


Multi-atlas segmentation of the skeleton from whole-body MRI—Impact of iterative background masking

Jakub Ceranka^{1,2}  | Sabrina Verga^{1,3} | Maryna Kvasnytsia^{1,2} | Frédéric Lecouvet⁴ | Nicolas Michoux⁴ | Johan de Mey⁵ | Hubert Raeymaekers⁵ | Thierry Metens⁶ | Julie Absil⁶ | Jef Vandemeulebroucke^{1,2}

¹Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

²IMEC, Leuven, Belgium

³Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

⁴Cliniques universitaires Saint Luc, Institut de Recherche Expérimentale et Clinique (IREC), Université catholique de Louvain (UCLouvain), Brussels, Belgium

⁵Department of Radiology, Universitair Ziekenhuis Brussel, Brussels, Belgium

⁶Department of Radiology, ULB-Hôpital Erasme, Université Libre de Bruxelles (ULB), Brussels, Belgium

Correspondence

Jakub Ceranka, Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, Brussels, Belgium.
Email: jceranka@etrovub.be

Purpose: To improve multi-atlas segmentation of the skeleton from whole-body MRI. In particular, we study the effect of employing the atlas segmentations to iteratively mask tissues outside of the region of interest to improve the atlas alignment and subsequent segmentation.

Methods: An improved atlas registration scheme is proposed. Starting from a suitable initial alignment, the alignment is refined by introducing additional stages of deformable registration during which the image sampling is limited to the dilated atlas segmentation label mask. The performance of the method was demonstrated using leave-one-out cross-validation using atlases of 10 whole-body 3D-T₁ images of prostate cancer patients with bone metastases and healthy male volunteers, and compared to existing state of the art. Both registration accuracy and resulting segmentation quality, using four commonly used label fusion strategies, were evaluated.

Results: The proposed method showed significant improvement in registration and segmentation accuracy with respect to the state of the art for all validation criteria and label fusion strategies, resulting in a Dice coefficient of 0.887 (STEPS label fusion). The average Dice coefficient for the multi-atlas segmentation showed over 11% improvement with a decrease of false positive rate from 28.3% to 13.2%. For this application, repeated application of the background masking did not lead to significant improvement of the segmentation result.

Conclusions: A registration strategy, relying on the use of atlas segmentations as mask during image registration was proposed and evaluated for multi-atlas segmentation of whole-body MRI. The approach significantly improved registration and final segmentation accuracy and may be applicable to other structures of interest.

KEY WORDS

atlas-based segmentation, bone segmentation, image registration, whole-body MRI

1 | INTRODUCTION

Multi-modal whole-body MRI is emerging as a new imaging standard for many diseases requiring detection and monitoring of skeletal and soft tissue involvement in systemic cancers. Whole-body MRI has been successfully used in several cancers for the detection of bone, lymph nodes, and visceral metastases and their monitoring under treatment.¹⁻⁶ Among these, metastatic cancers to bone and hematologic malignancies, mainly multiple myeloma, benefit from the excellent study of the skeleton by whole-body MRI. However, the radiologist reading time of the whole-body MRI scans may be long due to the large amount of information provided by the images, and to the limited contrast with adjacent tissues.

The principle challenge for integrating whole-body MRI in clinical routine comes from the large amount of data to be reviewed. Computer-aided image analysis may alleviate the workflow. Such automatic algorithms could ultimately facilitate the process of reading whole-body scans by reducing the reading time and improving the diagnostic accuracy of whole-body MRI.

Segmentation of anatomical tissue of interest is crucial for quantitative analysis of whole-body images. In metastatic bone disease (MBD), staging and treatment response monitoring together with medical image processing tools, like skeleton segmentation, could assist in automatic tumor detection, as well as volumetric lesion burden assessment.⁷ Skeleton segmentation can be used to remove false positive findings outside of the region of interest, thus directing the reader's attention directly to the bones. Additionally, in combination with spatially pre-aligned “functional” whole-body MRI sequences such as diffusion-weighted (DWI) whole-body MRI⁸ and apparent diffusion coefficient (ADC) maps, it enables focused, efficient, multi-parametric quantitative evaluation of the bone.

The segmentation of the skeleton from a mono-modal MRI anatomical sequence is a challenging process. The T₁-weighted MR images represent superior resolution and signal-to-noise ratio compared to “functional” modalities; however, osseous tissue is not easily distinguishable. The bone cortices do not generate signal and the visible part is mainly the bone marrow. Additionally, there is a significant intensity overlap between the values of the bones and extra-osseous fat tissue in T₁-weighted acquisition sequences. Therefore, in contrary to computed tomography (CT), a segmentation based only on the image intensity profile is not feasible.

Hedström⁹ proposed a skeleton segmentation method utilizing the contrast properties of whole-body multi-modal

MRI (water-fat separated images), by extracting regions with low bone signal, removing segmented lung and abdomen, and using various filtering techniques. The method resulted in 0.94 mean Dice coefficient over a dataset of 20 whole-body images; however, it requires multiple MR modalities with a specific MR parameter weighting.

An atlas-based segmentation approach could provide a viable solution applicable to both single and multi-modality skeleton segmentation problem. A classical atlas-guided approach treats segmentation as an image registration problem, where spatial correspondence is established between the atlas and target image. However, due to the complexity of the skeleton, large variations in patient anatomy (in terms of size, weight, and body composition) and differences in pose during scanning, spatial image registration of inter-patient whole-body images is a highly challenging task, limiting the accuracy of the segmentation algorithm.¹⁰

To date, few authors have reported the registration of whole-body MR inter-patients scans. Strand et al¹¹ proposed a sequential intersubject whole-body MRI registration method that relies on tissue-specific elasticity constraints. It requires multiple whole-body image modalities (i.e. Dixon water-fat imaging), used to automatically define tissue classes prior to the registration (i.e. water, fat, different bone regions⁹). The skeleton was registered with a region specific piecewise affine approach. The method performed well for inter-patient whole-body image alignment. Akbarzadeh et al¹² proposed a two-stage whole-body MRI to CT registration scheme, consisting of rigid and deformable registration. The scheme was later adapted by Arabi et al¹³ in a large label fusion comparison study for atlas-based skeleton segmentation.

In this study, we propose a novel multi-atlas skeletal segmentation approach based on registration with iterative tissue masking. Our approach consists of a multi-stage registration in which target patient and atlas are first roughly aligned using standard rigid and deformable registration. Next, background anatomy is progressively masked out using the label maps in the atlas, and registration on the target anatomical structure is refined further. Experiments are performed on whole-body MR acquisitions of patients with bone metastases and healthy volunteers.

2 | METHODS

2.1 | Atlas-based segmentation

Atlas-based segmentation is an automated segmentation technique which relies on spatial registration of the target image to be segmented with an atlas, a template intensity image for

which the segmentation (the atlas labels) are available. After alignment of the atlas template and the target image, the atlas labels are propagated to the target image. Often, multiple atlases are employed and target segmentations obtained from each atlas are combined using label fusion. So-called multi-atlas segmentation offers benefits in terms of robustness and accuracy.¹⁴

The proposed skeleton multi-atlas segmentation from whole-body MRI consists of multiple pairwise image registrations, performed between each atlas and the target image using iterative background masking, followed by a label fusion to obtain the final segmentation result.

2.2 | MRI and atlas skeleton label images

Anatomical whole-body 3D-T₁¹⁵ spin-echo images (Philips Ingenia 3T) were acquired as a routine follow-up examination of male patients with advanced prostate cancer with both focal and multi-focal bone metastases. The following parameters were used: echo time = 8 ms, repetition time = 382 ms, image station matrix size = 480 × 480, pixel spacing = 0.65 mm, slice thickness = 1.19 mm. The whole-body image composition from the four independent stations resulted in an image of a matrix size equal to 210 × 768 × 1612-1705. The same acquisition protocol was used for scanning healthy male volunteers. The study was performed as part of the Platform for Imaging in Clinical Research in Brussels (PICRIB), supported by the Brussels Region, and approved by the Institutional Ethics Board of the Cliniques Universitaires Saint-Luc, Universitair Ziekenhuis Brussels and ULB-Hôpital Erasme.

The skeleton atlases consisted of 10 whole-body images (5 male advanced prostate cancer patients and 5 male volunteers, mean age equal to 57 years ± 17.8 years). For each subject, the skeleton segmentation was performed manually

using the GrowCutEffect algorithm from Slicer¹⁶ and manually refined using the ITK-SNAP¹⁷ tool and mathematical morphology binary closing with a kernel radius equal to 2 voxels. The segmented bones were those most relevant for metastatic bone disease involvement and included the clavicle, vertebra from the second cervical up to sacrum, pelvis, and femur bones (see Figure 1). Both trabecular and cortical bone were included. On T₁-weighted MRI, bone metastases appear as a hypo-intense in comparison to healthy bone tissue and were included as the part of the skeleton mask.

2.3 | Data preprocessing

Prior to atlas registration, all whole-body images were corrected for the following geometric and intensity artifacts (see Figure 1):

- Inter-station spatial misalignment. Due to patient movement during scanning, separate anatomical station images are often misaligned at the station edge, mostly along anterior-posterior and right-left directions. Groupwise rigid image registration was applied to align all stations,¹⁸ using the average mutual information metric (AMI).^{19,20}
- Image noise and low-frequency bias field. A standard Insight Segmentation and Registration Toolkit²¹ (ITK) implementation of anisotropic diffusion filtering²² was applied to reduce the image noise while conserving the edge information. The default parameters available in the ITK implementation were applied: conductance = 4, iterations = 10, and time step = 0.01. Bias field was automatically reduced using the N4ITK nonparametric nonuniform intensity normalization algorithm.²³
- Inter-station intensity differences. Whole-body MRI stations do not represent comparable intensity profiles within

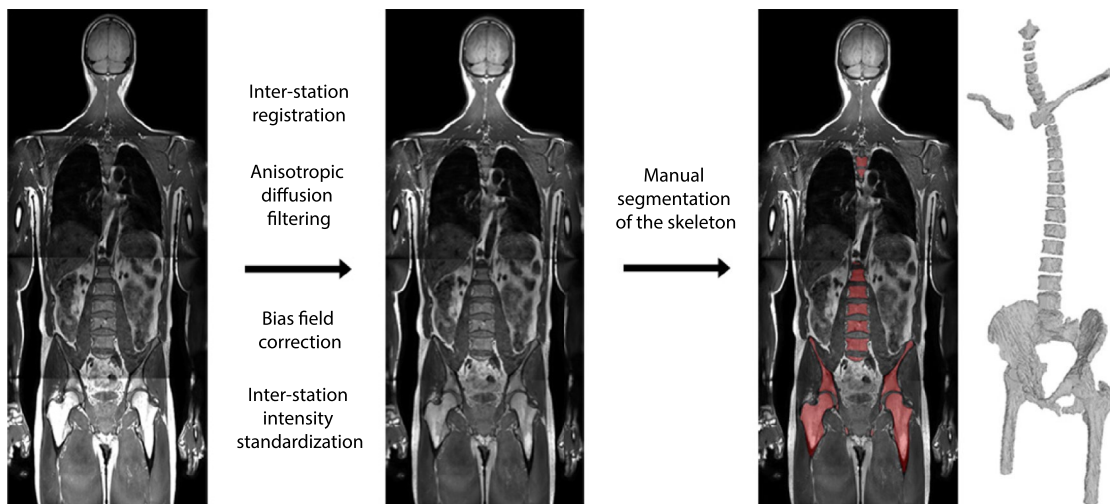


FIGURE 1 Raw whole-body T₁ anatomy image before (left) and after preprocessing steps (middle). Corrected image was used to obtain manual skeleton segmentation. Coronal view of a whole-body T₁ image in overlay with corresponding manual bone segmentation mask (red) and volume rendering from manual segmentation (right)

the same whole-body acquisition. Therefore, a linear intensity matching between 0th and 99th intensity percentile based on the station overlap region was applied sequentially prior to whole-body image composition, to align the whole-body intensity profile along the craniocaudal direction. The middle station was used as a reference intensity profile to minimize the influence of the cumulative intensity bias.

2.4 | Registration

The registration process of multi-atlas segmentation typically consists of pairwise multi-stage registrations between the 3D-T₁ target whole-body image t and each atlas template MR image a_n , $n = 1, \dots, N$, N being the number of considered atlases. It is commonly defined as the optimization problem over the parameters μ_n of the spatial transformation \mathcal{T} , guided by the cost function C

$$\hat{\mu}_n = \arg \min_{\mu_n} C(t(x), a_n(\mathcal{T}_{\mu_n}(x))). \quad (1)$$

In (1), the spatial coordinate x is taken from the overlapping region Ω , in which we assumed an intensity interpolation scheme for the discrete images t and a_n .

Arabi et al¹³ reported good segmentation accuracy using a registration method consisting of a rigid registration followed by a deformable B-Spline stage using a mutual information metric (see Table 1). For the ease of comparison, the proposed image registration protocol was implemented as an extension to the reported state-of-the-art registration

scheme^{12,13} by adding additional registration stages, that is, new registrations initialized from the previous solution.

Each additional stage consisted of a B-Spline registration using mutual information and a bending energy penalty (BEP).²⁴ In contrast to the previous stages, only regions in immediate vicinity of the region of interest to be segmented were considered. To this end, a dilated binary atlas-label image l_n was added as a registration mask. The degree of dilation (and the amount of neighboring tissue considered during registration) was decreased in each additional registration stage. The underlying assumption is that alignment of the target structures may be hampered by surrounding tissues, not of interest for the segmentation task, due to different relative position of these structures in the atlas and target patients. As the alignment after each additional iteration improves, a lower degree of dilation can be used, further focusing the registration on the region of interest.

The choice of dilation radius should provide sufficient overlap between the registration mask and the structures of interest, and take into account the image downsampling factor determined by the choice of multi-resolution image pyramid schedule. First, the initial dilation radius value was investigated, through quantitative evaluation of different dilation kernel sizes (range 5-100 mm). In the following experiments, the benefit of iterative masking with decreasing kernel sizes was investigated. To this end, the optimal dilation kernel was decreased by a factor two in subsequent iterations.

Registration was implemented in the elastix software package.²⁵ Contrary to (1), we redefined the transformation

TABLE 1 Overview of the registration parameters used in multi-atlas registration

Parameter	State-Of-The-Art ¹²		
	1 st stage	2 nd stage	Proposed additional stages
Transform	Rigid	B-Spline	Masked B-Spline
Metric	D_{MI} ^a	D_{MI}	D_{MI} , BEP ^b
Number of resolution levels	5	5	4
Image pyramid schedule	16 8 4 2 1	16 8 4 2 1	10 4 2 1
B-Spline grid spacing	-	32 16 8 4 2	4 2 1 1
Final B-Spline grid spacing (mm)	-	8	15
Number of histogram bins	32	32	32
Metric 1 weight	1	1	10
Metric 2 weight	-	-	1 10 10 25
Max iterations	4096 2048 1024 512 256	4096 4096 2048 1024 512	2000
Sampler	Random	Random	Random Sparse Mask
Number of samples	2000	5000	2000

^aMutual information similarity metric.

^bBending energy penalty regularizer.

from atlases a_n to the target t , allowing to use a fixed image mask which enables more efficient sampling strategies,

$$\hat{\mu}_n = \arg \min_{\mu_n} C(a_n(\mathbf{x}), t(\mathcal{T}_{\mu_n}(\mathbf{x}))), \quad (2)$$

where metric calculation was limited to the region of the dilated label image l_n^d .

The obtained set of transformation maps were used to deform each atlas label image l_n onto a target whole-body image t using the inverted transformation $\mathcal{T}_{\mu_n}^{-1}$. The inverse of a B-Spline transform cannot be computed analytically, as in case of a rigid transform, and has to be approximated using an iterative optimization process. Here, the inverse deformation field was found by iteratively minimizing the difference between the forward and inverse deformation field

$$\|\mathcal{T}_{\mu}^{-1}(\mathcal{T}_{\mu}(\mathbf{x})) - \mathbf{x}\|. \quad (3)$$

The approach was shown to provide accurate approximations.²⁶ The same optimization algorithm, multi-resolution pyramid schedule, and other registration parameters were used as for the initial deformable registration. The final segmentation is obtained through the application of a label fusion strategy (see Figure 2). All registration parameters are presented in Table 1.

2.5 | Label fusion strategies

Labels are combined using label fusion, aimed at maximizing the final skeleton segmentation accuracy. Arabi et al¹³ performed an extensive study on the influence of different

label fusion strategies on the accuracy of the atlas-based MRI skeleton segmentation. The local, voxel-wise weighted atlas fusion approach provided the best result achieving a Dice coefficient of 0.75-0.81, depending on the similarity measure used for ranking of the atlas voxels.

As the quality of registration may have an impact on the optimal strategy and settings of the label fusion, a few of common label fusion algorithms were reevaluated for current study.

- Majority Voting²⁷ (MV): all registered binary atlases are summed and normalized resulting in a probability map representing a probability for each voxel being a segmented structure. The probability map is later thresholded, allowing to obtain the final binary segmentation.
- Global Weighting²⁸ (GNCC): the pair-to-pair image registration quality in atlas segmentation may differ due to large differences between the target and atlas subject body size, weight, and pose during scanning and image quality. Therefore, for each pairwise registration, a similarity metric can be calculated between the reference image and the transformed atlas image, providing a quantitative value representing the registration accuracy for each image pair. First, the normalized cross-correlation metric (NCC) was calculated to rank the registration accuracy for each image pair. Later, only the first r best-ranked atlases were used in the MV scheme.
- Local Weighting²⁹ (LNCC): similarly to global weighting, the normalized cross-correlation is used to assess the similarity of the registered image with the reference. However, the calculation is performed locally. The similarity metric is calculated using a specific 3D image kernel of a given

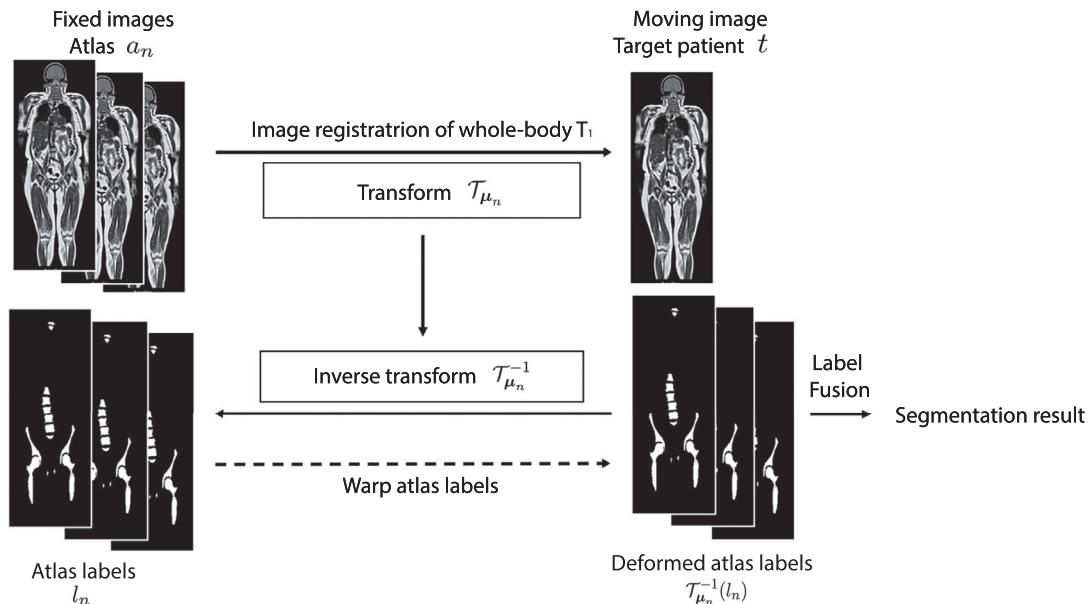


FIGURE 2 Schematic representation of the registration algorithm. Multiple pairwise registrations are performed between fixed atlas images a_n and a moving image t allowing for more efficient sampling using a dilated atlas-label fixed image mask l_n^d . Obtained transformations are inverted to warp atlas-label images on a common image space allowing for label fusion

radius k , from which the r highest ranked atlases are selected for local MV .

- **STEPS**: the Similarity and Truth Estimation for Propagated Segmentations (STEPS),³⁰ is an extension to the Simultaneous Truth and Performance Level Estimation (STAPLE)³¹ algorithm. It is an expectation-maximization algorithm, which computes the probabilistic estimate of the true segmentation from a collection of warped atlas label images, and a measure of the performance level represented by each segmentation. The optimal combination of the segmentations is obtained by facilitation of the local and global weighting of each deformed atlas label (using NCC metric), additionally incorporating a prior model of the spatial distribution of structures and a performance parameter bias improving the segmentation accuracy of structures much smaller than the background.

The label fusion strategies implementation was adapted from an open-source software toolbox NiftiSeg.³² The value of threshold for all label fusion strategies was set to 0.5, corresponding to a majority vote. For GNCC, LNCC, and STEPS, the influence of the number of highest ranked atlases ($r = 3-8$) was investigated. Additionally, for the locally weighted (LNCC) and STEPS label fusion methods, different kernel sizes were applied ($k = 3-33$ voxels).

2.6 | Evaluation criteria

Evaluation of the accuracy of obtained registration and segmentation results was performed using segmentation overlap and distance measures. The result of registration accuracy was calculated taking a mean of all atlas-target image pairs after registration. The result segmentation was compared with the ground truth manual skeleton segmentation using a leave-one-out cross-validation on the atlas of 10 whole-body images. The overlap measures included the Dice coefficient (DC) and false positive and false negative error volume fractions (FP, FN).³³

$$DC(V_R, V_S) = \frac{2|V_R \cap V_S|}{|V_R| + |V_S|} \quad (4)$$

$$FP(V_R, V_S) = \frac{|V_S \setminus V_R|}{|V_S|} \quad (5)$$

$$FN(V_R, V_S) = \frac{|V_R \setminus V_S|}{|V_R|} \quad (6)$$

where V_R is the skeleton segmentation binary ground truth image and V_S is the obtained segmentation image.

The distance criteria included the Hausdorff mean distance, the Hausdorff maximum distance, and SD . The Hausdorff distance (in mm) was computed from the Euclidean distance map of the ground truth manual segmentation and the surface of the segmentation obtained from atlas-based method, according to the formula:

$$H(V_R, V_S) = \max(h(V_R, V_S), h(V_S, V_R)), \quad (7)$$

where

$$h(V_R, V_S) = \max_{a \in V_R} \min_{b \in V_S} \|a - b\|. \quad (8)$$

Since not all of the data were normally distributed (Shapiro-Wilk normality test,³⁴ $P > .05$), nor equality of variance was observed (F -test,³⁵ $P > .05$), the Wilcoxon 2-tailed signed-rank test was used to check for statistical significance ($P = .05$) between the proposed method and the baseline method. The statistical difference between applied label fusion strategies was additionally investigated ($P = .05$).

3 | RESULTS

The proposed registration strategy based on the use of a dilated atlas mask was evaluated and compared to the state of the art through computation of the Dice coefficient between the atlas and target label images after registration. The influence of the size of the dilation mask radius was investigated for a range of 5-100 mm. The results of the Dice coefficient are presented in Table 2. The most accurate set of registrations was obtained

TABLE 2 Top: Influence of the size of the registration mask dilation radius (in mm) on registration accuracy

	Dilation of the registration mask in mm											
	SOTA	5	10	15	20	25	30	35	40	50	75	100
Registration DC	0.517	0.688	0.716	0.725	0.730	0.729	0.724	0.713	0.715	0.681	0.658	0.628
Segmentation DC	0.746	0.875	0.884	0.886	0.887	0.885	0.882	0.878	0.875	0.863	0.848	0.808
Mask dilation (mm)	DC	FP		FN		H_{mean}		H_{max}		H_{SD}		
20	0.887 ± 0.011	0.132 ± 0.025		0.093 ± 0.023		0.989 ± 0.189		32.23 ± 19.67		1.598 ± 0.890		
20 + 10	0.887 ± 0.010	0.131 ± 0.024		0.092 ± 0.021		1.013 ± 0.152		42.98 ± 20.83		1.774 ± 0.717		
20 + 10 + 5	0.885 ± 0.012	0.145 ± 0.025		0.080 ± 0.019		0.894 ± 0.125		26.04 ± 17.45		1.352 ± 0.649		

Notes: The mean Dice coefficient for all registration pairs and the segmentation Dice coefficient for STEPS-label fusion LOOCV ($r = 4, k = 5$) is given. Bottom: Segmentation accuracy ($\pm SD$) after additional stages with decreasing registration mask size. All results are given for STEPS-label fusion.

using the kernel dilation radius of 20 mm. For this specific application, the choice of the kernel size was not critical in the range 10-30 mm, all providing comparable results.

The detailed comparison of the Dice coefficient for the best performing kernel size (20 mm) values for all registration pairs are presented in Figure 3. The state-of-the-art methodology provided a mean Dice coefficient of 0.517 ± 0.159 for all registration atlas-target pairs. The proposed method improved registration accuracy on a vast majority of image pairs, providing a significantly better mean Dice coefficient equal to 0.730 ± 0.129 ($P < .001$). The proposed method can be characterized by higher robustness, providing less misregistrations across all possible target image to atlas image combinations. The majority of the reported Dice coefficients for different atlas pairs have improved; however, for some subject pairs the accuracy is still low (e.g., pair V_3 vs. P_3). After visual investigation of the registration result for this specific subject pair, the registration failed in the baseline stage, resulting in a large spatial deformations predominantly in the pelvic region.

The application of additional registration stages with decreasing mask dilation (mask with 10 mm and 10 mm + 5 mm dilation, respectively) did not further improve the registration mean Dice coefficient (0.733 ± 0.130 and 0.728 ± 0.129 , respectively, $P > .124$), see Figure 3.

In order to verify that the atlas registration accuracy improvement was caused by background masking and not due to increased sampling in the region of interest, the experiment was repeated using a full image sampler in the third registration stage, without using a sampling mask. The results showed no significant improvement ($P = .436$) in comparison to the state-of-the-art result (two-stage registration, STEPS). A similar experiment was conducted using the proposed three-stage registration without BEP regularizer, demonstrating that it is not significantly different from proposed methodology with BEP ($P = .241$, STEPS).

The choice of the optimal label fusion strategy with hyperparameters was additionally investigated for the proposed

dataset using state-of-the-art and proposed registration methodologies. The Dice coefficient results of STEPS label fusion strategy, significantly outperformed MV and GNCC label fusion strategies for the proposed registration scheme ($P < .001$). Label fusion hyperparameter optimization was performed to obtain the optimal value of parameters k (kernel size) and r (number of highest ranked atlases used). Figure 4 shows the accuracy of extracted skeleton using different kernel sizes and number of highest ranked atlases for different label fusion strategies. The best result for GNCC was achieved using $r = 7$ highest ranked atlases, irrespective of the chosen registration. The best results for LNCC and STEPS was achieved using a number of atlases equal to $r = 3$ and $r = 4$, respectively, and a kernel size $k = 5$ voxels for both registration approaches. The STEPS method proved to be least effected by the number of highest ranked atlases used. Additionally, the average Dice coefficient for varying number of atlases selected randomly (all combinations) from the entire dataset was investigated (see Figure 4, right), showing less DC variation, plateauing after six used atlases.

For the multi-atlas segmentation, using one additional registration stage, the mean Dice coefficient using STEPS label fusion technique was equal to 0.746 for the baseline method, whereas the proposed registration allowed for a mean 11% improvement in Dice coefficient (0.887, $P < .001$), using leave-one-out cross-validation. Additionally, all other validation criteria provided significantly better results (except of H_{max} , $P = .075$), showing 53% improvement for false positive and 114% for false negatives. Quantitative results for each validation criterion of the best performing label fusion strategy (STEPS) are presented in Table 3. The table with detailed performance of all of the label fusion strategies is included in the manuscript supplementary material (see Supporting Information Table S1).

An example of qualitative analysis for the state-of-the-art and proposed methods are depicted in Figures 5 and 6. Figure 5 shows the effect after one additional, masked registration

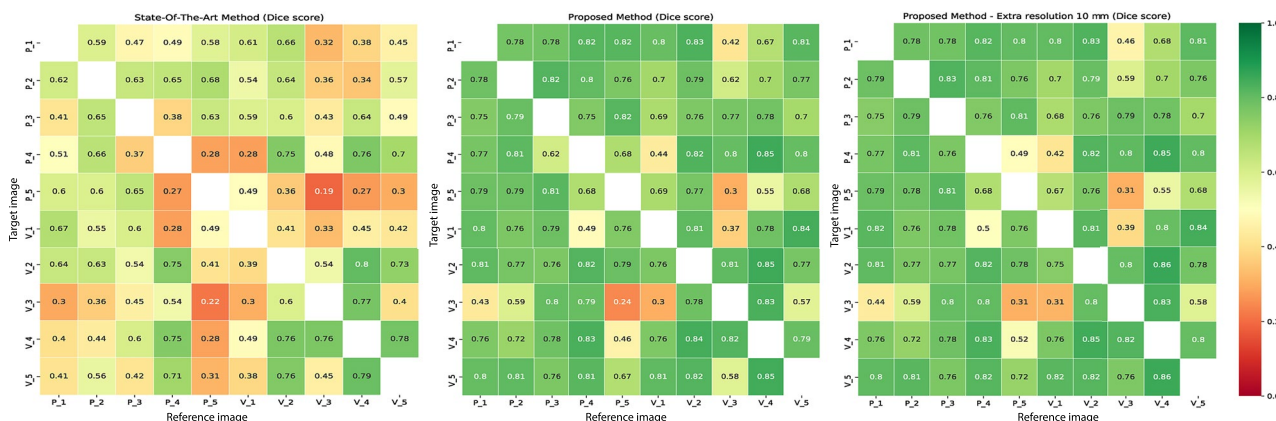


FIGURE 3 Atlas template registration confusion matrix results representing the Dice coefficient for the state-of-the-art (left), proposed (middle), and proposed with an extra resolution with a smaller mask size method for all registration combinations of patient and volunteer images

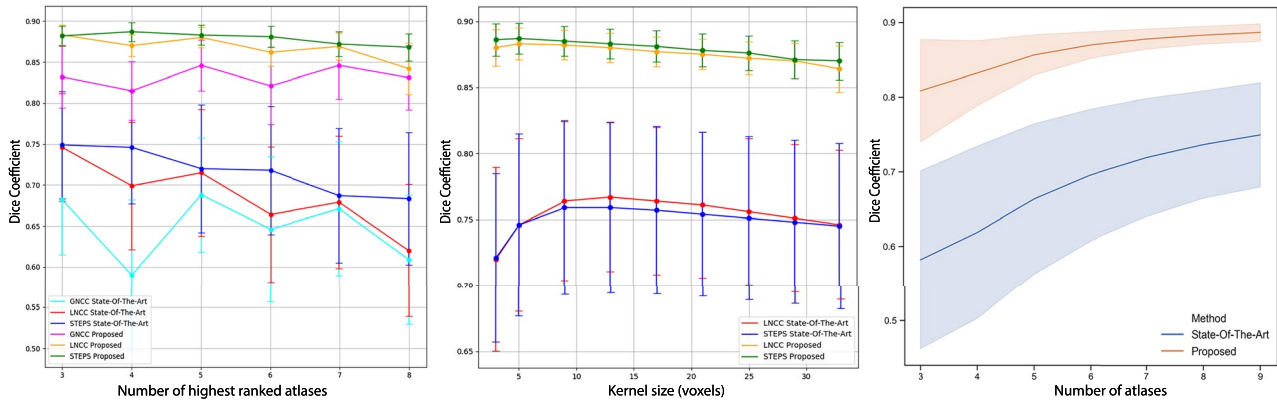


FIGURE 4 Left: Plot of Dice coefficient vs. the number of the highest ranked atlases parameter r for a fixed kernel size k equal to 5 voxels. STEPS label fusion method shows higher stability across a varying number of atlases used. Middle: Plot of Dice coefficient vs. size of the kernel parameter k , for a fixed r equal to 4. The influence of a kernel size and number of atlases has less influence when using the proposed registration methodology, showing higher robustness of the method. Right: Aggregated values of Dice coefficients: average with SD per number of atlases used for the best performing label fusion technique (STEPS) for the state-of-the-art and proposed registration methods

stage on the 3D reconstructed segmentation volumes and the Hausdorff distance between the ground truth and obtained result. A representative coronal slice of a whole-body T_1 in overlay with atlas-based segmentation, showing true positive, false positive, and false negative volume ratios, are presented in Figure 6. It is clearly visible that the false positive volume fraction has decreased (mostly in the intervertebral area) with a simultaneous increase of true positive volume fraction.

3.1 | Computation times

Processing was performed on a cluster node using Intel E5-2680v2 and 64 GB RAM. The registration procedure for one atlas took approximately 20 minutes (single-threaded execution), the total atlas registration time ($N=9$) was approximately 180 minutes. The execution of label fusion strategies took respectively: 1 minute for MV, 3 minutes for GNCC, and 30 minutes for LNCC and STEPS.

4 | DISCUSSION AND CONCLUSIONS

Skeletal segmentation from whole-body imaging may be of considerable benefit to longitudinal follow-up studies for patients with bone tumoral involvement, facilitating the assessment and tracking of changes in bone lesion distribution, total tumoral volume, and treatment-induced changes in ADC within segmented bone metastases. Due to its complex topology consisting of multiple structures with highly variable sizes, and large anatomical variations between different atlases, multi-atlas segmentation of the human skeleton remains a challenging task. The key step, having a large impact on the accuracy of the obtained segmentation, is the quality of spatial registration between the reference image and its atlas counterpart.

Arabi et al¹³ reported Dice coefficient of 0.78 ± 0.05 (LNCC), which is higher than the 0.75 ± 0.07 found here. However, the results cannot be compared directly due differences in the data. In particular, they used a higher number of atlases, which were obtained by intensity thresholding whole-body CT images, leading to detailed skeleton atlases including the ribs. Therefore, to provide a fair comparison between the methods, the state-of-the-art whole-body image registration was reimplemented and applied on the type of skeleton atlas used in this work.

We proposed an extension to a previously proposed registration scheme, which significantly improved both: the registration performance and obtained atlas segmentation quality without large increase in computational complexity. By adding the third stage to the registration scheme, the optimization is focused on the structures of interest (skeleton atlas label). The method provided higher stability with only a few low accuracy registration results (see Figure 3), which were primarily caused by large spatial deformation errors after the B-Spline deformable unmasked stage. The failure in the atlas pair V_3 and P_5 was probably caused by differences in femur bones positioning during scanning in V_3 ; and multiple focal metastases in upper pelvis with thin pelvic ilium in P_5 . The result could not be further improved by iterative background masking due to too large initial position differences.

The registration accuracy was visually assessed in three separate skeleton regions (spine, pelvis, and femur bone). The pelvic region showed low inter-patient variations in shape and pose, and was usually aligned with high accuracy providing a precise atlas-based segmentation. Femur bones showed high registration accuracy for most cases. However, in some subject pairs, large differences in positioning of the legs during MR acquisition resulted in failed registrations. In our study, the label fusion procedure was able to eliminate these case. Spine registration proved to be a challenging task,

TABLE 3 Segmentation evaluation criteria averaged over leave-one-out cross-validation results for the state-of-the-art and proposed registration strategies ($\pm SD$), STEPS label fusion

Label-fusion	Validation criterion	Registration		
		State-Of-The-Art	Proposed	<i>P</i> value
STEPS	<i>DC</i>	0.746 \pm 0.069	0.887 \pm 0.011	<.001
$r = 4^a$	<i>FP</i>	0.283 \pm 0.083	0.132 \pm 0.025	<.001
$k = 5^b$	<i>FN</i>	0.214 \pm 0.095	0.093 \pm 0.023	<.001
	H_{mean}	2.986 \pm 1.378	0.989 \pm 0.189	<.001
	H_{max}	49.75 \pm 19.47	32.23 \pm 19.67	.075
	H_{SD}	4.328 \pm 2.164	1.598 \pm 0.890	<.001

Note: Hausdorff distance criteria are presented in mm. *P* value between the proposed and state-of-the-art method is given.

^aNumber of highest ranked atlases.

^bSize of the kernel (voxels).

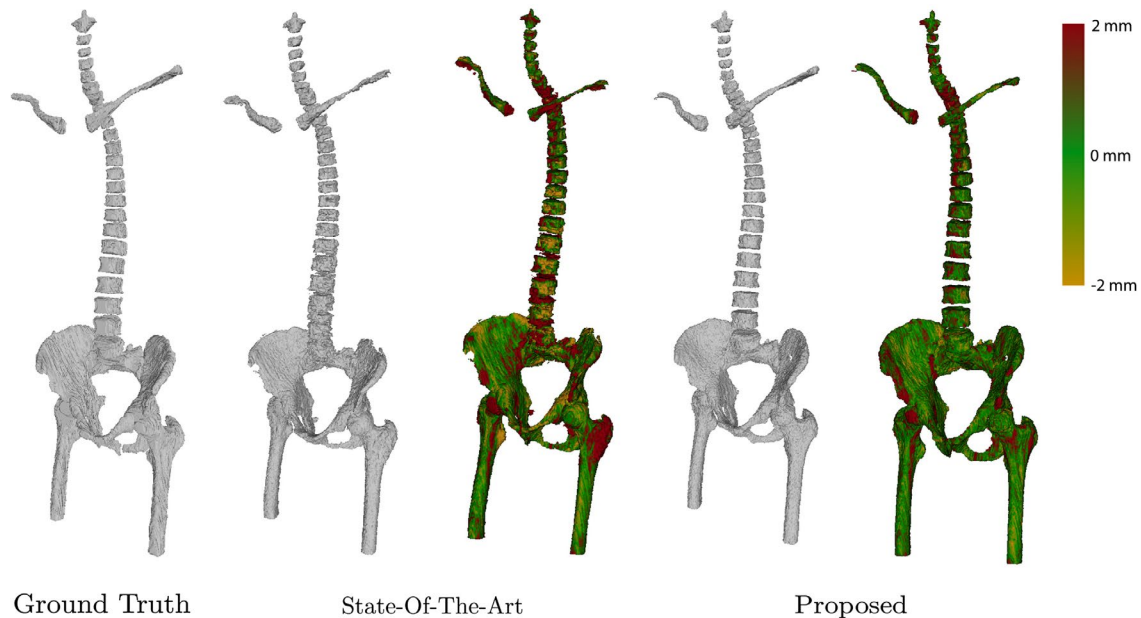


FIGURE 5 Comparison of the 3D volume rendering (gray) and Hausdorff distance between the ground truth manual skeleton segmentation, state-of-the-art method and proposed method with the best performing label fusion strategy (STEPS). Green indicates true positive voxels, red—false positive, and yellow—false negative. The proposed registration method visibly improves vertebral separation

with inter-patient spinal deformations and initial misalignment of vertebra along craniocaudal direction. The iterative background masking stage had the highest impact on the vertebral registration and segmentation accuracy.

Latter results lead us to believe the increase in registration accuracy is due to excluding the neighboring tissue from the registration. Skeletal structures may appear in different relative position with respect to surrounding organs from one patient to another. Jointly registering skeleton and neighboring tissues leads to a compromise in alignment of both types of structures. Registration with a background masking allows for partly independent deformations between the masked out region (organs, fat, muscle) and skeleton, compensating for differences in inter-patient posture, skeleton size, and body composition (see Figure 7).

A choice of the dilation radius in registration mask plays an important role in registration performance and should be individually optimized for specific atlas-based segmentation application. In the case of investigated dataset, the dilation of 20 mm gave the best registration accuracy in terms of mean Dice coefficient for all independent registration pairs. The further dilation of the registration mask lead to a worse result since less registration points were sampled from the direct proximity of the skeleton. Excessive dilation of the mask leads to further deterioration of the registration pairs mean Dice score, finally becoming equivalent to the state-of-the-art result.

For this study, using additional iterations of background masking did not lead to a significant improvement in registration mean Dice coefficient nor segmentation accuracy

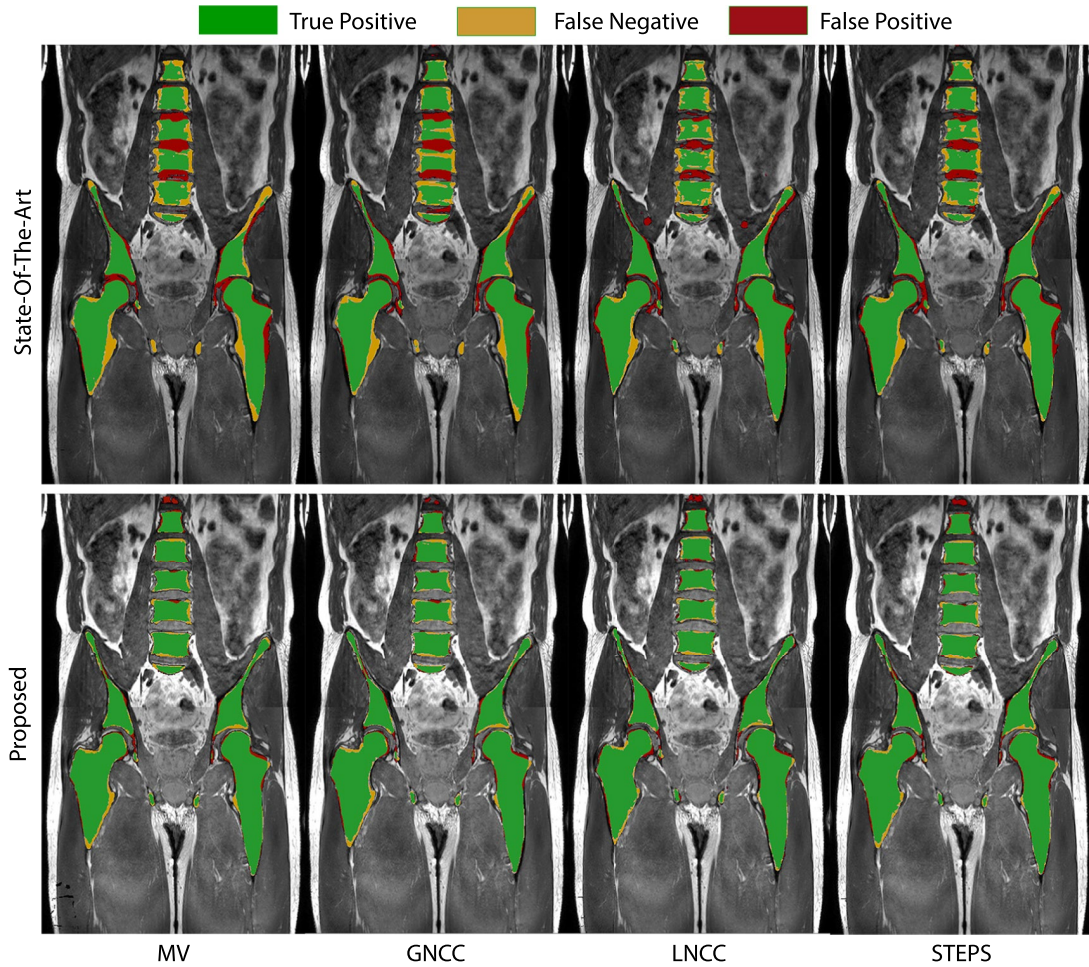


FIGURE 6 Coronal view of a whole-body T_1 image of subject V_2 in overlay with a skeleton segmentation result for different atlas registration and label fusion algorithms, indicating the true positive, false positive, and false negative segmentation fraction. Top row: State-of-the-art registration scheme. Bottom row: Proposed registration scheme with the best Dice coefficient of 0.894 (STEPS)

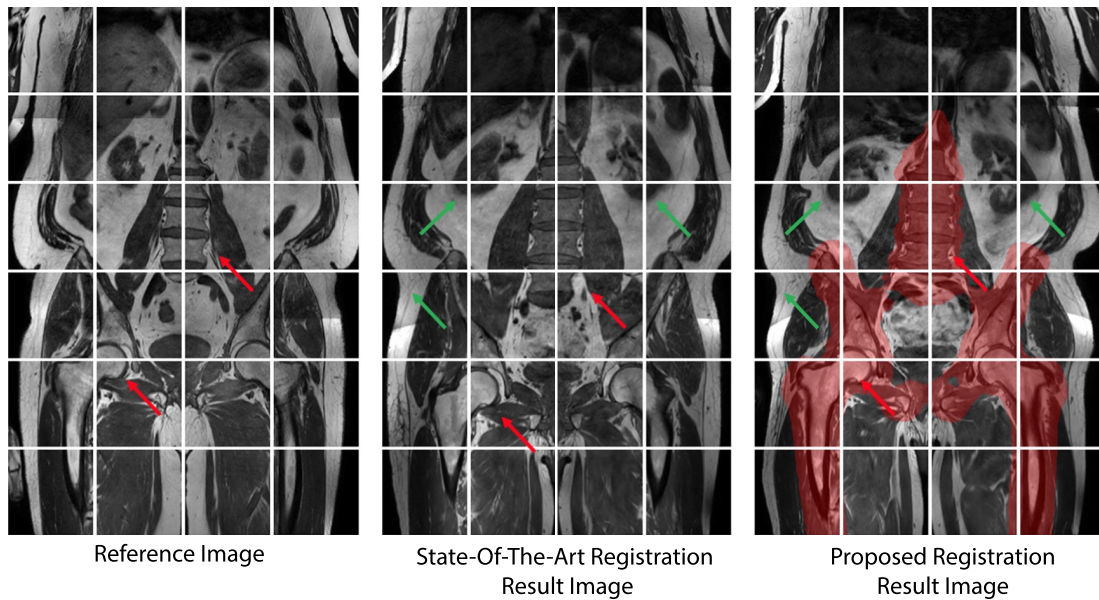


FIGURE 7 Example of a registration result between P_3 reference image (left) and warped target image V_3 for the state-of-the-art (middle) and proposed (right) methodologies with background masking (red volume). State-of-the-art registration led to an overall approximate alignment of bones and soft tissues (kidneys, fat, green arrows). The background masking procedure allowed for further improvement of the alignment of the skeletal structures (red arrows) without the negative effect of the surrounding soft tissue

in terms of Dice coefficient. Additional stages provided improvement for some segmentation evaluation metrics ($P > .05$) but the benefit was marginal and not deemed to outweigh the additional processing time (see Table 2). However, for other structures of interest, additional iterations with a decreased sampling mask dilation could provide a better segmentation accuracy.

Strand et al¹¹ proposed a piecewise affine registration of each of the nine independent body regions. The method, shows accurate whole-body MR inter-subject registration results, however, requires a series of additional preprocessing steps (multi-atlas body part segmentation, bone anatomical landmarks location). The approach is specifically tuned to whole-body image applications, potentially rendering it more robust. Background masking, however, is a simple and generic approach which can be applied to any multi-atlas segmentation task.

An increased number of atlases may improve the segmentation accuracy for the baseline method. The number of atlases used to build the skeleton atlas was, however, sufficient for the proposed method with Dice coefficient plateauing around 0.88, with only a marginal improvement between six and nine atlases (Figure 4). The behavior of different registration approaches has been additionally compared using four common label fusion strategies. STEPS showed the highest accuracy, provided a smooth result segmentation image (see Figure 5) and was stable across to the number of highest ranked atlases and the kernel size used (see Figure 4).

The atlas to reference image registration accuracy showed to have a key influence on the performance of the label fusion algorithms. It is clearly seen that the influence of atlas-fusion parameters r and k have less impact on the result when applied to an accurately registered atlas-label maps. Out of the two investigated label fusion parameters, the influence of the parameter r was more important for the performance of weighted label fusion strategies, showing higher accuracy variations across different values of r . The influence of the kernel size was found to have less impact on the segmentation accuracy (deviations of $\pm 1\%$ in Dice coefficient for the proposed registration scheme). A different choice of a label fusion strategy and voxel fusion similarity metric could possibly further improve the atlas-based skeleton segmentation algorithm accuracy. However, a broad investigation of the performance of the label fusion techniques was not the main scope of this work and has been already presented in the past.¹³ Our label fusion results are inline with the work of Arabi et al, showing superior performance of locally weighted fusion techniques (LNCC) and techniques using modification of a locally weighted approach (STEPS), over globally weighted techniques (GNCC) and straightforward majority voting for multi-atlas skeleton segmentation.

Composition of an atlas is an important factor in multi-atlas segmentation. Atlas should be represented by images showing

high statistical diversity, representing subjects of different size, body weight, and positioning during image acquisition. Such atlas, is less prone to large registration errors happening mutually for all atlas images, however, the registration between some atlas—target image pairs may fail due to large anatomical differences or present abnormalities. In the presented study, a registration between a volunteer (V_3)—patient (P_3) pair failed, resulting in a low Dice coefficient equal to 0.22. It was caused by large positioning differences of femur bones, anatomical differences in pelvis ilium thickness and several focal bone metastases present in the patient image. The low Dice was, however, compensated by label fusion techniques and should have a marginal influence on the segmentation result.

In conclusion, we have presented a modified registration protocol for atlas-based segmentation of the skeleton from whole-body MRI using background masking. The method outperformed existing state of the art, providing a robust solution for accurate whole-body bone segmentation.

ORCID

Jakub Ceranka  <http://orcid.org/0000-0002-0241-7737>

REFERENCES

1. Laurent V, Trausch G, Bruot O, Olivier P, Felblinger J, Régent D. Comparative study of two whole-body imaging techniques in the case of melanoma metastases: advantages of multi-contrast MRI examination including a diffusion-weighted sequence in comparison with PET-CT. *Eur J Radiol.* 2010;75:376–383.
2. Baur-Melnyk A, Buhmann S, Dürr H, Reiser M. Role of MRI for the diagnosis and prognosis of multiple myeloma. *Eur J Radiol.* 2005;55:56–63.
3. Weininger M, Lauterbach B, Knop S, et al. Whole-body MRI of multiple myeloma: comparison of different MRI sequences in assessment of different growth patterns. *Eur J Radiol.* 2009;69:339–345.
4. Lecouvet FE, El Mouedden J, Collette L, et al. Can whole-body magnetic resonance imaging with diffusion-weighted imaging replace Tc 99m bone scanning and computed tomography for single-step detection of metastases in patients with high-risk prostate cancer? *Eur Urol.* 2012;62:68–75.
5. Blackledge MD, Collins DJ, Tunariu N, et al. Assessment of treatment response by total tumor volume and global apparent diffusion coefficient using diffusion-weighted MRI in patients with metastatic bone disease: a feasibility study. *PLoS one.* 2014;9:e91779.
6. Padhani AR, Makris A, Gall P, Collins DJ, Tunariu N, de Bono JS. Therapy monitoring of skeletal metastases with whole-body diffusion MRI. *J Magn Reson Imaging.* 2014;39:1049–1078.
7. Yip S, Perk T, Jeraj R. Development and evaluation of an articulated registration algorithm for human skeleton registration. *Phys Med Biol.* 2014;59:1485.
8. Takahara T, Imai Y, Yamashita T, Yasuda S, Nasu S, Van Cauteren M. Diffusion weighted whole body imaging with background body signal suppression (DWIBS): technical improvement using free breathing, STIR and high resolution 3D display. *Radiation Medicine* 2004; 22:275–282.

9. Hedström A. Automatic segmentation of skeleton in whole-body MR images. 2013. <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A665983&dsid=1452>. Accessed November 21, 2013.
10. Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CR. Quo vadis, atlas-based segmentation? *Handbook of Biomedical Image Analysis*. Boston, MA: Springer; 2005:435–486.
11. Strand R, Malmberg F, Johansson L, et al. A concept for holistic whole body MRI data analysis, imiomics. *PLoS one*. 2017; 12:e0169966.
12. Akbarzadeh A, Gutierrez D, Baskin A, et al. Evaluation of whole-body MR to CT deformable image registration. *J Appl Clin Med Phys*. 2013;14:238–253.
13. Arabi H, Zaidi H. Comparison of atlas-based techniques for whole-body bone segmentation. *Med Image Anal*. 2017;36:98–112.
14. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey. *Med Image Anal*. 2015;24:205–219.
15. Pasoglou V, Michoux N, Peeters F, et al. Whole-body 3D T₁-weighted MR imaging in patients with prostate cancer: feasibility and evaluation in screening for metastatic disease. *Radiology*. 2014;275:155–166.
16. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. 2012;30:1323–1341.
17. Yushkevich PA, Piven J, Cody Hazlett H, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 2006;31:1116–1128.
18. Ceranka J, Polfliet M, Lecouvet F, Michoux N, de Mey J, Vandemeulebroucke J. Registration strategies for multi-modal whole-body MRI mosaicing. *Magn Reson Med*. 2018;79:1684–1695.
19. Bhatia KK, Hajnal J, Hammers A, Rueckert D. Similarity metrics for groupwise non-rigid registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Heidelberg: Springer; 2007:544–552.
20. Polfliet M, Klein S, Huizinga W, de Mey J, Vandemeulebroucke J. The pythagorean averages as group images in efficient groupwise registration. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague: IEEE; 2016:1261–1264.
21. Yoo TS, Ackerman MJ, Lorensen WE, Schroeder W, Chalana V, Aylward S, Metaxas D, Whitaker R. Engineering and algorithm design for an image processing API: a technical report on ITK—the insight toolkit. *Studies in health technology and informatics*. 2002 January 1;85:586–592.
22. Perona P, Shiota T, Malik J. Anisotropic diffusion. In: *Geometry-driven Diffusion in Computer Vision*. Dordrecht: Springer; 1994; 73–92.
23. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29:1310–1320.
24. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18:712–721.
25. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2010;29:196–205.
26. Metz C, Klein S, Schaap M, van Walsum T, Niessen WJ. Nonrigid registration of dynamic medical imaging data using nd+ t b-splines and a groupwise optimization approach. *Med Image Anal*. 2011;15:238–249.
27. Xu L, Krzyzak A, Suen CY. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans Syst Man Cybernet*. 1992;22:418–435.
28. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*. 2009;46:726–738.
29. Artaechevarria X, Munoz-Barrutia A, Ortiz-de Solórzano C. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans Med Imaging*. 2009;28:1266–1277.
30. Cardoso MJ, Leung K, Modat M, et al. STEPS: similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcellation. *Med Image Anal*. 2013;17:671–684.
31. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004;23:903–921.
32. Cardoso M, Clarkson M, Modat M, Ourselin S. Niftyseg: open-source software for medical image segmentation, label fusion and cortical thickness estimation. In: *IEEE International Symposium on Biomedical Imaging*. Spain: Barcelona; 2012.
33. Tustison N, Gee J. Introducing dice, jaccard, and other label overlap measures to ITK. *Insight J*. 2009 July-December;1–4.
34. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591–611.
35. Snedecor GWC, William G. *Statistical methods*. 8th ed. Ames: Iowa State University Press; 1989.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

TABLE S1 Evaluation criteria averaged over leave-one-out cross-validation results for the state-of-the-art and proposed registration strategies ($\pm SD$). Hausdorff distance criteria are presented in mm. The best performing method parameters are highlighted for each label fusion method. *P* value between the proposed and state-of-the-art methods is given

How to cite this article: Ceranka J, Verga S, Kvasnytsia M et al. Multi-atlas segmentation of the skeleton from whole-body MRI—Impact of iterative background masking. *Magn Reson Med*. 2019;00: 1–12. <https://doi.org/10.1002/mrm.28042>