

Robust Multi-modal Image Registration Based on Prior Joint Intensity Distributions and Minimization of Kullback-Leibler Distance

Albert C. S. Chung^{a,*}, Rui Gan^a and William M. Wells III^{b,c}

^a*Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,*

The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

^b*Harvard Medical School and Department of Radiology at Brigham and Women Hospital, Boston, MA 02115, U.S.A.*

^c*Computer Science and Artificial Intelligence Laboratory,*

Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Abstract

Robust registration is essential for image-guided therapy as well as structural and functional analysis. In this work, we present a new method that can give superior robustness in multi-modal image registration. This method is based on the *a priori* knowledge of the joint intensity distributions between image pairs at different image resolutions and the Kullback-Leibler distance (KLD) similarity measure. Expected joint distributions are estimated from pre-aligned training images. Two image volumes are registered when the value of KLD is minimized. Two thousand randomized registration experiments on clinical brain CT – T1 image pairs from the Retrospective Image Registration Evaluation (RIRE) project have been performed and evaluated independently by the project. The results demonstrate that, as compared with the conventional Mutual Information (MI)-based method and the Normalized Mutual Information (NMI)-based method, the proposed KLD-based method can significantly increase the registration success rates. To increase the registration accuracy, we further propose a refinement as the last step of the KLD-based method. The refinement step can be based on either MI or NMI, namely the KLD-MI-based method and KLD-NMI-based method respectively. Experimental results on CT – T1 image pairs show that the KLD-MI-based and KLD-NMI-based methods consistently give higher registration accuracy than the KLD-based method. The success rates of the KLD-MI-based and KLD-NMI-based methods are high. In addition, the effects on the performance of KLD-based methods under different histogram bin sizes, intensity inhomogeneity and different noise levels are analyzed using simulated BrainWeb T1 – T2 image pairs.

Key words: Multimodal Image Registration; Joint Intensity Distributions; Kullback-Leibler Distance; Mutual Information; Normalized Mutual Information; Robustness; Accuracy.

1. Introduction

Multi-modal image registration is a key enabling technology that facilitates many applications of medical images [Pluim et al., 2003, Maintz and Viergever, 1998]. Many registration problems have been solved by the use of Minimum Entropy or Maximum Mutual-Information-based methods [Wells et al., 1996, Maes et al., 1997, Pluim et al., 2003, Hajnal et al., 2001]. One of the appeals of these approaches lies in their generality – they do not require the use of domain-

specific modeling. However, as will be demonstrated in this paper, it is likely that more robust solutions to specific applications can be obtained by the use of stronger, domain-specific models. A robust multi-modal image registration method is particularly important for conducting a large-scale image analysis project.

As the use of multi-modal rigid body registration spreads, the number of aligned image pairs that are available in clinical image databases (obtained either by manual or by automatic means) is increasing. These image pairs can serve as training data, from which the statistical joint intensity properties or models can be observed and learned in order to acquire useful *a priori* knowledge for future registration tasks, e.g. mappings for structural and functional

* Corresponding author.

Email addresses: achung@cse.ust.hk (Albert C. S. Chung), raygan@cse.ust.hk (Rui Gan), sw@csail.mit.edu (William M. Wells III).

analysis of brains.

There is a growing body of methods that exploit domain-specific prior models while solving medical image analysis problems [Leventon, 2000, Osher and Paragios, 2003]. For example, given a trained statistical shape model of a target object, the model can be deformed and moved during the registration process so that the results of registration and segmentation can be more robust in the situations where the images are noisy or objects are unclear and subtle [Nikou et al., 1998, Wang and Staib, 2000, Corouge and Barillot, 2001, Vemuri et al., 2003, Tsai et al., 2003b,a, Yang et al., 2003]. However, in these methods, the complex shapes of the target object need to be delineated via segmentation, analyzed, parameterized and modeled, which is non-trivial.

Rather than using statistical shape models, we exploit the use of known probabilistic models of the joint intensity properties among two images in registration. Our method can increase registration robustness but does not require detailed segmentation. It is based on our prior work [Chung et al., 2002, Soman et al., 2003, Gan et al., 2004], which was motivated in part by the work of Leventon and Grimson [Leventon and Grimson, 1998]. The proposed method is a multi-resolution, multi-modal image registration method that uses the *a priori* knowledge of the joint intensity distributions. These joint distributions can be estimated from aligned training images at different resolutions. Our method makes use of the expected joint intensity distribution between two pre-aligned training images as a reference distribution. Two novel images of the same or different acquisitions are aligned when the reference and observed joint intensity distributions are well matched. The difference between distributions is measured using the Kullback-Leibler distance (KLD). The registration procedure is a multi-resolution iterative process. The procedure at the current image resolution is terminated when the KLD value between the observed and reference joint intensity distributions becomes sufficiently small. Then, given the current estimated transformation, the next higher resolution registration continues until the original image resolution is reached.

In this work, the registration performance evaluation is based on experiments that used clinical brain T1-weighted MRI (T1) and X-ray computed-tomography (CT) image volumes, and simulated brain T1-weighted MRI and T2-weighted MRI (T2) image volumes. In total, there are 2000 randomized registration experiments on clinical CT – T1 image pairs from the RIRE project. In this intensive study, all registration results were evaluated independently by the RIRE project. Based on the evaluation results, it is experimentally shown that the KLD-based method can give significantly higher registration success rates than the conventional Mutual Information (MI)-based method [Maes et al., 1997, Wells et al., 1996] and the Normalized Mutual Information (NMI)-based method [Studholme et al., 1999].

Note that the intensity variation between the training and testing image pairs can lead to a slight mismatch be-

tween the expected and observed joint distributions, which can affect the registration performance. To minimize its adverse effect on registration accuracy, after implementing the KLD-based method, we further propose a refinement step based on MI or NMI, namely the KLD-MI-based and KLD-NMI-based methods. The experimental results show that the KLD-MI-based and KLD-NMI-based methods consistently give a higher registration accuracy than the KLD-based method, and obtain high success rates.

The organization of the paper is as follows. In Section 2, we present our registration method. In Section 3, we show the experimental results. We give the conclusion and outline future research directions in Section 4.

2. Methods

2.1. Estimation of the expected and observed joint intensity distributions

Let I_f and I_r be the sets of intensity values of two images acquired from the same or different acquisitions of a patient, where the subscripts f and r represent respectively the floating and reference images. Assume that the intensity values of image voxels are independently and identically distributed (i.i.d.) in space, and let X_f and X_r be the image domains of I_f and I_r respectively.

The joint distribution $\hat{P}(i_f, i_r)$ describes the underlying statistical relationship that we expect to observe on floating and reference images, that is, $i_f \in I_f$ and $i_r \in I_r$ respectively, when they are properly aligned. This *expected* distribution can be estimated from a pre-aligned training image pair that can be obtained from manual registration by experienced clinicians or other automatic image registration methods, for example, a multi-resolution MI-based method.

Given two precisely aligned training image volumes, samples of intensity pairs $\hat{\mathcal{I}} = \{i_f(x), i_r(x_r) \mid i_f \in I_f, i_r \in I_r\}$ can be drawn jointly from I_f and I_r , where x are the grid point coordinates in X_f (this means that the sampling domain is equal to X_f) and x_r are the corresponding coordinates of x in X_r . Partial volume (PV) interpolation is used to obtain smooth histograms [Maes et al., 1997]. The expected joint intensity distribution \hat{P} can be approximated by either Parzen windowing or histogramming [Bishop, 1995]. Histogramming is employed in this work because it is simple and efficient when estimates of derivatives are not needed.

To model the statistical relationship of the image intensities at different resolutions, a Gaussian-based multi-resolution (or coarse-to-fine) image representation is used [Burt and Adelson, 1983], namely Gaussian pyramid. Images at the original resolution give full details of the intensity characteristics, while at the lower resolutions the reduced resolution images give less noisy and low intensity variation representations of the original images [Gonzalez and Woods, 2002]. For further illustration, Figures

1a and 1b respectively show slices from T1-weighted MRI image volumes with 3% and 5% noise levels, which were obtained from the BrainWeb Simulated Brain Database [Collins et al., 1998]. The intensity distributions of the two volumes are plotted in Figures 1e and 1f. Note that the intensities are scaled to 32 bins. For comparison, Figures 1c and 1d, 1g and 1h give slices from the lower resolution image volumes and the corresponding intensity distributions. Empirically, hierarchical approach of this sort have demonstrated some resistance to becoming trapped in incorrect local extrema of the objective function, while preserving the accuracy of transformation estimates [Wells et al., 1996].

From Figures 1c and 1d, 1g and 1h, it is observed that the slices and distributions from the lower resolution images at different levels are more similar because of smoothing and reduced resolution, when compared to those of the original images. In the same way, the joint distributions of different image pairs at lower resolution can be more similar than those of the original image pairs. This property can increase the robustness of the proposed registration method with respect to the intensity variation among different image pairs. In our method, the expected joint intensity distribution is estimated and serves as the reference distribution at each resolution for a multi-resolution based optimization, which is described in Section 2.3.

Given a novel testing image pair with a hypothesized transformation T , the *observed* joint intensity distribution $P_o^T(i_f, i_r)$ models the current statistical intensity relationship between images. Note that the observed joint intensity distribution P_o^T is dependent on the transformation T and changes during the registration. At each resolution, samples of intensity pairs $\mathcal{I}_o = \{i_f(x), i_r(T(x)) | i_f \in I_f, i_r \in I_r\}$ can be drawn jointly from I_f and I_r , where x are the coordinates in X_f . If x falls outside the domain of floating image X_f , then we adopt the following convention: an arbitrary constant intensity value in the background of reference image X_r is assigned to i_r . Similar to the estimation of distribution \hat{P} , the histogramming approach is used to estimate the distribution P_o^T .

2.2. Kullback-Leibler distance (KLD)

Given the expected \hat{P} and observed P_o^T joint intensity distributions, the Kullback-Leibler distance (KLD) between the two distributions is given by [Cover and Thomas, 1991]

$$D(P_o^T || \hat{P}) = \sum_{i_f, i_r} P_o^T(i_f, i_r) \log \frac{P_o^T(i_f, i_r)}{\hat{P}(i_f, i_r)}. \quad (1)$$

KLD measures the similarity between two distributions, and gives an asymmetrical, or *directed* distance from the observed distribution P_o^T to the expected distribution \hat{P} [Ripley, 1996]. According to [Cover and Thomas, 1991, Kullback, 1968], the KLD value is non-negative and becomes zero if and only if two distributions are equivalent. Intuitively, and empirically, when the two testing images I_f and

I_r are not perfectly registered, the value of KLD, D , will be positive and relatively large because P_o^T and \hat{P} are not similar, $P_o^T \neq \hat{P}$. On the other hand, if the testing images are well registered, then the value of KLD becomes small or is equal to zero (that is, P_o^T is very similar or equal to \hat{P}). To avoid zero value in \hat{P} and P_o^T , a very small value ($\epsilon = 1.4 \times 10^{-45}$) is added to each bin before histogram normalization is performed.

2.3. Estimation of the optimal transformation \hat{T}

The goal of KLD-based registration is to find an estimate of the optimal transformation \hat{T} . We assume that the similarity between the observed P_o^T and expected \hat{P} joint intensity distributions is maximum when the optimal transformation is reached,

$$\hat{T} = \arg \min_T D(P_o^T || \hat{P}). \quad (2)$$

The proposed KLD-based method is conceptually different from the MI-based registration method, where the optimal transformation is estimated by

$$\hat{T} = \arg \max_T MI(I_f, I_r; T), \quad (3)$$

where

$$MI(I_f, I_r | T) = D(P_o^T(i_f, i_r) || P_o^T(i_f) \cdot P_o^T(i_r)), \quad (4)$$

$P_o^T(i_f)$ and $P_o^T(i_r)$ are the marginal distributions of floating and reference images respectively. According to Equation 4, mutual information is the Kullback-Leibler distance between the joint distribution and the product of its two marginal distributions. The MI-based method encourages functional dependence between the two image random variables, I_f and I_r , by maximizing the information-based discrepancy (the Kullback-Leibler distance) between the joint distribution $P_o^T(i_f, i_r)$ and the product of the two marginal distributions, $P_o^T(i_f) \cdot P_o^T(i_r)$. On the other hand, the proposed KLD-based method determines the transformation T based on the similarity between the expected \hat{P} and observed P_o^T joint intensity distributions. In other words, the optimal transformation is based on the expected outcomes learned from the training data. For a taxonomical review of several commonly used information theoretic similarity measures, see [Zöllei et al., 2003].

In order to accelerate the registration process and ensure the robustness of the proposed method, as mentioned in the previous section, we exploit a multi-resolution approach [Wells et al., 1996, Cole-Rhodes et al., 2003]. Rough estimates of \hat{T} can be derived from downsampled images and used as starting values for optimization at higher resolutions. Then, the fine-tuning of the solution can be derived at the original image resolution. In this paper, the observed and expected joint distributions are estimated at each resolution so that the prior knowledge about the intensity mappings at different resolutions can be better captured. The value of KLD at each resolution is minimized by Powell's

method with a multi-resolution strategy [Press et al., 1992] because it does not require calculations of gradient values and, hence, is simple in terms of implementation. Powell’s method uses Brent’s 1D line minimizations to iteratively search along a set of directions that are appropriate for the local structure of the objective function.

2.4. Refinement

In our method, we expect that the registration quality depends on the similarity between the joint distributions of the training (expected) and testing (observed) image pairs. Due to the non-linear bias field or noise, the induced intensity variation between the training and testing image pairs at the original resolution may affect the registration accuracy because of the slight mismatches between joint distributions.

In order to further improve registration accuracy, a refinement step is introduced as an extension to the proposed algorithm. It can be based on Mutual Information (MI) or Normalized Mutual Information (NMI). The refinement uses the transformation obtained by the aforementioned algorithm based on KLD (see Sections 2.1 – 2.3) as the starting alignment between the reference and floating images at the original resolution. Then, the conventional MI-based or NMI-based methods can be applied to derive the optimal registration, as the last registration step. This refinement is usually completed after a few iterations. As such, the computational complexity is only slightly higher.

In Figure 2, a flow chart is shown that describes the proposed multi-resolution registration method, which consists of the KLD-based registration and the refinement step. Experimental results and further illustration are given in Section 3.3.

3. Results

To evaluate the Kullback-Leibler distance (KLD)-based similarity measure and the proposed registration method, we have performed experiments on a set of clinical brain X-ray computed-tomography (CT) and T1-weighted MRI images (T1), and a set of simulated brain T1-weighted and T2-weighted MRI images (T2). Section 3.1 describes the image datasets, ground truth and sampling region. Section 3.2 then shows the performance comparisons on robustness and accuracy between the proposed method and the conventional MI-based and NMI-based methods. In Section 3.3, we present the experimental results for justifying the algorithmic extension based on the refinement step. Finally, we demonstrate the effects on registration robustness and accuracy under different histogram bin sizes, intensity inhomogeneity and different noise levels in Section 3.4.

3.1. Image datasets, ground truth and sampling region

In the experiments described below, we used a set of real CT – T1 data from the Retrospective Image Registration Evaluation (RIRE) Project. Note that all the T1 images have been rectified by the RIRE project. CT and T1 image volumes were the reference and floating volumes respectively. In general, the size of a CT image volume is $512 \times 512 \times 34$ voxels and the voxel size is $0.65 \times 0.65 \times 4\text{mm}^3$, and a T1 image contains $256 \times 256 \times 26$ voxels of dimensions $1.25 \times 1.25 \times 4\text{mm}^3$. The registration results were examined based on the evaluations of accuracy performed by the RIRE project.

To build expected (training) joint intensity distributions, we determined the *ground truth* for registration experiments as follows. The training image pair was first registered by the multi-resolution MI-based and NMI-based registration methods. For both methods, the evaluations of accuracy, measured as the target registration errors (TREs) in ten volumes of interest (VOIs) [West et al., 1997], were obtained independently from the RIRE project. TRE is defined as the Euclidean distance between the registered target position of the centroid voxel of a VOI and that of the gold standard [West et al., 1997]. After examination, four datasets (Datasets pt-001, pt-003, pt-005 and pt-007) with less than 1mm median TRE values were selected and used in the experiments. Then the corresponding transformations, whose median TRE values were 0.5077mm (for pt-001), 0.7200mm (for pt-003), 0.7807mm (for pt-005) and 0.6179mm (for pt-007) respectively, were used as *ground truth* registrations for estimating the expected \hat{P} joint intensity distributions and performing randomized experiments.

We also used six pairs of BrainWeb T1 and T2 image volumes [Collins et al., 1998] ($181 \times 217 \times 181$ voxels, $1 \times 1 \times 1\text{mm}^3$) to study the effects of different histogram bin sizes, intensity inhomogeneity and different noise levels on the registration performance. For three pairs, the intensity non-uniformity (RF) level was 0%, and the noise levels were 3%, 5%, 7% and 9% respectively. The other two pairs were with 3% noise and 20% and 40% RF respectively. T1 and T2 image volumes were the reference and floating volumes respectively.

We find that, in all training pairs from the RIRE project and BrainWeb, more than 70% and 42% of the total voxels belong to the background region respectively. This makes the joint distribution of the “interesting” foreground voxels have less importance in the entire distribution. Therefore, for estimating the expected joint intensity distribution \hat{P} , samples were generated jointly from the foreground regions alone (brain regions in this application) and the background region was left out. Note that the brain regions can be identified by using the active contours or brain detection techniques [Boesen et al., 2004, Brummer et al., 1993]. To preserve the original intensity range, the maximum and minimum of the expected joint intensity distribution were the

entire image volume maximum and minimum respectively before the background region was left out. With regard to the estimation of the observed joint intensity distribution P_o^T , since the input images are unknown, the background region was not left out in the testing image pairs. As such, the entire testing image volumes were included in the registration process. Figures 3a and 3b show the translation probes of KLD values when the background region is left out and included respectively in estimating the expected \hat{P} joint intensity distribution. Two pairs (pt-003 and pt-005) of CT – T1 RIRE image volumes were used as testing and training image pairs respectively. It is shown that the capture range is significantly longer when the background is excluded in the training process. Note that the background of the testing image pairs is not excluded in the registration process.

3.2. Performance comparisons on robustness and accuracy

In order to study and compare the registration robustness and accuracy of the conventional MI-based and NMI-based methods, and the KLD-based method, randomized experiments were performed on a set of CT – T1 RIRE image pairs. Hereafter the aforementioned three methods are referred to as MI, NMI, KLD respectively.

A total of four CT – T1 image pairs (pt-001, pt-003, pt-005 and pt-007) were used in these experiments and each image pair was treated as either testing or training image pairs. Therefore, there were 20 experiments in total (i.e., for each testing image pair, there were 1 experiment for MI and for NMI respectively, and 3 experiments for KLD) and each experiment took 100 trials. At each trial, the *ground truth* registration parameters of the testing image pair (see Section 3.1) were perturbed by six uniformly distributed random offsets for all translational and rotational axes. The perturbed parameters were then treated as the starting alignment. In order to show the high registration capability of KLD with respect to initial alignment, random offsets for X and Y translational axes were drawn between around [-150, 150]mm and random offsets for Z translational axis were drawn between around [-70, 70]mm. Note that these ranges were set so that two brains in CT and T1 images have at least 10% overlapping region. While, random offsets for each rotational axis were respectively drawn between [-0.52, 0.52] radians (i.e., [-30, 30] degrees). As a fair comparison, for any testing image pair, the same set of randomized starting alignments was used for different methods.

For implementation, four resolution levels are used for all the three methods (MI, NMI and KLD). The definition of resolution levels in the Gaussian Pyramid representation follows the same convention as in [Burt and Adelson, 1983], that is, Level 0 image represents the highest and original resolution and Level 3 image represents the lowest resolution. Smoothing was performed via the binomial filter with coefficients [1, 4, 6, 4, 1] [Wells et al., 1996]. For the ease of implementation, all voxels in the downsampled floating vol-

umes (i.e., T1 images) were used at Levels 1 – 3. At Level 0, 1/4 (25%) of all voxels were sampled (one voxel randomly picked from every 2×2 matrix in each slice). In the experiments for both MI and NMI, the image intensity values were linearly scaled to 64 bins, which have been used in the conventional MI-based and NMI-based methods [Studholme et al., 1999, Knops et al., 2006, Zhu and Cochoff, 2002], and the same number of bins was used for comparing the performance between different refinement methods. In the experiments for KLD, to reduce noise and intensity variation, the image intensity values in both testing and training image pairs were scaled to 32 bins by performing histogram normalization based on a histogram transformation function which linearly mapped an input histogram from the input intensity range to a discrete intensity range (e.g. 0 - 31). To be more robust to the outliers and image artifacts, the input intensity maximum and minimum of the histogram transformation function are respectively the median minimum I_{min} and median maximum I_{max} of all datasets in each experiment set. When registering a novel testing image pair, the same median values, I_{min} and I_{max} , are used for histogram normalization. For the Brent and Powell optimization methods, the fractional precision convergence parameters were set to 10^{-3} and 10^{-4} respectively.

With regard to the registration experiments, we examined the results based on the evaluations of accuracy measured as TREs in ten VOIs, which were obtained independently from the RIRE project. To evaluate each derived registration, the median TRE value was then taken for assessing registration success. A registration was judged to be successful if the median TRE value was smaller than 4mm, which was the largest voxel dimension of the CT – T1 image pair; otherwise, it was considered a misregistration.

Table 1 lists the success rates for MI, NMI and KLD for all testing image pairs (pt-001, pt-003, pt-005 and pt-007), together with the means and standard deviations of the median TRE values for the *successful* registrations. Note that the results of KLD were grouped based on the testing image pair, by summarizing results from different training image pairs for each testing image pair. For example, if the testing image pair was from dataset pt-001, then the training image pair would be from datasets pt-003, pt-005 and pt-007. Note that the **Total** row summarizes the results from all testing image pairs. (The detailed results for KLD will be presented in Section 3.3.) It is shown in Table 1 that KLD gives the highest success rates as compared with MI and NMI.

For registration accuracy, it is observed that, across different testing image pairs, the median TRE values of the successful registrations for KLD are slightly larger than those for MI and NMI. But the registration accuracy of KLD can be improved by way of a refinement step (see next section for details).

3.3. Justification of the refinement step

As described in Section 2.4, the quality of the KLD-based registration method can be further improved if a refinement step is added as the last step of the registration process. To justify the extension, we have performed a series of similar randomized experiments described Section 3.2 on CT – T1 registration with the KLD-based method and the extended methods respectively. We have considered two possible refinement steps based on MI (hereafter referred to as KLD-MI) and NMI (hereafter referred to as KLD-NMI). The same image pairs (pt-001, pt-003, pt-005 and pt-007) obtained from the RIRE project were tested. Therefore, there were 12 experiments on CT – T1 registration for KLD-MI (i.e., 3 experiments for each testing image pair), and 12 experiments for KLD-NMI. Note that each experiment took 100 trials, and each testing image pair used the same set of randomized starting alignments as those in Section 3.2.

For each derived solution, the registration was judged to be successful using the same criterion ($< 4\text{mm}$) based on the median TRE values obtained from the RIRE project. As listed in Table 2, the **KLD**, **KLD-MI** and **KLD-NMI** rows show the breakdown success rates and registration accuracies when different training pairs (the columns **pt-001**, **pt-003**, **pt-005** and **pt-007**) were used for each testing pair. In all experiments, training and testing pairs were different for each test. The column **All** lists the combined results when different testing pairs were used for KLD, KLD-MI and KLD-NMI, and the **Total** row summarized the results from all testing image pairs.

For registration accuracy, after either one of the refinement steps was implemented, it is shown in Table 2 (column **All** and row **Total**) that the median TRE values of KLD have been reduced for all testing image pairs. In addition, after refinement, KLD-MI and KLD-NMI can achieve the accuracy as high as those of MI and NMI respectively, as listed in Table 1. We would like to note that since neither MI-based nor NMI-based refinement step has improved the success rates for pt-001, pt-003, pt-005 and pt-007, then the success rates of KLD-MI and KLD-NMI are not shown in the table for a clear representation. The table reveals that (1) neither one of the refinement methods made any originally successful registration obtained from KLD unsuccessful; (2) neither one of the refinement methods could correct the misregistrations that had large median TRE values.

Table 2 also presents the detailed results for each testing image pair when different training pairs were used. It is observed that, for all testing image pairs, there was only slight variation of the success rates and registration accuracy of KLD, KLD-MI and KLD-NMI, while changing the training image pairs.

With regard to the registration speed, after averaging all experiments presented above (i.e., in Sections 3.2 and 3.3), KLD, KLD-MI and KLD-NMI respectively took 151s, 193s and 192s to register a CT – T1 RIRE image pair on a 3.2 GHz PC with 1.00 GB RAM. On the other hand, MI and

NMI took 148s and 143s respectively.

3.4. Effects on robustness and accuracy under different histogram bin sizes, intensity inhomogeneity and noise

To study the effects on registration performance of KLD under different histogram bin sizes, intensity inhomogeneity and different noise levels, extra sets of randomized experiments were performed on five pairs of BrainWeb T1 and T2 image volumes. The image pair with 3% noise level and 0% RF level was used as the training image pair, and the testing image pairs were the other five image pairs (see Section 3.1 for a detailed description). As the BrainWeb T1 – T2 image pairs have already been aligned precisely, the ground truth registrations were the null transformation. Note that each experiment took 100 trials and the randomized initial alignments were generated as described in Section 3.2.

To evaluate each derived registration with respect to the corresponding ground truth registration (i.e., the null transformation), similar to Maes et al. and Knops et al. [Maes et al., 1997, Knops et al., 2006], a tight bounding box was fitted around the brain for each T2 images. For each of the eight corner points, the Euclidean distance between the ground truth position and the position transformed by our solution was computed. The median value of the eight distances was then taken for assessing registration success. A registration was judged to be successful if the median error was smaller than 1mm, which was the largest voxel dimension of the BrainWeb T1 – T2 image pair; otherwise, it was considered a misregistration.

The experiments for different histogram bin sizes (i.e., 16, 32 and 64) took the T1 – T2 image pair with 5% noise level and 0% RF level as the testing image pair. Table 3 lists the success rates of KLD and the means and standard deviations of the median errors based on the 1mm criterion as described above. It is shown that, compared to the 1mm criterion, the registration accuracies of KLD are high and comparable between 16, 32 and 64 bins. The robustness of KLD increases from 93% to 98% when the number of bins increases from 16 to 32, and drops slightly from 98% to 96% when the number of bins further increases from 32 to 64. Therefore, for KLD, 32 bin histogram configuration was used in this paper.

To study the effects under intensity inhomogeneity, two T1 – T2 image pairs with 3% noise level and different RF levels (i.e., 20% and 40%) were used as the testing image pairs. The evaluation results (based on the 1mm criterion) of KLD are shown in Table 4, together with those of KLD-MI as a comparison. Note that the success rates of KLD-MI are exactly the same as those of KLD. It is observed that the robustness and accuracy decrease slightly when the RF level increases.

Three T1 – T2 image pairs with different noise levels (i.e., 5%, 7% and 9%) and 0% RF level were used as the testing image pairs in the experiments for studying the

effects under different noise levels. Table 5 presents the evaluation results (based on the 1mm criterion) of KLD and KLD-MI. From the table, an observation is that the success rates and accuracies of KLD and KLD-MI are both high but decrease when the noise level increases from 5% to 7%.

In Table 5, it is noticed that when the noise level increases to a high value, i.e. 9%, the success rate drops significantly to 25% for both KLD and KLD-MI, and the accuracy also decreases. The intensity distributions of the T1 image volumes with 9% noise level at the original resolution and the resolution Level 2 are plotted in Figures 4c and 4d respectively, and the image slices at the original resolution and Level 2 are shown in Figures 4a and 4b respectively. By comparing between Figure 1e (training pdf, 3% noise level) and Figure 4c, and between Figure 1g (training pdf, 3% noise level) and Figure 4d, it is observed that there is a relatively large discrepancy between the testing and training joint intensity distributions. Using the background class labels for the T1 images available on the BrainWeb website, the background regions were selected from the images with 3% and 9% noise levels at the original resolution level. The estimated background means and SDs were 5.88 ± 3.94 and 14.38 ± 7.62 for 3% and 9% noise levels respectively (intensity ranges from 0 to 255). This shift of the background intensity with a large number of voxels is one of the causes leading to a relatively large discrepancy between the testing and training joint intensity distributions. However, intensity shift of such large number of voxels was not observed in the other clinical datasets from the RIRE project. The results listed in Table 1 and Table 2 show that the success rates of KLD are consistently high. As a comparison, similar to the BrainWeb simulated image volumes, large VOIs in the RIRE T1 clinical image volumes were selected. The estimated background means and SDs were relatively consistent: 33.36 ± 15.61 for pt-001, 31.53 ± 14.34 for pt-003, 21.86 ± 10.10 for pt-005 and 27.01 ± 12.34 for pt-007 (intensity ranges from 0 to $2^{12} - 1$).

4. Conclusion and Future Work

This paper has proposed a multi-resolution, multi-modal image registration method that is based on minimizing the Kullback-Leibler distance (KLD) between the observed and expected joint intensity distributions. An extension of the proposed KLD-based method has also been presented. We have performed experiments on brain T1 and CT datasets from the RIRE project. The estimated transformations were evaluated independently by the RIRE project. The results show that the KLD-based method has a higher success rate (95.17%) than the Mutual Information (MI)-based method (70.50%) and the Normalized Mutual Information (NMI)-based method (13.75%). The MI-based method and NMI-based method give the accuracy of $1.0172\text{mm} \pm 0.3730\text{mm}$ and $0.7961\text{mm} \pm 0.3111\text{mm}$ respectively. However, the KLD-based method gives slightly lower accuracy

($1.4094\text{mm} \pm 0.5881\text{mm}$) than MI-based and NMI-based methods.

Non-linear bias fields or noise can bring about an adverse effect in registration accuracy because of the slight mismatches between joint intensity distributions. Therefore, a refinement has been suggested as the last step of the KLD-based registration process. The refinement step can be based on either MI-based or NMI-based method, namely the KLD-MI-based and KLD-NMI-based methods respectively.

The experimental results of the randomized registration tests on T1 and CT datasets from the RIRE project demonstrate that the accuracies of the KLD-MI-based method ($0.9319\text{mm} \pm 0.2418\text{mm}$) and KLD-NMI-based method ($0.7565\text{mm} \pm 0.2488\text{mm}$) are higher than those of KLD-based method ($1.4094\text{mm} \pm 0.5881\text{mm}$), and there is no difference in success rates (95.17%) between KLD-based, KLD-MI-based and KLD-NMI-based methods.

Experiments have been also carried out using a set of simulated brain T1 and T2 images from the BrainWeb Simulated Brain Database. The results show that the refinement step can improve the registration accuracy under intensity inhomogeneity and noise. For example, when the RF level was 40%, the registration accuracy increased from $0.0255\text{mm} \pm 0.0305\text{mm}$ to $0.0094\text{mm} \pm 0.0057\text{mm}$ after the MI-based refinement step was implemented. Similarly, when the noise level was 7%, the accuracy increased from $0.0601\text{mm} \pm 0.0637\text{mm}$ to $0.0141\text{mm} \pm 0.0053\text{mm}$ after the refinement step was implemented. When the noise level increased from 3% to a high value, i.e. 9%, it was observed that the mismatch between the expected and training joint intensity distributions became large, which can lead to a lower image registration quality in terms of the robustness and accuracy.

Future work will include a further study on the inclusion of useful features in the training process, e.g. edges, ridges, local patterns, or a combination of these cues and so on. Finally, we will investigate and compare different ways to combine information extracted from several training pairs or pre-aligned atlas images for constructing the expected joint intensity distributions.

Acknowledgements

The clinical images in our experiments were provided as part of the project, ‘‘Evaluation of Retrospective Image Registration’’, National Institutes of Health, Project Number 8R01EB002124-03, Principle Investigator, J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN. We would like to thank Dr. J. Michael Fitzpatrick and Ms. Ramya Balachandran for their great help in evaluation of registration accuracy of a very large number of estimated transformations. Also, we would like to thank Jue Wu for his help in the program implementation and testing in the early stages of this work, and Eric Grimson, Alexander Norbash, Salil Soman, Jue Wu and Simon Yu for useful dis-

cussions. W. Wells would like to acknowledge support from an NSF ERC grant, CISST (JHU Agreement #8810-274) and the NIH (grant #1P41RR13218). A. Chung would like to acknowledge support from the Research Grants Council (RGC) of Hong Kong under Grant HKUST6155/03E.

References

- C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford U. Press, 1995.
- K. Boesen, K. Rehm, K. Schaper, S. Stoltzner, R. Woods, E. Lüders, and D. Rottenberg. Quantitative comparison of four brain extraction algorithms. *NeuroImage*, 22:1255–1261, 2004.
- M.E. Brummer, R.M. Mersereau, R.L. Eisner, and R.R.J. Lewine. Automatic Detection of Brain Contours in MRI Data Sets. *IEEE Transactions on Medical Imaging*, 12(2):153–166, 1993.
- P.J. Burt and E.H. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540, 1983.
- A.C.S. Chung, W.M. Wells III, A. Norbash, and W.E.L. Grimson. Multi-Modal Image Registration by Minimising Kullback-Leibler Distance. In *The 5th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'02)*, pages 525–532, 2002.
- A.A. Cole-Rhodes, K.L. Johnson, J. LeMoigne, and I. Zavorin. Multiresolution Registration of Remote Sensing Imagery by Optimization of Mutual Information Using a Stochastic Gradient. *IEEE Transactions on Image Processing*, 12(12):1495–1511, 2003.
- D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and Construction of a Realistic Digital Brain Phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, 1998.
- I. Corouge and C. Barillot. Use of a Probabilistic Shape Model for Non-linear Registration of 3D Scattered Data. In *International Conference on Image Processing (ICIP'2001)*, pages 149–152, 2001.
- T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.
- R. Gan, J. Wu, A.C.S. Chung, S.C.H. Yu, and W.M. Wells III. Multiresolution image registration based on Kullback-Leibler distance. In *The 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'04)*, 2004.
- R.C. Gonzalez and R.E. Woods. *Digital Image Processing, 2nd Ed.* Prentice-Hall, Inc. New Jersey, 2002.
- J.V. Hajnal, D.L.G. Hill, and D.J. Hawkes. *Medical Image Registration*. CRC Press LLC, 2001.
- Z.F. Knops, J.B.A. Maintz, M.A. Viergever, and J.P.W. Pluim. Normalized mutual information based registration using k-means clustering and shading correction. *Medical Image Analysis*, 10(3):432–439, 2006.
- S. Kullback. *Information Theory and Statistics*. Dover Publications, Inc., 1968.
- M.E. Leventon. *Statistical Models for Medical Image Analysis*. MIT PhD Dissertation, 2000.
- M.E. Leventon and W.E.L. Grimson. Multi-Modal Volume Registration Using Joint Intensity Distributions. In *The 1st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'98)*, pages 1057–1066, 1998.
- F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality Image Registration by Maximization of Mutual Information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- J.B.A. Maintz and M.A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–16, 1998.
- C. Nikou, F. Heitz, and J.P. Armspach. Multimodal Image Registration using Statistically Constrained Deformable Multimodels. In *International Conference on Image Processing (ICIP'1998)*, pages 838–842, 1998.
- S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics (Chapter V)*. Springer-Verlag, New York, Inc., 2003.
- J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.
- W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C, 2nd Ed.* Cambridge University Press, 1992.
- B.D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.
- S. Soman, A.C.S. Chung, W.E.L. Grimson, and W.M. Wells III. Rigid Registration of Echoplanar and Conventional MR Images by Minimizing Kullback-Leibler Distance. In *The 2nd International Workshop on Biomedical Image Registration (WBIR'03)*, pages 181–190, 2003.
- C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.
- A. Tsai, W. Wells, C. Tempary, E. Grimson, and A. Willsky. Coupled Multi-shape Model and Mutual Information for Medical Image Segmentation. In *The International Conference on Information Processing in Medical Imaging (IPMI'03)*, pages 185–197, 2003a.
- A. Tsai, A. Yezzi, W. Wells, C. Tempary, D. Tucker, A. Fan, W.E. Grimson, and A. Willsky. A Shape-Based Approach to the Segmentation of Medical Imagery Using Level Sets. *IEEE Transactions on Medical Imaging*, 22(2):137–154, 2003b.
- B.C. Vemuri, J. Ye, Y. Chen, and C.M. Leonard. Image Registration via Level-set Motion: Applications to Atlas-based Segmentation. *Medical Image Analysis*, 7(1):1–20, 2003.
- Y. Wang and L.H. Staib. Physical Model-based Non-rigid Registration Incorporating Statistical Shape Information. *Medical Image Analysis*, 4(1):7–21, 2000.
- W.M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kiki-

- nis. Multi-Modal Volume Registration by Maximization of Mutual Information. *Medical Image Analysis*, 1(1): 35–51, 1996.
- J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assist. Tomogr.*, 21(4):554–566, 1997.
- J. Yang, L.H. Staib, and J.S. Duncan. Neighbor-Constrained Segmentation with 3D Deformable Models. In *The International Conference on Information Processing in Medical Imaging (IPMI'03)*, pages 198–209, 2003.
- L. Zöllei, J.W. Fisher III, and W.M. Wells III. A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration. In *The International Conference on Information Processing in Medical Imaging (IPMI'03)*, pages 366–377, 2003.
- Y.M. Zhu and S.M. Cochoff. Influence of Implementation Parameters on Registration of MR and SPECT Brain Images by Maximization of Mutual Information. *The Journal of Nuclear Medicine*, 43(2):160–166, 2002.



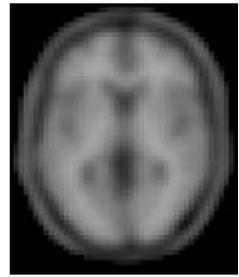
(a)



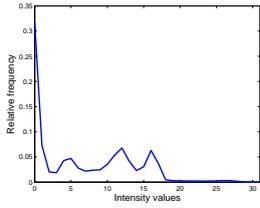
(b)



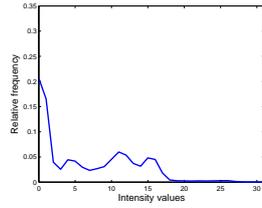
(c)



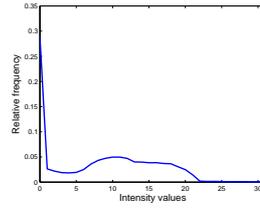
(d)



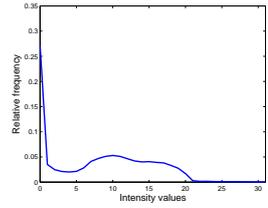
(e)



(f)



(g)



(h)

Fig. 1. (a) and (b) are slices of BrainWeb T1-weighted MRI (T1) image volume with 3% noise level and 5% noise level respectively. (e) and (f) are the intensity distributions of the T1 image volumes shown in (a) and (b) respectively. (c) and (d) are slices from the lower resolution T1 image volume with 3% noise level and 5% noise level respectively. (g) and (h) are the intensity distributions of the lower resolution T1 image volume shown in (c) and (d) respectively.

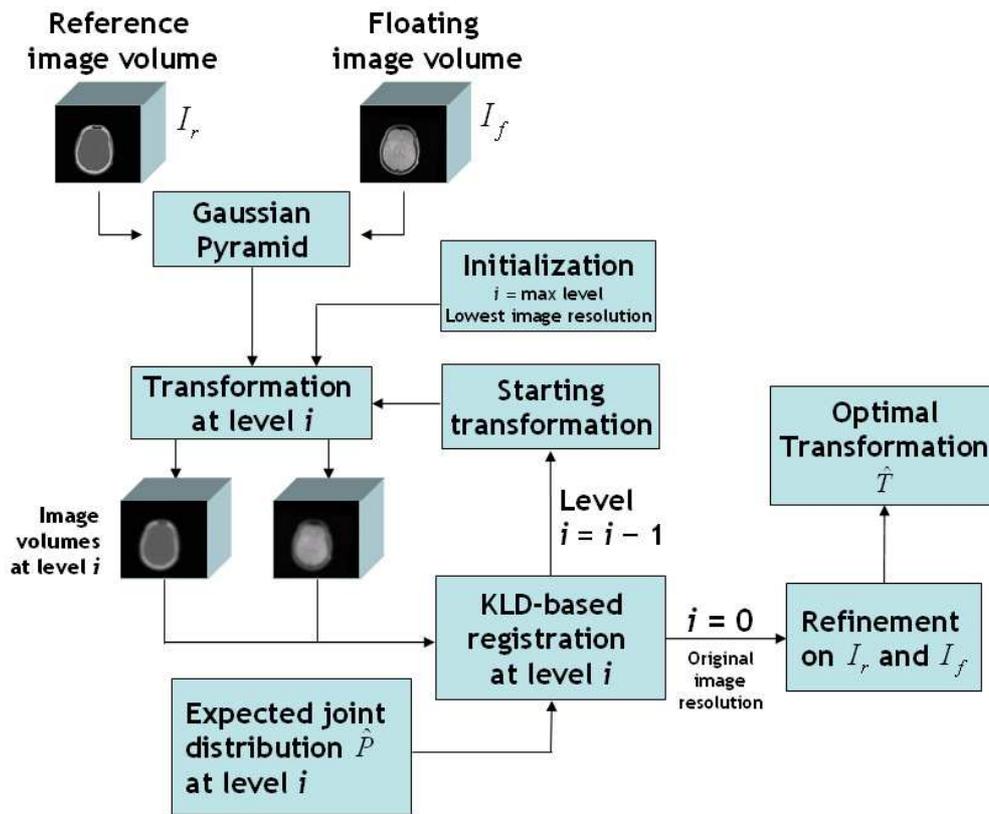
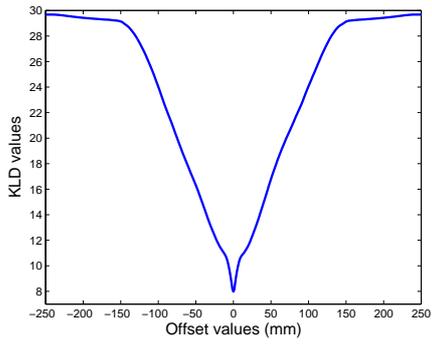
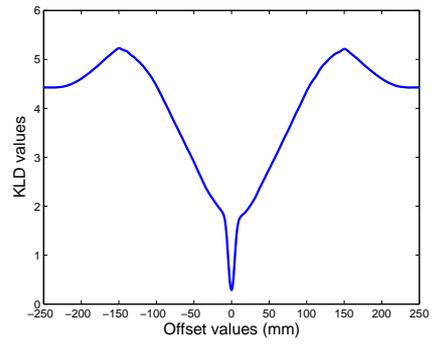


Fig. 2. This chart describes the implementation flow of the proposed multi-resolution KLD-based registration and the refinement step.



(a)



(b)

Fig. 3. The translation probes of KLD values when the background region is left out (a) and included (b) respectively in estimating the expected \hat{P} joint intensity distribution.

Table 1

The success rates of MI, NMI and KLD, and the means and standard deviations of the median TRE values of the successful registrations (based on 4mm criterion) for different testing image pairs, while varying the training image pairs for each testing image pair. Please see Section 3.2 for more details.

Dataset	MI		NMI		KLD	
	success%	mean \pm sd	success%	mean \pm sd	success%	mean \pm sd
pt-001	67.00%	0.7025 \pm 0.0196	13.00%	0.8162 \pm 0.4287	95.67%	1.1067 \pm 0.1250
pt-003	76.00%	0.9041 \pm 0.4431	12.00%	0.7351 \pm 0.0324	95.67%	1.3366 \pm 0.3737
pt-005	72.00%	1.1158 \pm 0.2359	16.00%	1.0480 \pm 0.2620	95.00%	2.0143 \pm 0.8369
pt-007	67.00%	1.3542 \pm 0.2603	14.00%	0.5421 \pm 0.0301	94.33%	1.1813 \pm 0.1254
Total	70.50%	1.0172 \pm 0.3730	13.75%	0.7961 \pm 0.3111	95.17%	1.4094 \pm 0.5881

Table 3

The success rates of KLD with different numbers of bins (i.e. 16, 32 and 64), and the means and standard deviations of the median errors (based on 1mm criterion). The training and testing image pairs are respectively BrainWeb T1 – T2 image pairs with 3% and 5% noise levels. Please see Section 3.4 for more details.

# Bins	KLD	
	success%	mean \pm sd
16	93%	0.0192 \pm 0.0114
32	98%	0.0236 \pm 0.0214
64	96%	0.0183 \pm 0.0089

Table 4

The success rates of KLD and KLD-MI on BrainWeb T1 – T2 image pairs with 3% noise level and different intensity non-uniformity (RF) levels, together with the means and standard deviations of the median errors (based on 1mm criterion). The BrainWeb T1 – T2 image pair with 3% noise level and 0% RF level was taken for training. Please see Section 3.4 for more details.

RF level	KLD		KLD-MI
	success%	mean \pm sd	mean \pm sd
20%	98%	0.0184 \pm 0.0221	0.0059 \pm 0.0058
40%	96%	0.0255 \pm 0.0305	0.0094 \pm 0.0057

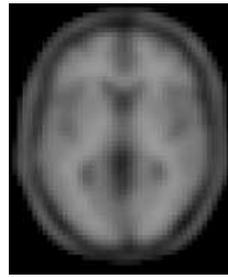
Table 5

The success rates of KLD and KLD-MI on the BrainWeb T1 – T2 image pairs with 0% RF level and different noise levels, together with the means and standard deviations of the median errors (based on 1mm criterion). The BrainWeb T1 – T2 image pair with 3% noise level and 0% RF level was taken for training. Please see Section 3.4 for more details.

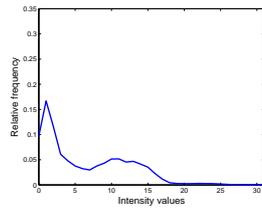
Noise level	KLD		KLD-MI
	success%	mean \pm sd	mean \pm sd
5%	98%	0.0236 \pm 0.0214	0.0089 \pm 0.0060
7%	97%	0.0601 \pm 0.0637	0.0141 \pm 0.0053
9%	25%	0.0942 \pm 0.0360	0.0213 \pm 0.0072



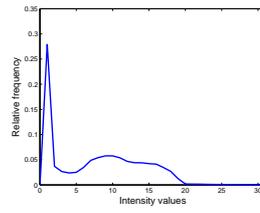
(a)



(b)



(c)



(d)

Fig. 4. (a) shows a slice of BrainWeb T1-weighted MRI (T1) image volume with 9% noise level. (c) is the intensity distribution of the T1 image volumes shown in (a). (b) shows a slice from the lower resolution T1 image volume with 9% noise level. (d) is the intensity distribution of the lower resolution T1 image volume shown in (b).