Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Validation and Improved Registration of Bone Segmentation Using Contour Coherency

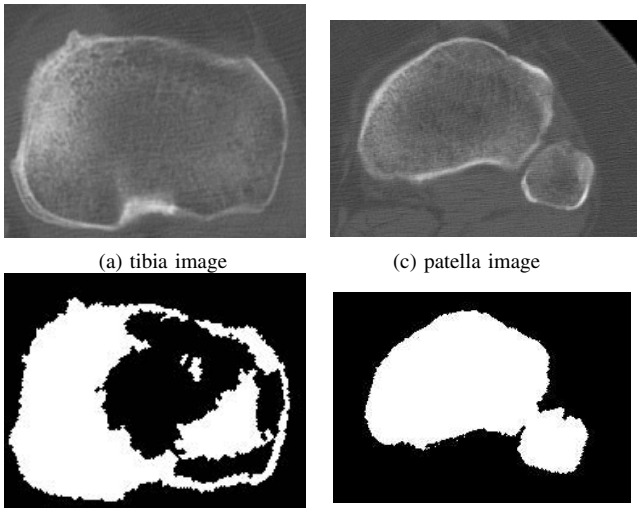
Michael Greenspan[1], Liping Ingrid Wang[2], and Randy Ellis[3]

***Abstract*— A method is presented to validate the segmentation of computed tomography (CT) image sequences, and improve the accuracy and efficiency of the subsequent registration of the 3D surfaces that are reconstructed from the segmented slices. The method compares the shapes of contours extracted from neighborhoods of slices in CT stacks of tibias. The bone is first segmented by an automatic segmentation technique, and the bone contour for each slice is parameterized as a 1-D function of normalized arc length versus inscribed angle. These functions are represented as vectors within a K-dimensional space comprising the first K amplitude coefficients of their Fourier Descriptors. The similarity or *coherency* of neighboring contours is measured by comparing statistical properties of their vector representations within this space. Experimentation has demonstrated this technique to be very effective at automatically identifying low coherency segmentations, the removal of which significantly improved the accuracy and time efficiency of the registration of 3-D bone surface models.**

I. Introduction

The automatic segmentation of bone tissue in a computed tomography (CT) image is an important component of image-based computer assisted orthopedic surgery. Segmentation is a difficult problem due to similarities in the radiological density of bone and surrounding tissues, and despite the large volume of research into this problem [7], [5], [3], [10], [9], [6], the development of a fully automated solution remains a significant challenge. Factors such as the inhomogeneous structure of bone, pathologies, and the inherent blurring of CT data all contribute to the difficulty of fully automated segmentation of bone from other tissues. Two examples of segmentation errors that can occur with fully automated segmentation methods are shown in Fig. 1. In cases where the output of segmentation is used to plan or even directly execute surgeries, segmentation errors could be critical.

We present a method to validate segmentation results that is based upon a comparison of the shapes of contours extracted from neighborhoods of slices in CT stacks of tibias. Medical image segmentation has proven to be a challenging task, and most current automatic methods rely upon manual post-processing [7], [5], [3], [10]. Previous work has shown that fully automatic segmentation sometimes fails due to constraints on image quality and variations of the anatomy of the object of interest. CT slices provide cross-sectional images that clearly distinguish dense cortical bones from soft tissues (mainly fat and muscle). However, difficulties are encountered in extracting contours near joints. This is because the bones may have injuries, calcium loss (osteoporosis), or inhomogeneous structures that cause the bones to be nearly indistinguishable from nearby soft tissues. Furthermore, limitations in the resolution (typically 512x512 pixels) of a CT scan can cause contours of adjacent bones to combine, as has been observed with the tibia and fibula, and at the hip joint [9].



(a) tibia image (c) patella image

(b) segmentation of (a) (d) segmentation of (c)

Fig. 1. Examples of imperfect segmentation results. (b) has a gap in the upper region, and (d) has merged two separate regions.

In this work, we used Fourier Descriptors (FDs) as a metric to characterize the bone contour shapes. The bone was first segmented automatically, following which the bone contours for each slice were parameterized as a 1-D function of normalized arc length versus inscribed angle. These functions were next represented as vectors within a K-dimensional space comprising the first K amplitude coefficients of their FDs.

We suspected that the shapes of neighboring contours would vary gradually, and that this coherence could be used to validate the segmentation. Any deviations from this assumption, resulting from segmentation errors or pathologies, could be measured by the statistical properties of the FDs of neighboring contours, and such segmentations were then flagged as having low coherency. These low coherency

[1]Dept. of Electrical and Computer Engineering/School of Computing, Queen's University, Kingston, Canada, michael.greenspan@queensu.ca

[2]School of Computing, Queen's University, Kingston, Canada, ingridlp@gmail.com

[3]Dept. of Radiology, Harvard Medical School, Boston, MA; School of Computing/Dept. of Surgery, Queen's University, Kingston, Canada, ellis@bwh.harvard.edu

segmentations could be flagged for human inspection or passed to subsequent specialized post-processing routines. Removal of the automatically detected low coherency segmentations was also found to improve both the accuracy and the efficiency of 3-D bone surface models generated from the data.

II. Contour Shape Description

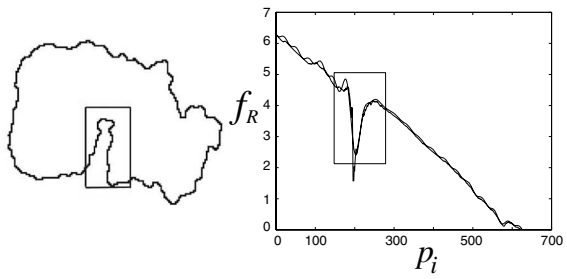
Our objective is to compare the shapes of segmented bone regions in neighborhoods of slices, and to identify those contours that differ sharply from their neighbors. It is possible to compare the segmented regions based solely upon their gross statistics such as size, moments, autocorrelation, etc. [4]. Such comparisons tend, however, to be insensitive to the typical segmentation errors that we wish to identify, such as small gaps and channel closings. Although these segmentation errors may be small, they can cause the merger of two contours or the splitting of a single contour, and can have a large affect on the registration accuracy of the reconstructed 3D surfaces.

We base our technique on a representation of the shape of the contour of each segmented region. Let $\mathcal{C}_R = \{p_i = (x_i, y_i)\}_{i=0}^{N-1}$ be the bounding contour of segmented region R. We assume that $\mathcal{C}_R$ is thin (i.e., only 1 pixel thick), and that the sequence of points has been permuted so that p_i and p_m lie adjacent on $\mathcal{C}_R$ for all $i \in 0 \ldots N-1$ and $m = (i+1) \bmod N$. Visiting the points $p_0 ... p_{N-1}$ in order therefore has the effect of traversing $\mathcal{C}_R$. The choice of p_0 can be arbitrary but, once chosen, the order of the remaining points is distinct; we assume that the permutation describes a clockwise traversal. There are a number of effective algorithms that generate such an ordered contour from a connected region [1].

The *contour function* f_R is defined as the relationship between the L_1 distance of p_i along the perimeter of $\mathcal{C}_R$, and the angle $\alpha(p_i) = \arctan(y_i/x_i)$ measured about the centroid $\mathcal{O}_R$ of $\mathcal{C}_R$ from p_i to the positive x-axis. By convention, p_0 is selected as the smallest point of intersection of $\mathcal{C}_R$ and the positive x-axis.

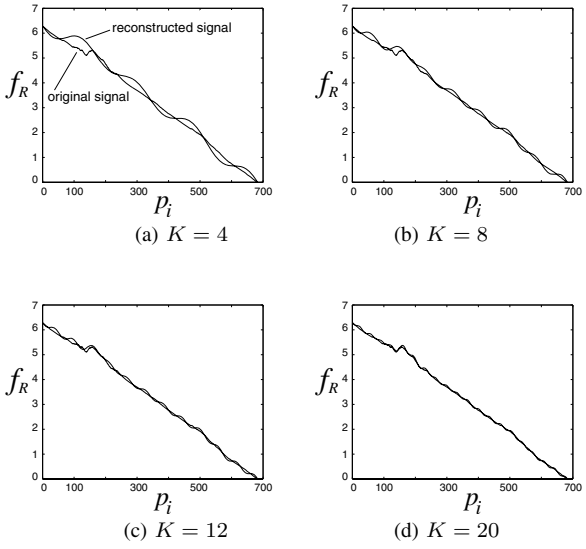
The contour function is a 1-D representation of the 2-D region boundary that can be used to determine the shape similarity between two regions. It is invariant to translations, which do not affect the centroid of $\mathcal{C}_R$, and can be made invariant to scale by normalizing the circumference of $\mathcal{C}_R$. An example of f_R for a segmented tibia is illustrated in Fig. 2. The protrusion at the bottom of the contour, boxed in Fig. 2(a), can be seen as a distinct ripple in the corresponding box in Fig. 2(b).

For the purpose of shape comparison, it is desirable to represent the contour function in a form that is both compact and convenient, and yet captures the salient features of the shape. To this end we turn to Fourier Descriptors (*FDs*), which encode a signal as a series of its frequency domain components. The FDs form a sequence $\{(A_i, \alpha_i)\}_{i=0}^{N-1}$ of amplitude and phase coefficients of the Discrete Fourier



(a) Contour (b) Contour function, $K = 20$

Fig. 2. Original and reconstructed contour function of a proximal-tibia image. Boxed areas are the corresponding regions of the contour and its contour function.



(a) $K = 4$ (b) $K = 8$

(c) $K = 12$ (d) $K = 20$

Fig. 3. Original and reconstructed contour functions for increasing number of Fourier coefficients K.

Transform $F(s)$ of $f_R(x)$, defined as:

$$F(s) = \sum_{x=0}^{N-1} f_R(x) e^{-j2\pi sx/N}, \; s = 0 \ldots N-1 \qquad (1)$$

The FDs are efficient to compute and can be used to exactly reconstruct f_R if all N amplitude and phase coefficients are employed, or to approximately reconstruct it if $K < N$ coefficients are used. As the coefficients are ordered by increasing frequency, using only the first K low-frequency components (the *harmonics*) to compare shapes will have the desirable effect of describing the overall shape of the contour while ignoring noise and small perturbations. It is also desirable to keep K small to improve efficiency.

We determined a reasonable value for K empirically by reconstructing a number of challenging tibia contours. An original and reconstructed contour function for a tibia image is illustrated in Fig. 3 using $K = 4, 8, 12,$ and 20 FDs.

We compared the similarity of two contour functions by considering their first K FDs as a vector in a K-dimensional space and measuring the distance between them. Let f_A and f_B be two contour functions for regions A and B, with respective FDs $\{(A_i, \alpha_i)\}_{i=0}^{N_A-1}$ and $\{(B_i, \beta_i)\}_{i=0}^{N_B-1}$. The similarity between the two contours was measured as the sum of the squared difference:

$$||A, B|| = \frac{1}{K-1} \sum_{i=0}^{K-1} (A_i - B_i)^2 \qquad (2)$$

Although in general $N_A \neq N_B$, the FDs are ordered by increasing frequency so a direct ordinal comparison of the sequence is meaningful. The phase coefficients α_i and β_i are required for signal reconstruction, but because they contribute little to the discrimination of the overall shape of the contour, only the amplitude coefficients are used for the similarity measure. The amplitude coefficients are also invariant to rotation, i.e., the choice of p_0.

III. Experimental Results

We have extensively tested and quantified the performance of the validation method in two separate experiments using two different segmentation methods. The first was a simplification of the method of Kang *et al.* [7], which used both global thresholds and adaptive thresholds based upon the statistics within a local region. The second method was an edge-based technique based upon the estimation and correction of the normal direction of bone edges [11].

In the first experiment, the detection accuracy was determined by comparing the automatic classification results against that of a human expert. In the second experiment, the effects on registration were measured by registering a manually validated clinically approved 3-D surface model of the bone which had been created using a semi-automatic method.

A. Experiment 1: Detection Accuracy

This experiment used 23 distinct sets of CT images of human lower legs (a total of $1,633$ slices) taken with two CT scanners. In each slice the tibia contours were extracted using the two segmentation methods. We especially sought to assess segmentations near the proximal tibia and those segmentations that had incomplete segmented contours due to bone pathologies or limitations of image quality. Starting from the diaphysis of a normal tibia, we observed that the shape changed gradually up to the region where the tibia and fibula join. By contrast, the shape of the tibia near the knee joint changed suddenly and dramatically. The CT slices were spaced less than 5mm apart, so the contour shape of each slice was close to that of its neighbors (except possibly near the knee).

Based on this observation, we used our similarity measurement (Eq.2) to compare each slice with its neighbors. We expected that the contour shape of a normal tibia would resemble those of its neighbors, except when abnormalities were present. The number of slices in the neighborhood over which the coherency is compared needed to be small enough so as to preserve the local details yet large enough to suppress noise. We found that a neighborhood size $S = 5$ was effective, where the neighborhood of the i_{th} slice included slices $i-2, i-1, \ldots\ i+2$.

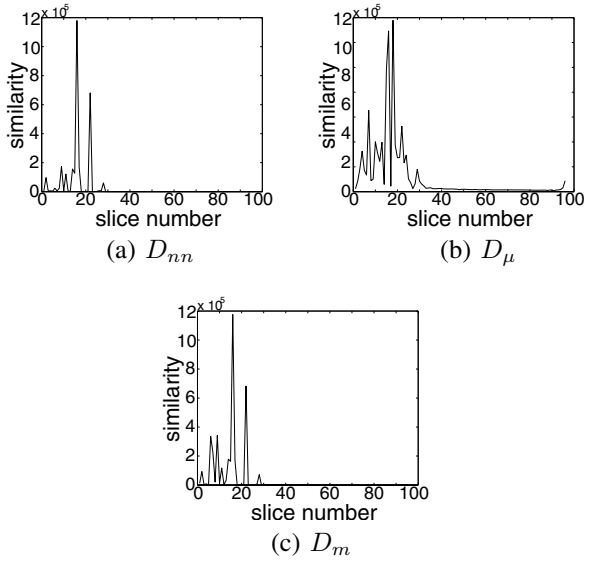


Fig. 4. Contour-function similarity measures for a tibia scan

For each slice, the similarities were calculated for all contours extracted over the neighborhood and the following three measures were determined: D_{nn}, the distance of a contour to its nearest neighbor; D_{μ}, the mean similarity; and D_m, the median similarity.

An example of these three measures for a typical tibia sequence is illustrated in Figure 4. The peaks in Figure 4 indicate that the corresponding slices contained dissimilar shapes, so these are suspected to be bone anomalies. For each measure, a threshold was determined by averaging all except the largest value. A slice was then classified as having low coherency if all three measures exceeded their threshold values for that slice.

Automatic detection was performed for all 23 image sequences and the detection accuracy was evaluated by comparing the results of the automatic validation method against the judgment of a human expert. For example, if the human expert identified an automatically segmented slice as being incorrect, and the automatic validation method also identified this slice as having low coherency, then this contributed to the accuracy of the detected low coherency segmentations. Similarly if an automatic segmentation was approved by the human expert and was also deemed as having a high coherency by the automatic validation method, then the accuracy of detected high coherency segmentations would increase.

For the segmentations by the simplified Kang method, in 11 of the 23 sequences, all of the low-coherency segmentations were correctly detected by the automatic method, and the accuracy of detection was above 80% in all but 4 sequences. For the segmentations by Yao's method, in 10

of the 23 sequences, all of the low-coherency segmentations were correctly detected by the automatic method, and the accuracy of detection was above 80% in all but 7 sequences. The segmentation-validation method therefore resulted in a high detection accuracy for both low-coherency and high-coherency segmentations. Overall, 87.5% and 80% of low-coherency segmentations and 97.7% and 95.5% of high-coherency segmentations were correctly detected for the two respective segmentation methods.

B. Experiment 2: Registration Enhancement

In the second experiment, we evaluated the effectiveness of segmentation validation for enhancing registration. Registration is the process of determining a transformation between two surfaces, and is typically executed on reconstructed surfaces using the Iterative Closest Point (*ICP*) algorithm [2]. In computer-assisted surgery systems, it is important to register the surfaces with a high degree of positional accuracy, particularly for orthopedic surgical procedures. Poorly segmented slices can produce outliers that bias the results, resulting in misregistrations. Detecting and removing these segmentations prior to surface reconstruction can improve the accuracy of the subsequent registration.

To evaluate the improvement to the registration accuracy, we registered bone models that were reconstructed from the contours generated by both segmentation methods. The surfaces were triangulated using an implementation of the Marching Cubes algorithm [8], and reconstructions were generated both before and after the removal of detected low-coherency segmentations. These automatically generated surface models were registered using the ICP algorithm to models that had been semi-automatically reconstructed from the same data by a human expert. We ran this experiment on 4 data sets that had been used clinically, i.e., the semi-automatic segmentations were used in actual computer-assisted surgeries.

Removing the automatically detected low-coherency contours improved the registration accuracy significantly for both segmentation methods, over all of the data sets. On average, the accuracy was improved by 570% (i.e., a factor of 5.7) for the simplified Kang segmentation and by 283% for Yao segmentation. The maximum improvement was 1151%, achieved for a simplified Kang segmentation, while the minimum was 129% for a Yao segmentation.

Much of the computational expense in ICP is in establishing correspondences between surfaces that are highly separated, so the removal of low-coherency segmentations had the added benefit of improving the execution time of the registration. The time performance was improved by 540% and 790% for the simplified Kang and Yao segmentations, respectively, with a maximum and minimum improvement of 1508% and 152%, both for Yao segmentation.

IV. Summary and Conclusions

Our experiments have shown that the proposed method can accurately detect certain classes of low-coherency segments, such as broken bones and the start of the knee bone, in automatically segmented images. We have also shown that the removal of these slices can improve both the accuracy and efficiency of the registration of the reconstructed 3-D bone surfaces against semi-automatically reconstructed models that were manually validated and were of clinical quality.

A number of avenues remain for future research. While the presented contour function is an effective parameterization of a region boundary, there are other methods that could be used to encode and compare contours, such as techniques based on principle-components analysis. Higher-order statistical properties of the contour functions could also be analyzed to further identify particular known classes of segmentation failure. In this way, contours could be identified as landmark features in an anatomical atlas. We have given one example of a particular segmentation method and the failures that can result, but different segmentation methods will likely fail in different situations and different ways. The relationship between a specific segmentation method and the statistical properties of its resulting contours needs to be established. Finally, the notion of shape analysis can be extended directly to the evaluation and comparison of 3-D surfaces, following reconstruction.

Acknowledgements

This work was supported in part by the Natural Sciences and Engineering Council of Canada, The Institute for Robotics and Intelligent Systems, the Ontario Research and Development Challenge Fund, the Center for Innovative Technology for Medicine, and NIH Contract U41RR019703.

References

[1] J. R. Bennett and J. S. Mac Donald. On the measurement of curvature in a quantized environment. *IEEE Transactions on computers*, c-24(8), August 1975.

[2] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on pattern analysis and machine intelligence*, 14(2), February 1992.

[3] A. Elmoutaouakkil, E. Peyrin, J. Elkafi, and A. Laval-Jeantet. Segmentation of cancellous bone from high-resolution computed tomography images: influence on trabecular bone measurements. *IEEE Transactions on medical imaging*, 21(4), April 2002.

[4] R. C. Gonzales and R. E. Woods. *Digital Image Prossing*, chapter 11. Prentice Hall, 2nd edition, 2002.

[5] V. Grau, U. U. J. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on medical imaging*, 23(4), April 2004.

[6] K. Haris, S. Efstratiadis, and A. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Transations on Image Processing*, 7(12), December 1998.

[7] Y. Kang, K. Engelke, and W. A. Kalender. A new accurate and precise 3-d segmentation method for skeletal structures in volumetric ct data. *IEEE Transactions on medical imaging*, 22(5), May 2003.

[8] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Computer Graphics (Proceedings of SIGGRAPH '87)*, volume 21. ACM, 1987.

[9] X.M. Pardo, M.J. Carreira, A. Mosquera, and D. Cabello. A snake for ct image segmentation integrating region and edge information. *Image and Vision Computing*, 19, 2001.

[10] T. B. Sebastian, H. Tek, J. J. Crisco, and B. B. Kimia. Segmentation of carpal bones from ct images using skeletally coupled deformable models. *Medical Image Analysis*, 7, 2003.

[11] W. Yao, P. Abolmaesumi, M. Greenspan, and R. E. Ellis. An estimation/correction algorithm for detecting bone edges in CT images. Submitted to IEEE Trans. Medical Image, January 2004.