Medical Engineering and Physics 70 (2019) 19–28

Contents lists available at ScienceDirect

Medical Engineering and Physics

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

# Automated segmentation of cortical and trabecular bone to generate finite element models for femoral bone mechanics

Sami P. Väänänen[a,b,c,d,1,*], Lorenzo Grassi[e,1], Mikko S. Venäläinen[a,f], Hanna Matikka[b], Yi Zheng[g], Jukka S. Jurvelin[a], Hanna Isaksson[e]

[a] Department of Applied Physics, University of Eastern Finland, POB 1627, FIN-70211 Kuopio, Finland
[b] Department of Clinical Radiology, Diagnostic Imaging Center, Kuopio University Hospital, POB 100, 70029 Kuopio, Finland
[c] Department of Orthopaedics, Traumatology and Hand Surgery, Kuopio University Hospital, POB 100, FIN-70029 Kuopio, Finland
[d] Department of Medical Physics, Central Finland Central Hospital, Keskussairaalantie 19, FIN-40620 Jyväskylä, Finland
[e] Department of Biomedical Engineering, Lund University, BMC D13, 221 84 Lund, Sweden
[f] Turku Bioscience Centre, University of Turku and Åbo Akademi University, Tykistökatu 6, FIN-20520 Turku, Finland
[g] Department of Physics, Technical University of Denmark, Fysikvej, building 311, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

*Article history:*
Received 20 September 2018
Revised 16 May 2019
Accepted 23 June 2019

*Keywords:*
Finite element modeling
Automated segmentation
Femur
Bone
Isotopology
Surface strains

ABSTRACT

Finite element (FE) models based on quantitative computed tomography (CT) images are better predictors of bone strength than conventional areal bone mineral density measurements. However, FE models require manual segmentation of the femur, which is not clinically applicable. This study developed a method for automated FE analyses from clinical CT images. Clinical *in-vivo* CT images of 13 elderly female subjects were collected to evaluate the method. Secondly, proximal cadaver femurs were harvested and imaged with clinical CT ($N = 17$). Of these femurs, 14 were imaged with μCT and three had earlier been tested experimentally in stance-loading, while collecting surface deformations with digital image correlation. Femurs were segmented from clinical CT images using an automated method, based on the segmentation tool Stradwin. The method automatically distinguishes trabecular and cortical bone, corrects partial volume effect and generates input for FE analysis. The manual and automatic segmentations agreed within about one voxel for *in-vivo* subjects ($0.99 \pm 0.23$ mm) and cadaver femurs ($0.21 \pm 0.07$ mm). The strains from the FE predictions closely matched with the experimentally measured strains ($R^2 = 0.89$). The method can automatically generate meshes suitable for FE analysis. The method may bring us one step closer to enable clinical usage of patient-specific FE analyses.



## 1. Introduction

Osteoporosis is characterized by reduced bone mineral density (BMD) and inferior bone quality, thereby leading to an increased bone fracture risk [1,2]. Osteoporosis is typically diagnosed by measuring BMD of the hip or lumbar spine using dual-energy X-ray absorptiometry (DXA) [3]. BMD is most commonly used alone or in combination with epidemiological risk factors to predict the fracture risk of a patient. However, only less than one third of all low-trauma fractures are explained solely by BMD [4]. This is partly because 3D geometry and internal architecture of the bone are not fully accounted for in BMD measurements from DXA images. Subject-specific finite element (FE) models from CT images can explain at least 20% more of the total strength of the femur [5]. However, CT-based FE analyses usually need manual intervention from a trained engineer, especially during the image processing phase [6]. Therefore, prediction of fracture risk based on subject-specific FE models is not yet applicable in current clinical practice.

Automatic methods to extract the femoral shape from clinical images have been developed [7–10] with the aim to increase the automation level of the FE modelling procedure. These solutions allow to automatically extract the periosteal surface of the femur with an error in the order of the pixel size. The majority of these methods rely on a statistical model of shape and density distribution. They require extensive training set to ensure generality in case of large variability of femoral anatomy as a function of age [11] and ethnicity [12]. Besides, all these methods can identify the

*List of abbreviations:* BMD, Bone mineral density; CPM, Cortical profile modelling; DIC, Digital image correlation; DSC, Dice similarity coefficient; DXA, Dual-energy X-ray absorptiometry; FEM, Finite element modeling; HU, Hounsfield unit.

* Corresponding author at: Department of Applied Physics, University of Eastern Finland, POB 1627, FIN-70211 Kuopio, Finland.
*E-mail addresses:* sami.vaananen@uef.fi (S.P. Väänänen), lorenzo.grassi@bme.lth.se (L. Grassi), mikko.venalainen@utu.fi (M.S. Venäläinen), hanna.matikka@kuh.fi (H. Matikka), yizhe@fysik.dtu.dk (Y. Zheng), jukka.jurvelin@uef.fi (J.S. Jurvelin), hanna.isaksson@bme.lth.se (H. Isaksson).

[1] The first two named authors contribute equally to this work.

https://doi.org/10.1016/j.medengphy.2019.06.015
1350-4533/

periosteal surface of the femur, but do not identify the endocortical surface, i.e., they cannot calculate the cortical thickness.

Recent studies showed that the cortical bone plays a crucial role in the fracture resistance of the proximal femur [13,14]. In addition, cortical and trabecular bone both contribute to proximal femur failure load [15], whilst deficits in trabecular and cortical bone density and structure independently contribute to fracture risk [16]. Therefore, the inclusion of separate contributions of trabecular and cortical bone to fracture resistance is one of the key areas where FE models can perform better than current clinical methods. However, the cortical thickness in the proximal femur is typically in the order of the pixel size of clinical CT images. Thus, extracting the cortical thickness with conventional manual segmentation is not feasible. Therefore, computational approaches based on image deblurring [17,18] or Cortical Profile Modelling (CPM) [19–21] have been proposed. The latter allows to identify periosteal and endocortical surfaces with sub-voxel accuracy by correcting the imprecision arising from the bone boundary blurring effect due to the moderate image resolution, known as "partial volume effect". While recent studies presented how to automate the CPM technique [22,23], a manual initialization is still needed.

The aim of the current study is to present a method for automated segmentation of the proximal femur including generation of FE meshes that are directly suitable for bone strength analysis. Specifically, we aim to ensure that the method can:

- Distinguish the periosteal and endocortical surfaces with sub-pixel accuracy
- Provide isotopological meshes for easy boundary condition definition and creation of statistical models
- Directly generate volume meshes for FE simulations of bone mechanics.

Finally, we validate the accuracy of the automatic segmentation, as well as the accuracy of the generated FE models against previously collected experimental measurements [24].

## 2. Material and methods

A method to automatically segment proximal femurs from clinical CT images and generate FE models for analysis of femoral mechanics was developed. The method pre-processes the images if needed, separates the femur from the pelvis, obtains the periosteal and endocortical surfaces of the femur, and warps a FE mesh to the obtained result (Fig. 1).

The validation was performed both using *in-vivo* and *ex-vivo* dataset, as depicted in Fig. 2. The segmentation accuracy was evaluated with an *in-vivo* dataset by comparing the outcome of the proposed automatic method with conventional semi-automatic segmentation. The *ex-vivo* dataset consisted of a set of cadaver femurs imaged both with clinical CT and μCT. μCT images were used as gold standard when segmentation accuracy of clinical CT was evaluated. Lastly, a second *ex-vivo* dataset comprised three femurs, for which experimental measurements were available from a previous study [24]. These were used to evaluate the actual accuracy of the generated FE models in predicting femoral strains.

### 2.1. Material

#### 2.1.1. In-vivo dataset

The hips of 13 female patients (age range 69–78, $74.5 \pm 2.7$ years) were imaged with a clinical CT system (Philips Precedence 6P, in-plane resolution between 0.7 and 0.8 mm, slice separation 2.0 mm, tube voltage 120 kV and tube current 100 mAs) [25,26]. Five of the patients had a previous hip fracture on one side, of which two had intramedullary nails. Right femurs were analyzed, with the exception that left femurs were analyzed for two subjects with intramedullary nails on right side. The study was approved by the local ethics committee (Kuopio University Hospital Ethical Committee, permission 80/2008).

#### 2.1.2. Ex-vivo dataset

Cadaver proximal femurs from 17 donors were collected (14 females, age $70 \pm 17$ (range 22–88), 3 males, ages 22, 58 and 58). The use of cadaveric tissue was approved by the National Authority for Medicolegal Affairs (TEO, 5783/04/044/07). None of the cadavers had any pre-existing conditions known to affect bone metabolism. All the 17 femurs were imaged with Siemens Somatom AS clinical CT-scanner (pixel size of 0.4–0.5 mm, slice separation of 0.6 mm, tube voltage 120 kV, tube current 210 mAs and $CTDI_{vol}$ ~16). The 14 female femurs were additionally imaged with a μCT system (Nikon XT H 225 scanner, 200 mAs, 100 kVp isotropic voxel size of 0.052–0.060 mm). The 3 male femurs lacked μCT images, as they had earlier been tested mechanically until failure in stance-loading configuration, and the deformations at the anterior surface were recorded and analyzed using digital image correlation technique (DIC) [24].

### 2.2. Methods

#### 2.2.1. Pre-processing of the CT images

Based on our experience, CPM [20] performs best with images which are reconstructed using medium or soft (neutral or low-pass) kernels. However, the CT images of the 3 male femurs in the *ex-vivo* dataset were reconstructed only with hard kernel, while the original sinograms were no longer available at the time of the study. Thus, we processed these three images with a Gaussian filter, which transformed the images reconstructed using a hard kernel into images reconstructed with a soft kernel, i.e., from high-pass filtered images to low-pass filtered images. We determined Gaussian filter parameters, i.e., filter size and filter's standard deviation based on the other 14 femurs of the *ex-vivo* dataset, for which reconstructions using both types of kernels were available. The parameters were determined using a genetic algorithm, which minimized the sum of absolute difference between the images reconstructed with soft kernel and Gaussian filtered hard kernel images.

Additionally, the three male cadaver femurs in the *ex-vivo* dataset have been imaged in air, whereas CPM assumes close-to-soft tissue Hounsfield unit (HU) values around bone. This was addressed by setting the minimum HU value to −107, based on the average HU value of ice (in which the rest of our samples were imaged) in the other subset of cadaver femurs.

#### 2.2.2. Template surface mesh and template volume mesh

One cadaver femur from another bone set [26] having an average size and shape, as based on visual evaluation and geometrical measurements, was chosen as the template femur. A triangular surface mesh was generated from the periosteal surface using Stradwin software (version 5.3, University of Cambridge, UK), which implements CPM [20,21]. Thereafter, Stradwin analysis captured both the periosteal and endocortical surfaces. The resulting periosteal surface and cortical thickness were corrected using the algorithm described in section 2.2.4. The generated periosteal and endocortical surfaces and cortical thickness were combined to form a template surface mesh (27,543 vertices and 55,082 triangle elements). The template tetrahedral volume mesh (25,958 vertices and 138,794 elements resulting in average edge length of $2.18 \pm 0.51$ mm) was obtained based on the template surface mesh using Abaqus (v6.14, Dassault Systèmes, Providence, RI, USA).

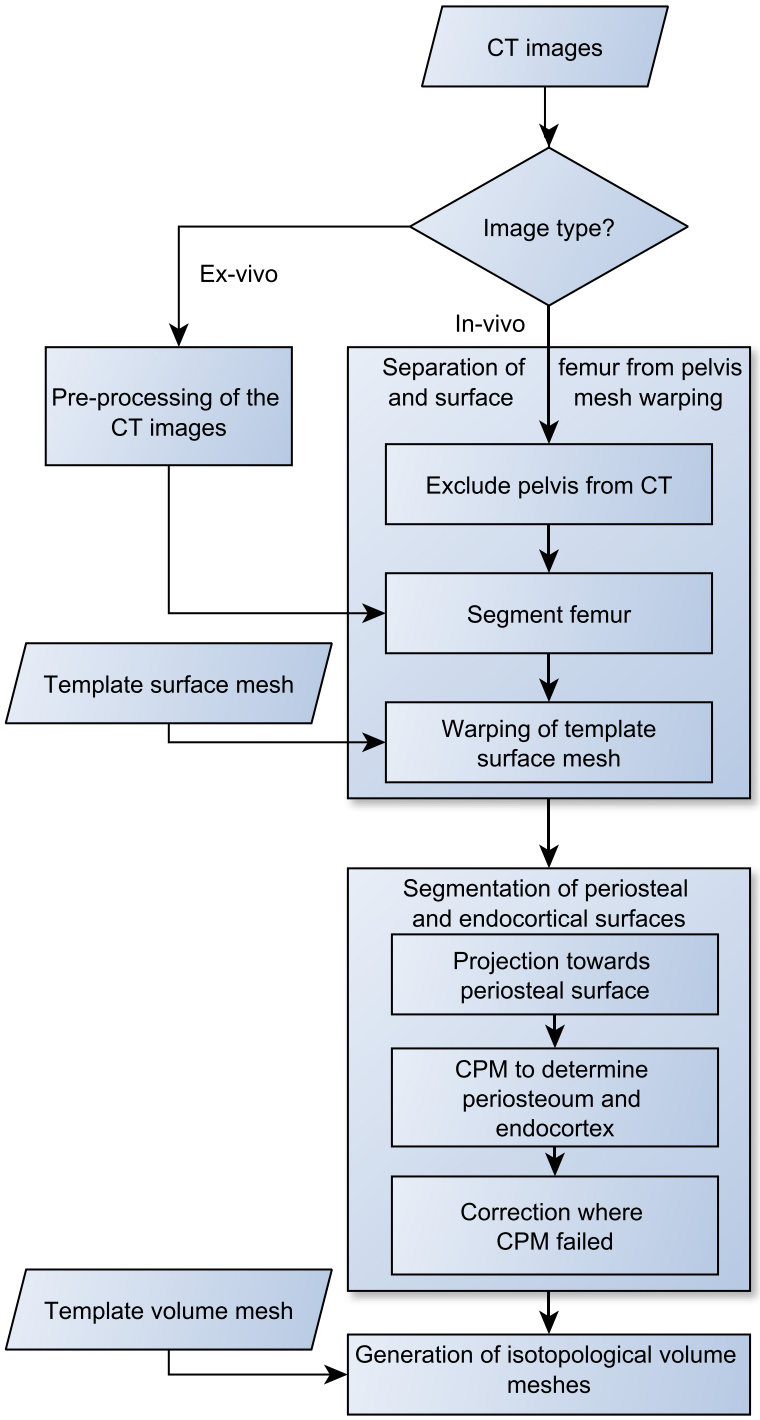


**Fig. 1.** Flowchart of the presented automatic segmentation and FE model creation from clinical CT images.

*2.2.3. Separation of femur from pelvis and surface mesh warping*

The first step to process the clinical images was to exclude the pelvis from the images. This was achieved by first roughly segmenting the CT images with multi Otsu's method [27]. The hip and femur from desired side was determined from the most inferior CT slice. Objects in the 3D binary image that were not connected to the femur in this slice were excluded. The first axial slice superior to the femoral neck was then determined by finding the first axial slice, going from distal to proximal, where the femoral neck is no longer visible and two distinct bony parts (i.e., the femoral head and the greater trochanter) were visible in the axial slice [28]. The center of the femoral head and its radius were determined from this slice [28]. The hip joint space location and the surface of the femoral head were determined from 17 radial intensity curves starting from the center of the head and directed towards the acetabulum, and towards lateral, inferior and posterior directions with even distribution. The intensity curves were generated similarly as done in CPM by using an in-house Matlab code. Joint space was determined from the derivative of these curves by finding the local minimum before the highest ascent, which localized the acetabular cortex, and the femoral head surface by finding the highest descent at lateral, inferior and posterior directions. A sphere was fitted to the points determined from the joint space and from femoral surface. The objects medially and superiorly from

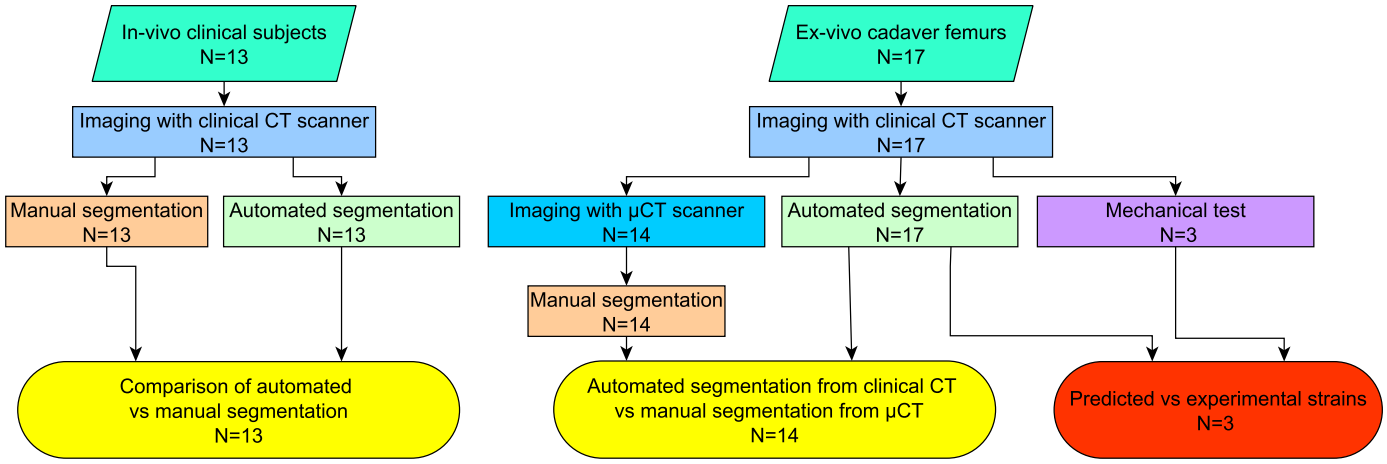


**Fig. 2.** Flowchart of the validation of the automated segmentation pipeline. CT scans of *in-vivo* subjects were used to directly compare the performance of manual and automated segmentation within the same clinical datasets. *Ex-vivo* cadaver femurs were used to compare the results of the automated segmentations from CT with the manual segmentation from uCT, as well as to validate the strain prediction accuracy against previously collected experimental measurements [24].

this sphere were excluded, which resulted in a CT image that only included the femur.

The femoral surfaces were then segmented from the CT image, which was filtered using cubic Gaussian filter having standard deviation 1, by applying the multi Otsu's method. The largest object (i.e., the femur) was chosen and it was closed with a spherical mask of 5 voxels, thus filling the bone marrow cavity and other holes. An initial surface mesh was generated using marching cubes algorithm.

The template mesh was then registered to the target femur in a multi-step procedure. First, the template surface mesh was registered to the initial surface mesh using affine registration [29,30]. Second, the minor trochanter, whose orientation can vary greatly from subject to subject [31], was aligned to target bone as follows. The minor trochanter region of the template bone was affinely registered to the target bone [29,30]. Then, the template bone was deformed non-linearly using thin-plate splines (TPS) by adjusting the tip of minor trochanter, while constraining the rest of the bone with 30 predefined automatically set landmarks [26]. Thirdly, non-rigid ICP algorithm with 7 iterations [32,33] was used to align the surface of template bone non-linearly on target bone. Fourthly, 2000 randomly distributed corresponding surface vertex points between target and template were used to define TPS, which warped the template surface mesh onto the target bone. This resulted in the warped template surface mesh.

### 2.2.4. Segmentation of the periosteal and endocortical surfaces of femur

The segmentation of periosteal and endocortical bone was performed automatically via the scripting interface of Stradwin. This process was performed by calling Stradwin in two consecutive iterations. The first iteration projected the warped template surface mesh accurately to the target femur's periosteum. This was aimed at accurately detecting bone surface in the acetabular region. The second iteration performed the actual CPM to detect periosteal and endodocortical surfaces. Before the second iteration, the surface normals were adjusted at the diaphysis to point towards the center of the bone, i.e., to intersect with the femur's 3D skeleton image [34], since the endocortical surface mesh may distort at regions where the diaphysis shape diverges from a cylindrical shape, such as at the linea aspera (as illustrated in Fig. 4(b) of the work by Zhang et al. [22]). Periosteal and endocortical surfaces were then identified. Since CPM results depend on the chosen search length, Stradwin was run with three different search line lengths: 10, 19 and 22 mm. This reduced the number of cases where the search resulted in errors. The width of each search line was set to 4 pixels (corresponding to ~4 mm).

The resulting output of the three analyses were then post-processed and re-assembled in Matlab. The analysis outcome from the search line length of 10 mm was used for the femoral head region. In the other regions, the estimation of the cortex was considered successful if the estimated standard deviation of the cortical thickness was $< 10$ mm. The two most similar successful analyses were chosen, and among them the one with the smallest standard deviation was used. Finally, the code looked for any residual irregularities in the obtained mesh. These were identified as vertices where cortical thickness differed over one mm from the neighboring vertices, and were smoothed by adjusting them to the median thickness of their neighborhood.

### 2.2.5. Generation of isotopological volume meshes

The template volume mesh was warped to the shape of the target bone by using TPS according to the detected periosteal and endocortical surfaces meshes. Distorted elements, if any ($<$100 in all subjects, as identified by an automatic script that calls Abaqus to check element quality), were corrected iteratively. The correction was done for each element by determining the vertex whose distance to the opposing face was smallest. This vertex was moved towards opposing face's negative normal by a distance 1/3 of the distance between the vertex and its neighboring elements' vertices. This was repeated until no element had insufficient quality according to Abaqus. The final element quality parameters were calculated and reported.

### 2.2.6. Validation

The validation of the method described above was performed according to the scheme depicted in Fig. 2. The automatic segmentation method was performed on the clinical CT images for both the *in-vivo* and the *ex-vivo* datasets.

Then, for the *in-vivo* dataset, the automatic segmentation of the clinical CT images was compared to the segmentation of the same clinical CT images as performed via conventional semi-automatic segmentation using Mimics (version 15.1, Materialize, Belgium). Differences in the periosteal surfaces and cortex thicknesses were compared by calculating mean absolute surface differences, dice similarity coefficient (DSC) [35], sensitivity and specificity for segmented whole bone, cortical bone, and trabecular bone regions.

For the *ex-vivo* dataset, the automatic segmentation of the clinical CT images for the 14 female femurs was compared to the

**Table 1**
Agreement between manually and automatically segmented femoral CT images of clinical subjects. The data presents mean ± standard deviation of 13 clinical subjects.

| | Average surface distance (mm) | Volume difference ($cm^3$) | Dice similarity coefficient | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Whole Femur | $0.99 \pm 0.23$ | $-16.5 \pm 5.8$ | $0.93 \pm 0.02$ | $0.88 \pm 0.28$ | $0.99 \pm 0.01$ |
| Trabecular | $0.99 \pm 0.20$ | $-8.9 \pm 5.8$ | $0.91 \pm 0.02$ | $0.88 \pm 0.04$ | $0.96 \pm 0.04$ |

**Table 2**
Agreement between automatically segmented clinical CT images and manually segmented µCT images of femoral cadaver samples. The data presents mean ± standard deviation of 14 cadaver femur samples.

| | Average surface distance (mm) | Volume difference ($cm^3$) | Dice Similarity coefficient | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Whole Femur | $0.21 \pm 0.07$ | $-1.68 \pm 2.99$ | $0.98 \pm 0.01$ | $0.98 \pm 0.02$ | $0.99 \pm 0.01$ |
| Trabecular | $0.35 \pm 0.05$ | $-1.43 \pm 1.82$ | $0.97 \pm 0.01$ | $0.97 \pm 0.01$ | $0.98 \pm 0.01$ |

manual segmentation of the corresponding µCT-images. This was in order to evaluate the level of sub-pixel accuracy achieved by the proposed automatic segmentation method. The same accuracy metric as for the *in-vivo* dataset was used for the comparison.

As for the 3 male femurs in the *ex-vivo* dataset, finite element analyses mimicking stance load were automatically formed from the automatically generated FE models [36]. Briefly, a BMD phantom (Model 3CT, Mindways Inc.) was imaged together with the bones, which enabled us to convert HU values into BMD values. BMD values were converted to Young's modulus using the relationships proposed by Schileo et al. [37]. We used density-elasticity relation $E(\text{GPa}) = 6.850 * \rho_{\text{BMD}}^{1.49}(\frac{\text{g}}{\text{cm}^3})$ from Morgan et al. [38], and Poisson's ratio 0.3. The code to generate the FE analyses, which is also automatic, adjusted boundary conditions patient-specifically, i.e., it adjusted the force magnitude and direction on the femoral head and the set of constrained nodes on the shaft. The code also generated input files for Abaqus and launched the Abaqus solver. These three male femurs were earlier tested experimentally in single-leg-stance loading until failure [24], with simultaneous recording of the surface deformations using the DIC technique with high temporal and spatial resolution. The predicted surface strains from the FE analyses were thus compared via robust linear regression analysis to the experimental measurements to evaluate the accuracy of the automatically generated FE models.

## 3. Results

### 3.1. In-vivo dataset

The surface and DSC difference between the automatic and manual segmentation of clinical CT images for the *in-vivo* dataset were $0.99 \pm 0.23$ mm and $0.93 \pm 0.02$, respectively (Table 1). The segmentation was the most accurate at the femoral neck and trochanteric areas, and less accurate at regions where the femur overlaps with the acetabulum (Fig. 3).

### 3.2. Ex-vivo dataset

When automatic segmentation of clinical CT images of the *ex-vivo* dataset was compared to manual segmentation of their corresponding µCT images, the surface difference decreased to $0.21 \pm 0.07$ mm, while the DSC agreement increased to $0.97 \pm 0.01$ (Table 2). The agreement between segmentation techniques was equally high at all anatomical locations (Fig. 4).

As for the three male femurs in the *ex-vivo* dataset, the mesh quality parameters were adequate, and all elements were within acceptance limit of Abaqus (Table 3) with only few elements indicating suboptimal element quality. The correlation between the calculated surface strains and the experimentally measured strains for the femurs was $R^2 = 0.89$ for first principal strains in three bones pooled (Fig. 5). This agreement is compared with agreement between the manual FE model from our previous study [36] and the experiment (Table 4). Anatomical locations of the strains loci agreed well between simulations and experiments both in tensile and compressive regions (Figs. 6 and 7).

**Table 3**
Mesh quality parameters calculated in Abaqus for the three automatically generated FE meshes. Table reports the average, 95% and 99% limits and the worst parameter value. Abaqus accepts elements within the acceptance limit.

| Shape metric | | Model | | | Acceptance limit in Abaqus |
|---|---|---|---|---|---|
| | | Bone #1 | Bone #2 | Bone #3 | |
| Shape factor | Average | 0.66 | 0.65 | 0.65 | > 0.0001 |
| | 95% | 0.36 | 0.34 | 0.35 | |
| | 99% | 0.22 | 0.20 | 0.20 | |
| | Worst | 0.007 | 0.006 | 0.006 | |
| Min tetra angle | Average | 38.0 | 37.5 | 37.8 | > 5 |
| | 95% | 27.0 | 26.1 | 26.6 | |
| | 99% | 19.7 | 19.0 | 19.3 | |
| | Worst | 7.1 | 7.7 | 7.0 | |
| Max tetra angle | Average | 88.8 | 89.3 | 89.1 | < 170 |
| | 95% | 107.4 | 108.5 | 108.1 | |
| | 99% | 117.8 | 119.4 | 118.8 | |
| | Worst | 158.2 | 156.1 | 158.8 | |
| Aspect ratio | Average | 1.7 | 1.7 | 1.7 | < 10 |
| | 95% | 2.2 | 2.2 | 2.2 | |
| | 99% | 2.9 | 3.0 | 2.9 | |
| | Worst | 6.9 | 7.0 | 6.9 | |

## 4. Discussion

We developed an automated pipeline to segment cortical and trabecular bone from *in-vivo* clinical CT images of the human femur, and to generate FE models for simulation of femur mechanics. The automated segmentations agreed well with manual segmentations on the *in-vivo* dataset (DSC = 0.93), especially in regions which are of highest interest for analysis of fracture mechanics, i.e., femoral neck and trochanter regions (Fig. 3). The segmentation accuracy was lower in the region of the acetabulum, since the narrow joint space between the femur and the pelvis was partly blurred due to the resolution of clinical CT images. This problem seems to be common to most segmentation methods [8,9,22]. Yet, the segmentation accuracy was sufficient for FE model generation also in that region and did not in any case affect the ability of the models to accurately predict the mechanical behaviour of bone, as indicated by the validation against experimentally measured principal strains. It has been proposed that most of the load is transferred through the centre of the femoral head [39,40], which may explain why the lower segmentation accuracy did not significantly affect our results. Besides, the most fracture-prone regions are located at the femoral neck, which was segmented with high accuracy (Fig. 3).

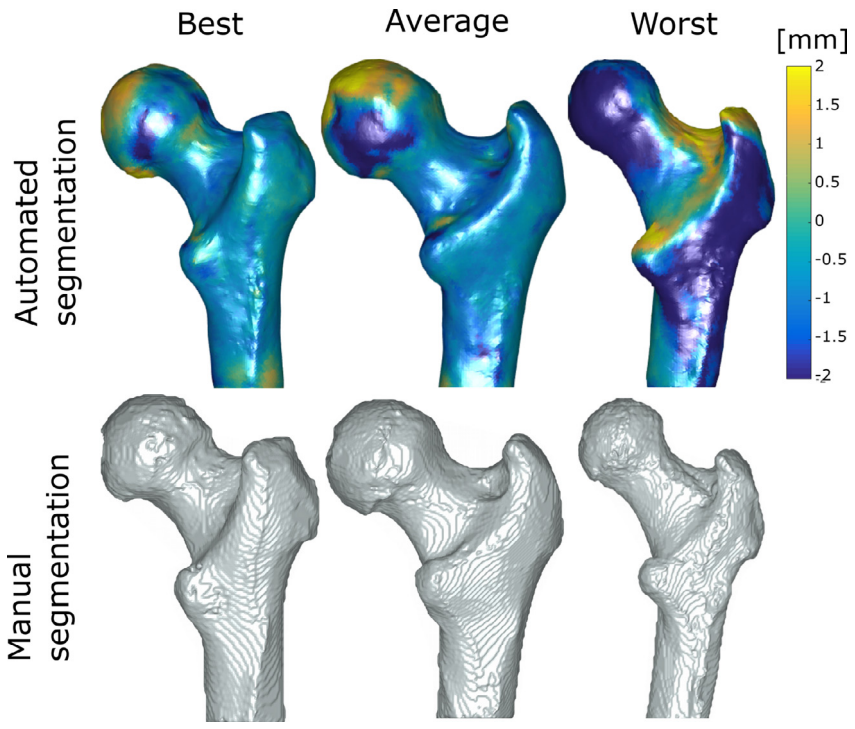


**Fig. 3.** The best, average and worst correspondence measured by surface agreement between the automatic and manual segmentations of clinical subjects. The colors on the automatically segmented femurs indicate the distance between the surfaces of the automated and manual segmentation.

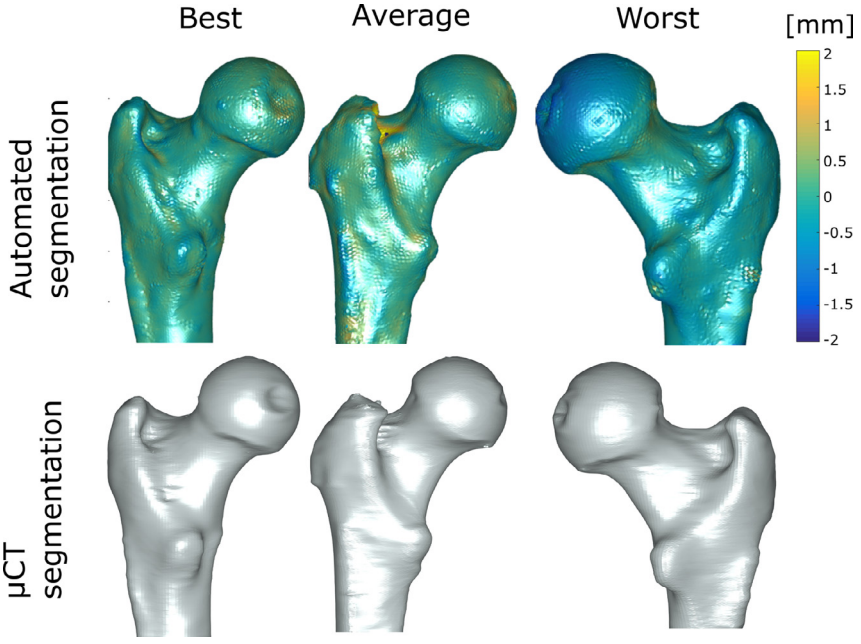


**Fig. 4.** The best, average and worst correspondence measured by surface agreement between the automatically segmented clinical CT and manually segmented µCT of cadaver femurs. The color on the automatically segmented femur indicates the distance between the surfaces of the automated clinical CT segmentation and manual µCT segmentation.

**Table 4**
Comparison of the strain prediction accuracy obtained using manually derived FE models [36] and using the automatically generated FE models proposed in this study.

| | Experimental data vs manually derived FE models (all data is from [36]) | | | | Experimental data vs new automatically derived FE models | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ | slope | intercept (µε) | NRMSE (%) | $R^2$ | slope | intercept (µε) | NRMSE (%) |
| Bone 1 | 0.92 | 0.92 | 144 | 10 | 0.87 | 0.91 | 174 | 9 |
| Bone 2 | 0.94 | 0.97 | 174 | 11 | 0.89 | 0.93 | 90 | 7 |
| Bone 3 | 0.95 | 1.01 | 79 | 11 | 0.90 | 1.06 | 86 | 7 |
| Pooled 3 bones | 0.94 | 0.96 | 133 | 9 | 0.89 | 0.96 | 121 | 6 |

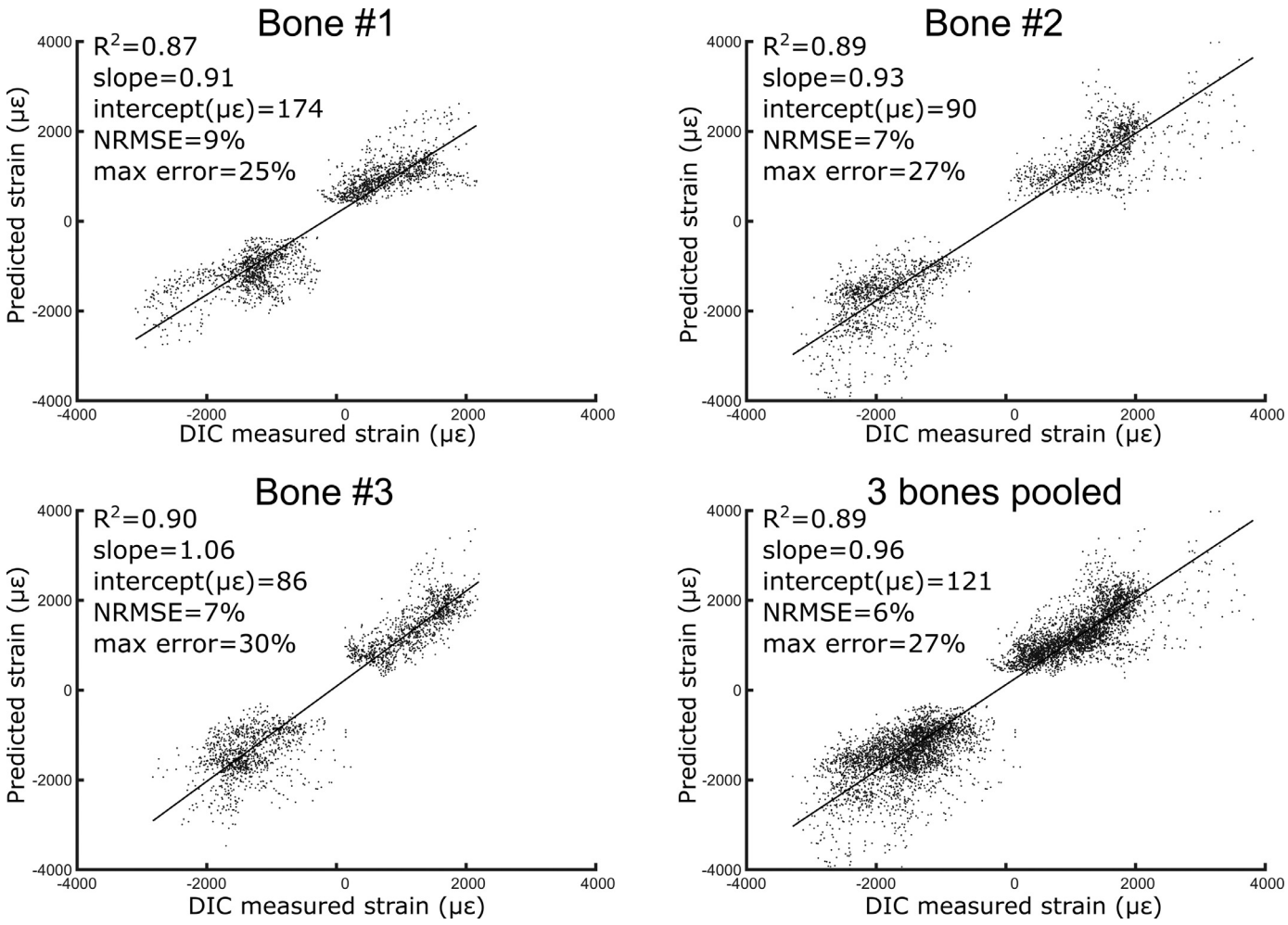


**Fig. 5.** Experimental principal tensile and compressive strains versus the predicted ones from automated FE analysis.

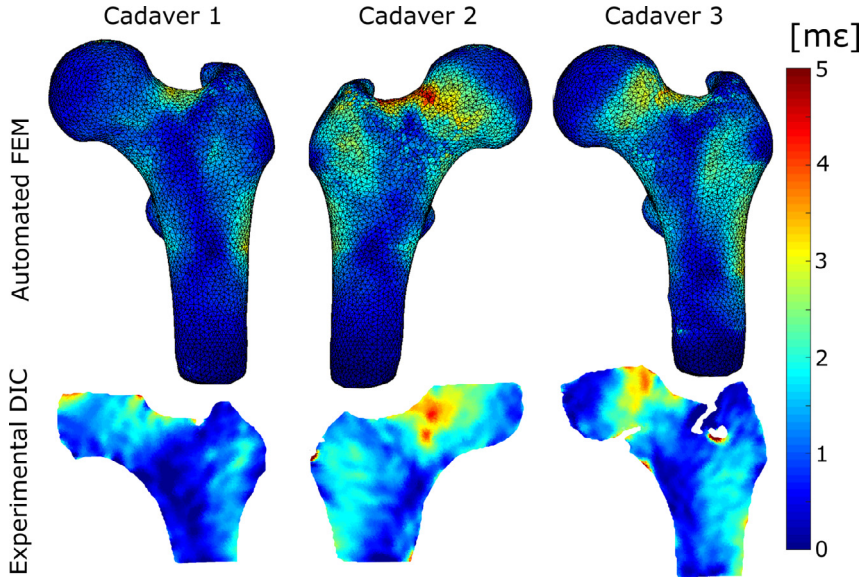


**Fig. 6.** First principal strain fields under a load four times the body weight for FE simulations and experimentally measured strains ($N = 3$). The experimental results are collected from our previous study [24].

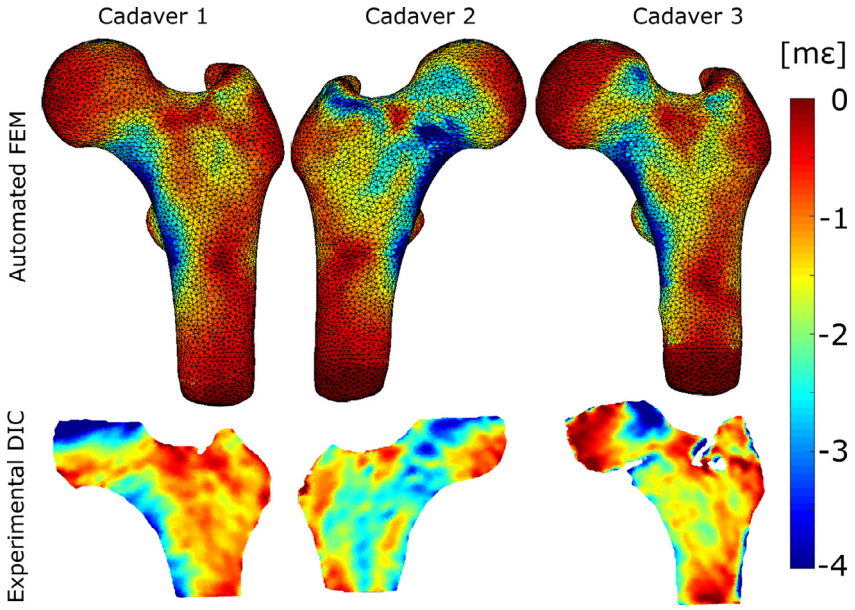


**Fig. 7.** Third principal strain fields under a load four times the body weight for FE simulation and experimentally measured strains ($N = 3$). The experimental results are collected from our previous study [24].

When the same automated segmentation pipeline was tested with cadaver femurs and compared to manual segmentation of μCT images, the DSC agreement between segmentations increased to 0.98. This indicates that the most significant error sources in segmentation of *in-vivo* CT images are the hip joint region, and if present in the image and in the used method, moderate image resolution and manual segmentation process.

Specificity was higher than sensitivity (0.99 vs. 0.88) for the *in-vivo* dataset. This suggests that manually segmented clinical CT images are enlarged compared to automatically segmented images, most likely since conventional manual segmentation lacks correction of partial volume artefacts. This speculation is corroborated by the fact that the specificity was as high as the sensitivity (0.98 vs. 0.99) when the automatic segmentation of clinical CT images was compared to μCT-based segmentation. Thus, the proposed automatic segmentation method can reach a sub-pixel accuracy in femoral shape extraction, both for the periosteal and the endocortical surface. Previous studies showed that the adopted deblurring technique is robust against changes in imaging parameters [20]. Besides, the present study tested the automatic segmentation algorithm on two datasets (*in-vivo* and *ex-vivo*) which were acquired using different CT scanners and with different scanning parameters, thus corroborating the robustness of the method against changes in imaging parameters.

The ability to distinguish between cortical and trabecular bone is advantageous, since it allows to independently model the mechanical behavior of the two components during FE simulations. It has been shown that both cortical and trabecular bone significantly contribute to the failure load of the proximal femur [13,15,41]. In addition, deficits in trabecular and cortical bone density and structure independently contribute to fracture risk [16]. A number of studies have already been presented where the cortical and trabecular bone are modelled as separate compartments [42–44]. By providing an automatic identification of these compartments, the proposed method allows to perform this kind of analysis in an automatic fashion.

The automatically generated FE models predicted femoral strains with high correlation ($R^2 = 0.89$, NRMSE = 6%), although slightly lower than that obtained by manually generated models ($R^2 = 0.94$, NRMSE = 9% [36]) (Table 4). This result may appear counterintuitive, since the automatic segmentation replicated the real femoral anatomy more closely than the manual segmentation. However, the relationship used in the FE analyses to convert bone density to elastic modulus was nonlinear [36] and obtained for trabecular bone only at the femoral neck region [38]. Such relationship has proven to be successful in traditional subject-specific FE models from clinical CT images [36,45], where partial volume effects were not addressed during model creation. It is likely that these new, more accurate geometrical models with cortical-trabecular separation would benefit from having dedicated density-elasticity relationships where CPM correction in FE models is accounted for. However, we chose to retain the same density-elasticity relationship for the models as in the previous study, as this study focuses on presenting an automated segmentation algorithm and automated FE modelling pipeline. The effects of choosing different density-elasticity relationships for FE models for different models and boundary conditions have been previously treated in literature [46,47], and finding the optimal one for these FE models remains as a scope for future studies.

The proposed method extracted the periosteal surface of the femurs with an accuracy comparable to previously published studies [7–10]. Thus, the single template morphing followed by surface extraction with the CPM algorithm was able to successfully capture the anatomical variability within our subjects of elderly females. Both in our current approach and in previous studies using statistical approaches [7–10], the femoral joint region has the lowest accuracy and the remaining parts of the femur show an accuracy of about half a voxel [48].

The proposed methodology is the first that allows to take a clinical CT scan and automatically perform all the steps required to obtain a FE model with the separation of cortical and trabecular bone, mapping of the material properties and isotopological meshes. The isotopology also means that it is suitable to create statistical models of variations, which further enables to identify boundary conditions at similar anatomical locations, including muscle contact forces. To the authors' best knowledge, only one study proposed a similar automatic pipeline including all these features [22]; however, that study validated the method only using *ex-vivo* images from cadavers, and no validation of the strain accuracy of the generated FE models was provided. The presented method was extensively validated in all its steps. The accuracy of the obtained femoral shape, both periosteal and endocortical, has

been validated first against semi-automatic segmentation of the same *in-vivo* clinical CT images. Second, the ability of reaching sub-pixel resolution was verified using the *ex-vivo* CT images, against μCT images. Finally, the ability of the FE models to effectively and accurately predict the mechanical behavior of bone was assessed by comparing the model predictions to an extensive set of experimental strain measurement collected using DIC [24].

The main limitation with this study is that a relatively low number ($N = 13$) of *in-vivo* clinical CT images was available. While the authors are open to anyone who wants to try the presented method on their set of clinical images, it has to be said that the CT images we collected for this study represented a quite challenging benchmark for the algorithm. This is due to our subjects being elder female with shown osteoporosis (several of them had a fractured contralateral femur), decreasing bone appearance, relatively low radiation dose (i.e., reduced signal-to-noise ratio) and relatively thick slices (2 mm), which results in blurring (especially at joint space). As a second limitation, the strain prediction accuracy achieved with the proposed automated segmentation and FE generation pipeline was slightly lower than the accuracy obtained by manually generated FE models on the same validation dataset (Table 4). We believe, however, that this aspect is greatly compensated by the benefit of the complete automation of the process.

In conclusion, the presented fully automatic segmentation method was able to segment proximal femurs from clinical low-dose CT images and generate finite element models for mechanical analysis. These FE models agreed with experimental measurements, showing similar accuracy as the current state-of-the-art FE models generated from a labor-intensive manual procedure. In a real clinical scenario, the method has potential for use in fracture risk prediction from hip CT images.

## Conflict of interest

None.

## Funding

The authors acknowledge the following funding sources: Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (project 5203101), Kuopio, Finland, Finnish Cultural Foundation, Swedish Research Council (2015-4795), Swedish Foundation for Strategic Research (IB2013-0021) and 3D Imaging Center at the Technical University of Denmark (DTU).

## Ethical approval

The study was approved by Kuopio University Hospital Ethical Committee, permission 80/2008). Permission for collection and analysis of cadaver proximal femur samples was granted by the National Authority for Medico-legal Affairs (TEO, 5783/04/044/07).

## References

[1] World Health Organization. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report, 843. Springer; 1994. doi:10.1007/BF01622200.
[2] Kanis JA, Melton LJ, Christiansen C, Johnston CC, Khaltaev N. The diagnosis of osteoporosis. J Bone Miner Res 2009;9:1137–41. doi:10.1002/jbmr.5650090802.
[3] Link TM. Osteoporosis imaging: state of the art and advanced imaging. Radiology 2012;263:3–17. doi:10.1148/radiol.12110462.
[4] Pasco JA, Seeman E, Henry MJ, Merriman EN, Nicholson GC, Kotowicz MA. The population burden of fractures originates in women with osteopenia, not osteoporosis. Osteoporos Int 2006;17:1404–9. doi:10.1007/s00198-006-0135-9.
[5] Cody DD, Gross GJ, Hou FJ, Spencer HJ, Goldstein SA, Fyhrie DP. Femoral strength is better predicted by finite element models than QCT and DXA. J Biomech 1999;32:1013–20. doi:10.1016/S0021-9290(99)00099-8.
[6] Zadpoor AA, Weinans H. Patient-specific bone modeling and analysis: the role of integration and automation in clinical adoption. J Biomech 2015;48:750–60. doi:10.1016/j.jbiomech.2014.12.018.
[7] Chu C, Chen C, Liu L, Zheng G. FACTS: fully automatic CT segmentation of a hip joint. Ann Biomed Eng 2015;43:1247–59. doi:10.1007/s10439-014-1176-4.
[8] Chu C, Bai J, Wu X, Zheng G. MASCG: multi-Atlas segmentation constrained graph method for accurate segmentation of hip CT images. Med Image Anal 2015;26:173–84. doi:10.1016/j.media.2015.08.011.
[9] Kim JJ, Nam J, Jang IG. Fully automated segmentation of a hip joint using the patient-specific optimal thresholding and watershed algorithm. Comput Methods Progr Biomed 2018;154:161–71. doi:10.1016/j.cmpb.2017.11.007.
[10] Nicolella DP, Bredbenner TL. Development of a parametric finite element model of the proximal femur using statistical shape and density modelling. Comput Methods Biomech Biomed Engin 2011;15:101–10. doi:10.1080/10255842.2010.515984.
[11] Travison TG, Beck TJ, Esche GR, Araujo AB, McKinlay JB. Age trends in proximal femur geometry in men: variation by race and ethnicity. Osteoporos Int 2008;19:277–87. doi:10.1007/s00198-007-0497-7.
[12] Marshall LM, Zmuda JM, Chan BKS, Barrett-Connor E, Cauley JA, Ensrud KE, et al. Race and ethnic variation in proximal femur structure and BMD among older men. J Bone Miner Res 2008;23:121–30. doi:10.1359/jbmr.070908.
[13] Koivumäki JEM, Thevenot J, Pulkkinen P, Kuhn V, Link TM, Eckstein F, et al. Cortical bone finite element models in the estimation of experimentally measured failure loads in the proximal femur. Bone 2012;51:737–40. doi:10.1016/j.bone.2012.06.026.
[14] Johannesdottir F, Turmezei T, Poole KES. Cortical bone assessed with clinical computed tomography at the proximal femur. J Bone Miner Res 2014;29:771–83. doi:10.1002/jbmr.2199.
[15] Manske SL, Cooper DML. Cortical and trabecular bone in the femoral neck both contribute to proximal femur failure load prediction. Osteoporos Int 2009;20:445–53. doi:10.1007/s00198-008-0675-2.
[16] Samelson EJ, Broe KE, Xu H, Yang L, Boyd S, Biver E, et al. Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the bone microarchitecture international consortium (BoMIC): a prospective study. Lancet Diabetes Endocrinol 2018;7:34–43. doi:10.1016/S2213-8587(18)30308-5.
[17] Falcinelli C, Schileo E, Pakdel A, Whyne C, Cristofolini L, Taddei F. Can CT image deblurring improve finite element predictions at the proximal femur? J Mech Behav Biomed Mater 2016;63:337–51. doi:10.1016/j.jmbbm.2016.07.004.
[18] Pakdel A, Hardisty M, Fialkov J, Whyne C. Restoration of Thickness, Density, and volume for highly blurred thin cortical bones in clinical CT images. Ann Biomed Eng 2016;44. doi:10.1007/s10439-016-1654-y.
[19] Museyko O, Gerner B, Engelke K. A new method to determine cortical bone thickness in CT images using a hybrid approach of parametric profile representation and local adaptive thresholds: accuracy results. PLoS ONE 2017;12:1–22. doi:10.1371/journal.pone.0187097.
[20] Treece GM, Gee aH, Mayhew PM, Poole KES. High resolution cortical bone thickness measurement from clinical CT data. Med Image Anal 2010;14:276–90. doi:10.1016/j.media.2010.01.003.
[21] Treece GM, Gee AH. Independent measurement of femoral cortical thickness and cortical bone density using clinical cT. Med Image Anal 2015;20:249–64. doi:10.1016/j.media.2014.11.012.
[22] Zhang J, Malcolm D, Hislop-Jambrich J, Thomas CDL, Nielsen P. Automatic meshing of femur cortical surfaces from clinical CT images. Lect. Notes Comput. Sci. 2012;7599:40–8 LNCS. doi:10.1007/978-3-642-33463-4_5.
[23] Zhang J, Hislop-Jambrich J, Besier TF. Predictive statistical models of baseline variations in 3-D femoral cortex morphology. Med Eng Phys 2016;38:450–7. doi:10.1016/j.medengphy.2016.02.003.
[24] Grassi L, Väänänen SP, Amin Yavari S, Jurvelin JS, Weinans H, Ristinmaa M, et al. Full-Field strain measurement during mechanical testing of the human femur at physiologically relevant strain rates. J Biomech Eng 2014;136:111010. doi:10.1115/1.4028415.
[25] Karjalainen JP, Riekkinen O, Töyräs J, Hakulinen M, Kröger H, Rikkonen T, et al. Multi-site bone ultrasound measurements in elderly women with and without previous hip fractures. Osteoporos Int 2012;23:1287–95. doi:10.1007/s00198-011-1682-2.
[26] Väänänen SP, Grassi L, Flivik G, Jurvelin JS, Isaksson H. Generation of 3D shape, density, cortical thickness and finite element mesh of proximal femur from a DXA image. Med Image Anal 2015;24:125–34. doi:10.1016/j.media.2015.06.001.
[27] Liao PS, Chen TS, Chung PC. A fast algorithm for multilevel thresholding. J Inf Sci Eng 2001;17:713–27.
[28] Liu B, Zhang H, Hua S, Jiang Q, Huang R, Liu W, et al. An automatic segmentation system of acetabulum in sequential CT images for the personalized artificial femoral head design. Comput Methods Programs Biomed 2015;127:318–35. doi:10.1016/j.cmpb.2015.12.012.
[29] Kroon D-J. Finite iterative closest point. Mathworks 2009. https://se.mathworks.com/matlabcentral/fileexchange/24301-finite-iterative-closest-point (accessed November 26, 2016).
[30] Kroon D-J. Segmentation of the mandibular canal in cone-beam CT data. University of Twente; 2011. doi:10.3990/1.9789036532808.
[31] Väänänen SP, Isaksson H, Waarsing JH, Zadpoor AA, Jurvelin JS, Weinans H. Estimation of 3D rotation of femur in 2D hip radiographs. J Biomech 2012;45:2279–83. doi:10.1016/j.jbiomech.2012.06.007.
[32] Manu (2016). nonrigidICP (https://www.mathworks.com/matlabcentral/fileexchange/41396-nonrigidicp), MATLAB Central File Exchange. (accessed November 26, 2016).

[33] Audenaert EA, Houcke JV, Almeida DF, Paelinck L, Peiffer M, Steenackers G, Vandermeulen D. Cascaded statistical shape model based segmentation of the full lower limb in CT. Comput Method Biomec 2019;22:644–57. doi:10.1080/10255842.2019.1577828.
[34] Lee TC, Kashyap RL, Chu CN. Building skeleton models via 3-D medial surface axis thinning algorithms. CVGIP Graph Model Image Process 1994;56:462–78. doi:10.1006/cgip.1994.1042.
[35] Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol 2004;11:178–89. doi:10.1016/S1076-6332(03)00671-8.
[36] Grassi L, Väänänen SP, Ristinmaa M, Jurvelin JS, Isaksson H. How accurately can subject-specific finite element models predict strains and strength of human femora? Investigation using full-field measurements. J Biomech 2016;49:802–6. doi:10.1016/j.jbiomech.2016.02.032.
[37] Schileo E, Dall'ara E, Taddei F, Malandrino A, Schotkamp T, Baleani M, et al. An accurate estimation of bone density improves the accuracy of subject-specific finite element models. J Biomech 2008;41:2483–91. doi:10.1016/j.jbiomech.2008.05.017.
[38] Morgan EF, Bayraktar HH, Keaveny TM. Trabecular bone modulus–density relationships depend on anatomic site. J Biomech 2003;36:897–904. doi:10.1016/S0021-9290(03)00071-X.
[39] Wolff J. Ueber die innere architectur der knochen und ihre bedeutung für die frage vom knochenwachsthum. Arch Für Pathol Anat Und Physiol Und Für Klin Med 1870;50:389–450.
[40] Wolff J. The classic: on the inner architecture of bones and its importance for bone growth. 1870. Clin Orthop Relat Res 2010;468:1056–65. doi:10.1007/s11999-010-1239-2.
[41] Holzer G, Von Skrbensky G, Holzer LA, Pichl W. Hip fractures and the contribution of cortical versus trabecular bone to femoral neck strength. J Bone Miner Res Off J Am Soc Bone Miner Res 2009;24:468–74. doi:10.1359/jbmr.081108.
[42] Nawathe S, Akhlaghpour H, Bouxsein ML, Keaveny TM. Microstructural failure mechanisms in the human proximal femur for sideways fall loading. J Bone Miner Res 2014;29:507–15. doi:10.1002/jbmr.2033.
[43] Luisier B, Dall'Ara E, Pahr DH. Orthotropic HR-pQCT-based FE models improve strength predictions for stance but not for side-way fall loading compared to isotropic QCT-based FE models of human femurs. J Mech Behav Biomed Mater 2014;32:287–99. doi:10.1016/j.jmbbm.2014.01.006.
[44] Enns-Bray WS, Bahaloo H, Fleps I, Ariza O, Gilchrist S, Widmer R, et al. Material mapping strategy to improve the predicted response of the proximal femur to a sideways fall impact. J Mech Behav Biomed Mater 2018;78:196–205. doi:10.1016/j.jmbbm.2017.10.033.
[45] Schileo E, Balistreri L, Grassi L, Cristofolini L, Taddei F. To what extent can linear finite element models of human femora predict failure under stance and fall loading configurations? J Biomech 2014;47:3531–8. doi:10.1016/j.jbiomech.2014.08.024.
[46] Helgason B, Perilli E, Schileo E, Taddei F, Brynjólfsson S, Viceconti M. Mathematical relationships between bone density and mechanical properties: a literature review. Clin Biomech (Bristol, Avon) 2008;23:135–46. doi:10.1016/j.clinbiomech.2007.08.024.
[47] Helgason B, Gilchrist S, Ariza O, Vogt P, Enns-Bray W, Widmer RP, et al. The influence of the modulus–density relationship and the material mapping method on the simulated mechanical response of the proximal femur in side-ways fall loading configuration. Med Eng Phys 2016:1–11. doi:10.1016/j.medengphy.2016.03.006.
[48] Whitmarsh T, Humbert L, De Craene M, Del Rio Barquero LM, Frangi AF. Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry. IEEE Trans Med Imaging 2011;30:2101–14. doi:10.1109/TMI.2011.2163074.