

Automatic Segmentation of the Articular Cartilage in Knee MRI Using a Hierarchical Multi-class Classification Scheme

Jenny Folkesson¹, Erik Dam^{1,2}, Ole Fogh Olsen¹,
Paola Pettersen², and Claus Christiansen²

¹ Image Analysis Group, IT University of Copenhagen, Denmark
jenny@itu.dk

² Center for Clinical and Basic Research, Ballerup, Denmark

Abstract. Osteoarthritis is characterized by the degeneration of the articular cartilage in joints. We have developed a fully automatic method for segmenting the articular cartilage in knee MR scans based on supervised learning. A binary approximate kNN classifier first roughly separates cartilage from background voxels, then a three-class classifier assigns one of three classes to each voxel that is classified as cartilage by the binary classifier. The resulting sensitivity and specificity are 90.0% and 99.8% respectively for the medial cartilage compartments. We show that an accurate automatic cartilage segmentation is achievable using a low-field MR scanner.

1 Introduction

Osteoarthritis (OA) is one of the major health concerns among the elderly today [1]. The main effects of OA is the degradation of the articular cartilage together with remodeling and overgrowth of bone, a process causing loss of mobility of the joints. It typically affects large weight bearing joints as hips and knees. Currently, the treatment of OA is restricted to symptom control, because as yet there are no disease-modifying drugs [2].

MRI allows for quantitative evaluation of the cartilage [3],[4], and cartilage deterioration can be detected using this technique [5]. MRI also has the advantage of being a non-invasive technique.

When assessing the cartilage, the MR scans can be manually segmented slice-by-slice by experts, but for clinical studies manual methods are too time consuming and are also prone to inter- and intra-observer variability. When automating the cartilage segmentation, the main challenges are the thin structure of the cartilage and the low contrast between the cartilage and surrounding soft tissues. The progression of OA is very often slow and it can take many years before the cartilage is reduced from its typical thickness of a few millimeters to possible total loss. It is therefore important to have high accuracy and precision of the cartilage assessment technique in order to detect statistically significant changes. This enables the correlation of the method with the effects of drugs, and the evaluation of their benefit to the joint in reducing the signs of the disease.

Several groups have developed automated methods for cartilage segmentation. 2D methods has limited continuation between slices and since they have to be converted into a 3D segmentation when finding for example thickness maps, it is advantageous to perform segmentation in 3D directly. Among the 3D techniques that have been developed, Grau et al. [6] use a semi-automatic segmentation method that is based on a watershed approach. The method is evaluated on 7 scans from 4 subjects and has an average sensitivity and specificity of 90.0% and 99.9% respectively. Pakin et al. [7] have developed an automatic segmentation method based on region growing followed by two-class clustering. It is evaluated on one scan with resulting sensitivity and specificity of 66.2% and 99.6%. The semi-automatic segmentation method of Warfield et al. [8], [9] iterates between a classification step and a template registration step, and has a lower variability compared to repeated manual segmentations on the scan it was evaluated on. Naish et al. [10] use a data set that consists of a longitudinal study of OA patients and local intensity changes over time is used as a measure of cartilage degradation. However, the cartilage is manually or semi-automatically segmented.

All of the methods mentioned (except for the one of Naish et al. but they have not focused on the segmentation part) have only been evaluated on a handful of scans, and the only fully automatic segmentation produces low sensitivity and specificity values compared to the semi-automatic methods.

In this paper, we present a method for segmenting the tibial and femoral medial cartilage in 3D MR scans of knees. The segmentation is based on an three class approximate kNN classification scheme and is improved by selecting the largest connected component from the result of the classification. The segmentation method works directly in 3D, not in 2D slices, and is fully automatic. This is an improvement of previous work [11] which was a method for locating tibial medial cartilage for the initialization of a shape model, a method based on a two class kNN classifier without any feature selection incorporated.

Our segmentation algorithm aids the automatization of cartilage assessment and is intended for clinical studies on a low-field MR scanner. Though the image quality of the scanner we are using is slightly lower compared to the conventional high-field scanners, we propose to examine if accurate automatic cartilage segmentation is achievable also on a low-field scanner. If such a scanner can be used in clinical studies it would reduce the costs significantly. It has been shown that low-field dedicated extremity MRI can provide similar information on bone erosions and synovitis as expensive high-field MRI units [12] comparing manual segmentations, but there has to our knowledge not been published any work on automatic segmentation of cartilage on low-field MRI. From the automatic segmentation, relevant quantitative measures such as the cartilage volume and thickness can be calculated either globally or locally in a point or a small area. In the latter case comparison between patients or temporal studies of the same patient will require establishing geometric or anatomical correspondence either by expert annotations or by automated modeling of landmarks. An automated approach for this is planned for future work and will not be part of this paper.

2 Methods

2.1 Image Acquisition

An Esaote C-Span low-field 0.18 T scanner dedicated to imaging of extremities acquires Turbo 3D T1 scans (40° flip angle, T_R 50 ms, T_E 16 ms). The scans are made through the sagittal plane with a voxel size in the range $0.7031 \times 0.7031 \times (0.7031/0.7813/0.8594) \text{ mm}^3$. Among the total of 71 scans, 50 have the resolution $0.7031 \times 0.7031 \times 0.7813 \text{ mm}^3$, 3 the resolution $0.7031 \times 0.7031 \times 0.7031 \text{ mm}^3$ and the remaining 18 scans have the resolution $0.7031 \times 0.7031 \times 0.8594 \text{ mm}^3$. The scans all have the size $256 \times 256 \times 104$ voxels, but we only use the central $170 \times 170 \times 104$ voxels because only they contain information.

The scans have been manually segmented on a slice-by-slice basis by a radiologist. A scan slice with the tibial and femoral medial cartilage manually segmented is shown in Figure 1.

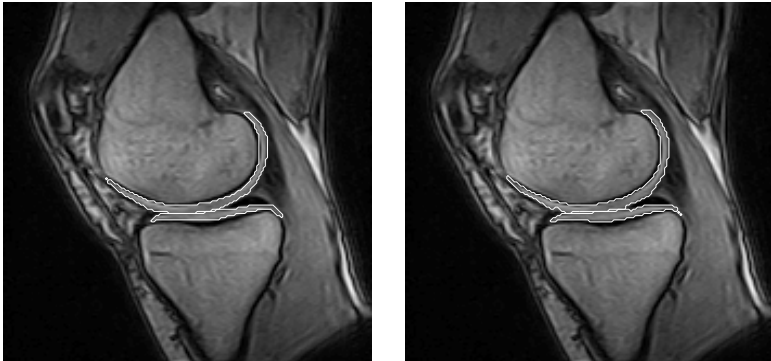


Fig. 1. To the left, a slice from a knee MR scan where the tibial medial and femoral medial cartilage is segmented manually by radiologists. The size of this slice is 170×170 pixels. To the right is the result from our automatic segmentation for the corresponding slice. The sensitivity and specificity for this scan are 92.52% and 99.82% respectively, with a dice similarity coefficient of 0.83.

The 71 scans in the data set are of both left and right knees. In order to treat all scans analogously, the right knees are reflected about the center in the sagittal plane. The test subjects are both males and females aged between 21 and 72 years. They have no or mild OA symptoms, diagnosed by radiologists as being between 0 and 3 on the Kellgren and Lawrence Index [13].

2.2 Cartilage Classification

For the segmentation of cartilage we use an approximate k NN classifier, which is implemented in an Approximate Nearest Neighbor (ANN) framework developed by Mount and colleagues [14]. The ANN classifier is in principle the same as a

k NN classifier, but with the modification that you can allow for a small amount of error in the search for nearest neighbors which may improve the run time significantly. An error bound, ϵ , is introduced, so instead of returning the k nearest neighbors from a data set, the ANN search algorithm returns k points such that the ratio of the distance between the i th reported point ($1 \leq i \leq k$) and the true i th nearest neighbor is at most $1 + \epsilon$. We have found empirically that examining the 100 nearest neighbors yields a good balance between computational complexity and accuracy, and we set $\epsilon = 2$, a value that only marginally lowers the accuracy while reducing computational time significantly.

In this work we examine the medial cartilage since OA is more often observed in this compartment [15] and in particular in the medial tibial part [16], thus these compartments are of major interest when it comes to finding disease markers for OA. In order to separate different types of cartilage from one another we use a three class classifier, where the classes are tibial medial cartilage, femoral medial cartilage and background.

The classification is hierarchical, and the first step is a two class classification where the voxels are roughly separated into cartilage or background. The k NN produces class probabilities for every voxel, and in this step we set the threshold at 0.65 yielding a sensitivity for medial cartilage close to 99%. This also results in a large amount of false positives, but since typically only a few percent of the total volume within the knee belongs to the cartilage, this first step is a way of reducing data significantly. In the second step, the voxels classified as cartilage in the first step are reconsidered. This time we use a three class classifier, where the three classes are tibial and femoral medial cartilage and background, and class membership is decided based on a majority vote. The three class classifier contains more features and the features are optimized to separate the three classes whereas the classifier in the first step has features optimized to separate cartilage from background. A sketch of the hierarchical classification scheme is illustrated in Figure 2.

We have also tested a direct partitioning into the three classes, but the hierarchical approach yields better results and is faster, since the first step has less features and thus lower computational complexity. The classifier in the first step has a set of 28 features compared to the three class classifier in the second step that contains 52 features.

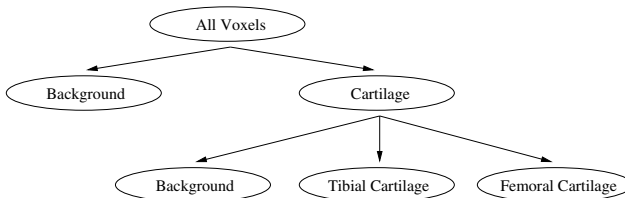


Fig. 2. Tree representation of the hierarchical classification scheme

2.3 Features and Feature Selection

In order to find a feature set that performs well for our classification scheme, we here introduce our set of candidate features and the subsets of the features that were found from our feature selection method [17], which consists of sequential forward selection followed by sequential backward selection.

When a radiologist examines an MR scan for cartilage, she or he takes the location and the intensity in the image into consideration. We therefore consider these as candidate features. Both the raw intensity and the Gaussian smoothed intensities on three different scales ($0.65mm$, $1.1mm$, $2.5mm$) are examined.

One can also consider features that are related to the geometry of the object in question. The 3-jet, which is all first, second and third order derivatives with respect to (x, y, z) forms a basis which can describe all geometric features up to third order [18] and are listed as candidate features. All the derivatives mentioned in this section are Gaussian derivatives and are defined as $I_{i_1, \dots, i_n} = \int \tilde{I}(\bar{x}) D_{i_1, \dots, i_n} g(\bar{x}, \sigma_1) d\bar{x}$, where g is a Gaussian, D a differential operator and σ_1 is the scale. All features are examined on the three scales, selected to cover the range of different cartilage thicknesses, mentioned above.

Cartilage can be described as a thin curved disc in 3D. The Hessian (H), which is the symmetric matrix containing second order derivatives with respect to the coordinates (x, y, z) ,

$$H = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix},$$

is therefore considered. The eigenvectors of the Hessian points in the directions of the principal curvatures and its eigenvalues corresponds to the curvature in those directions. A thin disc such as cartilage will locally yield one large and two small eigenvalues. The eigenvalues as well as the three eigenvectors are candidate features.

A feature that has been shown to be significant in the detection of thin structures such as fingerprints is the structure tensor (ST) [19]. It is a symmetric matrix containing products of the first order derivatives convolved with a Gaussian,

$$ST = G_{\sigma_2} * \begin{pmatrix} I_x I_x & I_x I_y & I_x I_z \\ I_y I_x & I_y I_y & I_y I_z \\ I_z I_x & I_z I_y & I_z I_z \end{pmatrix},$$

where σ is not necessarily the same scale as the one used for obtaining the derivatives. The ST examines the local gradient distribution at each location (x, y, z) . The directions of the eigenvectors depend on the variation in the neighborhood. The eigenvalues and eigenvectors of the ST were considered as potential features with a combination of three scales of σ_1 and three scales of σ_2 .

The third order derivatives with respect to (x, y, z) can be conveniently represented in the third order tensor I_{ijk} . Examining the third order structure in the local gradient direction (I_x, I_y, I_z) can be described using Einstein summation as

$$L_{www} = I_{ijk} I_i I_j I_k / (I_i I_i)^{3/2}.$$

The third order tensor is examined in the gradient direction on three different scales, which were considered as possible features.

2.4 Selected Features

The features used in the two class classifier are the position in the image, the Gaussian smoothed intensities on three different scales ($0.65mm$, $1.1mm$, $2.5mm$) and the raw intensities, the first order Gaussian derivatives on scales $0.65mm$ and $2.5mm$, the eigenvalues and the eigenvector corresponding to the largest eigenvalue of the structure tensor with $\sigma_1 = 0.65mm$ and $\sigma_2 = 2.5mm$, and the eigenvalues of the Hessian on scales $1.1mm$ and $2.5mm$.

The features in the three class classifier consist of combinations of first, second and third order Gaussian derivatives on the three different scales mentioned, the Gaussian smoothed intensities on three different scales ($0.65mm$, $1.1mm$, $2.5mm$) and the raw intensities, the position, the eigenvector corresponding to the largest eigenvalue of the ST with $\sigma_1 = 0.65mm$ and $\sigma_2 = 1.1mm$, the eigenvalues of the ST with $\sigma_1 = 1.1mm$ and $\sigma_2 = 2.5mm$, the eigenvalues of the Hessian on scales $1.1mm$ and $2.5mm$.

The features selected as most significant are the Hessian and the structure tensor along with the intensity and the position in the image. The features were normalized between zero and one. Normalization for unit variance was also examined, but the normalization of values between zero and one produces slightly better results.

3 Results

From our data set of 71 scans we use 25 for training and 46 for the evaluation of our algorithm. The results of our automatic segmentation is compared to the manual segmentation made by radiologists, resulting in an average sensitivity and specificity of 90.0% ($\pm 2.6\%$ st.d.) and 99.8% ($\pm 0.06\%$ st.d.) respectively



Fig. 3. On the left is the manually segmented medial cartilage from a knee MR scan. To the right is the corresponding automatic segmentation. For this scan, the sensitivity and specificity are 94.82% and 99.79% respectively, with a dice of 0.81.

for the test set for the medial cartilage compartments. A comparison between a golden standard segmentation and an automatically segmented knee MR scan can be seen in Figure 3. A slice by slice comparison is displayed in Figure 1. The dice similarity coefficient (DSC) measures spatial volume overlap between two segmentations, A and B , and is defined as $DSC(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}$. The Dice similarity coefficient between our automatic segmentation and the golden standard segmentation is for the test set on average 0.80 (± 0.03 st.d.).

4 Discussion

The average sensitivity and specificity of our method compared to the results of methods with similar evaluation though on different data is presented in Table 1. Comparing our method with the fully automatic segmentation algorithm (Pakin et al. [7]), we get a distinctly higher sensitivity and a slightly better specificity. Though slightly worse, the sensitivity and specificity of our method are comparable to those of Grau et al. [6]. They have a higher volume overlap ($DSC=0.90$) however their method is semi-automatic. We have evaluated our segmentation algorithm on more scans than the other two methods. Some of the semi-automated

Table 1

	Our method	Method of Pakin [7]	Method of Grau [6]
Sensitivity	90.01%	66.22%	90.03%
Specificity	99.80%	99.56%	99.87%
Data set	25+46	1	7
Interaction time	0	0	5-10 min

segmentation techniques described in section 1 have been evaluated in terms of inter- and intra-observer variability of the method compared to manual segmentation. As for our method, future work will involve inter-scan variability, and we will also examine intra-user variability for the manual segmentations.

Our segmentation algorithm performs well compared to two leading cartilage segmentation schemes, which leads us to the conclusion that accurate, fully automatic cartilage segmentation is achievable in low-field MR scanners.

References

1. J.Wyngaarden, H.Smith, L., Bennett, J.: Cecil Textbook of Medicine. 19 edn. Volume 2. W. B. Saunders (1992)
2. Creamer, P., Hochberg, M.C.: Osteoarthritis. Lancet **350** (1997) 503–509
3. Muensterer, O., Eckstein, F., Hahn, D., Putz, R.: Computer-aided three dimensional assessment of knee-joint cartilage with magnetic resonance imaging. Clinical Biomechanics **11** (1996) 260–266
4. Graichen, H., Eisenhart-Rothe, R.V., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. Arthritis and Rheumatism **50** (2004) 811–816

5. Loeuille, D., Olivier, P., Mainard, D., Gillet, P., Netter, P., Blum, A.: Magnetic resonance imaging of normal and osteoarthritic cartilage. *Arthritis and Rheumatism* **41** (1998) 963–975
6. Grau, V., Mewes, A., Alcaiz, M., Kikinis, R., Warfield, S.: Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging* **23** (2004)
7. Pakin, S.K., Tamez-Pena, J.G., Totterman, S., J.Parker, K.: Segmentation, surface extraction and thickness computation of articular cartilage. Volume 4684., SPIE (2002) 155–166
8. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis* (2000) 43–55
9. Warfield, S.K., Winalski, C., Jolesz, F.A., Kikinis, R.: Automatic segmentation of mri of the knee. SPL Technical Report 91, ISMRM Sixth Scientific Meeting and Exhibition, Sydney, Australia (1998)
10. Naish, J.H., Vincent, G., Bowes, M., Kothari, M., White, D., Waterton, J.C., J.Taylor, C.: A method to monitor local changes in mr signal intensity in articular cartilage: A potential marker for cartilage degeneration in osteoarthritis. Volume 3217., MICCAI (2004) 959–966
11. Folkesson, J., Dam, E., Pettersen, P., Olsen, O.F., Nielsen, M., Christiansen, C.: Locating articular cartilage in mr images. Volume 5747., SPIE (2005) 1484–1490
12. Ejbjerg, B., Narvestad, E., adn H.S. Thomsen, S.J., Ostergaard, M.: Optimised, low cost, low field dedicated extremity mri is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: a comparison with conventional high-field mri and radiography. *Annals of the Rheumatic Diseases* **13** (2005)
13. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases* **16** (1957)
14. Arya, S., Mount, D., Netanyahu, N., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. Number 5, ACM-SIAM. *Discrete Algorithms* (1994) 573–582
15. Dunn, T., Lu, Y., Jin, H., Ries, M., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. *Radiology* **232** (2004) 592–598
16. Kamibayashi, L., Wyss, U., Cooke, T., Zee, B.: Changes in mean trabecular orientation in the medial condyle of the proximal tibia in osteoarthritis. *Calcif Tissue Int.* **57** (1995) 69–73
17. Folkesson, J., Dam, E., Olsen, O.F., Pettersen, P., Christiansen, C.: A supervised learning approach to segmenting articular tibial cartilage in knee mri. Technical report, IT University of Copenhagen, Rued Langgaards Vej 7, 2300 Copenhagen S, Denmark (2005)
18. Florack, L.: The Syntactical Structure of Scalar Images. PhD thesis, University of Utrecht (1993)
19. Weickert, J.: *Anisotropic Diffusion in Image Processing*. B. G. Teubner (1998)