Published in final edited form as:

Simul Synth Med Imaging. 2017 September ; 10557: 33–40. doi:10.1007/978-3-319-68127-6_4.

A supervoxel based random forest synthesis framework for bidirectional MR/CT synthesis

Can Zhao¹, Aaron Carass¹, Junghoon Lee², Amod Jog¹, Jerry L. Prince¹

¹Dept. of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218

²Dept. of Radiation Oncology, The Johns Hopkins School of Medicine, Baltimore, MD 21287

Abstract

Synthesizing magnetic resonance (MR) and computed tomography (CT) images (from each other) has important implications for clinical neuroimaging. The MR to CT direction is critical for MRI-based radiotherapy planning and dose computation, whereas the CT to MR direction can provide an economic alternative to real MRI for image processing tasks. Additionally, synthesis in both directions can enhance MR/CT multi-modal image registration. Existing approaches have focused on synthesizing CT from MR. In this paper, we propose a multi-atlas based hybrid method to synthesize T1-weighted MR images from CT and CT images from T1-weighted MR images using a common framework. The task is carried out by: (a) computing a label field based on supervoxels for the subject image using joint label fusion; (b) correcting this result using a random forest classifier (RF-C); (c) spatial smoothing using a Markov random field; (d) synthesizing intensities using a set of RF regressors, one trained for each label. The algorithm is evaluated using a set of six registered CT and MR image pairs of the whole head.

Keywords

synthesis; MR; CT; JLF; segmentation; random forest; MRF

1 Introduction

Synthesizing computed tomography (CT) images from magnetic resonance (MR) images has proven useful in positron emission tomography (PET)-MR image reconstruction [16, 4] and in radiation therapy planning [5]. To overcome the lack of a strong MR signal in bone, one method [16] used specialized MR pulse sequences and another method [4] used multi-atlas registration with paired CT-MR atlas images. The synthesis of MR images from CT images is a new challenge that has not been reported until very recently [6, 18]. Potential uses for this process include 1) intraoperative imaging where visualization of soft tissue from cone-beam CT could be enhanced by generation of a synthetic MR image and 2) in multi-modal registration where use of both modalities can improve the accuracy of registration [7, 9]. The difficulty in CT-to-MR synthesis is the lack of a strong soft-tissue contrast in the source CT images. Given the duality that appears between these tasks, we have discovered a core organizing principle for bi-directional image synthesis and developed a new image synthesis approach.

To synthesize CT images from MR images, Burgos et al. [4] used multiple CT/MR atlas pairs, wherein the atlas MR images are deformably registered to the target MR image. The transformations are then applied to the atlas CT images and fused to form a single CT intensity. Although this approach can also be used to synthesize MR from CT, some degree of blurring can be expected due to the inaccuracies in registration due to poor soft-tissue contrast in the CT images. Machine-learning approaches that have been developed for image synthesis (cf. [8, 15]) can also be used for synthesizing MR from CT; but image patches by themselves do not contain sufficient information to distinguish tissue types without additional information about the location of the patches.

Image segmentation has long been used for image synthesis [14]. If the tissue type and physical properties are known, then given the forward model of the imaging modality, the corresponding tissue intensity can be estimated. However, in our framework, segmented regions are used to provide *context* wherein synthesis can be carried out through a set of learned regressions that relate the intensities of the input modality to those of the target modality. We demonstrate synthesis in both directions, MR to CT and CT to MR, using our method.

2 Methods

Given a subject image of modality 1 ($M1$), denoted I^{M1} , our goal is to synthesize an image of modality 2 ($M2$), $\hat{I}^{M2}$. To achieve this goal, we have a multi-atlas set, $\mathcal{A} = \{(A_n^{M1}, A_n^{M2}) | n = 1, \dots, N\}$, which contains N pairs of co-registered images of $M1$ and $M2$. An example of an atlas pair, where $M1$ is CT and $M2$ is MR (T1-weighted) is depicted in Fig. 1(a). The two intensities in atlas image pairs are examples of possible synthetic values, when synthesizing in either direction. It is well known that this relationship is not a bijection; given an intensity in $M1$ there may be multiple corresponding intensities in $M2$. However, given a particular tissue (e.g., white matter) the relationship is less ambiguous. We carry out a segmentation on the atlas images that divides them into distinct regions characterized by different paired intensities. Paired intensities from these regions are then used to train separate regressors that predict one modality from the other given the tissue class.

We start with the atlas image set $\mathcal{A}$. Each pair of atlas images is processed with the following steps with the eventual goal of learning regressions that predict the target modality given the input modality. The first step is a supervoxel over-segmentation process using a 3D version of the *simple linear iterative clustering* (SLIC) method [1] wherein the intensity feature space comprises the $M1$ and $M2$ intensity pairs. A result of SLIC on two atlas pairs is shown in Fig. 1(b). Multichannel k -means and fuzzy k -means have been previously used for tissue classification in neuroimaging [13]. However, it is difficult to obtain spatially contiguous regions using these simple methods. Super-voxel over-segmentation provides us with spatially contiguous regions that have homogeneous intensities.

We combine these homogeneous intensity regions by clustering them on the basis of their average supervoxel intensities taken jointly from both $M1$ and $M2$. These are clustered using the k -means clustering algorithm, which yields supervoxels that are labeled $z = 1, \dots, K$. The

voxels forming each supervoxel inherit the cluster label of the supervoxel and therefore yield an image of labels, which we call the z -field. Two examples of z -fields are shown in Fig. 1(c), where each label in the z -field is shown as a different color. A random selection of intensity pairs are plotted in the center of Fig. 1(c) (CT/MR on the horizontal/vertical axis), and colored by the z -field. These intensity pairs and their voxel-wise features, along with their labels provide the training data for regressors that predict the intensity of the target modality given the features of the input modality. Our features consist of $3 \times 3 \times 3$ image patches together with average image values in patches forming a constellation around the given voxel (“context features” similar to those in [2, 10]). We need $2 \times K$ regressors, one each per modality and cluster. For each label z , we extract features from M1 images and pair them with corresponding M2 intensities. This acts as the training data set for a random forest (RF) regressor. The training step is depicted in Figure 1(d).

Given the subject image I^M and the corresponding z -field that labels its voxels, we can apply the corresponding regressor based on the z value at that voxel to predict the synthetic M2 intensity. Thus, we next describe how to estimate the z -field for the subject image. The z -field of I^M is estimated by fusing two approaches. First, we predict an estimate of the z -field directly from the same image features that were noted above using a *random forest classifier* (RF-C). Shown in Fig. 1(e), are two random forests designed to synthesize K labels from either M1 or M2, which are trained in analogous fashion to the RF regressors described above. A second estimate of the z -field is generated using a multi-atlas segmentation. In this case, we augment the atlases to include the z -fields found using the supervoxel clustering approach (essentially augmenting the image pairs in Fig. 1(a) with the label fields in Fig. 1(c)), deformably register every atlas pair to I^M , apply the learned transformations to the corresponding z -fields, and combine the labels using *joint label fusion* (JLF) [17]. The registration between I^M and the atlas pair uses a two-channel approach in which the first channel uses the cross-correlation metric between I^M and A^M and the second channel uses the mutual information metric between I^M and A^{M2} .

We now have two estimates of the z -field for I^M , $\hat{z}_{\text{RF-C}}$ and $\hat{z}_{\text{JLF}}$, each provides a probability for each label at each voxel, $P_{\text{RF-C}}(z)$ and $P_{\text{JLF}}(z)$. Our experiments reveal that the RF-C yields inferior results in regions where intensities of the labels are ambiguous, while the JLF yields inferior results in areas where the registration is not accurate. We choose the label that maximizes the product of their probabilities at each voxel with a MRF spatial regularization.

Using a conventional MRF framework, we define the estimated z -field,

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \sum_i E_{\text{unary}}(z(i)) + \sum_{i,j} E_{\text{binary}}(z(i), z(j)), \quad (1)$$

where $E_{\text{unary}}(z(i))$ is the unary potential for voxel i and $E_{\text{binary}}(z(i), z(j))$ is the binary potential for adjacent (6-connected) voxels i and j . Since this energy will be used in a Gibbs distribution, the unary potential is defined as follows

$$E_{\text{unary}}(z(i)) = -\log P_{\text{RF-C}}(z(i)) - \log P_{\text{JLF}}(z(i)) \quad (2)$$

which yields the desired product of probabilities as the driving objective function for assigning labels to voxels.

Although the Potts model is often used in multi-label MRF models [11]—this is the model in which different labels have unity cost and similar labels have zero cost—we can exploit our atlas and its subsequent analysis to yield a cost function that is highly tailored to our application. Consider the z -fields produced by over-segmentation followed by k -means, as shown in Fig. 1(c), and consider adjacent voxels i and j . From the full collection of these images, we can compute the empirical joint probability mass function $P(z(i), z(j))$ for all adjacent voxels, as illustrated in Fig. 1(f). Some labels will almost never appear adjacent to each other and thus should be penalized heavily in the MRF we design. Accordingly, we define the binary potential as

$$E_{\text{binary}}(z(i), z(j)) = -\log P(z(i), z(j)) + \frac{1}{2}(\log P(z(i), z(i)) + \log P(z(j), z(j))). \quad (3)$$

When the labels are the same the cost is zero and when they are different, the cost increases according to their rarity of occurrence in the atlas. Given these definitions of unary and binary potentials (which is a semimetric), the estimated z -field is found by solving (1) using the α - β swap graph cut approach [3].

3 Experiments

MR images were obtained using Siemens Magnetom Espree 1.5 T scanner (Siemens Medical Solutions, Malvern, PA) and CT images were obtained using Philips Brilliance Big Bore scanner (Philips Medical Systems, Netherlands) under the routine clinical protocol from brain cancer patients treated by stereotactic-body radiation therapy (SBRT) or radiosurgery (SRS). Geometric distortions in MR images were then corrected using a 3D correction algorithm available in the Siemens Syngo console workstation. All MR images were then N4 corrected and normalized by aligning white matter peak identified by fuzzy C-means.

We applied our method to six subjects each having true CT and MR images to compare our results to. For algorithm comparison, we implemented [12] the intensity fusion method of Burgos et al. [4] using *structural similarity* (SSIM) as the local similarity measure instead of local normalized cross correlation (LNCC), which we refer to as Burgos+. Existing work on CT/MR synthesis [4] has focused on synthesizing CT from MR, so we can directly compare. Without a published method for synthesizing MR from CT, we simply applied Burgos+ in the reverse direction. To evaluate efficacy of synthesis, we computed SSIM and PSNR on the synthetic images with respect to the true images. The result is shown in Figure 2. In addition to the comparison with Burgos+, we have shown how well modifications of our own algorithm perform. The “JLF” result uses only the z -field computed from JLF, the “RF-

C” result uses only the z -field computed from RF-C, the “JLF+RF-C” result uses the product of the two z -field probabilities without MRF; and the “MRF” result is our proposed algorithm. We can see our method gives better synthetic MR in every respect, while the synthetic CT images are better than Burgos+ for SSIM and comparable for PSNR.

Figure 3 shows the estimated z -fields and final synthetic CT images for two subjects. It shows that our synthetic CT images have higher contrast and no blurry edges as compared to Burgos+, yet look somewhat artificial compared to the truth. Figure 4 shows the estimated z -fields and final synthetic MR images for the same two subjects. It shows that our synthetic MR images also have high contrast and no blurry edges as compared to Burgos+. We notice in Fig. 4(e), the result from Burgos+ cannot synthesize the soft tissues correctly. This is because the result depends on the accuracy of registration between atlas image pairs and subject CT images, which is relatively low in areas of soft tissues. Our method is more robust to registration inaccuracies because we use a MRF to predict the z -field and the K random forests used in synthesis overlap in their intensity coverage to some extent.

To evaluate whether our synthesis method improves multi-modal registration, we carried out a multi-modal registration experiment between the CT image of one subject and the MR image of another subject. The conventional approach for multi-modal registration uses mutual information (MI) as a similarity metric. With synthetic images, multi-modal registration can be carried out using a two-channel mono-modal registration process [7, 9]. In our case, for registration between Subject 1 and Subject 2, the first channel uses the original CT image of Subject 1 and the synthetic CT for Subject 2. The second channel uses the synthetic MR image of Subject 1 and the original MR image of Subject 2. The metric used in both channels is cross correlation (CC).

We used SyN deformable registration on 6 subjects yielding 30 pairs of registration experiments in all. The single MI registration and two-channel CC registration share the same parameters, including the number of iterations. As the true MR image is known, we compare the transformed MR image to the true MR image for each individual registration experiment. The difference between these two images is measured using both MSE and MI after either two-channel CC or single-channel MI (results are in Table 1). While the two images are not statistically different according to MI, the two-channel registration approach (which uses our synthetic images) is statistically better than the single-channel MI approach.

4 Conclusion

We have presented a bidirectional MR/CT synthesis method based on approximate tissue classification and image segmentation. The method synthesizes CT images from MR images with performance comparable to Burgos et al. [4] and is better than Burgos et al. [4] for synthesizing MR images from CT images. Our method reduces intensity ambiguity by estimating a z -field that is derived from both modalities and can be consistently created given just one modality as input.

Acknowledgments

This work was supported by NIH/NIBIB grant R01-EB017743.

References

1. Achanta R, et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Patt. Anal. Mach. Intell.* 34(11):2274–2282.2012;
2. Bai W, et al. Multi-atlas segmentation with augmented features for cardiac MR images. *Medical Image Analysis.* 19(1):98–109.2015; [PubMed: 25299433]
3. Boykov Y, et al. Fast approximate energy minimization via graph cuts. *IEEE Trans. Patt. Anal. Mach. Intell.* 23(11):1222–1239.2001;
4. Burgos N, et al. Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies. *IEEE Trans. Med. Imag.* 33(12):2332–2341.2014;
5. Burgos, N, , et al. *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2015. Robust CT synthesis for radiotherapy planning: application to the head and neck region; 476–484.
6. Cao X, et al. Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. *Medical Image Analysis.* 2017
7. Chen M, et al. Cross contrast multi-channel image registration using image synthesis for MR brain images. *Medical Image Analysis.* 36:2–14.2017; [PubMed: 27816859]
8. Huynh T, et al. Estimating CT image from MRI data using structured random forest and auto-context model. *IEEE Trans. Med. Imag.* 35(1):174–183.2016;
9. Iglesias, JE, , et al. *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2013. Is synthesizing mri contrast useful for inter-modality analysis?; 631–638.
10. Jog A, et al. Random forest regression for magnetic resonance image synthesis. *Medical Image Analysis.* 35:475–488.2017; [PubMed: 27607469]
11. Komodakis, N. *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.* Vol. 1. IEEE; 2006. Image completion using global optimization; 442–452.
12. Lee, J, , et al. *SPIE Medical Imaging. International Society for Optics and Photonics;* 2017. Multi-atlas-based CT synthesis from conventional MRI with patch-based refinement for MRI-based radiotherapy planning; 101331I–101331I–6.
13. Pham DL, et al. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering.* 2(1):315–337.2000;
14. Riederer SJ, et al. Automated MR image synthesis: feasibility studies. *Radiology.* 153(1):203–206.1984; [PubMed: 6089265]
15. Roy S, et al. Magnetic resonance image example-based contrast synthesis. *IEEE Trans. Med. Imag.* 32(12):2348–2363.2013;
16. Roy S, et al. PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging. *Journal of Nuclear Medicine.* 55(12):2071–2077.2014; [PubMed: 25413135]
17. Wang H, et al. Multi-atlas segmentation with joint label fusion. *IEEE Trans. Patt. Anal. Mach. Intell.* 35(3):611–623.2013;
18. Zhao, C, , et al. *8th International Workshop on Machine Learning in Medical Imaging. Lecture Notes in Computer Science.* Springer; Berlin Heidelberg; 2017. Whole brain segmentation and labeling from CT using synthetic MR images.

Fig. 1.

(a) Two CT/MR atlas pairs; (b) result of SLIC over-segmentation; (c) k -means clustering of supervoxels yields a z -field image; (d) training of $2 \times K$ RF regressors; (e) RF-Cs trained to estimate z -fields from single modalities; and (f) computation of pairwise potentials for a MRF.

Fig. 2. Evaluation of synthesis result
The six colors are for six subjects.

Fig. 3. Synthetic CT images

For two subjects, one in each row, we show the (a) input MR image, the (b) estimated z-field after MRF smoothing, the CT images generated by (c) our method, (d) Burgos+, and the (e) ground truth.

Fig. 4. Synthetic MR images

For two subjects, one in each row, we show the (a) input CT image, the (b) estimated z -field after MRF smoothing, the MR images generated by (c) our method, (d) Burgos+, and the (e) ground truth.

Table 1
Evaluation of registration results

Mean (and Std. Dev.) of MSE between reference MR and registered MR image; MI between target CT and registered MR image; p -value of paired-sample t-test for the MSE and MI of the two methods.

	MSE	MI
2 Channel CC	2.746(± 0.6492) $\times 10^4$	1.2314(± 0.0746)
Single Channel MI	3.375(± 0.6635) $\times 10^4$	1.2429(± 0.1018)
p-value over Single Channel MI	8.7637e-16	0.1962