

*Osteoarthritis and Cartilage* (2009) 17, 1589–1597

© 2009 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

doi:10.1016/j.joca.2009.06.003

# Osteoarthritis and Cartilage

**International  
Cartilage  
Repair  
Society**

## Intra- and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method

K. T. Bae†\*, H. Shim†, C. Tao†, S. Chang†, J. H. Wang†, R. Boudreau‡ and C. K. Kwoh§||

† Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA

‡ Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA

§ Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, USA

|| Section of Rheumatology, Pittsburgh VA Healthcare System, Pittsburgh, PA, USA

### Summary

**Objective:** We developed a semi-automated method based on a graph-cuts algorithm for segmentation and volumetric measurements of the cartilage from high-resolution knee magnetic resonance (MR) images from the Osteoarthritis Initiative (OAI) database and assessed the intra- and inter-observer reproducibility of measurements obtained via this method.

**Design:** MR image sets from 20 subjects of varying Kellgren–Lawrence (KL) grades (from 0 to IV) on fixed flexion knee radiographs were selected from the baseline double-echo and steady-state (DESS) knee MR images in the OAI database (O.B.1 Imaging Data set). Two trained radiologists independently performed the segmentation of knee cartilage twice using the semi-automated method. The volumes of segmented cartilage were computed and compared. The intra- and inter-observer reproducibility were determined by means of the coefficient of variation (CV%) of repeated cartilage segmented volume measurements. The subjects were also divided into the low- (0, I or II) and high-KL (III or IV) groups. The differences in cartilage volume measurements and CV% within and between the observers were tested with *t* tests.

**Results:** The mean ( $\pm$ SD) intra-observer CV% for the 20 cases was 1.29 ( $\pm$ 1.05)% for observer 1 and 1.67 ( $\pm$ 1.14)% for observer 2, while the mean ( $\pm$ SD) inter-observer CV% was 1.31 ( $\pm$ 1.26)% for session 1 and 1.79 ( $\pm$ 1.72)% for session 2. There was no significant difference between the two intra-observer CV%'s ( $P = 0.272$ ) and between the two inter-observer CV%'s ( $P = 0.353$ ). The mean intra-observer CV% of the low-KL group was significantly smaller than that for the high-KL group for observer 1 (0.83 vs 1.86%;  $P = 0.025$ ). The segmentation processing times used by the two observers were significantly different (observer 1 vs 2): (mean  $49 \pm 12$  vs  $33 \pm 6$  min) for session 1 and ( $49 \pm 8$  vs  $32 \pm 8$  min) for session 2.

**Conclusion:** The semi-automated graph-cuts method allowed us to segment and measure cartilage from high-resolution 3 T MR images of the knee with high intra- and inter-observer reproducibility in subjects with varying severity of OA.

© 2009 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

**Key words:** Knee, MR image, Cartilage segmentation, Volume measurement, Reproducibility.

### Introduction

Recent advances in magnetic resonance (MR) imaging and quantitative image analysis have increased our understanding of the pathophysiologic changes in articular cartilage and bone in osteoarthritis (OA), including the serial monitoring of changes over time<sup>1</sup>. Moreover, recognizing the necessity of new imaging and biochemical markers for OA, the Osteoarthritis Initiative (OAI) was recently launched ([www.niams.nih.gov/oe/oai](http://www.niams.nih.gov/oe/oai)). The OAI cohort consists of almost 4800 participants at four sites who undergo detailed clinical assessment annually that includes high-resolution, state-of-the-art 3 Tesla (3T) MR imaging of both knees over 4 years.

A critical and time-consuming process in the quantitative measurement of cartilage from MR images is the

segmentation of the cartilage region from the surrounding tissue. The task of cartilage segmentation is not trivial, however, because of the complexity of cartilage geometry, its small size, and the inherently low-contrast between the cartilage and surrounding tissues. The most laborious method, which is the technique most widely used in clinical studies, is the manual outlining of cartilage regions on each MR image by an analyst<sup>2–4</sup>. Although the process of manual segmentation is straightforward, it is resource intensive, time-consuming (requiring several hours to segment the cartilage in each knee) and is subject to analyst bias and error. To overcome these limitations, development of an automated approach is highly desirable<sup>5</sup>. To date, however, a fully automated technique for cartilage segmentation has not been established due to cartilage inhomogeneities, low tissue-contrast and shape irregularity, particularly in advanced OA cartilage<sup>1</sup>.

At present, semi-automated segmentation methods that combine the perceptual recognition of a human expert and the reliability of a computer seem the most appropriate and practical approach for cartilage segmentation. Although several semi-automated segmentation methods have been

\*Address correspondence and reprint requests to: K. T. Bae, Department of Radiology, University of Pittsburgh School of Medicine, 3362 Fifth Avenue, Pittsburgh, PA 15213, USA. Tel: 1-412/641-2657; Fax: 1-412/641-2582; E-mail: [baek@upmc.edu](mailto:baek@upmc.edu)

Received 17 December 2008; revision accepted 3 June 2009.

developed, they are still very time-consuming. An improved semi-automated segmentation method is critically needed to process a large quantity of MR imaging data such as that within the OAI data set. Recently, graph-cuts algorithms have been suggested as solutions to a wide range of image processing problems including segmentation<sup>6–8</sup>. While we recognize that graph-cut algorithms work well with the application of segmenting knee cartilage from MR images, to our knowledge no one has reported the use of graph-cut algorithms for this application. Thus, the purpose of this study is to propose a semi-automated method that is based on a graph-cuts algorithm for segmentation and volumetric measurements of the cartilage from high-resolution MR images and to assess the intra- and inter-observer reproducibility of the cartilage volume measurements obtained *via* this method from high-resolution OAI knee MR images in subjects with varying severity of OA.

## Materials and methods

### SEGMENTATION WITH THE SEMI-AUTOMATED GRAPH-CUTS METHOD

An overview of the segmentation process using the graph-cuts algorithm is described in the flowchart shown in Fig. 1.

#### Initialization of seeds

The observer reviews and scrolls through consecutive MR images and initializes the segmentation process by manually placing seeds that indicate cartilage vs non-cartilage (i.e., menisci, synovium, joint fluid) regions [Fig. 2(b)]. The initial seeding process typically starts in the mid-slice of the MR image data set. Algorithmically speaking, these seeds correspond to hard constraints that guide the computation of segmentation and they are particularly crucial for the segmentation of low-contrast, difficult-to-segment regions. In the current implementation, seeds consist of curvilinear lines of different color manually drawn by a radiologist using an electronic pen on a tablet pad. The exact position of the seeds does not affect the correct performance of the algorithm as long as the seeds are placed totally within the region to be segmented. It is not necessary to place seeds on every slice. While seeds are placed on every fifth to tenth slice in regions with relatively constant morphology and high tissue-contrast, they may need to be placed more frequently in order to demarcate and differentiate regions of relatively low tissue-contrast or in areas that are more difficult to segment. Intrinsically high-contrast interfaces such as the cartilage and bone marrow interface on fat-saturation MR images may not necessitate the placement of seeds for segmentation.

#### Propagation of seeds

The seeds placed by the observer over one slice are propagated by the computer program backward and forward to neighboring slices. This process eliminates the requirement of manually placing seeds in every slice and

substantially improves the efficiency of the segmentation. The placed seeds are copied and pasted automatically to the same location in the neighboring images. The copied seeds are thinned by one-pixel width as they propagate and are pasted to each subsequent neighboring image. The rationale for using this pixel-thinning (“decremental thickness”) scheme with the propagated seeds is to reflect the fact that the reliability of the initial seeding reduces with increased anatomical variation as images are further away from the initial seeding image. The propagation of seeds continues to subsequent images until the seeds become thinned out and disappear.

#### Graph-cuts computation

Using the seeds as hard constraints that narrow the search space for possible segmentations, the cartilage region is automatically segmented by means of a graph-cuts algorithm<sup>9</sup> [Fig. 2(c)]. The optimal segmentation is defined as a segmentation that satisfies global optimum criteria which are described mathematically in Appendix.

#### Revision of seeds

While reviewing the initial segmentation result, the observer can revise structures that are deemed to be incompletely or erroneously segmented by placing additional or revised seeds and re-running the graph-cuts computation [Fig. 2(d)]. This iterative reseeding-computation process can be repeated multiple times until the segmented outcome is satisfactory [Fig. 2(e)].

### INTRAS- AND INTER-OBSERVER REPRODUCIBILITY

To evaluate the intra- and inter-observer reproducibility of the segmentation method, MR image sets from 20 subjects were selected from the baseline double-echo and steady-state (DESS) MR images in the OAI database (0.1.1 Clinical Data set and 0.B.1 Imaging Data set). These subjects were chosen arbitrarily but with consideration of representing OA of varying severity as determined by Kellgren–Lawrence (KL) grade in Table I. The KL grade (0–IV) was determined by a central reading of each subject’s fixed flexion posterior–anterior knee radiograph: KL grade 0 = normal with absence of osteophyte or joint space narrowing, while KL grade IV = most severe OA. The number of the subjects in each group of KL grade 0, I, II, III, and IV was 2, 3, 6, 6, and 3, respectively. For the analysis, we divided the KL grades in to low- (0, I or II) and high-KL (III or IV). MR imaging was performed with a 3 T whole-body magnet (Trio; Siemens Medical Systems, Erlangen, Germany) with a quadrature transmit-receive knee coil (USA Instruments, Aurora, OH). Each MR image set is a three dimensional (3-D) sagittal DESS MR image data set with  $140 \times 140 \text{ mm}^2$  field of view (FOV),  $384 \times 384$  in-plane matrix size, 160 slices, and  $0.36 \times 0.36 \times 0.70 \text{ mm}^3$  voxel resolution that provides full coverage of the knee. Other MR acquisition parameters are repetition time (TR)/echo time (TE) = 16.32/4.71 ms, bandwidth = 185 Hz/pixel, flip angle =  $25^\circ$ , and number of averages = 1.

Two trained radiologists (CT, SC) independently performed the segmentation of knee cartilage twice using the semi-automated method. To reduce the memory effect, there was a minimum of a 4-week interval between the initial and repeat segmentation measurements. In addition, we documented the image analysis and processing time required for the segmentation of each case, including both the time spent placing seeds and the computation time with the graph-cuts algorithm. The segmented cartilage of each knee was further divided into three compartments [femur, tibia, and patella; Fig. 3(c)].

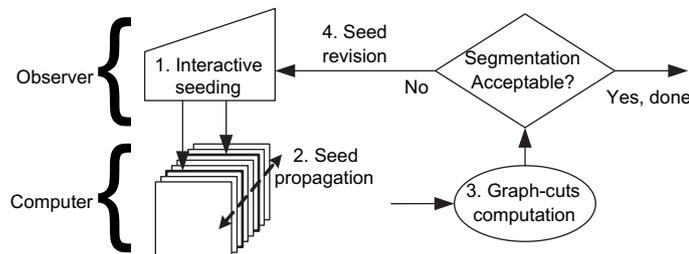


Fig. 1. Overview of the semi-automated graph-cuts segmentation method. An observer interactively places two different types of seeds (one type over the cartilage region and the other over the non-cartilage tissues) on every 5–10 slices. The seeds are automatically propagated to the neighboring slides by the computer program, and the cartilage region is segmented automatically based on the graph-cuts algorithm. While the placement of seeds relies on the expert’s perception and knowledge of knee anatomy, the second and third steps take advantage of the reliability of the computer. If the observer is not satisfied with the initial segmentation boundary result, the boundary can be revised by repeated placement of new seeds and re-execution of the automated segmentation until the results are satisfactory.

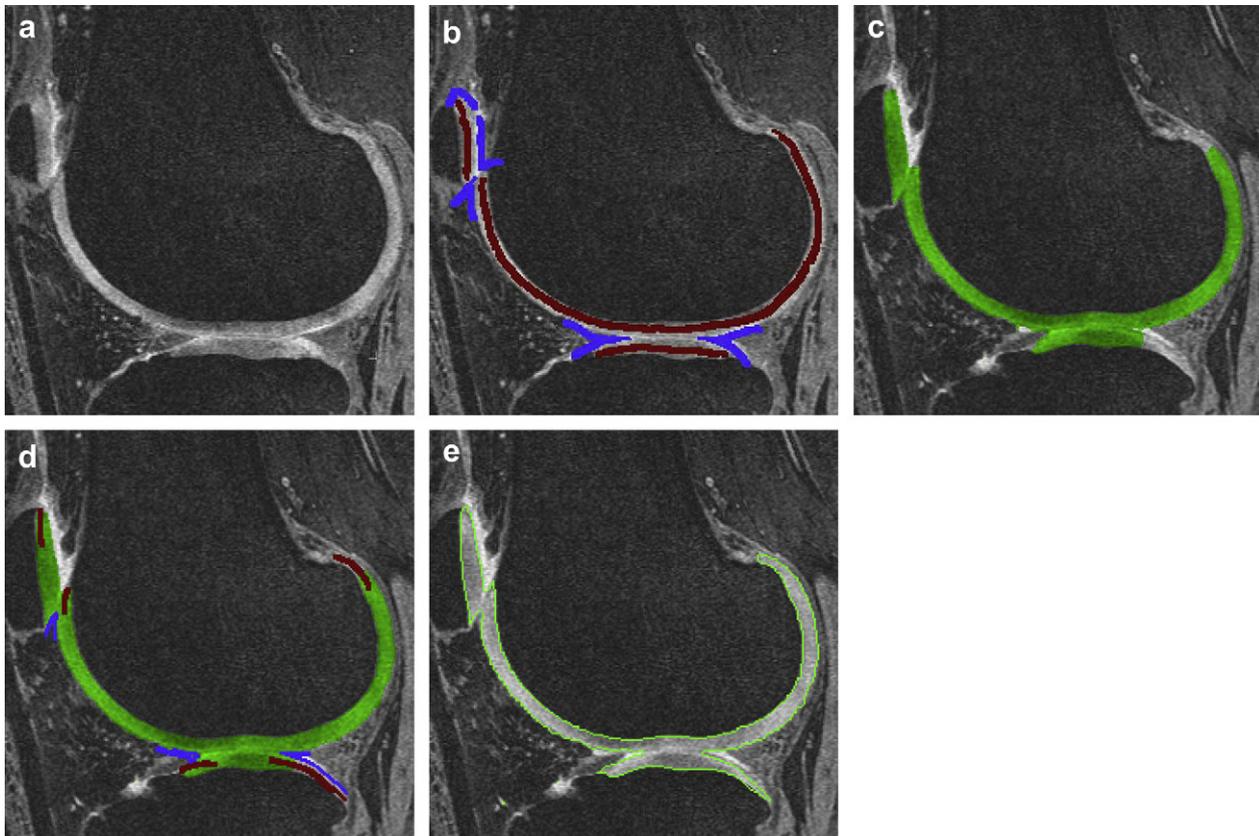


Fig. 2. Sagittal MR images of the knee showing the processing steps of the semi-automated segmentation: (a) original image at slice #47, (b) initial seeding at slice #47, (c) graph-cuts segmentation at slice #53, (d) revised seeding at slice #53, and (e) final segmented cartilage at slice #53. An observer placed seeds over the cartilage regions (red lines) and background regions (e.g., bone, synovium, joint fluid, menisci: blue lines) of the mid-slice at slice #47 (b). Then, the seeds were copied and propagated automatically by the program backward and forward to neighboring slices (images not shown). The cartilage regions were automatically segmented by the computer based on the graph-cuts algorithm (c). Some areas that were incompletely or erroneously segmented after the first seeding and computation of segmentation were detected at slice #53 and subsequently revised (red for cartilage and blue for synovium) (d). After reseeding and re-computation, the segmentation was revised and deemed appropriate (e). Note that, although only one 2-D image slice is shown for illustration, the image processing was carried out 3-D over a volumetric MR image set. The images are from a subject with KL-grade 2.

#### DATA ANALYSIS

For each of the 20 cases, cartilage was segmented twice by the two observers at two separate sessions. The volumes of segmented whole-knee and compartmentalized cartilage were computed. The intra- and inter-observer reproducibility were determined by means of the coefficient of variation (CV%) of repeated cartilage segmented volume measurements. The CV% for repeated measurements (i.e., the SD of paired measurements divided by the mean value computed for each subject, then averaged over subjects and multiplied by 100) is commonly used to describe precision errors in segmented cartilage volume measurements<sup>10,11</sup>. For each of the 20 cases, four CV%'s were calculated and compared: (1) the two repeated measurements by observer 1; (2) the two repeated measurements by observer 2; (3) the first measurements of observer 1 and 2; and (4) the second measurements of observer 1 and 2.

The differences between the repeated cartilage volume measurements within and between the observers were calculated and tested for statistical significance with paired *t* test. The differences in CV% within and between the two observers were also tested with paired *t* test. Furthermore, the 20 cases were divided into two groups according to the KL values, KL 0–II (low-KL group) and KL III–IV (high-KL group). The differences in the intra- and inter-observer volumes and CV% between these two groups were also tested for statistical significance with unpaired *t* test. The statistical analysis was performed with SPSS Statistical Software (Version 16.0, SAS Institute, Inc., Cary, NC). The level of significance was set at 0.05.

#### Results

Two examples of MR images with cartilage segmentation are shown in Fig. 2 (a subject with KL grade II) and Fig. 3

(a subject with KL grade IV). Successful segmentation of cartilage was performed in all 20 cases.

#### INTRA- AND INTER-OBSERVER DIFFERENCES IN VOLUME MEASUREMENTS

The cartilage volume measurements of whole knee by the two observers at two different sessions in the 20 cases are summarized in Table I. The mean  $\pm$  SD volume measurements of the segmented whole-knee cartilage were ( $24.85 \pm 5.76$ ,  $25.26 \pm 5.88$  cc) for (observer 1, 2) at session 1 and ( $25.18 \pm 5.87$ ,  $24.86 \pm 5.81$  cc) for (observer 1, 2) at session 2. The mean and percent intra-observer differences between the two sessions (i.e., session 1 minus session 2) were  $-0.33$  cc and  $-1.30\%$  for observer 1 and  $+0.40$  cc and  $+1.60\%$  for observer 2. These small differences between the sessions were significant in both observers ( $P=0.014$  for observer 1 and  $P=0.008$  for observer 2) and the differences were in opposite directions. In addition, the mean and percent inter-observer differences between the two observers (i.e., observer 1 minus observer 2) were  $-0.41$  cc and  $-1.62\%$  for session 1 and  $+0.32$  cc and  $+1.29\%$  for session 2. These differences between the observers, although small, were also significant in both sessions ( $P=0.004$  for session 1 and  $P=0.021$  for session 2).

Table I  
The whole knee cartilage volumes (cc) measured by the two observers at two separate sessions in the 20 cases of varying KL grades

Group	KL*	Observer 1 (cc)		Observer 2 (cc)		
		Session 1	Session 2	Session 1	Session 2	
Low-KL ( <i>n</i> = 11)	0	17.05	17.22	17.67	17.42	
	0	31.08	31.69	31.39	30.86	
	1	24.56	24.65	25.02	25.71	
	1	23.52	23.71	24.12	23.45	
	1	25.64	25.81	25.63	25.51	
	2	29.67	29.82	30.06	30.00	
	2	25.59	25.51	26.67	25.35	
	2	15.18	15.55	15.37	14.75	
	2	15.83	15.84	16.03	15.88	
	2	26.17	26.79	26.31	26.33	
	2	15.11	15.49	15.30	15.72	
	Mean		<b>22.67</b>	<b>22.92</b>	<b>23.05</b>	<b>22.82</b>
	SD		5.88	5.92	5.93	5.86
High-KL ( <i>n</i> = 9)	3	28.93	28.74	29.14	28.18	
	3	27.48	28.15	29.58	27.94	
	3	29.11	30.59	30.35	29.96	
	3	23.91	23.42	24.27	24.51	
	3	26.29	25.98	26.18	25.15	
	3	19.23	19.98	18.79	18.34	
	4	36.10	36.89	36.32	36.74	
	4	30.02	29.70	30.07	29.44	
	4	26.56	28.01	26.87	25.86	
	Mean		<b>27.51</b>	<b>27.94</b>	<b>27.95</b>	<b>27.35</b>
	SD		4.59	4.72	4.82	4.96
	All KL ( <i>n</i> = 20)	Mean	<b>24.85</b>	<b>25.18</b>	<b>25.26</b>	<b>24.86</b>
		SD	5.76	5.87	5.88	5.81
<i>P</i> -values	Low- vs high-KL	0.059	0.054	0.061	0.082	
Mean & % differences	Intra-observer (session 1 minus 2)	-0.33 (-1.30%); <i>P</i> = <b>0.014</b>		0.40 (1.60%); <i>P</i> = <b>0.008</b>		

\*KL grade of severity of OA (0–4: 0 = normal, 4 = most severe).

When the low- (0, I or II) and high-KL (III or IV) groups were compared, the mean cartilage volumes of whole knee of the low-KL group were consistently lower than those for the high-KL group in all comparisons (Table I). However, none of them reached statistical significance ( $P=0.054-0.082$ ). By compartment, the same trend of the mean cartilage volumes of the low-KL group lower than those for the high-KL group was observed in all comparisons. A statistical significance was observed in the femoral compartment ( $P=0.34-0.050$ ) not in the tibial ( $P=0.358-0.509$ ) and patellar ( $P=0.121-0.245$ ) compartments.

#### INTRA- AND INTER-OBSERVER REPRODUCIBILITY (CV%)

The intra- and inter-observer CV%'s of the whole knee cartilage volume measurements for the 20 cases are summarized in Table II. The mean ( $\pm$ SD) intra-observer CV% for the 20 cases was 1.29( $\pm$ 1.05)% for observer 1 and 1.67( $\pm$ 1.14)% for observer 2. These relatively small CV% values indicate that cartilage volumes measured with the semi-automated method were highly reproducible within each observer. The reproducibility for observer 1 was slightly higher than that of observer 2 but this difference was not statistically significant ( $P=0.272$ ). The mean ( $\pm$ SD) inter-observer CV% for the 20 cases was 1.31( $\pm$ 1.26)% for session 1 and 1.79( $\pm$ 1.72)% for session 2. This indicates that the whole knee cartilage volume measurements were highly reproducible between the two

observers with the semi-automated method. The differences between the two sessions were not statistically significant ( $P=0.353$ ).

By compartment, the mean ( $\pm$ SD) intra- and inter-observer CV% of each compartmental cartilage volume measurements is summarized in Table III. All mean values were within 5%. The maximum mean intra-observer CV% was 3.63% for the high-KL patella compartment, while the maximum mean inter-observer CV% was 4.7% for the same group. No statistical significant reproducibility differences or trends for each of the three compartments were detected between the two observers or between the sessions.

When the low- and high-KL groups were compared for the whole knee cartilage volume measurements, the mean intra-observer CV% of the low-KL group tended to be smaller than that for the high-KL group for both observers (low- vs high-KL): (0.83 vs 1.86%) for observer 1 and (1.43 vs 1.95%) for observer 2. This was statistically significant only for observer 1 ( $P=0.025$ ) but not for observer 2 ( $P=0.320$ ). Similar to the intra-observer CV%, the mean inter-observer CV% of the low-KL group tended to be smaller than that for the high-KL group for both sessions (low- vs high-KL): (1.20 vs 1.45%) for session 1 and (1.30 vs 2.39%) for session 2. However, neither the session 1 ( $P=0.676$ ) nor session 2 ( $P=0.164$ ) differences in inter-observer CV% for low-KL vs high-KL grade cases were statistically significant. The comparisons by compartment between the low- and high-KL groups are shown in Table III. The mean intra-observer CV% of the low-KL group

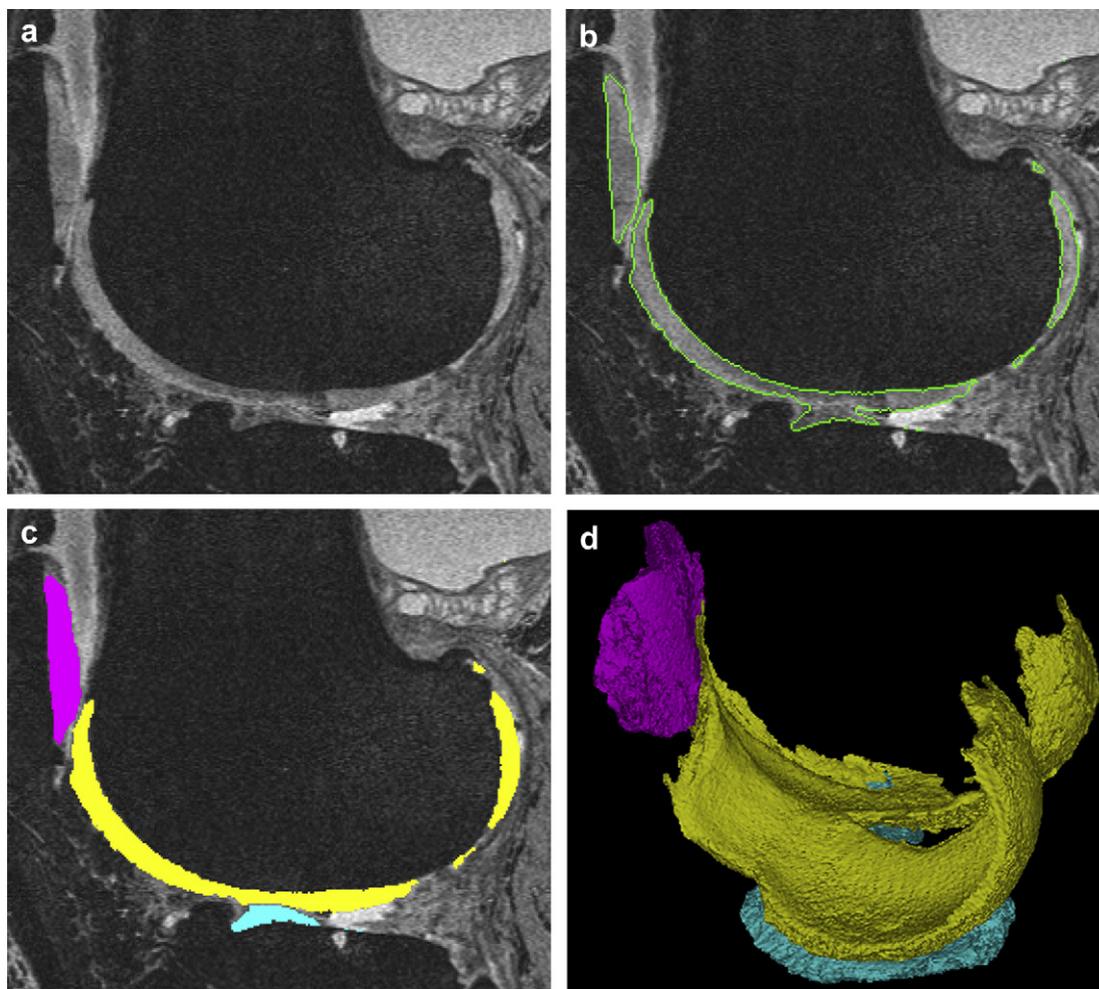


Fig. 3. Sagittal MR images of the knee: (a) original image, (b) segmented cartilage boundary, (c) cartilage compartmentalized into femur (yellow), tibia (cyan), and patella (pink), and (d) 3-D representation of the segmented cartilage. The images are from a subject with KL grade 4. The femoral cartilage shows increased thinning and focal defects, compared to that in Fig. 2 from a subject with KL grade 2.

was smaller than that for the high-KL group in all but one (in the tibial compartment by observer 1). Statistically significant differences were noted in two comparisons: the femoral compartment by observer 1 ( $P=0.013$ ) and the patellar compartment by observer 2 ( $P=0.017$ ). For the inter-observer CV%, the mean of the low-KL group was consistently smaller than that for the high-KL group without an exception. However, statistically significant difference was observed only in the patellar compartment in session 2 ( $P=0.017$ ).

When the intra- and inter-observer CV% were compared for the whole knee cartilage volume measurements within the low-KL or high-KL group, none of the comparisons were statistically significant. The  $P$ -values for the intra-observer CV% comparison between observers 1 and 2 were 0.154 for the low-KL and 0.872 for the high-KL group, while those for the inter-observer CV% comparison between sessions 1 and 2 were 0.835 for the low-KL and 0.362 for the high-KL group.

#### TIME USED FOR SEGMENTATION

The mean  $\pm$  SD segmentation processing times spent on the 20 cases by the two observers at two sessions were

(observer 1 vs 2) ( $49.1 \pm 11.7$  vs  $33.5 \pm 5.9$  min) for session 1 and ( $49.3 \pm 8.2$  vs  $31.8 \pm 7.9$  min) for session 2. In all but one case, observer 1 spent more time than observer 2 in segmentation ( $P < 0.001$ ). The mean  $\pm$  SD segmentation processing times broken down between the low- and high-KL groups were (low-KL vs high-KL,  $P$ -value) ( $50.2 \pm 14.0$  vs  $47.8 \pm 8.6$  min,  $P=0.660$ ) for observer 1 at session 1; ( $48.4 \pm 8.9$  vs  $50.3 \pm 7.5$  min,  $P=0.604$ ) for observer 1 at session 2; ( $30.6 \pm 4.8$  vs  $37.0 \pm 5.3$  min,  $P=0.011$ ) for observer 2 at session 1; and ( $31.5 \pm 8.5$  vs  $32.1 \pm 7.7$  min,  $P=0.860$ ) for observer 2 at session 2. Thus, the significant difference in the segmentation time between the low- and high-KL groups was noted only in observer 2 at session 1.

Since only 1–2 min was required for the computation of the segmentation by the graph-cuts algorithm in each case, more than 90% of the semi-automated segmentation processing time was used by the observers to place seeds. The number of iterations for reseeded and computation was tracked for each semi-automated segmentation process; it varied according to the observer and the complexity of the knee anatomy of the test cases. The mean number ( $\pm$ SD, min–max) of iterations for reseeded and computation was 4.7 ( $\pm 0.66$ , 2–5) for observer 1 and 3.5 ( $\pm 0.85$ , 2–5) for observer 2.

Table II  
CV%'s of repeated measurements of the whole knee cartilage volumes by the two observers at two separate sessions in the 20 cases of varying KL grades

Group	KL*	Intra-observer		Inter-observer	
		Observer 1	Observer 2	Session 1	Session 2
Low-KL ( <i>n</i> = 11)	0	0.70	1.01	2.53	0.82
	0	1.37	1.20	0.70	1.88
	1	0.26	1.92	1.31	2.98
	1	0.57	1.99	1.78	0.78
	1	0.47	0.33	0.03	0.83
	2	0.36	0.14	0.92	0.43
	2	0.22	3.59	2.90	0.44
	2	1.70	2.91	0.88	3.73
	2	0.04	0.66	0.89	0.18
	2	1.66	0.05	0.38	1.22
	2	1.76	1.91	0.88	1.04
	Mean	<b>0.83</b>	<b>1.43</b>	<b>1.20</b>	<b>1.30</b>
	SD	0.66	1.15	0.88	1.12
High-KL ( <i>n</i> = 9)	3	0.47	2.37	0.51	1.39
	3	1.70	4.03	5.20	0.53
	3	3.51	0.91	2.95	1.47
	3	1.46	0.70	1.06	3.22
	3	0.84	2.84	0.30	2.30
	3	2.71	1.71	1.64	6.05
	4	1.53	0.81	0.43	0.29
	4	0.76	1.51	0.12	0.62
	4	3.76	2.71	0.82	5.64
		Mean	<b>1.86</b>	<b>1.95</b>	<b>1.45</b>
	SD	1.20	1.12	1.66	2.17
All KL ( <i>n</i> = 20)	Mean	<b>1.29</b>	<b>1.67</b>	<b>1.31</b>	<b>1.79</b>
	SD	1.05	1.14	1.26	1.72
<i>P</i> -values	Low- vs high-KL	<b>0.025</b>	0.320	0.676	0.164
<i>P</i> -values	Intra-observer observer 1 vs 2	0.272			
<i>P</i> -values	Inter-observer session 1 vs 2			0.353	

\*KL grade of severity of OA (0–4: 0 = normal, 4 = most severe).

## Discussion

### SEMI-AUTOMATED SEGMENTATION OF CARTILAGE

Degradation and eventual loss of articular cartilage are believed to be crucial elements in the pathophysiology of OA<sup>12</sup>. Morphometric assessment of cartilage structure such as volume, thickness, and area using MR images has been shown to provide an accurate and precise measure for OA progression<sup>1</sup>. To derive these measurements, cartilage needs to be accurately and precisely segmented from surrounding tissues. The task of cartilage segmentation is not trivial because of the complexity of cartilage geometry, small size, and the inherently low-contrast between the cartilage and surrounding tissues.

The most laborious method is the manual outlining of cartilage regions on each MR image by an analyst<sup>2–4</sup>, which is the technique most widely used in clinical studies. Although the process of manual segmentation is straightforward, it is resource intensive, time-consuming (requiring several hours to segment the cartilage in each knee from high-resolution MR images) and is subject to analyst bias and error. To overcome these limitations, development of an automated approach is highly desirable<sup>5</sup>. To date, however, a fully automated technique for cartilage segmentation has not been established due to cartilage inhomogeneities, low tissue-contrast and shape irregularity, particularly in advanced OA cartilage<sup>1</sup>.

At present, semi-automated segmentation methods that combine the perceptual recognition of a human expert and the reliability of a computer seem the most appropriate and practical approach for cartilage segmentation. An optimally implemented semi-automated segmentation method will yield accurate and precise segmentation results in a reasonable amount of time with minimal variability. A number of semi-automated segmentation techniques have been proposed and published, including edge detection<sup>5</sup>, region growing<sup>4,13–17</sup>, active shape models<sup>16</sup>, b-spline snakes<sup>18–21</sup>, live wires<sup>22</sup>, and multispectral analysis<sup>23</sup>. An extensive review of these methods is provided in Eckstein *et al.*<sup>10</sup>. The success of these techniques has been limited, in part because they were developed prior to the advent of a high-resolution MR imaging (3 T) equipped with newer imaging sequences and superior image processing algorithms. As a result, despite a plethora of semi-automated methods, most large-scale studies have still relied on manual segmentation of cartilage structures. Some investigators<sup>10</sup> found that correction of the results of these semi-automated methods often took longer and was no more reliable than manual segmentation by skilled experts.

An improved semi-automated segmentation method is critically needed to process a large quantity of MR imaging data such as that within the OAI data set and other longitudinal studies. Recently, graph-cuts algorithms have been suggested as solutions to a wide range of image processing

Table III  
CV%’s of repeated measurements of the compartmentalized cartilage volumes of each knee by the two observers at two separate sessions in the 20 cases of varying KL grades

Group	Femur				Tibia				Patella			
	Intra-observer		Inter-observer		Intra-observer		Inter-observer		Intra-observer		Inter-observer	
	O1	O2	S1	S2	O1	O2	S1	S2	O1	O2	S1	S2
Low-KL (n = 11)	Mean	1.27	0.69	1.06	2.51	2.68	3.02	2.30	1.72	1.44	1.51	2.20
	SD	1.19	0.55	0.99	1.58	2.19	3.24	1.85	0.90	0.91	1.42	1.89
High-KL (n = 9)	Mean	1.84	1.25	2.82	2.13	3.02	4.23	2.71	2.78	3.63	2.68	4.70
	SD	0.83	0.88	2.58	2.23	3.21	3.37	1.84	2.18	2.57	2.88	2.34
All KL (n = 20)	Mean	1.53	0.94	1.85	2.34	2.84	3.57	2.49	2.20	2.42	2.04	3.32
	SD	1.06	0.75	2.03	1.86	2.62	3.27	1.81	1.65	2.11	2.21	2.41
P-values	0.013		0.095		0.664		0.422		0.156		0.253	
P-values	0.254		0.055		0.519		0.262		0.638		0.056	
P-values												

O1: observer 1, O2: observer 2, S1: session 1, S2: session 2.

problems including segmentation<sup>6-8</sup>. While we recognize that graph-cut algorithms work well with the application of segmenting knee cartilage from MR images, to our knowledge no one has reported the use of graph-cut algorithms for this application. Thus, based on graph-cut framework, we have specifically developed and implemented a semi-automated segmentation method for the segmentation of knee cartilage. In this method, a user provides a simple manual initialization of the bone, cartilage and non-cartilage regions on sampled MR image sections. Cartilage is then automatically segmented by the graph-cuts algorithm on the whole volume of sections without requiring further user input. The segmented cartilage boundaries are then reviewed and revised as necessary by an expert observer. This interactive semi-automated approach, requiring only simple initiation and minimal interaction by the user, improves the efficiency and reproducibility of the cartilage segmentation process.

INTRA- AND INTER-OBSERVER DIFFERENCES AND REPRODUCIBILITY

The results of our study demonstrate that knee cartilage can be segmented efficiently and reproducibly from 3 T high-resolution MR images using the semi-automated graph-cuts segmentation method that we have developed. The cartilage volume measurements were highly correlated and were closely matched within and between the readers. Nevertheless, there were small but statistically significant differences in the volume measurements between the observers and between the sessions. This shows the importance of measuring serial cartilage volumes by a single observer in a single session to reduce any precision errors and biases, particularly when the changes in cartilage volumes are small<sup>1</sup>. With our segmentation method, however, the differences were only on the order of 1.3–1.6% (in actual estimated volume) whether comparing two repeat segmentations of knee cartilage from the same MR image or comparing two readers segmenting the same MR image.

The cartilage volume measurements in our study were highly consistent and reproducible. The mean intra- and inter-observer CV%’s for the whole knee cartilage volume measurements were better than those for the compartmentalized (femur, tibia, patella) cartilage volume measurements. This is not surprising because additional variations are likely introduced during the cartilage compartmentalization process. When we compare the reproducibility of the three compartmentalized cartilage volume measurements in our study with the reported CV%’s (range of 2–8%) in various studies of knee cartilage segmentation and volumetric measurement from knee MR images<sup>4,10,11,20,24</sup>, our results for the mean intra- (1.18–3.57%) and inter-observer CV%’s (0.94–3.32%) are comparable or better than with the published results.

The mean image processing time of 30–50 min for the segmentation of the cartilage from the high-resolution MR image set of 160 slices is much faster than the segmentation by the manual delineation of cartilage on each of 160 slices. Although the comparison of the segmentation processing time with other semi-automated segmentation techniques is difficult because very few studies report this variable, our method seems twice faster than 75.4 min per knee published by Duryea *et al.*<sup>24</sup> and approximately 10 times faster than that reported (57–78 min for a single patella or medial tibial plateau) by McWalter *et al.*<sup>25</sup>. In addition, because over 90% of the semi-automated segmentation processing time was

used by the observer to place seeds, the processing time may be shortened further in the future with refinement of graph-cuts algorithm. Seed placement can be improved and automated with use of pattern recognition techniques (e.g., use of *a priori* information on shape, location and signal intensity of the cartilage).

Although the mean intra-observer CV% of observer 1 was lower than that for observer 2, no statistically significant differences were found in CV% within and between the observers. This suggests that the cartilage volume can be measured reproducibly using the semi-automated method without being significantly affected by the performance of individual trained observers. The intra-observer CV% may be reduced with more meticulous placement of seeding as evidenced by longer processing times spent by observer 1 (mean 49 min) than observer 2 (33 min). Such a trade-off between efficiency and reproducibility is not surprising but was not investigated in the current study.

The intra- and inter-observer CV%'s for the low-KL group were consistently lower than those for the high-KL group in both whole-knee cartilage and compartmentalized cartilage volume measurements. This finding may reflect that the segmentation of cartilage in advanced knee OA is less reproducible and more difficult than that for less advanced knee OA. No definite trend for the low- and high-KL groups was observed in the image processing time.

The intra- and inter-observer CV%'s of the semi-automated method were not 0%. In the manual segmentation method, each manually drawn boundary corresponds to the final segmentation outcome. Any variations in manual boundary delineation will likely result in reduced reproducibility and affect the precise quantitative measurement of knee cartilage. On the other hand, in our method, while a large portion of the segmentation operation was performed algorithmically by the computer based on the graph-cuts algorithm, the placement of seeds was determined subjectively by an observer. This subjective interpretation of what is considered cartilage vs background, particularly in the regions of low tissue-signal, would affect intra- and inter-reproducibility.

In conclusion, the semi-automated graph-cuts method allows us to segment and measure cartilage from high-resolution 3 T MR images of the knee with high intra- and inter-observer reproducibility in subjects with varying severity of OA.

## Conflict of interest

The authors have no conflict of interest. The author, CKK received research grant from AstraZeneca.

## Acknowledgment

This study was supported by a Research Grant from the Western Pennsylvania Chapter of the Arthritis Foundation.

## Appendix: Description of global optimum criteria in graph-cuts algorithm

A segmentation of a 3-D volume image  $V$  can be represented as a vector  $X = (x_1, \dots, x_v, \dots, x_{|V|})$ , where  $v$  represents a voxel within  $V$  and each binary component  $x_v$  is assigned to one of two values, either "cartilage" or "background". Then, the optimal segmentation is the one that

minimizes a cost function which is commonly described for the graph-cuts image segmentation as follows:

$$E(X) = \sum_{v \in V} R_v(x_v) + \sum_{(u,v) \in N} B_{uv}(x_u, x_v). \quad (1)$$

The first term  $R_v(x_v)$  is called as the data-dependent term or the regional cost and represents the individual cost that incurs when each voxel  $v$  is assigned to the "cartilage" or the "background", i.e.,  $R_v$ ("cartilage") or  $R_v$ ("background"). These regional costs are typically expressed by the negative logarithm value of the likelihood of voxel  $v$  being either "cartilage" or "background", which is the conditional probability that voxel  $v$  has intensity value of  $I_v$ , given that voxel  $v$  belongs to either "cartilage" or "background", i.e.,  $R_v$ ("cartilage") =  $-\ln(P(I_v | \text{"cartilage"})$ ) and  $R_v$ ("background") =  $-\ln(P(I_v | \text{"background"}))$ .

These two conditional probabilities are computed using the two histograms that have been acquired from intensity values of the two types of seed voxels ("cartilage" and "background")<sup>26</sup>.

The smoothness term or the boundary cost term,  $B_{uv}(x_u, x_v)$ , represents the cost associated with a discontinuity between neighboring voxels  $u$  and  $v$  as defined within a neighborhood system  $N$  (i.e., the 18-connectivity in the current implementation) and is described by the following *ad hoc* function<sup>6</sup>:

$$B_{uv}(x_u, x_v) = \exp\left(-\frac{(I_u - I_v)^2}{2\sigma^2}\right) \frac{1}{\text{dist}(u, v)}. \quad (2)$$

This cost function assigns a high value to discontinuity between the voxels of similar intensities ( $|I_u - I_v| < \sigma$ ) but a low value to discontinuity between the voxels of dissimilar intensities ( $|I_u - I_v| > \sigma$ ). Consequently, the optimal segmentation  $X_{opt}$  is designed to cut through edges whose voxels are of different intensity values. The  $\sigma$  represents the standard deviation of the noise calculated from the air region outside of the body. In addition, the boundary cost function is formulated inversely proportional to the spatial distance between the voxels,  $\text{dist}(u, v)$ .

## References

1. Eckstein F, Cicuttini F, Raynauld JP, Waterton JC, Peterfy C. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. *Osteoarthritis Cartilage* 2006; 14(Suppl A):A46–75.
2. Cibere J, Bellamy N, Thorne A, Esdaile JM, McGorm KJ, Chalmers A, *et al.* Reliability of the knee examination in osteoarthritis: effect of standardization. *Arthritis Rheum* 2004;50:458–68.
3. Cicuttini F, Forbes A, Morris K, Darling S, Bailey M, Stuckey S. Gender differences in knee cartilage volume as measured by magnetic resonance imaging. *Osteoarthritis Cartilage* 1999;7:265–71.
4. Peterfy CG, van Dijke CF, Janzen DL, Gluer CC, Namba R, Majumdar S, *et al.* Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: optimization and validation. *Radiology* 1994;192:485–91.
5. Kshirsagar AA, Watson PJ, Tyler JA, Hall LD. Measurement of localized cartilage volume and thickness of human knee joints by computer analysis of three-dimensional magnetic resonance images. *Invest Radiol* 1998;33:289–99.
6. Boykov Y, Jolly MP. Interactive graph cuts for optimal boundary and region segmentation of objects in N–D images. In: *International Conference on Computer Vision*. Volume 1. 2001, pp. 105–112.
7. Rother C, Kolmogorov V, Blake A. Grabcut – interactive foreground extraction using iterated graph cuts. In: *International Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH)*. Los Angeles, CA: ACM; 2004:309–14.
8. Kumar MP, Torr PHS, Zisserman A. OBJ CUT. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Volume 1. 2005, pp. 18–25.

9. Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. *Int J Comp Vis* 2006;70:109–31.
10. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. *NMR Biomed* 2006;19: 822–54.
11. Blumenkrantz G, Majumdar S. Quantitative magnetic resonance imaging of articular cartilage in osteoarthritis. *Eur Cell Mater* 2007;13: 76–86.
12. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, *et al.* Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med* 2000;133:635–46.
13. Piplani MA, Disler DG, McCauley TR, Holmes TJ, Cousins JP. Articular cartilage volume in the knee: semiautomated determination from three-dimensional reformations of MR images. *Radiology* 1996;198: 855–9.
14. Eckstein F, Gavazzeni A, Sittek H, Haubner M, Losch A, Milz S, *et al.* Determination of knee joint cartilage thickness using three-dimensional magnetic resonance chondro-crassometry (3D MR-CCM). *Magn Reson Med* 1996;36:256–65.
15. Eckstein F, Schnier M, Haubner M, Priebsch J, Glaser C, Englmeier KH, *et al.* Accuracy of cartilage volume and thickness measurements with magnetic resonance imaging. *Clin Orthop Relat Res* 1998;137–48.
16. Solloway S, Hutchinson CE, Waterton JC, Taylor CJ. The use of active shape models for making thickness measurements of articular cartilage from MR images. *Magn Reson Med* 1997;37:943–52.
17. Waterton JC, Solloway S, Foster JE, Keen MC, Gandy S, Middleton BJ, *et al.* Diurnal variation in the femoral articular cartilage of the knee in young adult humans. *Magn Reson Med* 2000;43:126–32.
18. Kauffmann C, Gravel P, Godbout B, Gravel A, Beaudoin G, Raynauld JP, *et al.* Computer-aided method for quantification of cartilage thickness and volume changes using MRI: validation study using a synthetic model. *IEEE Trans Biomed Eng* 2003;50:978–88.
19. Burgkart R, Glaser C, Hyhlik-Durr A, Englmeier KH, Reiser M, Eckstein F. Magnetic resonance imaging-based assessment of cartilage loss in severe osteoarthritis: accuracy, precision, and diagnostic value. *Arthritis Rheum* 2001;44:2072–7.
20. Stammberger T, Eckstein F, Michaelis M, Englmeier KH, Reiser M. Inter-observer reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. *Magn Reson Imaging* 1999;17:1033–42.
21. Lynch JA, Zaim S, Zhao J, Stork A, Peterfy CG, Genant HK. Cartilage segmentation of 3D MRI scans of the osteoarthritic knee combining user knowledge and active contours. In: *SPIE Medical Imaging*. Volume 3979. San Diego: 2000, pp. 925–935.
22. Gougoutas AJ, Wheaton AJ, Borthakur A, Shapiro EM, Kneeland JB, Udupa JK, *et al.* Cartilage volume quantification via Live Wire segmentation. *Acad Radiol* 2004;11:1389–95.
23. Tamez-Pena JG, Barbu-McInnis M, Totterman S. Knee cartilage extraction and bone-cartilage interface analysis from 3D MRI data sets. In: *SPIE: Medical Imaging 2004*. Volume 5370. San Diego, CA: 2004, pp. 1774–1784.
24. Duryea J, Neumann G, Brem MH, Koh W, Noorbakhsh F, Jackson RD, *et al.* Novel fast semi-automated software to segment cartilage for knee MR acquisitions. *Osteoarthritis Cartilage* 2007;15:487–92.
25. McWalter EJ, Wirth W, Siebert M, von Eisenhart-Rothe RM, Hudelmaier M, Wilson DR, *et al.* Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. *Osteoarthritis Cartilage* 2005;13:48–53.
26. Boykov Y, Veksler O, Zabih R. Markov random fields with efficient approximations. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 1998, pp. 648–655.