Available online at www.sciencedirect.com

SciVerse ScienceDirect

International Journal of Spine Surgery 6 (2012) 167–173

www.sasjournal.com

Variability of manual lumbar spine segmentation

Daniel J. Cook, MS [a], David A. Gladowski, BS [a], Heather N. Acuff [a], Matthew S. Yeager, BS [a], Boyle C. Cheng, PhD [a,b,*]

[a] *Department of Neurosurgery, Allegheny General Hospital, Pittsburgh, PA*
[b] *Department of Neurosurgery, Drexel University College of Medicine, Pittsburgh, PA*

Abstract

Background: The application of kinematic data acquired during biomechanical testing to specimen-specific, three-dimensional models of the spine has emerged as a useful tool in spine biomechanics research. However, the development of these models is subject to segmentation error because of complex morphology and pathologic changes of the spine. This error has not been previously characterized.
Methods: Eight cadaveric lumbar spines were prepared and underwent computed tomography (CT) scanning. After disarticulation and soft-tissue removal, 5 individual vertebrae from these specimens were scanned a second time. The CT images of the full lumbar specimens were segmented twice each by 2 operators, and the images of the individual vertebrae with soft tissue removed were segmented as well. The solid models derived from these differing segmentation sessions were registered, and the distribution of distances between nearest neighboring points was calculated to evaluate the accuracy and precision of the segmentation technique.
Results: Manual segmentation yielded root-mean-square errors below 0.39 mm for accuracy, 0.33 mm for intrauser precision, and 0.35 mm for interuser precision. Furthermore, the 95th percentile of all distances was below 0.75 mm for all analyses of accuracy and precision.
Conclusions: These findings indicate that such models are highly accurate and that a high level of intrauser and interuser precision can be achieved. The magnitude of the error presented here should inform the design and interpretation of future studies using manual segmentation techniques to derive models of the lumbar spine.
© 2012 ISASS - International Society for the Advancement of Spine Surgery. Published by Elsevier Inc. All rights reserved.

Keywords: Lumbar spine; Manual segmentation; Precision; Accuracy; Solid models

Modern medical imaging technology, such as computed tomography (CT) and magnetic resonance imaging, has made it possible to explore anatomic features in three dimensions (3D). Furthermore, advances in image processing have led to the development of specimen-specific models of certain anatomic features. These models are obtained by defining the portion of the image corresponding to the feature of interest, such as the brain, a tumor, or a single vertebra. This process is known as segmentation.

These models are becoming more prevalent in biomechanics research. They have been used to develop specimen-specific finite element models, and they have been used in conjunction with kinematic data acquired during biomechanical testing to investigate joint behavior. This technique provides a means of obtaining additional information with regard to the behavior of specific joint features under various loading conditions. The application of kinematic data to rigid body models has been used to investigate the region of contact at the facet joints, carpal bone interaction, and cartilage contact kinematics in the knee.[1–3] The fidelity of the analysis obtained from these techniques is dependent on the accuracy of the kinematic data acquired, the accuracy of registration between the reference frames of the motion capture system and the image data, and the accuracy of the solid models developed from the image data. Segmentation of the spine has proven to be a useful tool for several applications in medicine and medical research and requires varying levels of accuracy depending on which application is used. However, given the natural variation in morphology of spinal anatomy and limitations in imaging technology, segmentation of the human spine presents several challenges. Variation in the density of cortical bone across the surface of the vertebra can result in ambiguity in the boundary between bone and soft tissue in some regions, particularly in the spinous and transverse processes. The thickness of the articular cartilage of the facet is known to vary across

* Corresponding author: Boyle C. Cheng, PhD, 420 E North Ave, Ste 302, Pittsburgh, PA 15212; Tel: 412-359-4020; Fax: 412-359-8464.
E-mail address: boylecheng@yahoo.com

2211-4599 © 2012 ISASS - International Society for the Advancement of Spine Surgery. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijsp.2012.04.002

its surface and between different levels of the spine. The mean thickness of this layer has been shown to vary between 0.49 and 0.61 mm across the cervical spine.[4] The proximity of adjacent facet surfaces presents the greatest challenge, given the fact that the maximum resolution of medical CT scanners ranges between 0.6 and 1.0 mm in the axial direction depending on the device. Anatomic variation is exacerbated by the presence of pathology. Narrowed or hypertrophic facet joints, bony growths, poor bone density, partial or complete fusion between levels, and degenerated discs contribute to the difficulty of accurately segmenting the lumbar spine. These challenges eliminate the possibility of defining complete models of the vertebral surface for a vast majority of spines based on a simple intensity threshold. Therefore segmentation of the human spine has required either the intelligence of a knowledgeable operator or that of a sophisticated algorithm.

Manual segmentation requires the persistent input of an operator and often uses image filters, intensity thresholding, morphologic filters, and manual "painting" or outline definition. Automated segmentation routines are distinguished from manual ones in that they rely, at least in part, on some image or pattern recognition algorithm.

Several methods have been developed for automated spine segmentation relying on a wide variety, and often a combination, of distinct segmentation frameworks, including thresholding, edge detection, and various manifestations of deformable models coupled with optimization routines. The normalized cuts method of Carballido-Gamio et al[5] segments vertebrae from 2-dimensional magnetic resonance images. The average reported error of this method ranged from 14.44% to 19.34% in vertebral body area from a manual segmentation baseline, depending on input values. de Bruijne et al[6] used a shape particle filtering method on spine radiographs that yielded an average segmentation error of 1.4 mm from manual segmentation by a medical expert and under 2.0 mm in 88 of 91 cases. Kim and Kim[7] developed a fully automatic vertebral segmentation method using deformable 3D fences for CT images, but the method was only evaluated qualitatively. Furthermore, only 80% of the specimens were segmented successfully with this automated routine. A promising class of algorithms that has emerged, model-based algorithms, relies on the inclusion of prior shape information to the segmentation process.[8,9] These prior shape models are usually referred to as active or adaptive in that the location and shape of the model can be modified to achieve optimum correspondence of the model with shape information contained within the image. A physical metaphor of energy is generally used to explain these algorithms with external energy used to describe the attraction of the model to image features and internal energy used to describe the restriction of the adaptation to a known shape. A minimization of the total energy is used to optimize the segmentation process. Klinder et al[8] developed such a technique and applied it to CT images of the thoracic spine. The group reported an average segmentation accuracy of 1.0 mm when compared with segmentation achieved through a similar algorithm using more operator interaction. Using an active shape model to segment the lumbar spine from planar X-rays, Zamora et al[10] reported an average error below 6.4 mm in 50% of cases.

These methods have been developed for fast identification or segmentation of vertebrae in applications such as surgical planning, deformity assessment, and image fusion. However, most model-based investigations of joint articulation have required manual segmentation for some or all of the specimens used in their respective studies.[3,11–15] To our knowledge, no currently available automatic segmentation algorithm has been shown to be successful at segmenting lumbar vertebrae with submillimeter accuracy. These algorithms are subject to variability in the definition of initial conditions and may converge to local minima in some cases.[7] Given that articular cartilage on the facet may be approximately 1.0 mm at its thickest point, these techniques do not appear to be sufficiently accurate for such purposes.[4] Furthermore, as described earlier, many presentations of these automatic algorithms rely on manual segmentation as a standard of accuracy. Because the rate of segmentation is of no relevance compared with accuracy in biomechanical studies of this type, a manual segmentation process was developed for CT images with the aim to produce models with submillimeter accuracy and precision.

The focus of this study is to examine the accuracy, as well as the intrauser and interuser precision, of this technique on a series of human cadaveric lumbar spines.

Methods

Specimen preparation

Eight cadaveric lumbar spine segments from T12 through the sacrum (4 female and 4 male cadavers; mean age, 59.6 years; age range, 51–68 years) were cleaned of muscle, loose connective tissue, and the anterior longitudinal ligament with special care given to preserving the remaining intervertebral ligamentous structures. Each specimen underwent CT scanning with a slice thickness of 0.6 mm in a 64-slice CT scanner (Somatom; Siemens, Munich, Germany).

Vertebral segmentation

Commercially available medical image analysis software (ScanIP; Simpleware, Exeter, England) was used to generate a 3D model of each vertebra. Each specimen was segmented twice by each of 2 operators (operator A and operator B). For ease of comprehension, the following scheme will be used throughout the remainder of this article: segmentation sessions will be abbreviated with operator name (A or B) followed by session number (1 or 2); for example, the second segmentation session by operator B will be abbreviated B2. To ensure that each segmentation session was independent of bias related to memory, the first and

second segmentations of each specimen by each operator were separated temporally by at least 1 week. The strategy of this segmentation process was to isolate the cortical bone of each vertebra and use this outline to develop a closed model of the vertebral body surface.

The 2 operators had different levels of experience using the image analysis software. Operator A had completed more than 20 segmentations of the lumbar spine before the study. Operator B had been trained in using the segmentation software before the study but had not otherwise segmented any spines.

During importation of the CT data into the segmentation software, a custom window width and level were interactively applied to maximize contrast between bone and soft tissue. Before segmentation of the CT volumetric data, a curvature anisotropic diffusion noise filter was applied to the image background.

Masks are defined as a delineation of voxels that define the shape of an anatomic component, such as a lumbar vertebra. A mask was created to contain an operator-specified threshold, defined as voxels contained between an upper and lower boundary of grayscale units. This threshold was interactively applied by the operator with the intent to maximize bone included in the mask, minimize the soft tissue, and create well-defined borders at the facet joints.

Definition of individual vertebrae from the initial thresholding mask was achieved based on mask connectivity. However, the thresholding mask often contains areas of connectivity, or "bridges," between adjacent vertebrae in instances where no bony connectivity exists. These bridges are most commonly found at the facet joints where bone from adjacent vertebrae is in proximity and can be attributed to image blur.[14] Bridges were removed when the masked area did not correspond to bone per the judgment of the operator. None of the spines used in this study exhibited fusion at the facet joints.

Because models of the vertebral surface were desired, the interior of each mask defined by thresholding was filled by use of a combination of morphologic closing filters and filling operations based on mask boundaries. Manual definition and/or deletion of portions of masks was needed in instances where bone near the periphery of the vertebrae was not appropriately included in the initial threshold process. Finally, all masks were visually evaluated to verify that the definition of the masks corresponded with the boundary of the vertebrae in the CT image space.

A smoothing recursive Gaussian filter was applied to each mask. During exportation, each model was decimated to a maximum of 70,000 triangular surface patches per part to reduce file size and computation time in subsequent processing while maintaining sufficient resolution of the surface model. Because the models were limited to a maximum number of surface patches, the resolution was a function of the size of each vertebra. The average distance per side of each triangular surface patch across all vertebral levels was 0.70 mm. The minimum and maximum average distances for all vertebral levels were 0.66 mm (L1 level) and 0.72 mm (L5 level), respectively. For an L3 vertebra representative of the population, the average distance per side of each triangular surface patch in a non-decimated model was 0.38 mm. Figure 1 shows the models produced from the segmentation of a full lumbar spine.

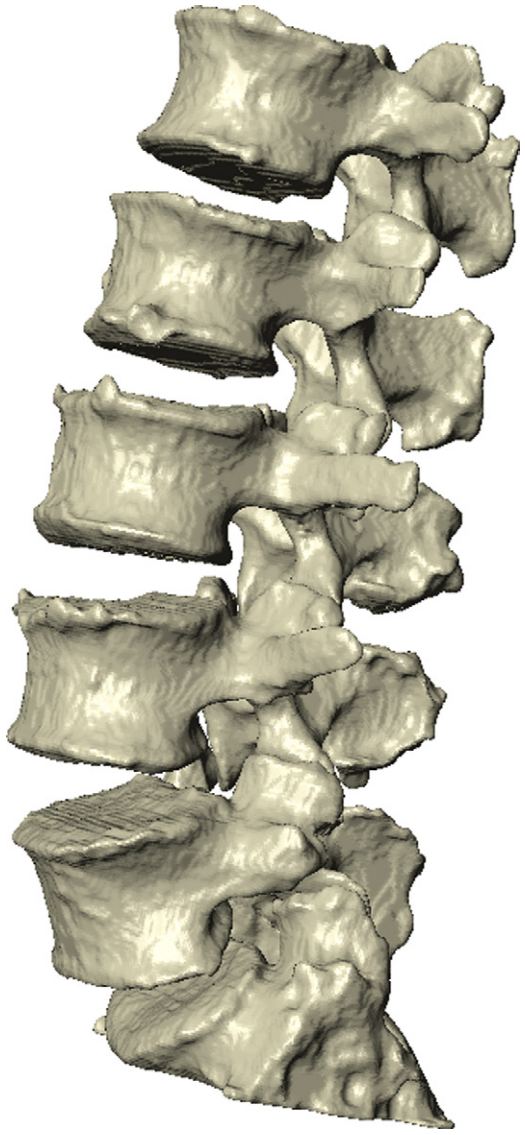

Fig. 1. Three-dimensional lumbar vertebrae models.

Precision of vertebral models

The precision of the segmentation process was determined by registering vertebral models from 1 segmentation session with the corresponding vertebral models from another segmentation session and then determining the distances from each point on 1 model to the nearest point on the other model. Although all segmentation sessions per specimen were completed using the same initial CT image, differences in cropping the image necessitated a translational registration process that was produced by the iterative closest point algorithm. Distance calculations were also

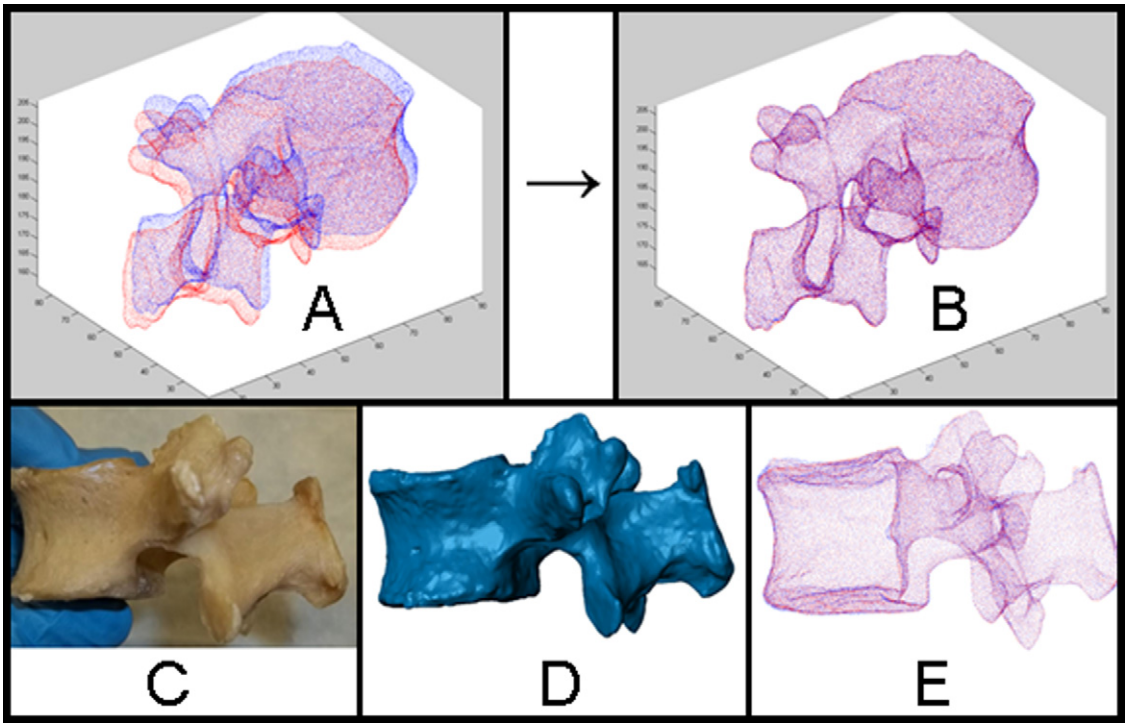

Fig. 2. (A) Unregistered segmentation point cloud models in red and blue. (B) Two models after registration. (C) Vertebra disarticulated from adjacent vertebrae and devoid of all soft tissue. (D) Three-dimensional surface-rendered model produced from the CT scan of the vertebra depicted in C. (E) Point cloud model of D registered with the corresponding operator A1 model.

determined with custom-written MATLAB software (MathWorks, Natick, Massachusetts). Figures showing the models before and after registration, as well as color distance maps, were produced to visually verify that the process was successful. Figure 2A and B show coupled segmentation sessions before and after registration. Intra-user precision was determined by comparing segmentation session A1 with session A2 and session B1 with session B2. Interuser precision was determined by comparing segmentation sessions A1 and B1.

Accuracy of vertebral models

To determine the accuracy of the segmentation technique, 5 vertebrae (L1 in 1, L2 in 2, and L3 in 2) from 2 lumbar segments that were used in the precision portion of this study were selected at random. The specimens were disarticulated from adjacent vertebrae, were completely removed of soft tissue, underwent CT scanning, and were then segmented an additional time. The soft tissue was removed as follows: After disarticulation, a portion of the soft tissue was manually and conservatively removed (to avoid damage to the bone) with common surgical tools. A maceration technique was then used. To loosen ligamentous tissue, the vertebrae were macerated in water that was kept at room temperature between 2 and 14 days. The vertebrae were then individually "simmered" in a solution of water with 2 tablespoons of sodium borate (20 Mule Team Borax; Dial, Scottsdale, Arizona) per 1.5 L followed by manual tissue removal. The total time each vertebra spent at simmering water temperature was approximately 2 hours. Both the endplates and articular cartilage were removed during this process. Finally, the vertebrae were subjected to another round of maceration at room temperature for between 2 and 7 days before CT scanning.

These steps were taken to produce 3D models that are less susceptible to segmentation error, particularly with regard to the facets. The lack of soft tissue and instrumentation allowed for the outer borders of the vertebrae to be easily identifiable without the use of judgment on the part of the operator. Facet surfaces are susceptible to error stemming from operator subjectivity. The methods used to determine segmentation accuracy alleviate this problem through disarticulation of the vertebrae. Thus the goal of this procedure was to determine the error associated with the segmentation process itself by removing those elements associated with user judgment.

This fifth "accuracy" segmentation was produced by operator A and compared with segmentation sessions A1 and B1 to determine segmentation accuracy. Because of the difference in orientation between the first and second CT scans of these vertebrae, it was necessary to apply a rigid body registration transformation to superimpose associated models. Models were registered by use of a freely available iterative closest point algorithm, written by Per Bergström, as part of custom-written MATLAB software. Figure 2C shows a vertebra that has undergone the tissue removal, Fig. 2D shows a 3D model produced by segmentation of the CT image, and Fig. 2E shows the

Table 1
Accuracy analysis

	Accuracy	
	Operator A	Operator B
RMS (mm)	0.39	0.36
Mean (mm)	0.37	0.35
SD (mm)	0.22	0.20
95th percentile (mm)	0.75	0.64
99th percentile (mm)	1.22	1.10

Table 2
Precision analysis

	Intrauser precision		Interuser precision: operators A1 and B1
	Operator A	Operator B	
RMS (mm)	0.32	0.33	0.35
Mean (mm)	0.31	0.32	0.34
SD (mm)	0.16	0.18	0.21
95th percentile (mm)	0.56	0.58	0.65
99th percentile (mm)	0.78	0.88	0.96

accuracy model registered with the corresponding A1 segmentation model.

Results

For each coupled session compared (ie, A1 and A2 for intrauser precision), 2 sets of distance calculations were made. For example, distances were calculated from the A1 model to the A2 model and from the A2 model to the A1 model. This was performed to ensure that a distance for each point in both models was included in the analysis. The results for each subsection presented later are shown with the distance calculation that yielded the maximum root-mean-square (RMS) between corresponding sets of points.

All accounts of vertebral accuracy coupled comparisons had submillimeter RMS values (range, 0.36–0.39 mm). The 95th percentile values were submillimeter values in all instances (range, 0.64–0.75 mm), whereas the 99th percentile values were near 1 mm (range, 1.10–1.22 mm). Comprehensive results are presented in Table 1 and compared with precision in Fig. 3.

All accounts of vertebral precision coupled comparisons had submillimeter RMS values (range, 0.32–0.35 mm) with the interuser precision RMS value being the largest. The 99th percentile values were also submillimeter values in all instances (range, 0.78–0.96 mm). Comprehensive results are presented in Table 2.

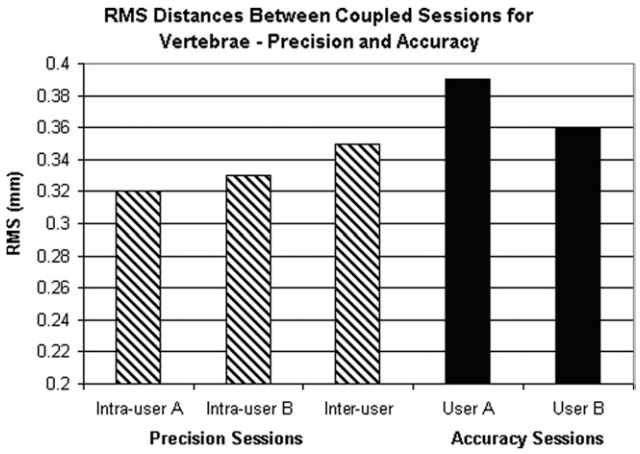

Fig. 3. The diagonal-lined bars show the maximum RMS values for the precision comparisons of the vertebral models, and the solid bars show the maximum RMS values for the accuracy comparisons of the vertebral models.

Discussion

3D solid modeling, using segmentation techniques similar to those described in this article, has become an important tool for biomechanical research as shown in the literature.[1-3,15-20] Applying kinematic data to solid models allows for the investigation of biomechanics and joint interaction in a noninvasive manner. As Cripton et al[21] outlined in an examination of the cervical spine, this technique allows researchers to visualize and interpret skeletal motion without obstruction from nonessential anatomic components. Furthermore, Cook and Cheng[1] used distance mapping between interacting facet surfaces that relied on the use of 3D solid models of the spine. Although the utility of 3D models is apparent, the validity of the results garnered from model-based kinematic studies is dependent in part on the quality of the models produced, which is the focus of this study.

The precision and accuracy of manual segmentation to create models of the lumbar spine have not been previously characterized. This information, in conjunction with error associated with kinematic tracking, is necessary for the analysis of data collected from model-based kinematic studies that use manual segmentation techniques. The accuracy of commercially available optoelectric tracking systems has been well characterized, and in many applications, it is on the submillimeter scale.[22,23] It is desirable then to achieve segmentation accuracy of approximately the same magnitude. Moreover, a priori knowledge of errors that are inherent to certain techniques equips researchers with necessary information for study design.

In terms of the precision of the segmented models, the intrauser comparisons yielded smaller RMS values than the interuser precision. This result was expected because manual corrections were made at the discretion of the operator and thus the resulting 3D models are dependent on the differences in judgment and manual segmentation techniques between operators. The low RMS values (range, 0.32–0.35 mm) for all precision sessions indicate that the use of manual segmentation is robust enough to allow for highly precise segmentation sessions between operators, regardless of experience, given that each operator has been properly trained using image analysis software. The accu-

racy results for the full vertebrae are similarly promising because the RMS values ranged from 0.36 to 0.39 mm. The accuracy segmentation sessions were a quality standard of accuracy because they incorporated the same imaging procedures as the precision models but did not contain soft tissue, instrumentation, or adjacent vertebrae and thus eliminated those components as potential sources of error. Most importantly, the lack of adjacent vertebrae eliminated operator judgment at the facet joints.

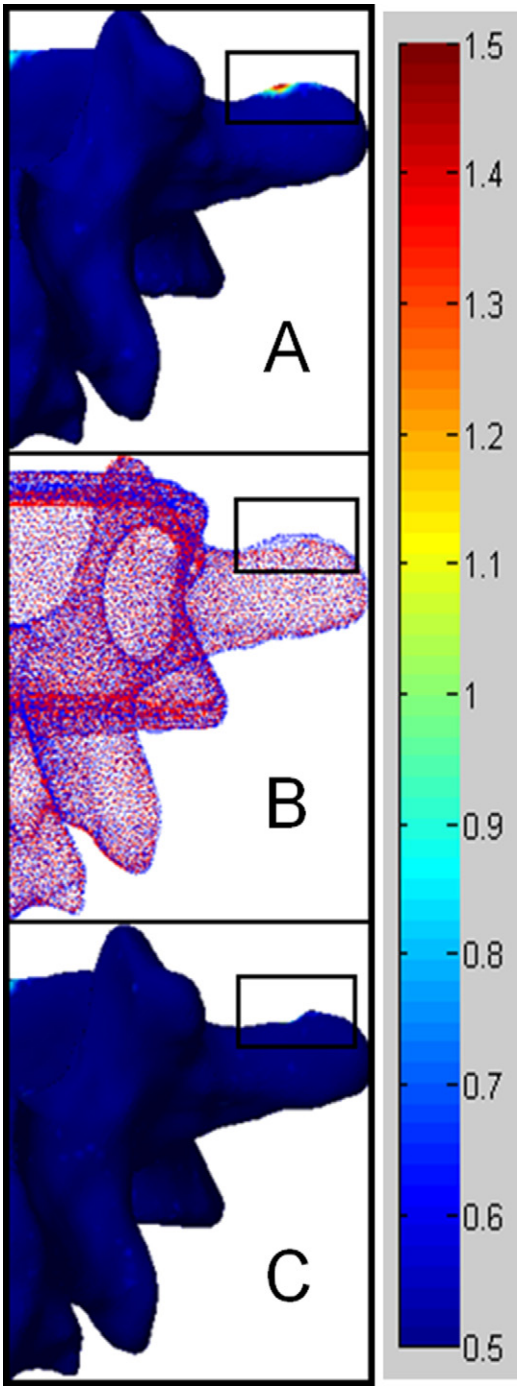

Fig. 4. (A) Distance map from blue model to red model. (B) Registered point cloud models of 2 coupled sessions. (C) Distance map from red model to blue model. For the distance maps, the bar on the right shows how the colors correspond to distances.

The technique presented for determining precision between 2 corresponding solid models encompasses variability across the entire model surface by investigating the distribution of distances between neighboring points on each surface after rigid body registration. To our knowledge, this is the first presentation of such a thorough technique in evaluating spinal segmentation. As a result, it should form the basis of future investigations in this area.

The importance of calculating distances for each set of points defining their respective models is illustrated in Fig. 4. Figure 4B shows registered point sets, with one depicted in red and the other in blue. It should be noted that part of the blue model exceeds the limits of the red model. Figure 4A shows a color map of the distances calculated from the blue model to the red model, and Fig. 4C shows a color map of the distances calculated from the red model to the blue model. The visible difference in the color maps indicates that results biased toward greater precision may be garnered if distances between both point sets are not included in analysis.

A limitation of this study is the use of maceration to remove soft tissue from the vertebrae used for the accuracy analysis. Liquid-based tissue removal techniques have been shown to alter skull shape and dimensions in rodents.[24] The amount of time that each vertebra used in the accuracy portion of the study spent in water for the purpose of maceration varied considerably. However, the use of the maceration techniques outlined in this article has not been characterized for human lumbar vertebrae. A further limitation is that each solid model was decimated before analysis. This process reduced surface resolution but was necessary to constrain file size and computation requirements. With an increase in computational power, analysis using non-decimated models should be considered. Finally, this study assumes that the CT scanner used had been properly calibrated and produced images that were true to the actual dimensions of the vertebrae for both the scans used for precision and those used for accuracy.

In conclusion, this study has shown that the reproducibility of lumbar vertebral models generated using manual segmentation is high and that these models accurately reflect the shape of the vertebra as indicated by submillimeter RMS accuracy values. The magnitude of the error presented here should inform the design and interpretation of future studies using manual segmentation techniques to derive models of the lumbar spine.

References

1. Cook DJ, Cheng BC. Development of a model based method for investigating facet articulation. *J Biomech Eng* 2010;132:064504.
2. Li G, DeFrate LE, Park SE, Gill TJ, Rubash HE. In vivo articular cartilage contact kinematics of the knee: an investigation using dual-

orthogonal fluoroscopy and magnetic resonance image-based computer models. *Am J Sports Med* 2005;33:102–7.
3. Green JK, Werner FW, Wang H, Weiner MM, Sacks JM, Short WH. Three-dimensional modeling and animation of two carpal bones: a technique. *J Biomech* 2004;37:757–62.
4. Womack W, Woldtvedt D, Puttlitz CM. Lower cervical spine facet cartilage thickness mapping. *Osteoarthritis Cartilage* 2008;16: 1018–23.
5. Carballido-Gamio J, Belongie SJ, Majumdar S. Normalized cuts in 3-D for spinal MRI segmentation. *IEEE Trans Med Imaging* 2004;23: 36–44.
6. de Bruijne M, Nielsen M. Image segmentation by shape particle filtering. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004). Volume 3. Los Alamitos, CA: IEEE Computer Society; 2004.
7. Kim Y, Kim D. A fully automatic vertebra segmentation method using 3D deformable fences. *Comput Med Imaging Graph* 2009;33:343–52.
8. Klinder T, Wolz R, Lorenz C, Franz A, Ostermann J. Spine segmentation using articulated shape models. *Med Image Comput Comput Assist Interv* 2008;11:227–34.
9. Boisvert J, Cheriet F, Pennec X, Labelle H, Ayache N. Articulated spine models for 3-D reconstruction from partial radiographic data. *IEEE Trans Biomed Eng* 2008;55:2565–74.
10. Zamora GA, Sari-sarraf HA, Long RB. Hierarchical segmentation of vertebrae from x-ray images. In: Proceedings of SPIE Medical Imaging 2003: Image Processing. Bellingham, WA, CA; 2003.
11. Van Sint Jan S, Giurintano DJ, Thompson DE, Rooze M. Joint kinematics simulation from medical imaging data. *IEEE Trans Biomed Eng* 1997;44:1175–84.
12. Van Sint Jan S, Salvia P, Hilal I, Sholukha V, Rooze M, Clapworthy G. Registration of 6-DOFs electrogoniometry and CT medical imaging for 3D joint modeling. *J Biomech* 2002;35:1475–84.
13. Anderst WJ, Tashman S. A method to estimate in vivo dynamic articular surface interaction. *J Biomech* 2003;36:1291–9.
14. Simon S, Davis M, Odhner D, Udupa J, Winkelstein B. CT imaging techniques for describing motions of the cervicothoracic junction and cervical spine during flexion, extension, and cervical traction. *Spine (Phila Pa 1976)* 2006;31:44–50.
15. Ishii T, Mukai Y, Hosono N, et al. Kinematics of the cervical spine in lateral bending: in vivo three-dimensional analysis. *Spine (Phila Pa 1976)* 2006;31:155–60.
16. Ishii T, Mukai Y, Hosono N, et al. Kinematics of the subaxial cervical spine in rotation in vivo three-dimensional analysis. *Spine (Phila Pa 1976)* 2004;29:2826–31.
17. Ishii T, Mukai Y, Hosono N, et al. Kinematics of the upper cervical spine in rotation: in vivo three-dimensional analysis. *Spine (Phila Pa 1976)* 2004;29:E139–44.
18. Kozanek M, Wang S, Passias PG, et al. Range of motion and orientation of the lumbar facet joints in vivo. *Spine (Phila Pa 1976)* 2009;34:E689–96.
19. McInerney T, Terzopoulos D. Deformable models in medical image analysis: a survey. *Med Image Anal* 1996;1:91–108.
20. Ochia RS, Inoue N, Renner SM, et al. Three-dimensional in vivo measurement of lumbar spine segmental motion. *Spine (Phila Pa 1976)* 2006;31:2073–8.
21. Cripton PA, Sati M, Orr TE, Bourquin Y, Dumas GA, Nolte LP. Animation of in vitro biomechanical tests. *J Biomech* 2001;34: 1091–6.
22. Barnes PJ, Baldock C, Meikle SR, Fulton RR. Benchmarking of a motion sensing system for medical imaging and radiotherapy. *Phys Med Biol* 2008;53:5845–57.
23. Maletsky LP, Sun J, Morton NA. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. *J Biomech* 2007; 40:682–5.
24. Williams SL, Smith HC. Effect of osteological processing treatments on dimensions and moisture absorption potential of rodent skulls. *Collection Foru* 1995;11:46–57.