


Article

# Nonrigid Medical Image Registration Using an Information Theoretic Measure Based on Arimoto Entropy with Gradient Distributions

Bicao Li <sup>1</sup> , Huazhong Shu <sup>2</sup>, Zhoufeng Liu <sup>1</sup>, Zhuhong Shao <sup>3</sup>, Chunlei Li <sup>1,\*</sup>, Min Huang <sup>4</sup> and Jie Huang <sup>1</sup>

<sup>1</sup> School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China; lbc@zut.edu.cn (B.L.); liuzhoufeng@zut.edu.cn (Z.L.); 3983@zut.edu.cn (J.H.)

<sup>2</sup> Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing 210096, China; shu.list@seu.edu.cn

<sup>3</sup> College of Information Engineering, Capital Normal University, Beijing 100048, China; zhshao@cnu.edu.cn

<sup>4</sup> School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China; huangmin1998@126.com

\* Correspondence: 5540@zut.edu.cn

Received: 12 December 2018; Accepted: 14 February 2019; Published: 18 February 2019



**Abstract:** This paper introduces a new nonrigid registration approach for medical images applying an information theoretic measure based on Arimoto entropy with gradient distributions. A normalized dissimilarity measure based on Arimoto entropy is presented, which is employed to measure the independence between two images. In addition, a regularization term is integrated into the cost function to obtain the smooth elastic deformation. To take the spatial information between voxels into account, the distance of gradient distributions is constructed. The goal of nonrigid alignment is to find the optimal solution of a cost function including a dissimilarity measure, a regularization term, and a distance term between the gradient distributions of two images to be registered, which would achieve a minimum value when two misaligned images are perfectly registered using limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) optimization scheme. To evaluate the test results of our presented algorithm in non-rigid medical image registration, experiments on simulated three-dimension (3D) brain magnetic resonance imaging (MR) images, real 3D thoracic computed tomography (CT) volumes and 3D cardiac CT volumes were carried out on *elastix* package. Comparison studies including mutual information (MI) and the approach without considering spatial information were conducted. These results demonstrate a slight improvement in accuracy of non-rigid registration.

**Keywords:** Arimoto entropy; free-form deformations; normalized divergence measure; gradient distributions; nonextensive entropy; non-rigid registration

## 1. Introduction

Volume registration is an essential task of image processing, especially in medical field, such as aiding diagnosis, surgical applications, and image-guided radiation therapy [1,2]. The images to be registered are generally obtained at different times and from different imaging sensors, namely, multi-modality imaging. Different medical imaging modalities could provide various and complementary information. For instance, CT and MRI display the anatomic structures of an organ, while positron emission tomography (PET) and single photon emission computed tomography (SPECT) provide the functional and metabolic information. Therefore, in clinic, these multi-modal volumes are often registered and fused together, in this way, much complementary information derived from

different modalities are supplied with the physicians to improve the diagnosis accuracy and assessment efficiency of lesion progression.

Image registration is related by the process to find the optimal mapping function between two images to be aligned [3]. In recent years, the registration algorithms using the similarity based on information theory have been attracted more and more attention in medical image registration, among which maximization of MI was early reported for registration of medical images from different modalities by Collignon et al. [4], Maes et al. [5], Wells et al. [6]. Studholme et al. [7] studied a normalized similarity measure called normalized MI (NMI) to tackle the problem of changing field of view (FOV). MI and NMI are both estimated by the probability distributions of images to be registered. Besides, the concept of cumulative probability distributions was introduced into image registration, and cumulative residual entropy (CRE) was investigated in [8]. Additionally, the relations between CRE and Shannon entropy were ulteriorly researched in [9], and a comparison with MI was reported in [10]. In this new measure, cumulative density functions (CDF) instead of probability density functions (PDF) were adopted to calculate values of the similarity measure, which illustrates a good robustness to noise.

The aforementioned measures based on information theory—such as MI, NMI, and CRE—are constructed by Shannon entropy. Nonetheless, the additivity of Shannon entropy signifies that it is an extensive entropy. However, Antolin et al. [11] pointed out that the extensive entropy does not consider the correlation of two variables. Consequently, they presented a similarity measure exploiting Tsallis entropy. Subsequently, this new divergence was employed to construct a non-rigid registration model for medical images [12,13].

Illuminated by the reference [11], a divergence measure based on Arimoto entropy was presented called the Jensen–Arimoto divergence (JAD) [14]. In [15], some properties, such as the concavity of Arimoto entropy and the boundedness of JAD, have been further investigated. However, the registration method based on JAD does not take the spatial information between voxels into account, which is of significance to medical image registration. This paper aims to present a novel nonrigid registration method of medical images adopting a normalized measure based on Arimoto entropy and gradient distributions. Firstly, the properties of Arimoto entropy and JAD are analyzed, and a distance of gradient distributions is constructed. Secondly, a nonrigid deformation model is chosen and the registration process is formulated by an optimization procedure. In the sequel, the continuous probability distributions are estimated using Parzen window method applying B-splines and the analytical gradient of objective function can be obtained. Finally, the L-BFGS optimization [16] is adopted to obtain the optimal deformation parameters. To assess the performance of our registration framework for medical images, several groups of non-rigid experiments on simulated volumes and real 3D data are implemented.

Our contributions are twofold. Firstly, the related measures based on Shannon entropy do not consider the correlation. Therefore, we present a normalized measure based on Arimoto entropy, a non-extensive entropy, as the dissimilarity measure. Secondly, in the existing measures, such as MI, NMI and JAD, the intensity values are directly exploited to calculate the similarity measure, while the spatial information has been not considered. To take the spatial information between voxels into account, a distance term of gradient distributions is constructed and incorporated into the objective function.

The rest of this work is arranged as follows. In Section 2, the knowledge of information theory was firstly reviewed, and then introduce the Arimoto entropy and JAD measure, constructing gradient distribution distance. We formulate the model of nonrigid registration and give the detailed description of the registration method in Section 3 adopting a normalized measure based on Arimoto entropy with gradient distribution. Section 4 demonstrates the nonrigid test results on 3D MR volumes and real 3D clinical datasets, with the compared results to other registration algorithms illustrated. Finally, we provide the conclusions and perspectives in Section 5.

## 2. Preliminaries

For this section, we briefly review the theoretical concept of information theory, and then introduce the Arimoto entropy and JAD, along with studying their properties. In addition, the gradient distributions of reference image and float image are constructed and a distance of them is derived.

### 2.1. Shannon Entropy and Mutual Information

For an arbitrary random variable  $X(x_1, x_2, \dots, x_N)$ , with its probability distributions  $p(x_1, x_2, \dots, x_N)$ , the Shannon entropy of  $X$  is used to measure the amount of average uncertainty included in this random variable,

$$H(X) = -\sum_{i=1}^N p(x_i) \log p(x_i) \quad (1)$$

which is also employed to measure the account of information provided by this random variable. Then, considering another random variable  $Y$ ,  $H(X|Y)$  is remarked as the conditional entropy of  $X$  when  $Y$  is known. The reduction in uncertainty due to  $Y$  is called the MI. MI of  $X$  and  $Y$  is defined by

$$I(X, Y) = H(X) - H(X|Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)} \quad (2)$$

where  $p(x)$  and  $p(y)$  denote marginal probability of two random variables, as well as  $p(x, y)$  being the joint distribution of them. MI is applied as a measure of the dependence between  $X$  and  $Y$ . MI is symmetric in  $X$  and  $Y$  and always nonnegative [17].

### 2.2. Arimoto Entropy

Arimoto [18] introduced a generalized form of Shannon entropy, Arimoto entropy is defined by

$$A_\alpha(X) = \frac{\alpha}{\alpha - 1} \left[ 1 - \left( \sum_{i=1}^N p_i^\alpha \right)^{\frac{1}{\alpha}} \right] \quad \alpha > 0, \alpha \neq 1 \quad (3)$$

Boekee et al. [19] investigated some significant properties of Arimoto entropy, here, we only exhibit several useful properties as follows.

Non-negativity:

$$A_\alpha(X) \geq 0 \quad \alpha > 0, \alpha \neq 1 \quad (4)$$

Pseudo-additivity:

$$A_\alpha(X, Y) = A_\alpha(X) + A_\alpha(Y) - \frac{\alpha - 1}{\alpha} A_\alpha(X) A_\alpha(Y) \quad \alpha > 0, \alpha \neq 1 \quad (5)$$

Concavity:

$$A_\alpha(tX_1 + (1 - t)X_2) \geq tA_\alpha(X_1) + (1 - t)A_\alpha(X_2) \quad t \in (0, 1), \alpha > 0, \alpha \neq 1 \quad (6)$$

Symmetry:

$$A_\alpha(\dots, p_i, \dots, p_j, \dots) = A_\alpha(\dots, p_j, \dots, p_i, \dots) \quad (7)$$

Upper bound:

$$A_\alpha(p_1, p_2, \dots, p_N) \leq A_\alpha\left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right) = \frac{\alpha}{\alpha - 1} \left[ 1 - N^{\frac{1-\alpha}{\alpha}} \right] \quad (8)$$

See [19] for the detailed proof of these properties. Property one ensures that Arimoto entropy is non-negative. Its pseudo-additivity illustrates that Arimoto entropy accounts for a non-extensive entropy. The parameter  $\alpha$  in (3) accounts for the degree of non-extensivity.

### 2.3. Jensen Arimoto Divergence

For a random variable  $X$ , and  $P(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)$  is probability distributions on  $X$ . JAD is defined to be [15]

$$JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) = A_\alpha\left(\sum_{i=1}^N \omega_i \mathbf{p}_i\right) - \sum_{i=1}^N \omega_i A_\alpha(\mathbf{p}_i) \quad \alpha > 0, \alpha \neq 1 \quad (9)$$

where  $A_\alpha(\cdot)$  denotes Arimoto entropy and  $\omega_i$  is a weighted vector to constrain  $\omega_i \geq 0$  and  $\sum_{i=1}^N \omega_i = 1$ . In the following, we review the properties of JAD.

**Proposition 1.** *JAD has these properties of non-negativity, symmetry. Also, JAD is identical to 0 when and only when all of the probability distributions are the same as each other. The proof has been reported in [15].*

Li et al. pointed that a distance must fulfill four requirements [20]. JAD does not meet the triangle inequality, so JAD is not a true distance metric. Nonetheless, JAD can still be adopted to measure the disparity among the probability distributions of two random variables.

**Proposition 2.** *The JA divergence has the maximum when  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$  are degenerate distributions, where  $\mathbf{p}_i = \delta_{ij} = 1$  when  $i = j$  and 0 otherwise. The proof had been provided in [21], and the maximum equals to  $A_\alpha(\omega)$ .*

According to Proposition 1 and 2, it is obviously observed that JAD is bounded,  $0 \leq JA_\alpha(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) \leq A_\alpha(\omega)$ .

### 2.4. Gradient Distributions Distance

Given the reference image  $R$  and float image  $F$ ,  $\nabla F$ , and  $\nabla R$  represent the gradients of  $F$  and  $R$ , along with  $q(\nabla F)$  and  $p(\nabla R)$  representing the gradient distributions of  $F$  and  $R$ . Kullback–Leibler divergence (KLD) can be used to calculate the distance of  $q(\nabla F)$  and  $p(\nabla R)$  as

$$KLD(q||p) = \sum_{\mathbf{x}} q(\nabla F(\mathbf{x})) \log \frac{q(\nabla F(\mathbf{x}))}{p(\nabla R(\mathbf{x}))} \quad (10)$$

where  $\mathbf{x}$  denotes any point of image gradient,  $\mathbf{x} = [x, y, z]^T$ . KLD is also known as relative entropy, measuring the diversity of two probability distributions. In (10), we use the convention that  $0 \cdot \log(0/0) = 0$ . In other word, when  $R$  and  $F$  are completely registered, KLD between gradient distributions defined in (10) achieves the minimum value. Considering the spatial transformation  $T_\mu$ ,  $\mu$  denoted by the parameters of transformation model, the gradient distribution distance of  $F$  and the transformed  $F$  is given by

$$KLD(\nabla F(T_\mu(\mathbf{x}))||\nabla R(\mathbf{x})) = \sum_d \sum_{\mathbf{x}} q(\nabla F_d(T_\mu(\mathbf{x}))) \log \frac{q(\nabla F_d(T_\mu(\mathbf{x})))}{p(\nabla R_d(\mathbf{x}))} \quad (11)$$

where  $\nabla F_d(T_\mu(\mathbf{x}))$  and  $\nabla R_d(\mathbf{x})$  represent the gradients of transformed float image and reference image, and  $d$  is the dimension of the images to be registered. Subsequently, the Parzen method is employed to estimate the gradient distributions.

Denote  $\beta^{(0)}$  and  $\beta^{(3)}$  by zero-order and three-order B-spline, respectively. In the process of image registration, the transformation parameters  $\mu$  does not affect the reference image, the gradient of reference image is also constant in registration process. Consequently, the gradient of reference image can be calculated before registration to improve the implementation efficiency. In addition, a zero-order

B-spline is exploited to estimate the gradient distributions  $\nabla R_d(x)$  of reference image  $R(x)$ , with the probability density function of  $\nabla R_d(x)$  defined by

$$p(\nabla R_d(x)) = \tilde{p}_d(r_i) = \frac{1}{V} \sum_{x \in \Omega} \beta^{(0)} \left( r_i - \frac{\nabla R_d(x) - \nabla R_d^0}{\Delta b_R} \right) \tag{12}$$

where  $\Omega$  is volume domain to estimate probabilities.  $V$  denotes the number of voxels of  $\Omega$  domain, as well as  $d$  being the image dimension.  $r_i$  represents intensity levels of  $\nabla R_d(x)$ , and ‘ $\nabla$ ’ is the gradient operator. The three-order B-spline is adopted to compute the gradient distributions  $\nabla F_d(T_\mu(x))$  of transformed float image, with the probability density function of  $\nabla F_d(T_\mu(x))$  shown as

$$q(\nabla F_d(T_\mu(x))) = \tilde{q}_d(f_j) = \frac{1}{V} \sum_{x \in \Omega} \beta^{(3)} \left( f_j - \frac{\nabla F_d(T_\mu(x)) - \nabla F_d^0}{\Delta b_F} \right) \tag{13}$$

In Equations (12) and (13),  $\Delta b_R$  and  $\Delta b_F$  are the widths of bins.  $\nabla R_d^0$  and  $\nabla F_d^0$  is the minimum values in two images, and  $f_j$  represents intensity levels of  $\nabla F_d(T_\mu(x))$ . Substituting (12) and (13) into (11), we can obtain the following formula.

$$\begin{aligned} KLD(\nabla F(T_\mu(x)) || \nabla R(x)) &= \sum_d \sum_x \tilde{q}_d(f_j) \log \frac{\tilde{q}_d(f_j)}{\tilde{p}_d(r_i)} \\ &= \sum_d \left[ H(\tilde{q}_d) - \sum_x \tilde{q}_d(f_j) \log \tilde{p}_d(r_i) \right] \end{aligned} \tag{14}$$

where  $H$  is the Shannon entropy. In this paper, KLD of gradient distributions will be applied as a distance term in medical image registration, regularizing that the gradient distribution of float image is similar to the gradient distribution of reference image.

### 3. Description of Proposed Nonrigid Registration Method

The details of the registration method using a normalized information theoretic measure based on Arimoto entropy with gradient distributions is described. Firstly, an appropriate transformation model needs to be selected, where the registration criteria (objective function or cost function) and optimization algorithm is applied to optimize the criteria. Finally, the images to be registered are aligned using an optimal solution obtained by the optimization scheme. Figure 1 displayed the block diagram of our nonrigid registration framework.

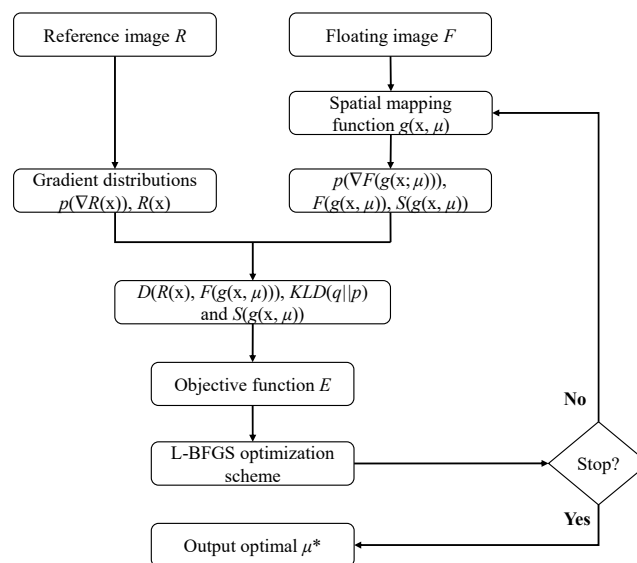


Figure 1. Block diagram of our registration algorithm.

### 3.1. Formulation

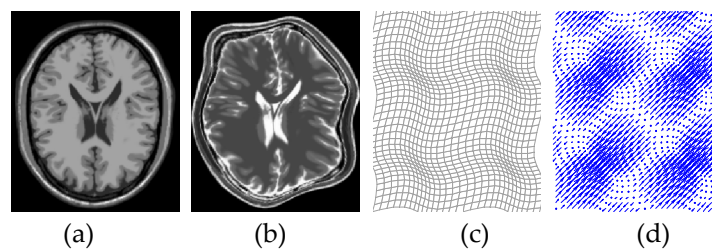
Figure 2a,b depict the corresponding planes in two 3D MR images, which account for T1-weighted MR and T2-weighted MR. Non-rigid registration is formulated as the process of searching for the optimal spatial deformation function of reference image  $R$  and float image  $F$  as

$$y = g(x; \mu) \quad (15)$$

where  $g(x; \mu)$  is the deformation function,  $\mu$  denotes the vector of deformation parameters, with  $x$  and  $y$  being the coordinates of arbitrary point in  $R$  and  $F$ , respectively. We can formulate the non-rigid registration of  $F$  to  $R$  as

$$T^* = \underset{\mu}{\operatorname{argmin}} D(F(x) \circ g(x; \mu), R(x)) = \underset{\mu}{\operatorname{argmin}} D(F(g(x; \mu)), R(x)) \quad (16)$$

where  $D$  represents a dissimilarity measure that can achieves its minimum in registration of  $R(x)$  and  $F(g(x; \mu))$ .



**Figure 2.** (a) MR T1 image; (b) MR T2 image; (c) deformation field; (d) deformation vector.

However, the process of image registration is an ill-posed issue, and a penalty term need to be incorporated to obtain a smooth transformation. Considering the regularization term, the cost function  $E$  is expressed by

$$E = D(F(g(x; \mu)), R(x)) + \lambda S(g(x; \mu)) \quad (17)$$

### 3.2. Transformation Model

Clinically, there exist some large deformations between medical images. Therefore, a nonrigid transformation is generally employed to deal with organ or tissue deformation. Figure 2c,d display the deformation fields and vectors of the nonrigid transformation between Figure 2a,b. The free-form deformations (FFD) model can preferably described local deformations between medical images. Therefore, cubic B-splines are employed to constructed FFD model [22] to simulate this elastic deformation.

In a 3D image,  $\Phi$  is a mesh with the size of  $n_x \times n_y \times n_z$ , as well as the control points  $[\omega_i, \omega_j, \omega_k]^T$ , and  $\delta$  is the size of spacing. Then, 3D deformation function of  $x = [x, y, z]^T$  could be defined as

$$g(x; \mu) = \sum_{ijk} \mu_{ijk} \beta^{(3)}\left(\frac{x - \omega_i}{\delta}\right) \beta^{(3)}\left(\frac{y - \omega_j}{\delta}\right) \beta^{(3)}\left(\frac{z - \omega_k}{\delta}\right) \quad (18)$$

where  $\mu_{ijk}$  is the vector of deformation coefficients, and  $\beta^{(3)}$  is the third-order B spine function.

### 3.3. Registration Criteria

The dissimilarity measure  $D$  is the most significant part of objective functions, which is used to measure the difference between two images to be registered. When  $D$  achieve the minimum value, the similarity of two volumes is maximized, resulting in the two volumes completely registered. We have employed JAD as similarity measure to register medical images in the presence of non-rigid

transformation, in which a negative sign is assigned to the JAD to construct the dissimilarity measure [15]. In this paper, according to Proposition 2 in Section 2.3, a normalized dissimilarity measure based on JAD is introduced. Its definition is given as

$$D(F(g(x;\mu)), R(x)) = 1 - \frac{JA_\alpha(F(g(x;\mu)), R(x))}{A_\alpha(\omega)} \quad (19)$$

In [15], JAD is expressed by

$$\begin{aligned} JA_\alpha(F(g(x;\mu)), R(x)) &= \frac{\alpha}{1-\alpha} \left\{ \left[ \sum_{j=1}^M \left[ \sum_{i=1}^M p(r_i) p(f_j | r_i) \right]^\alpha \right]^{\frac{1}{\alpha}} - \sum_{i=1}^M p(r_i) \left[ \sum_{j=1}^M p(f_j | r_i)^\alpha \right]^{\frac{1}{\alpha}} \right\} \\ &= \frac{\alpha}{1-\alpha} \left\{ \left[ \sum_{j=1}^M p(f_j)^\alpha \right]^{\frac{1}{\alpha}} - \sum_{i=1}^M \left[ \sum_{j=1}^M p(r_i, f_j)^\alpha \right]^{\frac{1}{\alpha}} \right\} \end{aligned} \quad (20)$$

where  $f = (f_1, f_2, \dots, f_M)$  and  $r = (r_1, r_2, \dots, r_M)$  are the intensity values in  $F(g(x; \mu))$  and  $R(x)$ . Also,  $M$  is bins number. Also,  $p(f_j | r_i)$  is the conditional probability. JAD and  $A_\alpha(\omega)$  are substituted into (19). In consequence, the dissimilarity measure  $D$  in (19) is rewritten as

$$D(F(g(x;\mu)), R(x)) = 1 - \left\{ \sum_{i=1}^M \left[ \sum_{j=1}^M p(r_i, f_j)^\alpha \right]^{\frac{1}{\alpha}} - \left[ \sum_{j=1}^M p(f_j)^\alpha \right]^{\frac{1}{\alpha}} \right\} / \left( 1 - M^{\frac{1-\alpha}{\alpha}} \right) \quad (21)$$

where  $M$  is the number of bins. To address the problem of nonrigid registration, the smooth deformation needs to be acquired by regularizing the deformation model. Incorporated with the regularization term, the objective function  $E$  is rewritten by

$$E(F(g(x;\mu)), R(x)) = D(F(g(x;\mu)), R(x)) + \lambda S(g(x;\mu)) \quad (22)$$

where  $D$  is the dissimilarity measure defined in (21), and  $S$  is the regularization term, with its expression given as follows.

$$\begin{aligned} S(g(x;\mu)) &= \frac{1}{V} \int_0^X \int_0^Y \int_0^Z \left[ \left( \frac{\partial^2 g(x;\mu)}{\partial x^2} \right)^2 + \left( \frac{\partial^2 g(x;\mu)}{\partial y^2} \right)^2 + \left( \frac{\partial^2 g(x;\mu)}{\partial z^2} \right)^2 \right. \\ &\quad \left. + 2 \left( \frac{\partial^2 g(x;\mu)}{\partial x \partial y} \right)^2 + 2 \left( \frac{\partial^2 g(x;\mu)}{\partial x \partial z} \right)^2 + 2 \left( \frac{\partial^2 g(x;\mu)}{\partial y \partial z} \right)^2 \right] dx dy dz, \end{aligned} \quad (23)$$

However, objective function shown in (22) does not take into account the spatial information between voxels. To deal with the issue, the distance between two gradient distributions  $q(\nabla F(g(x; \mu)))$  and  $p(\nabla R(x))$  displayed in (14) is introduced to (22). As a result, the nonrigid registration process is expressed by

$$\begin{aligned} \mu^* &= \underset{\mu}{\operatorname{argmin}} E(F(g(x;\mu)), R(x)) \\ &= \underset{\mu}{\operatorname{argmin}} \{ D(F(g(x;\mu)), R(x)) + \lambda_1 S(g(x;\mu)) + \lambda_2 \operatorname{KLD}(\nabla F(g(x;\mu)) || \nabla R(x)) \} \end{aligned} \quad (24)$$

where KLD represents gradient distribution distance, as well as  $\lambda_1$  and  $\lambda_2$  being weight parameters, balancing the tradeoff among a dissimilarity measure  $D$ , a regularization  $S$ , and a distance term KLD.

### 3.4. Optimization

Newton–Raphson algorithm has been widely exploited, in which second-order derivatives can show better convergence [23] compared with these strategies based on first-order gradient. L-BFGS [24]



does not calculate second-order information. Thus, a high computation efficiency can be achieved. A second-order Taylor approximation [16] of  $E$  with respect to  $\mu$  is given as

$$E(\mu + \Delta\mu) \approx E(\mu) + \Delta\mu^T \cdot \nabla E(\mu) + \frac{1}{2} \Delta\mu^T \cdot \nabla^2 E(\mu) \cdot \Delta\mu, \tag{25}$$

where  $\Delta\mu$  is the increment of  $\mu$ ,  $\nabla$  is gradient operation. The deformation parameter  $\mu$  of the L-BFGS optimization algorithm is updated as

$$\mu^{(k+1)} = \mu^{(k)} - (H^{(k)})^{-1} \cdot \nabla E(\mu^{(k)}), \tag{26}$$

In the sequel, the derivative of objective function  $E$  with respect to  $\mu$  need to computed.

$$\frac{\partial E}{\partial \mu} = \left[ \frac{\partial E}{\partial \mu_1}, \frac{\partial E}{\partial \mu_2}, \dots, \frac{\partial E}{\partial \mu_n} \right], \tag{27}$$

The pseudo code of our registration approach is displayed in Algorithm 1.

To solve the optimization process, we need calculate the analytical gradient of the objective function  $E$ . Traditionally, the probability distributions expressed in (21) was not continuous. Hence, the continuous probability density function (pdf) needs to be estimated by Parzen-window method. The continuous marginal and joint pdfs of two images to be registered have been calculated [15]. The continuous expression of gradient distribution distance has been also provided in Section 2.4. Equalization (18) is substituted into (23), the continuity of smoothness term is acquired. Consequently, the objective function  $E$  is continuous and its analytical derivative with respect to  $\mu$  can be calculated.

---

**Algorithm 1.** Nonrigid medical image registration with gradient distributions

---

**Input:** Reference image  $R$ , floating image  $F$

**Output:** Optimal deformation parameters  $\mu^*$

**Set**  $\lambda_1, \lambda_2, NMAX, \alpha, M, N, \delta, \varepsilon$

**Compute** the gradient of  $R$ , denote as  $\nabla R(x)$  and gradient distributions  $p(\nabla R(x))$

**Initialize** deformation parameters  $\mu^{(0)}$ , iteration  $k = 0$ ,  $F(g(x; \mu^{(0)})) = F$ ,  $E(\mu^{(0)}) = 0$

**While**  $|E(\mu^{(k+1)}) - E(\mu^{(k)})| > \text{threshold } \varepsilon$  or  $k < = NMAX$

Obtain the deformed float image  $F(g(x; \mu^{(k+1)}))$  and the regularization  $S(g(x, \mu^{(k+1)}))$

Compute  $\nabla F(g(x; \mu^{(k+1)}))$  and gradient distributions  $q(\nabla F(g(x; \mu^{(k+1)}))$

Estimate the dissimilarity measure  $D$  and gradient distributions distance  $KLD$

Calculate objective function  $E(\mu^{(k+1)}) = D(R(x), F(g(x; \mu^{(k+1)}))) + KLD(q^{(k+1)} | | p) + S(g(x, \mu^{(k+1)}))$

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

$k = k + 1$

**end**

---

### Derivative of the Objective Function

The objective function defined in (24) includes dissimilarity measure  $D$ , a regularization term  $S$ , and a distance term  $KLD$ . The derivative of  $D$  is deduced as

$$\frac{d[D(F(g(x;\mu)), R(x))]}{d\mu} = -\frac{1}{A_\alpha(\omega)} \frac{d[JA_\alpha(F(g(x;\mu)), R(x))]}{d\mu} \tag{28}$$

According to (20), we obtain

$$\frac{d[D(F(g(x;\mu)), R(x))]}{d\mu} = \frac{1}{\left(1 - M^{\frac{1-\alpha}{\alpha}}\right)} \cdot \left\{ \sum_i \sum_j Y \frac{\partial \tilde{p}(r_i, f_j)}{\partial \mu} \right\} \tag{29}$$



$$Y = \left( \sum_j \tilde{p}(f_j)^\alpha \right)^{\frac{1}{\alpha}-1} \tilde{p}(f_j)^{\alpha-1} - \left( \sum_j \tilde{p}(f_j|r_i)^\alpha \right)^{\frac{1}{\alpha}-1} \tilde{p}(f_j|r_i)^{\alpha-1} \tag{30}$$

where  $\partial \tilde{p}(f_j, r_i) / \partial \mu$  represents derivative of estimated joint probability. The derivative of  $p(f_j, r_i; \mu)$  is calculated by

$$\begin{aligned} \frac{\partial p(f_j, r_i; \mu)}{\partial \mu} = & -\frac{1}{N \cdot \Delta b_F} \sum_{x \in \Omega} \beta^{(0)} \left( r_i - \frac{R(x) - R^0}{\Delta b_R} \right) \\ & \times \beta'^{(3)} \left( f_j - \frac{F(g(x; \mu)) - F^0}{\Delta b_F} \right) \times \left( \frac{\partial F(s)}{\partial s} \Big|_{s=g(x; \mu)} \right) \times \frac{\partial (g(x; \mu))}{\partial \mu} \end{aligned} \tag{31}$$

with  $\beta^{(0)}$  and  $\beta'^{(3)}$  being the zero-order B-Splines and the derivative of the three-order B-Splines, respectively.  $R^0$  and  $F^0$  are the minimal intensities in  $R(x)$  and  $F(g(x; \mu))$ , as well as  $\partial F(t) / \partial t$  being the gradient of the deformed float image  $F(g(x; \mu))$ ,  $\partial (g(x; \mu)) / \partial \mu$  can be estimated by FFD model.

To obtain the derivative of the penalty term  $S$ , we rewrite (23) as

$$S(g(x; \mu)) = \frac{1}{V} \sum_x \sum_{i,j=1}^3 \left( \frac{\partial^2 g(x; \mu)}{\partial x_i \partial x_j} \right)^2 \tag{32}$$

where  $x$  represents the points in image region, and  $V$  denotes the number of pixels. The derivative of  $S$  has been provided by Staring and Klein [25],

$$\frac{\partial S(g(x; \mu))}{\partial \mu} = \frac{1}{V} \sum_x \sum_{i,j=1}^3 2 \left( \frac{\partial^2 g(x; \mu)}{\partial x_i \partial x_j} \right) \frac{\partial}{\partial \mu} \frac{\partial^2 g(x; \mu)}{\partial x_i \partial x_j} \tag{33}$$

where  $\partial^2 g(x; \mu) / \partial x_i \partial x_j$  denotes Hessian matrix of deformation function  $g(x; \mu)$ ,  $\partial (\partial^2 T / \partial x_i \partial x_j) / \partial \mu$  is the Jacobi of Hessian matrix.

Next, we calculate the derivative of KLD,

$$\frac{d[\text{KLD}(\nabla F(g(x; \mu)) || \nabla R(x))]}{d\mu} = \sum_d \sum_x \left( 1 + \log \frac{\tilde{q}_d(f_j)}{\tilde{p}_d(r_i)} \right) \frac{\partial \tilde{q}_d(f_j)}{\partial \mu} \tag{34}$$

where  $\tilde{p}_d(r_i)$  and  $\tilde{q}_d(f_j)$  represent the gradient distributions of the transforms float image and reference image, respectively. The derivative of  $\tilde{q}_d(f_j)$  is calculated as

$$\begin{aligned} \frac{\partial \tilde{q}_d(f_j)}{\partial \mu} = & -\frac{1}{V \cdot \Delta b_F} \sum_{x \in \Omega} \cdot \beta'^{(3)} \left( f_j - \frac{\nabla F_d(g(x; \mu)) - \nabla F_d^0}{\Delta b_F} \right) \\ & \cdot \left( \nabla^2 F_d(s) \Big|_{s=g(x; \mu)} \right) \cdot \frac{\partial (g(x; \mu))}{\partial \mu} \end{aligned} \tag{35}$$

where  $\beta'^{(3)}$  is derivative of three-order B-spline, and  $\nabla^2 F_d(t)$  represents the second-order gradient of  $F(g(x; \mu))$ , as well as  $\partial (g(x; \mu)) / \partial \mu$  being the derivative of deformation function  $g(x; \mu)$  with respect to the parameter  $\mu$ . Substituting (12), (13), and (35) into (34), we can obtain the derivative of gradient distribution distance. In the terms of (29), (33), and (34), the derivative of  $E$  will be easily calculated.

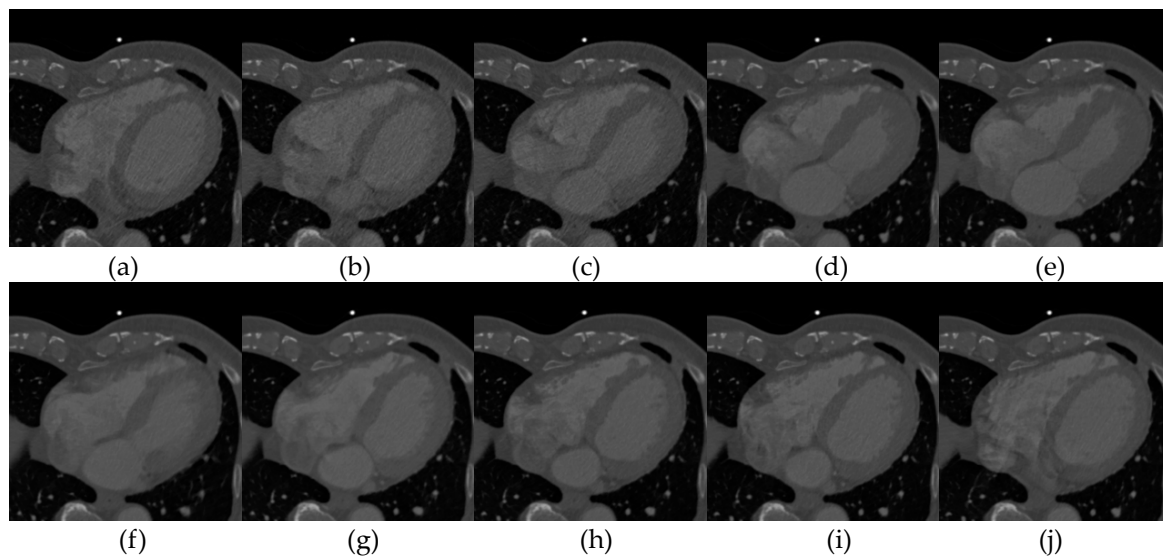
#### 4. Experiments and Results

To evaluate the registration method using the normalized JAD with gradient distribution (NJAD-GD), we designed several groups of tests and performed on simulated and real 3D data, respectively. In Section 4.1, the experimental data is depicted, including simulated and real medical images. The non-rigid registration of simulated MR volumes is performed in Section 4.2. The tests on real 3D thoracic CT images and 3D cardiac data are implemented, and the experimental results are

shown in Sections 4.3 and 4.4, respectively. Our nonrigid registration algorithm employing JAD and gradient distributions was implemented in the *elastix* package [26].

#### 4.1. Experimental Data

In this paper, simulated brain MR volumes, thoracic CT volumes and real 3D cardiac CT images were exploited as experimental data. The detailed descriptions of brain MR and 3D thoracic CT images have been reported in [15]. Additionally, non-rigid tests were also performed on twelve 4D cardiac CT sequences acquired from twelve patients. Each of 4D CT sequence consists of 10 3D cardiac CT images, which were obtained from one whole cardiac cycle of one patient. These CT images have  $256 \times 256$  pixels along axial direction. Figure 3 exhibits 10 3D cardiac CT volumes of one 4D CT sequence. It is obviously observed that some elastic deformations are existed between 10 images.



**Figure 3.** The axis slice of 10 3D cardiac CT images in one 4D sequence. (a–j) represent the 10 frames acquired from one whole cardiac cycle of one patient.

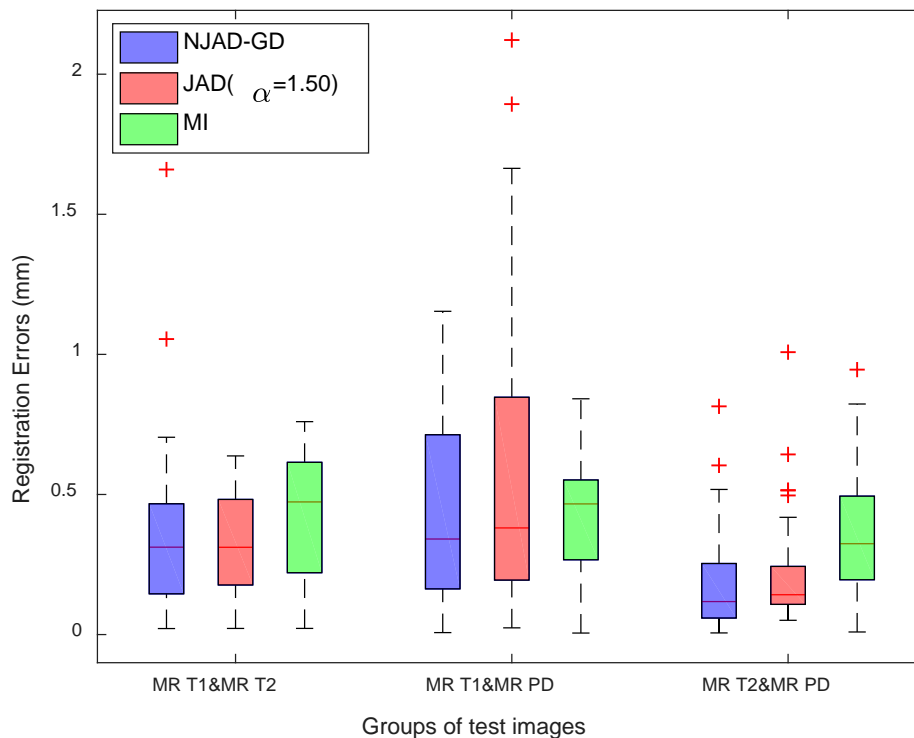
#### 4.2. Nonrigid Registration of Simulated Brain Images

Simulated brain volumes were firstly used to design the elastic alignment experiments. Furthermore, we employ a multiresolution hierarchical strategy with three levels to carry out these non-rigid tests. Also, a comparison with JAD without gradient distribution and MI is also reported.

We selected 60 warping indexes (see parameters  $m$  of the warping function in [15]), which were yielded randomly from the interval [1,7]. Consequently, 60 float images were produced based on the 60 deformations for each pair of test volumes and 540 nonrigid trials of three pairs of brain MR volumes for NJAD-GD, JAD, and MI algorithms in total.

To assess quantitatively test results of these trails, we exploit registration error as the evaluation standard. Here, the registration error is defined as the difference of true values that can be calculated by warping indexes and the obtained values by optimization strategy. In the registration trails of brain images, the involved parameters are set as follows: the nonextensive parameter  $\alpha = 1.5$ , bins  $M = 16$ , the number of random samples  $N = 2000$ ,  $\delta = 20 \times 20 \times 20$ . The weighting parameters  $\lambda_1 = 0.005$  and  $\lambda_2 = 0.001$  can provide a good tradeoff among three terms:  $D$ ,  $S$ , and gradient distribution term KLD in the objective function  $E$ .

Figure 4 shows the test results of all 540 non-rigid registrations. From Figure 4, the NJAD-GD registration algorithm could result in the lower errors of three pairs of test volumes compared to other two approaches.



**Figure 4.** The registration results of the simulated 3D brain MR T1 & MR T2, MR T1 & MR PD, and MR T2 & MR PD volumes using three algorithms. The red color crosses for each box represents these outliers.

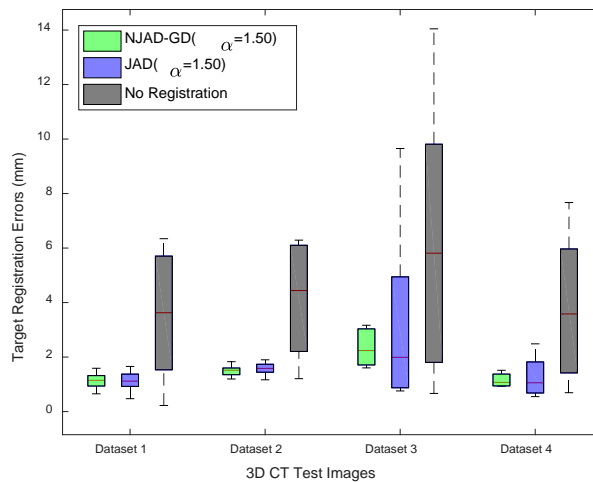
#### 4.3. Experiments of 3D Thoracic CT Images

3D thoracic CT volumes were chosen as the test images to carry out non-rigid registrations. These volumes consist of four 4D sequences, and each of them includes 10 3D volumes. A three-level implementation scheme was still employed to decrease registration accuracy and improve computation efficiency.

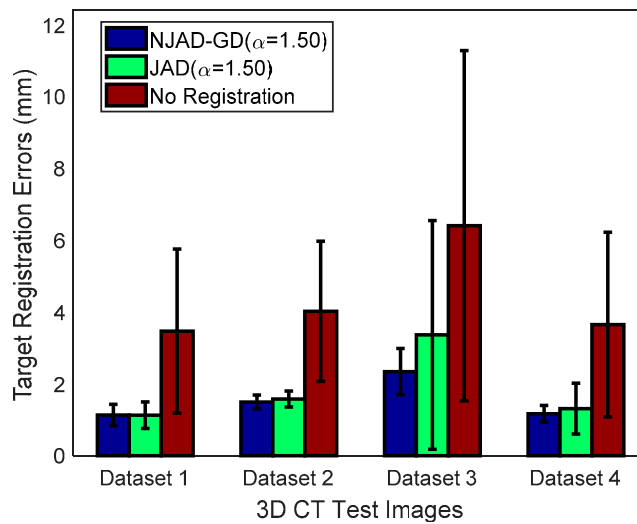
We denote the 10 volumes from each 4D sequence by T00-T90, in which the maximal inhalation and maximum exhalation are included, with indicated by T10 and T60, respectively. Then, we designed the following experiments: in each 4D sequence, the T60 frame is applied as reference image and the residual nine frames are chosen as float image, leading to nine non-rigid tests. Hence, 36 trails of elastic alignments were yielded in total for four 4D CT sequences. We also compared the results adopting NJAD-GD and JAD without considering spatial information. Finally, 72 elastic registration tests were conducted for two methods. In order to quantify the test errors, target registration error (TRE) and Hausdorff distance measure (HDM) were calculated. HDM is a widely-used measure to calculate the distance of two clouds of points. In the 3D thoracic CT registration, the manually marked landmarks can be applied to calculate HDM of two images.

Figures 5 and 6 demonstrate the registration results of 72 tests, along with TREs before registration and after alignment. Figure 7 illustrates the box-and-whisker plots of HDM values of four 4D CT sequences. It is observed from these results that the registration errors applying NJAD-GD algorithm are less than these obtained by the method based on JAD without gradient distribution.

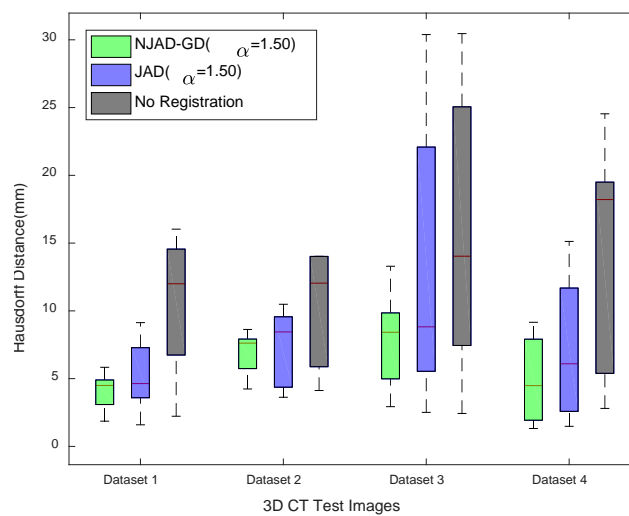
In implementation of experiments, the nonextensive parameter  $\alpha = 1.5$ , bins  $M = 16$ , the number of random samples  $N = 8000$ , the spacing of mesh points  $\delta = 20 \times 20 \times 20$ . The weighting parameters  $\lambda_1 = 0.005$  and  $\lambda_2 = 0.001$ .



**Figure 5.** The TREs obtained when employing NJAD-GD algorithm, the registration method based on JAD without gradient distribution.



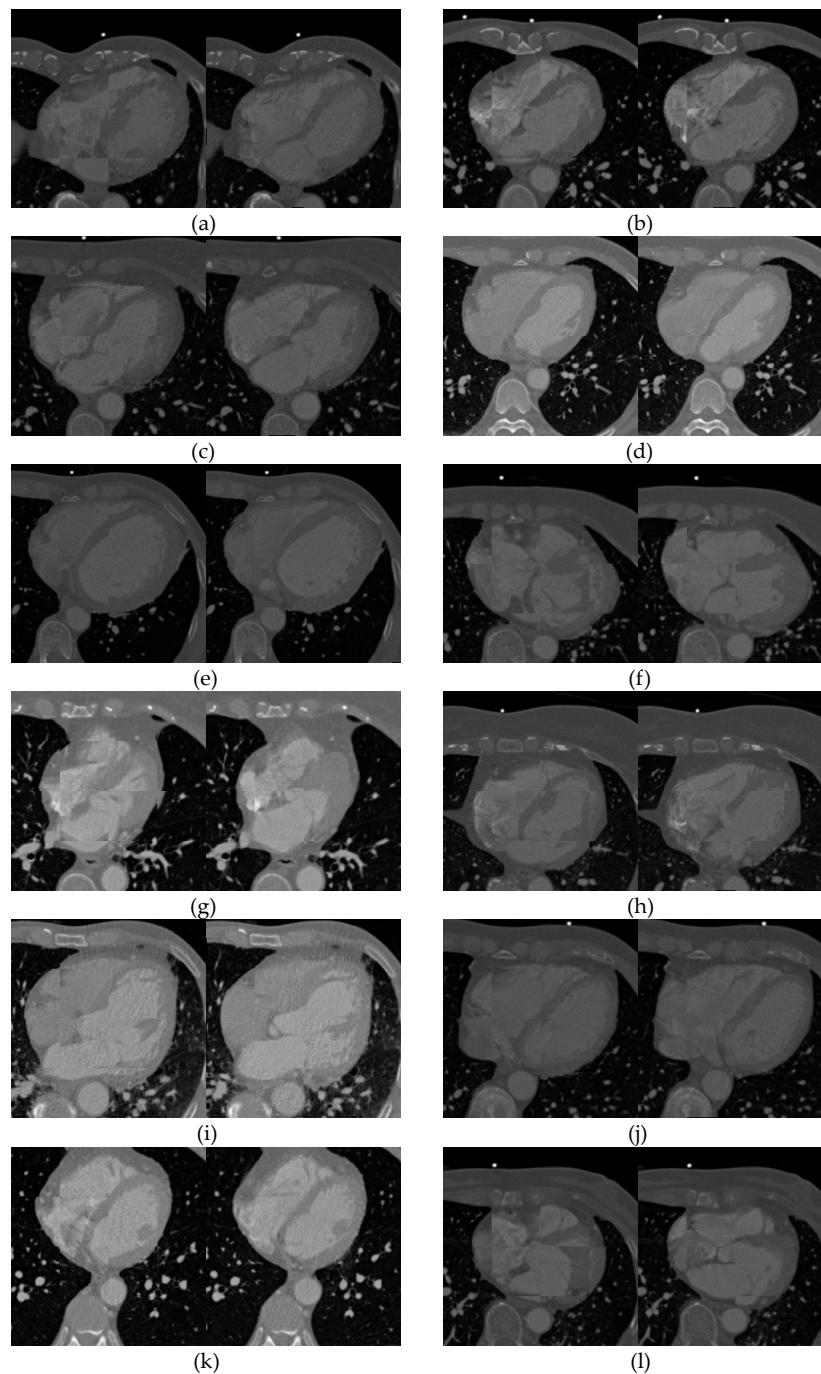
**Figure 6.** Statistics of TREs before registration and after alignment exploiting the NJAD-GD, JAD methods.



**Figure 7.** HDMs obtained when employing NJAD-GD algorithm, the registration method based on JAD without gradient distribution.

#### 4.4. Registration of 3D Cardiac CT Image

The cardiac CT data consists of 12 groups of 4D image sequence, and each of which includes 10 3D images acquired from one whole cardiac cycle. One cardiac cycle consists of the phase of systole and phase of diastole. In each 4D CT sequence, two 3D images with the maximum deformation were employed as the test images, and 12 nonrigid registration experiments were carried out adopting NJAD-GD approach. Figure 8 illustrates the checkboard of 12 examples adopting our non-rigid framework ( $\alpha = 1.50$ ) for 12 4D CT sequences. As it can be seen, the registration algorithm based on gradient distribution demonstrates the accuracy results.



**Figure 8.** Registration results of 12 groups of 3D cardiac images. (a–l) display the test results of patient 1 to 12, respectively. In each group, left image represents the checkboard before registration, and the right accounts for the result after registration.

In these tests, a multiresolution scheme with three levels was also exploited to implement these nonrigid registrations. Due to the large deformation, the number of random samples  $N$  and the spacing of mesh points  $\delta$  were set to 10,000 and  $10 \times 10 \times 10$ , with other parameters being as follows: bins  $M = 16$ , the weighting parameters  $\lambda_1 = 0.005$  and  $\lambda_2 = 0.001$ , the maximum number of iterations of the limited memory BFGS scheme  $NMAX = 200$ .

## 5. Conclusions

In this work, we review the definition and properties of Arimoto entropy, with an information measure based on Arimoto entropy, called JAD. The gradient distributions of reference image and float image are constructed and a distance between them is derived. Additionally, a normalized dissimilarity measure based on JAD was presented. A nonrigid registration method exploiting the normalized measure with gradient distributions is proposed.

Arimoto entropy is regarded as a generalized form of the classical Shannon entropy. In the aforementioned section, it is proofed that the JAD measure is equal to MI when  $\alpha$  tends to 1. We adopted FFDs as the parameter space for non-rigid registration, along with objective function  $E$  including three elements: the normalized JAD as the dissimilarity measure, a regularization to acquire the smooth deformation and a distance term of the gradient distributions.

**Author Contributions:** B.L. implemented the algorithm, validated the experiments, analyzed the data, and wrote the manuscript. B.L., H.S., Z.S., J.H., and M.H. investigated the project. B.L., H.S. and C.L. conceived and revised the manuscript. B.L., Z.L., Z.S., and C.L. acquired the funding. All authors have read and approved the final manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (no. U1804157, no. 61772576, no. 61601311), the Key Natural Science Foundation of Henan Province (no. 162300410338), Science and technology innovation talent project of Education Department of Henan Province (no. 17HASTIT019), the Henan Science Fund for Distinguished Young Scholars (no. 184100510002), Scientific and technological program projects of Henan Province (no. 192102210127, no. 172102210071), Key scientific research projects of Henan Province (no. 19B510011, no. 17A510006), Project of Beijing Excellent Talents (no. 2016000020124G088), Beijing Municipal Education Research Plan Project (no. SQKM201810028018), Program for Interdisciplinary Direction Team in Zhongyuan University of Technology.

**Acknowledgments:** The authors acknowledge the *BrainWeb*: Simulated Brain Database (<http://brainweb.bic.mni.mcgill.ca/brainweb/>) and DIR Validation Data (<https://www.creatis.insa-lyon.fr/rio/popu-model>) for the data used in this paper. Also, the authors would like to thank the editor-in-chief, the handling editors and the anonymous reviewers for their helpful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sadozye, A.H.; Reed, N. A review of recent developments in image-guided radiation therapy in cervix cancer. *Curr. Oncol. Rep.* **2012**, *14*, 519–526. [[CrossRef](#)] [[PubMed](#)]
2. Wang, L.; Gao, X.; Zhou, Z.; Wang, X. Evaluation of four similarity measures for 2D/3D registration in image-guided intervention. *J. Med. Imaging Health Inf.* **2014**, *4*, 416–421. [[CrossRef](#)]
3. Song, G.; Han, J.; Zhao, Y.; Wang, Z.; Du, H. A Review on Medical Image Registration as an Optimization Problem. *Curr. Med. Imaging Rev.* **2017**, *13*, 274–283. [[CrossRef](#)] [[PubMed](#)]
4. Collignon, A.; Maes, F.; Delaer, D.; Vandermeulen, D.; Suetens, P.; Marchal, G. Automated multi-modality image registration based on information theory. *Inf. Process. Med. Imaging* **1995**, *3*, 263–274.
5. Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; Suetens, P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* **1997**, *16*, 187–198. [[CrossRef](#)] [[PubMed](#)]
6. Wells, W.M., III; Viola, P.; Atsumi, H.; Nakajima, S.; Kikinis, R. Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* **1996**, *1*, 35–51. [[CrossRef](#)]
7. Studholme, C.; Hill, D.L.G.; Hawkes, D.J. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recogn.* **1999**, *32*, 71–86. [[CrossRef](#)]
8. Wang, F.; Vemuri, B.C.; Rao, M.; Chen, Y. Cumulative Residual Entropy, A New Measure of Information & its Application to Image Alignment. *IEEE Int. Conf. Comput. Vis.* **2003**. [[CrossRef](#)]



9. Rao, M.; Chen, Y.; Vemuri, B.C.; Wang, F. Cumulative residual entropy: A new measure of information. *IEEE Trans. Inf. Theory* **2004**, *50*, 1220–1228. [[CrossRef](#)]
10. Wang, F.; Vemuri, B.C. Non-rigid multi-modal image registration using cross-cumulative residual entropy. *Int. J. Comput. Vis.* **2007**, *74*, 201–215. [[CrossRef](#)]
11. Antolín, J.; LópezRosa, S.; Angulo, J.C.; Esquivel, R.O. Jensen-tsallis divergence and atomic dissimilarity for position and momentum space electron densities. *J. Chem. Phys.* **2010**, *132*, 131. [[CrossRef](#)]
12. Mohammed, K.; Hamza, A.B. Nonrigid image registration using an entropic similarity. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 681–690. [[CrossRef](#)]
13. Khader, M.; Hamza, A.B. An information-theoretic method for multimodality medical image registration. *Expert Syst. Appl.* **2012**, *39*, 5548–5556. [[CrossRef](#)]
14. Li, B.; Yang, G.; Shu, H.; Coatrieux, J.L. A New Divergence Measure Based on Arimoto Entropy for Medical Image Registration. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014. [[CrossRef](#)]
15. Li, B.; Yang, G.; Coatrieux, J.L.; Li, B.; Shu, H. 3d nonrigid medical image registration using a new information theoretic measure. *Phys. Med. Biol.* **2015**, *60*, 8767–8790. [[CrossRef](#)] [[PubMed](#)]
16. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes in C*, 3rd ed.; Cambridge Univ. Press: Cambridge, UK, 2007; Chapter 10; pp. 521–526.
17. Cover, T.M. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience: New York, NY, USA, 2017.
18. Arimoto, S. Information-theoretical considerations on estimation problems. *Inf. Control* **1971**, *19*, 181–194. [[CrossRef](#)]
19. Boekee, D.E.; Van der Lubbe, J.C. The r-norm information measure. *Inf. Control* **1980**, *45*, 136–155. [[CrossRef](#)]
20. Li, M.; Chen, X.; Li, X.; Ma, B.; Vitanyi, P.M.B. The similarity metric. *IEEE Trans. Inf. Theory* **2004**, *50*, 3250–3264. [[CrossRef](#)]
21. He, Y.; Hamza, A.B.; Krim, H. A generalized divergence measure for robust image registration. *IEEE Trans. Signal Process.* **2003**, *51*, 1211–1220. [[CrossRef](#)]
22. Mattes, D.; Haynor, D.R.; Vesselle, H.; Lewellen, T.K.; Eubank, W. PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging* **2003**, *22*, 120–128. [[CrossRef](#)]
23. Klein, S.; Staring, M.; Pluim, J.P.W. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Trans. Image Process.* **2008**, *16*, 2879–2890. [[CrossRef](#)]
24. Nocedal, J. Updating quasi-newton matrices with limited storage. *Math. Comput.* **1980**, *35*, 773–782. [[CrossRef](#)]
25. Staring, M.; Klein, S. Itk:: Transforms Supporting Spatial Derivatives. Available online: <http://hdl.handle.net/10380/3215> (accessed on 8 September 2010).
26. Klein, S.; Staring, M.; Murphy, K.; Viergever, M.A.; Pluim, J.P.W. Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **2009**, *29*, 196–205. [[CrossRef](#)] [[PubMed](#)]

