ELSEVIER

Bone 41 (2007) 505 – 515

BONE

www.elsevier.com/locate/bone

Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis

Helen R. Buie, Graeme M. Campbell, R. Joshua Klinck, Joshua A. MacNeil, Steven K. Boyd *

Department of Mechanical and Manufacturing Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

Received 6 March 2007; revised 12 June 2007; accepted 2 July 2007
Available online 18 July 2007

Abstract

The use of high resolution peripheral quantitative computed tomography (HR-pQCT) and in vivo micro-CT for studies of bone disease and treatment has become increasingly common, and with these methods comes large quantities of data requiring analysis. A simple, robust, and fully-automated segmentation algorithm is presented that efficiently segments bone regions. The dual threshold technique refers to two required threshold inputs that are used to extract the periosteal and endosteal surfaces of the cortex. The proposed method was tested against the gold standard, semi-automated hand contouring, using 45 datasets: mouse, rat, human, and cadaver data from the tibia or radius with nominal isotropic resolutions of 10–82 μm.

The performance of the proposed method to segment cortical and trabecular compartments was evaluated qualitatively from visualizations and quantitatively based on morphological measurements. Visual inspection confirmed successful segmentation of all datasets using the new method, with qualitatively better results when applied to the human and cadaver data compared to the gold standard. The dual threshold algorithm was able to extract thin and porous cortices, whereas some clipping and perforations occurred for the gold standard.

Morphological parameters measured for segmentation by the proposed method versus the gold standard agreed (95% confidence) for Tb.Th, Tb.Sp, and Tb.N, but not Ct.Th and BV/TV for the human and cadaver datasets. Nonetheless, correlations ranged from 0.95 to 1.00 for all morphological parameters except the cadaver Ct.Th because systematic errors were present. Poor agreement for Ct.Th and BV/TV was due to qualitatively incorrect segmentation by the gold standard when the cortex was thin compared to trabeculae, or operator bias during hand contouring. Since Tb.Th, Tb.Sp, and Tb.N were insensitive to segmentation method, despite operator bias, they are robust parameters for inter-site comparisons.

The dual threshold method offers a robust and fully-automated alternative to the gold standard that can efficiently segment bone regions with accurate and repeatable results. The algorithm can be easily implemented since it uses simple image analysis tools. Two input thresholds allow adjustment of the masked output, and are easily determined by trial and error. Using the same input thresholds for similar datasets assures maximal consistency while alleviating time consuming semi-automated contouring.

Keywords: Bone; Micro-computed tomography; Segmentation; Three-dimensional (3D) image analysis; High resolution peripheral quantitative computed tomography

Introduction

High resolution peripheral quantitative computed tomography (HR-pQCT) and micro-CT have recently been introduced as powerful imaging methods for assessing and quantifying the effects of bone disease and treatment on architecture and strength [1–3]. Micro-CT systems are typically used to examine bones of small animals in vivo, but can also be used to sample cores of larger animals or from human biopsies [4]. These non-destructive imaging modalities offer nominal isotropic resolutions ranging typically between 82 μm and 10 μm for clinical and animal use, respectively, which are sufficient to resolve both cortical and trabecular architecture [5–8]. Development of

* Corresponding author. Fax: +1 403 282 8406.
E-mail address: skboyd@ucalgary.ca (S.K. Boyd).

8756-3282/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.bone.2007.07.007

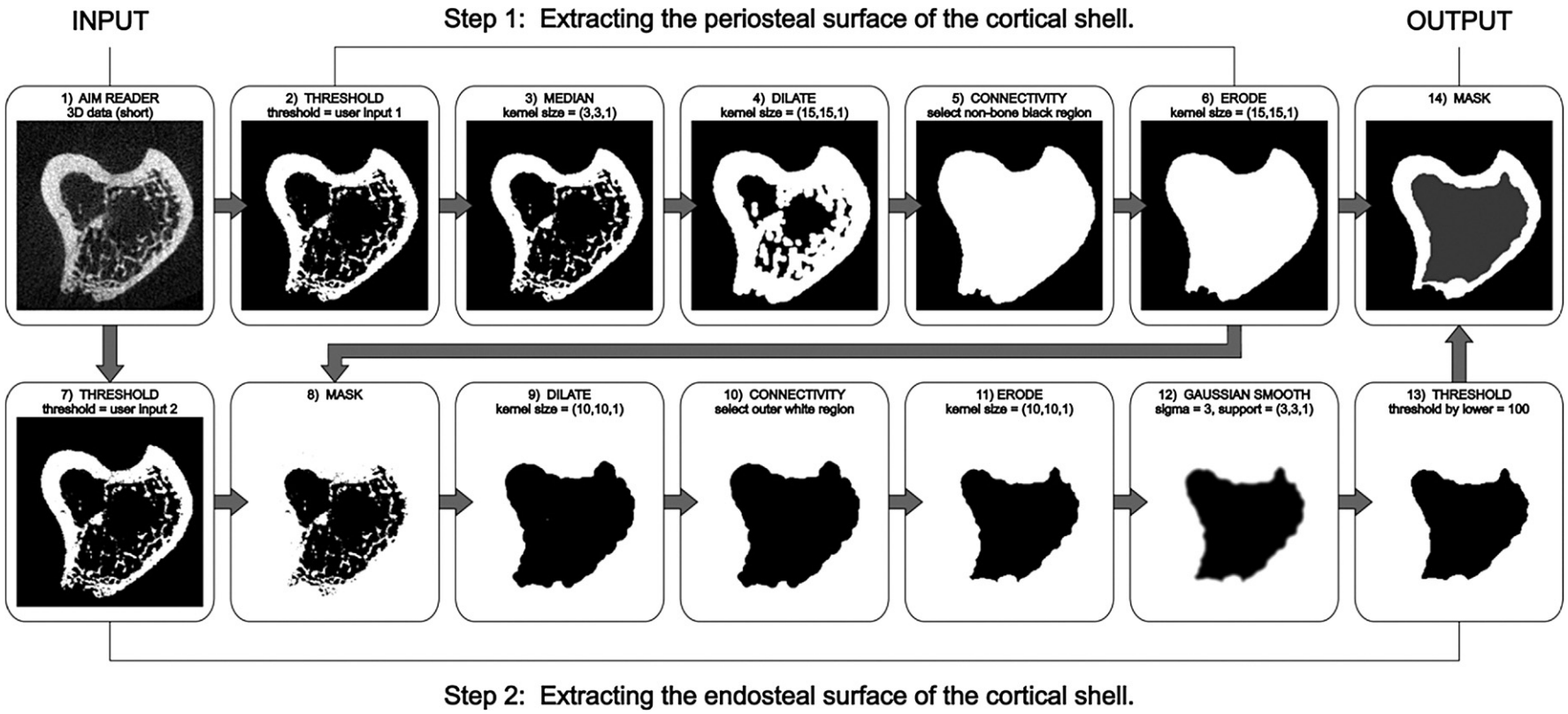


Fig. 1. Summary of the filters and parameters implemented in the dual threshold algorithm for analysis of a 3D dataset. The algorithm is shown here applied to a single CT slice to illustrate how the various stages of filtering are used to extract the cortical and trabecular regions; however, it should be noted that the algorithm is applied in 3D to the entire dataset.

Table 1
Properties of datasets analyzed and input thresholds used in the dual threshold analysis

Dataset		Scan Location	n	Resolution [μm]	Dual threshold inputs 1	and 2
Mouse	Balb/c (BAL)	Proximal tibia	5	10	11 500	10 500
	C3H	Proximal tibia	5	10	11 500	10 500
Rat	Ovariectomized (OVX)	Proximal tibia	10	12	11 000	12 000
	Sham operated (SHM)	Proximal tibia	10	12	11 000	12 000
Human		Distal tibia	5	82	3000	5000
		Distal radius	5	82	3000	5000
Cadaver		Distal radius	5	82	3000	4000

tools to accurately and efficiently quantify bone structure has become increasingly important as large numbers of three-dimensional (3D) datasets are generated with the use of these methods.

Segmentation of cortical and trabecular bone compartments is the first step in a structural analysis and is critical for accurate quantification of architectural parameters; for example, bone volume ratio would be overestimated by the inclusion of dense cortical shell in the analysis. The current segmentation gold standard is a semi-automated slice-by-slice hand contouring approach [9]. Snake algorithms that minimize spline energy based on large image gradients are used to "snap" the contours to edges of interest as directed by the operator [10]. In addition, an interpolating function may be used to reduce the number of slices that must be hand contoured. Although the snake algorithm and interpolating function reduce contouring time, they are not without limitations. The snake algorithm does not always perform as desired and may attach contours to unwanted regions, thereby slowing down the contouring process. Typically, patient datasets with low bone mineral density and high cortical porosity are problematic for the snake algorithm. Interpolating functions can be used successfully over small regions, where changes in the bone are minimal, but the operator must verify that the interpolated contours do not include the cortical region. Finally, being a semi-automated method, the gold standard is subject to operator error, affecting precision and bias. Thus, it is normally necessary to ensure that only one operator performs all the contouring in a study, which limits analysis efficiency.

Algorithms based on region growing, energy minimizing spline curves, and deformable model methodology have been implemented to enhance cortical segmentation and improve consistency, however, these methods are complex and mostly remain semi-automated [11]. The assumption that cortical shell is thick compared to trabeculae has led to segmentation methods based on high degrees of smoothing. The data is smoothed to "erase" trabeculae, then thresholded to extract the cortical region. Some problems with this method include loss of thin cortical shell, which are undistinguished from the trabeculae, inclusion of thick trabeculae as cortex, and reduction of the usable data because outer voxels must be cropped due to the smoothing. Another approach involves use of a Euclidean distance map combined with surface normals of the periosteal surface to find the regions of maximum thickness, which are assumed to be located at the centre of the cortical shell [12]. This method appears sound but has not been rigorously tested and requires some pre-processing of the data. In a commercially available standard patient analysis (Image Processing Language v. 4.29d, Scanco Medical AG, Bassersdorf, Switzerland), high degrees of smoothing are used in conjunction with hand contours of the periosteal surface and connectivity criteria to reduce the smoothing problems discussed. However, the method is limited by the use of hand contours, which make it semi-automated.

The objective of the current study was to develop a robust and fully-automated segmentation algorithm for in vivo micro-CT animal and patient analyses using simple image processing techniques. In vivo scans lend well to automation of the segmentation process because of consistency in: (1) image intensity from calibration of grayscale levels using phantoms, (2) bone orientation and resolution, and (3) imaging region of interest.

Materials and methods

Dual threshold segmentation algorithm

A fully-automated segmentation method was developed and given the name "dual threshold" because it requires two threshold inputs, which affect the periosteal and endosteal surfaces of the output mask. The proposed algorithm is broken down into two steps where first the periosteal surface of the cortex is identified, and second followed by the endosteal surface (Fig. 1). In the first step, the image is thresholded and a connectivity filter is used to extract the non-bone region based on the assumption that the trabecular region is enclosed by cortical shell. Therefore, for the connectivity filters to be effective, Volkmann's canals (i.e. tubular passages that perforate the surfaces of the cortex) must first be removed using morphological closing operations (dilation followed by erosion) prior to extraction. Noise is reduced by applying a median filter. In the second step, the original dataset is re-thresholded and masked with the non-bone region, found in step one, leaving only the marrow cavities. Dilation and erosion operations are then used to reconnect marrow cavities, thus eliminating trabeculae, allowing extraction of a trabecular region mask via a connectivity filter. The region is gaussian smoothed and thresholded to reduce roughness of the endosteal mask surface. A mask of the trabecular, cortical, and non-bone regions is created by combining the output of the two steps. The algorithm was implemented within the framework of an open source code (C+; Visualization Toolkit [13]) and is summarized with specific filter parameters provided (Fig. 1). To apply the algorithm, the user must adjust two threshold values to the specific dataset at hand, which impact the size of the mask's periosteal and endosteal surfaces. The remaining filter parameters are kept constant for all datasets and were determined experimentally by trial and error. The dilation kernels were set first to ensure that cortical porosity was removed and marrow cavities were reconnected for all datasets. Erosion kernels values were then set equal to the dilation kernels to preserve the original mask size. To avoid loss of detail due to dilation and erosion operations, the kernel sizes for these filters were made as small as possible while satisfying these requirements. Noise was minimal in all datasets and, therefore, a modest kernel for the median filter was chosen. Finally, the parameters for the gaussian smoothing and thresholding were kept constant for simplicity because a single set found by trial and error was found to provide a reasonable reduction of roughness of the mask's endosteal surface for all the datasets.

Application of the algorithm to 3D datasets

To test the robustness of the automated segmentation algorithm, it was applied to 3D datasets ($n=45$) that ranged in resolutions from 10 to 82 μm and in cortical thickness from 0.14 to 1.40 mm (Table 1). Included for analysis were in vivo measurements of young (24–35 years) human radii ($n=5$) and tibiae ($n=5$), BALB/C and C3H mice tibiae (BAL, $n=5$; C3H, $n=5$), and ovariectomized and

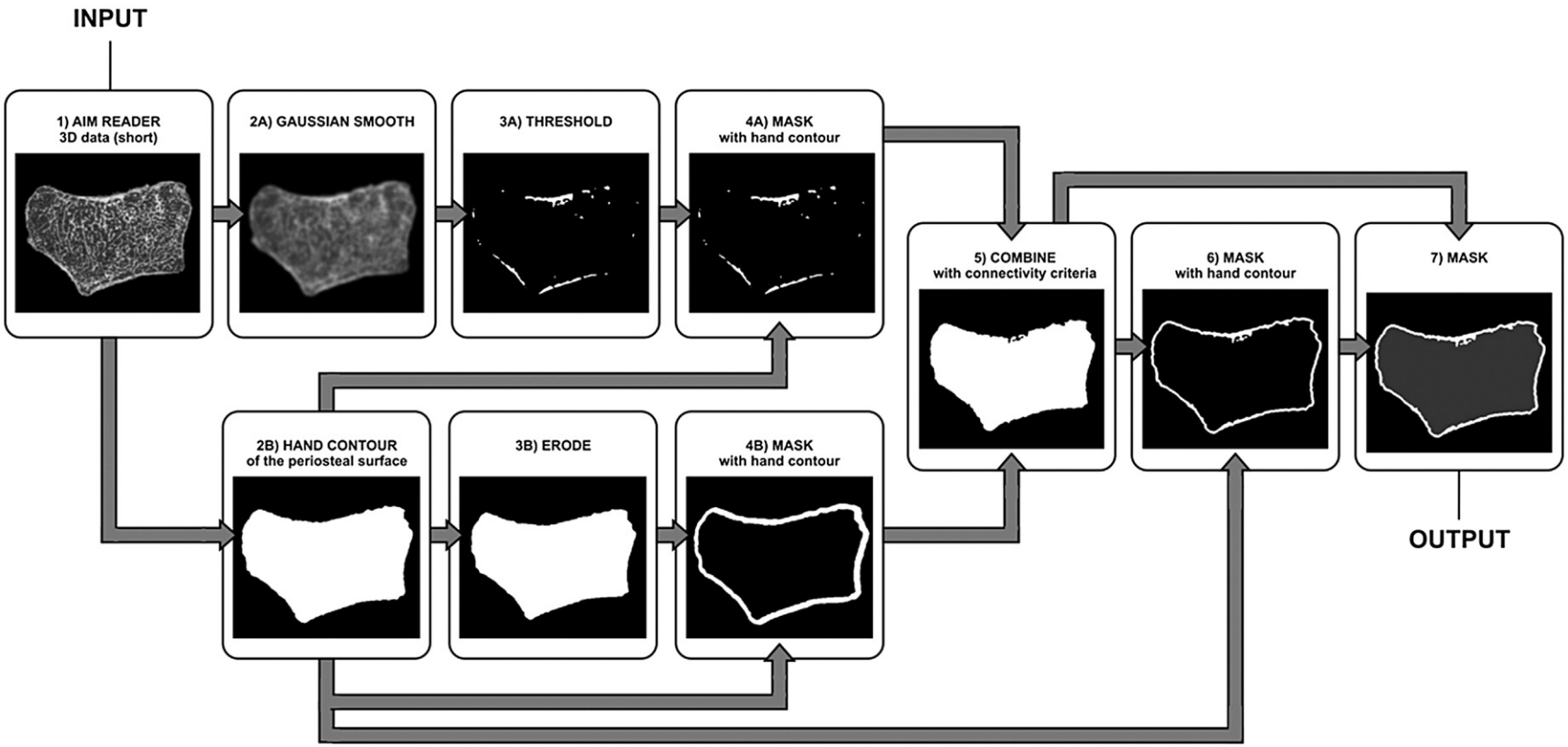


Fig. 2. General image processing used by a standard patient analysis (Scanco Medical AG) to segment cortical and trabecular compartments. For this method, hand drawn contours must be provided of the cortical's periosteal surface. A binary image is the output of the smoothed data (4A), which is used for area measurements to derive Ct.Th. The additional processing is then carried out prior to measurement of BV/TV, Tb.Th, Tb. Sp., and Tb.N.

Table 2
Estimates of agreement of morphological parameters based on the linear regression approach for the dual threshold method and gold standard

Measurement	Dataset	Parameter	Estimate	*P*-value	R^2
Ct.Th	Mouse	Slope	1.014±0.062	0.647	0.99
		Intercept	0.000±0.082	0.936	
	Rat	Slope	1.010±0.061	0.715	0.99
		Intercept	0.008±1.507	0.199	
	Human	Slope	0.801±0.005	0.014 [a]	0.95
		Intercept	0.222±0.030	0.011 [b]	
	Cadaver	Slope	0.915±0.029	0.700	0.87
		Intercept	0.285±0.113	0.001 [b]	
BV/TV	Mouse	Slope	0.972±0.180	0.442	0.99
		Intercept	−0.274±2.022	0.564	
	Rat	Slope	0.996±0.065	0.895	0.98
		Intercept	1.210±1.143	0.059	
	Human	Slope	0.982±0.030	0.792	0.96
		Intercept	0.170±0.030	0.878	
	Cadaver	Slope	0.945±0.193	0.004 [a]	1.00
		Intercept	0.296±0.016	0.026 [b]	
Tb.Th	All	Slope	1.007±0.015	0.629	0.99
		Intercept	−0.001±0.001	0.173	
Tb.Sp	All	Slope	0.991±0.006	0.131	1.00
		Intercept	0.001±0.002	0.478	
Tb.N	All	Slope	1.017±0.009	0.076	1.00
		Intercept	−0.023±0.037	0.542	

Significance is determined from a two-tailed Student's *t*-test with the null hypothesis slope = 1 and intercept = 0.
[a] Slope is significantly different from one ($\alpha = 0.05$).
[b] Intercept is significantly different from zero ($\alpha = 0.05$).

sham operated Wistar rat tibiae (OVX, $n = 10$; SHM, $n = 10$). Fresh human cadaver radii from elderly (80–85 years) individuals were also included ($n = 5$) for their reduced mineralization and thin cortex, similar to an osteoporotic condition, which makes segmentation more difficult. These samples were scanned with their surrounding soft tissue intact. Scanning protocols were approved by the Conjoint Health Research Ethics Board at the University of Calgary, and all human patients gave written informed consent before participating in the study. For the human and cadaver data, 110 slices were collected at an isotropic resolution of 82 μm (XtremeCT, Scanco Medical AG), and for the mice and rat data, 212 slices at 10 and 12 μm, respectively (VivaCT 40, Scanco Medical AG). However, only a subset of 125 slices was analyzed for the mice that corresponded to the contoured region of the gold standard approach, which purposely avoided the growth plate because it was not of interest for morphological analysis.

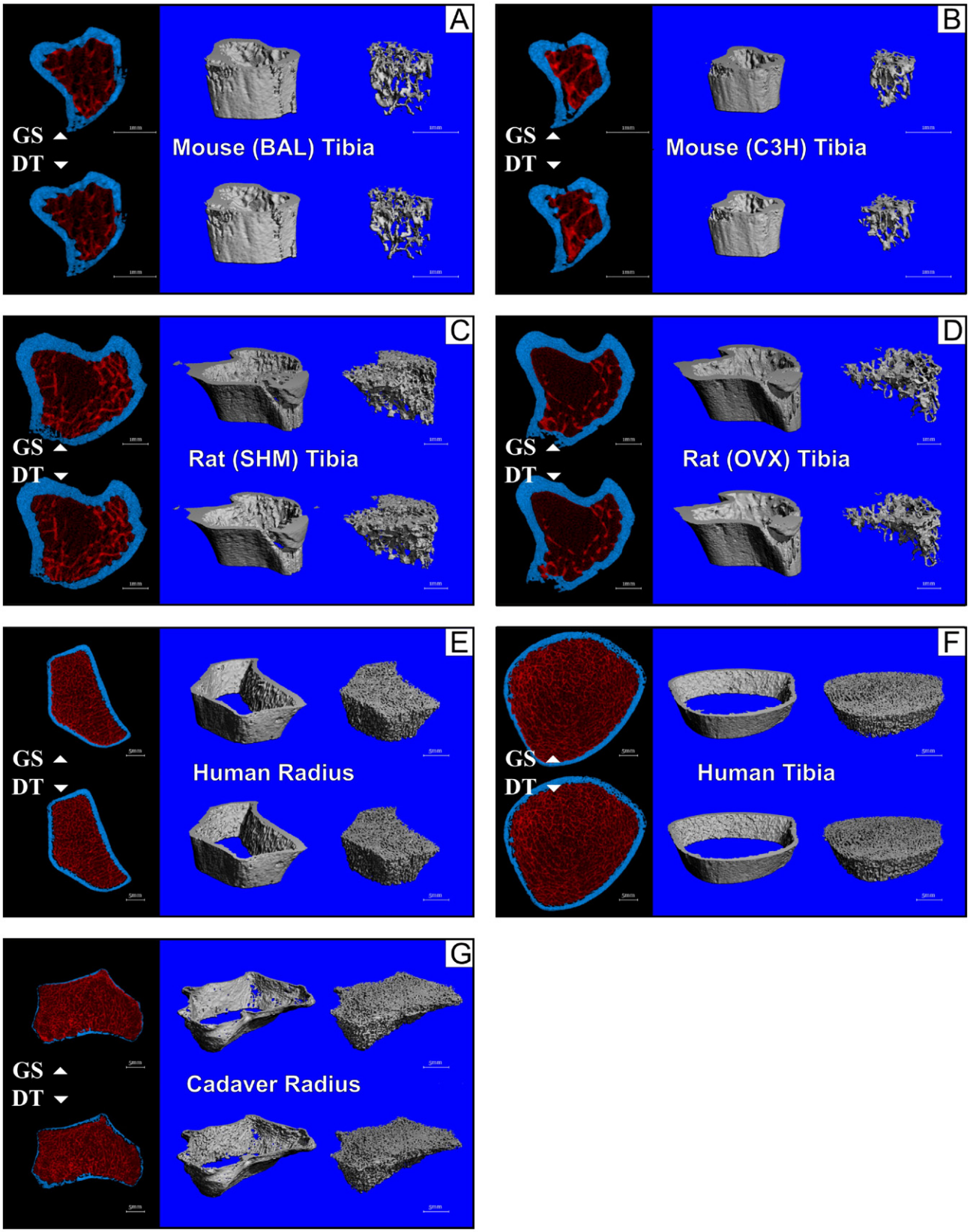
All datasets were pre-processed automatically with a custom-written program to remove the ulna or fibula from the forearm or shank datasets, respectively. In brief, connectivity filters were used to extract the largest bone component. For the gold standard method, three different operators performed the contouring based on their expertise: (1) human and cadaver, (2) rat, and (3) mice; a fourth operator set up the dual threshold algorithm for all the data. Contours were drawn adjacent to the endosteal surface for the animal data, which allowed direct segmentation of the bone compartments. For the patient data, contours were drawn adjacent to the periosteal surface for use with a standard patient analysis (Image Processing Language v. 4.29d, Scanco Medical AG; Fig. 2). Input thresholds for the dual threshold method were kept constant for similar types of data (Table 1). To determine the inputs, a threshold value for extracting the bone from the surrounding tissue was determined and then used for both inputs to the algorithm, which was applied to a single CT slice. The input thresholds were then optimized by visually assessing the quality of segmentation. The first and second threshold values were modified as necessary to adjust the periosteal and endosteal surfaces of the mask, respectively. Two independent threshold values are necessary to allow for modification of the mask's surfaces because all other filter parameters are fixed and may not be fully-optimized for the particular dataset at hand. In general, the periosteal threshold can be lowered to overestimate the size of the cortex and ensure that the resulting mask does not clip any of the bone. Similarly, the endosteal threshold can be lowered to underestimate the trabecular region of the mask. Typically, the best combination of threshold inputs will result in a periosteal threshold that is lower than the endosteal threshold; however, this is not required. For example, a moderately lower endosteal threshold for the mouse data (Table 2) improved segmentation because when the same threshold was used for both inputs, the final mask clipped the endosteal surface by approximately 1 voxel in thickness. Once appropriate inputs were found, they were tested on a single slice from a few different datasets of similar type. This process took approximately 5 min for each dataset (mouse, rat, human, and cadaver), 20 min in total.

To validate the proposed method, morphological parameters were determined (Image Processing Language v. 4.29d, Scanco Medical AG) and compared for segmentation by the dual threshold method versus the gold standard, semi-automated hand contouring approach. Morphological measurements were made using direct methods [14,15] for the mice and rats. For the patient data, semi-derived methods [2,5,9,16,17] were used because trabecular thickness is not accurately represented due to partial volume effects. In the patient analysis, BV/TV is derived from the CT attenuation data (i.e. density) of the trabecular region [5] and Ct.Th is calculated from the mean cortical area divided by the periosteal surface. For the gold standard, Ct.Th measurements are made from an intermediate segmentation step (Figs. 2 and 4A), as per the standard patient analysis. Total cortical and trabecular volumes, Ct.TV and Tb.TV, were also measured using direct methods for all datasets. The Bland–Altman method was used to qualitatively assess the agreement between the two methods, providing estimates of systematic and random errors [18]. A two-tailed Student's *t*-test with significance $\alpha = 0.05$ was then used to check if linear regression slopes and intercepts were significantly different from the line of equality (slope = 1, intercept = 0). Finally, two datasets (human cadaver and rat) were used to explore the algorithm's sensitivity to threshold inputs by calculating the change in cortical and trabecular volumes of the output mask.

Results

A qualitative comparison is provided first based on visualization of representative segmented datasets after masking the bone regions via the gold standard, semi-automated hand contouring, versus the fully-automatic dual threshold method (Fig. 3). For the mice and rat data, the methods produced comparable segmentations, with differences occurring primarily where trabeculae attach to the cortical shell and around the subchondral plate, where the distinction between cortical and trabecular regions is less clear. The operator who hand contoured the rat data was consistently conservative when drawing contours of the trabecular region around the subchondral plate. No visual differences could be detected between methods for the periosteal surface of the rodent data once the data had been masked. For the patient data, the endosteal surface was generally smoother for the gold standard. However, the dual threshold method produced a more uniform surface texture compared to the gold standard, which had larger protrusions for trabeculae that remained connected to the cortex. In addition, for thinner regions of the cortical shell, the gold standard segmentation had some clipping of the endosteal surface. The clipping was most apparent in the 2D segmentation images (Fig. 3) and from the 3D images of the cadaver samples, where clipping lead to perforations of the cortical shell. The dual threshold algorithm was able to fully extract thin cortices; perforations in this case occurred as a result of poor image quality (low image intensity in some regions of the cortical shell). Similar to the rodent data, no visual discrepancies could be detected between methods for the periosteal surface, aside from the aforementioned perforations. For the gold standard

segmentation used for the Ct.Th measurements of the patient analysis, perforations in the cortical shell occurred where it was thin and thick trabeculae were confused for cortex. A significant portion of the thinned cortical shell was missing for the cadaver samples (Fig. 4). In addition, the periosteal surface was overly smooth for both the human and cadaver datasets because the hand contours become the periosteal surface and are less accurate (Figs. 2 and 4).



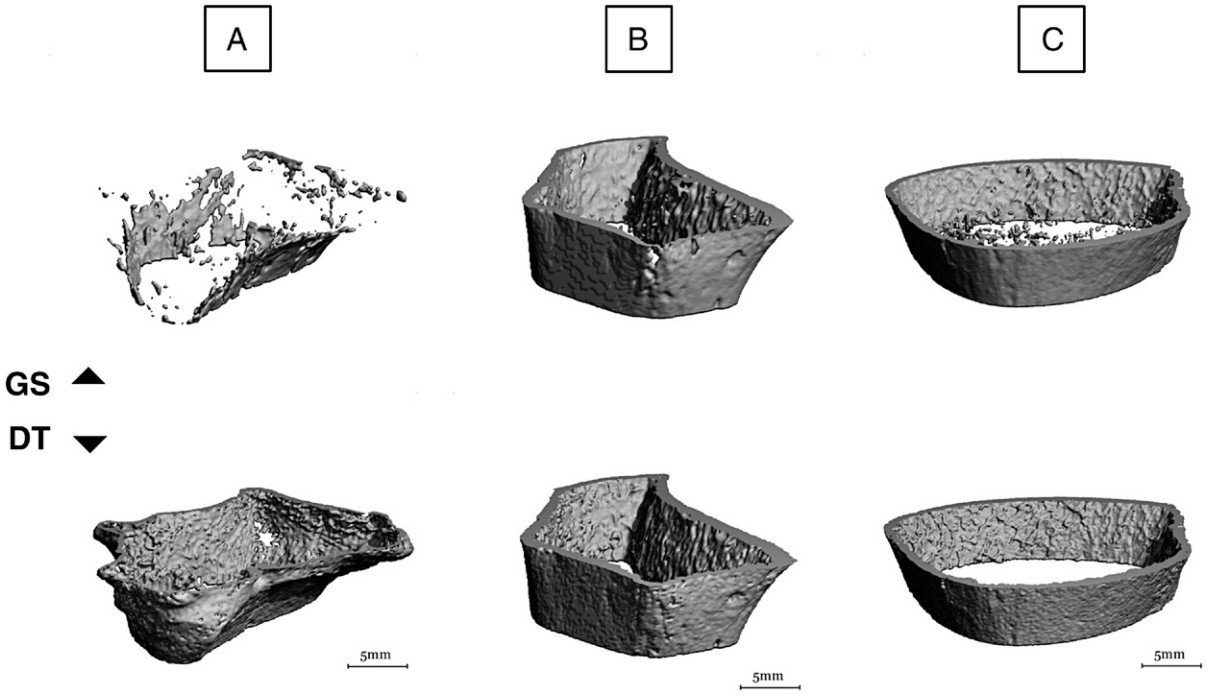


Fig. 4. Three-dimensional renderings of the segmented cortical shell used for Ct.Th and Ct.TV calculations of the patient data, gold standard (top) and dual threshold method (bottom): (A) cadaver radius, (B) human radius, and (C) human tibia. The segmented cortex for the gold standard is produced from the initial smoothing and thresholding (Fig. 2, and panel A of this figure) and appears smoother, with loss of detail at edges and where the cortex is thin; furthermore, thick trabeculae were extracted as part of the cortical shell. The dual threshold method, on the other hand, was not confounded by thick trabeculae and was able to maintain cortical detail. The difference in segmentation was most dramatic for the cadaver data.

For quantitative comparison, linear regression plots, Bland–Altman plots, and error histograms are provided (Fig. 5). Exact agreement occurs between two methods when the regression yields a slope of one and intercept of zero [18–20]. The estimates of agreement are provided (Table 2) for two-tailed Student's *t*-tests with the null hypothesis slope equals one and intercept equals zero. Both slope and intercept *P*-values were greater than the significance level ($\alpha = 0.05$) for trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), and trabecular number (Tb.N) indicating that the two methods agreed with 95% confidence. These morphological parameters had reasonably normal error distributions (Fig. 5), suggesting random errors. Cortical thickness (Ct.Th) on the other hand was subject to systematic errors (Bland–Altman plot, Fig. 5) for the human and cadaver data, and therefore each type of data was tested separately for agreement. The mice and rat data agreed for the dual threshold method and gold standard with 95% confidence (Table 2). In contrast, the two methods did not agree for the human and cadaver data ($P < 0.02$, Table 2) but were nonetheless highly correlated ($R^2 > 0.87$) because of a proportional bias (Fig. 5). Bone volume ratio (BV/TV) showed a constant bias in the rat data (Fig. 5), but it was not significant at the 95% confidence level (intercept $P = 0.06$, Table 2). The dual threshold method and gold standard agreed for measurements of BV/TV of the mice, rat, and human data. The cadaver data did not agree ($P < 0.03$, Table 2), however, only five data points were available when analyzing the datasets separately.

Agreement between methods for the masked regions followed similar patterns to that of the morphological parameters. Total cortical (Ct.TV) and trabecular (Tb.TV) volumes of the mice and rat datasets agreed for both methods with 95% confidence (Table 3). A non-significant constant bias ($P > 0.05$) could be observed for the rat data (Fig. 6) with the gold standard estimating a larger Ct.TV than the dual threshold method, and correspondingly a smaller Tb.TV. The methods did not agree for the human and cadaver data but were highly correlated ($R^2 > 0.88$, Table 3). The differences in Ct.TV and Tb.TV as measured for the two methods was less than 10% for the mouse and rat data, whereas Tb.TV varied by up to 15% for the human and cadaver data, and Ct.TV differences were as much as 20 and 193%, respectively.

Trabecular and cortical volumes of the mask generated by the dual threshold algorithm were insensitive to a 5% change in either threshold input. Trabecular volume varied by less than

Fig. 3. Two- and three-dimensional renderings of datasets segmented using the gold standard (GS, top), semi-automated hand contouring, versus the fully automatic dual threshold method (DT, bottom): (A) BAL mouse tibia, (B) C3H mouse tibia, (C) OVX rat tibia, (D) SHM rat tibia, (E) human radius, (F) human tibia, and (G) cadaver radius. The 2D slice was selected from the middle of the stack analyzed; the cortical and trabecular regions are coloured blue and red, respectively, to show the segmentation. Segmentation was similar for the mice and rats (A–D), with differences occurring primarily where trabeculae intersect the cortical shell and in the subchondral region. Several differences in segmentation can be observed for the human and cadaver data (E–G). The endosteal surface is generally rougher but more uniform for the dual threshold method. The dual threshold method also maintained thinner regions of the cortex whereas for the gold standard the segmentation some clipping of the endosteal surface occurred, which can be most clearly seen in the 2D slices. The clipping is most prominent for the cadaver data (G), where it leads to perforations of the cortex.

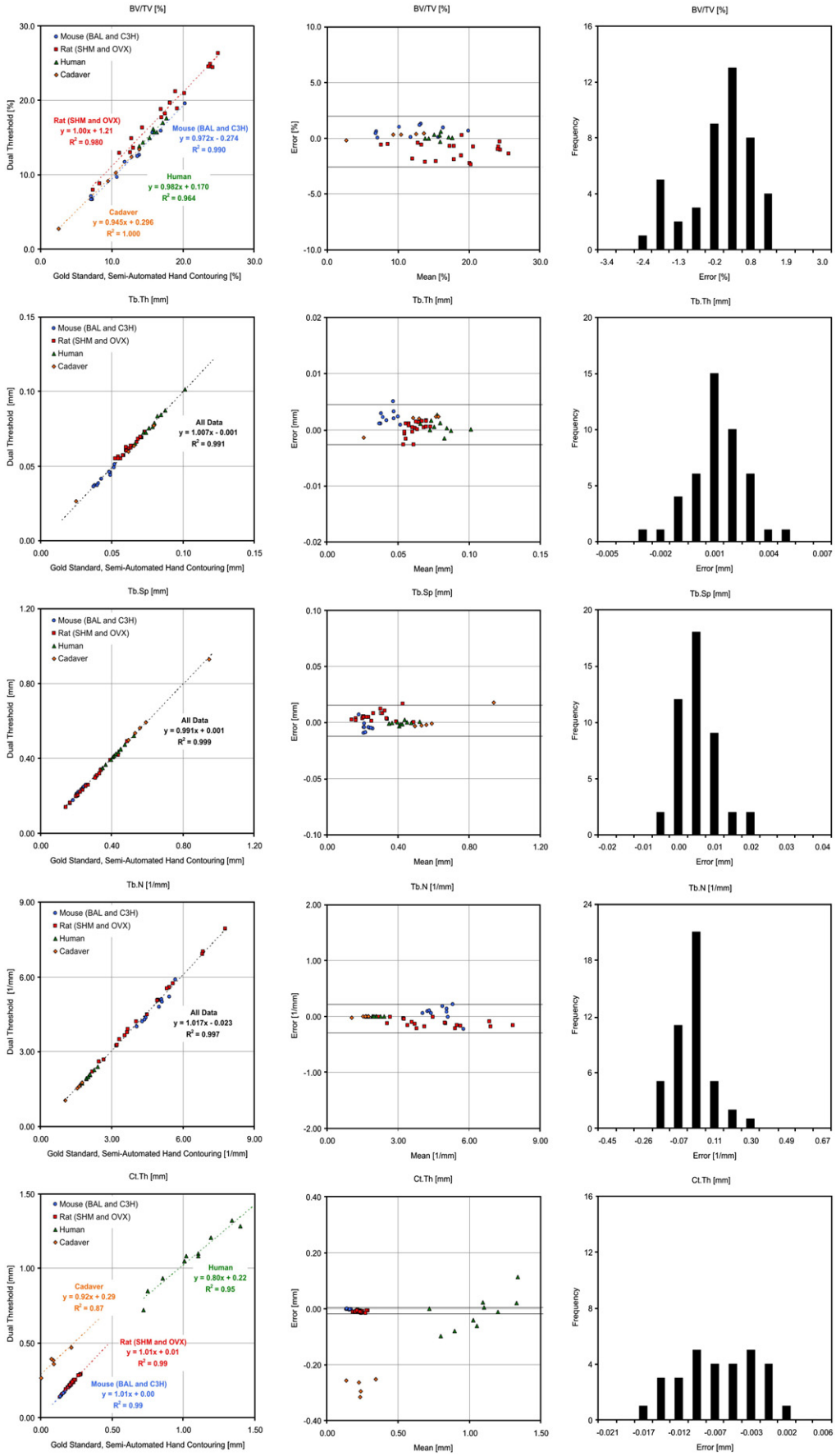

BV/TV [%]
Mouse (BAL and C3H)
Rat (SHM and OVX)
Human
Cadaver
Rat (SHM and OVX)
y = 1.00x + 1.21
R² = 0.980
Mouse (BAL and C3H)
y = 0.972x - 0.274
R² = 0.990
Human
y = 0.982x + 0.170
R² = 0.964
Cadaver
y = 0.945x + 0.296
R² = 1.000
Dual Threshold [%]
Gold Standard, Semi-Automated Hand Contouring [%]
Error [%]
Mean [%]
Frequency
Tb.Th [mm]
All Data
y = 1.007x - 0.001
R² = 0.991
Dual Threshold [mm]
Gold Standard, Semi-Automated Hand Contouring [mm]
Error [mm]
Mean [mm]
Tb.Sp [mm]
All Data
y = 0.991x + 0.001
R² = 0.999
Tb.N [1/mm]
All Data
y = 1.017x - 0.023
R² = 0.997
Dual Threshold [1/mm]
Gold Standard, Semi-Automated Hand Contouring [1/mm]
Error [1/mm]
Mean [1/mm]
Ct.Th [mm]
Cadaver
y = 0.92x + 0.29
R² = 0.87
Human
y = 0.80x + 0.22
R² = 0.95
Rat (SHM and OVX)
y = 1.01x + 0.01
R² = 0.99
Mouse (BAL and C3H)
y = 1.01x + 0.00
R² = 0.99

Table 3
Estimates of agreement of masked region based on the linear regression approach for the dual threshold method and gold standard

Measurement	Dataset	Parameter	Estimate	*P*-value	R^2
Tb.TV	Mouse	Slope	0.982±0.062	0.782	0.97
		Intercept	−0.034±0.082	0.691	
	Rat	Slope	1.005±0.061	0.942	0.94
		Intercept	0.972±1.507	0.527	
	Human	Slope	0.986±0.005	0.031 [a]	1.00
		Intercept	−0.208±0.030	0.000 [b]	
	Cadaver	Slope	0.959±0.029	0.246	1.00
		Intercept	−0.341±0.113	0.056	
Ct.TV	Mouse	Slope	1.092±0.180	0.624	0.99
		Intercept	−1.072±2.022	0.610	
	Rat	Slope	0.981±0.065	0.778	0.93
		Intercept	−0.773±1.143	0.507	
	Human	Slope	0.864±0.030	0.002 [a]	0.99
		Intercept	0.159±0.030	0.001 [b]	
	Cadaver	Slope	0.916±0.193	0.691	0.88
		Intercept	0.249±0.016	0.001 [b]	

Significance is determined from a two-tailed Student's *t*-test with the null hypothesis slope = 1 and intercept = 0.
[a] Slope is significantly different from one ($\alpha = 0.05$).
[b] Intercept is significantly different from zero ($\alpha = 0.05$).

1.4% for the rat and cadaver datasets, and cortical volume varied by less than 3.0%. A typical processing time for the dual threshold algorithm applied to a dataset containing 212 slices and file size of 120 MB was less than 10 min on a personal computer (2.91 GHz Athlon processor, 1.43 GB RAM) with connectivity filters implemented in two-dimensions (2D). The algorithm performed significantly faster with 2D connectivity applied to the 3D volume and generated the exact same results as with 3D connectivity in the typical case.

Discussion

A fully-automatic image analysis algorithm was described based on a dual threshold that allows efficient and reliable extraction of cortical and trabecular compartments for in vivo studies. The proposed method was shown to be robust through the successful segmentation of all 45 datasets, including those with highly thinned cortical shell. Problems associated with operator dependencies (precision and bias) are avoided and high consistency between datasets is ensured because the method is fully-automated and inputs are kept constant for similar datasets. The algorithm allows for some adjustment of the segmentation through the threshold inputs, which affect the size of the periosteal and endosteal surfaces of the cortical shell in the output mask. From a practical standpoint, the algorithm can be easily implemented since it makes use of simple image analysis tools.

An indirect result of this study is that it provides insight into operator bias and the sensitivity of morphological parameters to segmentation. Although not significant at the 95% confidence level, the rat data exhibited a constant bias that could be attributed to the hand contouring of the operator, who excluded highly dense (plate-like) trabeculae from contours of the subchondral plate region. This resulted in a constant bias in measurements of BV/TV, and Ct.Th to a lesser degree. These same parameters were sensitive to segmentation of the human and cadaver data, but were affected by a proportional bias. The differences in BV/TV for the two methods may be attributed to: (1) clipping of cortical shell and (2) inclusion of partial trabeculae as part of the cortex for the gold standard. The bias for Ct.Th was not constant because the gold standard (patient analysis) may be erroneous due to: (1) loss of thin cortical shell, (2) inclusion of thick trabeculae as cortex, and (3) smoothed over cortical porosity. Interestingly, Tb.Th, Tb.Sp, and Tb.N were not affected by the differences in segmentation and, therefore, are evidently more robust parameters for inter-study comparisons.

The success of the proposed method can be attributed to its underlying assumption, which is independent of cortical and trabecular thickness, and is most clearly demonstrated by its ability to extract highly thinned cortices of the cadaver datasets. Thin regions are successfully extracted because the method assumes that there is a cortex that fully surrounds the trabecular region. Segmentation problems with the gold standard (patient analysis) were most prominent when the underlying assumption that cortical shell is thick compared to trabeculae was poor. For example, the average ratio of cortical to trabecular thickness for the cadaver radii, which had the most severe segmentation problems, was only 6.7±1.9 (using values from the dual threshold analysis). In comparison, the ratio was approximately double for the younger, healthy human radii and tibiae (12.7±1.9 and 14.0±1.6, respectively), where the gold standard performed better.

Careful consideration should be given to implementation of the dual threshold algorithm to avoid segmentation problems. First, the dataset must be pre-processed to extract the bone of interest to ensure that adjacent bones in the dataset are not joined to it during dilation operations, which would interfere with segmentation. This process is easily fully-automated using connectivity filters. Second, if datasets are cropped tightly to the bone, clipping can occur during dilation leading to erroneous

Fig. 5. The Bland–Altman method for comparison of morphological parameters calculated for segmentation by the gold standard versus the dual threshold method. Parameters are direct measurements for the mice and rats and semi-derived for the human and cadaver data. Correlation plots (left column) are graphed using the same scale on both axes to aid the reader in assessing agreement between the methods. Bland–Altman plots (middle column) and error histograms (right column) provide an estimate of systematic and random errors. Dotted lines on the Bland–Altman plot represent 95% confidence intervals for agreement between the methods. For all parameters except Ct.Th of the human and cadaver data, there was a high correlation between methods and regression slopes were close to unity with intercepts close to zero, indicating agreement between methods. However, BV/TV exhibited some user bias; this was most prominent for the rat data, which had a regression intercept offset from zero (left column), and shifted error distribution (middle and right columns). User bias was negligible for Tb.Th, Tb.Sp, and Tb.N. Agreement for Ct.Th of the human and cadaver data was low because of significant differences in segmentation by the two methods (Fig. 2). Errors were not normally distributed for these datasets because the gold standard systematically missed thin regions of cortical bone during segmentation and included thick trabeculae as part of the cortex.

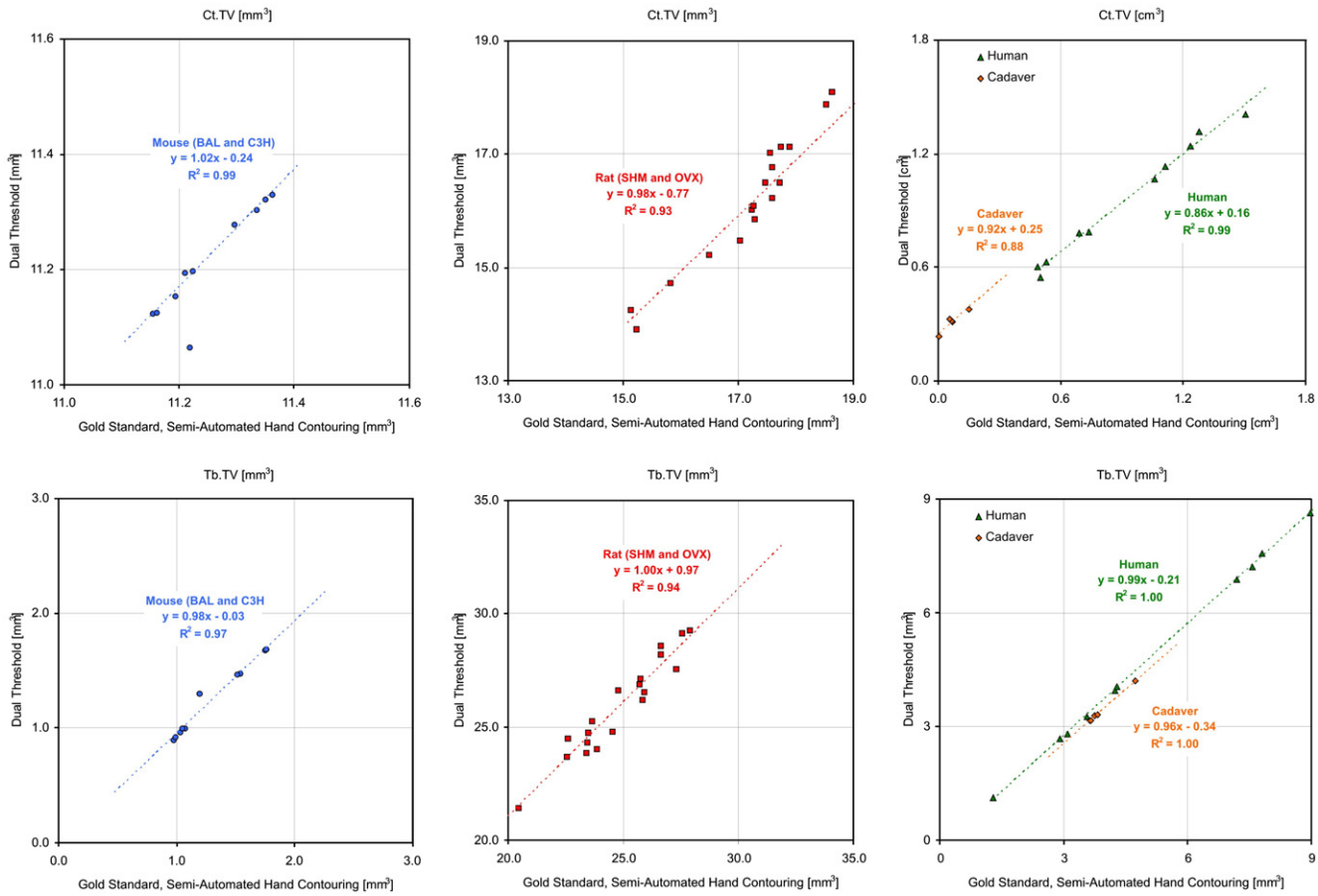


Fig. 6. Comparison of cortical volume (Ct.TV) and trabecular volume (Tb.TV) for the dual threshold method versus the gold standard. These parameters represent differences in the segmented regions, which subsequently affects morphological measures. Mouse data (left column) exhibits high agreement between methods for all but one dataset. The rat data (middle column) shows a constant bias between methods, with higher Ct.TV and lower Tb.TV for the gold standard. There is poor agreement between methods for Ct.TV of the human and cadaver data (right column), but fair agreement for Tb.TV. Despite variations in Ct.TV and Tb.TV, trabecular morphological parameters were in high agreement between the dual threshold method and gold standard.

results; however, this can be avoided by ensuring that the size of the dataset includes a border equal to the amount of dilation that will be used. Third, if Volkmann's canals are not fully closed after dilation, gaps will result in the trabecular region of the mask output. To correct this, the amount of dilation and erosion must be increased and, therefore, an option to override these parameters should be included. It should be noted, however, that high amounts of dilation and erosion can have the undesirable effect of smoothing out fine details. To overcome this problem, the periosteal threshold can be decreased to enlarge the masked region and avoid clipping the outer surface of the cortical shell when the mask is applied for analysis. Increasing the amount of dilation and erosion for reconnecting the marrow cavities will reduce roughness of the endosteal surface and generally will not pose segmentation problems unless extreme and unreasonable values are used. Dilation and erosion operations can be applied repeatedly with a small kernel size (5, 5, 1) to help preserve detail. Finally, preference of 2D or 3D connectivity can be included for efficient turnaround time of analyses. The algorithm is most robust when the connectivity filters are used in 3D, but processing time can be improved by 50% or more if 2D connectivity is calculated and typically yields identical results (data not shown). For only one of the 45 datasets tested, 3D connectivity was required because of an abnormal tunnel-like depression. In general, segmentation problems are obvious and, therefore, easy to identify and correct, making the dual threshold method simple to use.

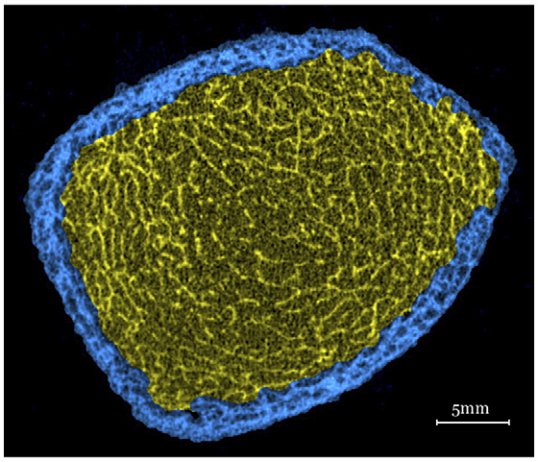


Fig. 7. Two-dimensional rendering of a human tibia exhibiting high cortical porosity segmented using the dual threshold method. In addition to the high porosity, a number of small perforations through the cortex can be observed, yet from a qualitative perspective the automated segmentation has provided a reasonable output.

Depending on the type of data being analyzed, it may be desirable to automate the selection of the threshold inputs from the histogram data. Although variation of thresholds must be performed cautiously to ensure data integrity within a study, automated selection could be advantageous for studies in which there is significant variability between subjects to ensure that all datasets are well segmented. Furthermore, automating the threshold selection would alleviate the need to determine input thresholds each time a new type of data is to be analyzed. On the other hand, if variability between subjects is low, as is typical of studies using murine models of osteoporosis, it may be preferable to use the same threshold inputs for all samples to ensure maximal consistency in segmentation of the bone compartments.

The dual threshold method was developed and is best suited for in vivo micro-CT studies in which cross-sections through a bone are analyzed; nonetheless, it should be possible to segment datasets of whole bones using this method. Some caveats to consider include potential problems with the growth plate region of long bones which when not fully closed may lead to erroneous results. It should be noted that the dual threshold method was intended for use with long bones and, therefore, setup of the connectivity filters may need to be modified slightly from the methods described here for irregular shaped bones, such as the vertebra.

Although not presented in this paper, the described dual threshold method has been tested on additional datasets. With slight modification to the method, over 100 whole mice vertebrae (19 μm resolution) have been successfully segmented. The method can also segment human data that contains high cortical porosity which presents a challenge to the gold-standard method (Fig. 7). The resulting cortical porosity could be calculated (it is not currently a standard output parameter) and further investigation is warranted. The measurement of cortical porosity rather than simply cortical thickness is of interest for studying patients undergoing treatment for osteoporosis (i.e., PTH).

Agreement of morphological parameters measured for the dual threshold method and gold standard depended on agreement of the initial segmentation, which emphasizes the importance of this process. The proposed method demonstrated a qualitative improvement over the gold standard, and therefore morphological parameters may be more accurate. However, this may pose problems for inter-site comparisons of morphological data unless the method is widely adopted. However, despite the limitations, the dual threshold algorithm offers a fully-automated alternative to the gold standard, semi-automated hand contouring, that can efficiently segment cortical and trabecular regions with accurate and repeatable results.

Acknowledgments

The authors would like to thank Dr. Andres Laib from Scanco Medical AG for his invaluable assistance with the image processing methodology of the gold standard patient analysis.

References

[1] Jiang Y, Zhao J, Liao EY, Dai RC, Wu XP, Genant HK. Application of micro-CT assessment of 3-D bone microstructure in preclinical and clinical studies. J Bone Miner Metab 2005;23:122–31 [Suppl].

[2] Laib A, Kumer JL, Majumdar S, Lane NE. The temporal changes of trabecular architecture in ovariectomized rats assessed by MicroCT. Osteoporos Int 2001;12(11):936–41.

[3] Müller R. The Zürich experience: one decade of three-dimensional high-resolution computed tomography. Top Magn Reson Imaging Oct 2002;13 (5):307–22.

[4] Rüegsegger P, Koller B, Müller R. A microtomographic system for the nondestructive evaluation of bone architecture. Calcif Tissue Int Jan 1996;58(1):24–9.

[5] Boutroy S, Bouxsein ML, Munoz F, Delmas PD. In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. J Clin Endocrinol Metab Dec 2005;90 (12):6508–15.

[6] Gasser JA, Ingold P, Grosios K, Laib A, Hammerle S, Koller B. Noninvasive monitoring of changes in structural cancellous bone parameters with a novel prototype micro-CT. J Bone Miner Metab 2005;23:90–6 [Suppl].

[7] Khosla S, Riggs BL, Atkinson EJ, Oberg AL, McDaniel LJ, Holets M, et al. Effects of sex and age on bone microstructure at the ultradistal radius: a population-based noninvasive in vivo assessment. J Bone Miner Res Jan 2006;21(1):124–31.

[8] Waarsing JH, Day JS, van der Linden JC, Ederveen AG, Spanjers C, De CN, et al. Detecting and tracking local changes in the tibiae of individual rats: a novel method to analyse longitudinal in vivo micro-CT data. Bone Jan 2004;34(1):163–9.

[9] Laib A, Häuselmann HJ, Rüegsegger P. In vivo high resolution 3D-QCT of the human forearm. Technol Health Care Dec 1998;6(5–6):329–37.

[10] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. Int J Comput Vis 1988;1(4):321–31.

[11] Gelaude F, Vander SJ, Lauwers B. Semi-automated segmentation and visualisation of outer bone cortex from medical images. Comput Methods Biomech Biomed Engin Feb 2006;9(1):65–77.

[12] Dufresne T. Segmentation techniques for analysis of bone by three-dimensional computed tomographic imaging. Technol Health Care Dec 1998;6(5–6):351–9.

[13] Schroeder W, Martin K, Lorensen B. The visualization toolkit an object-oriented approach to 3D graphics. 3rd ed. Uppersaddle River, NJ: Prentice Hall; 2003.

[14] Hildebrand T, Rüegsegger P. Quantification of bone microarchitecture with the structure model index. Comput Methods Biomech Biomed Engin 1997;1(1):15–23.

[15] Hildebrand T, Laib A, Müller R, Dequeker J, Rüegsegger P. Direct three-dimensional morphometric analysis of human cancellous bone: microstructural data from spine, femur, iliac crest, and calcaneus. J Bone Miner Res Jul 1999;14(7):1167–74.

[16] Laib A, Rüegsegger P. Comparison of structure extraction methods for in vivo trabecular bone measurements. Comput Med Imaging Graph Mar 1999;23(2):69–74.

[17] MacNeil JA, Boyd SK. Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. Med Eng Phys. 2007; in press.

[18] Njeh CF, Hans D. Instrument evaluation. In: Langton CM, Njeh CF, editors. The physical measurement of bone. Philadelphia, PA: Institute of Physics Publishing; 2004. p. 91–121.

[19] Linnet K. Necessary sample size for method comparison studies based on regression analysis. Clin Chem 1999;45(6):882–94.

[20] Magari RT. Statistics for laboratory method comparison studies. BioPharm 2002;5:28–32.