

Multi-atlas-based fully automatic segmentation of individual muscles in rat leg

Michael Sdika^{1,2} · Anne Tonson^{1,3} · Yann Le Fur¹ · Patrick J. Cozzone¹ · David Bendahan¹

Received: 3 July 2015 / Revised: 24 September 2015 / Accepted: 15 October 2015 / Published online: 8 December 2015
© ESMRMB 2015

Abstract

Objective To quantify individual muscle volume in rat leg MR images using a fully automatic multi-atlas-based segmentation method.

Materials and methods We optimized a multi-atlas-based segmentation method to take into account the voxel anisotropy of numbers of MRI acquisition protocols. We mainly tested an image upsampling process along Z and a constraint on the nonlinear deformation in the XY plane. We also evaluated a weighted vote procedure and an original implementation of an artificial atlas addition. Using this approach, we measured *gastrocnemius* and *plantaris* muscle volumes and compared the results with manual segmentation. The method reliability for volume quantification was evaluated using the relative overlap index.

Results The most accurate segmentation was obtained using a nonlinear registration constrained in the XY plane by zeroing the Z component of the displacement and a weighted vote procedure for both muscles regardless of the number of atlases. The performance of the automatic

segmentation and the corresponding volume quantification outperformed the interoperator variability using a minimum of three original atlases.

Conclusion We demonstrated the reliability of a multi-atlas segmentation approach for the automatic segmentation and volume quantification of individual muscles in rat leg and found that constraining the registration in plane significantly improved the results.

Keywords Magnetic resonance imaging · Skeletal muscle · Multi-atlas · Segmentation · Anisotropy · Rat · Reproducibility of results

Abbreviations

3D	Standard 3D nonlinear registration process
SBS	Slice-by-slice nonlinear registration process
2Dc	2D constrained nonlinear registration process
R-	Rescaled along Z
RO	Relative overlap index
CV	Coefficient of variation
ICC	Intraclass coefficient
mv	Majority vote procedure
wv	Weighted vote procedure
z0	No slice translation
z1	1 slice translation
z2	2 slice translation
DSC	Dice similarity coefficient

Michael Sdika and Anne Tonson have contributed equally to this work.

✉ Michael Sdika
michael.sdika@creatis.insa-lyon.fr

¹ Aix-Marseille Université, Centre National de la Recherche Scientifique (CNRS), Centre de Résonance Magnétique Biologique et Médicale (CRMBM) UMR 7339, 13385 Marseille, France

² Université de Lyon, CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon, Université Lyon 1, 7 Avenue Jean Capelle, 69621 Villeurbanne Cedex, France

³ Present Address: Department of Physiology, Michigan State University, East Lansing, MI 48824, USA

Introduction

Muscle quality is commonly defined as the ratio between muscle strength or power and muscle mass [1–3]. Muscle mass can significantly change in a variety of conditions such as growth [4, 5], aging [6, 7] or training [8]. It can also

decrease in pathological conditions including neuromuscular disorders, cardiovascular diseases and diabetes [1, 9]. In this context, muscle size has to be accurately quantified in order to properly investigate the corresponding functional changes. Moreover, given that these changes can be muscle-specific [10–13], the corresponding quantification must ideally be performed individually for each muscle.

Magnetic resonance imaging (MRI) has been recognized as the method of choice because of its noninvasive nature and capacity to distinguish fat, connective and muscle tissue. Considering the high contrast difference between muscle and fat signals, several automatic methods based on the signal intensity threshold have been successfully developed in order to quantify fat-free muscle area [14, 15]. However, muscle size has been considered as a whole so that individual muscle changes such as those occurring in aging [11, 16] or neuromuscular disorders [12] have not been investigated. The automatic segmentation of individual muscles within a limb or within the same muscle group is technically impossible using the same approach considering the poor contrast between connective and muscle tissues. In this context, the gold standard method is based on the manual delineation of regions of interest on contiguous individual slices, which is tedious and time-consuming and has questionable reproducibility. The corresponding results are largely operator-dependent, and it has been reported that measurement of a specific muscle volume using MRI with manual delineation in the forearm would not detect changes below 7 % in healthy adults [17]. In humans, a few alternative methods allowing a selective segmentation of muscles have been proposed in the last few years. A ‘pseudo muscle volume’ has been quantified on the basis of the manual segmentation of a limited number of slices by interpolating the missing data using geometrical [18] or mathematical models [5, 7]. Although these methods can reduce the processing time, they still involve extensive manual work so that a robust automatic method is highly desirable. However, automatic segmentation of individual muscles from MR images is technically challenging given that individual muscles share the same MR properties and that fasciae display very poor contrast. So far, very few methods targeting automatic segmentation of individual muscles have been reported, and to our knowledge none of them have been applied to small animal models despite their increasing use in the investigation of functional changes in skeletal muscle in various healthy and pathological conditions. The fully automatic method based on mesh modeling, which has been proposed recently, still requires integration of manual constraints [19]. Other methods based on the utilization of surface landmarks [20] or principal component analysis [21] have provided promising results, but no information related to the volume error has been reported. Interestingly,

tissue segmentation methods based on an atlas database have been successfully used in brain, bee brain, heart and prostate as well as whole body fat/water separated MR images [22–27]. In this context, the specific hurdles related to individual muscle segmentation are mainly related to the high voxel anisotropy, i.e., a different spatial resolution in at least one of the three dimensions.

The purpose of the present work was to develop an original segmentation method based on a multi-atlas database and allowing a complete automatic quantification of individual muscles in rat leg.

Materials and methods

Animals

Sixteen Wistar Han rats (2.5 months old, 8 males/8 females) were investigated according to the guidelines of the National Research Council Guide for the Care and Use of Laboratory Animals and the French law on the Protection of Animals. Animals were initially anesthetized in an induction chamber with 4 % isoflurane (Forene®; Abbott France, Rungis, France) mixed in 33 % O₂ (0.4 l/min) and 66 % N₂O (0.8 l/min) and then positioned supine in a home-built cradle designed to be accommodated within the MR superconducting magnet to investigate functional changes throughout a muscle exercise session [28].

MRI experiment

MRI of the right leg was performed using a fast spin echo sequence (RARE) at 4.7 T on a Biospec 47/30 superconductor magnet (Bruker, Germany). Eighteen images were recorded in order to cover the entire length of the leg from the knee to ankle using the following parameters: rare factor = 6; repetition time = 2000 ms; echo time = 16 ms; field of view = 30 × 32 mm², matrix size = 256 × 256 pixels; slice thickness = 1 mm, interslice space = 0.5 mm; slice orientation = axial; 8 accumulations for a total acquisition time of 8 min 32 s. Throughout the procedure rats were continuously anesthetized by gas inhalation of 2 % isoflurane mixed in 33 % O₂ and 66 % N₂O using a home-designed facemask. Body temperature was maintained at 37 °C using an electric blanket controlled by a temperature unit (Harvard Apparatus, MA, USA) connected to a home-made rectal probe [28].

Image analysis

Each set of images was both manually and automatically segmented.

Manual segmentation

Gastrocnemius and *plantaris* manual delineation was performed on each slice and for each data set (subject) using the FSL software (FMRIB Software Library, Analysis Group, FMRIB, Oxford, UK). The corresponding results were used as the gold standard measurements and were referred to the reference segmentation database (atlases). In order to assess the reproducibility of this process, manual delineation was performed twice by the same operator and once by two different operators, and the corresponding global muscle volume was compared between each process. The first version of the first operator manual segmentations was used as the ground truth segmentation of the atlas, resulting in 18 ground truth atlas images.

Automatic segmentation-general pipeline

The present automatic segmentation method has been tested using either one (Fig. 1) or multiple atlases (Fig. 2) for both muscles.

As indicated in Fig. 1, the set of images to be segmented (i.e., subject images) was initially corrected for bias inhomogeneities as previously described [29], and the background was removed [30]. Then, the images were

successively registered to each atlas (one or multiple) using an affine transformation (12 degrees of freedom) [31] followed by a nonlinear procedure [32]. Each step of this standardized procedure was optimized in order to account for the high voxel anisotropy (see below). When using N atlases (Fig. 2), the initial segmentation process provided N intermediate segmentation maps, which were combined using a vote procedure in order to provide the final segmentation map.

Each step of the segmentation process was optimized throughout different options of the nonlinear registration (more specifically the model of deformation used), the image upsampling, vote procedure and artificial atlas addition.

Automatic segmentation: specific implementations

Nonlinear registration

The standard full 3D registration method (3D) previously described by Sdika [32] was modified using a slice-by-slice nonlinear registration (SBS) or using an additional constraint, i.e., the deformation along Z was zeroed (2Dc), both constraining the deformation in the XY plan. Acquisition and post-processing constraints were related to the

Processing Pipeline

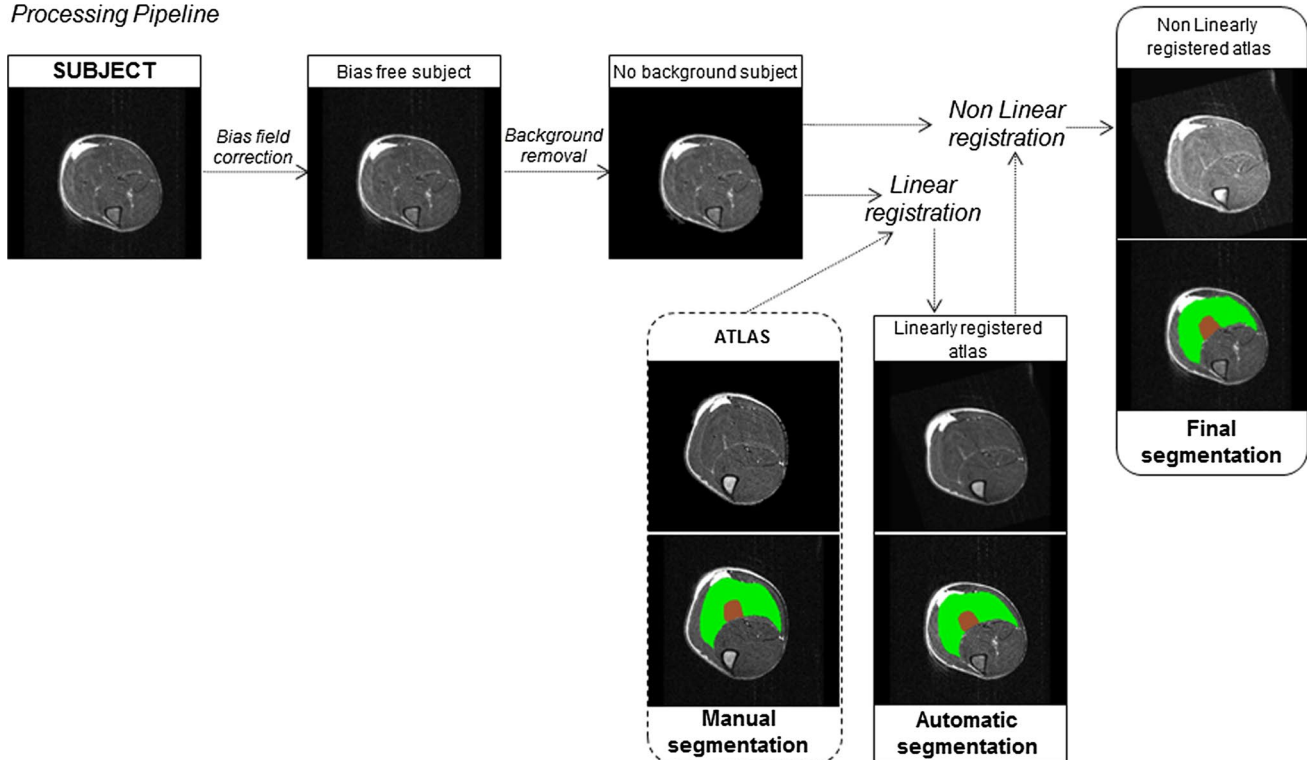


Fig. 1 Schematic representation of a standard processing pipeline used for automatic segmentation of gastrocnemius (green) and plantaris (brown) muscles

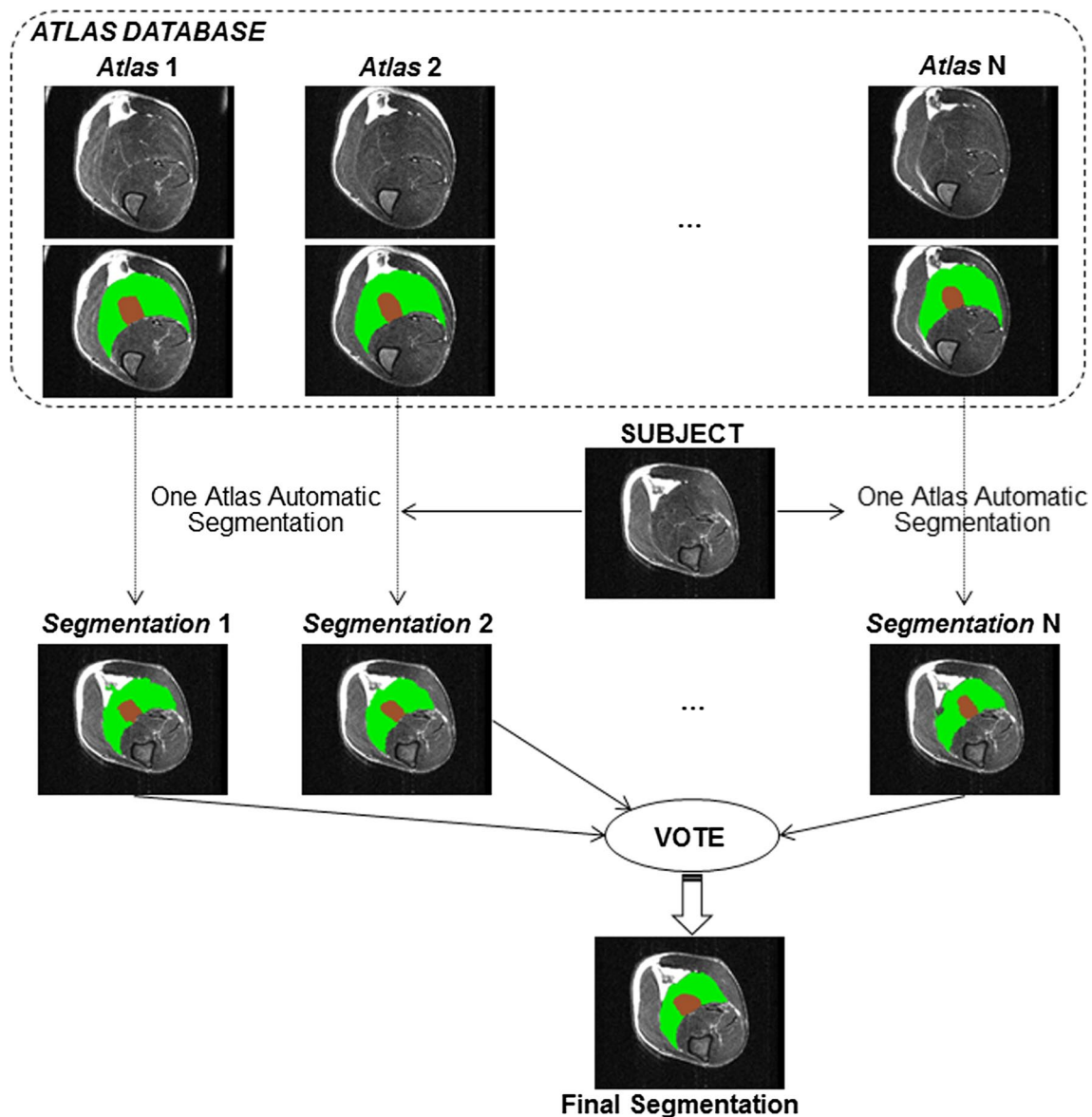


Fig. 2 Schematic representation of an automatic muscle segmentation procedure using multiple atlases

fact that the image resolution was larger in the XY plane as compared to the Z direction, thereby compromising the interpolation process in the Z direction. Both methods assume that the linear registration is reliable for a proper slice alignment and that nonlinear deformation does not occur along Z . These assumptions seem reasonable in our context given that structures in the limb are mostly tubular and do not usually present any drastic morphological changes along the limb length.

Image upsampling

Alternatively, we addressed the voxel anisotropy issue by upsampling the image along Z before the registration

processes. This process aimed at improving the resolution in the Z direction and thus the interpolation process. In this case, images were upsampled by a factor of three using a windowed cardinal sine kernel. This rescaling process has been combined with each of the previously described nonlinear registration models, i.e., with the standard 3D (**R-3D**), 2D constrained (**R-2Dc**) and slice-by-slice (**R-SBS**) models.

Vote procedure

The final segmentation of a given subject was obtained by combining the N segmentation maps produced by each atlas using a vote procedure. The final label was the one

with the largest number of votes (Fig. 2). A simple majority vote (mv) and a vote weighted by the registration residual (wv) as previously described [24, 33] have been implemented and evaluated.

The weighting function used was:

$$w(x) = G_{\sigma_s} * e^{-\frac{r^2}{2\sigma_r^2}} \quad (1)$$

where r is the residual of the registration at a given pixel, G_{σ_s} is a Gaussian function, $*$ is the convolution operator, and σ_s and σ_r are parameters to tune the weighting.

Artificial atlas addition

Given that the quality of a multi-atlas segmentation procedure is an increasing function of the number of atlases used, we tested an original approach of artificial atlas creation by shifting the original atlas slices up to \pm one ($z1$) (i.e., slice one became slice two, ...) or \pm two ($z2$) slices just before the nonlinear registration. As a result two (with $z1$) or four (with $z2$) artificial atlases were added to the database for each original atlas. This translation should compensate for linear registration errors along Z , especially when the nonlinear registration is 2Dc or SBS.

Automatic segmentation evaluation

The different settings proposed in this study were evaluated using a leave-one-out procedure, i.e., each atlas was sequentially used as the subject image and automatically segmented using each of the other subjects as atlases. A final segmentation was obtained from each possibility, and the results presented are related to the corresponding average. The standard deviations were not included in the figures because they were too small.

The segmentation quality for each method was quantified using the relative overlap (RO) defined as:

$$RO(L_g^s, L_a^s) = 100 \frac{v(L_g^s \cap L_a^s)}{v(L_g^s \cup L_a^s)} \quad (2)$$

where L_g^s is the mask of the structure s (*gastrocnemius* or *plantaris* muscles) in the ground truth (i.e., manual segmentation) and L_a^s is the mask of the structure s in the automated segmentation. V is the volume of voxels (in mm^3) inside the binary mask, i.e., the number of voxels multiplied by the voxel size ($0.12 \times 0.12 \times 1.5$ mm). As a result, a high RO is indicative of a high spatial overlap of the two given segmentations. We also calculated the RO values between the manual delineations performed by the same operator (intra) and two different operators (inter) in order to evaluate the accuracy of the present automatic

methods with respect to with the accuracy of the reference method.

Muscle volume quantification

Gastrocnemius and *Plantaris* muscle individual volumes were quantified from both the manual delineation (reference measurements) and fully automatic segmentation results using the setup resulting in the highest RO values. The relative corresponding error was calculated as the volume difference between the two segmentations and expressed in %.

Statistical analysis

The intra- and interoperator variability of the manual segmentation was quantified from the global volume measurements using the coefficient of variation (CV) and intraclass correlation coefficient (ICC) for both *gastrocnemius* and *plantaris* muscles specifically.

Results

Representative transverse images of the rat leg (Fig. 3a) and the corresponding manual (Fig. 3b) and fully automatic segmentations (Fig. 3c) of the *gastrocnemius* and *plantaris* muscles are presented in Fig. 3.

Reproducibility of the reference measurements

Manual delineation of both muscles by a trained expert required a minimum of 15 min per subject (i.e., 18 slices) and has been performed twice by the same operator and an additional time by a second operator. The corresponding volumes for each subject are presented in Table 1. The intraoperator coefficient of variation (CV) was 1.1 ± 0.8 % and 5.1 ± 4.2 % for the *gastrocnemius* and the *plantaris* muscles, respectively. The corresponding ICC was 0.99 (*gastrocnemius*) and 0.87 (*plantaris*). While measurements performed by two different operators provided comparable results for the *gastrocnemius* muscle as indicated by the small CV (2.3 ± 1.4 %) and high ICC (0.97), the corresponding measurements for the *plantaris* muscle were found to be less reproducible, with a CV of 18.7 ± 7.3 % and a poor ICC (0.41).

Accuracy of the automatic methods

We calculated the relative overlap index (RO) in order to evaluate the performance of each automatic procedure for both *gastrocnemius* and *plantaris* muscle segmentations and for different numbers of atlases, from 1 to 14.

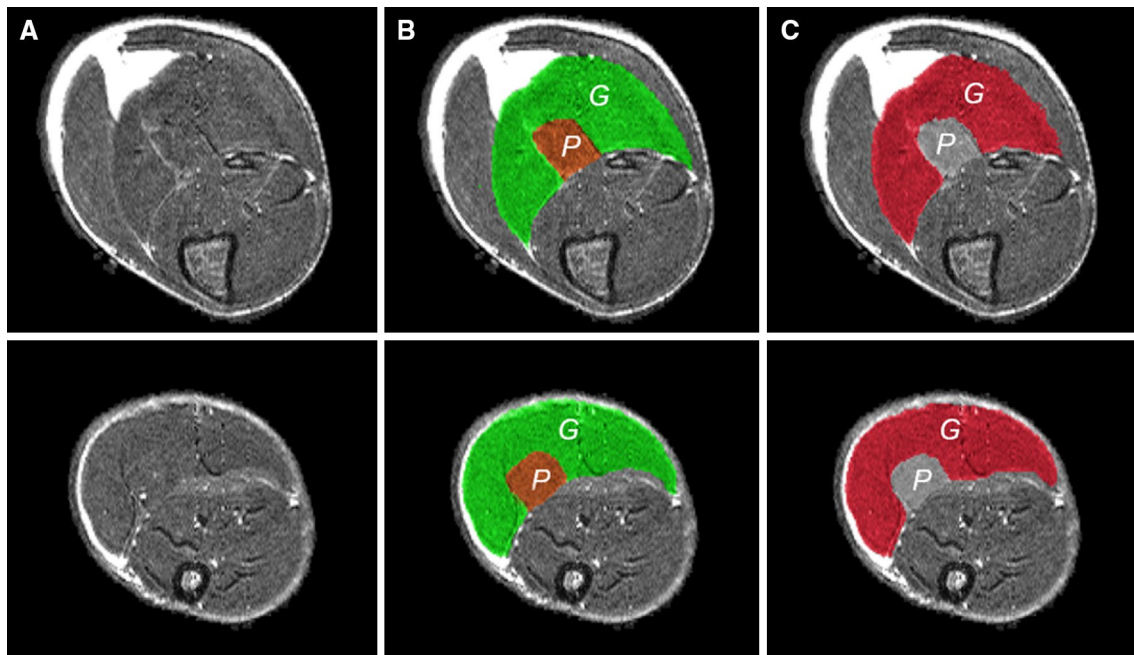


Fig. 3 **a** Representative images of right lower rat leg recorded at the proximal (*top*) and distal levels (*bottom*). **b** Corresponding manual and fully automatic segmentation of *gastrocnemius* (*G*) and *plantaris* (*P*) muscles

Table 1 Individual muscle volumes calculated from manual segmentation

Subject	<i>Gastrocnemius</i> muscle volume (mm ³)			<i>Plantaris</i> muscle volume (mm ³)		
	Operator 1_1	Operator 1_2	Operator 2	Operator 1_1	Operator 1_2	Operator 2
<i>F1</i>	1373.4	1415.0	1341.3	238.7	237.1	229.2
<i>F2</i>	1394.5	1374.0	1480.4	241.0	232.9	195.3
<i>F3</i>	1374.6	1373.2	1443.6	236.5	213.7	193.2
<i>F4</i>	1391.7	1363.0	1306.2	222.1	211.0	176.3
<i>F5</i>	1231.2	1212.8	1203.9	203.0	195.8	163.4
<i>F6</i>	1331.4	1355.8	1295.4	207.8	208.8	162.9
<i>F7</i>	1384.6	1378.8	1380.3	233.2	216.4	160.8
<i>F8</i>	1440.5	1388.8	1323.4	249.0	230.1	164.3
<i>M1</i>	1815.8	1811.6	1781.6	291.6	270.0	192.4
<i>M2</i>	1834.0	1817.5	1763.2	248.3	250.7	170.8
<i>M3</i>	1890.4	1896.9	1830.8	307.6	281.4	188.4
<i>M4</i>	1719.0	1770.5	1699.1	261.5	216.5	180.3
<i>M5</i>	1893.6	1834.2	1906.4	256.6	249.6	197.2
<i>M6</i>	1723.6	1750.7	1738.4	279.2	259.5	193.0
<i>M7</i>	1820.3	1804.1	1860.6	270.8	250.5	195.4
<i>M8</i>	1732.8	1747.3	1765.0	242.6	196.1	183.2
Average	1584.5	1580.9	1570.0	249.3	232.5	184.1
SD	235.4	236.7	242.6	28.3	25.9	17.8

F and *M* designate Female and Male subjects, respectively. Operator 1_1 and 1_2 refer to the results of two measurements performed by the same operator, and operator 2 shows the results obtained by a different operator. *SD* is standard deviation. ROs calculated for manual segmentations performed by the same operator and two different operators were, respectively, 88.3 and 83.2 for the *gastrocnemius* and 81.5 and 68.1 for the *plantaris* muscle segmentations

Processing time to complete both muscle segmentations for one subject ranged from 10 to 15 min, depending on the initial settings when one atlas was used. This time increased linearly with respect to the number of atlases (e.g., a minimum of 1 h was required to perform the same segmentation with 6 atlases). For the sake of comparison, the RO calculated for the manual segmentations performed by the same operator and two different operators were 88.3 and 83.2 for the *gastrocnemius* and 81.5 and 68.1 for the *plantaris* muscle segmentations.

Vote procedure

Table 2 presents the RO values obtained using the different nonlinear registration models (i.e., SBS, 2Dc and 3D) combined with either the majority or weighted vote with respect to the number of atlases for both muscle segmentation. Our results showed that, compared to the simple majority vote, the weighted vote improved the RO when the images were registered using the 2Dc or the SBS model and so regardless of the number of atlases used or the muscle segmented. However, the corresponding improvement was globally larger for the *plantaris* ($3.3 \pm 2.6 \%$ and $2.2 \pm 1.8 \%$ using the 2Dc and SBS models, respectively) as compared to the *gastrocnemius* ($1.1 \pm 0.4 \%$ and $0.6 \pm 0.2 \%$) muscle. Using the 3D model, no effect of the vote procedure was observed for the *gastrocnemius* segmentation while the RO was globally slightly improved (by $1.2 \pm 1.5 \%$) for the *plantaris* segmentation by the weighted vote procedure.

According to these results and for the sake of clarity, only the weighted vote was used for the remaining comparisons.

Images upsampling

We found a beneficial effect of the upsampling along the Z direction when the images were registered with the 3D model. Using this specific setting, the upsampling improved the RO associated with the *gastrocnemius* ($0.6 \pm 0.03 \%$) and *plantaris* ($1.5 \pm 0.07 \%$) muscle segmentation using 1–14 atlases. This indicates that the image anisotropy is a limiting factor for automatic segmentation using atlases and that the upsampling method can take this issue into account. When only an in-plane registration was performed, i.e., using the 2Dc and SBS models, this beneficial effect no longer existed. More specifically, combined with the SBS model, this procedure resulted in a decreased RO by $0.3 \pm 0.06 \%$ for both muscles regardless of the number of atlases used. Combined with the 2Dc model, we found a slight improvement for the *gastrocnemius* segmentation (RO improved by $0.3 \pm 0.09 \%$), while this setting was globally detrimental for the *plantaris* muscle ($-0.1 \pm 0.11 \%$). Accordingly, we discarded the R-SBS and R-2Dc of the nonlinear registration model comparison.

Nonlinear registration models

As illustrated by the higher RO values, our results clearly showed a higher reliability of the automatic segmentation

Table 2 Relative overlap (obtained) with the different nonlinear registration models combined with the majority or weighted vote procedure with respect to the number of atlases included in the database

Atlas, <i>n</i>	<i>Gastrocnemius</i> segmentation, RO						<i>Plantaris</i> segmentation, RO					
	SBS		2Dc		3D		SBS		2Dc		3D	
	mv	wv	mv	wv	mv	wv	mv	wv	mv	wv	mv	wv
1	75.8	75.8	80.2	80.2	80.2	80.2	57.7	57.7	66.1	66.1	64.5	64.5
2	76.8	78.5	81.0	81.8	81.6	81.4	55.5	61.7	63.7	68.4	62.5	66.0
3	80.1	80.7	84.0	84.1	83.8	83.8	63.6	64.5	71.2	71.6	69.1	69.1
4	80.6	81.8	84.5	85.0	84.6	84.4	62.8	66.5	70.2	73.0	68.2	70.0
5	81.9	82.6	85.4	85.8	85.0	85.2	66.1	67.6	73.4	74.3	70.9	71.0
6	82.3	83.3	85.7	86.2	85.7	85.6	66.0	68.8	72.7	74.8	70.3	71.7
7	82.3	83.2	85.8	86.2	85.4	85.6	66.9	68.5	73.8	74.8	71.4	71.6
8	82.6	83.6	86.1	86.6	86.1	86.0	66.9	69.3	73.7	75.7	71.3	72.3
9	82.9	83.8	86.2	86.8	85.8	86.1	68.0	69.5	74.7	75.8	72.1	72.3
10	83.2	84.1	86.5	87.1	86.3	86.3	67.9	70.1	74.4	76.2	71.8	72.6
11	83.2	84.2	86.6	87.2	86.1	86.4	68.6	70.1	75.2	76.3	72.5	72.7
12	83.3	84.3	86.7	87.3	86.5	86.5	68.3	70.4	74.8	76.5	72.2	72.9
13	83.4	84.3	86.7	87.3	86.3	86.5	68.9	70.4	75.6	76.7	72.7	73.0
14	83.6	84.5	86.9	87.4	86.6	86.7	68.9	70.7	75.3	76.8	72.5	73.2

SBS, 2Dc and 3D refer to the slice-by-slice, 2D constrained and standard 3D nonlinear registration models, respectively, and mv and wv are the majority and weighted vote, respectively. Atlas means the original atlas (i.e., from manual delineation)

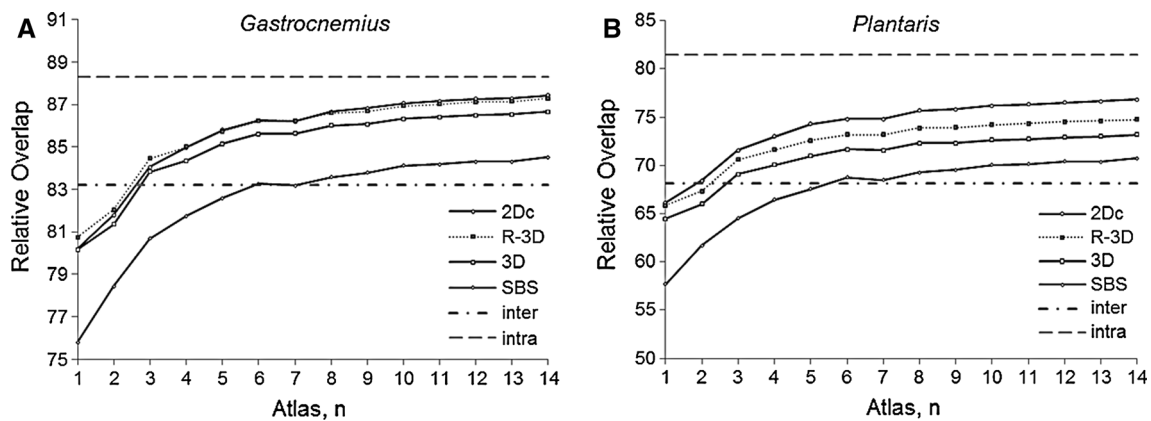


Fig. 4 Relative overlap (RO) expressed with respect to the number of atlases used when the images were registered using 3D, R-3D, 2Dc and SBS models. The dashed line represents RO calculated between

two manual segmentations performed by the same (intra) and by two different operator(s) (inter), respectively

for the *gastrocnemius* muscle as compared to the *plantaris* regardless of the initial setting. However, the accuracy levels differed according to the registration methods independently of the muscle segmented (Fig. 4).

According to the RO values, we found a lower performance of the SBS model compared to the other nonlinear registration models in our experimental conditions. More specifically, while RO increased from 80.2, 80.8 and 80.2 to 86.7, 87.3 and 87.4 with respect to the number of atlases using the 3D, R-3D and 2Dc models, respectively, for the *gastrocnemius* muscle, the corresponding values using the SBS model were significantly reduced to 75.8 and 84.5 % (Fig. 4a). The same difference was observed for the *plantaris* segmentation (Fig. 4b). For this specific muscle, the RO values ranged from 64.5, 65.9 and 66.1 to 73.2, 74.8 and 76.8 when the images were registered using the 3D, R-3D and 2Dc models, respectively, and from 57.7 to 70.7 using the SBS model (Fig. 4b). Finally, considering both *gastrocnemius* and *plantaris* muscle segmentation, the 2Dc model outperformed the other models in our experimental conditions.

Overall, with the exception of the SBS model, our results showed that a minimum of three atlases was necessary to automatically segment either muscle with a level of accuracy similar to the reference method when two operators were involved. However, despite an increased RO with respect to the number of atlases, none of the automatic methods reached the intraoperator threshold whatever the muscle (Fig. 4). In addition, our results showed that the corresponding improvement was significantly reduced beyond size atlases; from 1 to 6 atlases the RO was improved by 6.4, 6.3, 7.0 and 9.0 % for the *gastrocnemius* and by 10.0, 10.0, 11.6 and 16.1 % for the *plantaris* segmentations using the 3D, R-3D, 2Dc and SBS models, respectively, whereas the addition of eight more atlases resulted in a further RO

improvement lower than 1.5 and 3 % for the *gastrocnemius* and *plantaris* segmentation, respectively (Fig. 4).

Artificial atlas addition

According to the previous results, we evaluated the effect of the addition of artificial atlases on the automatic segmentation quality using the weighted vote combined with the 2Dc nonlinear registration model. As illustrated in Fig. 5, the addition of artificial atlases translated by \pm one ($z1$) or \pm two slices ($z2$) improved the RO when a low number of original atlases was used regardless of the muscle segmented. However, the $z1$ setup resulted in a larger improvement compared to $z2$ for both muscles (Fig. 5). In addition, using the ' $z2$ ' atlases we observed a degradation of the RO when more than 4 and 8 original atlases were used, while the ' $z1$ ' atlases improved the RO up to 8 and 14 original atlases for the *gastrocnemius* (Fig. 5a) and *plantaris* (Fig. 5b) segmentation, respectively, compared to the results obtained without (i.e., $z0$).

With respect to computation, the shift of up to k slices of the atlases and the addition to the vote data set multiply the number of atlases as well as the computation time by $2k + 1$. So, when $z1$ was used, the computation time was three times greater than when only the original atlases were used.

Muscle volume quantification

The relative volume error (expressed in % of the reference measurement) resulting from the automatic segmentation of both *gastrocnemius* and *plantaris* muscles with respect to the number of original atlases and using the following settings: the 2Dc nonlinear registration and weighted vote using (2Dc- $z1$) or not (2Dc) artificial atlases are displayed

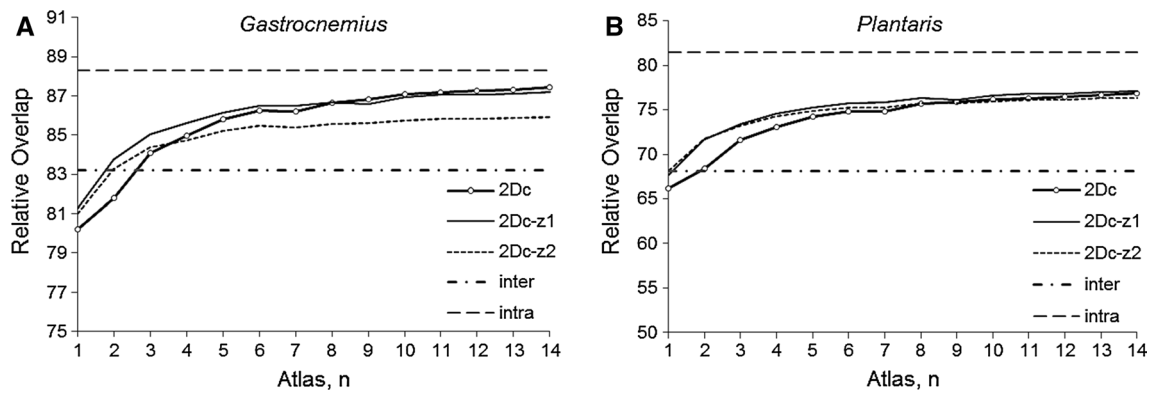


Fig. 5 Relative overlap (RO) calculated using the 2Dc nonlinear registration model expressed with respect to the number of atlases when only original atlases were used (2Dc) or when artificial atlases translated by one (2Dc-z1) or two slices (2Dc-z2) were added to the original

atlases database. The *dashed line* represents RO calculated between two manual segmentations performed by the same (intra) and by two different operator(s) (inter), respectively

Table 3 Absolute volume error from automatic segmentation with respect to the number of atlases included in the database

Atlas, <i>n</i>	Gastrocnemius muscle volume error, (%)		Plantaris muscle volume error, (%)	
	2Dc	2Dc-z1	2Dc	2Dc-z1
1	4.9	5.4	10.0	9.6
2	4.5	5.4	8.3	8.1
3	4.2	5.0	8.0	8.0
4	3.9	5.0	7.7	7.7
5	3.8	4.8	7.2	7.4
6	3.5	4.7	7.4	7.5
7	3.8	4.9	7.4	7.4
8	3.6	4.7	7.1	7.3
9	3.7	5.0	7.0	7.3
10	3.6	4.8	7.2	7.3
11	3.5	4.7	7.1	7.3
12	3.6	4.8	7.2	7.2
13	3.5	4.7	7.0	7.1
14	3.5	4.7	7.0	7.1

2Dc and 2Dc-z1 refer to the automatic segmentations using the 2D constrained nonlinear registration model when only original or with the addition of artificial atlases translated by one slice were used, respectively, combined with the weighted vote. The relative volume error (as percentage) is the averaged absolute error for the entire population

in Table 3. The corresponding absolute volumes in mm^3 when using six atlases as well as the percent error are presented in Table 4. We observed a systematic overestimation of the *gastrocnemius* muscle when the segmentation was performed automatically compared to the reference measurements (Table 4). Regarding the *plantaris* muscle, the absolute volume errors ranged from 0 to 19.5 % and 0.7 to 18.9 % using 2Dc or 2Dc-z1 procedures. However, despite

this larger variability, the average volume calculated from the automatic measurements did not significantly differ from the reference method as a result of individual volumes over and underestimations (Table 4). These errors are in agreement with the volume differences calculated from manual delineation performed by two different operators, i.e., 3.0 ± 2.2 % and 22.9 ± 8.6 % for the *gastrocnemius* and *plantaris* muscles, respectively, but did not reach the intra-operator variability (1.6 ± 1.1 % and 6.8 ± 5.3 % for the *gastrocnemius* and *plantaris*, respectively).

Finally, it is worth noting that in contrast to the improvement we observed for the RO values, the addition of artificial atlases did not improve the accuracy of the volume estimation for both muscles.

Discussion

In this study, we described for the first time a multi-atlas-based segmentation method that can be efficiently used to automatically quantify individual muscle volumes within the limb from rat MR images. We have investigated the effects of a number of modifications to the standard method in terms of accuracy (determined by the RO index) with a particular attention given to the voxel anisotropy of the images and reliability of individual muscle volume quantification.

We showed that an automatic process including a 2Dc nonlinear registration model, weighted vote and minimum of three atlases provided a better RO value than the value obtained from a comparison between manual segmentations performed by two different operators. On that basis, we demonstrated that the accuracy of the automatic segmentation of both *gastrocnemius* and *plantaris* muscles outperformed the interobserver manual segmentation

Table 4 Muscle volumes quantified from automatic segmentation when using six atlases and corresponding error compared to the reference measurements

Subject	<i>Gastrocnemius</i> muscle volume (mm ³)				<i>Plantaris</i> muscle volume (mm ³)			
	2Dc	% error	2Dc-z1	% error	2Dc	% error	2Dc-z1	% error
F1	1448.5	5.5	1474.8	7.4	223.1	-6.5	223.0	-6.6
F2	1422.0	2.0	1448.1	3.8	213.1	-11.6	213.2	-11.5
F3	1519.5	10.5	1498.6	9.0	201.2	-14.9	200.1	-15.4
F4	1436.9	3.3	1424.2	2.3	206.2	-7.2	206.2	-7.1
F5	1295.9	5.3	1323.7	7.5	199.7	-1.6	199.8	-1.6
F6	1357.8	2.0	1375.3	3.3	219.8	5.8	220.4	6.1
F7	1446.0	4.4	1469.7	6.1	224.1	-3.9	223.8	-4.0
F8	1506.1	4.6	1540.3	6.9	228.4	-8.3	227.1	-8.8
M1	1869.1	2.9	1871.2	3.1	286.6	-1.7	285.6	-2.0
M2	1862.8	1.6	1882.1	2.6	288.2	16.1	285.6	15.0
M3	1931.5	2.2	1940.9	2.7	270.6	-12.1	270.0	-12.2
M4	1821.9	6.0	1820.7	5.9	261.5	-0.01	259.7	-0.7
M5	1893.2	0.0	1919.0	1.3	306.6	19.5	305.0	18.9
M6	1767.6	2.6	1782.4	3.4	264.1	-5.4	265.0	-5.1
M7	1836.9	0.9	1860.1	2.2	302.6	11.6	301.5	11.4
M8	1771.4	2.2	1812.9	4.6	286.2	18.0	282.4	16.4
Average	1636.7	3.5	1652.7	4.5	248.9	-0.1	248.0	-0.5
SD	223.6	2.5	223.4	2.3	38.1	11.1	37.4	10.8

F and M designate female and male subjects. 2Dc and 2Dc-z1 refer to the automatic segmentations using the 2D constrained nonlinear registration model when only the original or with the addition of artificial atlases translated by one slice were used, respectively, combined with weighted vote. % Error is the corresponding difference with the manual segmentation. SD is standard deviation

variability. This result highlighted the robustness of the method, which only needs a very limited amount of manual work, i.e., on three atlases. The corresponding global muscle volume error was 4.2 and 8.0 % for the *gastrocnemius* and *plantaris* muscles, respectively. Interestingly, this error was lower than the interoperator variability for the reference method illustrating the reliability of the present fully automatic method. Finally, we showed that the quality of the automatic segmentation was improved by the addition of artificial atlases especially when a low number of original atlases were used. Of interest, this improvement did not affect the global muscle volume quantification.

The reference segmentation of both *gastrocnemius* and *plantaris* muscles was performed three times manually for each slice and each data set, twice by the same operator and an additional time by a new operator. The measurement variability was determined from the global volume quantification using the CV and ICC coefficients. Our results clearly disclosed a muscle-dependent reproducibility with CV much larger (18.7 ± 7.3 %) and ICC (0.41) and RO (68.1) much lower for the *plantaris* as compared to the *gastrocnemius* muscle. This inconsistency is likely related to the muscle size given that in our context the *plantaris* muscle volume was more than six times smaller than that of the *gastrocnemius* regardless of sex. Such a difference has already been suggested in humans. A <1 % interoperator

variability was reported for a large muscle group (i.e., quadriceps) [18], whereas a much larger variability has been reported for smaller muscle groups such as the *flexor* (3.3 %) and *extensor carpi ulnari* (5.7 %) muscles [17]. This difference can also be related to the unclear distinction of the surrounding *plantaris* muscle fascia, especially on the distal part of the leg, making the segmentation on the corresponding images highly operator-dependent. On this basis, it is important to keep in mind that the reproducibility of the manual delineation and as a matter of consequence the performance of the automatic segmentation method is closely linked to the ability to clearly distinguish individual muscles on the MR image, which is itself dependent on the inherent contrast and image resolution.

To our knowledge, no data related to the reproducibility of the manual delineation from small animal MR images have been reported so far. However, both intra- and interoperator CV (1.1 ± 0.8 % and 2.3 ± 1.4 %, respectively) and ICC (0.99 and 0.97, respectively) reported for the *gastrocnemius* muscle volume are in agreement with values previously reported in humans for different muscles [17, 18, 34]. Although frequently used for the segmentation of soft tissues [22, 24, 25, 33, 35], only one study so far has used a multi-atlas-based approach in order to automatically segment a single muscle [36]. More specifically, the authors reported an automatic method dedicated to the

delineation of the pectoral muscle from breast MR images using a standard 3D registration procedure but did not target the muscle volume quantification. The accuracy of this automatic segmentation [36] illustrated by a dice similarity coefficient (DSC) of 0.74 ± 0.06 [i.e., an RO of 0.59, $DSC, RO = DSC/(2-DSC)*100$] was actually inferior to the accuracy of our results. In addition, further supporting our initial conclusion, these authors reported a low reproducibility of the pectoral muscle manual delineation (RO = 0.54) and indicated that this poor reproducibility might be linked to the difficult distinction among the pectoral muscle, surrounding tissues and intercostal muscles on MR images [36].

Our results clearly showed that, regarding voxel anisotropy, a nonlinear registration constrained in two dimensions (i.e., 2Dc) was more beneficial to the automatic segmentation quality than the standard 3D model, especially given that the image resolution did not allow a clear distinction of a small muscle like the *plantaris* muscle. Indeed, in the context of an atlas-based segmentation, voxel anisotropy should result in a biased image interpolation along Z, thereby affecting the segmentation mask and deformation itself. Several factors can explain the superiority of the 2Dc registration as compared to the other models. In the 3D model, the parameters for the deformation are estimated in all three directions, whereas the 2Dc model restrained the deformation within the atlas slices where the resolution is the highest while the Z component is set to zero providing a *de facto* regularization beneficial to the registration. As a result, this model still takes advantage of the 3D nature of the data and is able to generate a more coherent and robust deformation during the registration as compared to the SBS model for which no regularization is imposed slice to slice. Following the same idea, image upsampling along Z before the registration was, as expected, beneficial when it was combined with the standard 3D model. This improvement clearly indicates that image resolution and more specifically voxel anisotropy are limiting factors for the segmentation process using atlas-based approaches. Interestingly, this limitation can be artificially reduced using an upsampling procedure, though at the expense of a larger processing time. Another possibility would be the acquisition of fully isotropic 3D data sets as previously described for the automatic quantification of fat-free muscle volumes [26, 27]. The ICCs between automatic and manual segmentation reported by these authors were above 0.89 for all muscle groups and showed that under those acquisition conditions age and BMI (from lean to obese) differences did not affect the multi-atlas method performance [27]. Interestingly, using full isotropic voxels, Karlsson et al. [26] reported a similar accuracy regarding automatic muscle volume estimates as compared to the manual segmentation, and so despite an eight-fold resolution

difference when data were acquired at 1.5 T (voxel dimensions = $3.5 \times 3.5 \times 3.5 \text{ mm}^3$) or at 3 T (voxel dimensions = $1.75 \times 1.75 \times 1.75 \text{ mm}^3$). Although this approach might provide a more accurate registration and, on that basis, an improved segmentation quality, it would be at the expense of both post-processing and acquisition times. In addition, the resulting gain in terms of volume quantification accuracy should only be marginal given that no drastic change in anatomical shape is expected between two consecutive slices. The improvement resulting from the upsampling procedure combined with the standard 3D model was actually not better than the automatic segmentation using the 2Dc model when considering both muscles together.

Based on the RO index, the 2Dc nonlinear registration also provided the most accurate segmentation for both muscles when it was associated with a weighted vote procedure and no upsampling. The corresponding RO values increased from 80.2 to 87.4 and from 66.1 to 76.8 for the *gastrocnemius* and *plantaris* segmentation with respect to the number of atlases used (from 1 to 14). The RO values reported for the *gastrocnemius* segmentation were in agreement with the results reported by Andrews et al. [21] in much larger muscle groups (i.e., knee extensor and flexor muscles) in humans. On the basis of a principal component analysis, they reported good reliability of the automatic segmentation illustrated by an average DSC value of 0.92 ± 0.03 (i.e., RO = 0.85) [21]. However, it has to be kept in mind that the level of accuracy expected from the automatic method has to be considered with respect to the reproducibility of the standard method. Unfortunately, these authors did not report the variability of their manual delineation process [21]. In the present study, although the accuracy of the automatic segmentation was largely lower for the *plantaris* than for the *gastrocnemius* muscle, the RO values achieved for both muscles outperformed the interoperator RO value using a minimum of three atlases. Similarly, the DSC values from the automatic segmentation reported by Gubern-Merida et al. [36] for the pectoral muscle also outperformed the DSC value calculated from the computed interobserver variability ($DSC = 0.70 \pm 12$), thereby highlighting the reliability of the multi-atlas approach for the individual muscle segmentation. It is worth pointing out that in spite of an increased RO value with respect to the number of atlases included in the database, the addition of further atlases beyond six resulted in a minor improvement of the RO values in both muscles, thereby suggesting that the manual work can be minimized to the delineation of six subjects for a high-quality level according to our data (i.e., RO = 86.2 and 74.8 for the *gastrocnemius* and *plantaris* segmentation, respectively).

From a physiological point of view, the accuracy of a segmentation method can also be evaluated on the basis of the error related to the muscle volume measurement. Here,

the global volume error was 4.2 and 8.0 % and slightly decreased to 3.5 and 7.4 % using 3 or 6 original atlases for the *gastrocnemius* and *plantaris* muscles, respectively. Similarly to what we observed for the RO values, the benefit associated with the addition of more than six atlases was negligible for the volume quantification of both muscles. More specifically, using 14 original atlases the global volume error was 3.5 and 7.0 % for the *gastrocnemius* and *plantaris* muscles, respectively. The present errors were consistent with those previously reported from the commonly used alternative methods based on a reduced number of slices manually delineated [7, 18, 37, 38]. In these models, the truncated cone formula has been particularly used to describe the limb shape between two consecutive, manually delineated slices (as the limb describes a global conic shape). However, it has been shown that the related bias significantly increased with the gap between slices. Tracy et al. [38] showed in humans that a minimum of 5–6 slices (depending on the thigh length) evenly distributed along the quadriceps muscle was necessary to estimate the corresponding volume within a 4 % error range independent of age or sex. Bias reduction to a value of <2 % [38] or to a point lower than the variability of the reference measure [18] required the manual delineation of 12 slices. This was reduced to seven slices using the Cavalieri method (assuming a cylindrical shape between two consecutive slices) for this specific muscle group. Although considerably reduced compared to the reference method (i.e., manual delineation from ~100 contiguous slices), the corresponding manual work remains tedious, especially for large cohorts. In this respect, Morse et al. [37] reduced the manual work to a single slice segmentation by developing a regression model describing the evolution of the cross-sectional area specific to each muscle composing the quadriceps as a function of muscle length in humans. Depending on the location of the segmented slice, the global error ranged from 10 to 27 %. Although these errors have been considered acceptable for the purpose of individual muscle size quantification, their use in a different context is highly questionable. The investigation of any new population would require a new model validation and would lose the benefit of any manual work previously performed. The present automatic method provides a high level of accuracy despite the large morphological heterogeneity of the population, thereby strongly suggesting the possibility of using a given database of atlases for the automatic segmentation of individual muscles from new populations. However, it should be noted that the present method, similarly to the previous automatic work [21, 36], does not identify intramuscular fat so that any comparison of populations with a different relative proportion of intramuscular fat such as in aging [11, 39] and in various pathological conditions [12] would require an additional thresholding process [14].

Conclusion

We reported in the present study an original multi-atlas-based automatic segmentation process dedicated to the skeletal muscle offering high accuracy and reliability. We demonstrated that automatic quantification of individual muscle volume in rat leg can be performed in a population characterized by a wide range of muscle size, thereby opening up promising opportunities in humans.

Author's contribution Michael Sdika: project development, data management and data analysis; Anne Tonson: project development, data collection and data analysis; Yann Le Fur: data collection; Patrick J. Cozzone: data analysis; David Bendahan: data analysis.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards All procedures involving animals were performed according to the guidelines of the National Research Council Guide for the care and use of laboratory animals and the French law on the Protection of Animals.

References

1. Park SW, Goodpaster BH, Strotmeyer ES, de Rekeneire N, Harris TB, Schwartz AV, Tylavsky FA, Newman AB (2006) Decreased muscle strength and quality in older adults with type 2 diabetes: the health, aging, and body composition study. *Diabetes* 55:1813–1818
2. Barbat-Artigas S, Rolland Y, Zamboni M, Aubertin-Leheudre M (2012) How to assess functional status: a new muscle quality index. *J Nutr Health Aging* 16:67–77
3. Metter EJ, Talbot LA, Schrager M, Conwit R (2002) Skeletal muscle strength as a predictor of all-cause mortality in healthy men. *J Gerontol A Biol Sci Med Sci* 57:B359–B365
4. Tonson A, Ratel S, Le Fur Y, Cozzone P, Bendahan D (2008) Effect of maturation on the relationship between muscle size and force production. *Med Sci Sports Exerc* 40:918–925
5. Morse CI, Tolfrey K, Thom JM, Vassilopoulos V, Maganaris CN, Narici MV (2008) Gastrocnemius muscle specific force in boys and men. *J Appl Physiol* 104:469–474
6. Jubrias SA, Odderson IR, Esselman PC, Conley KE (1997) Decline in isokinetic force with age: muscle cross-sectional area and specific force. *Pflug Arch Eur J Physiol* 434:246–253
7. Hasson CJ, Kent-Braun JA, Caldwell GE (2011) Contractile and non-contractile tissue volume and distribution in ankle muscles of young and older adults. *J Biomech* 44:2299–2306
8. Hug F, Marqueste T, Le Fur Y, Cozzone PJ, Grélot L, Bendahan D (2006) Selective training-induced thigh muscles hypertrophy in professional road cyclists. *Eur J Appl Physiol* 97:591–597
9. Kalyani RR, Corriere M, Ferrucci L (2014) Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. *Lancet Diabetes Endocrinol* 2:819–829
10. Brown M, Hasser EM (1996) Complexity of age-related change in skeletal muscle. *J Gerontol A Biol Sci Med Sci* 51:B117–B123
11. Frontera WR, Reid KF, Phillips EM, Krivickas LS, Hughes VA, Roubenoff R, Fielding RA (2008) Muscle fiber size and

- function in elderly humans: a longitudinal study. *J Appl Physiol* 105:637–642
12. Kesper K, Kornblum C, Reimann J, Lutterbey G, Schröder R, Wattjes MP (2009) Pattern of skeletal muscle involvement in primary dysferlinopathies: a whole-body 3.0-T magnetic resonance imaging study. *Acta Neurol Scand* 120:111–118
 13. Akagi R, Takai Y, Ohta M, Kanehisa H, Kawakami Y, Fukunaga T (2009) Muscle volume compared to cross-sectional area is more appropriate for evaluating muscle strength in young and elderly individuals. *Age Ageing* 38:564–569
 14. Mattei JP, Fur YL, Cuge N, Guis S, Cozzone PJ, Bendahan D (2006) Segmentation of fascias, fat and muscle from magnetic resonance images in humans: the DISPIMAG software. *Magn Reson Mater Phys* 19:275–279
 15. Positano V, Christiansen T, Santarelli MF, Ringgaard S, Landini L, Gastaldelli A (2009) Accurate segmentation of subcutaneous and intermuscular adipose tissue from MR images of the thigh. *J Magn Reson Imaging* 29:677–684
 16. Klein CS, Rice CL, Marsh GD (2001) Normalized force, activation, and coactivation in the arm muscles of young and old men. *J Appl Physiol* 91:1341–1349
 17. Smeulders MJC, van den Berg S, Oudeman J, Nederveen AJ, Kreulen M, Maas M (2010) Reliability of in vivo determination of forearm muscle volume using 3.0 T magnetic resonance imaging. *J Magn Reson Imaging* 31:1252–1255
 18. Nordez A, Jolivet E, Südhoff I, Bonneau D, de Guise JA, Skalli W (2009) Comparison of methods to assess quadriceps muscle volume using magnetic resonance imaging. *J Magn Reson Imaging* 30:1116–1123
 19. Gilles B, Magnenat-Thalmann N (2010) Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. *Med Image Anal* 14:291–302
 20. Wang C, Teboul O, Michel F, Essafi S, Paragios N (2010) 3D knowledge-based segmentation using pose-invariant higher-order graphs. *Med Image Comput Comput-Assist Interv* 13:189–196
 21. Andrews S, Hamarneh G, Yazdanpanah A, HajGhanbari B, Reid WD (2011) Probabilistic multi-shape segmentation of knee extensor and flexor muscles. *Med Image Comput Comput-Assist Interv* 14:651–658
 22. Sdika M (2010) Combining atlas based segmentation and intensity classification with nearest neighbor transform and accuracy weighted vote. *Med Image Anal* 14:219–226
 23. Rohlfing T, Brandt R, Menzel R, Maurer CR (2004) Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* 21:1428–1442
 24. Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B (2009) Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE Trans Med Imaging* 28:1000–1010
 25. Klein S, van der Heide UA, Lips IM, van Vulpen M, Staring M, Pluim JPW (2008) Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med Phys* 35:1407–1417
 26. Karlsson A, Rosander J, Romu T, Tallberg J, Grönqvist A, Borga M, Dahlqvist Leinhard O (2015) Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water-fat MRI. *J Magn Reson Imaging* 41:1558–1569
 27. Thomas MS, Newman D, Leinhard OD, Kasmai B, Greenwood R, Malcolm PN, Karlsson A, Rosander J, Borga M, Toms AP (2014) Test-retest reliability of automated whole body and compartmental muscle volume measurements on a wide bore 3T MR system. *Eur Radiol* 24:2279–2291
 28. Giannesini B, Izquierdo M, Le Fur Y, Cozzone PJ, Fingerle J, Himber J, Künnecke B, Von Kienlin M, Bendahan D (2005) New experimental setup for studying strictly noninvasively skeletal muscle function in rat using 1H-magnetic resonance (MR) imaging and 31P-MR spectroscopy. *Magn Reson Med* 54:1058–1064
 29. Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 17:87–97
 30. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57
 31. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* 5:143–156
 32. Sdika M (2008) A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. *IEEE Trans Med Imaging* 27:271–281
 33. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C (2009) Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans Med Imaging* 28:1266–1277
 34. Elliott MA, Walter GA, Gulish H, Sadi AS, Lawson DD, Jaffe W, Insko EK, Leigh JS, Vandenborne K (1997) Volumetric measurement of human calf muscle from magnetic resonance imaging. *Magn Reson Mater Phys* 5:93–98
 35. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB (2011) A review of atlas-based segmentation for magnetic resonance brain images. *Comput Methods Programs Biomed* 104:e158–e177
 36. Gubern-Mérida A, Kallenberg M, Martí R, Karssemeijer N (2012) Segmentation of the pectoral muscle in breast MRI using atlas-based approaches. *Med Image Comput Comput-Assist Interv* 15:371–378
 37. Morse CI, Degens H, Jones DA (2007) The validity of estimating quadriceps volume from single MRI cross-sections in young men. *Eur J Appl Physiol* 100:267–274
 38. Tracy BL, Ivey FM, Jeffrey Metter E, Fleg JL, Siegel EL, Hurley BF (2003) A more efficient magnetic resonance imaging-based strategy for measuring quadriceps muscle volume. *Med Sci Sports Exerc* 35:425–433
 39. Kent-Braun JA, Ng AV, Young K (2000) Skeletal muscle contractile and noncontractile components in young and older women and men. *J Appl Physiol* 88:662–668