Model-based Auto-Segmentation of Knee Bones and Cartilage in MRI Data

Heiko Seim, Dagmar Kainmueller, Hans Lamecker, Matthias Bindernagel, Jana Malinowski, and Stefan Zachow

Zuse Institute Berlin (ZIB), Medical Planning Group, Takustr. 7, 14195 Berlin, Germany
{seim,kainmueller,lamecker,bindernagel,malinowski,zachow}@zib.de
http://www.zib.de/visual/medical

Abstract. We present a method for fully automatic segmentation of the bones and cartilages of the human knee from MRI data. Based on statistical shape models and graph-based optimization, first the femoral and tibial bone surfaces are reconstructed. Starting from the bone surfaces the cartilages are segmented simultaneously with a multi object technique using prior knowledge on the variation of cartilage thickness. We validate our method on 40 clinical MRI datasets acquired before knee replacement.

Keywords: Segmentation, Statistical shape model, Magnetic resonance imaging, Knee, Cartilage, Femur, Tibia, Joint

1 Introduction

Osteoarthritis (OA) is a disabiling disease affecting more than one third of the population over the age of 60. Monitoring the progression of OA or the response to structure modifying drugs requires exact quantification of the knee cartilage by measuring e.g. the bone interface, the cartilage thickness or the cartilage volume [4]. Manual delineation for detailed assessment of knee bones and cartilage morphology, as it is often performed in clinical routine, is generally irreproducible and labor intensive with reconstruction times up to several hours [10].

Due to the increasing availability of MRI scanners in clinical routine the demand for automatic segmentation of knee bone and cartilage tissue is growing. Though semi-automatic approaches remain necessary and useful in their own right, we focus here on fully automatic methods. To this end, an approach for segmentation of articular cartilage based on supervised learning was presented in [1], where an evaluation on 46 MRI datasets with no or mild OA symptoms resulted in an average Dice similarity coefficient (DSC) of 0.80, a sensitivity of 90.0% and specificity of 99.8%. A similar approach in combination with an elastic registration scheme was presented in [11] but only evaluated on a single MRI dataset. Fitting of a probabilistic atlas to MRI data exploiting linear programming was presented in [3] to segment the cartilage of the patella, achieving a DSC

215
of 0.84, a sensitivity of 94.1%, a specificity of 99.9% and an average surface distance of 0.49 mm evaluated on 28 MRI datasets from unspecified sources. In [2], statistical shape models (SSMs) are used to extract the bone-cartilage interface (BCI). 20 MRI datasets from healthy subjects were segmented with DSCs of 0.96, 0.96 and 0.89 for the femur, the tibia and the patella, respectively, and an average point-to-surface error of 0.16 mm on the BCI. A multi-object graph-based approach to segment the bones and the cartilage of the knee with minor manual interaction (≈ 30 sec) was proposed in [12], where an evaluation on 16 MRI datasets showed average surface positioning errors of 0.2 to 0.3 mm for the bones and 0.5 to 0.8 mm for the cartilage.

Comparing the performance of these different approaches is difficult, because evaluation results strongly depend on varying properties and origins of the employed image data (e.g. MR sequences, varying types of pathologies) as well as different evaluation metrics. Very recently, a standardized benchmark for automatic knee segmentation systems has been introduced [5]. In this paper, we propose a new model-based segmentation approach and validate it using the data and evaluation metrics put forth in [5].

2 Auto-Segmentation System for Knee MRI

The general outline of our automatic segmentation system is shown in Fig. 1. It consists of two major parts: (1) An SSM-based reconstruction method for bone surfaces [9] applied to knee bones, and (2) a method for simultaneous segmentation of adjacent structures via shared displacement directions [6] applied to the tibial/femoral cartilage. In the following, due to space limitations we only describe the modifications with which we adapted these two methods to the application-specific situation of bone and cartilage in knee MRI. For more detail we refer the reader to [9] and [6].
2.1 Statistical Shape Models

The statistical shape models of proximal tibia and distal femur used in this work were generated from 60 MRI datasets provided by the MICCAI 2010 Workshop Medical Image Analysis for the Clinic - A Grand Challenge [5]. For each of these datasets an expert segmentation was available for the bones and the associated cartilage. We fitted existing SSMs of the femur and the tibia to the gold-standard segmentation using the method in [9], thus extrapolating the femoral and tibial shafts not included in the field-of-view of the MRI datasets. We used the resulting reconstructed surfaces to generate a new SSM for each bone covering the range of bone lengths occurring in the 60 MRI datasets.

2.2 Bone Segmentation

Side Selection and Positioning. We compute the initial transformation parameters (rigid transformation + uniform scaling) using a low resolution surface template of the left distal femur (500 vertices, mean shape of the training data) via the Generalized Hough Transform (GHT) [7]. To detect the correct side of the body, i.e. left or right knee, this method is repeated with a shape template mirrored on the mid-sagittal plane. Accepting the transformation with the maximum vote count after both detection cycles yields the side of the body and the initial transformation \( T_0 \) to position the SSMs of the bones in the data.

Parameter Initialization and Model Adaptation. The model adaptation procedure described in the following is performed independently for the femur and for the tibia. Given the transformation \( T_0 \) of the SSM, the parameters of the image feature based adaptation strategy are initialized. To this end, we rate image features along surface normals with a new cost function that is particularly suitable for MRI: Locations on the surface normals that exhibit intensities within an intensity interval \([t_1, t_2] \) and directional derivatives along the normal that exceed a gradient threshold \( g_{min} \) are equipped with low costs \( g_{min}/g \), where \( g \) denotes the respective directional derivative. All other locations are equipped with high costs (some constant larger than all low costs).

The parameters \( t_1, t_2 \) and \( g_{min} \) are determined automatically via histogram-analyses: The bone intensity threshold \( t_1 \) is computed by analyzing the histogram of intensities inside the initialized SSM (see Fig. 2). It is set such that 5\% of the voxels that contribute to the histogram have higher intensities. The upper threshold \( t_2 \) is set to the maximum intensity that occurs within the initialized SSM. To determine the gradient threshold \( g_{min} \), a histogram of the gradient magnitude of the entire image is computed, and \( g_{min} \) is set to the value above which 15\% of the gradient magnitudes lie.

In addition we check for a special case to handle large, bright artifacts (see Fig. 3 left): A weighted sum of 5 gaussians

\[
\sum_i w_i \cdot g_i \quad \text{with} \quad g_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-0.5 \left( \frac{x - \mu_i}{\sigma_i} \right)^2 \right]
\]
is fitted to the histogram of intensities via the Expectation Maximization (EM) algorithm. We determine the highest gaussian \( g_a \) (i.e. \( a = \arg \max_i \frac{w_i}{\sigma_i \sqrt{2\pi}} \)), and then we look for the highest gaussian \( g_b \) for which holds \( \mu_a + \sigma_a < \mu_b - \sigma_b \). If there is such a \( g_b \) and a value \( x_0 \) where \( g_a \) and \( g_b \) have equal height, i.e. \( w_a \cdot g_a(x_0) = w_b \cdot g_b(x_0) \), and furthermore \( \mu_a < x_0 < \mu_b \) and \( x_0 < t_1 \) holds, then \( t_1 \) is set to \( x_0 \) and \( t_2 \) is set to \( x_0 + 0.25\sigma_b \).

SSM adaptation followed by a graph-cuts based fine-grain adaptation is performed as in [9] employing the cost function described above. The results are surfaces of the femoral and tibial knee bone.

### 2.3 Cartilage Segmentation

The surfaces of femur and tibia are coupled with shared intensity profiles [6] and deformed with multi-object graph cuts [8] employing a new cost function that we specifically designed for cartilage in MRI: From the training data, we learn the minimum, mean and maximum cartilage thickness in normal direction per vertex of the bone surface mesh. Costs are generally high ("infinite") at locations that lie below the learnt minimum and above the learnt maximum cartilage thickness per vertex in surface normal direction. Locations closer to the mean thickness are slightly preferred over those further apart (but still within the minimum-maximum range). Locations within the minimum-maximum range are equipped with low costs \(-g_{\min\text{Cart}}/g\) if the directional derivative in normal direction, \( g \), satisfies \( g < -g_{\min\text{Cart}} \), and the location exhibits an absolute intensity within an intensity interval determined by histogram analysis. The respective histogram is computed from voxels in-between the bone surface and a second surface "bone + maximum cartilage thickness along normals". A weighted sum of five gaussians...
is fitted, and the mean ± standard deviation of the highest gaussian serves as the respective intensity interval. The gradient threshold $g_{\text{minCart}}$ is set to 10 (determined heuristically). Note, this can be seen as an “inverted” version of the cost function described above for bone segmentation.

To obtain the final segmentation, the resulting bone and cartilage surfaces are converted into binary voxel representations with the same extension and voxel size as the original MRI image.

3 Results and Discussion

For evaluation, 40 additional clinical MRI datasets acquired before knee replacement were made available by the workshop organizers. A detailed evaluation is presented in Table 1. For tibial and femoral bones the average (AvgD) and the roots mean square (RMS) surface distances were computed. Cartilage segmentation is quantified by volumetric overlap (VOE) and volumetric difference (VD) measures. For all four structures a score was computed indicating the agreement with human inter-observer variability. Reaching the inter-observer variability results in 75 points, while obtaining an exact match to one distinguished manual segmentation results in 100 points. An error twice as high as the human rater’s gets 50 points, 3x as high gets 25 points and if 4x as high or more receives 0 points (no negative points). All points are averaged for each image, which results in a total score per image. Details on the evaluation procedure will be published in an overview article of the Grand Challenge workshop [5].

The average performance of our auto-segmentation system for knee bones and associated cartilage was $54.4 \pm 8.8$ points. Note that this value corresponds to the status of the software at the time the first results were uploaded to the workshop organizers. New results including bugfixes and other improvements will be available by the time of the workshop and will be published on the workshop website [5]. The automatic segmentation of one dataset took approximately 6 minutes on a standard desktop PC (Core 2 Duo CPU, 3.00GHz, 16GB RAM). It is implemented as an extension to the ZIBAmira software (amira.zib.de).

With no failures the GHT proved to be a reliable method for the initialization of the bone segmentation (side selection and positioning). The bone segmentation achieves scores that indicate an error larger than that obtained by human experts. This may be due to relatively large mismatches of the SSM at the proximal and distal end of the MRI data due to missing or weak image features related to intensity inhomogeneities stemming from the MRI sequence (see Fig. 3). A strong artifact of unknown source (see Fig. 3) lead to the worst segmentation result for the tibia. The scores for cartilage segmentation are based on different error measures (volumetric) and are generally better, presumably due to a higher inter-observer variability.

For the reasons already discussed at the end of the introduction, it is difficult to compare our results to previous approaches. However, from now on the standardized evaluation system of [5] will allow for an unbiased comparison of our method with other systems.
Fig. 3. Problem cases: Artifacts introducing "false" image features (a), intensity inhomogeneities typical for MRI (b), similar intensities for cartilage and neighboring soft tissues (d bottom) and invisible bone-soft-tissue interfaces (d top).

In the future we will investigate how the intensity inhomogeneities in the MRI data can be dealt with. We are already examining the use of a combined active shape model of the femur and tibia to increase the robustness and accuracy of the bone segmentation. Currently, our results indicate that some manual post-processing is still required to draw level with human-rater performance. However, we are confident that with our proposed framework and the suggested improvements we will be able to increase the number of cases where manual postprocessing will become obsolete.

Acknowledgments. H. Seim and H. Lamecker are supported by the EU-FP7 Project MXL (FP7-ICT-2009.5.2-248693). H. Lamecker is also funded by the DFG Research Center MATHEON. D. Kainmueller is funded by the DFG Collaborative Research Center SFB760.

References

and further morphologic parameters. Arthritis and rheumatism 50(3), 811–6 (Mar 2004)


<table>
<thead>
<tr>
<th>Img</th>
<th>Femur bone</th>
<th></th>
<th></th>
<th>Tibia bone</th>
<th></th>
<th></th>
<th>Fem. cartilage</th>
<th></th>
<th></th>
<th>Tibial cartilage</th>
<th>Total</th>
<th>Score</th>
</tr>
</thead>
</table>
Fig. 4. Different views on selected test cases 8 (top), 16 (center), and 31 (bottom). The outline of the reference segmentation is displayed in green, the outline of the automatic method described in this paper in red.