HHS Public Access

Author manuscript

Comput Methods Biomech Biomed Engin. Author manuscript; available in PMC 2018 March 27.

Published in final edited form as:

Comput Methods Biomech Biomed Engin. 2016 December ; 19(16): 1693–1703. doi: 10.1080/10255842.2016.1181173.

Interactive Graph-Cut Segmentation for Fast Creation of Finite Element Models from Clinical CT Data for Hip Fracture Prediction

Yves Pauchard[1,*], **Thomas Fitze**[1], **Diego Browarnik**[1], **Amiraslan Eskandari**[2,3,4], **Irene Pauchard**[5], **William Enns-Bray**[2], **Halldór Pálsson**[3,4], **Sigurdur Sigurdsson**[6], **Stephen J. Ferguson**[2], **Tamara B. Harris**[7], **Vilmundur Guðnason**[4,6], and **Benedikt Helgason**[2]

[1]Institute of Applied Information Technology, Zurich University of Applied Sciences, Winterthur, Switzerland [2]Institute for Biomechanics, ETH Zürich, Zürich, Switzerland [3]Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, School of Engineering and Natural Sciences [4]University of Iceland, Reykjavik, Iceland [5]McCaig Institute, Cumming School of Medicine, University of Calgary [6]The Icelandic Heart Association Research Institute, Kopavogur, Iceland [7]Laboratory of Epidemiology and Population Sciences, National Institute on Aging

Abstract

In this study, we propose interactive graph cut image segmentation for fast creation of femur finite element (FE) models from clinical computed tomography scans for hip fracture prediction. Using a sample of N=48 bone scans representing normal, osteopenic and osteoporotic subjects, the proximal femur was segmented using manual (gold standard) and graph cut segmentation. Segmentations were subsequently used to generate FE models to calculate overall stiffness and peak force in a sideways fall simulations. Results show that, comparable FE results can be obtained with the graph cut method, with a reduction from 20 minutes to 2–5 minutes interaction time. Average differences between segmentation methods of 0.22 mm were not significantly correlated with differences in FE derived stiffness (R^2 = 0.08, p = 0.05) and weakly correlated to differences in FE derived peak force (R^2 = 0.16, p = 0.01). We further found that changes in automatically assigned boundary conditions as a consequence of small segmentation differences were significantly correlated with FE derived results. The proposed interactive graph cut segmentation software MITK-GEM is freely available online at https://simtk.org/home/mitk-gem.

Keywords

hip fracture prediction; image segmentation; clinical computed tomography; image-based finite element modeling; femur segmentation

[*]Corresponding author, present address: McCaig Institute, Cumming School of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta, Canada, T2N 4N1, yves.pauchard@ucalgary.ca.

INTRODUCTION

Subject specific, image-based finite element (FE) analysis of bone has received considerable attention in the biomechanics literature in recent years. One clinical application of interest using FE models is patient-specific prediction of the risk of a hip fracture (Amin et al., 2011; Bessho et al., 2009; Falcinelli et al., 2014; Keyak et al., 2011; Keyak et al., 2013; Kopperdahl et al., 2014; Lang et al., 2012; Orwoll et al., 2009). These tools are aimed at mitigating the economic burden of such fractures for our healthcare systems and the devastating consequences of these events for the patients involved. Current clinical screening tools, such as DEXA or FRAX, are not sensitive enough to identify individuals at risk, even though they can capture population trends (Leslie and Morin, 2011). Despite the potential for improving hip fracture risk prediction, FE analysis has not yet been implemented into clinical practice mainly due to: the cost of carrying out the simulations by trained engineers which are not a part of the typical clinical staff; the uncertain accuracy of these methods for prospective fracture prediction; and the radiation from x-Ray CT scans needed to construct the models. Justifying the extra radiation dosage beyond the current standard of care will be difficult until these methods can be used reliably to identify at risk individuals.

The workflow for constructing the FE models of bones is conceptually the same from study to study, consisting of extracting the bone geometry of interest from the CT image data, also known as segmentation; generating a FE mesh; applying heterogeneous gray level based material properties to the finite elements; applying boundary conditions; solving the FE equations; and interpreting the results (Pahr and Zysset, 2009; Poelert et al., 2013; Viceconti et al., 2004). Programming this workflow into a robust, automatic or a semi-automatic FE simulation pipeline is necessary for practical clinical application, allowing researchers to process larger cohorts, improving the accuracy of prospective FE-based fracture risk prediction.

The segmentation step is a key element in the FE pipeline and is especially challenging at joints and in aging cohorts (Wang et al., 2006), due to low bone quality and narrow joint space. Bone segmentation is recognized as time-consuming and labor intensive (Poelert et al., 2013), due to extensive manual interaction needed to guarantee suitable geometry for FE model creation. Numerous approaches have been developed for segmenting bones, ranging from manual tracing, thresholding (with morphological operations), contour extraction, region growing, to active shape based (snake and level set) methods (Poelert et al., 2013; Wang et al., 2006). For best results, these methods are often combined into hybrid methods (Cheng et al., 2013; Pahr and Zysset, 2009; Wang et al., 2006) and paired with manual correction (Kang et al., 2003; Orwoll et al., 2009). In fact, adopting a ‘user in the loop’ paradigm and allowing for manual correction is currently the most straightforward approach to obtain suitable segmentation for FE modeling. Recently, statistical shape models (Fritscher et al., 2007; Kainmueller et al., 2009; Yokota et al., 2013), multi-atlas registration (Whitmarsh et al., 2014) and graph cut segmentation (Krcah et al., 2011) have been applied to segment the proximal femur from CT images with promising results. While statistical shape models and multi-atlas registration require a database of gold standard segmentations, in itself an active area of research, graph cut segmentation was shown to work without such prior knowledge (Krcah et al., 2011). Additionally, graph cut segmentation is well suited for

fast interactive segmentation (Boykov and Funka-Lea, 2006), but its merit in image-based FE model creation has not been assessed. More importantly, while intra- and inter-operator differences of manual segmentation on FE results were recently investigated (Peleg et al. 2014), the influence of different segmentation methods on FE results has not been previously reported. Therefore, the current study investigated the potential of interactive graph cut segmentation for fast creation of FE femur models from clinical CT data and compares results to manual segmentation.

This study had two aims: to develop and release for use a highly automated tool for segmenting bone from clinical CT images, and to demonstrate its use in retrospective proximal femur strength prediction using a sample cohort data from the AGES-Reykjavik study. The outcome was compared to manual segmentation, the current gold standard.

MATERIAL AND METHODS

Image Data

Quantitative computed tomography (QCT) scans were obtained from the Age Gene/Environment Susceptibility Reykjavik Study (AGES-Reykjavik Study) as previously described (Keyak et al., 2011). Briefly, subjects were scanned at the pelvis using a 4-detector CT system (Sensation, Siemens Medical Systems, Erlangen, Germany) with a bone mineral reference standard (3-sample calibration phantom, Image Analysis, Columbia, KY USA) containing calibration cells of 0, 75 and 150 mg/cm^3 equivalent concentration of calcium hydroxyapatite. Scan parameters were selected (120 kVp, 140 mAs, 1-mm slice thickness, pitch=1) to produce images with voxel dimensions of 1mm × 1mm × 1mm and included the proximal femur from a point 1 cm superior to the acetabulum to a point 3–5 mm inferior to the lesser trochanter. For the current study using this acquisition protocol, CT scans of the proximal femur from 24 subjects were selected from the AGES-Reykjavik study cohort. We randomly selected 4 men and 4 women in each group of normal, osteopenic or osteoporotic based on t-scores. The t-scores were based on CT derived simulated DXA BMD of the femoral neck area. The t-scores and groups of normal, osteopenic and osteoporotic were based on the NHANES-III BMD data (Kaptoge et al., 2008; Looker et al., 1997). A total of 48 bones (24 left and 24 right) were available for processing by one operator. Additionally, a subset consisting of 12 left femur bones (2 men, 2 women, from each of the 3 groups: normal, osteopenic and osteoporotic) was selected for a reproducibility study evaluating inter-operator variability with two additional operators.

Manual Image Segmentation

An interactive segmentation method, combining region growing and manual correction was used as the gold standard. These methods are readily available in the open source image analysis tool The Medical Imaging Interaction Toolkit (MITK, www.mitk.org). This software framework served as a basis for software development in the current project. Custom plugins were implemented and integrated in MITK as explained below. This new analysis software, referred to as MITK-GEM, was employed in the current study. The 3D femur image was processed from proximal to distal end in 2D axial slices. If possible, region growing was initialized with a seed in high-density cortical bone and enlarging the inclusion

criteria to obtain an initial segmentation. Missing and misidentified regions were manually added and removed. If region growing did not provide a satisfactory initial guess, the contour was drawn entirely manually. This process was repeated in every second slice and interpolation between these contours was employed in the remaining slices. The output of the segmentation was a voxel mask, defining the voxels belonging to the femur. Prior to constructing a FE mesh, the voxel mask was converted to a closed surface by filtering with a Gaussian filter (sigma = 2.0, radius = 1.5), followed by marching cubes to provide a smooth surface with custom plugins integrated into MITK-GEM. Overall, manual segmentation took on the order of 20–35 minutes per bone.

Graph Cut Image Segmentation

The proposed interactive segmentation method is based on graph cut segmentation (Boykov and Funka-Lea, 2006). This method classifies each voxel in an image to belong either to the object or the background by finding the global minimum of the following cost function:

$$E(A)=\lambda \cdot \sum_{p \varepsilon P} R_p(A_p) + \sum_{\{p,q\} \varepsilon N} B_{p,q} \cdot \delta_{A_p \neq A_q} \quad (1)$$

where R_p is the regional term incorporating knowledge of individual voxels and $B_{p,q}$ is the boundary term exploring the voxel neighborhood. The Dirac delta function $\delta_{Ap \neq Aq}$ ensures that edge weights only contribute to the sum of the boundary term if neighboring voxels are not assigned the same class. The parameter λ (>0) defines the relative weighting of the two terms. To find the minimum of the cost function, a graph containing one node for each voxel in the image plus two additional nodes, the source (object) and sink (background) terminal, was constructed. Edge weights between voxel nodes encode the boundary term (n-links), while edge weights between voxel nodes and the source and sink nodes encode the regional term (t-links). The graph cut algorithm assigns each voxel to either object or background by cutting appropriate n- and t-links using a max-flow/min-cut algorithm (Boykov and Kolmogorov, 2004).

Traditionally, the regional term is composed of statistical information (e.g. the histogram) providing some likelihood of a voxel belonging to object or background. In the present study, the region term only encodes so-called hard constraints, i.e. voxels selected by the user that are considered object and background. Corresponding t-links from voxel to source and background terminal are set to L and 0, respectively, if the voxel was identified as bone, and 0 and L if the voxel was identified as background (L = largest float value).

The boundary term, i.e. n-link weight, encodes the relationship between neighboring voxels. The term is defined using the following relationship (Boykov and Funka-Lea, 2006):

$$B_{p,q}=\exp\left(\frac{(I_p - I_q)^2}{2 \cdot \sigma^2}\right) \quad (2)$$

Where σ^2 is interpreted as variance in homogenous object regions. In essence, the difference in Hounsfield units defines the edge weight: if differences are large compared to σ^2 (e.g. at object boundaries) the cost of cutting the edge is low, while if differences are small compared to σ^2, then the cost is high. In the proposed method, the boundary term is computed for each voxel with respect to its 6 direct neighboring voxels.

Interactive graph cut segmentation was implemented in C++ using the Insight Toolkit (www.itk.org) and the max-flow/min-cut library (Boykov and Kolmogorov, 2004) and integrated as a plugin into MITK-GEM. Graph cut implementation was inspired by an open source contribution (Doria D., 2010).

With this software, hard constraints for object and background were obtained by manually drawing on a single 2D coronal slice in the center of the femur (see Figure 1). In addition to these hard constraints, a σ = 50 HU in the boundary term was utilized to set the n- and t-weights in each image and obtain a 3D graph cut segmentation of the femur. If necessary, the resulting 3D voxel mask was manually corrected using the tools available in MITK-GEM. Similar to manual segmentation, the voxel mask was converted to a closed surface by filtering with a Gaussian filter (sigma = 2.0, radius = 1.5), followed by marching cubes with custom plugins available in MITK-GEM. Overall, graph cut segmentation took on the order of 2–5 minutes per bone.

Validation of Segmentation Methods

New graph cut segmentation and manual segmentation were analyzed in two sub-studies. In a full analysis study, all 48 bones were segmented with both approaches by a single operator. In the inter-operator reproducibility study, a subset of 12 left femur bones were segmented with both approaches by two additional operators (total of three operators).

In order to compare interactive graph cut segmentation to the manual gold standard, voxel masks were compared, and the absolute maximum difference (mm) (Hausdorff distance, Insight Toolkit, www.itk.org) and mean overlap (Dice coefficient, Insight Toolkit, www.itk.org) were computed for each sample. Additionally, differences in segmentations were assessed by computing the average, minimum and maximum signed surface-to-surface differences [mm] (Visualization Toolkit, www.vtk.org) of smoothed segmentation surfaces. These surface-to-surface differences from the full analysis study were subsequently compared to differences in FE results. In addition to comparing graph cut to manual segmentations, in the inter-operator reproducibility study, these metrics were calculated between graph cut segmentations and manual segmentations from the three operators in pair-wise manner.

Finite Element Modelling

All segmentations of the full analysis study (N=48) were utilized for FE analysis. The smoothed segmentation surface (STL-file) was meshed using ICEM CFD, (v.14. ANSYS Inc., Canonsburg, PA, USA). The ash density (ρ_{ash}) in each pixel was linked with calibrated bone mineral content (mgHA) using ρ_{ash} = (mgHA/1000 + 0.09)/1.14 g/cm^3 and related to apparent density (ρ_{app}) with $\rho_{app} = \rho_{ash}$ /0.6 (Schileo et al., 2008). Modulus of elasticity (E) was determined with $E = 6850(\rho_{app})^{1.49}$ (MPa) (Morgan et al., 2003). The ultimate stress

was defined based on element apparent density using the equation proposed by (Carter and Hayes, 1977) $\sigma_u = 68 \cdot (\dot{\varepsilon})^{0.06} \cdot \rho_{\text{app}}^2$ assuming strain rate $\dot{\varepsilon} = 0.005/s$ which was the strain rate used in the experimental testing by (Morgan et al., 2003). The yield stress of trabecular bone was defined as a fraction of the ultimate stress, specifically $\sigma_y = \sigma_u/1.1$(Helgason et al., 2008). Tension compression asymmetry ratio of 0.7 (Bayraktar et al., 2004) was implemented with a Drucker-Prager yield criterion. Strain at ultimate stress was set to 2% and beyond this strain the response was assumed to be perfectly plastic. Yield strain was defined by a 0.2% offset rule and the proportional limit defined as 80% of the yield stress. A sample of the stress-strain behavior in compression for the material mapping strategy is illustrated in Figure 2. Poisson's ratio was assumed to be 0.3. Five hundred different material cards were defined for each model and the elements assigned the properties of the closest material card.

The boundary condition of the FE simulations were defined to replicate a sideways fall on the hip, which is believed to be the most relevant loading mode for low trauma hip fractures (Parkkari et al., 1999). The angles of impact for a sideways fall loading mode vary from study to study in the literature (Amin et al., 2011; Bessho et al., 2009; Falcinelli et al., 2014; Keyak et al., 2011; Keyak et al., 2013; Kopperdahl et al., 2014; Lang et al., 2012; Orwoll et al., 2009). In the present study the FE models were simply aligned in accordance with their scanning position with the medial-lateral x-axis of the image plane defined as the direction of impact. Various ways of applying boundary conditions (BCs) to the FE models in a sideways fall loading configuration have been used in the retrospective fracture prediction in the literature. (Falcinelli et al., 2014) applied concentrated BCs at single nodes at the femoral head and greater trochanter. In contrast, distributed support and loading BCs over an assumed area of contact was used in other studies (Amin et al., 2011; Bessho et al., 2009; Keyak et al., 2011; Keyak et al., 2013; Kopperdahl et al., 2014; Lang et al., 2012; Orwoll et al., 2009). In the present work we supported the greater trochanter region of the FE models in the drop direction and allowed free sliding in the yz-plane (Figure 2). Contact was assumed to be distributed over the surface nodes found within a radius of 15 mm from the most laterally located FE node, in accordance with previously published work (Keyak et al., 2011; Keyak et al., 2013; Lang et al., 2012). The distal end of the models was supported in the z-direction and allowed to slide freely in the xy-plane in agreement with the study of (Amin et al., 2011). The FE models were loaded to an overall deformation of 4% nominal strain by displacing the surface nodes of the femoral head found within a radius of 15 mm of the most medially located FE node. The peak force was taken as the maximum force from the overall force-deformation curve of the whole bone in agreement with the work of (Amin et al., 2011). After segmentation and meshing, the creation of FE simulation files, application of boundary conditions and mapping of material properties, and launching of the simulations were carried in a fully automated fashion using in-house Matlab scripts (Matlab, Mathworks). The FE equation system was solved using ANSYS (v.15. ANSYS Inc., Canonsburg, PA, USA).

The principal outcome measures for the comparison of the FE results between the two segmentation methods were whole bone stiffness (K), calculated as the slope of the greater trochanter force-displacement response up to a displacement of 0.4 mm, and peak force (F),

taken as the reaction force at the greater trochanter support nodes at 4% nominal strain. In order to identify potential contributors to the difference in FE results between the two segmentation methods, simple linear regression analysis was performed. In addition to comparing effects of surface-to-surface differences of segmentations on FE results, the effect of differences in automatically determined boundary conditions on FE results was investigated. More precisely, the y- and z-distances (ry and rz) between the center of the greater trochanter support area and femoral head load application area, as determined by the automatic applications of the BCs, were used as independent variables and differences in whole bone stiffness and peak force were dependent variables in linear regression analysis. General linear hypothesis testing of the slope being different from unity and intercept different from zero, with $p < 0.05$ was performed on each dependent variable individually (R-Statistics, vers. 3.1.0, © The R Foundation for Statistical Computing).

RESULTS

Validating Segmentation Methods

When producing interactive graph cut segmentations, manual corrections were performed in 47 out of 48 femur segmentations. Corrections were generally small in magnitude and predominantly at the femur head, in the proximity of the acetabulum. Median percentage of voxels adjusted, amounted to 0.74%, with interquartile range spanning 0.44% – 1.41%, with a maximum of 10.16%. Comparing all N=48 femur segmentations, absolute maximum difference (Hausdorff distance) between manual segmentation and interactive graph cut segmentation was 3.75 ± 1.26 mm (mean ± SD) with absolute minimum and maximum values of 2.18 mm and 9.0 mm, respectively. Mean overlap (Dice coefficient) was 0.973 ± 0.005, ranging from 0.963 to 0.981. Surface-to-surface differences between manual and interactive graph cut segmentations are summarized in Table 1. On average, graph cut segmentation produces slightly smaller geometries, as indicated by the negative average distance. Due to smoothing, surface-to-surface differences are slightly smaller than the differences of voxel masks (compare to above). Results of the inter-operator reproducibility segmentation study (N=12) are summarized in Table 2. Segmentation metrics for interactive graph cut to manual segmentation comparison were comparable (slightly superior) to results in the full study with small variability between operators. Pair-wise comparison of segmentation methods between operators indicated that Hausdorff distances were consistently smaller between manual segmentations. In contrast, mean surface-to-surface differences were consistently smaller and mean overlap (Dice) was higher between graph cut segmentations.

Surface-to-surface differences between segmentation methods (full study, N=48) were not significantly correlated with differences in FE derived results except for the average and the min surface-to-surface difference being weakly albeit significantly correlated with differences in FE derived peak force with $R^2 = 0.16$, $p = 0.01$ and $R^2 = 0.09$, $p = 0.04$ respectively (see Table 3). In Figure 3, the surface-to-surface differences of two samples are visualized as color rendering on the graph cut surface and the corresponding histogram of values. Note that overall differences are small and there exist a few locations with large differences.

Comparing FE Results

The regressions and correlations between FE results from manual and automatic segmentation methods are illustrated in Figure 4 along with corresponding Bland-Altman plots. The bones selected from the cohort, with the material mapping strategy and boundary conditions used in the present study, represent a wide range of bone quality with FE-derived stiffness ranging from 5041 N/mm to 17556 N/mm N and FE-derived peak force of 3796 N to 12636 N, which are results based on the manual segmentation FE models. The regression slope for stiffness (0.97) and peak force (1.00) respectively (Figures 4 (a) and 4 (b)) were found not to be significantly different from unity. The intercept of the regression lines for both stiffness and force comparison were found not to be significantly different from zero. The Bland-Altman plots (Figure 4 (c) and (d)) did not reveal any particular trends in the difference between the two segmentation methods. Average absolute difference in stiffness was −3.4% (max difference 26.6%) and in peak force −3.4% (max difference 18.5%).

Further analysis of the source of the difference between the FE result for the two segmentation methods showed that the difference in stiffness can partially be explained by difference in r_z ($R^2 = 0.27$), which is the distances measured in the z-directions respectively between the center of the femoral head load application area and the center of the greater trochanter support area (Figure 2 and Table 3). The difference in peak force, which on average is smaller than the difference in stiffness, was also significantly correlated to the difference in r_z ($R^2 = 0.12$). The maximum difference in stiffness and peak force was found for the same bone, which was also the bone with the highest difference in r_z between the segmentation methods. Further study of the FE models for this specimen showed that a subtle difference in the shape of the femoral head resulting from the two segmentation methods had a large influence of the location of the most prominent node on FE models in the loading direction which again produced large difference in application of the BCs (Figure 5).

DISCUSSION

The aim of this study was to develop a highly automated tool for segmenting bone from clinical CT images, to quantify its geometric accuracy and to demonstrate its use in a retrospective proximal femur strength prediction using a sample cohort data from the AGES-Reykjavik study. Since ground-truth values for segmentation and FE were not available for this *in vivo* data, comparing segmentation and FE results to manual segmentation assessed segmentation accuracy and viability. We found that the proposed interactive graph cut segmentation corresponds well to manual segmentation, based on surface-to-surface comparison of both methods (average surface-to-surface differences = 0.22 mm), while reducing the user interaction time from 20–35 min (manual) to 2–5 min (proposed). In addition, we found good agreement in predicted whole bone stiffness ($R^2 = 0.95$) and peak force ($R^2 = 0.98$) when comparing the two segmentation methods used to create FE models of the proximal human femurs based on clinical CT scans.

This study attempted to use meaningful FE results to compare the two different segmentation methods. The regression analysis of the FE results (Figure 4) showed no statistically significant differences and a high correlation was shown between the

segmentation methods on the population of bones used in the present study. However, the resulting internal strength and stiffness ranking of the bones was not exactly the same for both segmentation methods. We identified that the difference in BCs between the FE models was the largest contributor to the differences in FE results, but the BCs in the present study, which were defined based on previous work in the literature, are highly sensitive to even small differences in the segmentation outcome. This is because the center of the support and load application areas are selected automatically based on the most laterally and medially placed nodes in the FE models and we found that the location of these nodes could vary considerably for the same bone segmented with the two different methods, even though other metrics for comparing the methods indicated insignificant differences. One way of decreasing the sensitivity of the BCs to the segmentation would be to define the center of the support area with respect to the center of the femoral head but it is less obvious what the approach should be at the greater trochanter. We concluded that the outcome of the FE is highly sensitive to small changes in BC application, and this is important consideration for future studies carrying out FE fracture prediction of human femoral specimens based on patient data. Since small variations in BCs can influence both the magnitudes of whole bone stiffness and strength as well as the internal ranking of a bone population based on these metrics, the BCs may potentially influence fracture predictions based on strength ((Bessho et al., 2009; Falcinelli et al., 2014; Keyak et al., 2011; Keyak et al., 2013; Lang et al., 2012)) as well as fracture predictions based on load-to-strength ratios (Amin et al., 2011; Orwoll et al., 2009).

Similar to the present study, other femur segmentation studies use manual segmentation as a gold standard (Ben Younes et al., 2014; Fritscher et al., 2007; Kainmueller et al., 2009; Krcah et al., 2011; Whitmarsh et al., 2014). These studies were based on different image resolutions and varying number of images analyzed. However, differences between segmentations here, i.e. average maximum (Hausdorff) distance 3.75 mm, mean overlap (Dice) 0.973 and average surface-to-surface distance of −0.22 mm, compare well to previously published results. Average maximum (Hausdorff) distances were reported to be 2.98 mm (Whitmarsh et al., 2014), 5.04 mm (Krcah et al., 2011) and 10.53 mm (Ben Younes et al., 2014). Mean overlap ranged from 0.915 (Cheng et al., 2013), 0.988 (Whitmarsh et al., 2014), 0.992 (Fritscher et al., 2007). Definition of average distances vary in previous studies and values reported range from 0.2 mm (Fritscher et al., 2007), 0.23 mm (Whitmarsh et al., 2014), 0.97 mm (Yokota et al., 2013), 1.1 mm (Kainmueller et al., 2009), 1.22 mm (Cheng et al., 2013) to 1.48 mm (Ben Younes et al., 2014).

Inter-operator reproducibility performed on a sub-set of femurs revealed that maximum errors (Hausdorff distance) between manual segmentations were slightly smaller for all three operators. However, segmentation bias was removed (mean surface-to-surface difference smaller than 0.006 mm) and a higher mean overlap (Dice) was measured when using interactive graph cut segmentation. Since Hausdorff distance measures worst-case error and in general only a small area is affected and average measures are in favor of the new segmentation method, interactive graph cut segmentation can be considered the method with higher reproducibility. These results further indicate that the exact seeding of the foreground and background voxels in graph cut segmentation does not affect reproducibility. However, it

was found that the seeding pattern and anatomical plane are important and the pattern displayed in Figure 1 produces good results.

Furthermore, we have not studied effects of partial volume artifacts known to blur edges, affecting geometric and intensity measurements (Hangartner and Short, 2007; Treece et al., 2010), and FE modeling results (Peleg et al., 2014). Methods to reduce partial volume artifacts have previously been reported, e.g. (Hangartner and Short, 2007; Peleg et al., 2014; Treece et al., 2010), and these methods can be applied as a post-processing step to the proposed interactive graph cut segmentation method. Applying partial volume artifact correction to segmentations renders bone surfaces and bone mineral density estimates closer to physical values (Hangartner and Short, 2007; Treece et al., 2010), as well as, reducing intra- and inter-operator differences (Peleg et al., 2014).

In summary, by using interactive graph cut image segmentation we found that it is possible to produce bounding surfaces and contained volumes of proximal femurs for hip fracture FE robustly in around 2 minutes. FE results are comparable to manual segmentation and enable analysis of large cohorts. The software is open-source and available online at https://simtk.org/home/mitk-gem.

Acknowledgments

This research was supported in part by the Intramural Research Program of the National Institute on Aging and National Eye Institute at the National Institutes of Health (NIH contract number N01-AG-1-2100), the Icelandic Heart Association, the Icelandic Parliament, the Icelandic Gerontological Institutes St. Joseph´ s PhD grant and the School of Engineering Zurich University of Applied Sciences Winterthur.

References

Amin S, Kopperdhal DL, Melton LJ, Achenbach SJ 3rd, Therneau TM, Riggs BL, Keaveny TM, Khosla S. Association of hip strength estimates by finite-element analysis with fractures in women and men. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2011; 26:1593–1600.

Bayraktar HH, Morgan EF, Niebur GL, Morris GE, Wong EK, Keaveny TM. Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue. Journal of biomechanics. 2004; 37:27–35. [PubMed: 14672565]

Ben Younes L, Nakajima Y, Saito T. Fully automatic segmentation of the femur from 3D-CT images using primitive shape recognition and statistical shape models. Int J Comput Assist Radiol Surg. 2014; 9:189–196. [PubMed: 24101434]

Bessho M, Ohnishi I, Matsumoto T, Ohashi S, Matsuyama J, Tobita K, Kaneko M, Nakamura K. Prediction of proximal femur strength using a CT-based nonlinear finite element method: differences in predicted fracture load and site with changing load and boundary conditions. Bone. 2009; 45:226–231. [PubMed: 19398043]

Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. Int J Comput Vision. 2006; 70:109–131.

Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell. 2004; 26:1124–1137. [PubMed: 15742889]

Carter DR, Hayes WC. Compressive Behavior of Bone as a 2-Phase Porous Structure. Journal of Bone and Joint Surgery-American. 1977; 59:954–962.

Cheng YZ, Zhou SJ, Wang YD, Guo CY, Bai J, Tamura S. Automatic segmentation technique for acetabulum and femoral head in CT images. Pattern Recogn. 2013; 46:2969–2984.

Doria D, C S. Interactive Image Graph Cut Segmentation. The Insight Journal July–December. 2010

Falcinelli C, Schileo E, Balistreri L, Baruffaldi F, Bordini B, Viceconti M, Albisinni U, Ceccarelli F, Milandri L, Toni A, Taddei F. Multiple loading conditions analysis can improve the association between finite element bone strength estimates and proximal femur fractures: A preliminary study in elderly women. Bone. 2014; 67:71–80. [PubMed: 25014885]

Fritscher KD, Grunerbl A, Schubert R. 3D image segmentation using combined shape-intensity prior models. Int J Comput Ass Rad. 2007; 1:341–350.

Hangartner TN, Short DF. Accurate quantification of width and density of bone structures by computed tomography. Medical physics. 2007; 34:3777–3784. [PubMed: 17985623]

Helgason B, Viceconti M, Runarsson TP, Brynjolfsson S. On the mechanical stability of porous coated press fit titanium implants: a finite element study of a pushout test. Journal of biomechanics. 2008; 41:1675–1681. [PubMed: 18471819]

Kainmueller D, Lamecker H, Zachow S, Hege HC. An Articulated Statistical Shape Model for Accurate Hip Joint Segmentation. Ieee Eng Med Bio. 2009:6345–6351.

Kang Y, Engelke K, Kalender WA. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. IEEE transactions on medical imaging. 2003; 22:586–598. [PubMed: 12846428]

Kaptoge S, da Silva JA, Brixen K, Reid DM, Kroger H, Nielsen TL, Andersen M, Hagen C, Lorenc R, Boonen S, de Vernejoul MC, Stepan JJ, Adams J, Kaufman JM, Reeve J. Geographical variation in DXA bone mineral density in young European men and women. Results from the Network in Europe on Male Osteoporosis (NEMO) study. Bone. 2008; 43:332–339. [PubMed: 18519175]

Keyak JH, Sigurdsson S, Karlsdottir G, Oskarsdottir D, Sigmarsdottir A, Zhao S, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. Male-female differences in the association between incident hip fracture and proximal femoral strength: a finite element analysis study. Bone. 2011; 48:1239–1245. [PubMed: 21419886]

Keyak JH, Sigurdsson S, Karlsdottir GS, Oskarsdottir D, Sigmarsdottir A, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Lang TF. Effect of finite element model loading condition on fracture risk assessment in men and women: the AGES-Reykjavik study. Bone. 2013; 57:18–29. [PubMed: 23907032]

Kopperdahl DL, Aspelund T, Hoffmann PF, Sigurdsson S, Siggeirsdottir K, Harris TB, Gudnason V, Keaveny TM. Assessment of incident spine and hip fractures in women and men using finite element analysis of CT scans. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2014; 29:570–580.

Krcah M, Szekely G, Blanc R. Fully Automatic and Fast Segmentation of the Femur Bone from 3D-CT Images with No Shape Prior. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2011:2087–2090.

Lang TF, Sigurdsson S, Karlsdottir G, Oskarsdottir D, Sigmarsdottir A, Chengshi J, Kornak J, Harris TB, Sigurdsson G, Jonsson BY, Siggeirsdottir K, Eiriksdottir G, Gudnason V, Keyak JH. Age-related loss of proximal femoral strength in elderly men and women: The Age Gene/Environment Susceptibility Study - Reykjavik. Bone. 2012; 50:743–748. [PubMed: 22178403]

Leslie WD, Morin S. Fracture Burden in Relation to Low Bone Mineral Density and FRAX (R) Probability. Journal of Clinical Densitometry. 2011; 14:279–285. [PubMed: 21723761]

Looker AC, Orwoll ES, Johnston CC Jr, Lindsay RL, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP. Prevalence of low femoral bone density in older U.S. adults from NHANES III. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 1997; 12:1761–1768.

Morgan EF, Bayraktar HH, Keaveny TM. Trabecular bone modulus–density relationships depend on anatomic site. Journal of biomechanics. 2003; 36:897–904. [PubMed: 12757797]

Orwoll ES, Marshall LM, Nielson CM, Cummings SR, Lapidus J, Cauley JA, Ensrud K, Lane N, Hoffmann PR, Kopperdahl DL, Keaveny TM. Osteoporotic Fractures in Men Study, G. Finite element analysis of the proximal femur and hip fracture risk in older men. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2009; 24:475–483.

Pahr DH, Zysset PK. From high-resolution CT data to finite element models: development of an integrated modular framework. Comput Methods Biomech Biomed Engin. 2009; 12:45–57. [PubMed: 18839383]

Parkkari J, Kannus P, Palvanen M, Natri A, Vainio J, Aho H, Vuori I, Jarvinen M. Majority of hip fractures occur as a result of a fall and impact on the greater trochanter of the femur: A prospective controlled hip fracture study with 206 consecutive patients. Calcified Tissue International. 1999; 65:183–187. [PubMed: 10441647]

Peleg E, Herblum R, Beek M, Joskowicz L, Liebergall M, Mosheiff R, Whyne C. Can a partial volume edge effect reduction algorithm improve the repeatability of subject-specific finite element models of femurs obtained from CT data? Comput Methods Biomech Biomed Engin. 2014; 17:204–209. [PubMed: 22452517]

Poelert S, Valstar E, Weinans H, Zadpoor AA. Patient-specific finite element modeling of bones. Proceedings of the Institution of Mechanical Engineers. Part H, Journal of engineering in medicine. 2013; 227:464–478.

Schileo E, Dall'ara E, Taddei F, Malandrino A, Schotkamp T, Baleani M, Viceconti M. An accurate estimation of bone density improves the accuracy of subject-specific finite element models. Journal of biomechanics. 2008; 41:2483–2491. [PubMed: 18606417]

Treece GM, Gee AH, Mayhew PM, Poole KE. High resolution cortical bone thickness measurement from clinical CT data. Medical image analysis. 2010; 14:276–290. [PubMed: 20163980]

Viceconti M, Davinelli M, Taddei F, Cappello A. Automatic generation of accurate subject-specific bone finite element models to be used in clinical studies. Journal of biomechanics. 2004; 37:1597–1605. [PubMed: 15336935]

Wang LI, Greenspan M, Ellis R. Validation of bone segmentation and improved 3-D registration using contour coherency in CT data. IEEE transactions on medical imaging. 2006; 25:324–334. [PubMed: 16524088]

Whitmarsh T, Treece GM, Poole KES. Automatic Segmentation and Discrimination of Connected Joint Bones from CT by Multi-atlas Registration. Lecture Notes in Computational Vision and Biomechanics. 2014:199–207.

Yokota, F., Okada, T., Takao, M., Sugano, N., Tada, Y., Tomiyama, N., Sato, Y. Lecture Notes in Computer Science. Springer-Verlag; Berlin-Heidelberg: 2013. Automated CT Segmentation of Diseased Hip Using Hierarchical and Conditional Statistical Shape Models; p. 190-197.

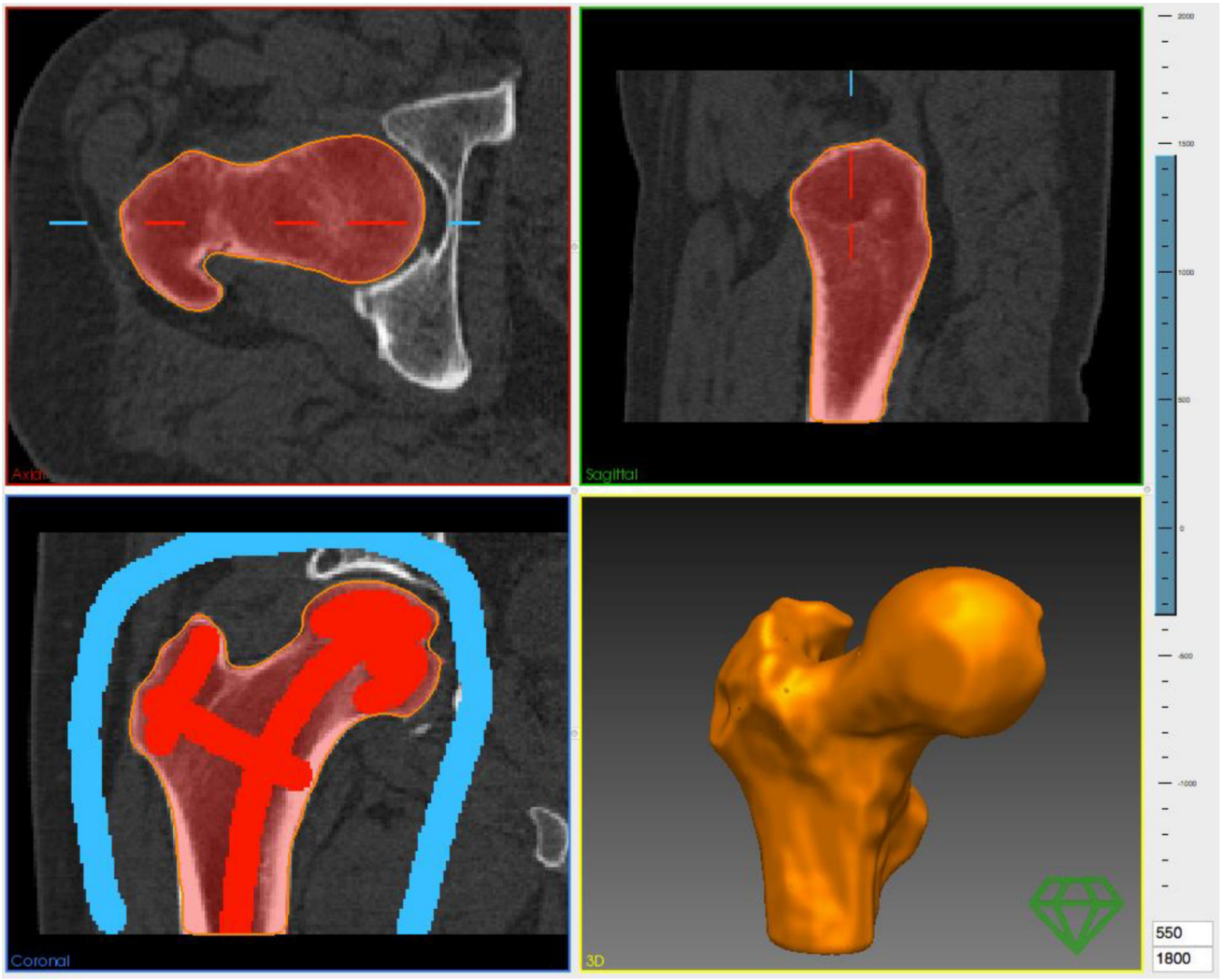

Figure 1.

Screenshot from graph cut segmentation tool MITK-GEM. Voxels belonging to the femur (red) and background (blue) are defined by the user in a single coronal slice (bottom, left). The resulting graph cut segmentation is visible in axial (top, left), sagittal (top, right) and coronal slices as red overlay. The resulting smoothed surface is visible as a 3D rendering (bottom, right).

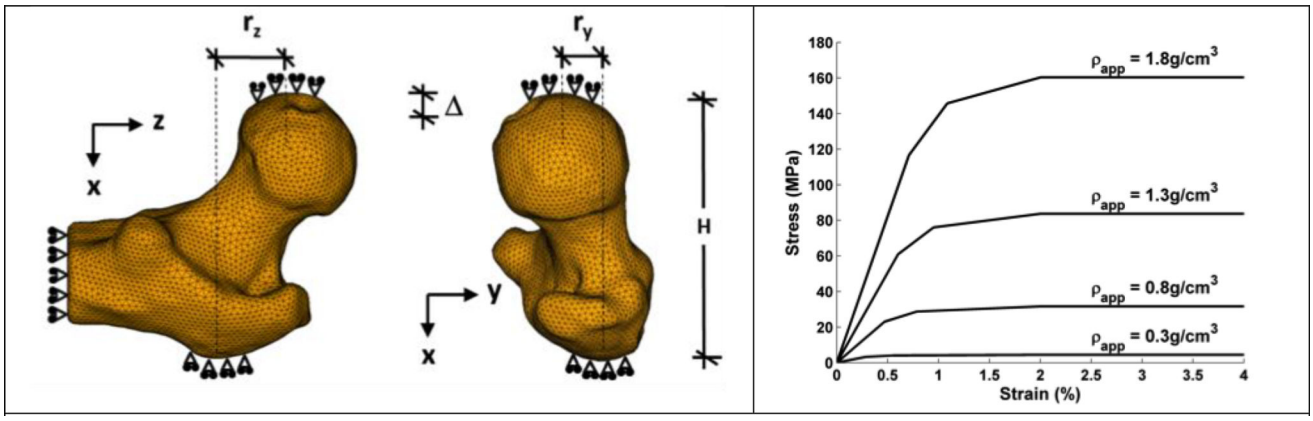

Figure 2.

Definition of boundary conditions and coordinate system used in the FE demonstrated on a sample FE model from the cohort data (left). Load was applied to the femoral head by displacing the surface nodes found with-in a radius of 15 mm of the most medially located node of the femoral head, by a distance Δ which was set as 4% of H, which was the distance in the x-direction from the most medially located node on the femoral head and the most laterally located node on the greater trochanter. The dimensions r_y and r_z are the distance between these nodes in the y- and z-directions respectively. Sample of the material properties assigned to the FE models (right). The curves show stress-strain relationship for compression. A tension-compression asymmetry ratio of 0.7 was assumed.

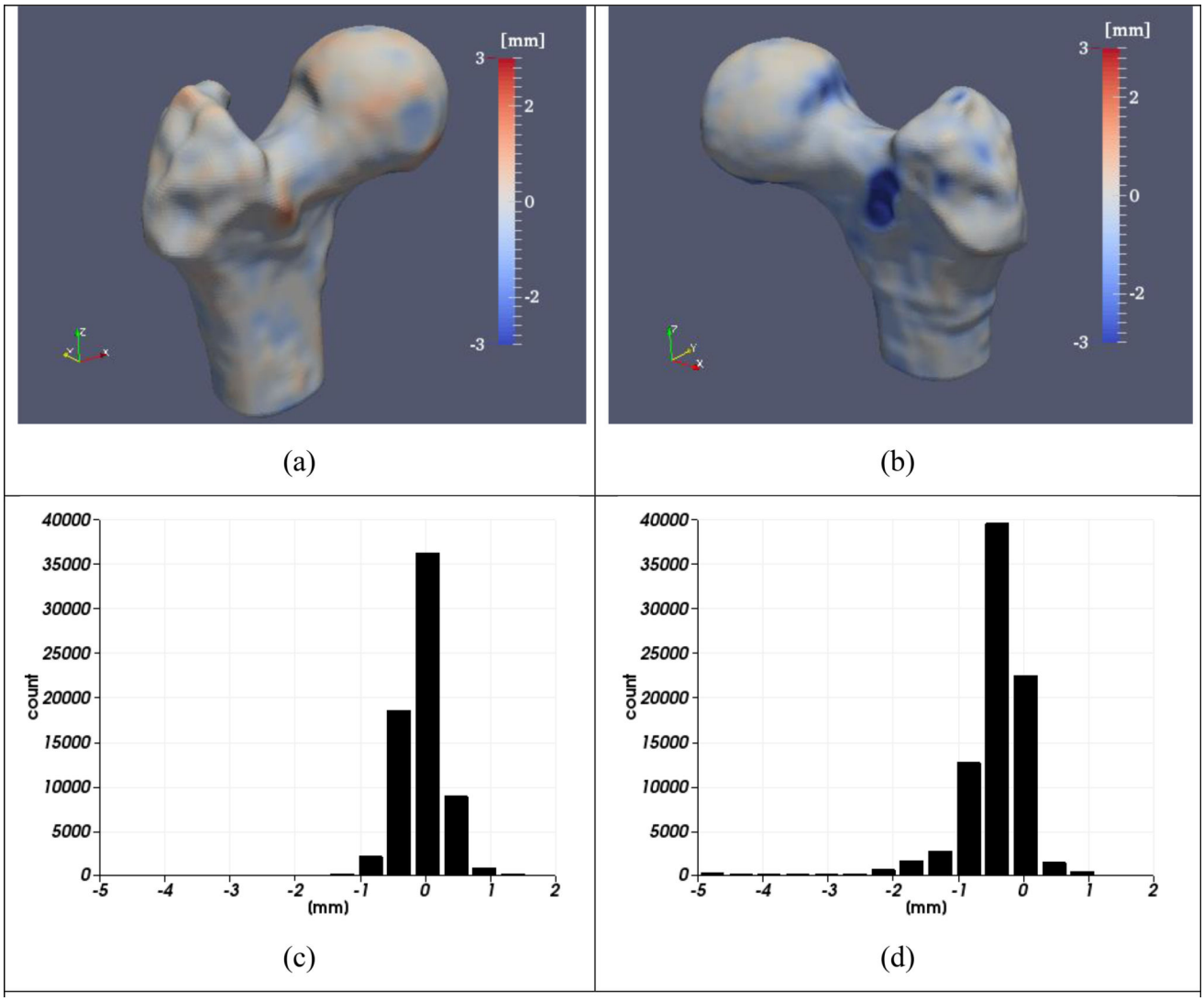

Figure 3.

Surface to surface differences of a bone segmentation with small average differences (mean = −0.05 mm, min = −1.23 mm, max = 1.52 mm) (a) and corresponding histogram (c); large average differences (mean = −0.43 mm, min = −7.23 mm, max = 1.10 mm) (b) and corresponding histogram (d).

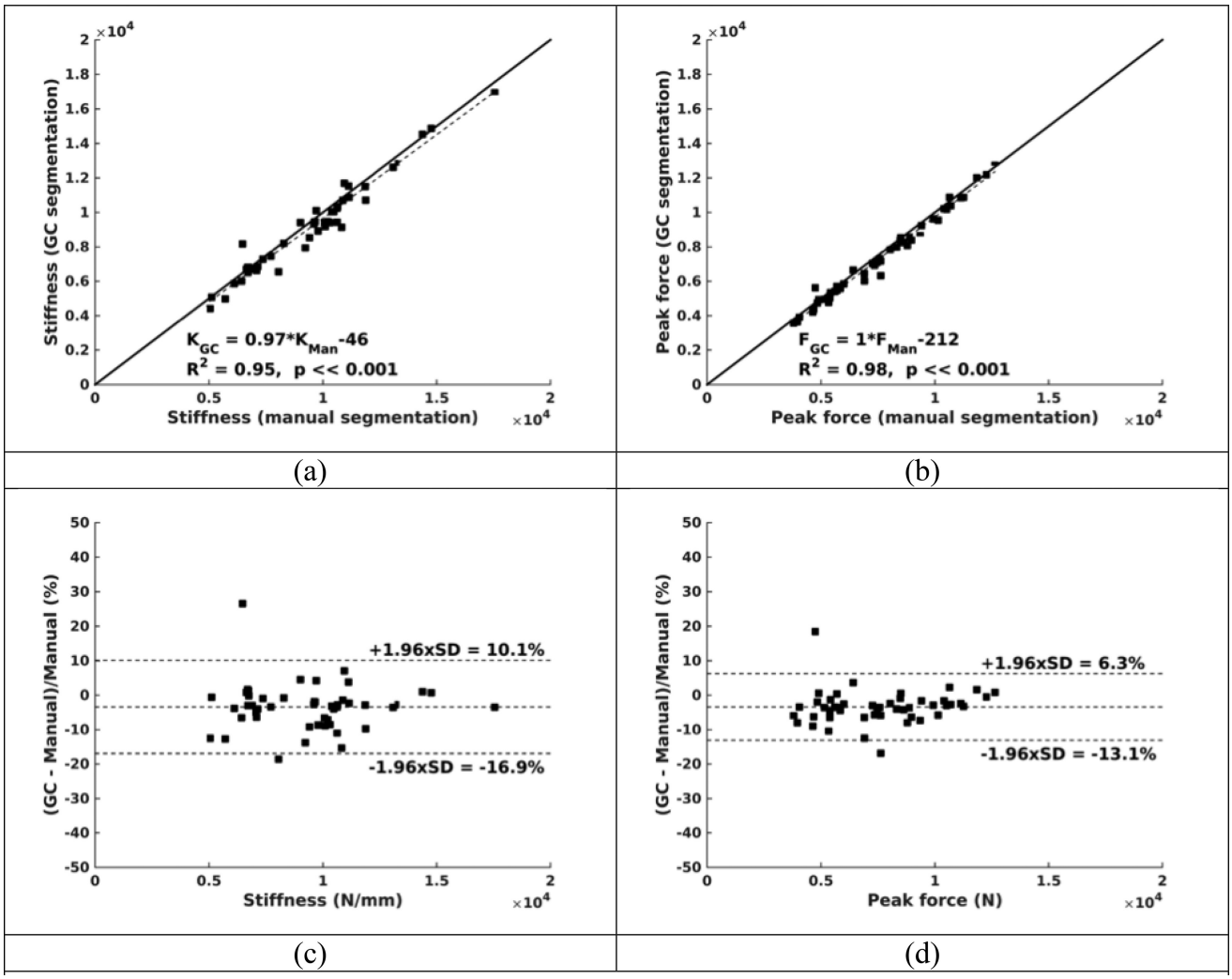

Figure 4.

Comparison between whole bone stiffness (K) in N/mm (a) and peak force (F) in N (b) with corresponding Bland-Altman plots in (c) and (d).

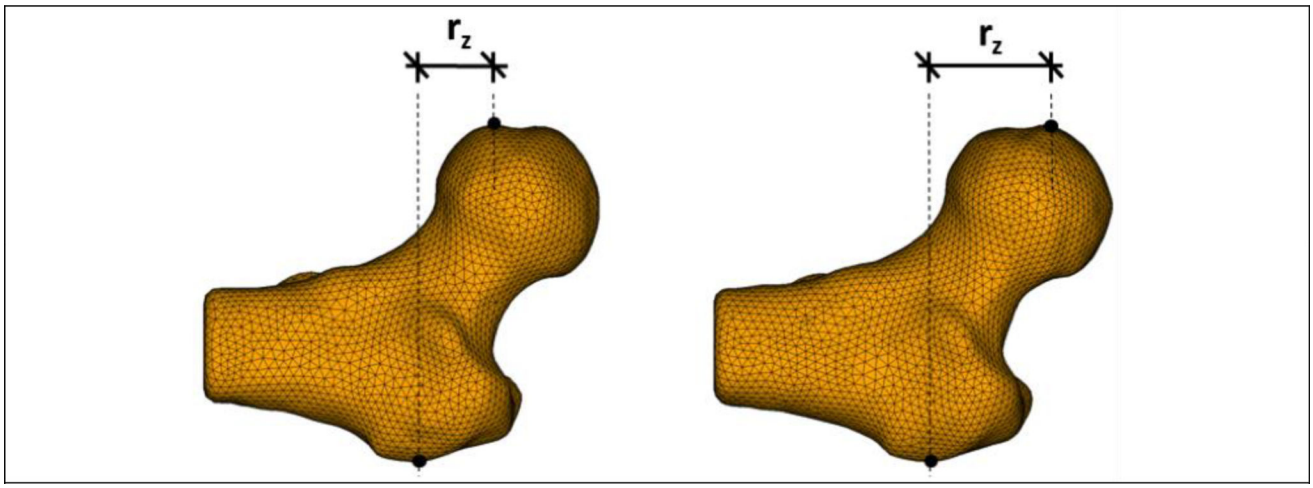

Figure 5.

FEMs (graph cut segmentation on the left and manual segmentation on the right) of the specimen that produced the largest error in stiffness and peak force comparison between the two segmentation methods. The black, solid circles indicate the most prominent FE nodes in the loading direction on the greater trochanter and the femoral head. The difference in r_z between the models was 11.52 mm.

Table 1

Surface-to-surface difference (N=48) between the two segmentation methods and its correlation to the difference in stiffness and peak force derived from the FE results. NS: not statistically significant.

	Surface to surface distance mean ± SD (range) (mm)	Correlation to difference in stiffness	Correlation to difference in peak force
Average difference	−0.22 ± 0.12 (−0.45 to 0.00)	$R^2 = 0.08^{NS}$ p = 0.05	$R^2 = 0.16$ p = 0.01
Min difference	−2.80 ± 1.30 (−7.23 to −1.12)	$R^2 = 0.04^{NS}$ p = 0.19	$R^2 = 0.09$ p = 0.04
Max difference	1.54 ± 0.54 (0.74 to 3.33)	$R^2 = 0.08^{NS}$ p = 0.05	$R^2 = 0.06^{NS}$ p = 0.10

Table 2

Comparing manual and graph cut segmentations created by three individual operators (N=12).

		Average mean surface-to-surface distance [mm]	Average mean overlap (Dice)	Average Haussdorff distance [mm]
Graph cut to manual	Op 1	−0.043	0.982	2.77
	Op 2	0.179	0.974	3.64
	Op 3	−0.199	0.974	3.30
	Op 3 (full study)	−0.219	0.973	3.75
Graph cut to graph cut	Op 1 – Op 3	−0.003	0.995	2.98
	Op 1 – Op 2	0.006	0.995	3.18
	Op 2 – Op 3	−0.002	0.993	3.49
Manual to manual	Op 1 – Op 3	−0.159	0.980	2.67
	Op 1 – Op 2	0.233	0.975	3.03
	Op 2 – Op 3	−0.378	0.969	3.09

Table 3

Difference in center-to-center distances between the greater trochanter support area and the femoral head load application area and its correlation to the difference in stiffness and peak force derived from the FE results. NS: not statistically significant.

	Difference in center-to-center distance between manual segmentation and GC segmentation (mm)	Correlation to difference in stiffness	Correlation to difference in peak force
Δr_y (mm)	0.06 ± 3.39 (−9.72 to 7.20)	$R^2 = 0.04^{NS}$ p = 0.15	$R^2 = 0.04^{NS}$ p = 0.17
Δr_z (mm)	−0.94 ± 4.02 (−11.52 to 9.36)	$R^2 = 0.27$ p < 0.001	$R^2 = 0.12$ p = 0.01