stated in (22).

4. Experiments and results

To evaluate the nonrigid registration method, two groups of experiments were conducted on simulated 3D brain MRI images and real 3D thoracic CT data, respectively. In section 4.1, these datasets are described. The simple registration experiments are carried out on several model images in section 4.2. The performance of our method in terms of registration accuracy and noise immunity is exemplified through simulations in section 4.3 and compared with MI technique. Its interest in solving the elastic registration of real CT data is shown in Section 4.4. Finally, the overall results are analyzed.

Our nonrigid registration method based on JA divergence was implemented in the registration package *elastix* (Klein *et al* 2010) available at <http://elastix.isi.uu.nl>. This package is based on the Insight Toolkit (ITK) (Ibáñez *et al* 2005). In this work, all experiments were performed on a 2.33 GHz and 6 GB memory PC.

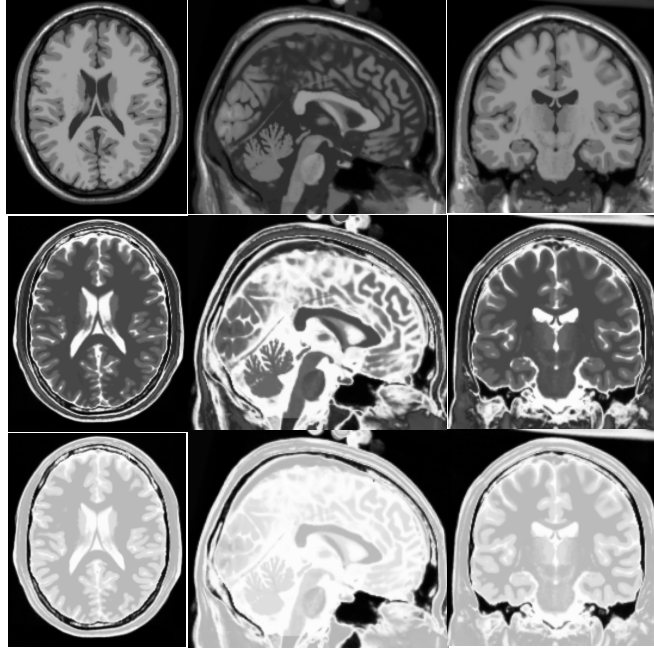


Figure 3. Three orthogonal central planes in the simulated MR brain volumes. Columns: axial, sagittal and coronal slices of the MR volumes. Rows 1 to 3 denote MR T1, MR T2 and MR PD images, respectively.

4.1. Test data

4.1.1 Simulated brain MRI images

The simulated data were obtained from the Brainweb database of the Montreal Neurological Institute (Cocosco *et al* 1997). They contain 3D brain MRI volumes simulated through three different protocols: T1-weighted, T2-weighted and Proton Density-weighted (PD-weighted) with various slice thicknesses, noise levels and intensity non-uniformities. The size of the volumes (with voxel coding on 8 bits) is $181 \times 217 \times 181$ voxels with a voxel size $1\text{mm} \times 1\text{mm} \times 1\text{mm}$. All corresponding slices in these volumes were aligned with each other. In the simulated experiments, the corresponding volumes from MR T1 & MR PD pairs were chosen for evaluation. MR T1 volume served as the reference image and MR PD volume as the float image to be aligned using a known 3D nonrigid transformation (with the true value T_0). Figure 3 depicts three orthogonal central slices (i.e. axial, sagittal and coronal) used in our protocol.

4.1.2 Real 3D thoracic CT data

Experiments were also conducted on forty 3D thoracic CT images acquired from four 4D CT sequences of four patients. These data were obtained on a Brilliance Big Bore 16-slice CT scanner (Philips Medical Systems, Cleveland, OH) and were part of a radiotherapy planning process at the Léon Bérard Cancer Center in Lyon, France. The 4D datasets were reconstructed to yield forty phases of 3D CT images (Vandemeulebroucke *et al* 2011, Vandemeulebroucke *et al* 2012). The first three datasets have 512×512 pixels in plane. The number of slices varies from 141 to 170, with an in-plane resolution ranging from $0.8789 \times 0.8789 \text{ mm}^2$ to $0.9766 \times 0.9766 \text{ mm}^2$ and 2mm interslice. The fourth dataset has the size of $482 \times 360 \times 141$ with the voxel dimension of $0.9766 \times 0.9766 \times 2\text{mm}^3$. For each phase image of the data, a number of landmarks features were labeled manually by an expert. The image information and landmarks for each case are shown in Table 1. More detailed descriptions of the data can be found in (Vandemeulebroucke *et al* 2011, Vandemeulebroucke *et al* 2012). Figure 4 displays three orthogonal planes of the maximal exhale and inhale phases of 4D CT image of the forth dataset.

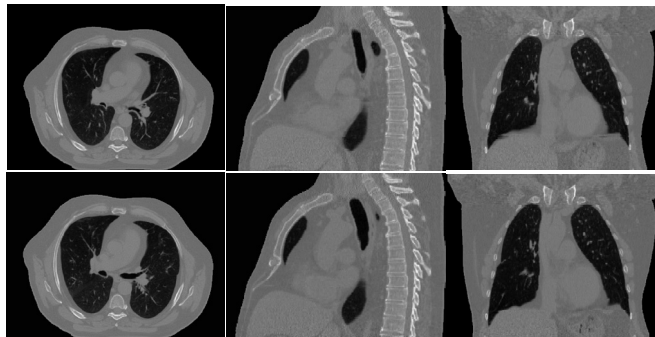


Figure 4. Two frames of 3D thoracic CT images extracted from a 4D CT scan. The top row represents the maximum exhalation frame and the bottom row, the maximum inhalation frame. Columns 1 to 3 in each row display the axis, coronal and sagittal slices of two phases of the CT volumes, respectively.

4.2. Illustration

To illustrate the effect of the pseudo-additivity property in image registration, we applied our method as well as the mutual information method to register the MR T1 images shown in figure 5, with different degrees of deformations (various levels of correlation) between two images to be registered. Here, we chose figure 5 (a) as the reference image and the others as float images. Note that when warping indexes are increased, the correlation coefficients between them are gradually decreased. Figure 6 exhibits the registration results using mutual information (MI) based on the standard Shannon entropy and our similarity measure based on the Arimoto nonextensive entropy. Only slight differences between our measure and MI are observed when registering figure 5 (a) and (b), (c), (d) with relative small deformations. However, due to the non-extensibility of the Arimoto entropy, our measure (figure 6 (d), (e) and (f)) gives a much more accurate alignment in the presence of large deformations than MI (figure 6 (g), (k) and (l)). This is particularly visible at the level of the cerebral ventricles, a major structure of the brain (see red arrows).

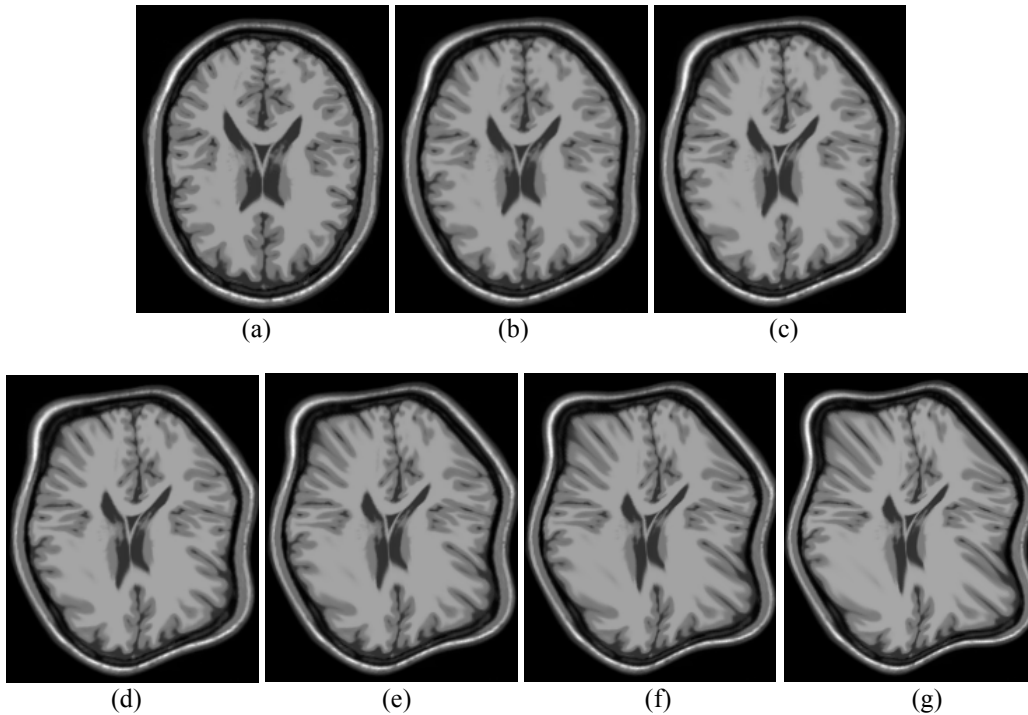


Figure 5. The MR T1 images. (a) Reference image, (b)-(g) float images, with different deformations with respect to (a), along with the correlation coefficients between them being 0.9055, 0.8492, 0.8059, 0.7745, 0.7543 and 0.7399.

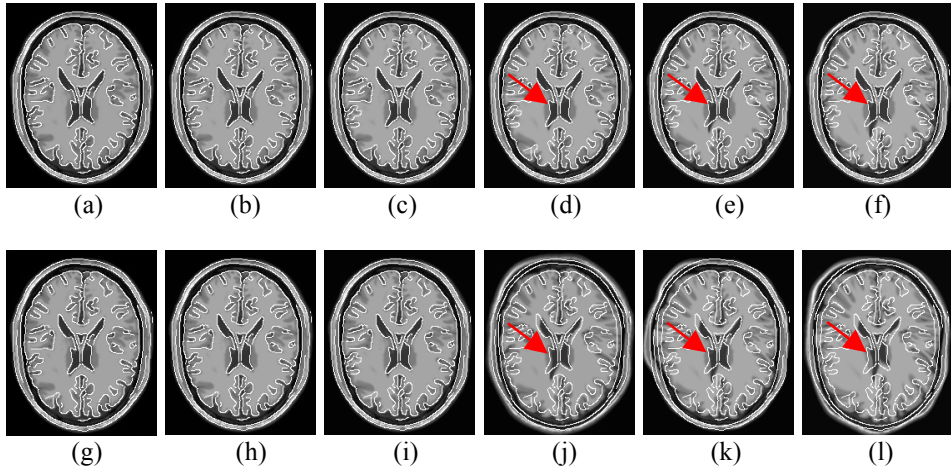


Figure 6. Registration results of the MR T1 images displayed in figure 5. (a)-(f) The transformed float images superimposed with the edges (Canny's operator) of the reference image (a) when registering figure 5 (a) and (b)-(g) using JA measure, (g)-(l) the corresponding results when employing MI method.

4.3. Simulated Data Experiments

The performance of our JA registration method is first assessed using the simulated 3D brain data through a three-level hierarchical multiresolution scheme. The role of the several parameters involved in our method is examined and a comparison with mutual information (MI) is also reported.

Table 1. CT images information and landmarks of four datasets.

CT properties	Volume sizes	Voxel dimensions(mm)	Landmarks
Dataset 1	$512 \times 512 \times 141$	$0.9766 \times 0.9766 \times 2$	100
Dataset 2	$482 \times 360 \times 141$	$0.9766 \times 0.9766 \times 2$	41
Dataset 3	$512 \times 512 \times 169$	$0.9766 \times 0.9766 \times 2$	100
Dataset 4	$512 \times 512 \times 170$	$0.8789 \times 0.8789 \times 2$	100

4.3.1. Parameters setting

The registration parameters were chosen by means of trial-and-error on the simulated data. To demonstrate the effect of these parameters, the experiment process was performed as follows. Using the pair of MR T1 and MR PD volumes mentioned in Section 4.1.1, the former volume was chosen as reference image while the latter volume was elastically deformed and considered as the float image. This deformation is based on the following warping function being analogous as that of (Lu *et al* 2008)

$$x' = x + m \left(\sin\left(\frac{\pi x}{48}\right) + \sin\left(\frac{\pi y}{48}\right) + \sin\left(\frac{\pi z}{48}\right) \right) \quad (24)$$

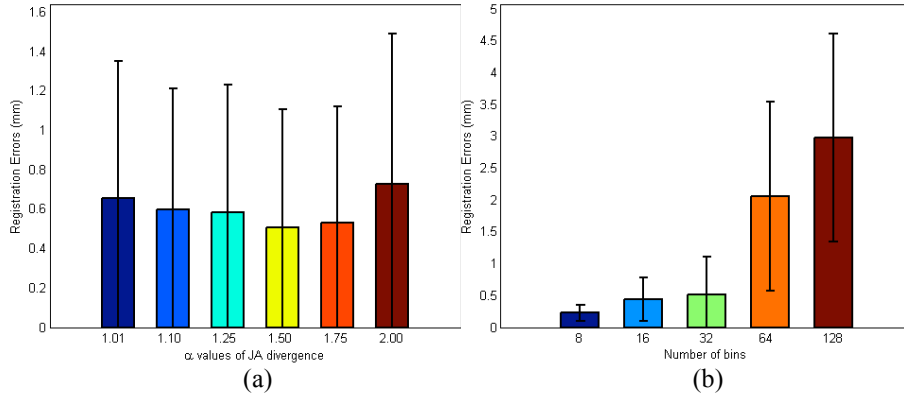
$$y' = y + m \left(\sin\left(\frac{\pi x}{48}\right) + \sin\left(\frac{\pi y}{48}\right) + \sin\left(\frac{\pi z}{48}\right) \right) \quad (25)$$

$$z' = z + m \left(\sin\left(\frac{\pi x}{48}\right) + \sin\left(\frac{\pi y}{48}\right) + \sin\left(\frac{\pi z}{48}\right) \right) \quad (26)$$

where (x, y, z) and (x', y', z') denote respectively the original coordinates before and after deformation. The parameter m accounts for the warping index, and also determines the magnitude of the deformation field.

To evaluate quantitatively the registration results, the difference between the true values and the estimated values (magnitude of the displacement vector) was calculated as the registration error. In all experiments of this section, sixty warping indexes were selected randomly from a wide-range interval $[1, 7]$, along with sixty pairs of test images being produced consequently. We set the weighting parameter $\lambda = 0.005$ that provides a good tradeoff between the dissimilarity measure D and the regularization term S in the objective function E .

These studied parameters include the nonextensive parameter α , the number of bins M , the spacing of mesh points δ and the amount of random samples N . Figure 7 displays the mean and standard deviation of the registration errors when varying these four parameters. As illustrated in figure 7 (a), the lowest mean of registration error is obtained when $\alpha = 1.50$ (here, $M = 32$, $\delta = 16 \times 16 \times 16$, $N = 2000$). The difference in mean running time (about 4min) is not so large when varying α values, hence, we set $\alpha = 1.50$ for the remaining experiments. The average calculation times when varying the other three parameters are tabulated in Table 2. It can be observed from figure 7 (b) and Table 2 that the errors and the computation times gradually increase with a higher number of bins M . Although a larger value of M leads to a better estimation of the probability density functions, a compromise has to be set. The value $M = 16$ has been retained here. The mesh spacing with value of $16 \times 16 \times 16$ provides the better registration results in figure 7 (c) and has been selected. Figure 7 (d) shows that the lowest registration error is obtained when $N = 5000$. However, the computation times increase rapidly with a higher number of random samples (Table 2), thus the value $N = 2000$ was preferred.



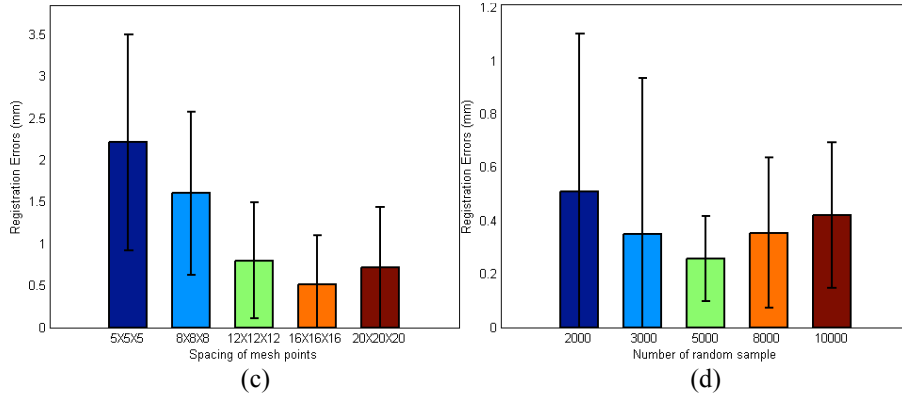


Figure 7. The non-rigid registration results of the simulated 3D brain MR T1 and MR PD images when choosing various values of four parameters, α , M , δ and N . The deformations parameters were generated randomly from the range $[1, 7]$, with the unit of deformation and errors being also millimeters. (a) Registration errors for several various α values, $\alpha \in \{1.01, 1.10, 1.25, 1.50, 1.75, 2.00\}$ and $M = 32$, $\delta = 16 \times 16 \times 16$, $N = 2000$; (b) The results of different bins size, $M \in \{8, 16, 32, 64, 128\}$ with $\alpha = 1.50$, $\delta = 16 \times 16 \times 16$, $N = 2000$; (c) The effect of mesh points with $\delta \in \{5 \times 5 \times 5, 8 \times 8 \times 8, 12 \times 12 \times 12, 16 \times 16 \times 16, 20 \times 20 \times 20\}$, $\alpha = 1.50$, $M = 32$, $N = 2000$; (d) Statistics of registration errors for the number of random samples, $N \in \{2000, 3000, 5000, 8000, 10000\}$ and $\alpha = 1.50$, $M = 32$, $\delta = 16 \times 16 \times 16$.

Table 2. Average calculation times of the registration shown in Figure.5. Here, cases 1 to 5 denote $\{8, 16, 32, 64, 128\}$ for the number of bins M , $\{5 \times 5 \times 5, 8 \times 8 \times 8, 12 \times 12 \times 12, 16 \times 16 \times 16, 20 \times 20 \times 20\}$ for the spacing of mesh points δ and $\{2000, 3000, 5000, 8000, 10000\}$ of random samples N .

Parameters	Case 1	Case 2	Case 3	Case 4	Case 5
Bins M	3min20s	3min30s	4min	5min	9min
Spacing δ	13min	6min	4min30s	4min	3min40s
Samples N	4min	6min	8min	14min	17min

4.3.2. Accuracy

To illustrate the registration accuracy of our method, MI is compared to the JA divergence on three groups of 3D brain data sets, MR T1 & MR T2, MR T1 & MR PD and MR T2 & MR PD. In each group, the former volume is used as reference image. We selected sixty warping indexes generated randomly from the same range $[1, 7]$ as in section 4.3.1, resulting in sixty pairs of test images produced for each group of data and 360 nonrigid registrations of these three groups of data sets for the two methods in total.

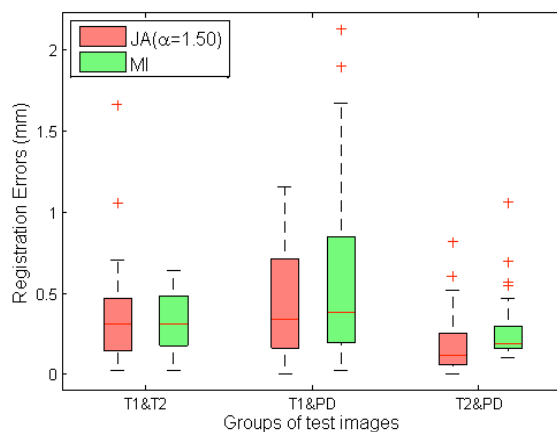


Figure 8. The registration results of the simulated 3D brain MR T1 & MR T2, MR T1 & MR PD and MR T2 & MR PD volumes using two algorithms. Experimental parameters are set by $\alpha = 1.50$, $M = 16$, $\delta = 16 \times 16 \times 16$, $N = 2000$.

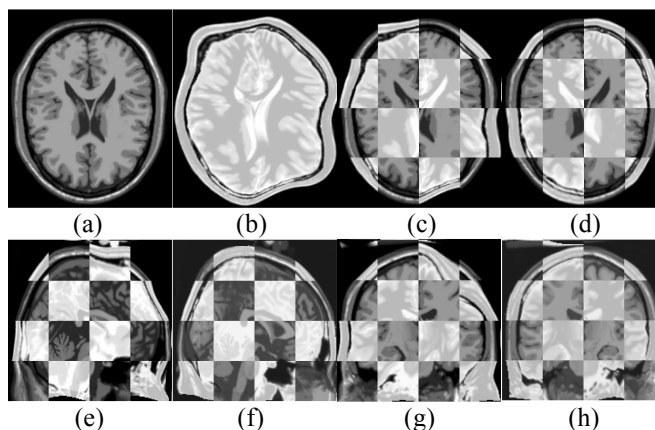


Figure 9. The checkboard of non-rigid registration results using JA method ($\alpha = 1.50$). (a) central slice of the reference MR T1 volume; (b) central slice of the deformed float MR PD volume; (c) checkboard of middle axial slice in reference and float volumes before registration; (d) checkboard of middle axial slice in reference and float volumes after registration; (e)-(f) and (g)-(h) shows the checkboard of the sagittal and coronal planes, respectively.

Figure 8 displays the resulting box-and-whisker plots of the registration errors for all 360 trials. For each box, the central red mark is the median of the registration errors of the ten tests, with two edges of each box denoting the 25% and 75%, the upper and below whisker being the maximum and minimum without considering outliers. In addition, these outliers are represented by the red color crosses for each box. As it can be observed, the registration errors for MR T1 & MR PD are relative high when compared to those of other two groups while the lowest errors are obtained for MR PD to MR T2. The JA method leads to a slight improvement in registration accuracy.

Figure 9 depicts the checkboard of one non-rigid registration example resulting from our method for a MR T1 & MR PD registration case when the warping index $m = 5$ ($\alpha = 1.50$). The shape of the registered float image appears visually close to the reference image.

4.3.3. Noise immunity

In this section, we assess the performance of our method and MI-based algorithms in presence of different levels of noise. The three groups of data sets are those already exploited in section 4.3.2. The warping index m has been set to 5. Sixty levels of Gaussian noise with zero-mean were randomly generated along with their standard deviations chosen from the range $[0, 60]$. These different levels of Gaussian noise were added to the deformed float images, and sixty pairs of test images were obtained for each group of data (MR T1 & MR T2, MR T1 & MR PD and MR T2 & MR PD).

Altogether, 360 registrations were performed. In order to accelerate the optimization process, we adopted a three-level multi-resolution strategy. For each resolution level, the L-BFGS optimization method was used to minimize the objective function with the registration parameters specified in section 4.3.1. The calculation times for registration are almost the same: 3min 20s for our method and 3min 30s for the MI method. Figure 10 shows the box-and-whisker plots of the resulting registration errors. As it can be seen, the average registration errors obtained using the two methods are less than one millimeter, i.e. a subvoxel accuracy. The difference between MI and our method is higher than the one reported figure 10 which means that the JA approach is more robust to noise effect.

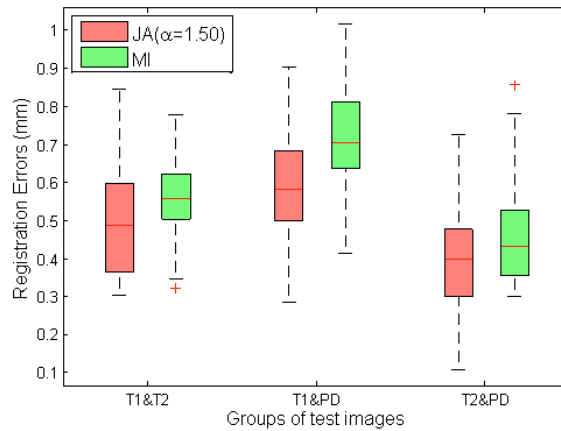


Figure 10. The registration errors of the simulated 3D brain MR T1 & MR T2, MR T1 & MR PD and MR T2 & MR PD volumes using MI and our approach in the presence of different levels of noise. The parameters have the same values than before ($\alpha = 1.50$, $M = 16$, $\delta = 16 \times 16 \times 16$, $N = 2000$).

4.4. Real Data Experiments

In this section, a series of experiments on real 3D thoracic CT data were carried out. Forty phases of four 4D CT from four patients were selected as test images. For each phase image, a certain number of landmark features (Table 1) were labeled manually by an expert, and these landmarks correspond to each other during the forty phase images. A three-level hierarchical method was used to achieve the best compromise of registration accuracy and calculation time.

The ten frames for each 4D CT, designated as T00-T90, cover the time sequence between the maximum inhalation (marked as T10) and the maximal exhalation (indicated by T60). To evaluate the performance of our method, the experiments were designed as follows: for each 4D CT set, the T60 phase served as the reference image and the float image was selected from the residual nine phases. Thus, 36 pairs of nonrigid registrations were performed for the

total four data sets, resulting in 108 registrations using our approach and the normalized cross correlation (NCC) and MI methods. The deformation registration accuracy was evaluated using the target registration error (TRE) (Vandemeulebroucke *et al* 2011, Vandemeulebroucke *et al* 2012) calculated as the 3D Euclidean distance between the manually marked landmarks in the maximum exhalation image and the ones estimated by exploiting the registration approach to the corresponding location in the inhale image (Castillo *et al* 2010).

Figure 11 illustrates the box-and-whisker plots of TREs for 108 registrations. For visual inspection, figure 12 displays an example of registration using our method for T10 and T60 images of the dataset 1. The mean and standard deviation of 3D TREs before and after registration are exhibited in figure 13. To evaluate the significance of differences in registration errors of JA, MI and NCC, a statistical test based on the Wilcoxon signed-ranks was conducted on the TREs of four real CT datasets. At the significance level of 0.05, a p value less than 0.05 indicates a rejection of null hypothesis, and a p value larger than 0.05 means an acceptance of null hypothesis. The results of Wilcoxon signed-ranks tests between JA and the other two methods on four datasets are reported in Table 3. They show that the registration errors of four groups of data when adopting JA are significantly lower than NCC ($p < 0.05$). From Table 3, a significant difference between JA and MI on TREs of datasets 1 and 3 is found ($p < 0.05$); no significance for datasets 2 and 4 ($p > 0.1$). However, for datasets 2 and 4, if we only consider those cases in which the deformations between the two 3D CT images to be registered are relatively large, a significant difference of JA with MI can be observed ($p < 0.05$).

Table 3. p -values of Wilcoxon signed-ranks tests of JA with MI and NCC on four sets of CT data. The significance level is 0.05. A p -value less than 0.05 means that the difference between two methods is significant, otherwise no significance is found.

Methods	Dataset 1	Dataset 2	Dataset 3	Dataset 4
JA and MI	0.0195	0.1641	0.0442	0.2031
JA and NCC	0.0477	0.0273	0.0391	0.0441

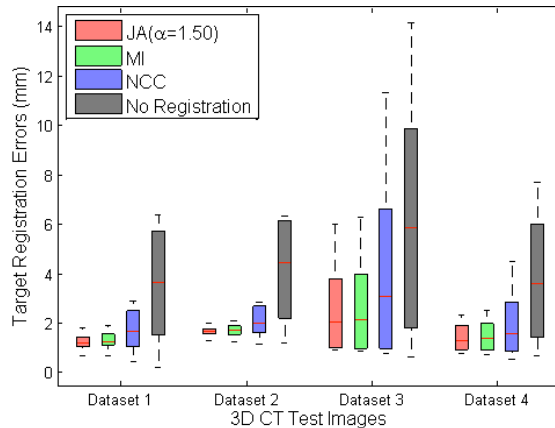


Figure 11. The target registration errors (TRE) obtained when employing JA algorithm, MI and NCC method for 3D thoracic CT images from four 4D CT data sets.

The limited memory BFGS scheme was applied as in section 4.3 with a three-level multiresolution strategy, the maximum number of iterations being 100 per level. Besides,

3000 samples were selected randomly by the image sampler at each iteration. The runtimes of our method, MI and NCC were approximately 6, 5.5 and 6.5 minutes, respectively.

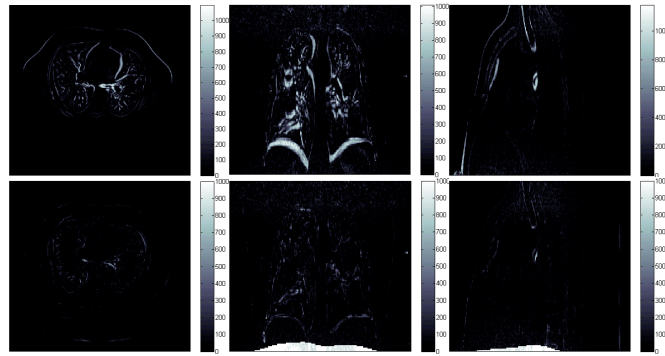


Figure 12. Difference images in orthogonal central planes before registration (row 1) and after registration (row 2) between T10 and T60 phase images of data set 1 using the JA algorithm ($\alpha = 1.50$). From left to right in each row, the transverse, coronal and sagittal slices.

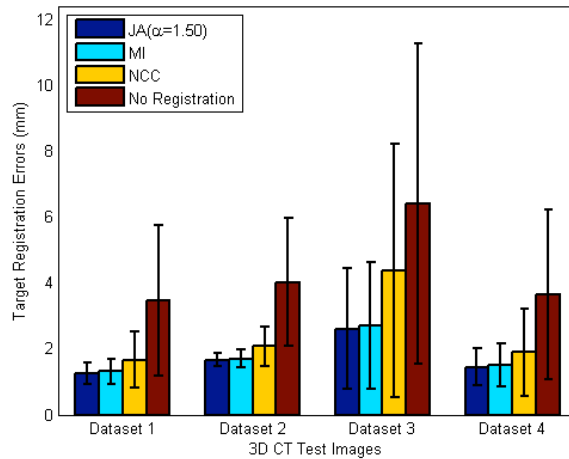


Figure 13. The mean and standard deviation of target registration errors before registration and after registration exploiting the JA, MI and NCC methods for the four groups of 3D thoracic CT images.

4.5. Analysis of Results

The objective of the nonrigid registrations performed on simulated data was to evaluate the effect of the parameters on the performance of the JA method. These parameters include α values, the number of bins, spacing of mesh points and random samples. Several experiments on three pairs of simulated MR volumes, MR T1 & MR T2, MR T1 & MR PD and MR T2 & MR PD were conducted and the results were compared with the classic information based measure (i.e. MI method). The test datasets were initially aligned with each other, that is to say, the true transformations (the “gold standard”) were known. In our case, we utilized in each image pair, the former MR T1 or MR T2 volume and the latter MR T2 or MR PD volume as the reference and float images, respectively. To compare the registration accuracy of our method, sixty trials were carried out based on deformations using distortion parameters randomly generated within a large range. Smaller errors and a better behavior with respect to convergence were observed for the JA method. To assess the immunity of noise, sixty levels

of noise were added to the float images and 60 trials were performed. For these sixty tests, in which we fixed the magnitude of elastic deformations, a better robustness to noise was observed, showing thus the effectiveness of our approach.

The goal of the second set of experiments was to assess its performance on real data. 3D thoracic CT data sets corresponding to an entire breath cycle were used. We considered the geometric transformation between two phases of 3D CT images. In such case, the so-called “ground truth” is unknown. To estimate the registration accuracy of our method, a number of landmarks were first defined in four groups of 4D CT data. This allowed exploiting the 3D target registration errors (TRE) between the corresponding landmark locations as a measure of performance. In all ten frames of the breath cycle of one 4D data, T60 was chosen as the reference image, the remainder frames being the float images. Nine groups of experiments were generated for each 4D CT data and 36 pairs of test images in total were considered. The experimental results show a better behavior of our JA algorithm with low TREs when compared to MI and NCC methods, achieving a sub-voxel registration accuracy in almost all cases.

5. Conclusion

In this work, a novel similarity measure based on the Arimoto entropy, called the Jensen-Arimoto divergence (JA) has been proposed for nonrigid registration of medical images. This new similarity measure is a generalization of the well-known Jensen-Shannon divergence measure. Based on the properties introduced in section 2.2 of the JA divergence, a new non-rigid registration technique has been presented. We employed free-form deformations as the parameter space along with the negative JA divergence as the dissimilarity measure and a penalty term to smooth the deformation constituting the entire registration criteria (objective function). To search the minima of this objective function, the L-BFGS optimization approach was used. We applied the Parzen window estimation algorithm with zero order and cubic order B-spline function as the kernel function to estimate the marginal probability distribution of the two volumes to be registered. The joint probability distribution between them was obtained in the same way.

Experiments were carried out on two groups of data sets, the former using simulated brain data (MR T1 & MR T2, MR T1 and MR PD and MR T2 & MR PD volumes), the latter including four time sequences of real 3D thoracic CT data highlighting an entire respiratory cycle. The results obtained on simulated sets have shown that our alignment technique leads to lower registration errors in the presence of various magnitudes of deformation and levels of noise when compared to the standard MI method. It has been assessed through the tests made on real 3D data and using the 3D target registration error as quality criterion that the JA method provides sub-voxel registration accuracy. As mentioned above, the parameter α in Arimoto entropy weights the nonextensive degree. The nonextensivity gradually becomes weak as α tending to 1, and the JA divergence also approximates to mutual information that accounts for the complete extensivity property. While in this paper α value is a self-defined constant by real experiments, its choice determines in some extent the quality of the registrations. So α value is a sensitive factor and how to select a suitable α value is an existing difficulty when choosing JA as registration criteria. In the next future, we will extend our applications to other multimodal medical images such as ultrasound and MR images and also to other organs.

Acknowledgments

This work was supported by the National Basic Research Program of China under Grant 2011CB707904, by the National Natural Science Foundation of China under Grants 61201344, 61271312, 61401085, 81101104 and 61073138, by the Ministry of Education of China under Grants 20110092110023 and 20120092120036, the Project-sponsored by SRF for ROCS, SEM, and by Natural Science Foundation of Jiangsu Province under Grant BK 2012329, BK2012743, DZXX-031, BY2014127-11, by the ‘333’ project under Grant BRA2015288 and by the Qing Lan Project.

References

- Antolin J, López-Rosa S, Angulo J C and Esquivel R O 2010 Jensen–Tsallis divergence and atomic dissimilarity for position and momentum space electron densities *J. Chem. Phys.* 132 044105.
- Arimoto S 1971 Information-theoretical considerations on estimation problems *Information and Control* 19 181–194
- Bentoutou Y, Taleb N, El Mezouar M C, Taleb M and Jetto L 2002 An invariant approach for image registration in digital subtraction angiography *Pattern Recognit.* 35 2853–2865
- Boekee D E and Van Der Lubbe J C A 1980 The R-norm information measure *Information and Control* 45 136–155
- Castillo E, Castillo R, Martinez J, Shenoy M and Guerrero T 2010 Four dimensional deformable image registration using trajectory modeling *Phys. Med. Biol.* 55 305–327
- Cocosco C A, Kollokian V, Kwan R K S, Pike G B and Evans A C 1997 Brainweb: Online interface to a 3D MRI simulated brain database *NeuroImage* 5
- Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P and Marchal G 1995 Automated multi-modality image registration based on information theory *14th Int. Conf. on IPMI*, Ile De Berder, France 263–274
- Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd ed., New York: Wiley
- Flusser J and Suk T 1998 Degraded image analysis: an invariant approach *IEEE Trans. Pattern Anal. Mach. Intell.* 20 590–603
- He Y, Ben Hamza A and Krim H A 2003 Generalized Divergence Measure for Robust Image Registration *IEEE Trans. Signal Process.* 51 1211–1220
- Hill D L G, Batchelor P G, Holden M and Hawkes D J 2001 Medical image registration *Phys. Med. Biol.* 46 R1–R45
- Hoh C K, Dahlbom M, Harris G, Choi Y, Hawkins R A, Philips M E and Maddahi J 1993 Automated iterative three-dimensional registration of positron emission tomography images *J. Nucl. Med.* 34 2009–2018
- Huber P J 1981 *Robust Statistics*. New York: Wiley
- Ibáñez L, Schroeder W, Ng L and Cates J 2005 *The ITK Software Guide* Kitware ISBN 1–930934-15–7
- Khader M and Hamza A B 2011 Nonrigid image registration using an entropic similarity *IEEE Trans. Inf. Technol. Biomed.* 15 681–690
- Kim J and Fessler J 2002 A Image registration using robust correlation in *Proc. IEEE Int. Symp. Biomedical Imaging* 353–356
- Kim J and Fessler J A 2004 Intensity-based image registration using robust correlation coefficients *IEEE Trans. Med. Imaging* 23
- Klein S, Staring M and Pluim J P W 2007 Evaluation of Optimization Methods for Nonrigid Medical Image Registration Using Mutual Information and B-Splines *IEEE Trans. Image Process.* 16 2879–2890
- Klein S, Staring M, Murphy K, Viergever M A and Pluim J P W 2010 elastix: A Toolbox for Intensity-Based Medical Image Registration *IEEE Trans. Med. Imaging* 29 196–205
- Kong Y Y, Deng Y and Dai Q H 2015 Discriminative Clustering and Feature Selection for Brain MRI Segmentation *IEEE Signal Process. Lett.* 22 573–577
- Lemieux L, Jagoe R, Fish D R, Kitchen N D and Thomas D G T 1994 A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs *Med. Phys.* 21 1749–1760
- Li B C, Yang G Y, Shu H Z and Coatrieux J L 2014 A New Divergence Measure Based on Arimoto Entropy for Medical Image Registration *22nd ICPR Stockholm Sweden Aug.* 24–28 3197–3202
- Li M, Chen X, Li X, Ma B and Vitányi P M B 2004 The Similarity Metric *IEEE Trans. Inf. Theory* 50 3250–3264
- Liese F and Vajda I 2006 On Divergences and Informations in Statistics and Information Theory *IEEE Trans. Inf. Theory* 52 4394–4412
- Lin J H 1991 Divergence Measures Based on the Shannon Entropy *IEEE Trans. Inf. Theory* 37 145–151
- Loeckx D, Slagmolen P, Maes F, Vandermeulen D and Suetens P 2010 Nonrigid Image Registration Using Conditional Mutual Information *IEEE Trans. Med. Imaging* 29 19–29
- Lou Y F, Irimia A, Vela P A, Chambers M C, Van Horn J D, Vespa P M and Tannenbaum A R 2013 Multimodal Deformable Registration of Traumatic Brain Injury MR Volumes via the Bhattacharyya Distance *IEEE Trans. Biomed. Eng.* 60 2511–2520
- Lowe D 2004 Distinctive Image Features from Scale-Invariant Keypoints *Int. J. Comput. Vis.* 60 91–110
- Lu X S, Zhang S, Su H and Chen Y Z 2008 Mutual information-based multimodal image registration using a novel joint histogram estimation *Comput. Med. Imaging Graph.* 32 202–209
- Maes F, Collignon A, Vandermeulen D, Marchal G and Suetens P 1997 Multimodality image registration by maximization of mutual information *IEEE Trans. Med. Imaging* 16 187–198
- Maintz J B A, van den Elsen P A and Viergever M A 1996 Comparison of edge-based and ridge-based registration of CT and MR brain images *Med. Image Anal.* 1
- Maintz J B A and Viergever M A 1998 A survey of medical image registration *Med. Image Anal.* 2 1–36

- Mattes D, Haynor D R, Vesselle H, Lewellen T K and Eubank W 2003 PET-CT image registration in the chest using free-form deformations *IEEE Trans. Med. Imaging* 22 120–128
- Maurer C and Fitzpatrick J M 1993 A review of medical image registration *Interactive Image-Guided Neurosurg.* Park Ridge, IL: American Association of Neurological Surgeons 17–44
- Nocedal J 1980 Updating quasi-Newton matrices with limited storage *Math. Comput.* 35 773–782
- Pluim J P W, Maintz J B A and Viergever M A 2000 Image registration by maximization of combined mutual information and gradient information *IEEE Trans. Med. Imaging* 19 809–814
- Pluim J P W, Maintz J B A and Viergever M A 2001 f-Information measures in medical image registration *Proc. SPIE Medical Imaging: Image Processing* M. Sonka and K. M. Hanson, Eds., San Diego, CA 2 579–587.
- Pluim J P W, Maintz J B A and Viergever M A 2003 Mutual-Information Based Registration of Medical Images: A Survey *IEEE Trans. Med. Imaging* 22 986–1004
- Pluim J P W, Maintz J B A and Viergever M A 2004 f-Information Measures in medical image registration *IEEE Trans. Med. Imaging* 23 1508–1516
- Press W H, Flannery B P, Teukolsky S A and Vetterling W T 2007 Numerical Recipes in C 3rd ed. Cambridge, U. K.: Cambridge Univ. Press., ch. 10 521–526
- Rao M, Chen Y, Vemuri B C and Wang F 2004 Cumulative residual entropy: A new measure of information *IEEE Trans. Inf. Theory* 50 1220–1228
- Rohlfing T, Maurer C R, Bluemke D A and Jacobs M A 2003 Volume-Preserving Nonrigid Registration of MR Breast Images Using Free-Form Deformation With an Incompressibility Constraint *IEEE Trans. Med. Imaging* 22 730–741
- Rueckert D, Sonoda L I, Hayes C, Hill D L G, Leach M O and Hawkes D J 1999 Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images *IEEE Trans. Med. Imaging* 18 712–721
- Smith S M and Brady J M 1997 SUSAN - a new approach to low level image processing *Int. J. Comput. Vis.* 23 45–78
- Studholme C, Hill D L G and Hawkes D J 1999 An overlap invariant entropy measure of 3D medical image alignment *Pattern Recognit.* 32 71–86
- Studholme C, Drapaca C, Iordanova B and Cardenas V 2006 Deformation- based mapping of volume change from serial brain MRI in the presence of local tissue contrast change *IEEE Trans. Med. Imaging* 25 626–639
- Szeliski R and Coughlan J 1997 Spline-based image registration *Int. J. Comput. Vis.* 22 199–218
- Thévenaz P and Unser M 2000 Optimization of mutual information for multiresolution image registration *IEEE Trans. Image Process.* 9 2083–2099
- Tustison N J, Awate S P, Song G, Cook T S and Gee J C 2011 Point set registration using Havrda-Charvat-Tsallis entropy measures *IEEE Trans. Med. Imaging* 30 451–460
- Vandemeulebroucke J, Rit S, Kybic J, Clarysse P and Sarrut D 2011 Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs *Med. Phys.* 38 166–178
- Vandemeulebroucke J, Bernard O, Rit S, Kybic J, Clarysse P and Sarrut D 2012 Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT *Med. Phys.* 39 1006–1015
- Viola P and Wells III W M 1997 Alignment by maximization of mutual information *Int. J. Comput. Vis.* 24 137–154
- Wang F, Vemuri B C, Rao M and Chen Y 2003 Cumulative residual entropy, a new measure of information & its application to image alignment *Proc. of 9th IEEE ICCV*, Nice, France, Oct. 13–16, 548–553
- Wang F and Vemuri B C 2007 Non-rigid multi-modal image registration using cross-cumulative residual entropy *Int. J. Comput. Vis.* 74 201–215
- Wells W M, Viola P, Atsumi H, Nakajim S and Kiknis R 1996 Multi-modal volume registration by maximization of mutual information *Med. Image Anal.* 1 35–51