

# **HHS Public Access**

Author manuscript *Med Image Anal.* Author manuscript; available in PMC 2018 April 12.

Published in final edited form as:

Med Image Anal. 2017 October; 41: 18-31. doi:10.1016/j.media.2017.05.004.

# Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis

Xiaohuan Cao<sup>a,b</sup>, Jianhua Yang<sup>a</sup>, Yaozong Gao<sup>b,c</sup>, Yanrong Guo<sup>b</sup>, Guorong Wu<sup>b</sup>, and Dinggang Shen<sup>b,d,\*</sup>

<sup>a</sup>School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, China

<sup>b</sup>Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>c</sup>Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>d</sup>Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

# Abstract

In prostate cancer radiotherapy, computed tomography (CT) is widely used for dose planning purposes. However, because CT has low soft tissue contrast, it makes manual contouring difficult for major pelvic organs. In contrast, magnetic resonance imaging (MRI) provides high soft tissue contrast, which makes it ideal for accurate manual contouring. Therefore, the contouring accuracy on CT can be significantly improved if the contours in MRI can be mapped to CT domain by registering MRI with CT of the same subject, which would eventually lead to high treatment efficacy. In this paper, we propose a bi-directional image synthesis based approach for MRI-to-CT pelvic image registration. First, we use patch-wise random forest with auto-context model to learn the appearance mapping from CT to MRI domain, and then vice versa. Consequently, we can synthesize a pseudo-MRI whose anatomical structures are exactly same with CT but with MRIlike appearance, and a pseudo-CT as well. Then, our MRI-to-CT registration can be steered in a dual manner, by simultaneously estimating two deformation pathways: 1) one from the pseudo-CT to the actual CT and 2) another from actual MRI to the pseudo-MRI. Next, a dual-core deformation fusion framework is developed to iteratively and effectively combine these two registration pathways by using complementary information from both modalities. Experiments on a dataset with real pelvic CT and MRI have shown improved registration performance of the proposed method by comparing it to the conventional registration methods, thus indicating its high potential of translation to the routine radiation therapy.

### Keywords

Non-rigid registration; Multi-modality; Image synthesis; Radiation therapy

<sup>\*</sup>Corresponding author. dgshen@med.unc.edu (D. Shen).

# 1. Introduction

Prostate cancer (Partin et al., 1993; Shen et al., 2004) is a common type of cancer among men worldwide. Clinically, external beam radiation therapy (EBRT) is one of the most efficient ways for treating prostate cancer. In EBRT (de Crevoisier et al., 2005), Computed Tomography (CT) is of great importance as it provides Hounsfield unit (HU) values, which can be used for dose planning purpose. In the planning stage of EBRT, in order to deliver prescribed dose to the prostate volume, while spare nearby normal organs from the harmful radiation, three critical pelvic organs, i.e., prostate, bladder and rectum, need to be contoured precisely. This is quite a difficult task due to low tissue contrast of the prostate, bladder and rectum in the CT image, as illustrated in Fig. 1. Also, the low delineation accuracy could limit the efficacy in EBRT for prostate cancer treatment and lead to potential side effects (Dearnaley et al., 1999). Currently, magnetic resonance imaging (MRI) is often used jointly with CT in the MRI-assisted radiation therapy (Roach et al., 1996; Sannazzari et al., 2002; Dowling et al., 2012; Metcalfe et al., 2013), since MRI provides high soft tissue contrast and makes manual organ delineation accurate. However, CT and MR images are always acquired at separate time points, for the same patient. Due to intra-patient anatomical changes, accurate soft tissue delineation in MRI cannot be transferred directly to the CT image space. Therefore, it is of clinical importance to register the pelvic MR image to the CT image of the same patient, for the purpose of accurate contour propagation.

However, there are two main challenges for accurate and robust pelvic MRI-to-CT registration. 1) The first challenge comes from *local anatomical deformation*. This is because the CT and MR images of the same patient are always scanned at different time points, which may change positions, shapes and appearances of pelvic organs dramatically due to possible bladder filling and emptying, bowel gas, and irregular rectal movement. Thus, basic linear registration can hardly obtain good registration performance with such a challenge. As shown in Fig. 1, the deformations on three crucial pelvic organs are still obvious, even after linear alignment of the CT and MR images of the same patient. This necessitates the use of non-rigid registration to correct local deformations. 2) The second challenge comes from significant appearance difference between the CT and MR images. For example, from Fig. 1, there are no obvious intensity differences among the organs of the prostate, bladder and rectum in the CT image, and the boundaries between three organs are hard to distinguish. But, in the MR image, the bladder has relatively brighter intensity than both the prostate and rectum regions, and the boundaries between these three organs are easily detectable. Additionally, the texture patterns of the prostate in the MR image are much richer than those in CT. Thus, these different characteristics of image appearance make it hard to design a common similarity metric in MRI-to-CT image registration, which also makes existing available non-rigid registration approaches less efficient in this multi-modality registration case.

#### 1.1. Related work

Many approaches have been developed for non-rigid multi-modal image registration, as described below.

**1.1.1. Information theory based registration methods**—The first category is the information theory based methods, which use mutual information (MI) as the similarity metric for multi-modal image registration (Collignon et al., 1995; Wells et al., 1996; Maes et al., 1997; Viola and Wells III, 1997). Based on this similarity metric, rigid and affine registration results can be efficiently obtained between multi-modal images. However, MI is a good global similarity measurement, which has limited power in capturing local anatomical details, and thus has limited capacity in tackling local deformations. To address this issue, normalized mutual information (NMI) (Studholme et al., 1999) and localized mutual information (LMI) (Klein et al., 2008) have been proposed to measure the similarity between local anatomical structures. However, insufficient number of voxels within the image patch undermines the estimation of intensity distribution when computing MI. As a result, local anatomical structures are often not well aligned by MI-based registration methods. A good review of MI-based image registration methods can be found in (Pluim et al., 2003).

Besides MI, Chung et al. (2002) proposed using Kullback–Leibler distance (KLD) as the similarity metric to guide multi-modal image registration, by enforcing the joint intensity distribution of the to-be-aligned image pair following the priors of learned distributions from the well-aligned images. In addition, So and Chung (2011) further proposed using Bhattacharyya distance (BD) as the similarity metric, which obtained better registration performance compared with KLD-based registration.

However, either MI-based or KLD-based (BD-based) registration methods have a common limitation that these methods are all based on the intensity probability distribution, and ignore spatial information of anatomical structures. This means that different anatomical regions will be found as being matched as long as they have similar intensity distributions.

**1.1.2. Image synthesis based registration method**—For this kind of methods, one image modality is synthesized from the other modality in order to reduce the large appearance gap between different modalities. Afterwards, the multi-modal image registration problem is simplified to a mono-modal registration problem, where most existing registration methods can then be applied. Roche et al. (2001) synthesized an ultrasound image and then estimated rigid registration between the ultrasound and the MR images. Wein et al. (2008) simulated the ultrasound image from the CT image to solve the rigid and affine registration problems.

There are two main problems with the current image synthesis based registration methods. The first problem is that the image synthesis is always performed in a single direction. This will often introduce bias when performing registration only based on one modality image, by ignoring all the anatomical details in the other modality. The second problem is that most of the existing works focus on rigid/affine registration. Few work tackles the non-rigid registration problem. Since existing methods often synthesize from the image modality with rich anatomical details (e.g., MRI) to the image modality with limited anatomical details (e.g., ultrasound/CT), it is quite difficult to accurately estimate local deformations based on the image modality with limited anatomical details, especially for the case of pelvic images where the main pelvic organs often have large local deformations. Therefore, an accurate

and robust image synthesis method, which can preserve adequate anatomical details, is essential for image synthesis based multi-modal non-rigid registration.

**1.1.3. Image synthesis methods**—To date, many image synthesis methods are proposed although most of them are not for registration purpose.

**1.1.3.1. Examplar-based image synthesis:** By using sparse representation methods and its extension, Ye et al. synthesized T2-weighted MRI and FA maps from T1-weighted MRI (Ye et al., 2013). Roy et al. estimated spoiled gradient recalled (SPGR) MRI sequence from magnetization prepared rapid gradient echo (MPRAGE) MRI sequence and vice versa (Roy et al., 2011, 2013). He also predicted FLAIR image from the combination of T1- and T2-weighted MRI (Roy et al., 2010). Iglesias et al. synthesized T1-weighted MRI from PD-weighted MRI (Iglesias et al., 2013). Cordier et al. utilized the brain label image to predict FLAIR image (Cordier et al., 2016). The main idea in these methods is using a patch-wise prediction process. That is, the new source image can be sparsely represented by the patches extracted from an atlas dataset of the same modality, and then the computed sparse coefficients can be utilized to estimate the target image patches of another modality. However, the optimization is needed in the application stage, which often makes the prediction process less efficient. Also, the synthesized results are sometimes noisy due to inaccurate predictions.

**1.1.3.2. Learning-based image synthesis:** In learning-based methods, a training dataset with pre-aligned multi-modal image pairs is needed to help build an accurate and robust mapping model from one image modality to another. Different machine learning techniques are applied. 1) Gaussian mixture regression model. Based on Gaussian mixture regression model, Johansson et al. estimated CT image from three MRI sequence (Johansson et al., 2011) and Roy et al. proposed to synthesize CT image from ultrashort echo-time MRI (Roy et al., 2014). Since the anatomical information is ignored in this method, the synthesized image is always over-smoothed and not accurate. 2) Deep learning based regression model. Convolutional neural networks (CNN) have been applied in Li's work to predict PET from MRI (Li et al., 2014) and shown promising results. However, for training a robust and accurate CNN model, a large dataset is needed and also the computational time is long. 3) Random forest based regression model. Random forest is a non-linear classification and regression method, which can be trained efficiently. Jog et al. built the high resolution T2weighted MRI from low resolution T2-weighted MRI with the help of T1-weighted MRI (Jog et al., 2014a). He also synthesized FLAIR MRI from the combination of T1, T2 and PD weighted MRI (Jog et al., 2014b) by using the concept of random forest. In order to preserve the anatomical information and structural continuity during prediction, the patchwise manner is incorporated into random forest. Jog et al. synthesized T2-weighted MRI from T1-weighted MRI using patch-wise random forest (Jog et al., 2013). Tri et al. proposed to apply structured random forest to estimate CT image from MRI data, where he also utilized the patch as the regression target (Huynh et al., 2016).

One problem with all aforementioned learning based image synthesis methods is that the direction of synthesis model is always from the image modality with rich anatomical details to the image modality with limited anatomical details, e.g., synthesizing CT from MRI.

Although some works have obtained synthetic MRI (Jog et al., 2014), the source modality must have equivalent anatomical details, e.g., multiple modalities are required. This is not applicable in many clinical applications. However, by using image synthesis to improve multi-modal non-rigid registration problem, we argue that the image synthesis should be performed in bidirections: *not only* synthesizing CT from MRI, *but also* synthesizing MRI from CT. In this way, all the anatomical details can be used to accurately steer the non-rigid registration. Thus, an effective image synthesis method, which can accurately build the mapping relationship from both CT-to-MRI and MRI-to-CT, is under high demand.

**1.1.4. Other multi-modal registration methods**—Some feature-based registration methods (Shen and Davatzikos, 2002; Zacharaki et al., 2009; Ou et al., 2011; Heinrich et al., 2012; Toews et al., 2013) estimate deformation fields by leveraging different artificial features (e.g., geometric moments (Shen and Davatzikos, 2002), Gabor attributes (Ou et al., 2011)), rather than directly using image intensities. In this way, the influence of appearance differences between modalities can be minimized during registration. However, it is difficult to find effective features that can describe common anatomical details and are also invariant to the appearance differences. Additionally, some methods conduct the registration and segmentation jointly (Ashburner and Friston, 1997; Jia et al., 2012), or use the segmentation images (Shen and Davatzikos, 2002) to help improve the multi-modal registration. But, accurate segmentation is needed for accurate non-rigid registration, which is often not easy to achieve.

#### 1.2. Contributions

In this paper, we propose a novel image-synthesis-based multi-modal non-rigid registration method to tackle the challenging problem of pelvic CT/MRI registration. The image contrast of the MR pelvic image is better than the counterpart CT image in certain regions. For example, the bone in the CT image is more observable than in the MR image. In contrast, the MR image has much richer details in soft tissue parts (See Fig. 1). Thus, in order to improve image registration for the entire image domain using complementary information from both modalities, it is necessary to perform image synthesis in bi-directions. Therefore, each original image has its synthesized pseudo image where their anatomical shapes remain intact, but with different appearances. Based on the synthesized MR and CT images, we propose a dual-core multi-modal image registration method to simultaneously estimate 1) one deformation pathway from the synthesized CT image of the counterpart MR image to the actual CT image, and 2) the other deformation pathway from the actual MR image to the synthesized MR image of the counterpart CT image. It is apparent that the registration in each core is free of the appearance gap. Finally, we present a dual-core deformation fusion method to obtain a unified deformation pathway, which is then used to propagate the segmented contours of the prostate, bladder and rectum from the MRI to the CT domain in EBRT of prostate cancer. The contributions of our work can be summarized as follows:

• In order to address the large appearance gap between the CT and MR images, we propose to use improved patch-wise random forest with auto-context model for bi-directional image synthesis, i.e., synthesizing MRI from CT and also synthesizing CT from MRI. Here, we choose random forest as the basic learning

method since it has the capacity to build a robust non-linear relationship in an efficient way. Inspired by the patch-wise random forest (Jog et al., 2013; Huynh et al., 2016), besides CT image synthesis from MRI, we try to tackle the challenging problem of MRI synthesis from single CT modality. It is worth noting that, synthesizing MRI from single CT modality is a "simple-to-complex" image synthesis problem, which is not proposed and solved in previous publications. It is of great importance for the following multi-modal pelvic registration since it can provide more anatomical details.

• To fully utilize the complementary image information from both modalities, a dual-core deformation fusion framework is proposed. This method can effectively estimate the deformation pathway between the CT and MR images, by iteratively fusing two deformation pathways: 1) from the synthesized CT of MRI to the actual CT, and 2) from the actual MRI to the synthesized MRI of CT. Experimental results show that the registration accuracy can be boosted under this dual-core steered registration framework, compared with the previous methods that are often based on single-directional image synthesis.

This paper is a significant extension of our previous conference paper (Cao et al., 2016). In particular, we present more details for both the proposed learning-based image synthesis and the dual-core deformation fusion framework. Moreover, we also give more detailed experimental results on both bi-directional image synthesis and multi-modal image registration tasks. In order to evaluate the effectiveness of the proposed method, we further use various measurements to fully evaluate both image synthesis performance and registration performance, and analyze how the image synthesis will influence or contribute to the multi-model registration. A detailed discussion section is also added to analyze the advantages and clinical significance of the proposed method. In addition, the limitation and future work are discussed.

The rest of this paper is organized as follows. Section 2 presents the learning-based dualcore multi-modal image registration, including the overview, learning-based bi-directional image synthesis and dual-core steered MRI-to-CT registration. Section 3 provides the experimental description and results based on a real pelvic dataset. Finally, the discussion and conclusion are provided in Section 4 and Section 5, respectively.

### 2. Method

#### 2.1. Overview

Fig. 2 gives a systematic flowchart of the proposed dual-core steered multi-modal image registration method. This method consists of two major steps, as indicated by two asterisks in Fig. 2.

• **Bi-directional Image Synthesis.** A learning-based bi-directional image synthesis method is proposed to overcome the appearance gap between CT and MRI in multi-modal image registration. Since MRI synthesis from CT is much more challenging and novel, our method is particularly described in the context of CT-to-MRI synthesis. The same manner can also be performed to MRI-to-CT

synthesis. Specifically, in our method, an improved patch-wise random forest is first used to directly predict the entire MR image patch from the corresponding CT patch, where Haar-like features are employed to better characterize the CT patch. Compared to the voxel-wise random forest that only estimates one MRI value from a CT patch, the patch-wise random forest can preserve the neighboring structured information of the predicted MRI. To increase the overall accuracy of image synthesis, we further adopt an auto-context model to deeply refine the synthesized MRI by considering the context features in the neighborhood. With the improved quality of the synthesized image, the non-rigid registration performance can be boosted accordingly. The details of this step are described in Sections 2.2 and 2.3.

• **Dual-core Steered MRI-to-CT Registration.** In the beginning of registration, a synthesized MR image  $\hat{I}_{MR}$  is obtained from the CT image  $I_{CT}$ , and also a synthesized CT image  $\hat{I}_{CT}$  is obtained from the MR image  $I_{MR}$ . Then, the deformation pathway between MRI and CT is estimated in two ways: 1) registering  $\hat{I}_{CT}$  to  $I_{CT}$ , and 2) registering  $I_{MR}$  to  $\hat{I}_{MR}$ , as shown in Fig. 2. Eventually, the MR image is warped to the CT image space by following the fused deformation pathway  $\varphi$ , which is also refined iteratively under *dual-core deformation fusion* framework. The details of this step are described in Section 2.4.

#### 2.2. Learning-based image synthesis

**2.2.1. Patch-wise random forest (P-RF) regression**—The basic theory of random forest can be found in (Liaw and Wiener, 2002). It can be used for non-linear classification and regression, and has been widely adopted in disease diagnose (Zhang et al., 2016), image registration (Wei et al., 2016) and segmentation (Wang et al., 2015; Gao et al., 2016). Here, we directly introduce the improved patch-wise random forest for the application of image synthesis. Overall, in MR image synthesis from CT, it can be used to regress the MRI patch intensity from the corresponding CT patch and build a nonlinear mapping to bridge the appearance gap between the two modalities based on multiple decision trees. Fig. 3 illustrates the training and testing stages of patch-wise random forest for the MR image synthesis from CT.

In the **training step** of patch-wise random forest, the CT and MR images of the same patient are already pre-aligned (with the details of the CT and MRI pre-alignment are described in Section 3.2). Based on this pre-aligned dataset, we first randomly select *N* voxels to obtain *N* training samples. Each sample consists of two components: 1) the input appearance feature vector extracted from a single CT patch, which can be denoted by  $\mathbf{x}$ , and 2) the target MRI patch intensity value corresponding to the center of the CT patch, which can be represented by vector  $\mathbf{y}$ . For training the non-linear mapping from CT to MRI, the input is *N* CT feature vectors  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$ , along with the corresponding *N* target MRI patch intensity values  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]$ . Random forest consists of multiple decision trees, with each one trained independently. For each tree, its training is conducted by learning a set of splitting nodes to recursively partition the training sample set. Specifically, in each split node, for a feature indexed by k, its optimal threshold  $\tau$  is found to best split the training set

into left and right subsets  $S_L$  and  $S_R$  with consistent target MRI patch intensity values. Mathematically, it is to maximize the variance reduction by a split:

$$\underset{k,\tau}{\operatorname{argmax}} V(S) - \frac{N_{\rm L}}{N} V(S_{\rm L}) - \frac{N_{\rm R}}{N} V(S_{\rm R}), \quad (1)$$

$$S_{\mathrm{L}} = \left\{ (x,y) \in S \mid x^k < \tau \right\}, \ S_{\mathrm{R}} = \left\{ (x,y) \in S \mid x^k \geq \tau \right\}, \quad (2)$$

where  $V(\cdot)$  computes the averaged variance of target MRI intensity values across each element of the patch vector in the training set,  $x^k$  indicates the *k*-th feature, and *S* indicates a training set. Once the optimal combination of feature index and its threshold is found for a split node, the training set *S* is partitioned into  $S_L$  and  $S_R$  with the sample numbers  $N_L$  and  $N_R$ , respectively. The same split operation is recursively conducted on  $S_L$  and  $S_R$ , until 1) the tree reaches the maximum tree depth, or 2) the number of remaining training samples is too few to be split.

In the **testing step**, given a testing sample with feature vector  $x_{new}$  extracted from a new CT patch sample (see the right part of Fig. 3), it is pushed to the root split node of each tree in the forest. Under the guidance of the split node (i.e., go left if  $x_{new}^k < \tau$ , and go right

otherwise), the testing sample will arrive at a leaf node of each tree, where the averaged target MRI patch intensity values of training samples in that leaf are used as a prediction of the tree. The final output of random forest is the average of patch predictions from all trees. Obviously, by going through all the voxels in the new CT image, the predicted patches are highly overlapping. Thus, by averaging all those overlapping values, the predicted MRI can be finally obtained.

Since the target MR image has strong spatial clues, we combine the patch-wise representation with random forest for the MR image prediction. The advantage of patch-wise random forest regression is that, by predicting a MRI patch as a whole, the neighborhood information can be naturally preserved and lead to better image synthesis eventually. Moreover, since more target values can be obtained as the output for each tree, the number of trees in the whole forest can be reduced, which improves the efficiency of the prediction.

**2.2.2. Feature extraction**—Feature extraction is a crucial step for random forest training. An effective feature should have the capacity to distinguish different anatomical structures, as well as robustness to noise. Typically, it is not effective to directly use patch intensities as appearance features, since soft tissues have low contrast in the CT image, and also the CT image is always noisy. Therefore, the individual CT values are insufficient to characterize different pelvic structures, as they are also sensitive to noise. To solve this problem, we extract Haar-like features from CT patch to serve as appearance features for random forest training. Specifically, a Haar-like feature (Gao et al., 2016) describes 1) the averaged

intensity within a sub-block in the patch, or 2) the averaged intensity difference between two sub-blocks in the patch, as shown in Fig. 4.

By using the integral image (Viola and Jones, 2004), in which each position contains the summation of the voxel intensity values from the original position to the current location, Haar-like feature can be calculated with the constant time complexity. To generate more Haar-like features, we randomly sample the information within the patch. Since random forest has the inherent capability of feature selection, only the meaningful Haar-like features will be retained. It is worth noting that, to cover both local and global appearances for the underlying voxel, Haar-like features are extracted from coarse, medium and fine image resolutions by using the same patch size, respectively.

2.2.3. Auto-context model (ACM) based image synthesis refinement—In the

patch-wise random forest, the MR image patch of each voxel is estimated from the Haar-like features extracted from the corresponding CT image patch. To build better non-linear mapping between the CT patch features and the MR image patch, we further incorporate the neighboring prediction results to enhance the patch-wise random forest training. In particular, an auto-context model (Tu and Bai, 2010) is adopted to refine the synthesized MR image iteratively. In this paper, we perform three layers as illustrated in Fig. 5. Note that, the auto-context model is used only in the last two layers, as indicated by a dashed box.

In the first layer, the appearance features (i.e., Haar-like features) from CT images are extracted to train a patch-wise random forest. Then, the trained forest is used to generate the initial synthesized MR images. In the second layer, to coordinate the prediction information in the neighborhood, the Haar-like features (namely the context features, as shown in Fig. 5) are further extracted from those initial synthesized MR images, similar to the extraction of appearance features. By combining these context features with the CT appearance features, a second patch-wise random forest can be trained, which also leads to the update of the synthesized MR images and their context features. This process iterates until it reaches the maximum number of layers.

#### 2.3. Bi-directional image synthesis

In order to achieve unbiased MRI-to-CT registration and use the complementary information from both modalities in registration, the image synthesis should be performed in bidirections. Similar to Section 2.2, which proposes a learning-based method for synthesizing the MR image from CT, the CT image is also synthesized from MRI by using the same implementation: 1) The patch-wise random forest is applied to regress the CT patch from the corresponding MR image patch (represented by Haar-like features); 2) Further apply the auto-context model to refine the performance of the synthesized CT image. The learning-based bi-directional image synthesis method is summarized in Algorithm 1.

#### Algorithm 1

Learning-based Bi-directional image synthesis.

Purpose: Synthesize Modality-A from Modality-B

Author Manuscript

Data: Pre-aligned original *Modality-A* and *Modality-B* image pairs

**Result:**  $\hat{I}_A$  - synthesized *Modality-A* image

**Initialization:** Predicted *Modality-A* image map:  $P_A(0) = 0$ ; Context feature:  $X_{\text{context}}(0) = \emptyset$ 

#### Training stage:

If Synthesizing MRI from CT

• *Modality-A*  $\leftarrow$  MRI; *Modality-B*  $\leftarrow$  CT;

#### Else If Synthesizing CT from MRI

• *Modality-A*  $\leftarrow$  CT; *Modality-B*  $\leftarrow$  MRI;

#### End If

- Extract appearance feature X<sub>appearance</sub> from Modality-B image
- Extract target *Modality-A* intensity patch y

For i ← 1:MaxIteration

- Extract context feature  $X_{\text{context}}$  (i-1) from the predicted *Modality-A* image map  $P_A(i-1)$
- i-th layer structured random forest P training  $\leftarrow \{(X_{\text{appearance}}, X_{\text{context}}, (i-1)), Y\}$
- Update predicted *Modality-A* image map  $P_A(i)$
- *i* = *i* + 1

#### End For

Testing stage:

Extract appearance feature  $X_{appearance}^{new}$  from new *Modality-B* image  $I_B$ 

#### For $i \leftarrow 1$ :MaxIteration

- Extract context feature  $X_{\text{context}}^{\text{new}}(i-1)$  from predicted *Modality-A* image  $P_A^{\text{new}}(i-1)$
- $i \text{th layer structured random forest } F^{i} \text{ testing } \leftarrow \{X_{\text{appearance}}^{\text{new}}, X_{\text{context}}^{\text{new}}(i-1)\}$
- Update predicted *Modality-A* image map  $P_A^{new}(i)$
- *i* = *i* + 1

End For

 $\hat{I}_A = P_A^{\text{new}}(i)$ 

#### 2.4. Dual-core steered MRI-to-CT image registration

**2.4.1. Intensity-based non-rigid registration**—After the CT and MR images are synthesized, we have a total of four images: actual CT, actual MRI, synthesized CT from MRI, and the synthesized MRI from CT. Based on the actual and synthesized images, we can utilize the existing non-rigid registration methods to estimate the deformations 1) from the synthesized CT to the actual CT, and 2) from the actual MRI to the synthesized MRI. Among various existing registration methods, we choose two popular non-rigid registration methods for evaluation: 1) Diffeomorphic Demons (D. Demons) (Vercauteren et al., 2009), which is a non-parametric registration method widely used in various applications; 2)

Symmetric Normalization (SyN) (Avants et al., 2008), which is a top-ranked method among the 14 state-of-the-art non-rigid registration methods evaluated for brain image registration in (Klein et al., 2009). For SyN, we use the Advanced Neuroimaging Tool (ANTs) (Avants, Tustison et al., 2011) for its implementation.

#### 2.4.2. Dual-core deformation fusion (DDF) for MRI-to-CT registration—To

overcome the limitation of multi-modal image registration methods that utilize only singledirectional image synthesis, we utilize both synthesized CT and synthesized MRI in MRI-to-CT image registration, which can take the advantage of complementary information from both modalities. Let  $I_{CT}$ ,  $\hat{I}_{CT}$ ,  $I_{MR}$  and  $\hat{I}_{MR}$  denote the actual CT, synthesized CT, actual MRI, and synthesized MRI, respectively. Traditionally, a general objective function for the conventional MRI-to-CT image registration can be given as:

$$\arg\min_{\varphi} \mathcal{M}(I_{\text{CT}}, \varphi(I_{\text{MR}})) + \lambda \mathcal{R}(\varphi), \quad (3)$$

where  $\varphi$  is the deformation field to be estimated,  $\mathfrak{M}$  is the dissimilarity metric, and  $\varphi(I_{MR})$  means that deforms the subject MR image to the CT image space with the estimated deformation field  $\varphi$ .  $\mathfrak{R}$  is a regularization term, used to constrain the smoothness of the estimated deformation field  $\varphi$ .

By using the bi-directional synthesized images, the deformation field can be estimated by simultaneously minimizing 1) the dissimilarity between the CT modality core and 2) the dissimilarity between the MRI modality core. Then, the registration objective function can be rewritten as:

$$\underset{\varphi}{\operatorname{argmin}} \frac{1}{2} \mathscr{M} \left( I_{\text{CT}}, \varphi(\hat{I}_{\text{CT}}) \right) + \frac{1}{2} \mathscr{M} \left( \hat{I}_{\text{MR}}, \varphi(I_{\text{MR}}) \right) + \lambda \mathscr{R}(\varphi) \,. \tag{4}$$

To solve Eq. (4) and reuse the available registration tools, we apply the alternative optimization method. Specifically, we optimize the registration between the CT modality core ( $I_{\text{CT}}$  and  $\hat{I}_{\text{CT}}$ ) and the registration between the MRI modality core ( $\hat{I}_{\text{MR}}$  and  $I_{\text{MR}}$ ) separately as below:

$$\underset{\varphi_{\text{CT}}}{\operatorname{argmin}} \frac{1}{2} \mathcal{M} \left( I_{\text{CT}}, \varphi_{\text{CT}} (\hat{I}_{\text{CT}}) \right) + \frac{1}{2} \lambda \mathcal{R}(\varphi_{\text{CT}}), \quad (5)$$

$$\underset{\varphi_{\rm MR}}{\operatorname{argmin}} \frac{1}{2} \mathscr{M}(\hat{I}_{\rm MR}, \varphi_{\rm MR}(I_{\rm MR})) + \frac{1}{2} \lambda \mathscr{R}(\varphi_{\rm MR}).$$
(6)

Eqs. (5) and (6) are used to minimize the differences 1) between the CT image  $I_{CT}$  and the warped synthesized CT image  $\varphi_{CT}$  ( $\hat{I}_{CT}$ ), and 2) between the synthesized MR image  $\hat{I}_{MR}$  and

the warped MR image  $\varphi_{MR}$  ( $I_{MR}$ ), respectively. In order to use the complementary information from both modalities, the dual-core deformation fusion is proposed accordingly:

$$\underset{\varphi}{\operatorname{argmin}} \frac{1}{2} \|\varphi - \varphi_{\text{CT}}\|_{2}^{2} + \frac{1}{2} \|\varphi - \varphi_{\text{MR}}\|_{2}^{2}.$$
 (7)

Eq. (7) is used to ensure that the final deformation pathway  $\varphi$  to be close to both separately estimated  $\varphi_{CT}$  and  $\varphi_{MR}$ .

Both  $\varphi_{CT}$  and  $\varphi_{MR}$  can be optimized by using *either* D. Demons *or* ANTs-SyN, although the objective functions are slightly different in Eqs. (5) and (6). After estimating  $\varphi_{CT}$  and  $\varphi_{MR}$ , the final deformation  $\varphi$  can be effectively optimized by letting the gradient of Eq. (7) equal to zero, which brings to:

$$\varphi = \frac{1}{2}(\varphi_{\rm CT} + \varphi_{\rm MR}). \quad (8)$$

However, simple average of deformation pathways may make the well-aligned structures in one modality become misaligned after averaging. To alleviate this issue, we approximate the optimal solution of Eq. (4) by alternating the dual-core deformation fusion procedure (Eqs. (5)–(8)) based on the warped synthesized CT and the original MR images until convergence, as summarized in Algorithm 2.

For iteration *i*, the tentatively warped images  $\hat{I}_{CT}^{i-1}$  and  $I_{MR}^{i-1}$  are used to estimate a next set of deformations  $\varphi_{CT}^{i}$  and  $\varphi_{MR}^{i}$ . The estimated deformations are then merged to form a combined deformation  $\varphi^{i}$ , which is used to update the currently estimated deformation pathway  $\varphi = \varphi \circ \varphi^{i}$ . Here, " $\circ$ " denotes deformation field composition (Vercauteren et al., 2009), which means updating deformation field  $\varphi$  by  $\varphi^{i}$ . Using this deformation field composition can help preserve the quality of the subject image, by warping the original moving image  $\hat{I}_{CT}$  and  $I_{MR}$  for just one time. The whole registration procedure iterates until the incremental deformation  $\varphi^{i}$  is small enough.

#### Algorithm 2

Optimization of Eq. (4): Iterative dual-core deformation fusion (DDF).

**Result:**  $\varphi$  - the final deformation field and approximated solution of Eq. (4)

**Initialization:** i = 0;  $\hat{I}_{CT}^0 = \hat{I}_{CT}$ ;  $I_{MR}^0 = I_{MR}$ ;  $\varphi = 0$ ;

While  $(\varphi_2^i > \varepsilon)$ 

- i = i + 1;
- Register CT image core:  $\varphi_{CT}^{i} = \text{Register} (\hat{I}_{CT}^{i-1}, I_{CT})$

**Data:** CT:  $I_{CT}$ ; synthesized CT:  $\hat{I}_{CT}$ ; MRI:  $I_{MR}$ ; synthesized MRI:  $\hat{I}_{MR}$ 

- Register MR image core:  $\phi_{MR}^{i} = \text{Register}(I_{MR}^{i-1}, \hat{I}_{MR})$
- Dual-core deformation fusion:  $\varphi^{i} = \frac{1}{2}(\varphi^{i}_{CT} + \varphi^{i}_{MR})$
- Update the estimated deformation:  $\varphi = \varphi \circ \varphi^i$
- Deform the subject image:  $\hat{I}_{CT}^{i} = \varphi(\hat{I}_{CT}, ), \hat{I}_{MR}^{i} = \varphi(I_{MR})$

End While

# 3. Experiments

#### 3.1. Dataset description

The experimental dataset consists of 20 pairs of CT and MR images acquired from 20 prostate cancer patients. For this dataset, the oncologists have manually labeled three crucial pelvic organs in each image, including the prostate, bladder and rectum. We use these labels as the ground-truth. The CT image size is  $512 \times 512 \times 443$  with voxel size of  $1.172 \times 1.172 \times 1 \text{ mm}^3$ . The MR image size is  $256 \times 256 \times 144$  with voxel size of  $1 \times 1 \times 1 \text{ mm}^3$ . Intrasubject and inter-subject linear alignments are done with FLIRT (Jenkinson and Smith, 2001). Then, the unnecessary regions are removed to make the experiments more efficient. After preprocessing, both CT and MR images have the same image size ( $200 \times 180 \times 80$ ) and resolution ( $1 \times 1 \times 1 \text{ mm}^3$ ). Note that the cropped image size is large enough to include the prostate, bladder and rectum in each patient's data.

#### 3.2. Image synthesis experiments

For the data used in patch-wise random forest training, both intra-subject and inter-subject need to be well aligned. The intra-subject alignment is to register each pair of CT and MR images belonging to the same patient. Since it is quite difficult to well-align the CT and MR image accurately due to large deformation on soft tissue, we use the manual label image of crucial pelvic organs (ground-truth image) to guide accurate pre-alignment. Specifically, affine registration is first performed with FLIRT (Jenkinson and Smith, 2001). Then, non-rigid registration (ANTs-SyN (Avants et al., 2008)) is conducted based on careful parameter tuning for each individual subject. For these affine and non-rigid registration steps, we use mutual information as the similarity metric and perform the registration using intensity images. To further refine the registration results on the crucial pelvic organs between the CT and MR images, we use D. Demons to register the respective manual labels of the prostate, bladder and rectum in the CT and MR images. The detailed parameter settings of each registration step are summarized in Table 1.

Finally, for inter-subject alignment, linear registration (again with FLIRT (Jenkinson and Smith, 2001)) is further applied to roughly register all subjects to a common space. Note that, for the above registration procedures, all respective transformations are composed into a single transformation to warp the image (under registration) for just one time, just avoiding multiple interpolations, which often make the final warped image fuzzy. For this pre-aligned dataset, the Dice Similarity Coefficients (DSC) values of the manually-segmented prostate,

bladder and rectum are 92.8%, 95.7% and 92.3%, respectively. For better evaluating the image synthesis performance, the intensity range of MR image is normalized to [0, 4000], which is the intensity range for the CT image.

**3.2.1. Parameter setting**—In image synthesis, the aligned CT and MR image pair is used to train our image synthesis models: 1) synthesizing CT from MRI (**MRI-to-CT**) and 2) synthesizing MRI from CT (**CT-to-MRI**). Table 2 provides parameters of patch-wise random forest (**P-RF**) training for bi-directional image synthesis. Specifically, 2-layer auto-context model (**ACM**) and 10-fold cross validation are applied. For P-RF training, the input feature sample is extracted from the local  $15 \times 15 \times 15$  patch and the output is the  $3 \times 3 \times 3$  patch intensity. Here, more trees and deeper tree depth are used for synthesizing MRI from CT, since this simple-to-complex mapping is more difficult. The voxel-wise random forest (**V-RF**) is also applied as the baseline method. Note that V-RF utilizes the same parameter setting as P-RF, but its output is a single voxel value, not the  $3 \times 3 \times 3$  patch.

**3.2.2. Evaluation**—Two measurements are used to quantitatively evaluate the image synthesis performance: 1) mean absolute error (MAE) and 2) peak signal-to-noise ratio (PSNR), which can be defined as:

$$MAE = \frac{1}{lmn} \sum_{a=1}^{l} \sum_{b=1}^{m} \sum_{c=1}^{n} \left| I(a, b, c) - \widehat{I}(a, b, c) \right|, \quad (9)$$

$$PSNR = 10 \log_{10} \left( \frac{MAX_I^2}{\frac{1}{lmn} \sum_{a=1}^{l} \sum_{b=1}^{m} \sum_{c=1}^{n} \left| I(a, b, c) - \hat{I}(a, b, c) \right|^2} \right), \quad (10)$$

where I and  $\hat{I}$  denote the original image and the synthesized image, and the image size represents by I, m, and n in three dimensions. Note that a high-quality synthesized image should have lower MAE and higher PSNR.

**3.2.3. Bi-directional image synthesis results**—Table 3 provides the quantitative performance of CT image synthesis from MRI, and Table 4 presents the quantitative performance of MR image synthesis from CT. From these two tables, we can observe that the patch-wise random forest is effective in building non-linear regression models to fill up the appearance gap between CT and MRI modalities. Furthermore, the image synthesis performance is improved by using more layers with the auto-context model. For both CT and MRI synthesis, the improvement achieved by adding the first-layer auto-context model is more significant than adding the second layer. This can qualitatively demonstrate the fast convergence of the auto-context model. In practice, two-layer auto-context model always leads to the convergence.

Figs. 6 and 7 provide detailed visualization of the synthesized CTs and the synthesized MRIs for one subject, along with the error maps between the actual and the synthesized

images. The results are consistent with the performance in Tables 3 and 4. From Figs. 6 and 7, we can observe that the boundaries of both bone regions and soft tissue regions are clearer and less noisy by using P-RF, compared with results by the V-RF based image synthesis method. Moreover, the quality of the synthesized image is further improved by auto-context model refinement. The final synthesized image is quite similar to the actual image, by also preserving the smoothness and continuity. Note that, to achieve this promising performance, a relatively small number of trees (as shown in Table 2) is used to train our image synthesis model. This well demonstrates the effectiveness of the P-RF method, which allows obtaining adequate output values by using only a small number of trees.

It is reasonable that the performance of the synthesized CT image is much better than the synthesized MR image. This is because 1) the texture patterns are much more complex in the MR image than in the CT image, thus difficult to estimate, and 2) also the appearance in the MR image is more diverse than in the CT image. As shown in Figs. 6 and 7, the error maps of synthesized MRI are fuzzy, and also the synthesized MR images are smoother than the actual ones. However, the boundaries of pelvic tissues in the synthesized MR image are clear, and the appearance relationship is consistent between the synthesized MR image and the actual MR image, which provide crucial reference information for guiding the subsequent image registration.

#### 3.3. Image registration experiments

In this section, the experiments are performed by registering the CT and MRI of each subject from the total 20 subjects. At each time, the image synthesis model is trained from the 18 selected subjects (using the pre-aligned dataset). Then, the registration is tested upon the remaining 2 subjects (using their original data, instead of pre-aligned data). After repeating 10 times, all subjects are used as testing subjects to perform the registration experiment. The averaged results are reported in the following section.

**3.3.1. Evaluation**—After the MR image is registered to the CT image, we use two kinds of measurements to evaluate the registration performance: 1) Volume overlap. Specifically, Dice Similarity Coefficient (DSC) between manual labels on CT ( $L_{CT}$ ) and aligned MRI ( $L_{MR}$ ) is used as the main measurement to evaluate the registration performance. Besides, sensitivity (SEN) and positive predictive value (PPV) are also used to compare the proposed method with the existing methods. 2) Surface distance. Specifically, the distance measurement is used to account for the boundary discrepancies between the CT labels and the aligned MRI labels. Here, symmetric average surface distance (SASD) and Hausdorff distance (HAUS) are calculated to further demonstrate the registration performance. Eqs. (11)–(14) below provide the definition of all these measurements.

$$DSC = \frac{2 |L_{CT} \cap L_{MR}|}{|L_{CT}| + |L_{MR}|}, \quad (11)$$

Page 16

Cao et al.

$$\text{SEN} = \frac{\mid L_{\text{CT}} \cap L_{\text{MR}} \mid}{\mid L_{\text{CT}} \mid}, \text{ PPV} = \frac{\mid L_{\text{CT}} \cap L_{\text{MR}} \mid}{\mid L_{\text{MR}} \mid}, \quad (12)$$

$$SASD = \frac{1}{2} \left\{ \max_{u \in L_{CT}} \min_{v \in L_{MR}} d(u, v) + \max_{u \in L_{MR}} \min_{v \in L_{CT}} d(u, v) \right\}, \quad (13)$$

HAUS = max 
$$\left\{ \max_{u \in L_{\text{CT}}} \min_{v \in L_{\text{MR}}} d(u, v) + \max_{u \in L_{\text{MR}}} \min_{v \in L_{\text{CT}}} d(u, v) \right\}.$$
 (14)

Here, *L* means a binary label mask from prostate, bladder and rectum. d(u, v) is the Euclidean distance between voxels *u* and *v*. The unit of the distance measurement is presented by millimeter (mm). A good registration result should have higher volume overlap and lower surface distance.

**3.3.2. MRI-to-CT registration results**—As described in Section 2.4, two state-of-the-art non-rigid registration methods (D. Demons and ANTs-SyN) are used in our experiment, with their detailed parameter settings shown in Table 5. For the preprocessed data, the CT and MRI of the same patient are already linearly aligned. The DSC values of prostate, bladder, and rectum after linear registration are 85.66%  $\pm$  5.65%, 89.29%  $\pm$  2.18% and 78.44%  $\pm$  3.54%, respectively, and we regard these results as the baseline.

Fig. 8 illustrates the MRI-to-CT registration results under different layers of ACM for image synthesis, and different dual-core deformation fusion (DDF) iterations for image registration, respectively. The mean DSC values of prostate, bladder, and rectum are evaluated. Here, the registration results in Fig. 8(a) use 3 iterations (3-iter) DDF, while the registration results in Fig. 8(b) use 2 layers (2-layer) ACM.

As shown in Fig. 8(a), more layers of ACM lead to better registration accuracy due to better quality of the synthesized images. Fig. 8(b) demonstrates that the DDF framework improves registration performance of both D. Demons and ANTs-SyN iteratively. In practice, we found that it is enough to have 2-layer ACM in image synthesis and 3 iterations (3-iter) in DDF.

Table 6 provides the mean and standard deviation of DSC of the total 20 subjects for the three crucial pelvic organs. It can be observed that, for D. Demons, which is originally not applicable for multi-modal image registration, can now work well by introducing the synthesized image. For ANTs-SyN, using MI as similarity metric obtains reasonable registration results on the original CT and MRI. However, as shown in Table 6, better performance can be obtained by using the synthesized image. This demonstrates that using synthesized image can enhance the performance of multi-modal image registration method.

Moreover, the best performance is achieved by using our proposed dual-core deformation fusion algorithm. The consistently higher DSC and the lower standard deviation by our proposed method demonstrate both its robustness and accuracy in solving multi-modal image registration problem.

We further provide the mean SEN, PPV, SASD, and HUAS values of three pelvic organs (prostate, bladder and rectum) in Tables 7 and 8 to compare our proposed bi-directional image synthesis based registration method with single-directional image synthesis based registration method.

From Tables 7 and 8, we observe that our proposed bidirectional image synthesis based method consistently obtains higher SEN and PPV and also lower SASD and HAUS, compared with the single-directional image synthesis based registration methods. This demonstrates that, by using the complementary information from both modalities, the performance is significantly improved for multi-modal image registration. It is worth noting that, for all cases, the PPVs always have higher values than SENs. This phenomenon occurs when the organ contour is under-segmented compared with the ground truth. It is reasonable in our MRI-to-CT registration case, since the manual labels on MRI are always smaller than those on CT. Due to the low tissue contrast in CT, the manual labels on CT image always include more regions due to the difficulty of identifying the organ boundaries.

Fig. 9 provides the visualized registration results after ANTs-SyN registration for 3 subjects in the dataset. For each subject, the first row shows the actual CT and MR images and the corresponding synthesized MR and CT images, which are all used in our proposed iterative DDF registration framework. The second row provides the registration results, i.e., the warped MR image, by using the actual CT and MR image (**CT & MRI**), single-directional image synthesis (**CT & S-CT, S-MRI & MRI**), and our bi-directional image synthesis based registration method (**Proposed**). The visualized registration results are corresponding with those shown in Table 6. The yellow contours indicate CT's ground-truth labels of the prostate, bladder and rectum. The red contours denote the labels of warped MR image after registration. From the overlaps between the ground-truth labels and the warped (estimated) labels, we can also observe that using single-directional synthesized image can partially improve the registration result, compared to the case of only using actual CT and MR images. Furthermore, it is obvious that our proposed method can better preserve the structures of pelvic organs during registration, and also achieves much higher registration performance.

# 4. Discussion

We have presented a novel image-synthesis-based multi-modal pelvic registration method to help improve the accuracy and efficiency of prostate cancer radiation therapy. Compared with the traditional image synthesis based methods, which only perform single-directional image synthesis, we proposed a bi-directional image synthesis based registration method and for the first time tackled the challenging problem of the MR image synthesis from the single CT modality. The bi-directional image synthesis can provide adequate anatomical information from both modalities, thus boosting the performance of MRI-to-CT non-rigid

registration. The promising experimental results also show the potential of translation to the routine clinical application.

For image synthesis, the training time of each random forest model is about 3.5 h on a 2.93 GHz 12-core Intel processor, while the testing time is about 5 min for each synthesized image by using the 2-layer auto-context model. For the proposed registration method, each DDF iteration takes about 2 min by using D. Demons, while 3 min by using ANTs-SyN. In the application stage, if using three DDF iterations (i.e., the same setting in the experiment section), registering a new pair of CT and MRI costs less than 15 min, which can be further improved with the optimization of implementations.

The topic of the MR image synthesis from CT is worth to discuss individually. Besides affording a new idea for multi-modal image registration problem, it can also be widely used in other medical image analysis tasks. The drawback of current MRI synthesis methods is that more image modalities are needed to predict MRI, which limits their clinical applications. In this paper, we attempt to solve this problem by using a common machine learning technique, random forest, with the consideration of the characters from each image modality. Reasonable and promising MRI synthesis results are obtained, and we have already proved its positive role in solving multi-modal registration problem. For future work, we will try to develop more advanced approach to further improve the image synthesis performance.

It is obvious that, for image synthesis based multi-modal registration, the quality of the synthesized image will influence the registration performance. Based on machine learning techniques, a specific detail can be hardly predicted when the source modality has no information about this detail. An example is the subject 1 in Fig. 9. Some fills in the rectum region show different appearance with the regular rectum appearance in the MR image, while there is no difference in the CT image. Thus, this anatomical detail cannot be predicted accurately in the synthesized MRI since no information is provided by the source CT. This will affect the registration when only using single modality pair. The proposed iterative dual-core deformation fusion in this paper can help mitigate this problem. The fusion procedure can correct the mis-alignment in one modality core with the help of another modality core to largely avoid the registration errors.

Further work on this topic should consider the inter-subject MRI-to-CT pelvic registration problem. It is more difficult to tackle, compared with intra-subject registration presented in this paper, since the pelvic organs are quite variable across time and subjects. Also, a larger dataset is needed to further verify the effectiveness of the proposed image synthesis and registration method.

#### 5. Conclusion

In this paper, we have proposed a bi-directional image synthesis based dual-core multimodal image registration method to register the pelvic MR image to the CT image for facilitating more accurate prostate cancer radiation therapy. To obtain accurate MRI-to- CT registration, image synthesis is proposed to fill up the appearance gap between the two

modalities. In particular, the patch-wise random forest and auto-context model are used to 1) synthesize CT from MRI (complex-to-simple mapping), and also 2) synthesize MRI from CT (simple-to-complex mapping). This bi-directional image synthesis can effectively reduce the bias for registration, compared with single-directional image synthesis. Furthermore, we proposed an iterative dual-core deformation fusion framework to steer the estimation of the deformation pathway from the MR image to the CT image by fully utilizing the complementary information in multiple modalities. Experimental results show that our proposed method can achieve higher registration performance than the conventional methods under various evaluation measurements.

# References

- Ashburner J, Friston K. Multimodal image coregistration and partitioning—a unified framework. Neuroimage. 1997; 6(3):209–217. [PubMed: 9344825]
- Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal. 2008; 12(1):26–41. [PubMed: 17659998]
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011; 54(3):2033–2044. [PubMed: 20851191]
- Cao, X., Gao, Y., Yang, J., Wu, G., Shen, D. Learning-based multimodal image registration for prostate cancer radiation therapy. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2016.
- Chung, AC., Wells, WM., III, Norbash, A., Grimson, WEL. Multi-modal image registration by minimising kullback-leibler distance. Medical Image Computing and Computer-Assisted Intervention—MICCAI; 2002; Springer; 2002. p. 525-532.
- Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G. Automated multi-modality image registration based on information theory. Inf Process Med Imag. 1995
- Cordier NH, Delingette M, Lê, Ayache N. Extended modality propagation: image synthesis of pathological cases. 2016
- de Crevoisier R, Tucker SL, Dong L, Mohan R, Cheung R, Cox JD, Kuban DA. Increased risk of biochemical and local failure in patients with distended rectum on the planning CT for prostate cancer radiotherapy. Int J Radiat Oncol Biol Phys. 2005; 62(4):965–973. [PubMed: 15989996]
- Dearnaley DP, Khoo VS, Norman AR, Meyer L, Nahum A, Tait D, Yarnold J, Horwich A. Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial. Lancet North Am Ed. 1999; 353(9149):267–272.
- Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, Wratten C, Denham JW, Greer PB. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. Int J Radiat Oncol Biol Phys. 2012; 83(1):e5–e11. [PubMed: 22330995]
- Gao Y, Shao Y, Lian J, Wang A, Chen R, Shen D. Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests. IEEE Trans Med Imaging. 2016
- Heinrich MP, Jenkinson M, Bhushan M, Matin T, Gleeson FV, Brady M, Schnabel JA. MIND: modality independent neighbourhood descriptor for multi- modal deformable registration. Med Image Anal. 2012; 16(7):1423–1435. [PubMed: 22722056]
- Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, Shen D. Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Trans Med Imaging. 2016; 35(1): 174–183. [PubMed: 26241970]
- Iglesias, JE., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B. Is synthesizing MRI contrast useful for inter-modality analysis. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2013.

- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001; 5(2):143–156. [PubMed: 11516708]
- Jia H, Yap PT, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. Neuroimage. 2012; 59(1):422–430. [PubMed: 21807102]
- Jog, A., Carass, A., Pham, DL., Prince, JL. Random forest FLAIR reconstruction from T1, T2, and PD-weighted MRI. 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); IEEE; 2014a.
- Jog, A., Carass, A., Prince, JL. Improving magnetic resonance resolution with supervised learning. 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); IEEE; 2014b.
- Jog, A., Roy, S., Carass, A., Prince, JL. Magnetic resonance image synthesis through patch regression. 2013 IEEE 10th International Symposium on Biomedical Imaging; IEEE; 2013.
- Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. Med Phys. 2011; 38(5):2708–2714. [PubMed: 21776807]
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009; 46(3):786–802. [PubMed: 19195496]
- Klein S, van der Heide UA, Lips IM, van Vulpen M, Staring M, Pluim JP. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys. 2008; 35(4):1407–1417. [PubMed: 18491536]
- Li, R., Zhang, W., Suk, H-I., Wang, L., Li, J., Shen, D., Ji, S. Deep learning based imaging data completion for improved brain disease diagnosis. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2014.
- Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002; 2(3):18-22.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. Med Imag IEEE Trans. 1997; 16(2):187–198.
- Metcalfe P, Liney G, Holloway L, Walker A, Barton M, Delaney G, Vinod S, Tome W. The potential for an enhanced role for MRI in radiation-therapy treatment planning. Technol Cancer Res Treat. 2013; 12(5):429–446. [PubMed: 23617289]
- Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. Med Image Anal. 2011; 15(4):622–639. [PubMed: 20688559]
- Partin A, Yoo J, Carter HB, Pearson J, Chan D, Epstein J, Walsh P. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. J Urol. 1993; 150(1):110–114. [PubMed: 7685418]
- Pluim JP, Maintz JA, Viergever MA. Mutual-information-based registration of medical images: a survey. Med Imag IEEE Trans. 2003; 22(8):986–1004.
- Roach M, Faillace-Akazawa P, Malfatti C, Holland J, Hricak H. Prostate volumes defined by magnetic resonance imaging and computerized tomographic scans for three-dimensional conformal radiotherapy. Int J Radiat Oncol Biol Phys. 1996; 35(5):1011–1018. [PubMed: 8751410]
- Roche A, Pennec X, Malandain G, Ayache N. Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information. Med Imag IEEE Trans. 2001; 20(10): 1038–1049.
- Roy, S., Carass, A., Prince, J. A compressed sensing approach for MR tissue contrast synthesis. Biennial International Conference on Information Processing in Medical Imaging; Springer; 2011.
- Roy S, Carass A, Prince JL. Magnetic resonance image example-based contrast synthesis. IEEE Trans Med Imaging. 2013; 32(12):2348–2363. [PubMed: 24058022]
- Roy, S., Carass, A., Shiee, N., Pham, DL., Prince, JL. MR contrast synthesis for lesion segmentation. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; IEEE; 2010.
- Roy S, Wang WT, Carass A, Prince JL, Butman JA, Pham DL. PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging. J Nucl Med. 2014; 55(12):2071–2077. [PubMed: 25413135]
- Sannazzari G, Ragona R, Ruo Redda M, Giglioli F, Isolato G, Guarneri A. CT–MRI image fusion for delineation of volumes in three-dimensional conformal radiation therapy in the treatment of localized prostate cancer. Br J Radiol. 2002; 75(895):603–607. [PubMed: 12145134]

- Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging. 2002; 21(11):1421–1439. [PubMed: 12575879]
- Shen D, Lao Z, Zeng J, Zhang W, Sesterhenn IA, Sun L, Moul JW, Herskovits EH, Fichtinger G, Davatzikos C. Optimized prostate biopsy via a statistical atlas of cancer spatial distribution. Med Image Anal. 2004; 8(2):139–150. [PubMed: 15063863]
- So, RW., Chung, A. Learning-based multi-modal rigid image registration by using Bhattacharyya distances. Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE; IEEE; 2011.
- Studholme C, Hill DL, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. 1999; 32(1):71–86.
- Toews, M., Zöllei, L., Wells, WM, III. Feature-based alignment of volumetric multi-modal images. Information Processing in Medical Imaging: Proceedings of The... Conference, NIH Public Access; 2013.
- Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. Pattern Anal Mach Intell IEEE Trans. 2010; 32(10):1744–1757.
- Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. Neuroimage. 2009; 45(1):S61–S72. [PubMed: 19041946]
- Viola P, Jones MJ. Robust real-time face detection. Int J Comput Vision. 2004; 57(2):137-154.
- Viola P, Wells WM III. Alignment by maximization of mutual information. Int J Comput Vision. 1997; 24(2):137–154.
- Wang L, Gao Y, Shi F, Li G, Gilmore JH, Lin W, Shen D. LINKS: learning- based multi-source IntegratioN framework for segmentation of infant brain images. Neuroimage. 2015; 108:160–172. [PubMed: 25541188]
- Wei, L., Hu, S., Gao, Y., Cao, X., Wu, G., Shen, D. Learning appearance and shape evolution for infant image registration in the first year of life. International Workshop on Machine Learning in Medical Imaging; Springer; 2016.
- Wein W, Brunke S, Khamene A, Callstrom MR, Navab N. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Med Image Anal. 2008; 12(5):577–585. [PubMed: 18650121]
- Wells WM, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. Med Image Anal. 1996; 1(1):35–51. [PubMed: 9873920]
- Ye, DH., Zikic, D., Glocker, B., Criminisi, A., Konukoglu, E. Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2013.
- Zacharaki EI, Hogea CS, Shen D, Biros G, Davatzikos C. Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. Neuroimage. 2009; 46(3):762–774. [PubMed: 19408350]
- Zhang J, Gao Y, Gao Y, Munsell BC, Shen D. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. IEEE Trans Med Imaging. 2016; 35(12):2524–2533. [PubMed: 27333602]



#### Fig. 1.

Pelvic CT and MRI from the same patient, which scanned at different time points. Left two images: CT and labeled CT; Right two images: labeled MRI and MRI. The CT and MRI have already been linearly registered.



#### Fig. 2.

The framework of the proposed dual-core steered MRI-to-CT image registration based on bi-directional image synthesis. The actual CT and actual MRI are already linearly aligned.





An illustration of training and testing in patch-wise random forest for MR image synthesis from CT.



#### Fig. 4.

An illustration of Haar-like features extraction within a single patch. Two ways of Haar-like feature generation: (a) the averaged intensities within a sub-block; (b) the averaged intensity differences between two sub-blocks.





Iterative refinement of synthesized MRI with auto-context model.



Fig. 6. Visualization of the synthesized CT image (S-CT) from MRI for one subject.



**Fig. 7.** Visualization of the synthesized MR image (S-MRI) from CT for one subject.



# Fig. 8.

MRI-to-CT non-rigid registration results by using the proposed registration method. (a) The mean DSC values of prostate, bladder and rectum by different number of ACM layers in image synthesis. (b) The mean DSC values of prostate, bladder and rectum with respect to different DDF iterations in Algorithm 2. Note that, 3-iter DDF is applied in (a), while 2-layer ACM is used in (b).



#### Fig. 9.

Examples of the synthesized images and the registration results (with ANTs-SyN) for 3 subjects in the dataset. Yellow contours are the contours of the prostate, bladder and rectum of the original CT image, used as ground-truth. Red Contours are the warped contours of the prostate, bladder and rectum after respective registrations. (For interpretation of the

references to colour in this figure legend, the reader is referred to the web version of this article.)

Parameter settings of intra-subject CT and MRI registration for generating the pre-aligned dataset.

|                   | Affine (FLIRT) (Intensity image) | Non-rigid (ANTs-SyN) (Intensity image)                                | Non-rigid (D. Demons) (Label<br>image) |
|-------------------|----------------------------------|---|--|
| Parameter Setting | flirt                            | antsRegistration  | DemonsRegistration                     |
|                   | -in \$MR_intensity_image         | -d 3  | -f \$CT_label_image                    |
|                   | -ref \$CT_intensity_image        | -m MI [\$CT_intensity_image,  | -m \$MR_label_image                    |
|                   | -out \$registered_image          | \$MR_intensity_image, 1]  | -o \$warped_label_image                |
|                   | -omat transformation.mat         | -o [\$output_file]  | -O \$out_deformationfield              |
|                   | -cost MI                         | -t SyN [0.3, 3, 0]  | -s 2                                   |
|                   | -searchcost MI                   | -c $[80 \times 80 \times 60, 1.e-6, 10]$                              | -a 0                                   |
|                   | -searchrx -20 20                 | -s $4 \times 2 \times 1$ vox  | -i $30 \times 20 \times 20$            |
|                   | -searchry -20 20                 | -f $3 \times 2 \times 1$  |  |
|                   | -searchrz -20 20                 | *common setting, parameters fine-tuned for each<br>individual subject |  |

Parameter setting for patch-wise random forest and auto-context model based image synthesis.

| Parameters | Input patch size | Output patch size                                | Tree number | Tree depth | Auto-context |
|------------|------------------|--|-------------|------------|--------------|
| MRI-to-CT  | 15 15            |  | 20          | 19         | 2-layer      |
| CT-to-MRI  | CI × CI × CI     | $\mathbf{c} \times \mathbf{c} \times \mathbf{c}$ | 25          | 24         |              |

Quantitative results of *CT image synthesis from MRI* by voxel-wise random forest (**V-RF**) and patch-wise random forest (**P-RF**) with auto-context model refinements (**1-layer ACM** and **2-layer ACM**).

|      | V-RF             | P-RF           | 1-layer ACM (P-RF) | 2-layer ACM (P-RF) |
|------|------------------|----------------|--------------------|--------------------|
| MAE  | $58.61 \pm 4.63$ | $51.24\pm3.81$ | $40.97\pm3.64$     | 38.64 ± 3.42 *     |
| PSNR | $29.17\pm0.91$   | $31.68\pm0.89$ | $33.80\pm0.88$     | $34.34 \pm 0.91$ * |

\* indicates significant improvements via paired *t*-tests, p < 0.05.

Quantitative results of *MRI image synthesis from CT* by voxel-wise random forest (**V-RF**) and patch-wise random forest (**P-RF**) with auto-context model refinements (**1-layer ACM** and **2-layer ACM**).

|      | V-RF             | P-RF            | 1-layer ACM (P-RF) | 2-layer ACM (P-RF)           |
|------|------------------|-----------------|--------------------|------------------------------|
| MAE  | $149.81\pm10.32$ | $139.48\pm9.79$ | $128.11\pm8.81$    | $124.03 \pm 8.14  {}^{\ast}$ |
| PSNR | $24.47 \pm 0.94$ | $25.50\pm0.83$  | $26.01\pm0.82$     | $26.32 \pm 0.82$ *           |

\* indicates significant improvements via paired *t*-tests, p < 0.05.

The bold is used to show the best performance.

Implementation details for D. Demons and ANTs-SyN, respectively.

|                   | D. Demons                   | ANTs-SyN                      |
|-------------------|-----------------------------|-------------------------------|
| Parameter Setting | DemonsRegistration          | antsRegistration              |
|                   | -f \$intensity_image        | -d 3                          |
|                   | -m \$intensity_image        | -m MI [\$intensity_image,     |
|                   | -o \$warped_image           | \$intensity_image, 1]         |
|                   | -O \$out_deformationfield   | -o [\$output_file]            |
|                   | -s 2                        | -t SyN [0.4, 3, 0]            |
|                   | -a 0                        | -c $[30 \times 30 \times 20]$ |
|                   | -i $20 \times 15 \times 15$ | -s $4 \times 2 \times 1$ vox  |
|                   |                             | -f $3 \times 2 \times 1$      |

Author Manuscript

# Table 6

Comparison of DSC (%) values of three pelvic organs after non-rigid registration through the original CT and MRI (CT & MRI), single-directional image synthesis (CT & S-CT, S-MRI & MRI), and our proposed bi-directional image synthesis under 3-iter DDF (Proposed).

| Method        | Region   | CT & MRI       | CT & S-CT        | S-MRI & MRI      | Proposed             |
|---------------|----------|----------------|------------------|------------------|----------------------|
| D. Demons     | Prostate | N/A            | $86.29\pm5.01$   | $87.44\pm6.86$   | $88.88 \pm 4.34$     |
|               | Bladder  | N/A            | $91.09\pm1.01$   | $91.51\pm0.85$   | $93.21 \pm 0.48^{*}$ |
|               | Rectum   | N/A            | $84.29\pm3.06$   | $84.13\pm6.18$   | $86.63 \pm 2.54$     |
| ANTs-SyN (MI) | Prostate | $86.84\pm3.52$ | $87.88 \pm 2.93$ | $87.31\pm3.38$   | $89.21 \pm 2.82^{*}$ |
|               | Bladder  | $90.41\pm0.39$ | $91.36\pm0.46$   | $91.53\pm0.69$   | $93.04 \pm 0.32$     |
|               | Rectum   | $83.68\pm4.69$ | $85.03 \pm 4.24$ | $85.19 \pm 5.38$ | $87.22 \pm 3.23$ *   |

indicates significant improvement via paired *t*-tests (p < 0.05).

Comparison of mean **SEN** and **PPV** values of three pelvic organs after non-rigid registration based on singledirectional image synthesis (**CT & S-CT, S-MRI & MRI**), and our proposed bi-directional image synthesis under 3-iter DDF (**Proposed**).

| Metric  | Method    | Single-          | <b>Bi-directional</b> |   |
|---------|-----------|------------------|-----------------------|---|
|         |           | CT & S-CT        | S-MRI & MRI           | Proposed  |
| SEN (%) | D. Demons | $80.41 \pm 8.48$ | $83.55\pm7.31$        | $\textbf{86.59} \pm \textbf{6.28}^{\texttt{*}}$ |
|         | ANTs-SyN  | $83.87\pm7.10$   | $84.27\pm6.30$        | $\textbf{86.39} \pm \textbf{5.21}^{\texttt{*}}$ |
| PPV (%) | D. Demons | $93.43 \pm 3.22$ | $94.39 \pm 2.10$      | $\textbf{95.16} \pm \textbf{1.82}^{\texttt{*}}$ |
|         | ANTs-SyN  | $93.42\pm2.95$   | $94.29 \pm 2.28$      | $\textbf{95.40} \pm \textbf{1.69}^{\texttt{*}}$ |

indicates significant improvement via paired t-tests (p < 0.05).

Comparison of mean **SASD** and **HAUS** values of three pelvic organs after non-rigid registration based on single-directional image synthesis (**CT & S-CT, S-MRI & MRI**), and our proposed bi-directional image synthesis under 3-iter DDF (**Proposed**).

| Metric    | Method    | Single-         | Single-directional |  |
|-----------|-----------|-----------------|--------------------|--|
|           |           | CT & S-CT       | S-MRI & MRI        | Proposed                                       |
| SASD (mm) | D. Demons | $1.25\pm0.67$   | $1.68\pm0.79$      | $\textbf{1.03} \pm \textbf{0.64}^{\texttt{*}}$ |
|           | ANTs-SyN  | $1.38\pm0.77$   | $1.25\pm0.74$      | $\textbf{1.10} \pm \textbf{0.71}^{~*}$         |
| HAUS (mm) | D. Demons | $8.89 \pm 2.74$ | $8.57\pm3.01$      | $\textbf{6.71} \pm \textbf{2.30}^{*}$          |
|           | ANTs-SyN  | $7.63 \pm 2.73$ | $7.18 \pm 2.02$    | $\textbf{6.69} \pm \textbf{1.94}^{\texttt{*}}$ |

indicates significant improvement via paired t-tests (p < 0.05).