Medical Image Analysis 36 (2017) 98–112

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Comparison of atlas-based techniques for whole-body bone segmentation

Hossein Arabi[a], Habib Zaidi[a,b,c,d,*]

[a] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland
[b] Geneva Neuroscience Center, Geneva University, CH-1205 Geneva, Switzerland
[c] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, Netherlands
[d] Department of Nuclear Medicine, University of Southern Denmark, DK-500, Odense, Denmark

ARTICLE INFO

Article history:
Received 14 June 2016
Revised 7 November 2016
Accepted 10 November 2016
Available online 12 November 2016

Keywords:
MRI
PET/MRI
Bone segmentation
Atlas-based segmentation
Whole-body

ABSTRACT

We evaluate the accuracy of whole-body bone extraction from whole-body MR images using a number of atlas-based segmentation methods. The motivation behind this work is to find the most promising approach for the purpose of MRI-guided derivation of PET attenuation maps in whole-body PET/MRI. To this end, a variety of atlas-based segmentation strategies commonly used in medical image segmentation and pseudo-CT generation were implemented and evaluated in terms of whole-body bone segmentation accuracy. Bone segmentation was performed on 23 whole-body CT/MR image pairs via leave-one-out cross validation procedure. The evaluated segmentation techniques include: (i) intensity averaging (IA), (ii) majority voting (MV), (iii) global and (iv) local (voxel-wise) weighting atlas fusion frameworks implemented utilizing normalized mutual information (NMI), normalized cross-correlation (NCC) and mean square distance (MSD) as image similarity measures for calculating the weighting factors, along with other atlas-dependent algorithms, such as (v) shape-based averaging (SBA) and (vi) Hofmann's pseudo-CT generation method. The performance evaluation of the different segmentation techniques was carried out in terms of estimating bone extraction accuracy from whole-body MRI using standard metrics, such as Dice similarity (DSC) and relative volume difference (RVD) considering bony structures obtained from intensity thresholding of the reference CT images as the ground truth. Considering the Dice criterion, global weighting atlas fusion methods provided moderate improvement of whole-body bone segmentation (DSC = 0.65 ± 0.05) compared to non-weighted IA (DSC = 0.60 ± 0.02). The local weighed atlas fusion approach using the MSD similarity measure outperformed the other strategies by achieving a DSC of 0.81 ± 0.03 while using the NCC and NMI measures resulted in a DSC of 0.78 ± 0.05 and 0.75 ± 0.04, respectively. Despite very long computation time, the extracted bone obtained from both SBA (DSC = 0.56 ± 0.05) and Hofmann's methods (DSC = 0.60 ± 0.02) exhibited no improvement compared to non-weighted IA. Finding the optimum parameters for implementation of the atlas fusion approach, such as weighting factors and image similarity patch size, have great impact on the performance of atlas-based segmentation approaches. The voxel-wise atlas fusion approach exhibited excellent performance in terms of cancelling out the non-systematic registration errors leading to accurate and reliable segmentation results. Denoising and normalization of MR images together with optimization of the involved parameters play a key role in improving bone extraction accuracy.

1. Introduction

The emergence of hybrid imaging techniques, such as PET/CT and PET/MRI in clinical practice engendered a number of new clinical and research opportunities and improved the quantitative accuracy and diagnostic confidence of PET findings (Judenhofer et al., 2008). A number of active research groups are focusing their efforts on addressing the challenges of combined PET/MRI, encompassing instrumentation developments, optimization of workflow and data acquisition protocols and the improvement of the quantitative performance of both imaging modalities (Zaidi and Del Guerra, 2011). Beside the precious anatomical information provided by CT or MRI, additional information that can be extracted from these images, such as attenuation properties of body tissues and motion information can be exploited for correction of emission data and quantitative PET image reconstruction. However, MRI-guided attenuation correction in whole-body PET/MRI proved to be

[*] Corresponding author. Fax: +41 22 372 7169.
E-mail address: habib.zaidi@hcuge.ch (H. Zaidi).

http://dx.doi.org/10.1016/j.media.2016.11.003
1361-8415/

a challenging issue and has therefore remained an active and open research question during the last decade (Mehranian et al., 2016). Commercially available PET/MR scanners employ tissue classification methods, which rely on segmentation of MR images into tissue classes and assigning uniform linear attenuation coefficients to each tissue class (Martinez-Moller et al., 2009; Arabi et al., 2015). The major drawback of such methods, particularly in the context of whole-body imaging, lies in ignoring bones as a separate tissue class. Since bone tissue generates a void signal when using common MR sequences, it is indistinguishable from air. As such, bony structures are commonly replaced by soft-tissue in current methods, thus leading to significant underestimation of tracer uptake in the vicinity of bony structures (Bezrukov et al., 2013; Hofmann et al., 2011).

A number of techniques have been proposed to consider bone tissue during attenuation correction (AC) in whole-body PET/MRI. Basically two categories have emerged: atlas-guided attenuation map generation approaches (Hofmann et al., 2011; Bezrukov et al., 2013; Arabi and Zaidi, 2016a; Marshall et al., 2013; Arabi and Zaidi, 2016b) and emission-based approaches (Rezaei et al., 2012; Mehranian and Zaidi, 2015). Atlas-guided methods primarily rely on prior information provided by registration of an atlas into target image coordinates to allow classification of bone tissues. Direct segmentation of bones from MR images, particularly in whole-body imaging, is a difficult task owing to anatomical complexity, low quality and high noise level of dedicated MR sequences used for the purpose of AC (Hofmann et al., 2008). Atlas-guided segmentation has been successfully applied in various image segmentation tasks using a wide variety of imaging modalities (Lorenzo-Valdés et al., 2004). In principle, each individual atlas image transformed to the coordinates of the target image is regarded as potential candidate. It has, however, been proven that using the information from multiple atlas images leads to more accurate results (Svarer et al., 2005). The information obtained from several atlas images can be pooled into an average atlas or into a so called probabilistic atlas (Rohlfing et al., 2001; Svarer et al., 2005). However, there is a trend to take full advantage of multiple atlas images at hand by exploiting pattern recognition techniques to identify morphologically similar cases in the atlas dataset during the multi-atlas fusion process. This dramatically reduces non-systematic registration errors and improves the accuracy of the segmentation (Artaechevarria et al., 2009).

Various strategies were proposed to incorporate bone tissue in PET/MRI attenuation maps in whole-body imaging (Hofmann et al., 2011; Bezrukov et al., 2015; Ay et al., 2014; Arabi and Zaidi, 2016a; Bezrukov et al., 2013; Marshall et al., 2013; Paulus et al., 2015). In whole-body imaging, almost all proposed methods, except joint attenuation-activity reconstruction techniques, rely on prior knowledge present in atlas images to predict bone from MRI. Moreover, owing to long acquisition time, application of ultra-short echo time (UTE) (Keereman et al., 2010) or zero time echo (ZTE) (Delso et al., 2015) sequences are still limited to brain imaging (single bed position). Atlas-guided segmentation has been successfully applied in various image segmentation tasks using different imaging modalities, particularly for cases with very low contrast to the surrounding tissues (Lorenzo-Valdés et al., 2004). Atlas-based methods are of special interest since they have so far exhibited superior performance in terms of bone identification (Burgos et al., 2014) particularly in whole-body imaging (Hofmann et al., 2011). Burgos et al. (2014) demonstrated superior performance of atlas-based methods in CT synthesis and PET quantitative accuracy compared to a segmentation method using an UTE MRI sequence in brain imaging. Likewise, Mehranian et al. (2016) demonstrated that atlas-based methods provide the most accurate attenuation maps compared to simultaneous activity-attenuation estimation and state-of-the-art 3-class segmentation method. In whole-body imaging, Hofmann et al. (2011) proposed an atlas-based method combined with a pattern recognition technique, which resulted in less than 10% uptake error on average, thus outperforming standard segmentation methods in whole-body imaging. Marshall et al. (2013) evaluated a method enabling to incorporate bony structures into attenuation maps based on a fast atlas-based approach. By including bone, the magnitude of the relative error was reduced to a range acceptable in clinical setting.

Various atlas-based methods were independently developed and evaluated using different MRI sequences, different atlas datasets in terms of sample size, patient variability, field of view and body region, different MRI quality (noise level or acquisition time) and evaluation procedures and metrics. Although there is substantial literature reporting promising results achieved by atlas-based methods, the performance of these techniques still requires further investigation based on a common ground. Therefore, a comparison of various atlas-based strategies provides a valuable insight into their application to attenuation correction in PET/MRI.

Since the delineation of bones is the most challenging task in whole-body MRI-guided attenuation map generation, we focused our comparison of the various pseudo-CT generation approaches and atlas-based segmentation methods on the accuracy of extracted whole-body bone. To this end, we selected and implemented a number of conventional atlas-based segmentation methods, such as majority voting, intensity averaging, global and local weighting atlas fusion strategies together with Hofmann's algorithm (proposed for whole-body PET/MR attenuation map generation) and shape-based averaging (SBA) technique. In addition to the comparison of the different segmentation techniques, our goal is to select the most promising algorithm for attenuation correction in whole-body PET/MRI. The very preliminary results of this work have been previously published (Arabi and Zaidi, 2014). The present article presents a substantial extension of the previous work through the implementation and comparison of a higher number of algorithms using a larger database of clinical studies and reporting more detailed quantitative analysis of the data.

2. Materials and methods

2.1. Atlas-based segmentation

The objective of atlas based segmentation is to provide labeling of unknown tissue classes on the target image. Consider the segmentation of an image with potentially L different classes belonging to a label set $Label = \{1, 2, \ldots, L\}$. In the case of bone segmentation, the number of classes is confined to $L = 2$ where label 1 stands for background and label 2 represents bony structures. Here, a set of 3-D MR images Amr_n along with their corresponding aligned CT images Act_n are considered as atlas images. An atlas-based classifier is defined by a set of atlas images Amr_n $n = 1, \ldots, N$ and transformation matrices (M_n) which map coordinates from the target image T to the atlas images n: M_n: $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Since bone segmentation can be simply carried out by intensity thresholding of CT images, Act_n images act as candidates for tissue labeling of the target MR image of T. Applying a given transformation matrix M_n to an atlas image Act_n yields an estimated segmentation of the target subject T_{An} where a set of segmentation candidates T_{An} $n = 1, \ldots, N$ must be combined to form the final estimated bone segmentation T_s. Atlas-based segmentation can be regarded as the classification of X unordered samples where the candidate n assigns x to class $\boldsymbol{l} \in \boldsymbol{Label}$. The output of N independent classifiers can be combined to generate a single response of the combination strategy, $E(x)$. The aim of building an ensemble classifier is to achieve a higher probability of correctly classifying the voxels of the image than that obtained by using an individual classifier maximizing the probability given all classifier decisions T_{An} and a classifier performance model C (Eq. 1) (Rohlfing et al., 2004b).

$$\boldsymbol{E}(\boldsymbol{x}) = \boldsymbol{a}rg\max_{\boldsymbol{l}} \boldsymbol{P}(\boldsymbol{x} = \boldsymbol{l}\ |\boldsymbol{T}_{A1},\ \ldots,\ \boldsymbol{T}_{AN},\ C) \tag{1}$$

2.2. PET/CT and PET/MR data acquisition

The study population comprised $N = 23$ consecutive patients, 15 men and 8 women (mean age ± SD = 60 ± 8 y), refereed to our department for MRI of the head and neck, whole-body ^{18}F-FDG PET/MRI and whole-body ^{18}F-FDG PET/CT for staging of head and neck malignancies. The study protocol was approved by the institutional ethics committee and all patients gave their informed consent to participate in the study. ^{18}F-FDG PET/CT scans were performed on a Biograph 64 True Point scanner (Siemens Healthcare, Erlangen, Germany). The CT subsystem consists of a 40-row ceramic detector with 1344 channels per row using adaptive collimation and the z-sharp technique to acquire 64 slices per rotation. After a localization scout scan, an unenhanced CT scan (120 kVp, 180 mAs, 24 × 1.5 collimation, a pitch of 1.2, and 1 s per rotation) was performed for attenuation correction and localization. The typical acquisition time for whole-body CT was less than 10 s. PET/MRI examinations were performed on the Ingenuity TF PET/MR, a sequential system consisting of a whole-body time-of flight (TOF) GEMINI TF PET and a 3T Achieva TX MRI separated by a distance of 3 m sharing a common rotating table platform (Zaidi et al., 2011). The gradient system value and the slew rate are 40 mT/m and 100 mT/m/s, respectively. The coils used for MR imaging include a SENSE neurovascular 16-channel coil for head and neck and a quadrature body coil for total body scanning. Whole-body Dixon examinations were performed on the 3T Achieva TX MRI of the Ingenuity TF PET/MR scanner. The whole body Dixon 3D volumetric interpolated T1-weighted sequence (Dixon, 1984) was acquired using the following parameters: flip angle 10°, TE$_1$ 1.1 ms, TE$_2$ 2.0 ms, TR 3.2 ms, 450×354 mm^2 transverse FOV, $0.85 \times 0.85 \times 3$ mm^3 voxel size, and a total acquisition time of 2 min 17 s. Both MRI and CT acquisitions were performed in free shallow breathing. This sequence produced in-phase and opposed-phase images that are then added together to obtain water only images, and subtracted to get fat-only images. In-phase images were used for the assessment of whole-body bone segmentation.

Due to temporal separation between MRI and CT acquisitions, in-phase MRI were deformably registered to the corresponding CT images using the Elastix framework based on the ITK library (Klein et al., 2010) using a combination of rigid registration based on maximum mutual information and non-rigid registration as described previously (Akbarzadeh et al., 2013). MRI and CT acquisitions were performed with the same patient positioning to minimize non-rigid deformation. However, in case of alignment errors owing for instance to breathing motion, the registration parameters were adjusted to achieve acceptable results. In case of gross registration errors, the studies were excluded.

2.3. Data preprocessing

Clinical whole-body MR images contain a relatively high level of noise and are commonly corrupted by low frequency bias field and inter-patient intensity inhomogeneity (Lötjönen et al., 2010). As will be described in the following section, bone segmentation procedures entail direct handling of MR image intensity. As such, the presence of aforementioned sources of intensity variation in MR images might skew bone segmentation accuracy. To overcome these prospective sources of error, in-phase MR images of all patients underwent some pre-processing procedures in the following order:

- Gradient anisotropic diffusion filtering to suppress noise using the following parameters: conductance = 4, iterations = 10 and time step = 0.01. This algorithm smoothes regions of an image where the gradient magnitude is relatively small (homogenous regions) but diffuses little over areas of the image where the gradient magnitude is large (i.e., edges). Therefore, the central regions of objects are smoothed but their edges are blurred to a lower extent.
- N4 bias field correction (Tustison et al., 2010) to remove magnetic field inhomogeneity effect: Bspline grid resolution = 400, number of iteration = 200 (at each grid resolution), convergence threshold = 0.001, Bspline order = 3, Spline distance = 400, number of histograms = 256 and shrink factor = 3.
- Histogram matching (McAuliffe et al., 2001): Histogram level = 1024 and match points = 128. In order to get the best result from histogram matching, it is recommended to exclude background air voxels of both reference and target images before processing.

The bone segmentation procedure requires the binary mask of segmented background air to save processing time. To this end, the external body contour was determined by applying a 3D active snake contour algorithm on in-phase MR images (Kass et al., 1988). The segmentation process begins by manual selection of the initial seeds in the background using the ITK-SNAP image processing software (Yushkevich et al., 2006).

2.4. Label fusion strategies

This study contains 23 pairs of co-registered in-phase MRI Dixon and CT images. All MR images were processed according to the procedure described in Section 2.3. Using the leave-one-out cross-validation (LOOCV) method, for each subject, images of the remaining *N-1* (i.e. 22) patients are non-rigidly warped to the coordinates of the target image. Image registration was carried out using the Elastix package (based on the ITK library) (Klein et al., 2010) through a combination of affine and non-rigid alignment based on the advanced Mattes mutual information as described in previous work (Akbarzadeh et al., 2013). The following parameters were adopted: interpolate: Bspline, optimizer: standard gradient descent, image pyramid schedule: (16 8 4 2 2), grid spacing schedule (32.0 16.0 8.0 4.0 2.0), maximum number of iterations (4096 4096 2048 1024 512), number of histogram bins: 32. The obtained transformation matrices from the registration between atlas and target MR images were applied to the corresponding atlas CT images. For each target image, 22 candidate CT images are available from which bone can be segmented by intensity thresholding using a threshold of 180 HU. This work focuses on how well the label fusion strategies can pool the information from 22 segmentation candidates to maximize the final bone extraction accuracy. In the following sections, we describe in detail label fusion strategies commonly used in atlas-based segmentation.

2.4.1. General averaging

A commonly used approach for pseudo-CT generation and segmentation of anatomical structures is to simply calculate the arithmetic average of the aligned atlas images (Rohlfing et al., 2001; Rohlfing et al., 2004a). In our case, general arithmetic averaging is performed by computing the intensity average of $N = 22$ aligned atlas CT images (Eq. 2). There is no selective or weighting strategy in this approach and all atlas images (regardless of their morphological similarity to the target subject) contribute equally to bone extraction process.

$$T_{av} = \frac{1}{N}\sum_{n=1}^{N} T_{An} \qquad Bn(x) = \begin{cases} 1, & \text{if} \quad T_{av}\ (x) > 180 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Here T_{An} is the nth aligned atlas CT image to the target image T. As mentioned earlier, bone segmentation (B_n) can be performed by applying intensity thresholding to the average image, T_{av}. Hereafter, we call this approach intensity averaging (IA), meaning bone segmentation is performed after the averaging process.

The same task can be achieved by the well-known majority voting framework where instead of taking the average intensity of aligned atlas CT images, each CT image is converted to a binary bone mask (T_{Sn}) followed by the averaging process. The voxel the majority of classifiers agree on is labeled as bone (Eq. 3) (Heckemann et al., 2006; Artaechevarria et al., 2009; Yushkevich et al., 2010; Artaechevarria et al., 2008).

$$T_{Sav} = \frac{1}{N}\sum_{n=1}^{N} T_{Sn} \qquad B(x) = \begin{cases} 1, & \text{if} \quad T_{Sav}\ (x) > 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

T_{Sav} is also called bone probability map where values of 1 and 0 indicate that all the atlases unanimously predict bony and non-bony tissues for that voxel, respectively.

It is hypothesized that the number of atlas images N has a major impact on the accuracy of extracted bone (Heckemann et al., 2006). To evaluate this feature, bone segmentation was carried out for various numbers of atlases selected randomly among the 22 patient datasets.

Conventional multi-atlas segmentation approaches entails N online registrations between target and atlas images. A number of studies utilized only one single atlas image or template (obtained from taking the average of population) is utilized to delineate the anatomical structures in the target image after warping the atlas image to the target coordinates to reduce the computation time (Rohlfing et al., 2004a; Heckemann et al., 2006). Consequently, this approach requires only one online registration, which makes it computationally efficient. The performance evaluation of the single atlas approach is of special interest since it introduces a trade-off between computational time and the quality of the outcome compared to conventional multi-atlas approach. The single atlas approach, referred as “single atlas image” in Table 4, was compared to various multi-atlas approaches. To evaluate the accuracy of this approach, an iterative atlas generation framework was utilized via the LOOCV scheme (Rohlfing et al., 2001). In summary, an MR image belonging to the patient with the median body mass index of the population was selected as the initial atlas for atlas space alignment. The initial iteration contains the registration of other MR images to the selected atlas using the sequential affine and non-rigid registration procedure described in Section 2.4. At the end of each iteration, the new average atlas is generated and used in the subsequent iteration. Since the template obtained from the previous iteration serves better as common/reference spatial coordinate, after each iteration, the obtained template would be more representative for the target subject. In the present work, we used five iterations and the final transformation field was applied on the corresponding CTs to yield the average CT atlas. In the last step, the average MRI atlas is non-rigidly aligned to the target MRI and bone segmentation is carried out on the warped average CT image. As mentioned earlier, this approach requires only one online registration and the atlas creation is performed offline.

2.4.2. Global weighting

The methods described in Section 2.4.1 do not involve any strategy to detect and consequently discard miss-registration errors. Registration errors occur due to local minima, inter-patient anatomy variability and presence of noise, which might incur gross mismatch on the resulting images (Svarer et al., 2005). One strategy to overcome the misalignment error consists in assigning weights to the atlas images globally (as opposed to local or voxel-wise approach) on the basis of morphological similarity between target and atlas images. By this approach, aligned atlas images presenting the higher degree of anatomy and pose similarities contribute more effectively to the resulting segmentation (Artaechevarria et al., 2009; Chandra et al., 2012; Ying et al., 2013; Artaechevarria et al., 2008). The first step toward weighted atlas-based segmentation consists in developing a similarity criterion between the target image and aligned atlas images. Normalized mutual information (NMI), normalized cross correlation (NCC) and mean square distance (MSD) are the most common similarity measures used for implementation of weighted atlas-based segmentation (Yushkevich et al., 2010; Artaechevarria et al., 2008). These similarity measures are briefly described below. Normalized mutual information is defined as:

$$NMI = \frac{H(T) + H(M(Amr_n))}{H(T, M(Amr_n))} \tag{4}$$

where $H(T)$ is the entropy of image T and $H(T,M(Amr_n))$ indicates the joint entropy of both images. The entropy of an image can be computed from its histogram $h(x)$ as:

$$H(T) = -\sum_{i=1}^{F} h(c_i) log_2 h(c_i)$$

where F is the number of histogram bins and c_i corresponds to the centroid of the ith histogram bin (Wells et al., 1996).

The normalized cross-correlation between the two images is defined as:

$$NCC = \frac{Cov(T, M(Amr_n))}{\sqrt{Var(T)}.\sqrt{Var(M(Amr_n))}} \tag{5}$$

where Cov(T, M(Amr_n)) is the covariance of the images and $Var(T)$ indicates the variance of the image T.

The mean square distance is simply the intensity difference between two images. Here, we used the following formulation to measure the intensity similarity between the target MR image (T) and the co-registered atlas MR images ($M(Amr_n)$). X denotes the total number of image voxels.

$$MSD = \frac{X}{\sum_{x=0}^{X} |T(x) - M(Amr_n(x))|^2} \tag{6}$$

Previously published works in the realm of multi-atlas based segmentation employed various ways of incorporating weighting factors in either majority voting (MV) or intensity averaging (IA) label fusion schemes. In this work, we examined three most commonly used schemes for whole-body bone segmentation through global weighting. Each of these schemes can be performed using either of three above introduced similarity criteria. Ying et al. (2013) exploited NMI similarity measure to identify similar atlas images via the following equation for the purpose of bone elements segmentation of hip and femur from MR images.

$$w_n = \frac{SM(\ T, M(Amr_n)) - min_m[SM(\ T, M(Amr_m))]}{max_m[SM(\ T, M(Amr_m))] - \ min_m[SM(\ T, M(Amr_m))]}$$
$$\text{subject to} \quad w_n \geq \Phi \tag{7}$$

Here SM can be any similarity measure criterion (NMI, NCC and MSD) between the target MR image and transformed atlas images Amr_n. The min and max of the SM are calculated among all atlases to normalize the weighing factor w. After obtaining the weighting factor w, the next step is to select the atlases which satisfy the condition $w \geq \Phi$ ($0 \leq \Phi \leq 1$), where Φ is the threshold used to discard poorly performing atlases. Therefore, the weighted average of atlases T_{av} can be calculated using the following formulation:

$$T_{av} = \frac{1}{Nr}\sum_{n=1}^{N} w_n.T_{An} \qquad Bn(x) = \begin{cases} 1, & \text{if} \quad T_{av}\ (x) > 180 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Nr is the normalization factor obtained by $Nr = \sum_{n=1}^{N} w_n$. The majority voting scheme (Artaechevarria et al., 2009) can be adapted for this purpose as:

$$T_{Sav} = \frac{1}{Nr}\sum_{n=1}^{N} w_n . T_{Sn} \qquad Bn(x) = \begin{cases} 1, & \text{if} \quad T_{Sav}\ (x) > 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Ying et al. (2013) utilized only NMI similarity measure along with a fixed threshold $\Phi = 0.9$ while in our work all three similarity criteria and a variable threshold were examined for the both MV and IA schemes in order to determine the optimal threshold value and the most efficient similarity measure.

The second approach to incorporate the similarity weights in the atlas fusion process is through gain exponent. In this case, the weighting factor is defined as $w = SM(T,M(Amr_n))^P$, where the gain exponent P might be increased if the similarity measure is not sensitive enough to provide appropriate differences between weights (Artaechevarria et al., 2009). The weighting factor w can be incorporated either in Eqs. (8) or (9). Presently, our aim is to find the optimum value of the gain exponent P for the three similarity criteria via IA and MV schemes.

The third atlas weighting scheme is based on Yushkevich et al. (2010) work which assumed that the range of similarity measures can vary quite dramatically between subjects and locations. The same scheme was used by Burgos et al. (2013) for pseudo-CT generation in the head region. As such, a ranking scheme is proposed whereby the similarity measure value for each transformed atlas is ranked across all atlases. Let's suppose that ranked atlases are denoted as R_n. The conversion to the weight is performed by applying an exponential decay function.

$$w_n = e^{-aR_n}\ 10 \tag{10}$$

where R_n denotes the ranked atlas indices (e.g. 1, 2, 3, …) and a is a weighting parameter to be optimized. By adopting the ranking scheme, the training subject that best matches the target subject is given a weight of 1. The training subject with the second best match is assigned a weight e^{-a} and so on. Thus, the segmentation can be performed by applying the weighting factor w_n to Eqs. (8) and (9). Here, the ranking process was repeated three times using the NMI, NCC and MSD similarity measures and for each one the optimum parameter a, which maximized the accuracy of segmented bone was determined.

In some studies, the most similar subject is selected for either MRI segmentation or attenuation map generation in PET/MRI to reduce the computation time (Rohlfing et al., 2004a; Marshall et al., 2013). The most similar atlas can be determined before the registration process on the basis of meta-data and image processing features (Marshall et al., 2013). In our work, the most similar subject to the target image was determined after the registration process using the three aforementioned similarity measures and the extracted bone was validated for each one.

2.4.3. STAPLE

A well-established approach aiming at maximizing multi-atlas based segmentation accuracy is Simultaneous Truth and Performance Level Estimation (STAPLE) (Warfield et al., 2004). A number of studies using multi-atlas based segmentation employed STAPLE algorithm to find the optimal combination of segmentations suggested by the different classifiers (Artaechevarria et al., 2009; Artaechevarria et al., 2008). STAPLE is an expectation-maximization algorithm for simultaneous truth and performance level estimation that considers a collection of segmentations and computes a probabilistic estimate of the true segmentation and a measure of the performance level represented by each segmentation. The source of each segmentation in the collection may be an appropriately trained human rater (or raters), or an automated segmentation algorithm, such as registered atlas classifiers. The probabilistic estimate of the true segmentation is formed by estimating an optimal combination of the segmentations, weighting each segmentation depending upon the estimated performance level, and incorporating a prior model of the spatial distribution of structures being segmented as well as spatial homogeneity constraints (Warfield et al., 2004). The STAPLE algorithm estimates a ground truth bone map from given bone atlas binary maps (T_{Sn}). Let θ_n be a matrix where each element describes the probability that atlas n labels a voxel as bone (b) when the true label is s ($\theta_n(b,s)$). The perfect atlas will have a probability matrix (θ_n) equal to the identity matrix. Let $\theta = [\theta_1 \ldots\ \theta_N]$ be the unknown set of all probability matrices characterizing all atlas images (N) and $B = [B_1 \ldots B_N]$ be a vector representing the unknown ground truth bone label map and D be an $V \times N$ matrix (V is the number of image voxels) whose columns indicate the N unknown segmentations. STAPLE estimates the ground truth bone segmentation (B) as well as the parameter matrix (θ) by maximizing the log likelihood $f = (D,\ B|\theta)$ using the expectation maximization algorithm (Warfield et al., 2004).

Since the implementation of STAPLE algorithm is not very straightforward and is computationally demanding, Martin-Fernandez et al. (2005) proposed Williams' index whereby the classifiers are assigned weights based on mutual similarity with other classifiers and the general consensus agreed on by all classifiers. Williams' index is defined as:

$$I_n = \frac{(N-2)\sum_{i\neq n}^{N} a(T_{An}, T_{Ai})}{2\sum_{i\neq n}^{N}\sum_{k\neq n}^{i} a(T_{Ai}, T_{Ak})} \tag{11}$$

where N is the number of classifiers or atlases, T_{An} denotes the segmented bone provided by the nth atlas and $a(T_{An},T_{Ai})$ is the agreement between the classifier T_{An} and T_{Ai} over all image voxels. Various agreement measures can be used; a few of them will be defined in Section 2.5. We used the Dice similarity coefficient (Dice, 1945) for this purpose. In case the atlas n generates an index (I_n) greater than one, it can be concluded that the performance of this atlas coincides with the majority of the other atlases. Therefore, this index can be used to select effective atlases (Williams, 1976). The evaluation performed in Martin-Fernandez et al. (2005) demonstrates that the output of STAPLE analysis and Williams' index are similar. In this work, we implemented both algorithms to compare their performance in terms of segmentation accuracy.

2.4.4. Local weighting

The voxel-wise weighting procedure is carried out similarly to the global weighting scheme, except that the similarity measure between the target image and transformed atlas is obtained independently for each voxel within its surrounding image patch (D). The same image similarity criteria (NMI, NNC and MSD) used in global weighting are utilized here, except that the searching window parameter D (patch size) introduced above needs to be optimized. As such, the NMI similarity measure between the target MR image T and the nth transferred atlas image $M(Amr_n)$ for voxel x considering its D neighborhood is defined as:

$$NMI_D(x) = \frac{H_D(T) + H_D(M(Amr_n))}{H_D(T, M(Amr_n))} \tag{12}$$

The fast convolution-based approach proposed by Cachier et al. (2003) is used to compute the local normalized cross-correlation (LNCC).

$$LNCC_D(x) = \sum_{D} \frac{T,\ M(Amr_n)_x}{\sigma(T)_x\ .\sigma(M(Amr_n))_x}$$

$$\text{where}\ \ \sigma(T)_x = \sqrt{\overline{T_x^2} -\ \overline{T_x}^2}\ \ \bar{T}_x = K_G\ *\ T_x$$

$$\langle T,\ M(Amr_n)_x\rangle = \overline{T.M(Amr_n)}_x -\ \bar{T}_x.\overline{M(Amr_n)}_x \tag{13}$$

where K_G and * denote the Gaussian kernel and convolution operator, respectively. A Gaussian kernel with standard deviation equal to 3 voxels (4 mm) was adopted in this study. The MSD image similarity over the image patch D is defined as:

$$MSD_D(x) = \frac{D}{\sum_{x\in D} |T(x) - M(Amr_n(x))|^2} \tag{14}$$

Voxel-wise weighting atlas fusion using the gain exponent was used in Artaechevarria et al. (2009) for brain MR image segmentation. The gain exponent is used to boost the sensitivity of the similarity measure across the atlas dataset. The weighing factor would have the form $w_n(x)_D = SM_D(T,M(Amr_n))_x{}^P$. SM_D could be any of the image similarity criteria (NMI, LNCC and MSD) calculated over the block D centered at voxel x. The obtained weighting factor could be incorporated in IA or MV schemes as:

$$T_{av}(x)_D = \frac{1}{Nr}\sum_{n=1}^{N} w_n(x)_D.T_{An}(x)$$

$$Bn(x) = \begin{cases} 1, & \text{if} \quad T_{av}\ (x)_D > 180 \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$T_{Sav}(x)_D = \frac{1}{Nr}\sum_{n=1}^{N} w_n(x)_D.T_{Sn}(x)$$

$$Bn(x) = \begin{cases} 1, & \text{if} \quad T_{Sav}\ (x)_D > 0.5 \\ 0, & \text{otherwise} \end{cases}$$

$$Nr = \sum_{n=1}^{N} w_n(x)_D \tag{16}$$

The second approach for calculating the weighting factors is similar to that described in Section 2.4.2 as ranking scheme. The only difference is that the ranking step must be performed for each image voxel (considering the surrounding voxels in the window D) rather than the entire atlas image across the whole dataset. After calculating the voxel-wise ranking vector $R(x)_D$ on the basis of image similarity criteria, the weighting factor is obtained via:

$$w_n(x)_D = e^{-aR(x)_D} \tag{17}$$

Again, this weighting factor can be replaced either in Eq. (15) or (16) to perform the final segmentation step. The same local weighting atlas fusion strategy was exploited by Burgos et al. for attenuation map synthesis in brain PET/MRI (Burgos et al., 2013, 2014).

Another strategy for utilizing voxel-wise similarity considers only the information of the most similar voxel. To this end, after computing the voxel-wise ranking vector $R(x)_D$, only the intensity information (or segmentation label) of the foremost voxel is assigned to the final segmented image. From now on, this is referred to as the most similar voxel (MSV).

In this section, three voxel-wise atlas fusion schemes were introduced with the aim to seek the optimal value of their free parameters, namely P, a and D. To fulfill this endeavor, we first calculated the image similarity measure between the target and any of the atlas images using NMI, LNCC and MSD formula for a searching window $D = 10$ mm (in each direction x, y and z). Then, at a fixed value of D, the optimal values of the parameters P and a were determined. In the next step, the obtained optimal values of the parameters P and a were kept fixed to find the optimal value of D.

2.4.5. Hofmann's approach

Hofmann et al. (2011) proposed an approach of generating whole-body pseudo-CT images from MRI. This method relies on a combination of atlas registration and pattern recognition via Gaussian process regression (GPR) (Hofmann et al., 2008). Atlas registration process might fail to match the target patient perfectly because of local minima of non-rigid deformation energy function. To alleviate the adverse effect of the local signal mismatch, the nearby texture information of a given voxel was fed into a GPR via the patch of surrounding voxels to predict more accurate pseudo-CT values. To this end, a set of MRI/CT pairs are non-rigidly aligned to the target MR image and then the GPR kernel is formed using the local image patches on target and atlas images. In addition, 5-class segmentation (background air, lung, fat, fat & non-fat mixture and non-fat tissue) is performed on in-phase MR images (Bezrukov et al., 2013; Hofmann et al., 2011) and the corresponding patch information is used into the GPR kernel through Eq. (18).

$$k(d_i, d_j) = \exp\left(\frac{\left\| -W(P_{MR,i}) - W(P_{MR,j})\right\|^2}{2\sigma^2_{MR,patch}}\right) \times \exp\left(\frac{\left\| -X_i - X_j\right\|^2}{2\sigma^2_{pos}}\right) \times \exp\left(\frac{\left\| -W(P_{Seg,i}) - W(P_{Seg,j})\right\|^2}{2\sigma^2_{Seg,patch}}\right) \tag{18}$$

where $d = (P_{MR}, P_{Seg}, X)$ while P_{MR} and P_{Seg} are sub-volume patches from the in-phase MR image and 5-class segmented MR image, respectively. W is a weighting vector with higher value for central voxels in the patch relative to surrounding voxels, X is the training and test patch center position. The parameters $\sigma_{pos,}$ $\sigma_{MR,patch}$ and $\sigma_{Seg,patch}$ determine how the overall kernel value is influenced by similarity in position and patch intensity value in MRI and 5-class segmentation image. The training is performed on the samples d drawn from random locations in the atlas database. Finally, Eq. (19) is used to calculate the pseudo-CT value of a given voxel.

$$c_l = k^T C^{-1} y \tag{19}$$

where c_l denotes the calculated pseudo-CT value of the voxel of interest l. $k_l = k(d_i,d_l)$ stands for a $(n \times 1)$ vector where $d_i = (P_{MR,i}, P_{Seg,i}, X_i)$ is the information extracted from the patches of the MRI atlases and $d_l = (P_{MR,l}, P_{Seg,l}, X_l)$ indicates the information obtained from the patches of the target MRI. $C = k(d_i,d_j)$ represents the covariance matrix $(n \times n)$ obtained from Eq. (18) using d_i and d_j patches on the MRI atlases. y is an $(n \times 1)$ vector containing CT values corresponding to the central voxel of training patches d_i.

2.4.6. Shape-based averaging

Shape-based averaging (SBA), categorized as an atlas-based segmentation technique, is a voting scheme where each vote is weighted by the signed Euclidean distance computed for each input label. SBA voting is the only method incorporating spatial information in the label fusion process (Rohlfing and Maurer, 2007). Let $d_n(x)$ denote the signed Euclidean distance of voxel x from the nearest surface voxel with bone label in the nth atlas segmentation. A negative value of $d_n(x)$ corresponds to the inside bony structure of the nth atlas while a positive value implies that x is located outside. A value equal to zero is obtained if and only if voxel x is on the surface of bony structure. In effect, the signed Euclidean distance provides a probability map for the presence of bone based on every single atlas segmentation. By computing the distance maps of bony structures in all aligned atlas images, the average distance of a given voxel x from the bone surface is obtained from:

$$AD(x) = \frac{1}{N}\sum_{n=1}^{N} d_n(x) \tag{20}$$

Interested readers are referred to Rohlfing and Maurer Jr (2005) for more details on implementation of the SBA algorithm.

In addition to the spatial weight that is assigned to each voxel using the SBA algorithm on the basis of the Euclidean distance, the

Table 1
Comparison of validation measures (mean±SD), including Dice similarity (DSC), relative volume distance (RVD), Jaccard similarity (JC), sensitivity (S) and mean absolute surface distance (MASD) between the bone extracted from different methods of global weighting atlas fusion using intensity averaging (IA) and majority voting (MV) approaches together with the optimum weighting parameters Φ, *P* and *a* for MI, NCC and MSD image similarity measures. (∗) indicates *P*-value < 0.05 according to the paired *t*-test analysis.

Similarity measure	Weighting parameter	DSC	RVD(%)	JC	S	MASD(mm)
NMI						
IA	$\Phi = 0.50$	0.64 ± 0.06	−36.5 ± 05.6	0.47 ± 0.05	0.53 ± 0.06	06.4 ± 01.5
MV	$\Phi = 0.55$	0.64 ± 0.05	−40.1 ± 04.8	0.47 ± 0.06	0.51 ± 0.07	06.8 ± 01.7
IA	$P = 5$	0.63 ± 0.06	−41.5 ± 05.8	0.46 ± 0.05	0.50 ± 0.05	06.9 ± 01.5
MV	$P = 4$	0.63 ± 0.07	−43.1 ± 06.0	0.46 ± 0.06	0.49 ± 0.06	06.9 ± 01.8
IA	$a = 1$	0.63 ± 0.05	−41.6 ± 05.7	0.46 ± 0.05	0.50 ± 0.04	07.1 ± 01.6
MV	$a = 1$	0.63 ± 0.06	−41.7 ± 05.9	0.45 ± 0.06	0.49 ± 0.05	07.2 ± 01.7
NCC						
IA	$\Phi = 0.75$	0.64 ± 0.06	−39.9 ± 05.6	0.47 ± 0.06	0.51 ± 0.06	06.7 ± 01.5
MV	$\Phi = 0.8$	0.64 ± 0.06	−39.2 ± 05.9	0.47 ± 0.07	0.51 ± 0.07	06.9 ± 01.7
IA	$P = 6$	0.63 ± 0.05	−42.0 ± 6.0	0.46 ± 0.05	0.50 ± 0.05	06.9 ± 01.6
MV	$P = 5$	0.63 ± 0.06	−43.9 ± 6.0	0.45 ± 0.06	0.49 ± 0.06	07.0 ± 01.7
IA	$a = 1$	0.62 ± 0.05	−43.0 ± 6.1	0.45 ± 0.05	0.49 ± 0.06	07.1 ± 01.6
MV	$a = 2$	0.62 ± 0.05	−43.0 ± 6.3	0.45 ± 0.06	0.49 ± 0.07	07.1 ± 01.7
MSD						
IA	$\Phi = 0.9$	0.65 ± 0.05	−34.0 ± 04.8	0.49 ± 0.04	0.55 ± 0.04	05.7 ± 01.2
MV	$\Phi = 0.9$	0.64 ± 0.05	−36.9 ± 05.2	0.47 ± 0.06	0.53 ± 0.06	05.9 ± 01.2
IA	$P = 10$	0.64 ± 0.05	−37.5 ± 04.0	0.47 ± 0.05	0.52 ± 0.04	05.9 ± 01.3
MV	$P = 10$	0.64 ± 0.06	−39.5 ± 04.9	0.47 ± 0.06	0.52 ± 0.05	06.1 ± 01.4
IA	$a = 2$	0.63 ± 0.05	−41.0 ± 06.1	0.46 ± 0.06	0.50 ± 0.06	06.9 ± 01.6
MV	$a = 2$	0.63 ± 0.06	−41.3 ± 06.2	0.46 ± 0.06	0.50 ± 0.07	07.0 ± 01.7

local weight corresponding to the image similarity measure can also be incorporated in Eq. (20). Sabuncu et al. (2010) included voxel-wise similarity weighting factors in the SBA algorithm to enhance its performance in the context of brain image segmentation. As an extension to this work, we used identical weighting factors defined in Section 2.4.4 and included them in Eq. (20):

$$AD(x) = \frac{1}{N}\sum_{n=1}^{N} w_n(x)_D d_n(x) \tag{21}$$

Applying image similarity measure weighting factor to the SBA method introduces the same optimization parameters, namely *P*, *a* and *D*, for each image similarity criteria (NMI, LNCC and MSD). Since the SBA technique is computationally intensive and time-consuming (Rohlfing and Maurer Jr, 2005), the optimal value of *D* obtained from experiments described in Section 2.4.4 was used to optimize the rest of contributing parameters.

2.5. Evaluation metrics

The evaluation of the accuracy of extracted bone using the various atlas-based segmentation strategies described in Section 2.4 was carried out by comparing the segmentation output to the bone segmented on the corresponding reference CT images using five volume/distance-based measures: Dice similarity (DSC) (Dice, 1945), relative volume difference (RVD) (Uh et al., 2014), Jaccard similarity (JC) (Uh et al., 2014), sensitivity (S) (Ying et al., 2013) and mean absolute surface distance (MASD) (Heckemann et al., 2006).

$$DSC(A,M) = \frac{2|A\cap M|}{|A|+|M|},\quad RVD(A,M) = 100\times\frac{|A|-|M|}{|M|},$$

$$JC(A,M) = \frac{|A\cap M|}{|A\cup M|}.,\quad S(A,M) = \frac{|A\cap M|}{|M|}.,$$

$$MASD(A,M) = \frac{d_{ave}\ (S_A,\ S_M)\ +\ d_{ave}\ (S_M,\ S_A)}{2}$$

where *A* is the segmented bone from the reference CT image and *M* denotes the extracted bone by the atlas-based segmentation technique. $d_{ave}(S_A,S_M)$ is the average direct surface distance from all points on the reference bone surface S_A to the segmented bone surface S_M.

The Shapiro-Wilk test was used to examine the null hypothesis that the calculated evaluation metrics follow a normally distributed population and the calculated *p*-values were reported for each individual segmentation scheme. The differences were considered statistically significant if the *p*-value was less than 0.05.

3. Results

Whole-body bone segmentation through non-weighting averaging was performed for varying number of atlases selected randomly from the entire dataset. Fig. 1 illustrates the accuracy of extracted bone in terms of DSC and RVD validation measures using both IA and MV. The bars show the standard deviation at each measured point.

Fig. 2 depicts the accuracy of extracted bone using the weights defined in Eq. (7) for varying threshold levels. The top and bottom rows depict the results obtained using IA and MV frameworks, respectively, for NMI, NCC and MSD image similarity measures. A similar analysis was repeated to obtain the optimal value of parameters *P* and *a* (Table 1). The comparison was made using the five validation measures described in Section 2.5.

Fig. 3 depicts the accuracy of extracted bone based on DSC and RVD validation measures calculated at different values of *P* and *a* for NMI, LNCC and MSD similarity criteria using a searching window of $D = 10$ mm. The results illustrated in Fig. 3 are obtained using the IA framework.

The best result at $D = 10$ mm is achieved by the MSD similarity measure with $P = 3.5$ using the IA framework, yielding a DSC of 0.75, thus demonstrating significant improvement compared to the global weighting strategy (DSC = 0.65). After determining the optimal value of *P* and *a*, these parameters were kept fixed and the optimum size of the searching window *D* was calculated. Fig. 4 depicts the impact of varying size of the searching window *D* on the accuracy of extracted bone for different image similarity criteria. The top row corresponds to the ranking scheme obtained from Eq. (17) at $a = 1$ whereas the bottom row corresponds to the MSV

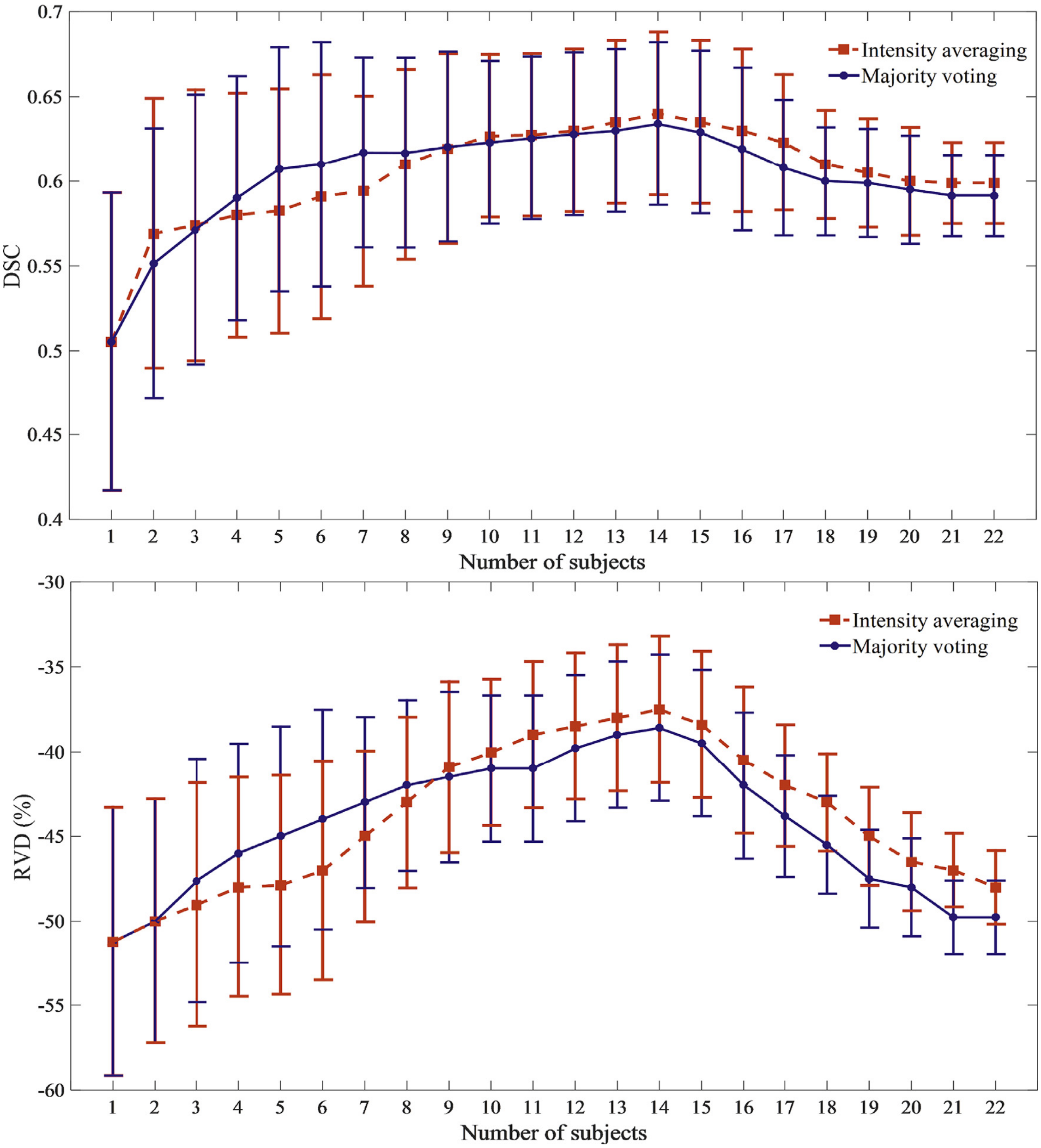

Fig. 1. DSC (top) and RVD (bottom) similarity measures vs. the number of subjects using the intensity averaging and majority voting frameworks.

approach using the IA framework. Table 2 summarizes voxel-wise atlas fusion results together with optimal parameter values. The best results were achieved when applying voxel-wise weighting ranking scheme (using $a = 1$) and the MSV approach (using $D = 5$ and MSD similarity measure) with a DSC = 0.81 (Table 2).

Although the SBA method was the most time consuming approach among those studied in this work, this technique exhibited poor performance without local weighting (Table 3). However, incorporating voxel-wise weighting improved the DSC from 0.56 to 0.76. Fig. 5 illustrates the performance of SBA at varying values of a obtained using different image similarity criteria.

A comparison of the performance of the various segmentation techniques is provided in Table 4. The techniques incorporating optimization parameters are reported at their optimal values. Figs. 6–8 illustrate a representative slice of segmented bone from a whole-body MR image together with corresponding error distance map using a combination of methods presented in Table 4.

4. Discussion

Bone segmentation from whole-body MR images proved to be a challenging task. We investigated the accuracy of a number of atlas-guided segmentation approaches. Our primary motivation for conducting this work is to identify the most promising algorithms for atlas-guided attenuation correction in PET/MRI. Since the identification and segmentation of bony structures for MRI-guided attenuation map generation, particularly in whole-body imaging, we focused our evaluation on metrics reflecting the accuracy of bone extraction among the various approaches.

A commonly used approach to combine the information provided by deformed atlas images is through IA or MV label fusion schemes (Chakravarty et al., 2013). In theory, in multiple atlas segmentation, increasing the number of input atlases would improve the outcome. As such, the quality of segmentation is expected to improve monotonically by adding more atlases. However, in prac-

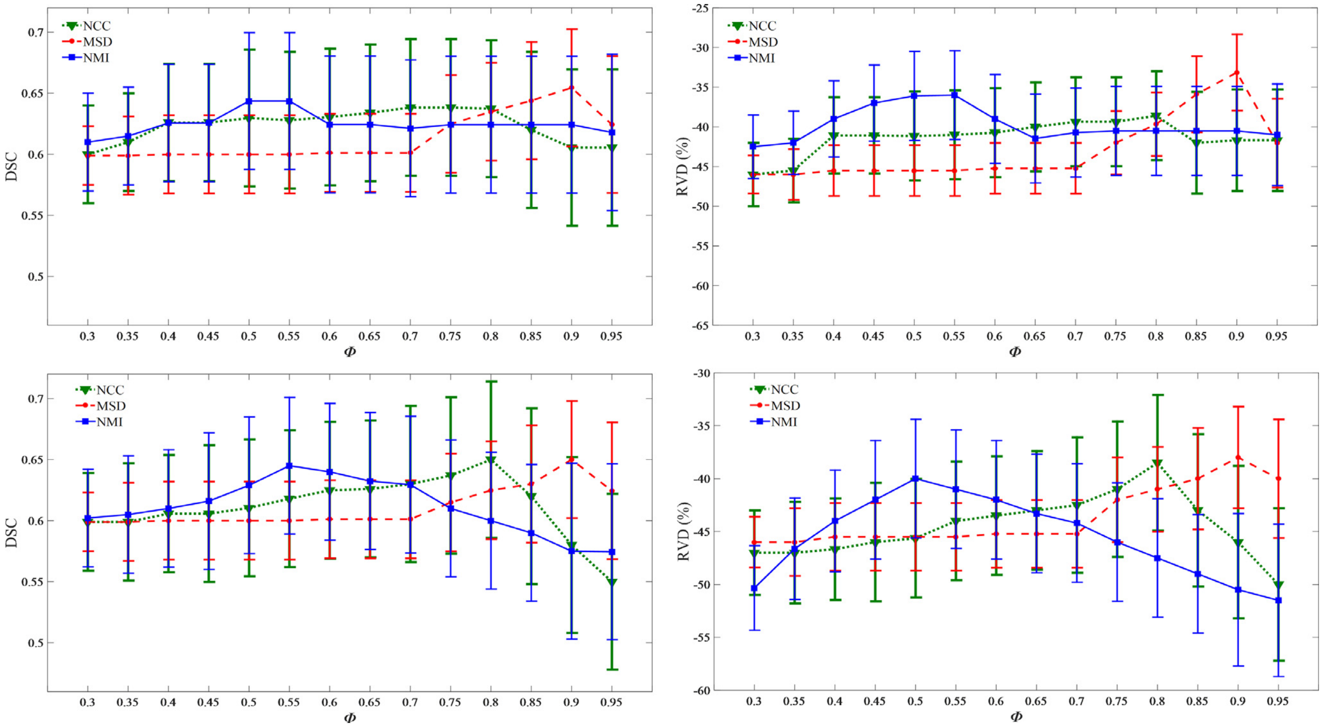

Fig. 2. Plots of DSC and RVD vs. global atlas weighting parameter Φ measured using NCC, MSD and NMI similarity criteria for intensity averaging (top row) and majority voting (bottom row) frameworks.

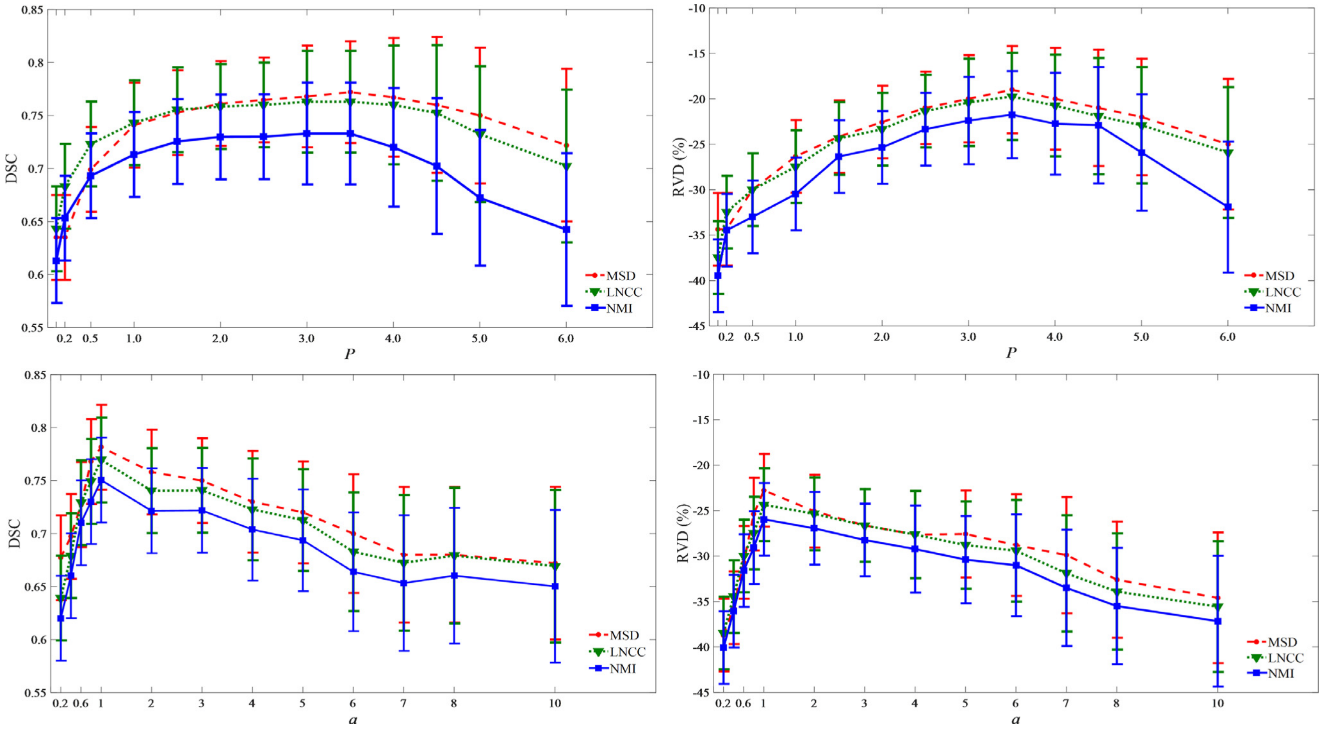

Fig. 3. The effect of varying voxel-wise label weighting parameters (*P* and *a*) on DSC and RVD validation measures obtained from IA segmentation framework using LNCC, MSD and NMI similarity criteria for $D = 10$ mm.

tice at a certain number of input atlases, the improvement reaches a peak (at a number of 14 in Fig. 1). The rising part of the DSC plot (from 1 to 14 subjects) can be justified by the nonsystematic misalignment cancelation due to uncorrelated error between atlases (Artaechevarria et al., 2009; Heckemann et al., 2006). By increasing the number of atlases beyond the peak, the resulting segmentations tend to approach the population mean and the segmentation accuracy will reach an asymptotic value. An overly increased number of atlases would degrade the segmentation accuracy because of the high level of smoothness and the lack of patient-specific details. Assuming that input atlases are of similar quality and are selected randomly, adding more atlases after reaching the peak (here

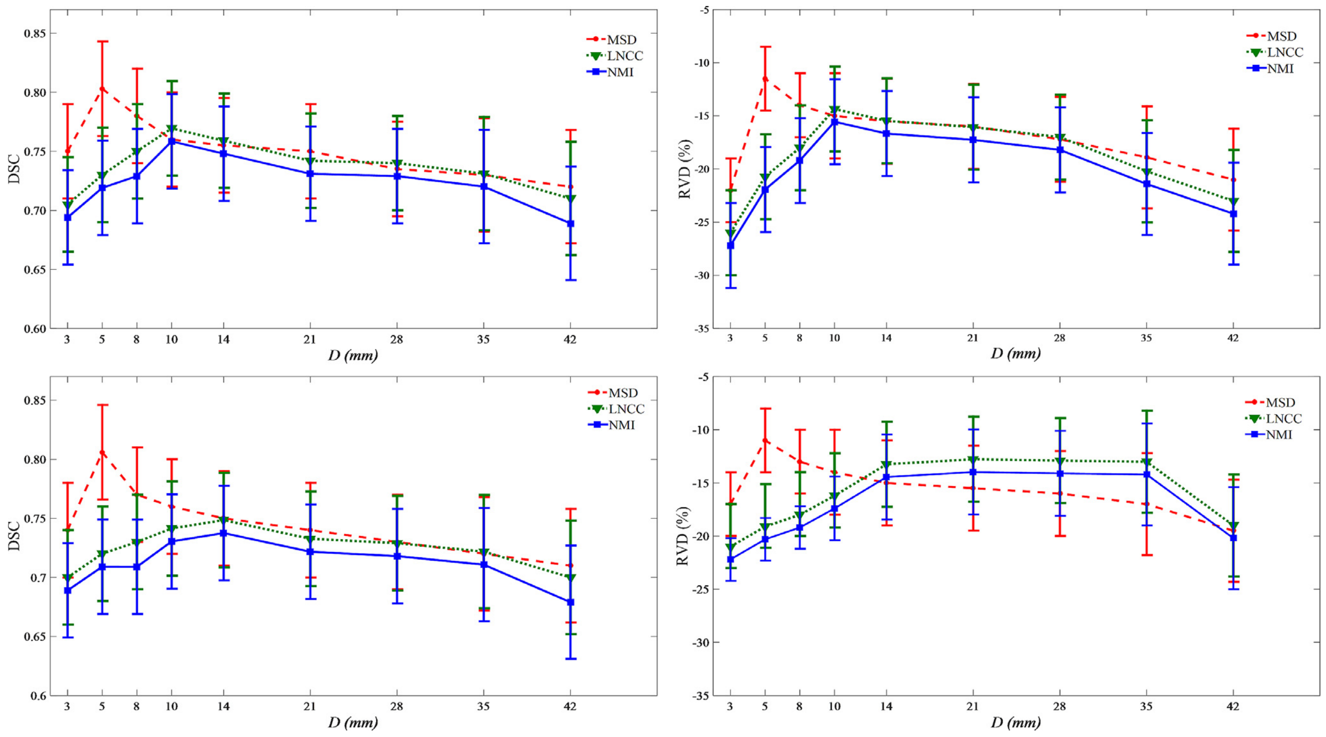

Fig. 4. Plots of DSC and RVD similarity measures vs. neighborhood window *D* using local ranking voxel-wise weighting (top row) and most similar voxel (bottom row) label fusion approaches through the IA framework for LNCC, MSD and NMI similarity criteria.

Table 2
Comparison of validation measures (mean±SD), including Dice similarity (DSC), relative volume distance (RVD), Jaccard similarity (JC), sensitivity (S) and mean absolute surface distance (MASD) between the bone extracted from different methods of voxel-wise weighting calculated using intensity averaging (IA) and majority voting (MV) approaches at the optimum neighborhood window *D* and weighting parameters. (*) indicates *P*-value < 0.05 according to the paired *t*-test analysis.

NMI	Weighting parameter	D (mm)	DSC	RVD(%)	JC	S	MASD(mm)
IA	*P* = 3.5	14	0.75 ± 0.04	−18.1 ± 05.2	0.61 ± 0.05	0.68 ± 0.05	07.3 ± 01.9
MV	*P* = 3.5	14	0.75 ± 0.05	−18.6 ± 05.1	0.61 ± 0.04*	0.69 ± 0.05*	07.4 ± 01.8
IA	*a* = 1	10	0.75 ± 0.04	−15.3 ± 04.6	0.60 ± 0.05	0.68 ± 0.04	07.3 ± 01.6
MV	*a* = 1	10	0.74 ± 0.04	−15.2 ± 04.9	0.59 ± 0.06	0.68 ± 0.05	07.9 ± 01.7
	MSV	14	0.74 ± 0.05	−14.1 ± 03.7	0.59 ± 0.05	0.67 ± 0.05	08.0 ± 01.8
LNCC							
IA	*P* = 3.5	14	0.77 ± 0.04	−15.6 ± 04.8	0.63 ± 0.04	0.70 ± 0.04	05.2 ± 01.5
MV	*P* = 3.0	14	0.76 ± 0.05	−16.9 ± 04.7	0.62 ± 0.04	0.69 ± 0.04	05.5 ± 01.7
IA	*a* = 1	10	0.78 ± 0.05	−14.0 ± 04.9	0.62 ± 0.05	0.69 ± 0.04	05.0 ± 01.9
MV	*a* = 1	10	0.77 ± 0.05	−14.5 ± 06.3	0.61 ± 0.06	0.68 ± 0.05	04.8 ± 01.7
	MSV	14	0.75 ± 0.04	−13.7 ± 06.3	0.60 ± 0.05	0.70 ± 0.06	05.5 ± 01.9
MSD							
IA	*P* = 3.5	5	0.80 ± 0.03	−12.4 ± 04.3	0.74 ± 0.05	0.77 ± 0.05	03.3 ± 01.3
MV	*P* = 3.0	5	0.80 ± 0.04	−12.0 ± 04.4	0.74 ± 0.05	0.77 ± 0.05	03.4 ± 01.4
IA	*a* = 1	5	0.81 ± 0.03	−11.7 ± 4.1	0.75 ± 0.04	0.77 ± 0.04	03.0 ± 01.1
MV	*a* = 1	5	0.80 ± 0.03	−11.9 ± 5.2	0.74 ± 0.04	0.77 ± 0.05	03.3 ± 01.2
	MSV	5	0.81 ± 0.04	−10.9 ± 4.7	0.74 ± 0.03	0.77 ± 0.04	04.9 ± 01.0

Table 3
Comparison of validation measures (mean±SD), including Dice similarity (DSC), relative volume distance (RVD), Jaccard similarity (JC), sensitivity (S) and mean absolute surface distance (MASD) for SBA method with ranking voxel-wise weighting approach set at optimum parameter of *a*. (*) indicates *P*-value < 0.05 according to the paired *t*-test analysis.

	a	DSC	RVD(%)	JC	S	MASD(mm)
SBA (without weighting)	–	0.56 ± 0.05	−15.3 ± 04.7	0.39 ± 0.05	0.52 ± 0.06	11.1 ± 04.1
SBA local weighting (NMI) for *D* = 10 mm	5	0.74 ± 0.05	–26.2 ± 06.7	0.57 ± 0.05	0.62 ± 0.06	08.5 ± 03.6
SBA local weighting (LNCC) for *D* = 10 mm	5	0.75 ± 0.06	–24.8 ± 06.4	0.58 ± 0.06	0.64 ± 0.07	09.7 ± 03.9
SBA local weighting (MSD) for *D* = 5 mm	2	0.76 ± 0.05	–24.4 ± 03.2	0.76 ± 0.07	0.79 ± 0.06	07.6 ± 03.3

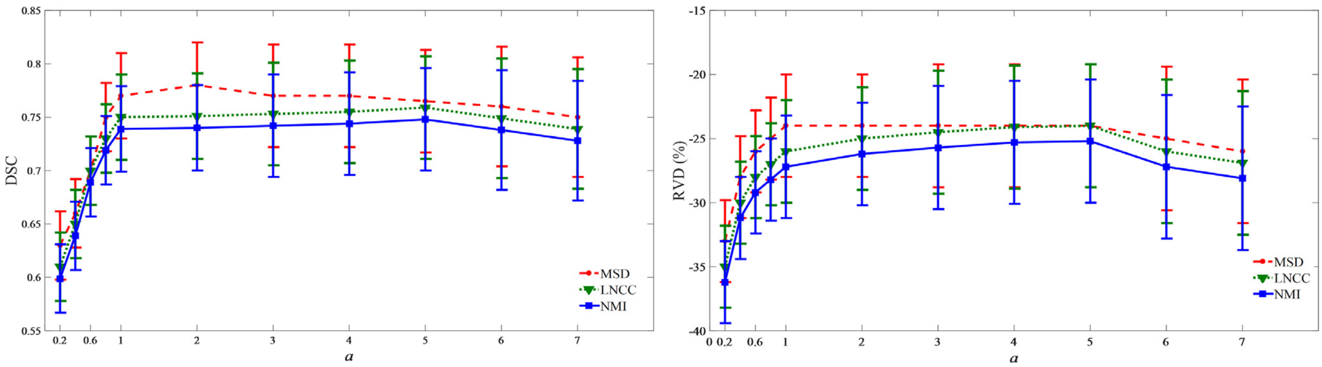

Fig. 5. Impact of varying the voxel-wise weighting parameter (*a*) on DSC and RVD validation measures obtained from the SBA segmentation techniques using an optimum neighborhood window (*D*) of 5 mm for MSD and 10 mm for LNCC and NMI.

Table 4
Comparison of validation measures (mean±SD), including Dice similarity (DSC), relative volume distance (RVD), Jaccard similarity (JC), Sensitivity (S) and mean absolute surface distance (MASD) for all combinations of atlas-based segmentation methods at optimum weighting parameters and neighborhood windows. (*) indicates *P*-value < 0.05 according to the paired *t*-test analysis.

Methods	DSC	RVD(%)	JC	S	MASD(mm)
Average of all atlas images	0.60 ± 0.02	−46.0 ± 02.4	0.43 ± 0.02	0.46 ± 0.02	10.7 ± 03.9
Single atlas image (template)	0.60 ± 0.02	−46.4 ± 02.5	0.44 ± 0.02	0.47 ± 0.02	11.1 ± 03.9
General Majority voting	0.59 ± 0.02	−49.8 ± 02.4	0.42 ± 0.02	0.44 ± 0.02	09.8 ± 03.6
Williams' index	0.61 ± 0.05	−46.2 ± 06.3	0.43 ± 0.06	0.45 ± 0.04	10.4 ± 04.1
STAPLE	0.62 ± 0.05	−4.7 ± 05.8	0.44 ± 0.05	0.49 ± 0.03	08.6 ± 03.8
Hofmann	0.61 ± 0.02	−45.5 ± 02.4	0.42 ± 0.02	0.45 ± 0.02	10.1 ± 03.3
SBA (without weighting)	0.56 ± 0.05	−55.3 ± 04.7	0.39 ± 0.05	0.52 ± 0.06	11.1 ± 04.1
Global weighting (NMI)	0.64 ± 0.06	−39.9 ± 05.6	0.47 ± 0.05	0.53 ± 0.06	06.4 ± 01.5
Global weighting (NCC)	0.64 ± 0.06	−41.5 ± 05.6	0.47 ± 0.06	0.51 ± 0.06	06.7 ± 01.5
Global weighting (MSD)	0.65 ± 0.05	−34.0 ± 04.8	0.49 ± 0.04	0.55 ± 0.04	05.7 ± 01.2
Most similar subject	0.58 ± 0.09	−39.2 ± 08.1	0.41 ± 0.10	0.52 ± 0.11	06.2 ± 02.0
Local weighting (NMI)	0.75 ± 0.04	−15.3 ± 04.6	0.60 ± 0.05	0.68 ± 0.04	07.3 ± 01.6
Local weighting (LNCC)	0.78 ± 0.05	−14.0 ± 04.9	0.62 ± 0.05	0.69 ± 0.04	05.0 ± 01.9
Local weighting (MSD)	0.81 ± 0.03	−11.7 ± 04.1	0.75 ± 0.04	0.77 ± 0.04	03.0 ± 01.1
MSV (MSD)	0.81 ± 0.04	−10.9 ± 04.7	0.74 ± 0.03	0.77 ± 0.04	04.9 ± 01.0

more than 15 in Fig. 1) would average out fine details leading to non-patient-specific and biased segmentation. Moreover, the quality of added atlases is not the same (even though they are chosen randomly), and as such, some fluctuations may be observed with added new atlases. Similar results have been reported in Aljabar et al. (2007) in the context of brain imaging. However, the optimal number of input atlases may vary from one experiment to another since it strongly depends on the shape of the target organ/tissue and quality of atlases. This trend do not seem to be a standard behavior of atlas-based methods since in many studies monotonically increasing or a plateau curve reaching an asymptotic value was reported (Collins and Pruessner, 2010; Heckemann et al., 2006; Wu et al., 2007). It should be noted that this trend holds for non-selective atlas fusion schemes since an increased number of atlases would increase the likelihood of finding more similar cases to the target image, for instance in local weighting atlas fusion schemes which leads to a asymptotically rising curve. In some studies, only one atlas image is used to carry out the segmentation procedure, which involves a single online registration procedure (Greer et al., 2011; Paulus et al., 2015). Marshall et al. (2013) proposed to select the most similar subject for PET/MRI attenuation correction on the basis of available metadata, such as sex and age and some image-derived features, such as body volume, lung volume, etc. The rationale behind using a single atlas or template is to avoid the computational burden of multiple atlas registration. In this work, rather than using metadata, the most similar subject was selected after pair-wise atlas registration using the aforementioned image similarity criteria. Since the most similar atlas is selected after registration, the outcome would be comparable to the original scheme even though it employs a larger database of atlases. The results shown in Table 4 indicate that single patient registration results in large error bias owing to the variability of patients' anatomy while using a single average atlas (template) led to slightly better bone extraction accuracy. Basically, the templates are close to the mean of patient population (compared to single atlas), which reduces non-rigid registration errors and consequently improve the outcome.

The global atlas weighting strategy exhibited moderate improvement compared to the general averaging method as transformed atlases with large misalignment errors are excluded or are at least given relatively low weights during the atlas fusion process. Due to the large axial field-of-view in whole-body imaging, local miss-match between the target and atlases might occur in some cases. Global strategies are not capable of evaluating the registration performance locally and only atlases with gross miss-matches are discarded as demonstrated by Artaechevarria et al. (2009). Therefore, global atlas weighting is much less effective when applied for large axial field-of-view. The similar approach was exploited by Ying et al. (2013) for automated bone segmentation from MR images of the hip joint using the NMI similarity criterion, which resulted in a DSC of 0.95. The marked difference between these results and those reported in our study stems from the different MRI sequence and image quality, registration algorithm and on top of all the smaller field-of-view, which led to better registration outcome and less local miss-matches. According to Table 1, the best performance was achieved by the IA segmentation framework using the MSD image similarity measure with $\Phi = 0.9$, which led to bone segmentation accuracy with a DSC of 0.65.

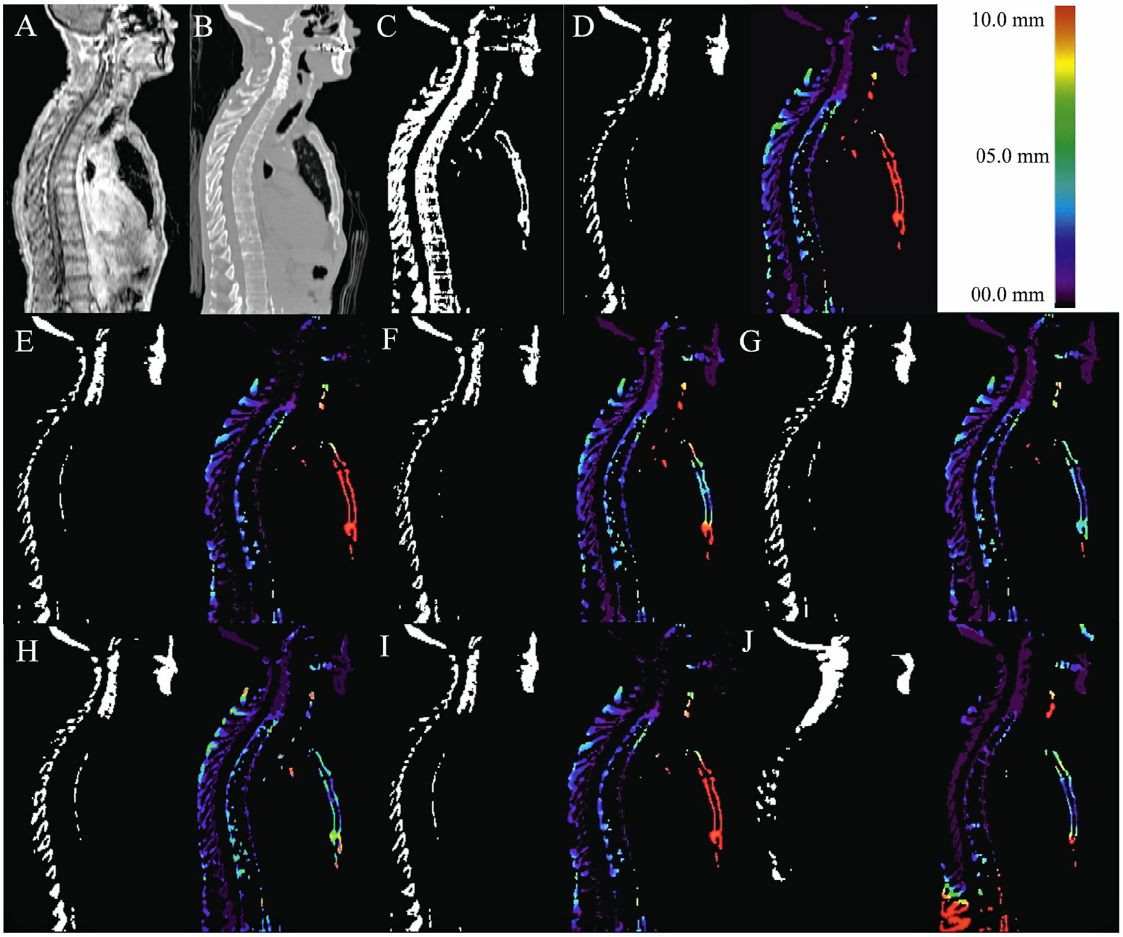

Fig. 6. Representative slice illustrating the performance of the different MRI bone segmentation techniques (left) along with the corresponding error distance map (right) showing: (A) In-phase MRI, (B) corresponding CT image, (C) binary image of reference bone extracted from CT, (D) general intensity averaging, (E) single atlas image, (F) general majority voting, (G) Williams' index, (H) STAPLE, (I) Hofmann's method and (J) SBA.

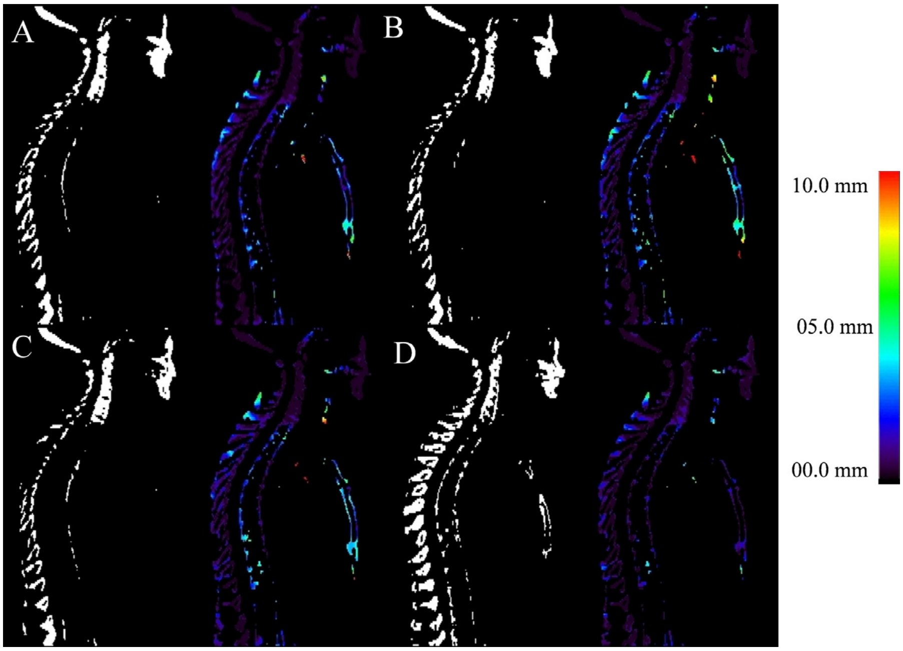

Fig. 7. Representative slice illustrating the performance of the different MRI bone segmentation techniques (left) along with the corresponding error distance map (right) showing: (A) global weighting using NMI, (B) global weighting using NCC and (C) global weighting using MSD, and (D) most similar subject.

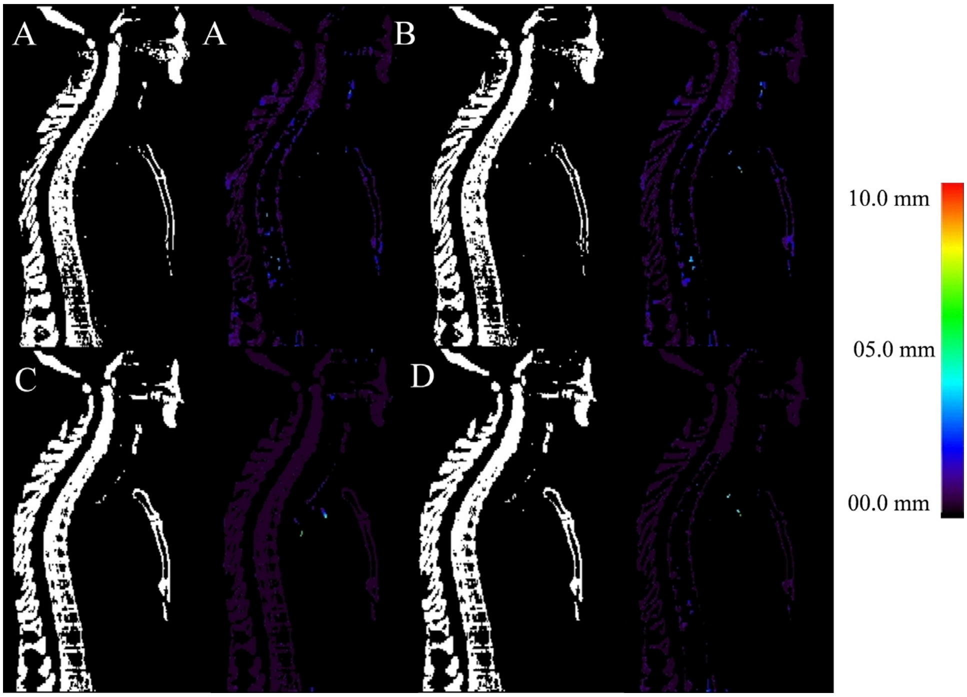

Fig. 8. Representative slice illustrating the performance of the different MRI bone segmentation techniques (left) along with the corresponding error distance map (right) showing: (A) local weighting using NMI, (B) local weighting using LNCC, (C) local weighting using MSD, and (D) most similar voxel using MSD.

There is still a remarkable difference between voxel-wise and global weighting strategies resulting in DSCs of 0.81 and 0.65, respectively. This is mainly due to locally discarded miss-matches between target and atlas images. The transformed atlas image might be aligned almost perfectly in one region whereas other regions bear massive misalignment errors, for instance because of anatomical variability that can properly dealt with it using local weighting strategies. The optimization of free parameters in local strategies (such as the neighborhood window in Fig. 4) plays a key role in local miss-match cancellation. Too large neighborhood windows would result in similar outcome to that of global weighting whereas too small windows would be severely affected by noise and factious local intensity/pattern similarities. The optimization of the searching window *D* seems to be essential since it had significant impact on the accuracy of extracted bone (Fig. 4) (Xie and Ruan, 2014).

MSD achieves the best performance among other image similarity criteria presumably owing to proper intensity normalization of MR images as described in Section 2.3. Implementing the MSD similarity measure adds inconsequential extra computation time to the segmentation procedure, as opposed to NMI which is prohibitively time-consuming when it comes to the voxel level processing (Lötjönen et al., 2010). NMI exhibited the poorest performance as image similarity measure in this particular study. However, in other experiments such as in Burgos et al. (2014), the LNCC similarity measure outperformed other techniques. This issue largely depends on the employed MR sequence, level of noise, inter-subject intensity normalization and registration algorithm. In case the registration between target and atlas images is performed for instance using MI image similarity measure, image alignment is already optimized on the basis of MI. Thus, employing MI after registration did not improve the outcome. On the other hand, the other image similarity measures, particularly MSD, provided more useful information about the similarity between target and atlas images.

The STAPLE framework, regarded as a state-of-the-art atlas fusion method, barely improved bone segmentation accuracy. The competing William's index method performed even slightly worse, nevertheless, it converges up to 4 times faster. The plausible reason of the sub-optimal performance is that both methods solely rely on the correlation between different atlases to determine fusion weights rather than using a similarity measure between target and atlas images. Moreover, the fusion weights are defined in global fashion (similar to global weighting strategies). Therefore, local miss-matches degrade the quality of the outcome. Similar observations were reported elsewhere in the context of prostate and brain segmentation (Artaechevarria et al., 2009, 2008).

Overall, voxel-wise weighting label fusion provided dramatic improvements to the accuracy of segmentation owing to local cancelation of misalignment errors. It is strongly recommended to employ a similarity measure different from the one used for the registration process. Apart from that, the performance of the image similarity criterion largely depends on the type and quality of images under study. In this light, for each study, the optimization step to determine the most effective image similarity criteria and associated optimal parameters, such as image similarity patch size (*D*), is indispensable to reach the best performance (Artaechevarria et al., 2009).

In essence, MRI sequences commonly used to generate PET attenuation maps suffer from high noise level and partial volume effect owing to short acquisition time. Using high quality MR images may possibly improve the registration outcome and consequently the segmentation accuracy; however, fast sequences should be used on PET/MRI systems in the clinic. Using MRI sequences other than the Dixon sequence used in this study would not significantly affect the quality of registration as far as they have similar signal to noise and provide similar anatomical details.

In this work, we focused on conventional atlas-based methods. Patch-based methods rely on a database of atlas images to find similar patches to predict segmentation labels or attenuation values for the target image. However, the characteristic difference between patch-based methods and the methods evaluated in this work is that atlas registration is not performed in patch-based methods while this process is the heart of multiple atlas segmentation.

The major drawback of atlas-based segmentation techniques is the relatively long computation time taken mostly by the image registration process. Future work will focus on reducing the overall computation time and on evaluating the performance of the obtained synthetic pseudo-CT images in the context of attenuation correction in whole-body PET/MRI.

5. Conclusion

We evaluated the accuracy of whole-body bone extraction from MR images using a number of atlas-based segmentation techniques. In particular, global and local weighted atlas fusion strategies as well as some commonly used atlas-based pseudo-CT generation methods were implemented and optimized for the task of whole-body bone segmentation. The voxel-wise weighted atlas fusion approach based on the MSD morphological similarity measure outperformed other segmentation approaches (provided proper MR image denoising and normalization are performed) by achieving a DSC of 0.81. This is in contrast to the non-weighted atlas fusion framework, which yielded a DSC of 0.60. Overall, the voxel-wise weighted atlas fusion approach is capable of canceling out the non-systematic registration errors. Optimization of contributing factors is crucial to reach optimal performance since they are largely determined by the type and quality of images under study.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Swiss National Science Foundation under grant SNFN 31003A-149957 and the Swiss Cancer Research Foundation under Grant KFS-3855-02-2016.

References

Akbarzadeh, A., Gutierrez, D., Baskin, A., Ay, M.R., Ahmadian, A., Riahi Alam, N., Lovblad, K., Zaidi, H., 2013. Evaluation of whole-body MR to CT deformable image registration. J. Appl. Clin. Med. Phys. 14, 238–253.

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2007. Classifier selection strategies for label fusion using large atlas databases. In: Ayache, N., Ourselin, S., Maeder, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. Springer, Berlin Heidelberg, pp. 523–531.

Arabi, H., Rager, O., Alem, A., Varoquaux, A., Becker, M., et al., 2015. Clinical assessment of MR-guided 3-class and 4-class attenuation correction in PET/MR. Mol. Imaging Biol. 17, 264–276.

Arabi, H., Zaidi, H., 2014. Comparison of atlas-based bone segmentation methods in whole-body PET/MRI. IEEE Nuclear Science Symposium & Medical Imaging Conference.

Arabi, H., Zaidi, H., 2016a. Magnetic resonance imaging-guided attenuation correction in whole-body PET/MRI using a sorted atlas approach. Med. Image. Anal. 31, 1–15.

Arabi, H., Zaidi, H., 2016b. One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI. Eur. J. Nucl. Med. Mol. Imaging 43, 2021–2035.

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de-Solorzano, C., 2009. Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE Trans. Med. Imaging 28, 1266–1277.

Artaechevarria, X., Muñoz-Barrutia, A., Ortiz-de-Solorzano, C., 2008. Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle. In: Proc. SPIE, Medical Imaging 2008: Image Processing 69141W-69141W-9.

Ay, M.R., Akbarzadeh, A., Ahmadian, A., Zaidi, H., 2014. Classification of bones from MR images in torso PET-MR imaging using a statistical shape model. Nucl. Instrum. Meth. A 734, 196–200 Part B.

Bezrukov, I., Schmidt, H., Gatidis, S., Mantlik, F., Schafer, J.F., et al., 2015. Quantitative evaluation of segmentation- and atlas-based attenuation correction for PET/MR on pediatric patients. J. Nucl. Med. 56, 1067–1074.

Bezrukov, I., Schmidt, H., Mantlik, F., Schwenzer, N., Brendle, C., et al., 2013. MR-based attenuation correction methods for improved PET quantification in lesions within bone and susceptibility artifact regions. J. Nucl. Med. 54, 1768–1774.

Burgos, N., Cardoso, M., Modat, M., Pedemonte, S., Dickson, J., et al., 2013. Attenuation correction synthesis for hybrid PET-MR scanners. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer, Berlin Heidelberg, pp. 147–154.

Burgos, N., Cardoso, M.J., Thielemans, K., Modat, M., Pedemonte, S., et al., 2014. Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies. IEEE Trans. Med. Imaging 33, 2332–2341.

Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N., 2003. Iconic feature based nonrigid registration: the PASHA algorithm. Comp. Vis. Image. Understand 89, 272–298.

Chakravarty, M.M., Steadman, P., Eede, M.C., Calcott, R.D., Gu, V., et al., 2013. Performing label-fusion-based segmentation using multiple automatically generated templates. Hum. Brain. Mapp. 34, 2635–2654.

Chandra, S.S., Dowling, J.A., Kai-Kai, S., Raniga, P., Pluim, J.P.W., et al., 2012. Patient specific prostate segmentation in 3-D magnetic resonance images. IEEE Trans. Med. Imaging 31, 1955–1964.

Collins, D.L., Pruessner, J.C., 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage 52, 1355–1366.

Delso, G., Wiesinger, F., Sacolick, L., Kaushik, S., Shanbhag, D., Hullner, M., Veit-Haibach, P., 2015. Clinical evaluation of zero echo time MRI for the segmentation of the skull. J. Nucl. Med. 56, 417–422.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Dixon, W.T., 1984. Simple proton spectroscopic imaging. Radiology 153, 189–194.

Greer, P.B., Dowling, J.A., Lambert, J.A., Fripp, J., Parker, J., et al., 2011. A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer. Med. J. Aust. 194, S24–S27.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126.

Hofmann, M., Bezrukov, I., Mantlik, F., Aschoff, P., Steinke, F., et al., 2011. MRI-based attenuation correction for whole-body PET/MRI: Quantitative evaluation of segmentation- and Atlas-based methods. J. Nucl. Med. 52, 1392–1399.

Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., et al., 2008. MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and Atlas registration. J. Nucl. Med. 49, 1875–1883.

Judenhofer, M.S., Wehrl, H.F., Newport, D.F., Catana, C., Siegel, S.B., et al., 2008. Simultaneous PET-MRI: a new approach for functional and morphological imaging. Nat. Med. 14, 459–465.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. Int. J. Comput. Vision 1, 321–331.

Keereman, V., Fierens, Y., Broux, T., De Deene, Y., Lonneux, M., Vandenberghe, S., 2010. MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. J. Nucl. Med. 51, 812–818.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W., 2010. elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205.

Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Elkington, A.G., Mohiaddin, R.H., Rueckert, D., 2004. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Med. Image. Anal. 8, 255–265.

Lötjönen, J.M.P., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., et al., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 49, 2352–2365.

Marshall, H.R., Patrick, J., Laidley, D., Prato, F.S., Butler, J., et al., 2013. Description and assessment of a registration-based approach to include bones for attenuation correction of whole-body PET/MRI. Med. Phys. 40, 082509.

Martin-Fernandez, M., Bouix, S., Ungar, L., McCarley, R.W., Shenton, M.E., 2005. Two methods for validating brain tissue classifiers. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005. Springer, pp. 515–522.

Martinez-Moller, A., Souvatzoglou, M., Delso, G., Bundschuh, R.A., Chefd'hotel, C., et al., 2009. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. J. Nucl. Med. 50, 520–526.

McAuliffe, M.J., Lalonde, F.M., McGarry, D., Gandler, W., Csaky, K., et al., 2001. Medical image processing, analysis and visualization in clinical research. In: *Proc 14th IEEE Symposium on Computer-based Medical Systems*, 2001, pp. 381–386.

Mehranian, A., Arabi, H., Zaidi, H., 2016. Vision 20/20: magnetic resonance imaging-guided attenuation correction in PET/MRI: challenges, solutions, and opportunities. Med. Phys. 43, 1130–1155.

Mehranian, A., Zaidi, H., 2015. Joint estimation of activity and attenuation in whole–body TOF PET/MRI using constrained Gaussian mixture models. IEEE Trans. Med. Imaging 34, 1808–1821.

Paulus, D.H., Quick, H.H., Geppert, C., Fenchel, M., Zhan, Y., et al., 2015. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based MR attenuation correction method including bone. J. Nucl. Med. 57, 1061–1066.

Rezaei, A., Defrise, M., Bal, G., Michel, C., Conti, M., et al., 2012. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. IEEE Trans. Med. Imaging 31, 2224–2233.

Rohlfing, T., Brandt, R., Maurer Jr., C.R., Menzel, R., 2001. Bee brains, B-splines and computational democracy: generating an average shape atlas. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, 2001, pp. 187–194.

Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr, C.R., 2004a. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 21, 1428–1442.

Rohlfing, T., Maurer Jr, C.R., 2007. Shape-based averaging. IEEE Trans. Image Process. 16, 153–161.

Rohlfing, T., Maurer Jr, C.R., 2005. Shape-based averaging for combination of multiple segmentations. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005. Springer, pp. 838–845.

Rohlfing, T., Russakoff, D.B., Maurer Jr, C.R., 2004b. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Trans. Med. Imaging 23, 983–994.

Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. IEEE Trans. Med. Imaging 29, 1714–1729.

Svarer, C., Madsen, K., Hasselbalch, S.G., Pinborg, L.H., Haugbøl, S., et al., 2005. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. Neuroimage 24, 969–979.

Tustison, N.J., Avants, B.B., Cook, P.A., Yuanjie, Z., Egan, A., et al., 2010. N4ITK: improved N3 Bias Correction. IEEE Trans. Med. Imaging 29, 1310–1320.

Uh, J., Merchant, T.E., Li, Y., Li, X., Hua, C., 2014. MRI-based treatment planning with pseudo CT generated through atlas registration. Med. Phys. 41, 051711–051718.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921.

Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. Med. Image Anal. 1, 35–51.

Williams, G.W., 1976. Comparing the joint agreement of several raters with another rater. Biometrics 619–627.

Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C.S., Aizenstein, H.J., 2007. Optimum template selection for atlas-based segmentation. Neuroimage 34, 1612–1618.

Xie, Q., Ruan, D., 2014. Low-complexity atlas-based prostate segmentation by combining global, regional, and local metrics. Med. Phys. 41 041909-041909.

Ying, X., Jurgen, F., Shekhar, S.C., Raphael, S., Craig, E., et al., 2013. Automated bone segmentation from large field of view 3D MR images of the hip joint. Phys. Med. Biol. 58, 7375–7390.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., et al., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116–1128.

Yushkevich, P.A., Wang, H., Pluta, J., Das, S.R., Craige, C., et al., 2010. Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. Neuroimage 53, 1208–1224.

Zaidi, H., Del Guerra, A., 2011. An outlook on future design of hybrid PET/MRI systems. Med. Phys. 38, 5667–5689.

Zaidi, H., Ojha, N., Morich, M., Griesmer, J., Hu, Z., et al., 2011. Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system. Phys. Med. Biol. 56, 3091–3106.