

## Whole-body bone segmentation from MRI for PET/MRI attenuation correction using shape-based averaging

ARABI, Hossein, ZAIDI, Habib

### Abstract

The authors evaluate the performance of shape-based averaging (SBA) technique for whole-body bone segmentation from MRI in the context of MRI-guided attenuation correction (MRAC) in hybrid PET/MRI. To enhance the performance of the SBA scheme, the authors propose to combine it with statistical atlas fusion techniques. Moreover, a fast and efficient shape comparison-based atlas selection scheme was developed and incorporated into the SBA method.

### Reference

ARABI, Hossein, ZAIDI, Habib. Whole-body bone segmentation from MRI for PET/MRI attenuation correction using shape-based averaging. *Medical physics*, 2016, vol. 43, no. 11, p. 5848-5861

PMID : 27806602

DOI : 10.1118/1.4963809

Available at:

<http://archive-ouverte.unige.ch/unige:90679>

Disclaimer: layout of this document may differ from the published version.



# Whole-body bone segmentation from MRI for PET/MRI attenuation correction using shape-based averaging

Hossein Arabi

*Division of Nuclear Medicine and Molecular Imaging, Department of Medical Imaging, Geneva University Hospital, Geneva 4 CH-1211, Switzerland*

Habib Zaidi<sup>a)</sup>

*Division of Nuclear Medicine and Molecular Imaging, Department of Medical Imaging, Geneva University Hospital, Geneva 4 CH-1211, Switzerland; Geneva Neuroscience Center, Geneva University, Geneva CH-1205, Switzerland; Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, Netherlands; and Department of Nuclear Medicine, University of Southern Denmark, Odense DK-500, Denmark*

(Received 20 June 2016; revised 15 September 2016; accepted for publication 18 September 2016; published 5 October 2016)

**Purpose:** The authors evaluate the performance of shape-based averaging (SBA) technique for whole-body bone segmentation from MRI in the context of MRI-guided attenuation correction (MRAC) in hybrid PET/MRI. To enhance the performance of the SBA scheme, the authors propose to combine it with statistical atlas fusion techniques. Moreover, a fast and efficient shape comparison-based atlas selection scheme was developed and incorporated into the SBA method.

**Methods:** Clinical studies consisting of PET/CT and MR images of 21 patients were used to assess the performance of the SBA method. In addition, the authors assessed the performance of simultaneous truth and performance level estimation (STAPLE) and the selective and iterative method for performance level estimation (SIMPLE) combined with SBA. In addition, a local shape comparison scheme (L-Shp) was proposed to improve the performance of SBA. The SIMPLE method was applied globally (G-SIMPLE) while STAPLE method was employed at both global (G-STAPLE) and local (L-STAPLE) levels. The evaluation was performed based on the accuracy of extracted whole-body bones, fragmentation, and computation time achieved by the different methods. The majority voting (MV) atlas fusion scheme was also evaluated as a conventional and commonly used method. MRI-guided attenuation maps were generated using the different segmentation methods. Thereafter, quantitative analysis of PET attenuation correction was performed using CT-based attenuation correction as reference.

**Results:** The SBA and MV methods resulted in considerable underestimation of bone identification (Dice  $\approx$  0.62) and high factious fragmentation error of contiguous structures. Applying global atlas selection or regularization (G-STAPLE and G-SIMPLE) to the SBA method enhanced bone segmentation accuracy up to a Dice = 0.66. The best results were achieved when applying the L-STAPLE method with a Dice of 0.76 and the L-Shp method with a Dice of 0.75. However, L-STAPLE required up to five-fold increased computation time compared to the L-Shp method. Moreover, both L-STAPLE and L-Shp methods resulted in less than 3% SUV mean relative error and 6% SUV mean absolute error in bony structures owing to superior bone identification accuracy. The quantitative analysis using joint histograms revealed good correlation between PET-MRAC images using the proposed L-Shp algorithm and the corresponding reference PET-CT images.

**Conclusions:** The performance of SBA was enhanced through application of local atlas weighting or regularization schemes (L-STAPLE and L-Shp). Bone recognition, fragmentation of the contiguous structures, and quantitative PET uptake recovery improved dramatically using these methods while the proposed L-Shp method significantly reduced the computation time. © 2016 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4963809>]

Key words: PET/MRI, attenuation correction, bone segmentation, atlas-guided, quantification

## 1. INTRODUCTION

Image segmentation and anatomical regions delineation, aiming at classifying a subject image into volumes of interest and background, play a major role in medical image processing. In many situations, such as bone segmentation from magnetic resonance images (MRI), nonideal image

quality, or lack of contrast challenges the segmentation task. Due to the high demand for (semi-) automated segmentation techniques, multiatlas based framework has attracted considerable attention because of its promising performance, robustness, and flexibility.<sup>1</sup> Technically, multiatlas refers to a set of images, preferably from the same body region of the test image together with approved target segmentation

information. Multiatlas based segmentation typically entails the following steps: (i) individual registration of atlas images to the test image followed by propagation (using the obtained deformation fields) of the corresponding label information to the test subject coordinates, and (ii) final decision on labeling of test image through atlas fusion schemes. Inclusion of multiple atlases is considered to be an effective procedure, which has been found to be more accurate than any of the individual atlases.<sup>1,2</sup>

The atlas fusion task is typically carried out using voxelwise decision merging techniques, such as majority voting (MV)<sup>3</sup> or assigning likelihoods to each of the outcome possibilities via Bayesian classifiers.<sup>4</sup> Moreover, the information provided by each of the raters can be preferably assessed and weighted according to performance and relevance to enhance the effectiveness of the atlas fusion task.<sup>5</sup> To achieve the best possible accuracy, this task is commonly carried out locally or at the voxel level, which might lead to undesirable fragmentation of the structures owing to local noise and spatial uncertainty.<sup>6</sup> To address this issue, shape-based averaging (SBA), which exploits the natural distance relationship between the voxels to combine different raters decisions based on the signed distance maps of the atlas label maps, was introduced.<sup>7</sup> Basically, the SBA technique relies on the Euclidean distance transform, which assigns the distance from the nearest boundary of the target object to each voxel of the label map. The decision about the final segmentation is made by computing the average of all distance maps calculated on rater label maps.

Atlas fusion SBA performed as well as or even slightly better than MV in terms of segmentation accuracy and object identification on human brain MRI data.<sup>7</sup> Besides, SBA resulted in less specious fragmentation, which led to significantly more natural and contiguous structures compared to the popular majority voting approach.<sup>6</sup> Moreover, SBA exhibited more robustness to poor resolution and small number of input atlases.<sup>7</sup> Robitaille and Duchesne<sup>8</sup> compared a number of effective atlas fusion schemes, namely SBA, simultaneous truth and performance level estimation (STAPLE)<sup>9</sup> and MV for brain MRI segmentation using datasets from the International Consortium for Brain Mapping (ICBM).<sup>10</sup> In almost all simulations and studies involving segmentation of 2D and 3D neuroanatomical structures, the SBA atlas fusion scheme outperformed the other methods in terms of recognition rate.

The outstanding merit of SBA, in addition to its robustness to noise and homogeneity of output labeling, is thought to be the good performance on a small number of input atlases. In this regard, we chose the SBA method for bone volume identification from whole-body MR images with the aim to incorporate bone in the derivation of attenuation maps for attenuation correction in hybrid PET/MRI. Correcting the PET emission data for attenuation is still a major challenge in PET/MRI owing to the lack of direct correlation between MRI signal and attenuation properties of biological tissues.<sup>11</sup> The complexity and variability of the human anatomy combined with the high level of noise and partial volume effect make the segmentation of bone from MRI a challenging task. In addition, application of dedicated MR sequences enabling to delineate bony structures, such as ultra-short echo time

(UTE)<sup>12</sup> or zero time echo (ZTE),<sup>13</sup> in whole-body imaging is not yet clinically feasible owing to the long acquisition time and susceptibility to artifacts when using a large field-of-view. As such, the use of atlas-based methods for identification of bone tissue is common practice in MRI-guided attenuation generation in brain<sup>14–17</sup> and more importantly in whole-body imaging.<sup>18–20</sup>

Despite promising potential reported in the literature, we observed that the SBA technique tends to underestimate bone volume identification in a comparative assessment of common atlas-based bone segmentation techniques from MRI using a set of MR/CT image pairs.<sup>21</sup> Although SBA exhibited poor performance compared to other techniques, its best performance was achieved on a relatively small number of input atlases.

In this work, we aim to propose the SBA method for whole-body bone segmentation from MRI motivated by its remarkable merits, in particular robustness to a small number of input atlases and minimal factitious fragmentation of contiguous structures. These features could be appealing to PET/MRI attenuation map generation, thus allowing the derivation of artifact free attenuation maps using gender/body/posture-specific atlas dataset, which may contain only few atlas images. The main contribution of this work is the combination of the promising SBA atlas fusion scheme with state-of-the-art statistical atlas selection/weighting methods, such as STAPLE (Ref. 9) and selective and iterative method for performance level estimation (SIMPLE),<sup>22</sup> to adapt the SBA technique and enhance its performance for the purpose of whole-body bone segmentation from MRI. In addition, a simple atlas weighting scheme based on shape similarity with a competitive performance was proposed for incorporation within the SBA technique. The performance of different atlas fusion schemes was evaluated in terms of whole-body bone extraction and PET attenuation correction accuracy.

## 2. MATERIALS AND METHODS

### 2.A. Data acquisition

The study population comprised 21 patients who underwent whole-body <sup>18</sup>F-FDG PET/CT and MRI scanning for staging of head and neck malignancies. The study protocol was approved by the institutional ethics committee and the patients signed written informed consent. PET/CT imaging was performed on a Biograph 64 True Point scanner (Siemens Healthcare, Erlangen, Germany). CT images were acquired for attenuation correction using an unenhanced low dose CT scan (120 kVp, 60 mAs, 24×1.5 collimation, pitch of 1.2 and 1 s/rotation) after a localization scout scan. PET data acquisition started at 146.2 ± 20 min post-injection of <sup>18</sup>F-FDG (371 ± 23 MBq) with 3 min per bed position for a total of 4–5 beds, resulting in a total acquisition time of 15–18 min. MRI examinations were performed on the Ingenuity TF PET/MR (Philips Healthcare, Cleveland, USA)<sup>23</sup> using a whole-body MRI Dixon volumetric interpolated T1-weighted sequence<sup>24</sup> using the following parameters: flip angle 10°, TE<sub>1</sub> 1.1 ms,

TE<sub>2</sub> 2.0 ms, TR 3.2 ms, 450 × 354 mm<sup>2</sup> transverse FOV, 0.85 × 0.85 × 3 mm<sup>3</sup> voxel size, and a total acquisition time of 2–3 min. The slight time difference is due to different axial coverages depending on the patients' length. Some patients underwent one or two more bed-positions acquisition. The in-phase image generated by the MRI Dixon sequence was used for the segmentation process. Due to temporal shift between MRI and CT acquisitions, the in-phase MR image of each patient was nonrigidly registered to the corresponding CT image to ensure spatial alignment of MR and CT image pairs. The segmented bone from CT images (produced by applying an intensity threshold of 140 HU) can be considered as the ground truth bone label map for corresponding MR images. The MR and CT image pairs after registration had 512 × 512 matrix resolution in 207 axial slices.

Fast whole-body MRI protocols acquired for attenuation map generation suffer from a relatively high level of noise, corruption due to the low frequency bias field and interpatient intensity inhomogeneity.<sup>25–27</sup> To remove or minimize these effects, in-phase MR images of all patients underwent the following corrections: (i) gradient anisotropic diffusion filtering<sup>28</sup> for statistical noise removal using a conductance = 4, number of iterations = 10, and time step = 0.01, (ii) N4 bias field correction<sup>29</sup> for canceling out intrasubject intensity inhomogeneity using B-spline grid resolution = 400, iteration number = 200 (at each grid resolution), convergence threshold = 0.001, B-spline order = 3, spline distance = 400, number of histogram bins = 256, and shrink factor = 3, (iii) histogram matching<sup>30</sup> to minimize intersubject intensity nonuniformity using a histogram level = 512 and match points = 64.

## 2.B. MRI image segmentation algorithms

### 2.B.1. Shape-based averaging

In principle, the combination of multiple atlas information or segmentations would result in a more accurate outcome than any of the individual segmentations. Unlike other segmentation fusion schemes, SBA exploits a natural distance relationship between the voxels of an image to achieve the segmentation fusion task. Usually, this distance relationship is expressed in terms of signed Euclidean distance maps of the bone label in each of the input atlas images. Let us suppose that  $L_j(x_i)$  stands for the label map of atlas ( $j$ ) at voxel ( $x_i$ ), for instance with a value of 0 for the background (BKG) and 1 for bone, and  $O_j(x_i)$  denotes its corresponding signed Euclidean distance of voxels ( $x_i$ ) from the nearest voxel with the bone label in  $L_j$ .  $O_j(x_i)$  is positive inside bone volume, negative outside, and equal to zero if and only if the voxel resides on the surface of bony structures. In other words, each voxel in the bone label map is weighted based on its distance from the surface of the bony structures. The voxels deep inside the bony structures receive higher positive weights and voxels further outside the bony structures receive higher negative weights [ $O_j(x_i)$ ].

In the first step, the distance maps of all structures in all input atlases/segmentations are computed [ $O_j(x_i)$ ] and then, the average distance ( $D_{ave}$ ) of voxel ( $x_i$ ) from nearest bone

label is defined as

$$D_{ave}(x_i) = \frac{1}{A} \sum_j O_{ji}, \quad (1)$$

where  $A$  indicates the number of input atlas images. Thereafter, the segmentation outcome for each voxel  $T(x_i)$  is determined through minimization of the average distance from combined label maps. Equation (2) tends to assign labels (bone or BKG) to each voxel, such that it minimizes the average distance map ( $D_{ave}$ ) to the assigned labels,

$$T(x_i) = \arg \min_{\text{Bone or BKG}} D_{ave}(x_i). \quad (2)$$

This minimization problem can be iteratively solved by applying Algorithm I below.

Compared to other atlas fusion schemes, SBA produces a smoother, more regular and natural looking output segmentations as well as comparable recognition rate.<sup>7,8</sup> However, the SBA method resulted in a dramatic underestimation of bone recognition in our previous study.<sup>21</sup> This stems from the complexity of structures and the relatively large number of fragmentations of whole-body bone tissue. Figure 1 depicts an example of the SBA fusion scheme, in which three atlas images contributed to the segmentation task while atlas No. 3 exhibited poor performance. Since in the SBA scheme, the fusion task is performed based on distance maps, the poor performance of atlas No. 3 is also magnified in the corresponding distance map at the indicated point, consequently leading to underestimation of the segmentation despite the fact that the majority of the atlases voted differently. Due to the large anatomical variability, different body postures, nonrigid motion and far from perfect accuracy of atlas registration in whole-body imaging (compared, for instance, to brain imaging), this issue is very likely to occur in multiatlas whole-body segmentation when applying the SBA method.

Motivated by the remarkable advantages of the SBA technique, we aimed to improve its performance in the context of whole-body bone segmentation through incorporation of SIMPLE and STAPLE procedures. To this end, the SBA scheme was combined with SIMPLE and STAPLE methods at both local and global levels. Moreover, a simple but effective atlas selection procedure based on shape comparison was

ALGORITHM I. Shape-based averaging algorithm.<sup>7</sup>

1. For all $x$	Initialization: loop over all voxel
$T(x_i) = \text{unknown}$	Unknown label for target label map
$D_{aux}(x_i) = \infty$	Auxiliary distance map
End	
2. For all labels	Main loop: over all labels
For all $x$	
$D_{ave}(x_i) = \frac{1}{A} \sum_j O_j(x_i)$	Average distance map for a specific label
If $D_{aux}(x_i) < D_{ave}(x_i)$	
$T(x_i) = \text{bone}$	
$D_{aux}(x_i) = D_{ave}(x_i)$	
End	
End	
End	

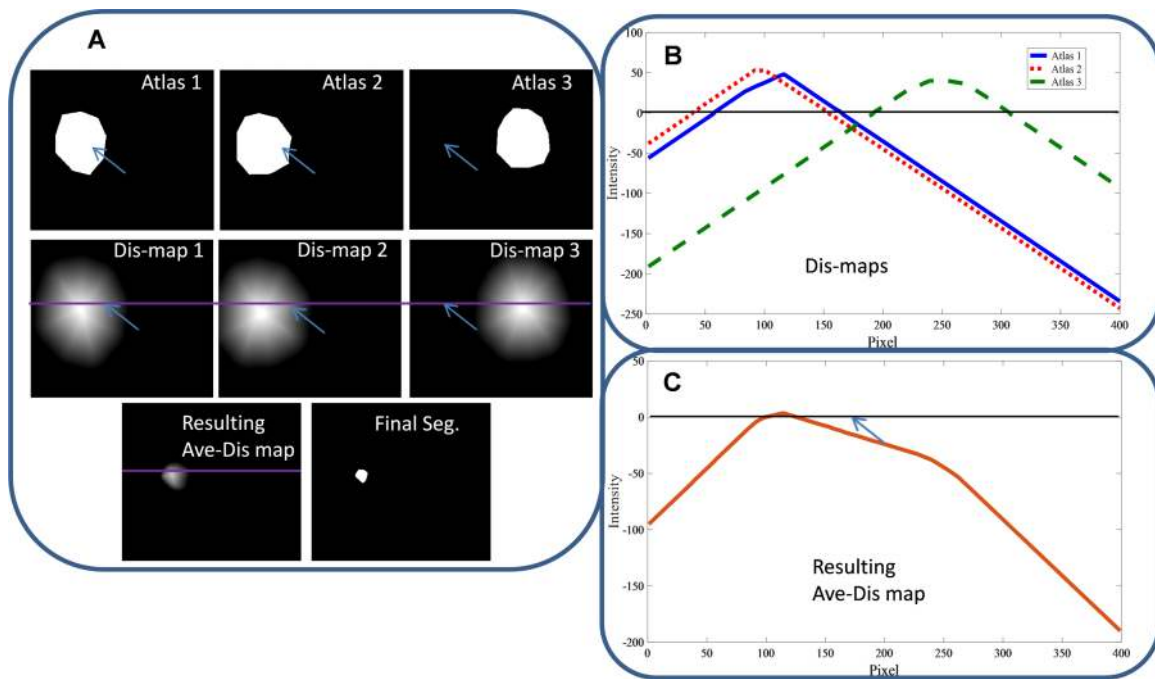


FIG. 1. Illustration of SBA atlas fusion framework. The top left panel (A) depicts three binary label images generated by three different atlases and the middle row shows the corresponding distance maps. The lower row presents the combined distance map together with the final labeling result. The right panel depicts the profiles of distance maps corresponding to the lines indicated on the atlases (B) and resulting (C) distance maps. The poor performance of one of the raters (here atlas 3) led to significant underestimation of the segmentation output.

proposed. In the following, we elaborate on the mathematical foundations and algorithmic implementation of the proposed methods.

**2.B.2. Global SIMPLE (G-SIMPLE)**

SIMPLE employs an atlas selection scheme after registration through iterative performance estimation and uses this performance as a weight in the label fusion process.<sup>22</sup> Poorly performing atlases act as noise in the SBA scheme as illustrated in Fig. 1. At each iteration, SIMPLE discards badly performing atlases such that they no longer contribute to the final segmentation. In the first step, all registered atlases are combined (here using the SBA scheme) to create a rough estimation of the ground truth segmentation ( $T^0$ ).

In the next step, the performance of each atlas ( $\emptyset_j = f(L_j, T^k)$ ,  $j \in SI^k$ ) can be estimated via binary overlap measures (such as Dice =  $\frac{2|L_j \cap T^k|}{|L_j| + |T^k|}$  or SP =  $\frac{TPR(L_j, T^k) + TNR(L_j, T^k)}{2}$ ), where  $SI^k$  indicates the selected atlases at the  $k$ th iteration, TPR (true positive rate) and TNR (true negative rate) indicate sensitivity and specificity, respectively, and SP indicates a binary overlap measure composed of the average of sensitivity and specificity (see Eq. (3) below). The estimated performance of the atlases is used to identify and discard poor performing ones in the next iteration to generate ( $T^{k+1}$ ), which in turn enhances the accuracy of the new estimated segmentation. This process is repeated iteratively to optimize the selection of atlases (Algorithm II). In Algorithm II, we propose to combine the SIMPLE atlas selection/weighting method with the SBA atlas fusion scheme.

To ensure that an atlas image is not excluded prematurely in the first few iterations (here we used three), all atlas images contribute to the segmentation. Technically, the atlas selection threshold ( $\theta$ ) plays a key role as it determines both the convergence rate and performance of the algorithm. To find the optimal value of  $\theta$ , we employed the procedure described in Ref. 22 where the atlas selection threshold is determined as a function of the mean ( $\bar{\emptyset}^k$ ) and standard deviation ( $\sigma^k$ ) of the performance of all atlases ( $j \in SI^k$ ):  $\theta^k = \bar{\emptyset}^k - c\sigma^k$ . The free parameter ( $c$ ) produced the best results when defined as an increasing function of the number of selected atlases ( $c/\text{Num}(SI^k)$ ) because in early iterations a larger number of atlases should be discarded and as the algorithm proceeds, the selection threshold should decrease to discard atlases less quickly. At the end of each iteration, the estimation of the ground truth segmentation is achieved by applying the SBA fusion scheme on the selected atlas images. The binary overlap measures (such as Dice and combination of sensitivity and specificity introduced earlier) between atlas images and the estimated ground truth segmentation ( $T^k$ ) are used to discard poor atlases. In our experiments, the sensitivity performed slightly better than Dice. Therefore, we selected it

ALGORITHM II. Global SIMPLE algorithm (G-SIMPLE).

1. Segmentation ( $T^k$ ) using the SBA scheme weighted by estimated performances ( $\emptyset_j = f(L_j, T^{k-1})$ ) using the selected atlases ( $j \in SI^k$ ).
2. Performance assessment of each input atlas ( $\emptyset_j = f(L_j, T^k)$ ,  $j \in SI^k$ ).
3. A set of atlases is determined ( $SI^k$ ) for the next segmentation provided  $j \in SI^k$  if  $\emptyset_j > \theta$  (atlas selection threshold)

as the similarity measure for threshold selection. The binary overlap measure used to estimate the performance of atlases is estimated globally (whole image). We call this approach G-SIMPLE.

### 2.B.3. Global STAPLE (G-STAPLE)

STAPLE considers a set of label maps, provided by atlas registration, and computes a probabilistic estimate of the ground truth segmentation as well as a criterion of the performance level achieved by each of the atlases.<sup>9</sup> The task of probabilistic estimate of the ground truth segmentation is carried out through combining an optimal subset of the atlases (or label maps) weighted by the estimated performance level. Here, we consider  $Y(x_i)$  as the target MR image with unknown label map  $T(x_i)$  containing  $N$  voxels. Given  $A$  atlas images or label maps ( $L(x_i)$ ) registered to the target image ( $Y$ ), a  $D_{ij}$  matrix of size  $N \times A$  can be formed where the rows ( $D_j$ ) represent  $L_j$ . Assuming an estimate of the ground truth segmentation ( $T$ ) is available, the sensitivity and specificity of each atlas can be calculated using

$$\begin{aligned} \text{Sensitivity (true positive ratio): } & \text{TPR} = \text{TP} / (\text{TP} + \text{FN}), \\ \text{Specificity (true negative ratio): } & \text{TNR} = \text{TN} / (\text{TN} + \text{FP}), \\ p_j = \text{Probability}(D_{ij} = 1 | T(x_i) = 1), \\ q_j = \text{Probability}(D_{ij} = 0 | T(x_i) = 0), \end{aligned} \quad (3)$$

where TP and FN are the true positive and false negative rates, respectively. Then, the sensitivity  $P = (p_1, p_2, \dots, p_A)^T$  and specificity  $Q = (q_1, q_2, \dots, q_A)^T$  matrices can be generated where the elements  $p$  and  $q$  indicate the sensitivity and specificity of one of the atlas images, respectively. Given a matrix  $D$  containing all the atlas label maps and hidden true target label map  $T$ , the probability mass function of the complete data would be  $f(D, T | P, Q)$ , where the goal is to estimate the performance level parameters of the atlases characterized by  $P$  or  $Q$ , which maximize the log likelihood function of the complete data,

$$(\hat{P}, \hat{Q}) = \arg \min_{P, Q} \log(f(D, T | P, Q)). \quad (4)$$

Based on the assumption that atlas label maps are all conditionally independent, given the estimated ground truth segmentation and performance level parameters, a version of the expectation–maximization (EM) algorithm was employed to estimate the solution of the maximization problem. A detailed description of the employed EM algorithm is provided in Refs. 9 and 31. In the *expectation-step* an estimation of the unknown target label map is derived through computation of the conditional probability at each voxel

$$w_i^k \equiv f(T(x_i) = 1 | D_i, P^k, Q^k) = \frac{a_i^k}{b_i^k + a_i^k}, \quad (5)$$

where  $w_i$  indicates the probability of the true target segmentation at voxel ( $i$ ) subject to

$$\text{if } w_i \geq 0.5 \text{ then } T(x_i) = 1 \text{ (bone) else } T(x_i) = 0 \text{ (BKG)}. \quad (6)$$

To merge the SBA atlas fusion scheme and STAPLE atlas selection method, the signed Euclidean distance map of all

atlas labels were computed ( $O_{ij}$ ) considering positive values inside bony structures and negative values outside. Then, the calculated distance map was incorporated into the conditional probability [Eq. (5)] as a weighting factor for each voxel

$$\begin{aligned} a_i^k & \equiv f(T(x_i) = 1) \prod_{j:D_{ij}=1} O_{ij} \times p_j^k \prod_{j:D_{ij}=0} \frac{1}{|O_{ij}|+1} \times (1 - p_j^k), \\ b_i^k & \equiv f(T(x_i) = 0) \prod_{j:D_{ij}=0} |O_{ij}| \times q_j^k \prod_{j:D_{ij}=1} \frac{1}{O_{ij}+1} \times (1 - q_j^k). \end{aligned} \quad (7)$$

Here,  $f(T(x_i) = 1) / f(T(x_i) = 0)$  serves as prior knowledge for the presence/absence of bone in the target image, where a constant value of 1 is considered for all voxels. Equation (7) differs from the original STAPLE method in the sense that it is weighted with signed distance map of bone label.

In the *Maximization-step*, given the estimated weight variables ( $w_i$ ), performance level parameters that maximize the conditional expectation of the complete data log likelihood function can be computed considering

$$\begin{aligned} p_j^{k+1} & = \frac{\sum_i w_i^k D_{ij}}{\sum_i w_i^k}, \\ q_j^{k+1} & = \frac{\sum_i (1 - w_i^k)(1 - D_{ij})}{\sum_i (1 - w_i^k)}. \end{aligned} \quad (8)$$

Regarding Eq. (8), since the sensitivity and specificity of each atlas are estimated using all voxels of the image (as opposed to local estimation) we call this method G-STAPLE.

### 2.B.4. Local STAPLE (L-STAPLE)

So far, the described approaches (G-SIMPLE and G-STAPLE) combined with the SBA scheme were designed to assign weights to each atlas based on the global measurement of performance level. In this section, the same concept is employed for local assignment of weights based on the atlas performance measured on a subset or small patch of images. The procedure (*expectation-* and *maximization-steps*) is quite similar as the one described in Sec. 2.B.3 but the sensitivity and specificity are computed locally and as such, a spatially varying (voxelwise) sensitivity-specificity matrix can be formed as described in Ref. 32,

$$\begin{aligned} \gamma & = (p, q), \\ \gamma & = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \dots & \gamma_{A1} \\ \gamma_{12} & \gamma_{22} & \dots & \gamma_{A2} \\ \dots & \dots & \dots & \dots \\ \gamma_{1N} & \gamma_{2N} & \dots & \gamma_{AN} \end{bmatrix}, \end{aligned} \quad (9)$$

where  $\gamma_{ij}(p_{ij}, q_{ij})$  indicates the sensitivity and specificity pair measured for the atlas  $j$  at the voxel  $i$ . Given a matrix  $\gamma$ , the aim is to estimate the voxelwise performance level parameters of the atlases which maximize the log likelihood function of the complete data,

$$\hat{\gamma} = \arg \max_{\gamma} \log(f(D, T | \gamma)). \quad (10)$$

In a way similar to the one described previously, the *expectation-step* deals with an estimation of the unknown target label map derived from the computation of the conditional probability at each voxel

$$w_i^k \equiv f(T(x_i) = 1 | D_i, \gamma_i^k) = \frac{a_i^k}{b_i^k + a_i^k}, \quad (11)$$

if  $w_i \geq 0.5$  then  $T(x_i) = 1$  (bone) else  $T(x_i) = 0$  (BGK).

The SBA atlas fusion scheme is then combined with spatially varying STAPLE approach via incorporation of the signed Euclidean distance map ( $O_{ij}$ ) into the conditional probability [Eq. (11)] as a weighting factor

$$a_i^k \equiv f(T(x_i) = 1) \prod_{j:D_{ij}=1} O_{ji} \times p_{ji}^k \prod_{j:D_{ij}=0} \frac{1}{|O_{ji}|+1} \times (1 - p_{ji}^k),$$

$$b_i^k \equiv f(T(x_i) = 0) \prod_{j:D_{ij}=0} |O_{ji}| \times q_{ji}^k \prod_{j:D_{ij}=0} \frac{1}{O_{ji}+1} \times (1 - q_{ji}^k). \quad (12)$$

These equations differ from the original spatially varying STAPLE algorithm since they are weighted with the signed distance map of bone label. The *Maximization-step* computes the performance level parameters that maximize the conditional expectation of the complete data log likelihood function

$$p_{ji}^{k+1} = \frac{\sum_{m \in R_i} w_m^k D_{jm}}{\sum_{m \in R_i} w_m^k},$$

$$q_{ji}^{k+1} = \frac{\sum_{m \in R_i} (1 - w_m^k)(1 - D_{jm})}{\sum_{m \in R_i} (1 - w_m^k)}. \quad (13)$$

By employing the spatially varying or L-STAPLE approach, a new parameter is introduced ( $R$ ), which represents the window size centered at the target voxel ( $i$ ) for local measurement of the atlas performance.

### 2.B.5. Local shape comparison (L-Shp)

The methods described so far tend to improve the quality of the segmentation outcome through iterative refinement of the atlas selection task. The iterative nature of these methods makes them computationally demanding, particularly when implemented locally (voxelwise). To address this issue, a robust and computationally efficient similarity measure is proposed based on local shape comparison (L-Shp). In this approach, the average distance map of the SBA scheme is computed through voxelwise weighting of the atlases according to their similarity to the others

$$D_{\text{ave}}(x_i) = \frac{1}{A} \sum_j (aw_{ji})^{-\beta} O_{ji}. \quad (14)$$

To calculate the weighting factor  $(aw_{ji})^{-\beta}$ , the distance map of the bone label should be created for all atlas images. Thereafter, the weight for atlas ( $j$ ) at voxel ( $i$ ) is obtained

from

$$aw_j^2(x_i) = \frac{1}{A} \sum_s \sum_{i \in \text{patch}_i} (O_{ji} - O_{si})^2 \frac{hO_{ji} + hO_{si}}{2},$$

$$hO_j = \frac{H(O_j)}{\sum_{\text{patch}_i} H(O_j)}, \quad (15)$$

where  $H(\cdot)$  denotes the Heaviside step function.<sup>33</sup> In a similar manner to L-STAPLE, a local window centered at the target voxel ( $\text{patch}_i$ ) is defined to calculate the local weights. This procedure implicitly defines the weights based on MV since the value of weights not only depends on the shape similarity but also on the number of similar atlases in the dataset.

## 2.C. Parameter optimization

The methods described in Secs. 2.B.2–2.B.5 contain free parameters that need to be optimized to reach optimal performance. The atlas selection threshold in the G-SIMPLE method plays a crucial role since a conservative threshold will discard fewer atlases at each iteration, which results in slow convergence rate. On the other hand, fast convergence can be achieved by choosing a strict threshold at the risk of unjustly excluding atlases. As mentioned earlier, the atlas selection threshold ( $\theta^k$ ) was determined as a function of the mean and standard deviation of the performance of all atlases  $\theta^k = \bar{\theta}^k - c\sigma_k$ . The best results were achieved when the free parameter ( $c$ ) was set as an increasing function of the number of selected atlases (ranging from 0.5 to 1.8) and updated after every second iteration. On average, 12 iterations were used for each data set while in the 3 first iterations no thresholding was applied.

For methods tending to refine atlas selection locally, the most influential parameter is the similarity window size ( $R$  for L-STAPLE and  $\text{patch}$  for L-Shp). These free parameters were optimized through a leave-one-out cross-validation (LOOCV) scheme. In the first step, for each individual patient, all the remaining ( $N - 1$ ) MR images were deformably registered to the target image. Therefore, for each subject, the remaining 20 patients act as atlas data set. The registration process was performed between in-phase MR images of the target and atlas data set. Thereafter, the obtained transformation matrices were used to transform the corresponding atlas CT images to the spatial coordinate of the target MR image. The registration process was carried out through a combination of affine and nonrigid alignment based on the advanced Mattes mutual information<sup>34</sup> implemented in the Elastix package<sup>35</sup> as described in previous studies.<sup>36,37</sup> The following parameters were adopted: interpolate: B-spline, 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 sizes of  $R$  and  $\text{patch}$  windows were optimized by varying the window size from 4 to 30 mm. The resulting bone segmentation was then evaluated

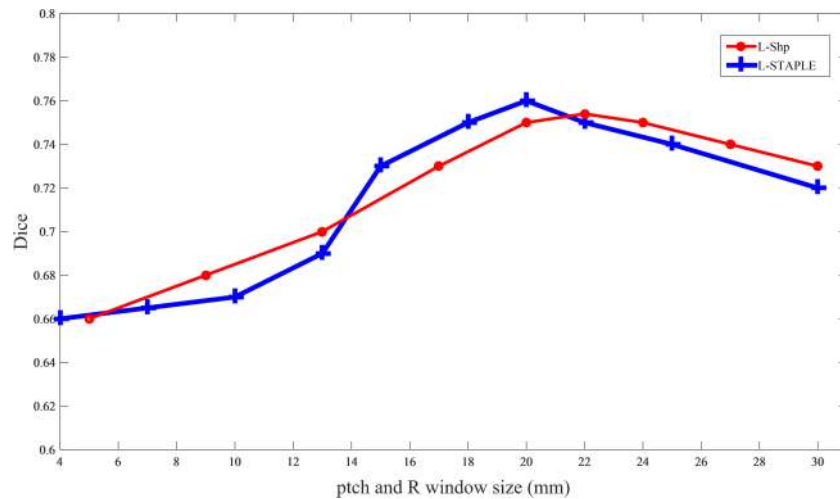


Fig. 2. Optimization of  $R$  window size for local atlas performance assessment in L-STAPLE method (top) and patch window size for local shape similarity assessment in the L-Shp method (bottom).

in terms of Dice similarity coefficient. This procedure was repeated for the whole dataset and the optimal sizes of  $R$  and patch windows were chosen for the rest of the study (Fig. 2).

## 2.D. Quantitative evaluation

Bone segmentation was carried out for the 21 clinical studies on a PC equipped with Intel Xeon CPU (2.3 GHz) running MATLAB (The MathWorks, Inc., Natick, MA). For each patient, the segmentation was repeated five times using the different methods, namely SBA, G-STAPLE, G-SIMPLE, L-STAPLE, L-Shp. Moreover, MV was considered as a conventional atlas fusion scheme. The assessment of the accuracy of the extracted bones was performed through comparison with a bone segmented from the corresponding reference CT images. Bone segmentation from CT images was performed by applying an intensity threshold of 140 HUs. The validation of bone segmentation was reported using seven volume/distance-based metrics: Dice similarity,<sup>38</sup> relative volume difference (RVD),<sup>39</sup> Jaccard similarity (JC),<sup>40</sup> sensitivity (S),<sup>41</sup> mean absolute surface distance (MASD),<sup>42</sup> Hausdorff distance (HD),<sup>43</sup> and distance error (DE),<sup>44</sup>

$$\text{Dice}(\text{Ref}B, T) = \frac{2|\text{Ref}B \cap T|}{|\text{Ref}B| + |T|},$$

$$\text{RVD}(\text{Ref}B, T) = 100 \times \frac{|T| - |\text{Ref}B|}{|\text{Ref}B|},$$

$$\text{JC}(\text{Ref}B, T) = \frac{|\text{Ref}B \cap T|}{|\text{Ref}B \cup T|},$$

$$S(\text{Ref}B, T) = \frac{|\text{Ref}B \cap T|}{|T|},$$

$$\text{MASD}(\text{Ref}B, T) = \frac{d_{\text{ave}}(S_{\text{Ref}B}, S_T) + d_{\text{ave}}(S_T, S_{\text{Ref}B})}{2},$$

$$\text{DE}(\text{Ref}B, T) = \frac{1}{\text{bp}} \sum_{p=1}^{\text{bp}} \text{mindist}(\text{Ref}B_p, T_p),$$

$$\text{HD}(\text{Ref}B, T) = \max_{\text{Ref}B} \{ \min_T \{ d(\text{Ref}B, T) \} \}, \quad (16)$$

where  $\text{Ref}B$  is the bone segmented from the reference CT image and  $T$  denotes the extracted bone from the target MRI.  $d_{\text{ave}}(S_{\text{Ref}B}, S_T)$  is the average direct surface distance from all points on the reference bone surface  $S_{\text{Ref}B}$  to the segmented bone surface  $S_T$ . The distance error is equal to the minimum distance from each boundary point of the source

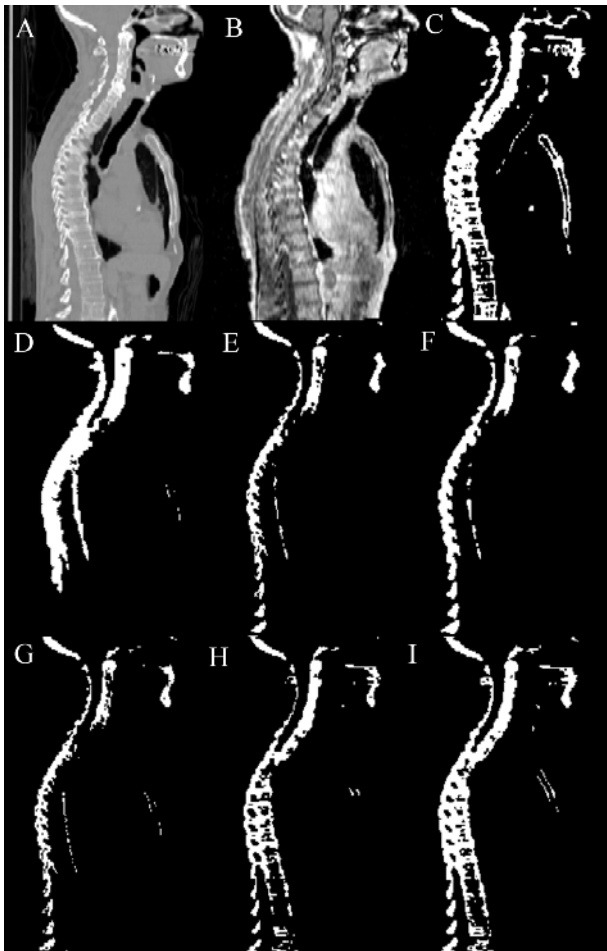


Fig. 3. Representative sagittal slice of whole-body bone segmentation using the different methods. (A) Reference CT, (B) target in-phase MR image, (C) reference bone map obtained from CT shown in (A), (D) SBA, (E) MV, (F) G-STAPLE, (G) G-SIMPLE, (H) L-Shp, and (I) L-STAPLE.



TABLE I. Comparison of bone segmentation accuracy (mean  $\pm$  SD) between the different approaches using the selected evaluation metrics, including the Dice, relative volume distance (RVD), Jaccard similarity (JC), sensitivity (S), mean absolute surface distance (MASD), Hausdorff distance (HD), distance error (DE).

	SBA	MV	G-STAPLE	G-SIMPLE	L-STAPLE	L-Shp
Dice	0.62 $\pm$ 0.07	0.63 $\pm$ 0.08	0.66 $\pm$ 0.06	0.66 $\pm$ 0.06	0.76 $\pm$ 0.05	0.75 $\pm$ 0.05
RVD (%)	-31.8 $\pm$ 9.1	-31.0 $\pm$ 9.2	-30.0 $\pm$ 8.2	-29.8 $\pm$ 8.4	-25.5 $\pm$ 6.5	-25.2 $\pm$ 6.9
JC	0.38 $\pm$ 0.06	0.39 $\pm$ 0.06	0.41 $\pm$ 0.05	0.40 $\pm$ 0.06	0.47 $\pm$ 0.05	0.46 $\pm$ 0.06
S	0.42 $\pm$ 0.10	0.42 $\pm$ 0.11	0.44 $\pm$ 0.08	0.45 $\pm$ 0.09	0.55 $\pm$ 0.07	0.54 $\pm$ 0.08
MASD (mm)	6.94 $\pm$ 3.10	6.51 $\pm$ 2.82	4.46 $\pm$ 2.15	5.56 $\pm$ 2.21	3.24 $\pm$ 1.95	3.17 $\pm$ 2.01
HD (mm)	14.5 $\pm$ 4.70	13.4 $\pm$ 3.95	11.6 $\pm$ 3.64	11.3 $\pm$ 3.70	07.3 $\pm$ 3.10	07.2 $\pm$ 3.22
DE (mm)	02.5 $\pm$ 1.70	02.2 $\pm$ 1.61	02.0 $\pm$ 1.55	02.1 $\pm$ 1.61	01.3 $\pm$ 1.33	01.2 $\pm$ 1.40

region ( $RefB_p$ ) to the entire set of points of the target region<sup>45</sup> averaged across the  $T_p$  boundary points. The Hausdorff distance measures the maximum distance one would need to move the boundaries of the source region ( $RefB$ ) to completely cover the target region ( $T$ ).

To assess the sensitivity of the different methods to the number of input atlases, the accuracy of bone segmentation and computation time were also evaluated for varying number of input atlases. In addition to bone extraction accuracy, fragmentation of bony structures was evaluated by computing the number of connected regions in the bone label maps. To this end, face-edge-vertex connectivity (two pixels are considered connected provided they share at least one vertex in the pixel grid) was employed to calculate the number of fragments in the target label maps compared to the reference obtained from CT images.

We further extended our validation through quantitative analysis of PET images corrected for attenuation using the bones segmented from MRI (PET-MRAC). To this end, the segmented bone from the target MRI was superimposed onto the MRI-guided 3-class attenuation map implemented on the Philips Ingenuity TF PET/MR scanner. The latter is obtained by segmenting MR images into the surrounding air, lung, and soft-tissue followed by assigning a predefined attenuation coefficient to each tissue class (air:  $0 \text{ cm}^{-1}$ , lung:  $0.022 \text{ cm}^{-1}$ , soft-tissue:  $0.098 \text{ cm}^{-1}$ ).<sup>23,36,46</sup> The 3-class attenuation maps

with inserted bony structures from reference CT images were taken as the reference in PET attenuation correction analysis (PET-CT). The described segmentation methods generate target bone label maps in binary format. However, a bone label map with continuous linear attenuation coefficients is desired for PET attenuation correction. To this end, we generated continuous valued bone map for each method. The described methods generate the final bone label map through weighted average of the atlas' votes. The obtained weights were multiplied by the values of the original atlas CTs in HUs (instead of binary votes) to recalculate continuous valued bone maps.

PET image reconstruction was performed by means of the e7 tool (Siemens Healthcare, Knoxville, TN)<sup>47</sup> using the ordinary Poisson ordered subset-expectation maximization (OP-OSEM) iterative reconstruction algorithm and default parameters (four iterations, eight subsets, and a postprocessing Gaussian kernel with a FWHM of 5 mm). The differences between bone segmentation techniques were quantified in terms of change in the standardized uptake value (SUV). The SUVs were calculated by dividing the activity concentration by the injected activity divided by body weight. The impact of whole-body bone segmentation on PET attenuation correction accuracy was assessed through calculation of the relative mean SUV error [Eq. (17)] and relative mean absolute SUV error [Eq. (18)] in bony structures (voxelwise evaluation)

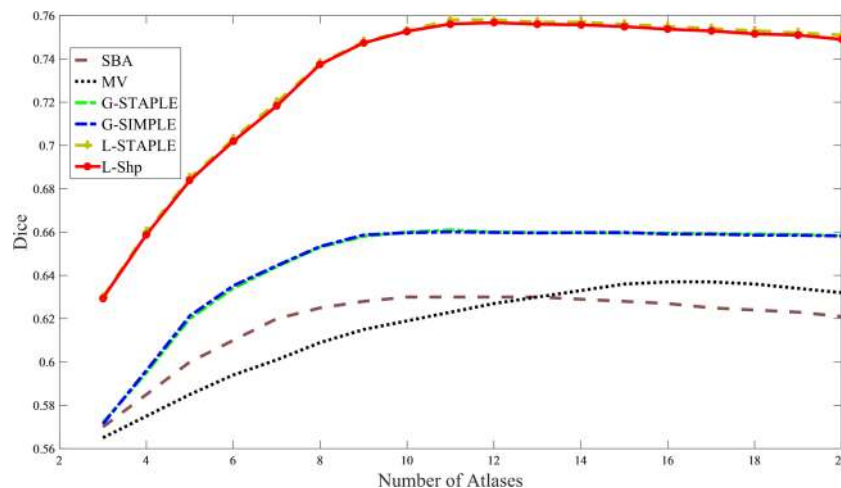


Fig. 4. Performance evaluation of the different segmentation methods based on the Dice metric as a function of the number of input atlases.

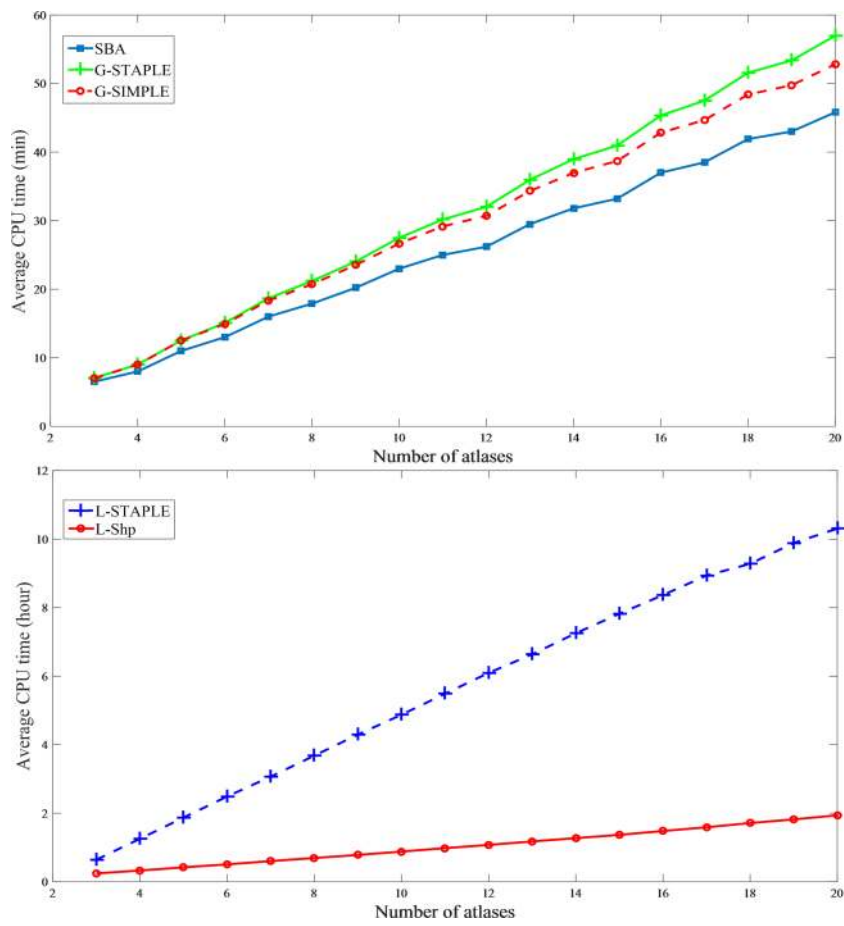


FIG. 5. Average computation time for performing whole-body bone segmentation (top) for SBA, G-STAPLE, and G-SIMPLE methods (minute). (Bottom) for L-STAPLE and L-Shp methods (hour) measured at varying number of input atlases.

and lesions in/near bony structures (region of interest-based evaluation). The relative errors of SUV estimates from PET-MRAC images are calculated relative to the SUV measured on PET-CT images used as reference. The relative absolute error [Eq. (18)] gives information about reconstruction errors and

deviations from the expected values, whereas the relative error [Eq. (17)] reflects the inherent bias in the methodology. The regions of interest (ROIs) were manually drawn on malignant lesions located inside bony structures or in close proximity such that their activity recovery can be affected by bones,

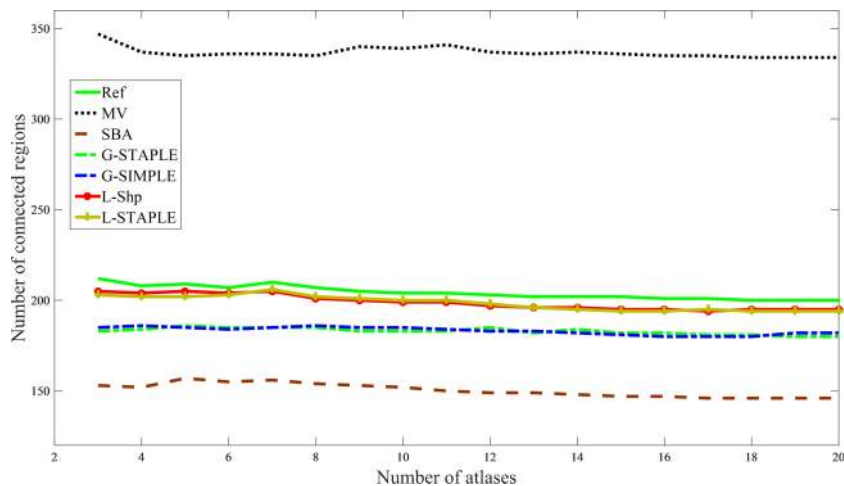


FIG. 6. Performance of the different segmentation methods in terms of number of connected regions vs the number of input atlases. Ref indicates the number of connected regions obtained from reference CT images.

$$\text{Relative error}(\%) = \frac{\text{MRAC}(\text{SUV}) - \text{Reference}(\text{SUV})}{\text{Reference}(\text{SUV})} \times 100\%, \quad (17)$$

$$\text{Relative absolute error}(\%) = \frac{\text{ABS}[\text{MRAC}(\text{SUV}) - \text{Reference}(\text{SUV})]}{\text{Reference}(\text{SUV})} \times 100\%. \quad (18)$$

In addition to the calculated SUV bias, linear regression analysis was performed on joint histograms of PET-MRAC and PET-CT images in bony structures for each segmentation method.

### 3. RESULTS

Figure 3 depicts a representative sagittal slice of the segmented whole-body bones from Dixon in-phase MRI together with reference CT image and bone label map. Visual inspection revealed superior bone identification when using locally regularized methods (L-Shp and L-STAPLE).

Table I summarizes the quantitative analysis of bone identification accuracy using seven standard volume/distance-based metrics. In agreement with Fig. 3, considerable bone detection enhancement was achieved using L-Shp and L-STAPLE methods compared to original SBA and MV methods based on the results achieved using 21 patients. Since the implemented methods exhibited different convergence rates in terms of number of input atlases, the representative segmented bones (Fig. 3) and the quantitative evaluation (Table I) show the best outcomes for each method. Figure 4 depicts the convergence trend of the different methods based on the Dice metric as the number of input atlases varies between 3 and 20.

In addition to the evaluation of the convergence rate, the required computation time for performing whole-body bone segmentation was assessed for different methods as a function of the number of input atlases (varying between 3 and 20). Figure 5 (top) shows the processing time (min) for the SBA and globally regularized methods where most of the computation time is taken by the calculation of the signed Euclidean distance maps on bone labels. Similarly, Fig. 5 (bottom) presents the processing time for the locally regularized methods where most of the processing time is taken by local iterative performance estimation (for L-STAPLE) and local similarity measurement (for L-Shp).

The fragmentation of contiguous structures is one of the major performance measures used in this study to evaluate the efficacy of different methods. The number of connected regions in the segmented images is plotted in Fig. 6, where those obtained from the target CTs are taken as reference. The number of connected regions achieved by MV is substantially higher than the other methods while application of local regularization makes the number of connected regions follow nicely the reference values.

Table II (top) summarizes the relative  $\text{SUV}_{\text{mean}}$  errors measured on PET images corrected for attenuation using the six different whole-body bone segmentation methods (PET-MRAC). The errors were measured voxelwise for all voxels in

bony structures as well as in ROIs within high uptake regions in and near bone tissue. Similarly, the  $\text{SUV}_{\text{mean}}$  relative absolute errors are shown in Table II (bottom) averaged over 21 patients.

Figure 7 depicts the joint histogram graphs obtained by plotting the correlation between PET images corrected for attenuation using the different bone segmentation methods (PET-MRAC) and the reference PET-CT image. The joint histograms are plotted by taking only voxels corresponding to bony structures into account. The L-STAPLE and L-Shp schemes yielded the best correlation for bone tissue  $y = 1.01x + 0.05$  with  $R^2 = 0.99$  and  $y = 1.02x + 0.08$  with  $R^2 = 0.99$ , respectively, while the SBA method deviated from the identity line ( $y = 0.87x + 0.10$  with  $R^2 = 0.97$ ). Figure 8 shows the error maps (relative bias) between reference PET-CT images and PET-MRAC images corrected for attenuation using the six different segmentation methods for one representative clinical study together with attenuation bias maps.

### 4. DISCUSSION

We assessed the performance of the SBA atlas fusion scheme for whole-body bone segmentation from Dixon MR images. Despite the promising performance reported in the literature,<sup>6,8</sup> the SBA method significantly underestimated bone identification in the context of whole-body imaging, even compared to MV atlas fusion scheme (Table I and Fig. 3). Due to the complexity of anatomical structures, large number of bone fragments, considerable anatomical variations across

TABLE II. Mean relative and absolute errors (bias) between  $\text{SUV}_{\text{mean}}$  estimated using the different PET-MRAC methods and PET-CT images in bony structures and lesions located in and near bone tissue.

Relative error	Bony structures	Lesions in/near bony structures
PET_SBA	$-09.3 \pm 7.2$	$-05.3 \pm 5.3$
PET_MV	$06.1 \pm 6.9$	$04.4 \pm 5.0$
PET_G-STAPLE	$-04.2 \pm 5.9$	$-03.3 \pm 4.2$
PET_G-SIMPLE	$-04.1 \pm 4.9$	$-03.5 \pm 4.0$
PET_L-STAPLE	$02.7 \pm 4.3$	$01.4 \pm 3.8$
PET_L-Shp	$02.8 \pm 3.9$	$01.5 \pm 3.5$
Relative absolute error	Bony structures	Lesions in/near bony structures
PET_SBA	$15.9 \pm 6.2$	$12.1 \pm 5.0$
PET_MV	$12.3 \pm 6.0$	$10.5 \pm 4.8$
PET_G-STAPLE	$09.8 \pm 3.5$	$08.8 \pm 3.2$
PET_G-SIMPLE	$09.6 \pm 3.2$	$08.3 \pm 3.0$
PET_L-STAPLE	$06.1 \pm 2.9$	$05.6 \pm 2.8$
PET_L-Shp	$05.9 \pm 2.5$	$05.5 \pm 2.7$

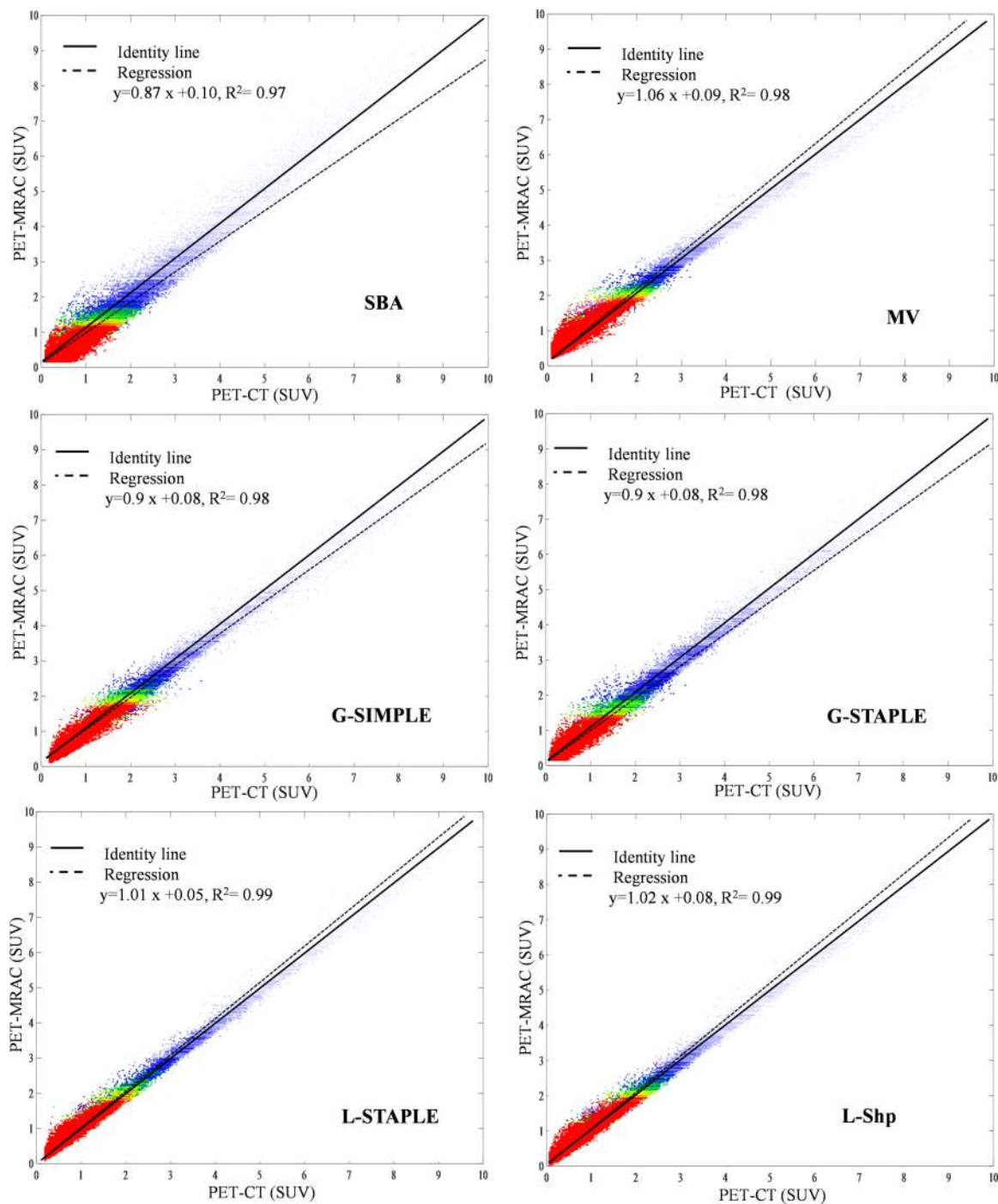


FIG. 7. Joint histograms correlating tracer uptake values in bony structures from PET-MRAC images using the different whole-body bone segmentation methods with CT-based attenuation correction (PET-CT).

patients, and greater likelihood of gross registration errors in whole-body imaging (compared to prostate segmentation), the performance of the SBA technique in whole-body bone identification is challenged.

The bone identification accuracy did not exceed 0.63 in terms of Dice metric when applying either the SBA or MV methods. SBA tended to overlook small fractions of bone tissues whereas the MV scheme led to dramatically increased factitious fragmentation (artificially disconnected regions in the resulting bone segmentation) (Figs. 3 and 6).

The MV technique operates solely on binary label maps and reaches a conclusion based on the consensus of the atlas' votes, while SBA implicitly ignores the consensus of the raters by incorporating the labels' distance map (Fig. 1). As such, in the SBA method, one atlas might have greater impact than a combination of several other atlases, which can result in underestimation/missing of a small fraction of the target segmentation. Fragmentation of the continuous structures plays a minor role in the task of PET attenuation correction considering the processing steps applied to generate

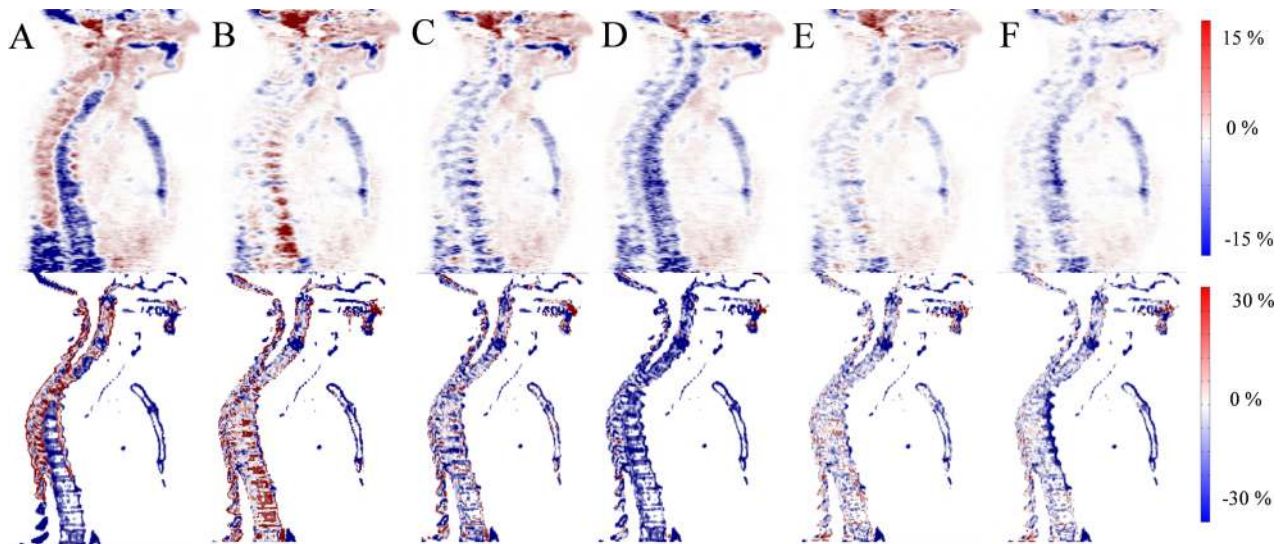


FIG. 8. SUV bias maps between PET-CT and PET-MRAC images (top row) and attenuation error maps (bottom row) between reference CT images and attenuation maps generated by (A) SBA, (B) MV, (C) G-STAPLE, (D) G-SIMPLE, (E) L-STAPLE, and (F) L-Shp methods.

attenuation maps (downsampling, Gaussian smoothing). However, this performance measure is an important factor in other applications requiring bone segmentation, such as vertebral fracture and morphometric analysis of the osteochondral elements. Despite suboptimal performance, the SBA method exhibited better convergence rate in terms of the number of input atlases, particularly compared to the MV method.

We proposed to combine the SBA scheme with an atlas selection/refinement technique to enhance the performance of the SBA method and at the same time sustain its superior convergence rate. To this end, the STAPLE method was chosen owing to its proven efficiency.<sup>25</sup> Moreover, the SIMPLE method exhibited similar performance when applied globally (G-SIMPLE) with relatively simpler implementation and lower computational complexity. However, the STAPLE method performed better when applied locally (L-STAPLE) compared to local SIMPLE, thus justifying to report only the results achieved by L-STAPLE. In addition, the proposed local shape-based similarity measure (L-Shp) substantially reduced the computation time by up to five times (Fig. 5) while enhancing the performance of the SBA method. The L-Shp method employed local atlas refinement scheme based on the measurement of the structure/shape similarity across the atlas label maps. In the L-STAPLE method, the atlas performance should be updated after each iteration and the whole algorithm takes on average 11 iterations to converge (in this study), while the L-Shp method computes the interatlas similarity matrix only once to generate weights to be inserted into the SBA atlas fusion scheme. In addition to the improved recognition rate when applying local regularization, the number of connected regions in the resulting label maps follows the reference true values, independent of the number of input atlases. Considering Fig. 4, combining local regularization methods (L-STAPLE and L-Shp) with the SBA scheme sustained the convergence rate of the original SBA to some extent.

The quantitative evaluation of the SUV bias demonstrated a similar trend to that of bone segmentation accuracy while PET

images showed less sensitivity to errors in bone identification owing to lower spatial resolution and higher level of noise. Both L-STAPLE and L-Shp methods resulted in less than 3% SUV mean relative and less than 6% mean absolute errors in bony structures due to superior bone identification. The joint histogram plot and regression analysis also support the voxelwise and ROI analysis of activity recovery. In this work, we used the 3-class MRI segmentation-based attenuation map for all methods followed by superimposition of only the bone tissue obtained using the different methods. The same 3-class attenuation map with bone tissue derived from reference CT was taken as reference attenuation map. However, a small SUV bias was observed in soft-tissue regions (even far from bony structures). This may be partially due to differences in scatter correction and the inherent smoothing during the generation of attenuation maps.

The results presented in this work demonstrate that the developed methods are promising and can potentially be used in the clinic. However, atlas-based segmentation techniques consist of some key steps where each step plays a critical role in the quality of the ultimate outcome. The first step is the creation of the atlas data set that should cover a reasonably wide variability of the human anatomy, with a sufficient number of samples for each category based for instance on gender, pathology, body weight and length, body mass index, sitting height/stature ratio,...etc. The richness of the atlas data set greatly influences the quality of outcomes since it increases the likelihood of finding an atlas similar to the target subject. The registration algorithm and optimization of their associated free parameters are also important determining factors. The final step is the atlas fusion task, which determines the segmentation output while trying to minimize the nonsystematic errors arising from the previous steps. In this light, atlas-based methods entail many steps and free parameters that need to be optimized prior to potential use in the clinic. These issues question the robustness of the atlas-based techniques to be utilized as the method

of choice. Therefore, despite the superior performance of atlas-based methods, as reported in this study and numerous previous works, the shift toward clinical implementation of these techniques is only gradual.

It is the intention of the authors to make the algorithms presented in this work available to the research community under the terms of the GNU Lesser General Public License version. This also applies to the CT/MRI atlases provided approval from our institution/IRB is secured.

## 5. CONCLUSION

In this work, we evaluated the performance of the SBA atlas fusion scheme in the context of whole-body bone segmentation from MR images. The SBA technique exhibited suboptimal performance compared to the conventional MV atlas fusion scheme. To enhance the performance of the SBA method, the technique was combined with STAPLE and SIMPLE atlas selection (refinement) methods. Local (voxelwise) application of the STAPLE method led to a dramatic improvement of bone fragmentation and recognition accuracy despite the high computation time. Moreover, a shape comparison based atlas weighting scheme (L-Shp) was proposed, which achieved similar results to L-STAPLE while considerably diminishing the computational burden. The quantitative SUV analysis revealed good correlation between PET images corrected for attenuation using both L-STAPLE and L-Shp methods and the corresponding reference CT images. The technique is promising for potential application in MRI-guided attenuation correction in whole-body PET/MRI.

## ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under Grant No. SNFN 31003A-149957 and the Swiss Cancer Research Foundation under Grant No. KFS-3855-02-2016.

## CONFLICT OF INTEREST DISCLOSURE

The authors have no COI to report.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: [habib.zaidi@hcuge.ch](mailto:habib.zaidi@hcuge.ch); Telephone: +41 22 372 7258; Fax: +41 22 372 7169.

<sup>1</sup> T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage* **21**(4), 1428–1442 (2004).

<sup>2</sup> X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imaging* **28**(8), 1266–1277 (2009).

<sup>3</sup> T. Rohlfing and C. R. Maurer, "Multi-classifier framework for atlas-based image segmentation," *Pattern Recognit. Lett.* **26**(13), 2070–2079 (2005).

<sup>4</sup> J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(3), 226–239 (1998).

<sup>5</sup> P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage* **46**(3), 726–738 (2009).

<sup>6</sup> T. Rohlfing and C. R. Maurer, Jr., "Shape-based averaging for combination of multiple segmentations," in *Medical Image Computing and Computer-*

*Assisted Intervention—MICCAI 2005* (Springer, Berlin Heidelberg, 2005), pp. 838–845.

<sup>7</sup> T. Rohlfing and C. R. Maurer, Jr., "Shape-based averaging," *IEEE Trans. Image Process.* **16**(1), 153–161 (2007).

<sup>8</sup> N. Robitaille and S. Duchesne, "Label fusion strategy selection," *Int. J. Biomed. Imaging* **2012**, 1–13.

<sup>9</sup> S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imaging* **23**(7), 903–921 (2004).

<sup>10</sup> J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: Theory and rationale for its development the international consortium for brain mapping (ICBM)," *NeuroImage* **2**(2), 89–101 (1995).

<sup>11</sup> A. Mehranian, H. Arabi, and H. Zaidi, "Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities," *Med. Phys.* **43**(3), 1130–1155 (2016).

<sup>12</sup> V. Keereman *et al.*, "MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences," *J. Nucl. Med.* **51**(5), 812–818 (2010).

<sup>13</sup> G. Delso *et al.*, "Clinical evaluation of zero echo time MRI for the segmentation of the skull," *J. Nucl. Med.* **56**(3), 417–422 (2015).

<sup>14</sup> N. Burgos *et al.*, "Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies," *IEEE Trans. Med. Imaging* **33**(12), 2332–2341 (2014).

<sup>15</sup> S. D. Wollenweber *et al.*, "Evaluation of an atlas-based PET head attenuation correction using PET/CT & MR patient data," *IEEE Trans. Nucl. Sci.* **60**(5), 3383–3390 (2013).

<sup>16</sup> D. Izquierdo-Garcia *et al.*, "An SPM8-based approach for attenuation correction combining segmentation and nonrigid template formation: Application to simultaneous PET/MR brain imaging," *J. Nucl. Med.* **55**(11), 1825–1830 (2014).

<sup>17</sup> A. Mehranian, H. Arabi, and H. Zaidi, "Quantitative analysis of MRI-guided attenuation correction techniques in time-of-flight brain PET/MRI," *NeuroImage* **130**, 123–133 (2016).

<sup>18</sup> H. Arabi and H. Zaidi, "Magnetic resonance imaging-guided attenuation correction in whole-body PET/MRI using a sorted atlas approach," *Med. Image Anal.* **31**, 1–15 (2016).

<sup>19</sup> H. Arabi and H. Zaidi, "One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI," *Eur. J. Nucl. Med. Mol. Imaging* **43**, 2021–2035 (2016).

<sup>20</sup> M. Hofmann *et al.*, "MRI-based attenuation correction for whole-body PET/MRI: Quantitative evaluation of segmentation- and Atlas-based methods," *J. Nucl. Med.* **52**(9), 1392–1399 (2011).

<sup>21</sup> H. Arabi and H. Zaidi, "Comparison of atlas-based bone segmentation methods in whole-body PET/MRI," *IEEE Nuclear Science Symposium and Medical Imaging Conference, Seattle, USA, 2014*.

<sup>22</sup> T. R. Langerak, U. A. Van Der Heide, A. N. Kotte, M. A. Viergever, M. Van Vulpen, and J. P. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Trans. Med. Imaging* **29**(12), 2000–2008 (2010).

<sup>23</sup> H. Zaidi *et al.*, "Design and performance evaluation of a whole-body Ingenuity TF PET-MRI system," *Phys. Med. Biol.* **56**(10), 3091–3106 (2011).

<sup>24</sup> W. T. Dixon, "Simple proton spectroscopic imaging," *Radiology* **153**(1), 189–194 (1984).

<sup>25</sup> J. M. Lotjonen *et al.*, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage* **49**(3), 2352–2365 (2010).

<sup>26</sup> Y. Zhuge, J. K. Udupa, J. Liu, and P. K. Saha, "Image background inhomogeneity correction in MRI via intensity standardization," *Comput. Med. Imaging Gr.* **33**(1), 7–16 (2009).

<sup>27</sup> L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging* **19**(2), 143–150 (2000).

<sup>28</sup> J. Weickert, *Anisotropic Diffusion in Image Processing* (Teubner Stuttgart, Teubner, 1998).

<sup>29</sup> N. J. Tustison *et al.*, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imaging* **29**(6), 1310–1320 (2010).

<sup>30</sup> M. J. McAuliffe *et al.*, "Medical image processing, analysis and visualization in clinical research," in *14th IEEE Symposium on Computer-Based Medical Systems* (CB MS, Bethesda, Maryland, 2001), pp. 381–386.

<sup>31</sup> A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. B* **1–38** (1977).

<sup>32</sup> A. J. Asman and B. A. Landman, "Formulating spatially varying performance in the statistical fusion framework," *IEEE Trans. Med. Imaging* **31**(6), 1326–1336 (2012).

- <sup>33</sup>D. Cremers, S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation," *Int. J. Comput. Vision* **69**(3), 335–351 (2006).
- <sup>34</sup>D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imaging* **22**(1), 120–128 (2003).
- <sup>35</sup>S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010).
- <sup>36</sup>H. Arabi, O. Rager, A. Alem, A. Varoquaux, M. Becker, and H. Zaidi, "Clinical assessment of MR-guided 3-class and 4-class attenuation correction in PET/MR," *Mol. Imaging Biol.* **17**(2), 264–276 (2015).
- <sup>37</sup>A. Akbarzadeh *et al.*, "Evaluation of whole-body MR to CT deformable image registration," *J. Appl. Clin. Med. Phys.* **14**(4), 238–253 (2013).
- <sup>38</sup>L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
- <sup>39</sup>T. Heimann *et al.*, "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Trans. Med. Imaging* **28**(8), 1251–1265 (2009).
- <sup>40</sup>D. L. Collins and J. C. Pruessner, "Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion," *NeuroImage* **52**(4), 1355–1366 (2010).
- <sup>41</sup>Y. Xia *et al.*, "Automated bone segmentation from large field of view 3D MR images of the hip joint," *Phys. Med. Biol.* **58**(20), 7375–7390 (2013).
- <sup>42</sup>G. Gerig, M. Jomier, and M. Chakos, "Valmet: A new validation tool for assessing and improving 3D object segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Berlin Heidelberg, 2001), pp. 516–523.
- <sup>43</sup>W. R. Crum, O. Camara, and D. L. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imaging* **25**(11), 1451–1461 (2006).
- <sup>44</sup>A. Klein *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage* **46**(3), 786–802 (2009).
- <sup>45</sup>J. Zaknun *et al.*, "Comparative analysis of MR imaging, ictal SPECT and EEG in temporal lobe epilepsy: A prospective IAEA multi-center study," *Eur. J. Nucl. Med. Mol. Imaging* **35**(1), 107–115 (2008).
- <sup>46</sup>V. Schulz *et al.*, "Automatic, three-segment, MR-based attenuation correction for whole-body PET/MR data," *Eur. J. Nucl. Med. Mol. Imaging* **38**(1), 138–152 (2011).
- <sup>47</sup>J. P. Carney, D. W. Townsend, V. Rappoport, and B. Bendriem, "Method for transforming CT images for attenuation correction in PET/CT imaging," *Med. Phys.* **33**(4), 976–983 (2006).