Published in final edited form as:

Phys Med Biol. ; 64(8): 085001. doi:10.1088/1361-6560/ab0b66.

MRI-based synthetic CT generation using semantic random forest with iterative refinement

Yang Lei¹, Joseph Harms¹, Tonghe Wang¹, Sibao Tian¹, Jun Zhou¹, Hui-Kuo Shu¹, Jim Zhong¹, Hui Mao², Walter J Curran¹, Tian Liu¹, Xiaofeng Yang^{1,3}

¹Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA, 30322, United States of America

²Department of Radiology and Imaging Sciences and Winship Cancer Institute, Emory University, Atlanta, GA, 30322, United States of America

Abstract

Target delineation for radiation therapy treatment planning often benefits from magnetic resonance imaging (MRI) in addition to x-ray computed tomography (CT) due to MRI's superior soft tissue contrast. MRI-based treatment planning could reduce systematic MR-CT co-registration errors, medical cost, radiation exposure, and simplify clinical workflow. However, MRI-only based treatment planning is not widely used to date because treatment-planning systems rely on the electron density information provided by CTs to calculate dose. Additionally, air and bone regions are difficult to separate given their similar intensities in MR imaging. The purpose of this work is to develop a learning-based method to generate patient-specific synthetic CT (sCT) from a routine anatomical MRI for use in MRI-only radiotherapy treatment planning. An auto-context model with patch-based anatomical features was integrated into a classification random forest to generate and improve semantic information. The semantic information along with anatomical features was then used to train a series of regression random forests based on the auto-context model. After training, the sCT of a new MRI can be generated by feeding anatomical features extracted from the MRI into the well-trained classification and regression random forests. The proposed algorithm was evaluated using 14 patient datasets with T1-weighted MR and corresponding CT images of the brain. The mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and normalized cross correlation (NCC) were 57.45 ± 8.45 HU, 28.33 ± 1.68 dB, and 0.97 ± 0.01 . We also compared the difference between dose maps calculated on the sCT and those on the original CT, using the same plan parameters. The average DVH differences among all patients are less than 0.2 Gy for PTVs, and less than 0.02 Gy for OARs. The sCT generation by the proposed method allows for dose calculation based MR imaging alone, and may be a useful tool for MRI-based radiation treatment planning.

Keywords

MRI-based treatment planning; synthetic CT; random forest

³Author to whom any correspondence should be addressed. xiaofeng.yang@emory.edu.

Disclosures

The authors declare no conflicts of interest.

1. Introduction

Magnetic resonance imaging (MRI) has several important advantages over computed tomography (CT) for radiation treatment planning. The principle advantage is more accurate and reliable organ-at-risk (OAR) and target delineation with superior soft tissue contrast offered by MRI over CT (Schmidt and Payne 2015). However, because treatment planning systems rely on electron density maps generated from CT images for calculating dose, MRIs are typically used as an adjunct to CTs for radiation treatment planning (van der Heide *et al* 2012). This can lead to CT/MR image co-registration errors; an inherent uncertainty of approximately 2 mm in cranial CT/MR image registration has been reported by Ulin *et al* (2010). A potential treatment planning process with MRI as the sole imaging modality could eliminate systematic CT/MR image co-registration errors, reduce cost, minimize patient radiation exposure, and simplify clinical workflow (Khoo and Joon 2006, Devic 2012). However, MRIs cannot provide the key electron density information that is needed for accurate dose calculation. Additionally, daily patient setup for radiotherapy is based on either cone-beam CTs or orthogonal planar x-ray images which are then compared to either planning CTs or digitally reconstructed radiographs (DRRs) generated from the planning CT. This setup process inherently relies on a CT image taken at the beginning of the treatment planning process.

Since electron density information and CT images are vital to the treatment-planning workflow, a method by which electron density information can be inferred from MR images, called synthetic CT (sCT) generation, has been investigated recently (Edmund and Nyholm 2012, Johnstone *et al* 2018). One common methodology, known as atlas-based methods, consists of performing a deformable registration between previously developed single or multiple MRI atlases and a new arrival MRI. The sCT image is estimated via image warping of the corresponding CT in the MRI-CT atlas pairs (Uh *et al* 2014, Sjölund *et al* 2015). The efficiency of these methods is limited by the accuracy of registration. Another class of methods first segments the MR image into different anatomic groups, such as bone, air, fat, and soft tissue, and then assigns a uniform electron density value to each group (Yang and Fei 2013, Korhonen *et al* 2014, Kim *et al* 2015). These methods are prone to misclassification of bone and air because the two materials produce similar MR intensities. Another approach to visualize the bony anatomy from MRI relies on the use of specialized MR sequences such as ultra-short echo time (UTE), or a mixture of the traditional and non-traditional MR acquisition schemes. Post-processing can then be performed on these images to generate a sCT (Johansson *et al* 2011, Edmund *et al* 2014, Cabello *et al* 2015). However, the current image quality of UTE sequences is inadequate for accurate delineation of blood vessels from bone, as they both appear dark (Hsu *et al* 1995, Gudur *et al* 2014). Moreover, the use of non-standard MR sequences may introduce additional scanning time to the existing MR scanning workflow and may increase the patient discomfort, leading to motion artifacts.

With the application of machine learning approaches to medical imaging, more sophisticated methods for sCT synthesis have been proposed. The model provided by machine learning learns the relationship between MR and CT images for several representative sets of CT/MR

image pairs in a process known as training. After training, sCT images can be generated by feeding new MR images into the model. Such sCT images will share structural information with MRI, but the intensity values will be scaled to typical Hounsfield units (HU) seen on CT. The electron density can then be derived from this sCT image. Based on different training models, these methods can be broadly classified into three categories: dictionary learning-based methods (Andreasen *et al* 2015, Aouadi *et al* 2016, Torrado-Carvajal *et al* 2016, Lei *et al* 2018), random forest-based methods (Huynh *et al* 2016, Andreasen *et al* 2016a, Yang *et al* 2017a, 2017b) and deep learning-based methods (Han 2017, Emami *et al* 2018, Nie *et al* 2018). Dictionary learning-based methods rely on a large set of paired MR and CT image patches. When a new patient's MR image is put into the model, similarity between the new MRI patches and training MRI patches is calculated. A sCT is then generated from a linear combination of the most similarly paired CT images, with the weights calculated based on the MR similarity. These methods are sensitive to MR intensities which can vary as a function of scanning parameters for a given tissue. As they also rely on large datasets, and the accuracy of the method is inherently dependent on the size of the dictionary. Random forest-based methods train a set of binary decision trees, allowing for flexible CT intensities. Each decision tree learns the best way to separate a set of paired MRI and CT patches into smaller and smaller subsets to predict the CT intensity. When a new MRI patch is put into the model, the sCT intensity is estimated as the combination of the predicted results of all decision trees. However, these methods can lead to ambiguous results when the MRI-to-CT mapping learned from decision trees is not one-to-one. Deep learning-based methods provide complex non-linear mapping from MRI to CT images through a multi-layer and fully trainable model. However, because of the need to optimize multiple parameters in the training model, and each iteration of optimization is only based on a small group of patches, these methods suffer from longer training times compared with random forest-based methods.

Recently, we developed a novel random forest-based method which considers intensity, frequency, and texture features as the representation of MRI (Yang *et al* 2017b). Compared with other random forest-based methods, the main advantage of the proposed method is the use of discriminative feature selection. The method has previously been shown to provide sCT images with similar dosimetric quality to planning CT images (Wang *et al* 2018). However, the lack of a one-to-one relation between MR features and CT intensity still exists. Thus, the method suffers from decreased image sharpness in sCT areas where large anatomical variation exists (such as nasal cavity and auditory meatus).

The purpose of this work is to develop a random forest-based method with anatomical signatures to derive sCT images from standard T1-weighted MRIs. In contrast to the previous method (Yang *et al* 2017b), where only texture features were used, we proposed to generate semantic information by using a classification random forest with an iterative refinement model (auto-context model). The semantic information was integrated into regression random forest-based method to enhance the model's ability for learning one-to-one MRI-CT mapping. We evaluated the method's accuracy of sCT synthesis and its ability to identify air, soft tissue, and bone in the head and skull base.

2. Methods and materials

2.1. System overview

For a pair of acquired MR and CT images, the CT image was used as the regression target of the MR image. Image pre-processing was applied for image registration. Our proposed algorithm, outlined graphically in figure 1, consisted of two stages, a training stage to learn the MRI-to-CT mapping and a synthesizing stage to generate a sCT from a new arrival MRI. In the training stage, material labels, i.e. air, soft tissue, and bone, were clustered from each CT images. Different features including discrete cosine transform (DCT), local binary pattern (LBP), and pairwise voxel difference features (Lei *et al* 2018) were extracted from each MR image. The most informative features were selected by a feature selection strategy. The informative features together with material labels were used to train a segmentation model using a classification random forest method. The posterior probability of material (called probability map) under MR image was generated by feeding the informative features to the well-trained model. Then, we trained a new segmentation model, not only on the informative features but also on the features extracted from a probability map (called semantic features). Once a new segmentation model was learned, the algorithm repeated the same procedure until convergence. The semantic features were derived from the probability map of the last iteration. These, together with informative features, were used to learn MRI-to-CT mapping by regression random forests. The sCT image was estimated by feeding these features to the well-trained mapping. Then, we trained a new mapping by the concatenation of original informative features, semantic features, and context features (Huynh *et al* 2016) extracted from the SCT image. Once a new mapping was learned, the algorithm repeated the same procedure until convergence. In synthesizing stage, we extracted features from a new MRI and fed them into the sequence of well-trained models and mappings for the sCT estimation.

2.2. Image acquisition

We retrospectively analyzed the MRI and CT data acquired during treatment planning for 14 patients who received cranial irradiation. The main patient selection criterion was that their MRIs were acquired with the same 3D sequences, had fine spatial resolution, and the entire head was imaged. Standard T1-weighted MRI was captured using a GE MRI scanner with magnetization-prepared rapid gradient echo (MP-PAGE) sequence and $1.0 \times 1.0 \times 1.4 \text{ mm}^3$ voxel size (TR/TE: 950/13 ms, flip angle: 90°) and CT was captured with a Siemens CT scanner with $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ voxel size with 120 kVp and 220 mAs.

2.3. Image pre-processing

MRI data were first resampled to match the resolution of CT data. For each patient, all training MR and CT images were first rigidly registered by an intra-subject registration using Velocity AI 3.2.1 (Varian Medical Systems, Palo Alto, CA). Inter-patient registration of MR images consisted of a rigid registration followed by a B-spline deformable image registration. The transformation matrix obtained during this registration process was applied to the CT images to generate the deformable template CT. CT images were first segmented by a previously described thresholding method (Huang and Wang 1995). Segmentation holes were filled by morphological erosion with a 5 voxel spherical kernel to eliminate any objects

that were not physically part of the patient. To improve the MR image quality, we performed denoising, inhomogeneity correction, and histogram matching for MRIs as recommended by our previous work (Lei *et al* 2018).

2.4. Semantic information

Previous random forest-based sCT synthesis methods have trained a series of regression forests by using patch-based MRI features and their associated CT patches as the training target. The MRI features and corresponding CT patches were extracted from paired training MR and CT images. After training, the sCT value of a new MRI was then estimated by feeding the new patch-based MRI features into a series of models (Huynh *et al* 2016, Yang *et al* 2017b). However, by setting a limited patch size, for example $15 \times 15 \times 15$ (Huynh *et al* 2016)), MRI patches often had similar structures or intensities around regions where CT patches were dissimilar. The extracted features on these MRI patches were similar, which can potentially lead to ambiguity in the model. Mapping similar MR patches to dissimilar CT patches can lead to erroneous sCT generation.

Figure 2 shows examples of regions where inclusion of semantic information can enhance the performance of random-forest based sCT models. Inserts (a1) and (a2) show axial brain MR images, with regions of interest (ROI) highlighted in red, and inserts (c1) and (c2) show the paired CT images. Inserts (b1) and (b2) show the MR ROIs while inserts (d1) and (d2) show the paired CT ROIs. While the MR patches have similar structures and intensities, they also have similar features, such as Haar-like patterns, pairwise-voxel differences, and discrete cosine transform (DCT) (Huynh *et al* 2016). However, the corresponding CT patches are clearly different. Previous random forest-based methods relying only on the features listed above can lead to erroneous prediction of the patches. Regardless of what splitting was learned by decision tree, the two different CT patches cannot be separated into different leaves by these similar MRI features. This means the model learned from MR to CT is not one-to-one. This feature-based approach to sCT generation does not give the model sufficient information to accurately differentiate these patches, which limits the applicability of the model.

In order to solve this issue, we propose to incorporate semantic information from MRIs to enforce the random forest-based model to train a one-to-one mapping. As the features of MRI patches are concatenated with label information, such as material type (i.e. bone, air, and soft tissue), this allows the training model to more closely approach one-to-one mapping. For example, in figure 2(b1) and (b2), if the features of these two MRI patches are combined with their labeling knowledge, it would be easier to accurately estimate the CT patches (d1) and (d2).

A simple threshold-based method was used for the segmentation of CT images and generation of label information. In the training stage, air, soft tissue, and bone labels were dictated by CT HU values: $[-\infty, -400]$ corresponds to air, $[-400, 300]$ is soft-tissue, and $[300, +\infty]$ is bone. Multi-level features were extracted as recommended from previous random forest-based methods (Huynh *et al* 2016). The multi-level features were concatenated to a feature vector which represents a given MRI patch. Then, as previously recommended (Lei *et al* 2018), we identified the most salient and informative features from

this feature vector using a logistic LASSO-based feature selection strategy. We utilized selected anatomical features together with the corresponding CT labels to train a segmentation model based on a classification random forest. Rather than obtaining the binary segmentation results when feeding the anatomical features to the well-trained segmentation model, the posterior probability of air, bone, and soft-tissue under the feeding features can also be estimated by maximizing a posterior as previously suggested (Tu and Bai 2010). We called the posterior volume of these three labels probability maps in this work. The probability maps were updated with each iteration, and their accuracy was increased by leveraging the information surrounding the object of interest (Tu and Bai 2010). To better approximate the margins of air, bone, and soft tissue material, we aimed to include a set of semantic information, i.e. the surrounding voxels' posterior (probability) under given MRI anatomical features to the objective voxel of interest. The surrounding voxels were located by a number of context locations which were chosen by the learning algorithm. Here, we proposed to use 27 locations for each voxel of interest. Figure 3 shows examples of semantic features extracted on context locations from a probability map. Insert (b1) shows an axial MR image, which we denoted x . Insert (a1) is the corresponding air label's probability map. The colormap window is scaled from 0 (blue) to 1 (red). (a1) Was generated by a thresholding method on the corresponding CT image and is regarded as ground truth. (a2) Shows the air probability map obtained from the first classification random forest-based segmentation model by feeding anatomical features. (b2) Shows the semantic feature extraction for the central voxel in highlighted ROI of (a2). The ROI size was set to $15 \times 15 \times 15$. As shown axially, we see nine highlighted smaller ROIs as context locations in (b2). The semantic features were generated by a concatenation of the means within these ROIs. Insert (a3) show the air probability map obtained from the second segmentation model. This model was trained not only using the features from the MR image, but also incorporating the semantic features as is given in (b2). Once a new segmentation model is well-trained, the probability map generation and semantic feature extraction procedure repeats the same procedure until convergence, as is shown in (a4), (b4) and (a5), (b5). The probability maps of soft tissue and bone are also given in figures 3(c1)–(c5) and (d1)–(d5). For the shown region of interest, the central voxel value of the highlighted ROI in (a5) has a relatively high posterior for air, while the posterior for soft tissue or bone is low in (c5) and (d5). Thus, this value can clearly separate the image labels, i.e. the separation of materials was best in the fourth segmentation model.

For implementation, the semantic information was concatenated to the original anatomical features, and this was stored as a feature vector. Then, together with a corresponding CT patch, we trained a series of classification and regression random forests to generate refined probability maps and a sCT patch that is very similar to the corresponding CT patch. This iterative refinement concept was called an auto-context model, as originally proposed by Tu and Bai (2010). In the synthesis stage, a new MRI patch follows the same sequence of training to generate a new sCT patch. Finally, a new sCT is generated by patch fusion, i.e. weighted averaging of the overlapping voxels. We termed this proposed method a semantic random forest-based method.

2.5. Validation and evaluations

We performed leave-one-out cross validation to evaluate the proposed semantic random forest-based method. Institutional review board approval was obtained with no informed consent required for this HIPAA-compliant retrospective analysis. We chose one of the 14 patients as the test, or new arrival, patient. The semantic random forest-based method was trained on the other 13 patients, and the 14th patient was used to generate a sCT. This procedure was repeated on all 14 patients' datasets. These synthetic sCT images were then compared with the original planning CT images, which served as ground truth for validation. Mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and normalized cross correlation (NCC) were used to quantify the absolute difference, relative difference, and image similarity within the body outline, respectively. To evaluate the ability of the semantic random forest-based approach in identifying bone, soft tissue, and air regions, Dice similarity coefficient (DSC) analysis (Chaibi and Nourine 2018) was performed on all test patients. True bone, air, and soft tissue were labeled by intensity thresholds as referred in section 2.3. DSC was used as a statistical validation metric to evaluate the spatial overlap accuracy of the sCT's air, bone, and soft tissue detection.

We further evaluated the sCT images by comparing their dosimetric performance to the planning CT, as previous authors have done for similar sCT generation algorithms (Andreasen *et al* 2016b, Aouadi *et al* 2017, Chen *et al* 2018, Largent *et al* 2019, Maspero *et al* 2018). Each of the 14 patients included in this retrospective study had at least one stereotactic radiosurgery (SRS) treatment plan based on their CT images. We compared the difference between dose maps calculated on the sCT and those on the original CT, using the same plan parameters. Note that in this study, MRI scans were not acquired for the purpose of simulation and thus the imaging parameters and patient setup differed from the CT simulation. Therefore, a rigid registration was needed between the original CT volume and its corresponding sCT (generated from MRI) to align the patient anatomy. In this study, we performed a rigid registration using commercial image registration software, Velocity AI 3.2.1 (Varian Medical Systems, Palo Alto, CA). The aligned sCT images were interpolated into the same voxel size as the original CT images and were then imported into the treatment planning system. The structures and treatment plan from the original CT volumes were duplicated onto the sCT volumes for dose calculation using the same algorithm (analytical anisotropic algorithm, AAA), grid size (0.1 cm) and HU calibration.

For each plan, we visually checked the similarity of the dose distributions calculated on sCT and original CT. Quantitatively, clinically relevant DVH metrics were extracted for comparison of dose to PTVs and relevant OARs.

2.6. Comparison with existing methods

To demonstrate the advantages of the proposed semantic random forest-based method, we compare it to a random-forest-based method previously proposed in Huynh *et al* (2016). We also compare it to a generative adversarial network (GAN)-based method, a recent deep learning-based method proposed by Nie *et al* (2018). GAN-based methods incorporate an adversarial learning strategy into an end-to-end fully convolutional network to train nonlinear mapping from MRI patches to CT image patches. Considering the tradeoff

between computational cost and spatial information, the patch size in Nie *et al* (2018) was set to $32 \times 32 \times 32$. All the comparison algorithms were performed using their optimal parameter settings. Paired two-tailed *t*-tests between our semantic random forest-based method and comparison methods were performed to quantify the statistical difference between each of the evaluated metrics above.

3. Results

In order to test the influence of semantic information, we compared sCT results generated by a previous random forest-based method to the proposed method. Figure 4 shows axial, sagittal and coronal views of MR (a1)–(a3), corresponding CT (b1)–(b3), and sCT (c1)–(c3) and (e1)–(e3) images. The difference between sCT and CT images are also shown (d1)–(d3) and (f1)–(f3). The previous random forest-based method captures the gross anatomy and preserves the structural details in soft tissue. However, it distorts structural details in regions with large anatomical variation, especially in the regions marked by the highlighted rectangles in figures 4(c1)–(c3). Combining both semantic information and multi-level features used in a random forest-based method, the proposed method improves the estimation of CT intensity and better preserves the tissue structural details. The difference images in figures 4(f1)–(f3) demonstrate the superior accuracy of the proposed method with less bias around bony anatomy.

Semantic information was added to the random forest model to more closely approximate one-to-one mapping from the MRI domain to the sCT domain. To show the impact of including this semantic information on a voxel-by-voxel basis, we compared the classification capabilities of semantic information to the multi-level feature extraction proposed by a previous random forest-based method (Lei *et al* 2018). First, a representative number of voxels in bone and air regions were chosen from the original CT image. Principle component analysis with the highlighted voxels are shown on CT and MR in figures 5(a) and (b). Green circles represent air voxels while red crosses represent bone voxels. Principle component analyses were then carried out on the multi-level feature vectors and the semantic feature vectors separately for each of the highlighted voxels. Figures 5(c) and (d) show the magnitude of the first 3 principle components for the multi-level features and semantic features, respectively. The strong clustering of air and bone voxels seen in figure 5(d) illustrates the strength of semantic classification. Since a random forest is a set of binary decision trees, each binary decision node separates samples by feature thresholds. Thus, by using semantic information, the binary decision node can clearly separate samples belonging to bone and air. On the other hand, the anatomical features selected by the previous method lack the discriminative power to differentiate air, bone, and soft tissue labels, leading to some of the ambiguities seen in figures 5(c1)–(c3).

Figures 6(b1)–(b3), (c1)–(c3), (d1)–(d3) and (e1)–(e3) show the air, soft tissue, and bone probability maps of the first, second, third, and fourth iterations in the auto-context model. In the first iteration of the auto-context model, the probability maps for bone, soft tissue, and air show relatively poor differentiation of materials, as shown in figures 6(b1)–(b3). As the model iterates, the classifier becomes more accurate, showing clear differentiation between the three materials by the fourth iteration in figures 6(e1)–(e3). In early iterations, this is

caused by the ambiguity mentioned above, namely that similar features were mapped to dissimilar CT patches during training of the first classification random forest. By training the other three more classification random forests with context features (each classification random forest is trained with each iteration), the accuracy of probability maps generated from classification random forests can be improved iteratively. Semantic information based on final probability maps can more easily differentiate air, soft tissue and bone.

Figure 7 shows the original planning CT image of a test patient and sCT images using comparison methods and the proposed method. Figures 7(b3)–(d3) show the difference maps between the sCT and the original CT. The proposed method can better preserve fine structural details; and compared to the other two methods, the magnitude of errors is smaller. The difference image based on the previous random forest-based method in figure 7(b3) and the GAN-based method in figure 7(c3) both show sharp prediction errors, especially at the nasal cavity and auditory meatus. Results from the proposed method, shown in figure 7(d3), shows the error is greatly reduced at these areas. The MAE of original CT and sCT generated by random forest-based method, GAN-based method and the proposed method within highlighted ROIs in (a2) and (a3) are 206.30, 251.39 and 139.59 HU, respectively. The inserts (a2)–(d2) show a zoomed-in view of the region of interest outlined in inserts (a1)–(d1). This region is highlighted because it contains large anatomic variations between air, soft tissue, and bone. The previous random forest-based method, shown in figure 7(b2), produces a sCT with significant noise in this region, resulting from a misclassification of the materials. The GAN-based method, shown in figure 7(c2), however, is smooth within this region while better preserving the bony anatomy than our previous random forest method. This is because the GAN-based method relies on deeper features than the previous random forest method. However, the GAN-based method is not without errors, for example the bone structure is shadowed within the soft tissue region. This is an artifact induced because patch-based deep learning methods tend to neglect spatial information when the patch size is small. Insert (d2) shows that the sCT generated by the proposed method most closely represents the original CT. The highlighted voxels in figures 7(a1)–(d1) are two illustrative areas where the proposed method succeeds and the other methods fail. In first row of figure 7, two small regions (1, 2) within the orbit are chosen to illustrate the performance of comparison methods and our proposed method. In region 1, the previous random forest-based method, figure 7(b1) predicts that the voxel should be soft tissue while it is bone in the original CT. The GAN-based method figure 8(b2) successfully identifies region 1, but in region 2, it predicts soft tissue as air. At both of these 2 regions, the proposed method estimates the sCT intensities correctly.

The MAE, PSNR, NCC for each sCT in comparison to the original CT are shown in table 1. The DSC of each tissue region was calculated between the original CT and each sCT for all 14 test patients. The proposed method outperformed all comparison methods in all metrics. Paired *t*-tests between the proposed method and previous random forest-based method/the GAN-based method for PSNR, NCC, and DSC of 14 test patients showed that the proposed method significantly improved sCT generation ($p < 0.005$). On average, the proposed semantic random forest approach improved bone detection by 12%–13%, as estimated by the DSC. Of note, in the worst case patient, the proposed method still achieved a bone DSC of 0.80. In the previous random forest and GAN-based methods, the bone DSC were 0.61

and 0.63. No significant improvement was observed in MAE with proposed method compared to the random forest-based method (57.45 ± 8.45 HU versus 61.74 ± 16.64 HU), and compared to the GAN-based method (57.45 ± 8.45 HU versus 62.51 ± 9.73 HU).

The differences in DVH metrics of PTVs and OARs among all patients are shown by the box plots in figures 8 and 9, respectively. The central mark of each box indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. The statistics of DVH metric differences are summarized in table 2. STD stands for standard deviation, and the *p*-values we calculated are with null hypothesis of zero dose difference between sCT and original CT. For PTVs, the average differences among all patients are less than 0.2 Gy for all metrics. Although the *p*-values of some metrics are less than 0.05 which indicate statistically significant differences, such differences with absolute values less than 0.2 Gy are negligible when compared to the clinically-prescribed dose of around 20 Gy. For selected OARs, the average differences are less than 0.02 Gy, which demonstrate the high accuracy of dose calculated for OARs based on sCT.

The DVH differences from ground truth were also evaluated on RF and GAN method and compared with SRF method. The dose differences on OARs were very minimal (<0.05 Gy) for the two methods as for the proposed SRF method since the absolute values of dose on OARs are very small. The dose difference between GAN/RF-based synthetic CTs and planning CT on PTVs are shown in table 3. Compared with the PTV part in table 2, the SRF method outperformed RF and GAN methods by 56% and 43% respectively among all DVH metrics. The two-paired t-test of dose errors between SRF and RF/GAN demonstrated that the dosimetric improvement of SRF is statistically significant on D5, Dmean and Dmax.

Patient #12 has the lowest DSC in bone, which may indicate the highest discrepancy in bone prediction. This patient also demonstrates highest DVH error in PTV (shown in figure 10). In this case, the Dmin and Dmean on synthetic CT are 0.486 Gy and 0.213 Gy higher than ground truth in PTV, respectively. Figure 11 shows the corresponding dose distributions at three orthogonal views around PTV of that case. These views show that on sCT, the dose in the skull in the beam path is lower than the ground truth, while the dose in the PTV, which is behind skull bone, is higher than ground truth. This is caused by underestimation of HU number on skull, which effectively decreases the attenuation and dose deposition in bone while increasing the beam fluence and dose deposition behind it.

4. Discussion

The proposed semantic random forest-based method is novel in several aspects. First and foremost, it incorporates semantic information generated from air, bone, and soft tissue probability maps of MRIs into a random forest-based method. Random forest-based methods rely on a feature's ability to classify MRI patches into different tissue regions and accurately predict sCT intensities. Each of these functions is enhanced by semantic information. The proposed method thus improves upon previous random-forest based methods for synthetic CT generation by including semantic information. In addition to

labeling each MRI patch with various features, each patch is also assigned a probability of being bone, soft tissue, or air. Inclusion of this information allows for mapping from MRI to CT that approximates a one-to-one relationship. Second, an auto-context model incorporated the semantic information and original features from the previous probability maps in the classification random forest to provide iterative refinement. This significantly enhanced the semantic information's ability to train the MRI-to-CT mapping.

Currently, computational cost for training a model is a challenge for machine learning-based sCT synthesis methods. In the present study, the training time for our method was about 11.5 h, while the prediction time was around 30 min for each patient. Our previous proposed random forest-based method (Yang *et al* 2017b), which relied on training a series of regression forests without training any classification forests, had an average training time of 7 h. Huynh *et al* (2016) reported that their random-forest based method takes about 4 h to train, while the GAN-based method (Nie *et al* 2018) reported a training time of 19 h. The main steps that increase the computational cost of the proposed method are as follows: first, Both the classification and regression random forests we used are CPU-based methods, although multicore parallel computing was utilized to enhance the efficiency, training and testing were still time consuming as compared with recent graphical processing unit (GPU)-based parallel computing. Second, since the auto-context model needed several rounds of training for segmenting and predicting the image, this further adds to the computational time of the proposed method. Recently, since the development of GPU-based parallel computing, a GPU-based parallelized algorithm has been proposed to accelerate classification random forests (Jayaraj *et al* 2016). In future work, we hope to modify the proposed algorithm to run on a GPU framework to decrease computational time. Although the computational efficiency of the proposed method is affected by these additional steps, the tradeoff is that they improve both the robustness in sCT generation and overall image quality. As shown in table 1, not only are the numerical results of the proposed method better than the comparison methods, but the standard deviation of MAE for all patients is reduced to 8.45 HU, lower than the RF- (16.64 HU) and GAN-based (9.73 HU) methods. In clinical evaluation, since HU accuracy is the main factor that impacts the dosimetric calculation, MAE improvement is especially important for MRI-only based sCT synthesis. The low standard deviation of MAE of different patients means that in some extreme cases our algorithm still achieves comparable performance.

The accuracy of the MRI-only based sCT synthesis is influenced by a number of factors, such as magnetic field inhomogeneities and distortions that appear in MRimages. In our study, all MRI images were pre-processed using an N3 Algorithm (Tustison *et al* 2010) to effectively reduce distortion before training or synthesizing. Other novel methods such as a real-time image distortion correction method (Crijns *et al* 2011) have been reported to have excellent performance, and combining these preprocessing methods with our method could increase the accuracy of the sCTs. We have also found that structural differences caused by rigid registration errors are minimal for the brain, but can pose a problem in other anatomical regions. Another limitation of the present work is that the random forest combines the predictions of a set of binary decision nodes. If a median or averaging decision node were used, the random forest-based would be more robust to biases caused by structural differences. When MR-CT image registration is performed in other anatomic site

such as pelvis and whole body, followed by a sCT synthesis task, the structural difference will be noticeable. A deformable registration method would be required for a whole body sCT generation, and the choice of the registration could impact results.

5. Conclusion

We have investigated a novel learning-based approach to generate patient-specific sCTs from routine MRIs. Experimental validation demonstrated its clinical feasibility and reliability. This CT image generation technique could be a useful tool for MRI-based radiation treatment planning.

Acknowledgments

This research was supported in part by the National Cancer Institute of the National Institutes of Health under Award No. R01CA215718 and the Emory Winship Cancer Institute pilot grant.

References

- Andreasen D, Edmund JM, Zografos V, Menze BH and Van Leemput K 2016a Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features Proc. SPIE 9784 978417
- Andreasen D, Van Leemput K and Edmund JM 2016b A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis Med. Phys 43 4742 [PubMed: 27487892]
- Andreasen D, Van Leemput K, Hansen RH, Andersen JAL and Edmund JM 2015 Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain Med. Phys 42 1596–605 [PubMed: 25832050]
- Aouadi S, Vasic A, Paloor S, Hammoud RW, Torfeh T, Petric P and Al-Hammadi N 2016 Sparse patch-based method applied to mri-only radiotherapy planning Phys/Med. 32 309
- Aouadi S, Vasic A, Paloor S, Torfeh T, McGarry M, Petric P, Riyas M, Hammoud R and Al-Hammadi N 2017 Generation of synthetic CT using multi-scale and dual-contrast patches for brain MRI-only external beam radiotherapy Phys. Med 42 174–84 [PubMed: 29173912]
- Cabello J, Lukas M, Forster S, Pyka T, Nekolla SG and Ziegler SI 2015 MR-based attenuation correction using ultrashort-echo-time pulse sequences in dementia patients J. Nucl. Med 56 423–9 [PubMed: 25678486]
- Chaibi H and Nourine R 2018 New pseudo-CT generation approach from magnetic resonance imaging using a local texture descriptor J. Biomed. Phys. Eng 8 53–64 [PubMed: 29732340]
- Chen S, Qin A, Zhou D and Yan D 2018 Technical note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning Med. Phys 45 5659–65 [PubMed: 30341917]
- Crijns SPM, Raaymakers BW and Lagendijk JJW 2011 Real-time correction of magnetic field inhomogeneity-induced image distortions for MRI-guided conventional and proton radiotherapy Phys. Med. Biol 56 289–97 [PubMed: 21149949]
- Devic S 2012 MRI simulation for radiotherapy treatment planning Med. Phys 39 6701–11 [PubMed: 23127064]
- Edmund JM, Kjer HM, Van Leemput K, Hansen RH, Andersen JAL and Andreasen D 2014 A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times Phys. Med. Biol 59 7501–19 [PubMed: 25393873]
- Edmund JM and Nyholm T 2012 A review of substitute CT generation for MRI-only radiation therapy Radiat. Oncol 12 199–217
- Emami H, Dong M, Nejad-Davarani SP and Glide-Hurst CK 2018 Generating synthetic CTs from magnetic resonance images using generative adversarial networks Med. Phys 45 3627–36

- Gudur MSR, Hara W, Le QT, Wang L, Xing L and Li RJ 2014 A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning *Phys. Med. Biol* 59 6595–606 [PubMed: 25321341]
- Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med. Phys* 44 1408–19 [PubMed: 28192624]
- Hsu SH, Cao Y, Huang K, Feng M and Balter JM 1995 Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy *Phys. Med. Biol* 58 8419–35
- Huang LK and Wang MJJ 1995 Image thresholding by minimizing the measures of fuzziness *Pattern Recognit.* 28 41–51
- Huynh T, Gao YZ, Kang JY, Wang L, Zhang P, Lian J, Shen DG and Alzheimer's Disease Neuroimaging Initiative 2016 Estimating CT image from MRI data using structured random forest and auto-context model *IEEE Trans. Med. Imaging* 35 174–83 [PubMed: 26241970]
- Jayaraj PB, Ajay MK, Nufail M, Gopakumar G and Jaleel UCA 2016 GPURFSCREEN: a GPU based virtual screening tool using random forest classifier *J. Cheminform* 8 12 [PubMed: 26933453]
- Johansson A, Karlsson M and Nyholm T 2011 CT substitute derived from MRI sequences with ultrashort echo time *Med. Phys* 38 2708–14 [PubMed: 21776807]
- Johnstone E, Wyatt JJ, Henry AM, Short SC, Sebag-Montefiore D, Murray L, Kelly CG, McCallum HM and Speight R 2018 Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy *Int. J. Radiat. Oncol. Biol. Phys* 100 199–217 [PubMed: 29254773]
- Khoo VS and Joon DL 2006 New developments in MRI for target volume delineation in radiotherapy *Br. J. Radiol* 79 S2–15 [PubMed: 16980682]
- Kim J, Glide-Hurst C, Doemer A, Wen N, Movsas B and Chetty IJ 2015 Implementation of a novel algorithm for generating synthetic ct images from magnetic resonance imaging data sets for prostate cancer radiation therapy *Int. J. Radiat. Oncol. Biol. Phys* 91 39–47 [PubMed: 25442341]
- Korhonen J, Kapanen M, Keyrilainen J, Seppala T and Tenhunen M 2014 A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer *Med. Phys* 41 011704 [PubMed: 24387496]
- Largent A, Barateau A, Nunes JC, Lafond C, Greer PB, Dowling JA, Saint-Jalmes H, Acosta O and de Crevoisier R 2019 Pseudo-CT generation for MRI-only radiotherapy treatment planning: comparison between patch-based, atlas-based, and bulk density methods *Int. J. Radiat. Oncol. Biol. Phys* 103 479–90 [PubMed: 30336265]
- Lei Y. et al. 2018; Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning. *J. Med. Imaging.* 5:034001.
- Maspero M, Savenije MHF, Dinkla AM, Seevinck PR, Intven MPW, Jurgenliemk-Schulz IM, Kerkmeijer LGW and van den Berg CAT 2018 Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy *Phys. Med. Biol* 63 185001 [PubMed: 30109989]
- Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q and Shen D 2018 Medical image synthesis with deep convolutional adversarial networks *IEEE Trans. Biomed. Eng* 65 2720–30 [PubMed: 29993445]
- Schmidt MA and Payne GS 2015 Radiotherapy planning using MRI *Phys. Med. Biol* 60 R323–61 [PubMed: 26509844]
- Sjölund J, Forsberg D, Andersson M and Knutsson H 2015 Generating patient specific pseudo-CT of the head from MR using atlas-based regression *Phys. Med. Biol* 60 825–39 [PubMed: 25565133]
- Torrado-Carvajal A, Herraiz JL, Alcain E, Montemayor AS, Garcia-Canamaque L, Hernandez-Tamames JA, Rozenholc Y and Malpica N 2016 Fast patch-based Pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in brain studies *J. Nucl. Med* 57 136–43 [PubMed: 26493204]
- Tu Z and Bai X 2010 Auto-context and its application to high-level vision tasks and 3D brain image segmentation *IEEE Trans. Pattern Anal. Mach. Intell* 32 1744–57 [PubMed: 20724753]
- Tustison NJ, Avants BB, Cook PA, Zheng YJ, Egan A, Yushkevich PA and Gee JC 2010 N4ITK: improved N3 bias correction *IEEE Trans. Med. Imaging* 29 1310–20 [PubMed: 20378467]

- Uh J, Merchant TE, Li Y, Li X and Hua C 2014 MRI-based treatment planning with pseudo CT generated through atlas registration *Med. Phys* 41 051711 [PubMed: 24784377]
- Ulin K, Urie MM and Cherlow JM 2010 Results of a multi-institutional benchmark test for cranial CT/MR image registration *Int. J. Radiat. Oncol. Biol. Phys* 77 1584–9 [PubMed: 20381270]
- van der Heide UA, Houweling AC, Groenendaal G, Beets-Tan RGH and Lambin P 2012 Functional MRI for radiotherapy dose painting *Magn. Reson. Imaging* 30 1216–23 [PubMed: 22770686]
- Wang T, Manohar N, Lei Y, Dhabaan A, Shu H-K, Liu T, Curran WJ and Yang X 2018 MRI-based treatment planning for brain stereotactic radiosurgery: dosimetric validation of a learning-based pseudo-CT generation method *Med. Dosim* (accepted) (10.1016/j.meddos.2018.06.008)
- Yang X, Lei Y, Shu HKG, Rossi PJ, Mao H, Shim H, Curran WJ Jr and Liu T 2017a A learning-based approach to derive electron density from anatomical MRI for radiation therapy treatment planning *Int. J. Radiat. Oncol. Biol. Phys* 99 S173–4
- Yang XF and Fei BW 2013 Multiscale segmentation of the skull in MR images for MRI-based attenuation correction of combined MR/PET *J. Am. Med. Inform. Assoc* 20 1037–45 [PubMed: 23761683]
- Yang XF, Lei Y, Shu HK, Rossi P, Mao H, Shim H, Curran WJ and Liu T 2017b Pseudo CT estimation from MRI using patch-based random forest *Proc. SPIE* 10133 101332Q

Figure 1.

Schematic flow chart of the proposed algorithm for MRI-based sCT generation. The left part of this figure shows the training stage of our proposed method, which consists of classification random forest training, semantic feature extraction, context feature extraction, and regression random forest training. The right part of this figure shows the synthesizing stage, where a new MR image follows a similar sequence to the training stage to generate a sCT image.

Figure 2.

An example of a MRI (a1) and (a2) and its corresponding CT image (c1) and (c2) of the same patient with highlighted patches (patch size = 15×15 , highlighted by red rectangles). (b1), (b2) and (d1), (d2) show the highlighted patches of (a1), (a2) and (c1), (c2), respectively, in greater detail.

Figure 3.

An example illustrating the generation of semantic information for incorporation into the random forest model. (b1) Is the MRI image, (a1), (c1) and (d1) show the binary segmentation for each material (air, soft tissue, and bone) based on a simple thresholding method in the CT image. (a2)–(a5), (c2)–(c5) and (d2)–(d5) Show the calculated material probability with increasing iterations. Material probabilities are calculated with a maximum likelihood method conditioned on Hounsfield unit (HU) values and image features. (b2)–(b5) Show the zoomed in regions of the nasopharynx. The context locations are listed in dotted rectangles of these regions. The semantic information is a concatenation of mean values of these rectangles. Originally, much of the air in this region is calculated to be a

mixture of multiple materials. As the probability maps are iteratively refined, the classifier correctly determines that many of the pixels in this region are composed of air.

Figure 4.

(a1)–(a3) MRI images shown in axial, sagittal and coronal planes. (b1)–(b3) Are the corresponding CT images. (c1)–(c3) Are sCT obtained by previous random forest-based method, and (d1)–(d3) are difference image between the original CT and the sCT generated by the previous method. (e1)–(e3) Are sCT obtained by the proposed semantic random forest-based method, and (f1)–(f3) are difference images between the original CT and the sCT generated by the proposed method.

Figure 5.

An example illustrating the benefit of semantic information in the random forest model. (a) and (b) Show axial CT and MRI images, where the voxels belonging to air are highlighted by green circles, and voxels belonging to bone are highlighted by red asterisks. (c) Shows the scatter plots of first 3 principle components of original extracted features generated from MRI patches which centered on corresponding samples. (d) Shows the scatter plots of first 3 principle components of semantic features generated from probability maps. These scatter plots show that the inclusion of semantic information allows the random forest to more clearly differentiate air from bone voxels.

Figure 6. Multiple iterative results based on auto-context method. (a1)–(a3) Are the original MRI, CT image and corresponding CT label (air, soft tissue and bone). (b1)–(b3), (c1)–(c3), (d1)–(d3) and (e1)–(e3) Show the probability maps of air, soft-tissue and bone with different iterations, respectively. The probability maps of air, soft-tissue and bone improved with each iteration.

Figure 7. Axial view of original CT scan (a1), corresponding MRI scan (a3) and synthetic sCT by random forest-based method (b1), GAN-based method (c1) and the proposed semantic random forest-based method (d1). (a2) and (d2) Show the highlighted region in greater detail (a1) and (d1), respectively. (b3) and (d3) Are the difference images between (b1)–(d1) and (a1).

Figure 8. DVH metrics comparison for the PTVs for all 14 patients. The red plus signs are statistical outliers. Note that the prescription dose for these treatments is around 20 Gy.

Figure 9. DVH metrics comparison for the OARs for all 14 patients. The red plus signs are statistical outliers. Note that the prescription dose for these treatments is around 20 Gy.

Figure 10.

The DVH curves for the PTV, chiasm, brain stem, left optic nerve (OpticNerve_L), and right optic nerve (OpticNerve_R) of patient #12.

Figure 11.
Dose distributions on original CT (upper), sCT (middle), and dose difference (bottom) for patient 12 at three orthogonal views about isocenter.

Table 1.

The numerical results of comparing methods and the proposed semantic random forest-based method.

Mean $\pm$ Std	MAE (HU)	PSNR (dB)	NCC	Air	DSC	
					Soft tissue	Bone
RF	61.74 $\pm$ 16.64	24.76 $\pm$ 1.77	0.95 $\pm$ 0.01	0.96 $\pm$ 0.01	0.88 $\pm$ 0.02	0.72 $\pm$ 0.04
GAN	62.51 $\pm$ 9.73	23.02 $\pm$ 1.47	0.94 $\pm$ 0.01	0.95 $\pm$ 0.01	0.89 $\pm$ 0.03	0.71 $\pm$ 0.05
Our SRF	57.45 $\pm$ 8.45	28.33 $\pm$ 1.68	0.97 $\pm$ 0.01	0.98 $\pm$ 0.01	0.93 $\pm$ 0.02	0.84 $\pm$ 0.06
<i>p</i> -value (SRF versus RF)	0.22	<0.001	0.002	<0.001	<0.001	<0.001
<i>p</i> -value (SRF versus GAN)	0.15	<0.001	<0.001	<0.001	<0.001	<0.001

Table 2.

Differences in DVH metrics of PTVs and OARs among all 14 patients.

DVH metrics difference (Gy)	Dmin	D5	D50	D95	D99	Dmean	Dmax
PTV							
Mean	0.059	0.072	0.071	0.061	0.059	0.070	0.123
Std	0.128	0.112	0.107	0.100	0.101	0.107	0.160
Pvalue	0.061	0.012	0.010	0.017	0.020	0.011	0.003
Brain stem							
Mean	−0.004	0.009	0.003	−0.002		0.003	0.015
Std	0.011	0.019	0.009	0.005		0.007	0.027
Pvalue	0.118	0.053	0.178	0.110		0.067	0.028
Chiasm							
Mean	−0.006	0.000	−0.004	−0.004		−0.002	0.020
Std	0.014	0.019	0.015	0.016		0.013	0.049
Pvalue	0.076	0.952	0.279	0.251		0.468	0.077
Left optic nerve							
Mean	0.004	0.006	0.006	0.004		0.006	0.003
Std	0.022	0.014	0.015	0.016		0.013	0.024
Pvalue	0.441	0.106	0.130	0.305		0.072	0.590
Right optic nerve							
Mean	0.009	0.006	0.007	0.004		0.006	0.016
Std	0.016	0.011	0.011	0.013		0.009	0.033
Pvalue	0.026	0.031	0.011	0.206		0.010	0.052

Table 3.

Differences in DVH metrics of PTVs among all 14 patients for RF and GAN methods.

PTV DVH metrics difference (Gy)		Dmin	D5	D50	D95	D99	Dmean	Dmax
RF	Mean	0.103	0.158	0.138	0.135	0.131	0.143	0.230
	Std	0.159	0.181	0.188	0.189	0.173	0.185	0.258
	<i>P</i> value (RF versus SRF)	0.147	0.007	0.063	0.066	0.051	0.036	0.009
GAN	Mean	0.135	0.266	0.241	0.187	0.146	0.239	0.300
	Std	0.482	0.414	0.393	0.428	0.493	0.397	0.440
	<i>P</i> value (GAN versus SRF)	0.485	0.017	0.029	0.142	0.389	0.031	0.031