Knee Cartilage Segmentation Algorithms:
a Critical Literature Review

Universiteit Utrecht

Biomedical Image Sciences
Supervisors: Dr. K.L. Vincken, Dr. C. Bos

Maxim D. Ryzhkov

Contents

Abstract

The objective of this study was to give a review of the knee cartilage segmentation methods presented in the literature, from a critical perspective. Segmentation allows quantitative and qualitative analysis of the cartilage from the morphological point of view. This is important in clinical applications, because it allows early diagnosis of pathological changes in the cartilage, such as osteoarthritis. Cartilage extraction can be done in various approaches, which can be grouped into edge tracking, intensity-based (thresholding, texture analysis, watershed transform), supervised learning (kNN classifiers, support vector machines, ensemble learning), and energy minimization (active contours, statistical shape models and surfaces, graph-cuts). Deciding which method is more applicable for particular situation greatly depends on what trade-off the user is willing to accept with respect to the robustness to the quality of the input data, computational time, and accuracy. It is often required to specify the region-of-interest or to indicate the most typical candidates for the cartilage and non-cartilage tissues. This step may require user interaction or can be fully automatic. Most of the automatic segmentation methods contain an additional pre-processing step, where the bones or bone-cartilage interfaces are extracted, since bone is an important landmark when searching for the cartilage.

Aim and method

The aim of this paper is to provide general information about the problems associated with knee cartilage segmentation, and to give a critical review of the relevant methods presented in the literature in the past two decades. For this purpose, 15 key publications in this field have been reviewed (highlighted in bold in the references), and their segmentation methods described. The applicability of the selected methods for particular situations are also discussed. The reviewed articles were selected based on the term "*segmentation of the knee cartilage*", and cover a broad variety of segmentation methods. The majority of the sources were retrieved from Google Scholar via Utrecht University UBU proxy.

Abbreviations

ASM	Active shape models
BCI	Bone-cartilage interface
BMI	Body mass index
CT	Computed tomography
DESS	Double-echo Steady-State
dGEMRIC	Delayed gadolinium-enhanced MRI of cartilage
DSC	Dice-Sørensen coefficient
GAG	Glycosaminoglycan
GUI	Graphics user interface
JSW	Joint space width
kNN	K-nearest neighbors
LCS	Local coordinate system
MRI	Magnetic resonance imaging
OA	Osteoarthritis
ROI	Region of interest
SSM	Statistical shape models
SVM	Support vector machine
VOI	Volume of interest

1. Introduction

Structure of human knee

The knee is the largest joint in the human body. Its purpose is to provide a pivotal point between the thigh and the lower leg during movement. It consists of bones (femur, tibia and patella), which interact with one another via articular surfaces of the hyaline cartilage (tibio-femoral and patello-femoral) [1]. Femur interacts with tibia in two major areas: medial (closer to the midline of the body) and lateral (away from the midline). The inner structures are held in place by connective tissue and are covered with adipose tissue. Figure 1.1 shows a magnetic resonance image of the knee in sagittal orientation with an anatomical atlas for reference. Cartilage is visible as a bright thin layer covering the bones. The fat in this image has been suppressed, in order to make the cartilage more prominent in contrast with the surrounding tissues. Bright mass adjacent to the femoral and tibial cartilages are the muscles.

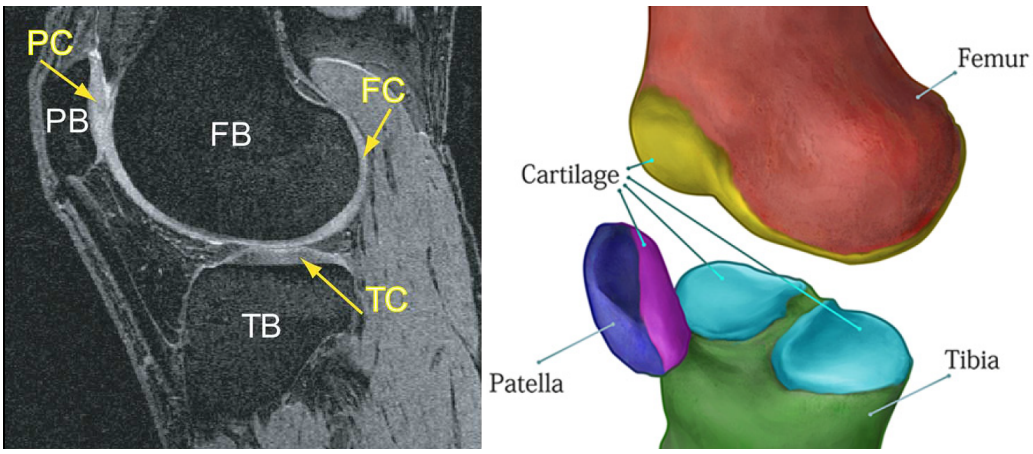


Figure 1.1: Yin et al [2] and Lee et al [3]. Illustration of the knee showing femoral, tibial and patellar bones with the associated cartilage surfaces. FB = femoral bone, TB = tibial bone, PB = patellar bone, FC = femoral cartilage, TC = tibial cartilage, PC = patellar cartilage.

Cartilage composition

Cartilage consists of extracellular matrix and chondrocytes. The matrix has two components: type II collaged fibers, which give cartilage its elastic properties, and proteoglycans containing high negatively charged polysaccharide chains called glycosaminoglycans, or shortly GAG [4]. It is possible to visualize the concentration of these molecules using a method called dGEMRIC (delayed gadolinium-enhanced MRI of cartilage). After the injection of contrast agent Gd-DTPA^{2-}, a T1 relaxation time map is constructed, where the signal is proportional to the amount of GAG in the cartilage [5].

Osteoarthritis

Hyaline cartilage has a limited capacity for repair. Injuries and extensive wear leads to degeneration of the tissue, which triggers a number of processes with various degree of severity. These include formation of osteophytes and subchondral cysts, fibrillation of the cartilage matrix, remodeling and sclerosis of the subchondral bone [6]. When the cartilage has eroded, the bone tissue is exposed and is highly prone to physical wear, causing pain and inflammation of the joint. This condition is called osteoarthritis (OA) and is the leading cause of musculoskeletal disability in developed countries. It is believed that as many as 30% of the people older than 65 will develop OA at some point in time [7]. It is also estimated that by the year 2020, 18% of the population in the US will be affected by OA [8].

Screening methods

Unfortunately, the standard treatment of OA nowadays does not fully cure the disease. It is therefore of most importance to identify the degeneration of cartilage at early stage before it becomes irreversible. There are number of ways to determine the degree of cartilage degeneration in patients.

Arthroscopy

One of the most common techniques used in the past was arthroscopy, which allowed direct observation of the surface morphology, and surgical intervention such as placement of cartilage allografts [9]. The biggest disadvantage of this method is that it is highly invasive and may itself contribute to the progression of the disease [10]. Also, it does not give an estimation of the cartilage thickness, which is an important marker in evaluating the stage of OA.

Radiography

Over the past decades, X-ray of the joint space width (JSW) has been the conventional method for OA screening. It offers a substantial advantage over arthroscopy because it is non-invasive and can be safely repeated if necessary. Among the disadvantages of X-ray imaging is the lack of precision for short-term studies, due to the fact that changes in the cartilage can only be detected on X-ray images after 2-3 years [10].

MRI

Magnetic resonance imaging is becoming the new standard technique for screening of the cartilage, because it does not use ionizing radiation, is non-invasive, repeatable, and gives a good image quality with high contrast and resolution. MRI delivers images in digital format, which can be stored and easily retrieved, and offers a variety of parameters for optimal image acquisition. The disadvantages include high cost of the apparatus (especially for high field strength magnets), long scanning times, and proneness to imaging artifacts. Additionally, it lacks signal intensity normalization, compared with CT where the signal is presented in Hounsfield units. This often has an impact on the robustness of the image processing algorithms [11].

Kellgren-Lawrence score

There are several ways to determine the stage of OA progression qualitatively. Kellgren-Lawrence (KL) score is one of the most common evaluation techniques used in clinical practice. In KL score, the severity of the damage is stated by a discrete score ranging from asymptomatic (grade 0) to severe OA (grade 4).

The KL score reflects three key features of OA [12]:

- Joint space narrowing
  Occurs when the cartilage thickness reduces to extreme levels. Normally, the cartilage is not visible on radiographic images, and the bones appear to be distinctly separated. When the cartilage is degenerating, the bones move closer to each other, which can be detected and measured.
- Osteophyte formation
  These appear as bright bony projections of high signal intensity (see Figure 1.2), which are believed to be one of the cause of chronic pain in osteoarthritis and limited joint movement.
- Bone tissue sclerosis
  Can be identified as brightening of the bone edges, which is the result of an increase in bone density.

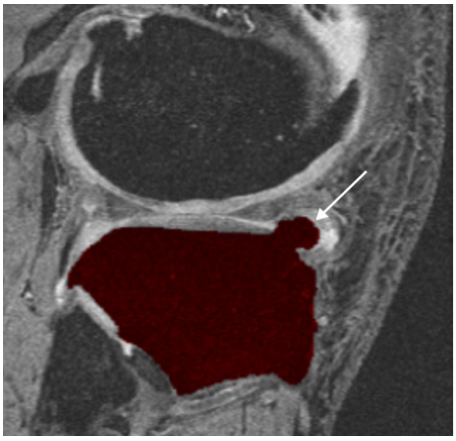


Figure 1.2: Lee et al., 2011. White arrow indicates an osteophyte.

MRI sequences

Among the MRI sequences used nowadays for imaging of cartilage, several approaches can be used. T1-weighted sequences are typically used to show the contrast between cartilage and the surrounding tissues. T2-weighted sequences are sometimes used to quantify the amount of synovial fluid in the joint, or to give an estimation of the amount of glycosaminoglycans in the extracellular matrix. Spoiled gradient echo (SPGR) sequence, also known as Fast Low Angle Shot (FLASH), has long been the preferred method to detect osteoarthritis [9, 13]. It provides a good signal-to-noise and contrast-to-noise ratios, and spatial resolution, especially with fat suppression. However, it does not provide enough contrast between cartilage and other non-bone structures such as meniscus and synovial fluid, making it hard to detect the local cartilage borders [14]. Recently, Dual-Echo in Steady-State (DESS) sequence has become popular specifically for orthopedic imaging. In DESS, T1/T2*-w signal (e.g. FISP) is combined with T2-w signal (e.g. mirrored FISP) in a single TR period to form a composite image. It produces high-resolution images with bright fluid and provides contrast between the fluid and the cartilage [15].

Osteoarthritis Initiative

Several studies mentioned in this paper used the data from Osteoarthritis Initiative (OAI), which is a multicenter longitudinal study of biomarkers for the development of osteoarthritis. It includes almost 4800 participants, which undergo annual screening of both knees using x-ray and 3T MRI (DESS sequence) for a total of 9 (previously 4) years [16, 17]. It is of high interest, not only because it provides the testing data for evaluating the algorithm, but also because the data has been acquired with the same imaging protocol. This makes it easier to compare methods that have used OAI data because it eliminates image quality discrepancy as a factor.

2. Cartilage segmentation

Overview

Qualitative assessment of the knee is only able to identify the disease progression in late stages. Recent studies suggest, however, that it is possible to predict the degeneration of the cartilage by monitoring its condition over a continuous period of time. The targets for monitoring include cartilage thickness, volume and surface [18]. In order to measure these accurately, we first need to segment the cartilage, i.e. extract its shape from the input image [19].

There may be various reasons for visualizing the knee cartilage, apart from osteoarthritis prevention. It can be, for instance, a cross study to find variation in cartilage structure across given population groups, such as gender, ethnicity, BMI, age etc. [20]. One can also think of a study to find out if there are any differences in cartilage size when using different MR sequences, as seen in van Leersum et al., 1995 and Eckstein et al., 2002 [21, 22]. This helps determining the optimal acquisition parameters for accurate and precise estimation of the cartilage volume.

Manual segmentation

Prior studies have shown that developing a fully automatic algorithm for cartilage segmentation is not a trivial task [17, 16]. Until now, manual delineation of the cartilage boundaries by a trained expert still remains the most common technique. It is a highly laborious and time-consuming process, which can easily take up to several hours per knee, depending on the image size, quality, and stage of the disease in particular patient. Additionally, this method suffers from low intra- and inter-observer reproducibility due to human error and requires special training. Nevertheless, none of the existing (semi)automatic algorithms have been able to deliver visually superior results, thus manual segmentation still remains the golden standard for validation of segmentation algorithms.

Challenges - automatic segmentation

In practice, there must always be a balance between the algorithm performance, accuracy, robustness, and the level of user interaction. An ideal segmentation algorithm would maximize the precision and accuracy for a large variety of images, while keeping user interaction at minimum [23]. The ultimate goal is to develop a fully automatic "one-click-do-it-all" segmentation algorithm, which demonstrates high accuracy, specificity, reproducibility and robustness, with a reasonably fast computational time. Unfortunately, the task is complicated by a number of factors.

First of all, the cartilage layer is very thin, especially in OA patients. In some areas the thickness reaches sub-millimeter levels. High resolution images are therefore required to achieve good-quality segmentation, although they are generally associated with longer acquisition times. Here lies the risk: the longer the patient has to stay in the gantry, the higher the chance of involuntary movements, leading to image artifacts [1].

Also, whereas it is usually easy to distinguish cartilage from the adjacent bones, the contrast between cartilage and the surrounding soft tissues is generally much worse. Even after fat suppression, the boundary between cartilage and ligaments or adipose tissue remains very vague. Also, within

the cartilage itself, the tissue is not homogeneous [24]. The algorithm therefore has to be trained to recognize these patterns and treat them accordingly. Image noise is another factor that has a direct effect on the outcome of the segmentation. As a rule of thumb, the output of the algorithm is only as good as the quality of the input data.

Cartilage segmentation is a complex process, and is typically approached in sequential manner. The researchers would break the task down into several steps, in order to have control over the intermediate results and make adjustments if necessary. For example, if a segmentation component fails because it cannot handle a particular type of input data, one of the solutions would be to replace that component with a more robust alternative. In addition to that, a component may or may not require user interaction. However, this does not only mean that the user has to be trained to use the software, but also that the algorithm has to be under constant supervision. While this may not be critical for most situations, there might come a time when a large collection of datasets has to be processed. The user might not always be available to examine every single case one-by-one. A fully automatic algorithm, on the other hand, can be left to run autonomously. One of the downsides of using automated methods is that they often suffer from lack of robustness due to the fact that they have been tuned for specific types of input data [23]. On the other hand, it is not uncommon to have both automatic and semi-automatic components in the segmentation chain depending on how critical the component is, or whether the process is virtually automatable or not.

Validation metrics

The segmentation results are usually validated on a ground truth, i.e. manual delineation by an expert or the output of an existing segmentation algorithm with high reported accuracy. The success rate of an algorithm is defined as a ratio of the outcomes which met the acceptance criteria to the total number of cases.

Pearson's correlation coefficient describes how two variables are linearly correlated. It is calculated by dividing the covariance of the variables by the product of their standard deviations:

$$r = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right) \cdot \left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}} \tag{2.1}$$

Sensitivity (SE), specificity (SP), positive predictive value (PPV) and Dice-Sørensen coefficient (DSC) [25] are defined as follows:

$$\begin{aligned} SE &= \frac{TN}{TN + FP} \\ SP &= \frac{TP}{TP + FN} \\ PPV &= \frac{TP}{TP + FP} \\ DSC &= \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \end{aligned} \tag{2.2}$$

where TP - true positive, TN - true negative, FP - false positive, FN - false negative.

Intra- and inter-observer variability is often measured by computing the coefficient of variation (C_v), which is defined as the ratio of standard deviation of the results to their mean value:

$$C_v = \frac{\sigma}{\mu} \tag{2.3}$$

Simultaneous truth and performance level estimation (STAPLE) is a common technique for evaluating the accuracy of segmentations when no ground truth is available or when multiple expert segmentations are present. This expectation-maximization approach estimates the "true" object mask and assigns a weight for each given segmentation based on how well it corresponds with the other data [26].

3. Literature review

This section gives on overview of the literature published in the past two decades, describing various methods of knee cartilage segmentation. The segmentation algorithms presented in the articles are sorted by the level of automation (manual, semi-automatic, automatic) and grouped by specific image processing techniques. A short summary can be found in Table 1 of the Appendices section.

3.1 Semi-automatic segmentation

Semi-automatic methods are aimed at reducing the amount of user interaction during segmentation process, while providing a level of supervision in order to guide the algorithm in the right direction. A semi-automatic method usually requires initialization, meaning that the user is asked to specify a set of landmarks on the image as a starting point for the segmentation. We can divide the literature into the following groups:

1. Intensity-based: thresholding [13], watersheds [27]
2. Edge detection [9, 1, 16, 28]
3. Energy-minimization: Live Wire, graph-cuts [29], active contours [30]

3.1.1 Intensity-based segmentation

In the work by Dupuy et al [13], delineation of the region-of-interest was done manually with a mouse on a workstation, separately on each slice. The software then determined the mean intensity in the specified region, and the user could specify a threshold within two standard deviations from the mean intensity. All voxels with intensities higher than the threshold would then be classified as cartilage. The algorithm was validated on physical specimen, which included two cadaveric knees with no signs of osteoarthritis and four knees from OA patients scheduled for full knee arthroplasty (replacement of the knee with prosthesis).

It was not clear whether the task was carried out by a trained radiologist or an average user. It the future, the experiment should be repeated with at least one extra observer who is a trained specialist, to determine the inter-observer variability. Furthermore, the thresholding step relies on the mean intensity, which is calculated from the volume-of-interest. It is therefore sensitive to the local intensity distribution. For example, if the area happens to include more bone (low intensity voxels) the mean intensity would be biased towards it. And lastly, the slice thickness in some images was as high as 3.5mm, which can potentially result in a large accuracy error.

A modified version of watershed transform is presented in the article by Grau et al [27]. The image is first smoothed using anisotropic filtering followed by manual annotation of 3 types of seeds: bone, cartilage and other tissues (ligaments, muscle etc). Using given seeds as local minima, the algorithm attempts to find intersections of the so-called catchment basins, regions that are upstream of the local minimum. The problem of oversegmentation, which is typical for watershed method, is overcome by applying geodesic reconstruction [31] and optimized using ordered queues. Instead of using gradient image as input for watershed transform, the authors introduce a function that

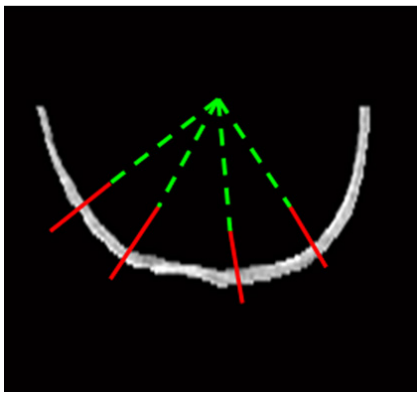


Figure 3.1: Akhtar et al., 2007 [9]. Cartilage border is found by plotting N rays from a center point and finding the gradient.

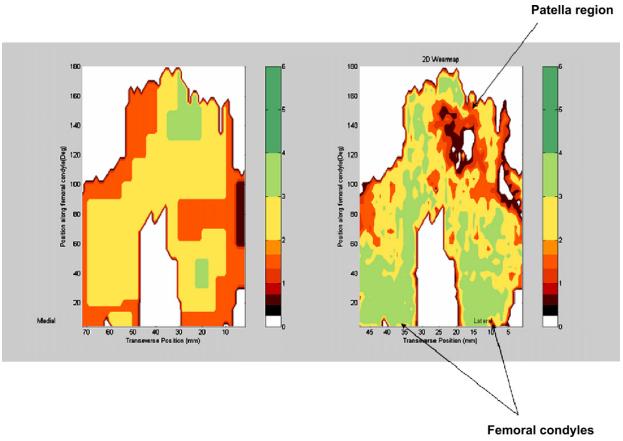


Figure 3.2: Akhtar et al., 2007 [9]. 2D wear map of cartilage from OA patient (left) and normal subject (right). Color map corresponds to cartilage thickness from lowest (black) to highest (green).

calculates the probability of finding an edge between labeled and unlabeled voxels. The authors claim that the intensities between different tissues have a normal distribution and intrinsically exhibit spatial homogeneity, which is in fact specific for the images used in validation. It would be helpful if the study was repeated using a different imaging sequence, in order to test the robustness.

The algorithm was validated on a set of manual segmentations defined by two experts independently. The accuracy of the semi-automatic segmentation was found to be as good as the manual segmentation. In the initialization step, 50 seeds had to be defined for each tissue class, which took on average 5-10 minutes per dataset. While this is significantly lower than segmenting the images manually (2 hours per dataset), the authors should consider testing the algorithm with a lower number of seeds, or selecting the seeds automatically. It should also be noted that geodesic reconstruction will smooth out all catchment basins which have not been marked, thus it is important to consider the location when placing the seeds.

3.1.2 Edge detection

Akhtar et al [9] describes a method of visualizing the femoral cartilage thickness on 3D mesh. Although cartilage segmentation was not the first priority in this paper, their method still relies on accurate delineation of the cartilage borders. The algorithm was implemented in MATLAB (Mathworks Inc). First, the user defines a region of interest around the femoral cartilage, for each slice. This region of interest can be copied to the next slices if the cartilage is still covered by it. From a specified center point, a number of rays are projected in the direction of the ROI, with a step of 4°. Each line is searched until the cartilage borders are identified, using edge detection (Figure 3.1). The edge points are then combined into a 3D mesh using a cubic B-spline interpolation. Lastly, the thickness is measured by taking the mesh width along the normals to the bone-cartilage boundary (Figure 3.2). The accuracy of the algorithm was determined by comparing the resulting segmentation with manually delineated contours.

Unfortunately, the algorithm has only been tested on femoral cartilage of healthy individuals, and no information was given about its performance on OA cartilage. The algorithm also relies on careful selection of the region of interest, which has to be repeated for each slice individually and therefore takes a considerable amount of time. The method in general lacks robustness, since edge detection is likely to fail when the borders are not defined clearly, whether due to poor image quality or low contrast between the tissues for given imaging sequence. In the future, it would be interesting to see the inter-observer variability, since the study was only validated by one expert.

In the method developed by Carballido-Gamio et al [1], the user defines a series of points on a

slice inside the cartilage, which are combined into a Bezier spline. Then the algorithm finds bone-cartilage interface and articulating interface using the following procedure. First, anisotropic diffusion filtering is applied with different parameters for each interface. Then, a number of rays are projected 15 voxels away from the spline, and the maximum first-order derivative value is taken as the cartilage border. At some parts of the cartilage, such as tibio-femoral and patello-femoral, the edges are not clearly defined, thus an extra post-processing step is applied. The image is processed again using anisotropic filtering with very weak smoothing factor, and the edge is extracted once more, using the points along the surface where cartilage thickness is greater than mean + $\frac{1}{4}$ of the standard deviation. If no local maximum is found, the old edge position is taken as the final. The user has a chance to manually fix the edges before accepting the slice. The contiguous slices are segmented using information from the previous slice. Bones are segmented in a similar manner to cartilage, using the bone-cartilage interface found in the previous step. The resulting splines are then combined and the mask is interpolated linearly using distance fields to produce isotropic voxels with size equal to the in-plane resolution.

There were several hard-coded parameters in the algorithm, for example searching distance along the rays, maximum cartilage intensity threshold, etc. Despite good performance for this particular study, using preset parameters does not guarantee optimal results for all types of input data. In the future, the initialization of these variables should be done automatically using statistical data. Furthermore, the proposed method of separating tibial and femoral cartilages needs to be improved, since their intensity distributions are very similar and the boundary is extremely faint. The Pearson's correlation coefficient between the water displacement of porcine femoral cartilage and its segmented volume was r=0.83, leaving room for further improvement.

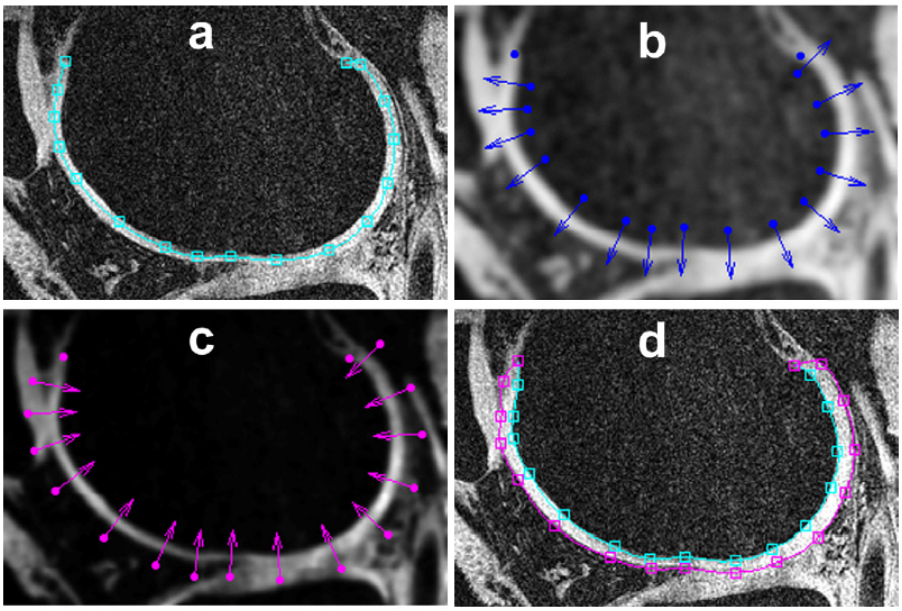


Figure 3.3: Carballido-Gamio et al., 2008 [1]. Femoral cartilage segmentation. (a) Bezier spline generated from control points. (b) Line profiles for finding the bone–cartilage interface. (c) Line profiles for finding the articular surface. (d) Bone–cartilage interface (cyan) and articular surface (magenta) as the result of the segmentation.

Duryea et al [16] developed a sophisticated graphics user interface (GUI) tool implemented in C programming language, which employs a number of edge-detection filters separately tuned for each type of cartilage. First, the user specifies a seed point on the bone-cartilage interface, and the algorithm then attempts to segment the cartilage using edge detection, separately for bone-cartilage, and cartilage-soft tissue edges. The user can then intervene to correct the mistakes, and the structure

is reshaped using active contours approach. The segmented shape can be transferred to the next slice and refined, once again using active contours. The process is repeated until 4 subsections of the knee cartilage (tibial cartilage was considered as 2 separate parts) have been segmented. The algorithm required 4 out of 20 ground truth segmentations to be used for training.

The authors provide a very broad set of tools for manipulating with the segmentation process. However, this process requires training, as the authors mention in the article. The amount of time it takes to segment the whole knee is comparable with manual segmentation, thus it is not clear what the actual benefit of this method is, compared with manual segmentation. On the other hand, the authors are confident that the segmentation time will reduce to 10-20 minutes per knee in the future, when the work flow has been improved.

A slightly modified version of this method was used in a longitudinal study by Iranpour-Boroujeni et al [28]. Here, the 24-month follow-up dataset is registered to the baseline, and cartilages are segmented and compared. When segmenting the cartilage, the user is instructed to specify two 10mm circular regions at the tips of the cartilage, where any segmentation errors can be ignored, because the algorithm consistently demonstrated inaccuracies. The BCI and the articulating edges are segmented as previously described in [16]. By overlaying the segmentations and subtracting them from each other, it is possible to visualize the gain and loss maps and give an estimate of the volume change. The applicability of this method for longitudinal studies has also been evaluated earlier by Brem at al [32].

3.1.3 Energy-minimization

Gougoutas et al [23] used a technique called Live Wire, where the user moves the mouse around the edge of the cartilage, while the algorithm computes the best path along it, in real time. The concept behind this method is to minimize the cost function composed of a number of elements, which are initialized by a short training session prior to the segmentation. During training the algorithm automatically calculates the boundary constraints from a reference image.

The user starts by selecting a seed point on the object edge, and drags the mouse in the direction of the object contour. The algorithm computes a joint cost, which includes several features such as intensity on both sides of the object border, gradient in different orientations, and distance from the previous boundary [33]. For complex shapes the process can be broken down into several steps where seed points are placed at the beginning of each subsequent section. The algorithm has been validated using bovine specimens, imaged with a 4T scanner. The volume of the cartilage was measured by water displacement. In order to test the precision, the experiment was repeated on scans of the right knee of healthy volunteers. Each subject was scanned 5 times with the knee fixed in a custom-built jig. The cartilages were segmented independently by 3 different observers.

The accuracy results can hardly be representative of in-vivo settings, since the imaging was done on phantoms. Despite the fact that the authors tried to imitate a realistic clinical image by introducing noise and blurring, in reality there are many more contributing factors, which cannot be simulated in vitro. The scanner strength was rather high for a real clinical situation, where 1.5T and 3T field strengths are still used. Also, there was only one observer who did the measurements on phantom images. Collecting more data from different experts would make the results statistically more significant.

Unfortunately, the algorithm was only tested on patellar cartilage, which has a relatively regular shape compared to femoral and tibial, as the authors point out. It is not yet clear whether the Live Wire method would perform equally well on images with cartilage degeneration, especially when lower field strength is used. Nevertheless, this method provides a good alternative to manual segmentation with the average operational time of 30 seconds per slice (15 minutes for the whole knee), and 5 minutes for training.

A graph-cuts approach has been proposed by Bae et al [29]. Initialization of the algorithm is done by placing several seed lines inside the cartilage and inside the surrounding tissues, close to each other.

The seeds are propagated to the neighboring slices, thinning with each slice by 1 voxel until the lines disappear and the user needs to place the seeds again. The graph-cuts algorithm described in Boykov et al [34] uses the seed information to separate object voxels from the background, where seeds act as hard constraints. The object is segmented by cutting the graph edges in such way that the cost function (Eq. 3.1) is minimized. The graph nodes are represented by voxels, and the edge costs are composed of regional properties and boundary properties. Regional property carries information such as probability of the voxel belonging to object or background judging by the intensity distributions found inside the seeds. Boundary cost incorporates information about the inter-voxel distance and difference in their intensities (Eq. 3.2).

$$E(A) = \sum_{p \in P} R_p\left(A_p\right) + \sum_{\{p,q\} \in N} B_{\{p,q\}} \tag{3.1}$$

$$B_{\{p,q\}} = \exp\left(-\frac{\left(I_p - I_q\right)^2}{2\sigma^2}\right) \cdot \frac{1}{dist\left(p,q\right)} \tag{3.2}$$

The method was tested by two independent experts using Osteoarthritis Initiative data with different OA severity, where the KL score ranged from 0 to 4. The intra-observer variability was measured by repeating the experiment after 4 weeks period. The inter-observer error in cartilage volume was higher for severe OA cases than for healthy subjects (1.86% vs 0.83%), However, the algorithm has not been validated due to the absence of ground truth data. Therefore, it is hard to judge whether or not it produces accurate results. The average processing time was between 30 and 50 minutes per scan, which is only a slight improvement from manual segmentation considering that the user also has to place the seeds several times. Automating the seed placement can be a topic for future research. This would eliminate the factor of subjectivity when initializing the algorithm, which is likely to improve the results, as it was mentioned in the discussion section.

Kauffmann et al [30] used local coordinate systems (LCS) and rigid object registration for evaluating the cartilage in longitudinal studies. The user places seed lines close to the cartilage, and the surface is extracted using active contours (snakes). This surface is used to initialize the segmentation on the contiguous slices.

To compare the cartilages from different timelines, the segmentations need to be registered to each other, because the position and orientation of the knee is likely to change from one scan to another. The authors claim that in longitudinal studies the articulating side of the cartilage undergoes more changes than the bone-cartilage side, thus it is logical to fix the local coordinate system to the BCI side. The femoral cartilage is transformed to a cylindrical LCS and the tibial cartilage is mapped onto a planar LCS, using least-squares fitting. Each point on the geometrical model (cylinder or plane) is defined as an offset to the nearest cartilage contour point. The thickness map is generated by computing the distance between offset point of BCI to the offset point of the articular edge. Knowing the bone-surface distance maps for master cartilage and the moving cartilage, it is possible to find a translation that would minimize the difference between the corresponding distances.

Despite the fact that the traditional validation approach is to use manually-defined ground truth, the authors decided to validate the algorithm using a synthetic MR image (Figure 3.4). The signal intensities in the synthetic image are derived from measuring the intensity distribution in the bone, synovium and cartilage across 100 images. The images have been degraded by noise, generated using Rician probability distribution function. It is not clear whether this method gives a true estimation of the algorithm performance, since it idealizes the situation by providing a clear contrast between cartilage and background. In real settings, the bones are surrounded by other non-cartilage tissues such as fat, muscle and synovial fluid, some of which have a very poor contrast with the cartilage depending on the image acquisition parameters.

The registration algorithm demonstrated good results in the longitudinal study context. The positional error was found to be significantly less than the minimum voxel size (0.02 mm vs 0.3mm

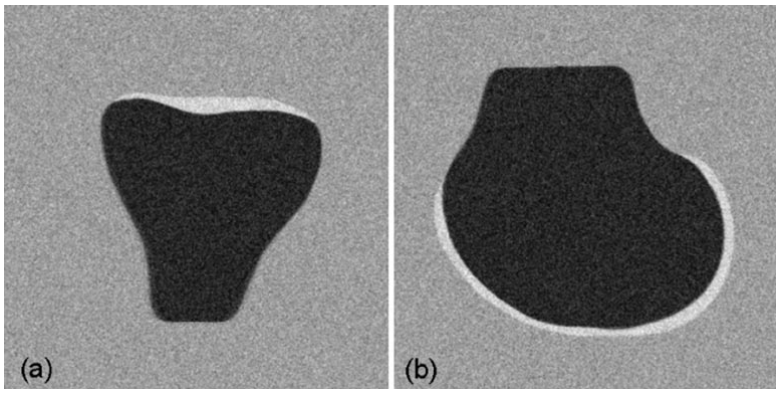


Figure 3.4: Kauffmann et al., 2003 [30]. Synthetic MR image of the knee based on a pool of 100 images. a) Tibial bone and cartilage. b) Femoral bone and cartilage

respectively), and the coefficient of variation for volume was as low as 0.1%, in synthetic MR images.

3.2 Fully automatic segmentation

The Osteoarthritis Initiative project offered a large amount of image data for the researchers. However, segmenting each image semi-automatically still requires substantial supervision, thus automatic segmentation has become a hot topic in the recent literature since it allows segmenting the cartilage without any user interaction. The following literature can divided into four groups:

- Texture analysis [35]
- Supervised learning: kNN classifier [19], SVM [36]
- Statistical models: Active shape models [37], Karhunen-Loeve expansion with adaptive template matching [38]
- Graph-cuts [2, 3]

3.2.1 Texture analysis

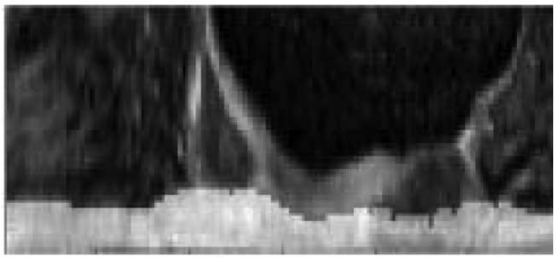
In Dodin et al [35], the cartilage is segmented by a classifier, which uses the texture of the tissue along normals to the BCI. The BCI is detected by an automated method described in an in-proceedings article by Lorenz and von Berg [39]. The bone segmentation algorithm consists of a ray generator, ray sampler and ray analyzer. Ray generator constructs a set of directions from a specified point, which can be arranged in a number of ways, e.g. circular, spherical, parallel etc. Ray sampler reads voxel intensities along the rays, applying smoothing to the image if necessary. Ray analyzer then determines whether the object is hit by looking if the search criteria are met. The results of the first ray analysis can be used to initialize the subsequent searches.

When a rough BCI contour has been found, a matrix of intensities along the normals to the BCI surface is constructed. For each sampled point, tissue in the vicinity is analyzed using histogram information. The authors distinguish two features of the tissue texture: brightness and homogeneity (Fig. 3.5-a). The dark and the bright tissues are separated by minimizing the variance between two corresponding modes on the histogram (Otsu thresholding). The bright regions are then further thresholded based on tissue homogeneity. When two cartilages overlap, their thicknesses are refined by dividing the combined thickness into two equal halves. Synovial fluid is excluded from the mask by employing the fact that water is more homogeneous than the cartilage (on DESS images), thus they can be separated from each other with Otsu thresholding. The outer cartilage border is found

TABLE I
CLASSES OF TISSUES

	Very homogeneous	Homogeneous	Inhomogeneous
Very Bright	Water	Water Cartilage	
Bright	Water Cartilage	Cartilage	Soft tissues
Dark		Dark cartilage Cartilage lesion	Cartilage lesion

a)



b)

Figure 3.5: Dodin et al., 2010 [35]. A) Classification of tissue based on brightness and homogeneity of the voxels in the neighborhood. B) The resulting cartilage mask superimposed into the normal profile map.

by looking at the edges along the normals, using the cartilage mask found in the previous step. The method is validated on semi-automatic segmentation results.

This method appears to be sensitive to the contrast of tissues, since Otsu thresholding most optimally when there is a clear separation between the two modes (dark and bright tissues). While DESS sequence provides a good contrast between cartilage and background, the fluid also appears bright on the images. The authors claim that fluid is on average more homogeneous than the cartilage, however this is subject to noise in the images. For the future, the algorithm has to be tested on a wider range of images.

The algorithm was validated by comparing the resulting segmentation with the output of a semi-automatic method presented by Kauffmann et al [30]. This is a major draw-back because manual segmentation is still considered to be the state-of-art for ground truth generation. Besides, the algorithm in Kauffmann et al was not validated on manual expert segmentations but on synthetic MR images instead.

3.2.2 Supervised learning

The article by Folkesson et al [19] describes a kNN-based approach to cartilage segmentation. The algorithm consists of two different binary classifiers, separately for tibial and femoral cartilages. The classifiers have been tuned to include features specific for the type of cartilage that is being segmented. For tibial cartilage the following features were included: position in the image, intensities after smoothing with Gaussian filter (3 different kernel sizes), Hessian, first order derivatives, eigenvalues of the Hessian, and eigenvalues of structure tensor generated from first-order derivatives convolved with a Gaussian. The femoral cartilage classifier had the following features: position, eigenvector of the largest eigenvalue of structure tensor, first-order derivatives on two different scales, intensities after smoothing, eigenvalues of the Hessian and eigenvalues of the tensor. The authors claimed that k=100 is the optimal number for the nearest neighbors parameter, based on empirical findings.

Due to the fact that kNN classifiers typically suffer from lack of performance [19], the authors have decided to optimize the algorithm using the following method. First, a small number of voxels located roughly in the center of the datasets are classified. If the voxel is labeled as cartilage, the algorithm then proceeds to examine the neighboring voxels. The process is repeated until there are no remaining neighbor voxels labeled as cartilage. The resulting connected components are then merged to produce a single component for each cartilage. Due to the fact that the position of the knee in the scans is not consistent, a post-processing step was introduced. After segmenting the cartilage, the image is shifted so that the center of mass of the segmentation matches with the one from the training data. The cartilage mask is dilated by 3 voxels, followed by kNN classification within the resulting VOI. The sensitivity of the algorithm (DSC=0.80) was comparable to the intra-observer error (DSC=0.87).

For the future, it would be useful to collect more ground truth data, at least from one more expert to test the inter-observer variability.

One of the major draw-backs of this method is that it has only been tested on 0.18T scans. It is not clear whether it would still be viable for more commonly used 1.5T images, where the resolution is higher. As far as the computational time is concerned, it should be noted that increasing the size of the image in its turn increases the computational time. This particularly applies to the kNN-type classifiers, which are known to be relatively slow, as mentioned by the authors. The algorithm was only tested on femoral and tibial cartilages. Also, it was mentioned that the position of the voxel was the most significant feature for both tibial and femoral classifiers, yet the authors claim that their method is robust to the position of the knee in the scans. Whether this statement is valid or not, the authors should consider registering the images to the template to eliminate the positioning errors.

Zhang et al [36] combined support vector machine (SVM) with discriminative random field (DRF) approach to segment femoral, tibial and patellar cartilage on multi-contrast MR images. Each dataset consisted of a combination of four images, scanned with the following sequences: FS SPGR, IDEAL-GRE water, FIESTA, IDEAL-GRE fat. The reason for using multi-contrast images is to combine information, which can only be obtained from specific sequences, such as bone-cartilage contrast, tissue-fluid contrast, and so on. SVM and DRF are combined to complement each other on their weaknesses. While SVM succeeds at classifying individual voxels, it does not take into account the surrounding voxels. DRF takes into account spatial dependencies between neighboring voxels, but is not suitable for binary classification.

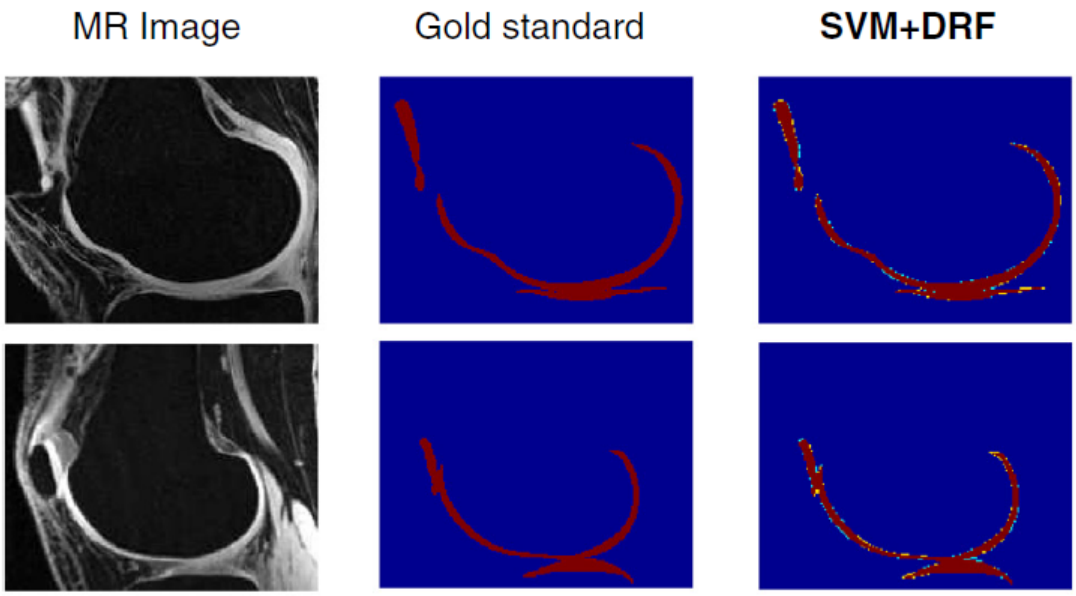


Figure 3.6: Zhang et al., 2013 [36] Comparison of ground truth with SVM-DRF automatic segmentation. Red zones are TP, cyan zones are FP, yellow zones are FN, and blue zones are TN.

The local dependency of the voxels was defined by a lattice graph with 10-neighbor connectivity. Femoral, tibial and patellar cartilages are segmented independently, e.g. femoral vs rest, then tibial vs rest, etc. The algorithm was optimized by loopy belief propagation, which takes into account the information from previously observed nodes. Leave-one-out method was used, where 1 sample was picked as the testing data and the rest were used as the training data. The following vectors have been picked for the SVM, based on the highest observed Dice coefficient: 4D normalized intensity values of multi-contrast MR, 24-D local image structure based features (included gradients of Gaussian-smoothed image, eigenvalues of Hessian), and 3D geometrical features. The latter has been obtained by an extra bone segmentation step, based on thresholding followed by connected component labeling. Thus, each voxel was represented by a 31-D vector.

Validation was carried out using manual segmentations from a trained medical expert. The resulting segmentation was also compared with the output of other methods presented in the literature,

including kNN classification by Folkesson et al [19], deformable models by Fripp et al [37], LOGISMOS method by Yin et al [2] and graph-cuts by Lee et al [3]. The authors claimed that their method produced the most accurate results among the listed algorithms, with DSC of 0.86, 0.88 and 0.84 for femoral, tibial and patellar cartilages, respectively. At the same time, the authors mention that their method exploits the multi-contrast information of the data, which is not done in the competing literature. It is not yet clear whether the method presented in this paper is superior when used in clinical settings, where the availability of multi-contrast data is limited. Also, no testing has been done to check the accuracy of the results on pathological knees. Lastly, the results indicate that the introduction of geometrical features to the SVN had a significant impact on the accuracy, therefore more attention has to be given to the bone segmentation step in the future.

3.2.3 Statistical models

Angelini et al [38] approached the task in two steps. First, training data was used to generate a cartilage VOI atlas with manual delineation of the cartilage borders. In the second step, the atlas was mapped on top of the patient image in such way that it minimizes the intensity variation in the corresponding voxels. The authors used a discrete version of the principal component analysis called Karhunen-Loeve expansion.

First, the images were normalized by bringing the datasets to zero mean energy. Here, the mean image was subtracted from each particular subject, and divided by the normalized mean energy. Then, for each image the cartilage interfaces were delineated manually by an expert. That included bone-cartilage interface and cartilage-soft tissue interface. A region-of-interest was constructed by sampling all voxels within 2.5 mm from the nearest cartilage border point. The regions of interest were then registered to each other using adaptive template matching, similar to affine registration. The degrees of freedom included translation, rotation and scaling. From here, each transformed ROI can be viewed as a multidimensional vector, where each axis corresponds to a voxel position in that ROI, in raster-scan order (top to bottom, left to right), and the elements are the gray scale intensities. Assuming that the anatomical variation between patients is minimal, the voxel intensities should have a normal distribution. From there, a covariance matrix can be computed (see equation 1, where: A - covariance matrix, e - eigen vector, λ - eigen value) followed by eigenvalue/vector decomposition.

$$Ae_k = \lambda_k e_k, \quad k = 1, n, \tag{1}$$

The largest eigenvalue and its corresponding eigenvector would then represent such a projection of data points onto the axes that would explain their spread the most. This approach is called Karhunen-Loeve expansion. To form a prototypical ROI, the authors took 3 largest eigenvalues and computed a weighted sum of their corresponding eigenvectors to form a single image.

In order to align the reconstructed ROI with the test images, the prototype was convolved over all possible starting positions, and the optimal location was found using two similarity metrics: cross-correlation (dot product) and the SSD (sum of squared differences). The best position of the prototype would be at the point where dot product is the highest, and SSD is the lowest. These points do not necessarily coincide, thus a weight has been assigned to each metric, which depended on the local gradient along the performance grid. The process stops when both metrics reach a common extremum. Once the location of the prototype has been optimized, the cartilage borders could then be found by looking at the gradient along normals to the prototypical ROI edges. A 2D spline is then fitted along the steepest gradient.

The result is validated by computing the Euclidean distance between the ground truth and the resulting curve. Although the errors (0.62 mm, 0.3 mm and 0.42 mm, for femur, tibia and cartilage, respectively) are smaller than the voxel size (0.47 mm), we must take into account the fact that the images have been interpolated, which contributes to the total error. Overall, this method comes with a number of limitations, most of which have already been identified by the authors.

The disadvantage of the Karhunen-Loeve expansion is that it requires intensities to have a normal distribution, which in its turn needs a large pool of data. The authors correctly point out that more

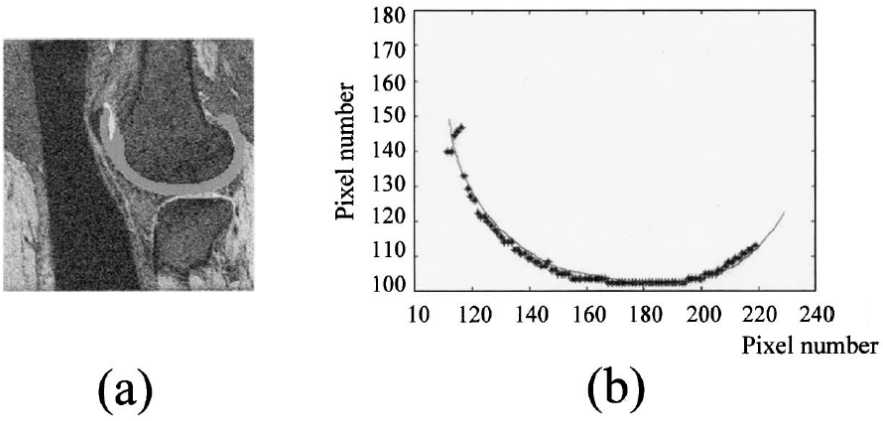


Figure 3.7: Angelini and Ciaccio, 2003 [38]. A) location of the prototype mask after adaptive translation; B) coincidence of the ground truth and the segmentation.

datasets need to be used to account for all possible inter-patient anatomy differences, to achieve a normal distribution. Furthermore, if we were to increase the resolution of the images, the number of voxels in the ROI would increase by the order of n^3 where n is the number of dimensions. This will have a drastic effect on the computational time. Unfortunately, the authors did not give any concrete numbers as to the performance of the algorithm time-wise. The computational time is determined by the size of the input images (number of starting locations) and how fast the ROI converges to the global extremum.

Another pitfall lies in the last step, which is the gradient computation. First of all, it relies on accurate prepositioning of the prototypical ROI, meaning that it has to end up on the correct position along the performance grid. Authors claim that this is not always the case, i.e. the ROI sometimes falls in the local minima and as the result only covers the cartilage partially. The gradient along the normals then does not guarantee accurate cartilage interface extraction. This happened in 8 cases out of 150. Also, this method in general would not work on outliers - patients with abnormally large or small knees, because the ROI would cover the wrong areas. Lastly, the gradient search is not likely to work on patients with severe OA where the cartilage has degenerated beyond detectable levels. It is not clear from the text how the algorithm handles such exceptions. As a suggestion for future work, the algorithm should be tested on a larger population, and using data from different scanning protocols.

In Fripp et al [37], a hierarchic approach was used, where the bone-cartilage interface is first found, since it gives a good initialization for cartilage detection. First, statistical shape models (SSM) of the bones (individual and collective) were constructed. The bones were parameterized by mapping the surface of a sphere that encompasses the object onto the object's surface, breaking down the meshes when details are needed. The probability of each mesh being the bone-cartilage interface was incorporated into the SSM. The bones on testing data were segmented using 3D active shape models. After initialization, the model was iteratively deformed along normals to the surface, and the shape was restricted by statistical constraints. Initialization was done by affine registration of an atlas with corresponding bone mesh models to the test image. The boundary was determined by walking along the normal to the pre-positioned surface and finding the strongest gradient along the profile. Internal tissue constraints were placed, ensuring that the tissue along the search path and around the strongest gradient resembles the bone.

The BCI was segmented using active shape models approach on the whole-knee SSM with 3-level multi-resolution Gaussian pyramid, where the image was smoothed with a median filter and re-sampled using linear interpolation. Initial estimate of the bone tissue was done by Otsu thresholding, which as authors point out is not optimal but good enough for initialization. Tissue properties were estimated by Gaussian distribution of intensities within the sample. The high-BCI probability points are then relaxed onto the boundary and corrected by looking at the tissue along the normals to the pre-BCI,

using pre-defined constraints.

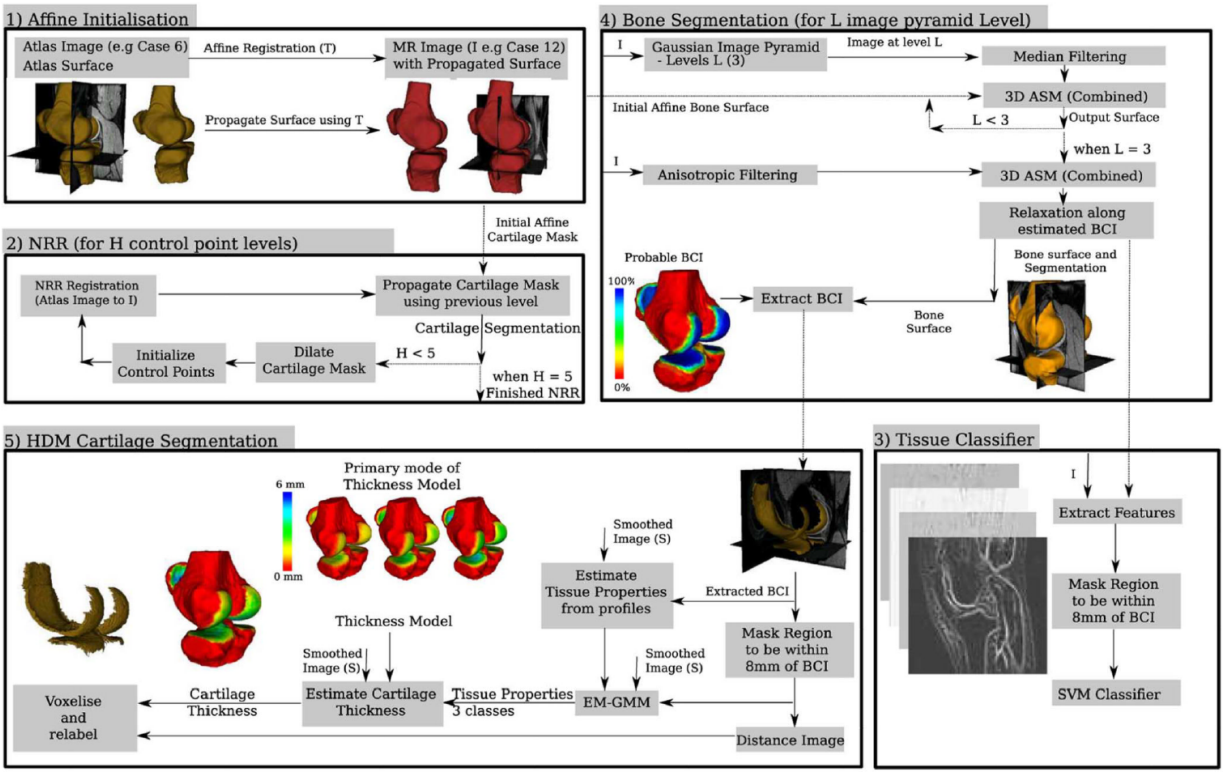


Figure 3.8: Fripp et al., 2007, 2010 [14, 37]. Automatic bone-cartilage interface extraction followed by cartilage segmentation.

The cartilage voxels are classified using expectation maximization Gaussian mixture model. First, a binary mask is created by covering the voxels within 8 mm from the BCI. The algorithm then looks at the highest gradients along the normal to the mask, taking into account statistical data from the learning set. The shape constraints are also enforced, using BCI information. The process is repeated several times, reducing the search range with each iteration until convergence. Any overlapping voxels are separated using distance fields.

Validation of the algorithm was done on expert manual segmentations (DSC=0.803 and 0.77 for femoral and tibial cartilages, respectively) and compared with the results of segmentation algorithms presented by Folkesson et al [19], Grau et al [27] and Glocker et al [40]. The DSC was worse than the watershed approach by Grau et al (0.895), but was comparable with the other two methods. As the authors mention in the discussion, the algorithm is sensitive to the bone segmentation, the failure of which accounted for all of the failed cartilage segmentation cases (3.6%). Therefore more attention has to be given to the BCI extraction step in future research. The algorithm is also sensitive to noise and partial volume effect, as the authors state in the text. Since the algorithm was only tested on healthy cartilage, this weakness is only likely to become more prominent in the future work when segmentation of degenerated cartilage will be tested. The computation time for the whole process was approximately 20-25 minutes, which is a decent performance taking into account the high accuracy of the results and no user interaction.

3.2.4 Graph-cuts

Yin et al [2] presented a method called LOGISMOS (layered optimal graph image segmentation of multiple objects and surfaces). This method requires a coarse bone pre-segmentation. First, VOIs for each bone have been found using Adaboost method, trained using learning set A, where the images undergo noise removal by median filter, and 3D Haar features are extracted for each VOI. After that, bone meshes are constructed using learning set B. The bones are extracted from the VOI using random forest approach with a set of features such as: intensity, gradients, second-order derivatives after Gaussian smoothing. This rough bone surface is adjusted using single-surface graph search, with optimization using electric field lines trick. The costs are derived from random forest likelihood probabilities. The cartilage is segmented using multi-layer graph cut with knowledge of surface constraints (location of the cartilage with respect to adjacent bones) and multiobject constraints (how the cartilages are related to each other). The costs are determined using random forest classifier trained from learning set B.

The authors state that their algorithm requires that the pre-segmentation is sufficiently close to the target surface, however no concrete information was provided to explain what is meant by the term "sufficiently close". The algorithm is also sensitive to the expectation statistics, for example one case failed due to the abnormally thick tibial cartilage. A larger training database is therefore needed to fully cover the cartilage properties variance. The algorithm took on average 20 minutes per scan, with the bone segmentation step taking roughly 6 minutes. For the future, the authors should consider optimizing this step, for instance by using affine registration of an anatomical atlas (see Fripp et al [37]).

Lee et al [3] developed a fully automatic method consisting of bone segmentation followed by BCI extraction and cartilage segmentation. The bone is segmented using constrained branch-and-mincut approach described in an earlier paper [41]. It is an optimized version of graph-cut method, which uses the fact that bones are found in relatively consistent position on MR scans from patient to patient. The BCI is found using a kNN classifier, based on position of the voxel and local appearances. The position of the BCI and non-BCI voxels is determined using distance transforms from an average of affinely registered training images. The appearance probability is found by computing Euclidean distance between the corresponding image patches in the test image and the nearest neighbor in the training set. The cartilage is segmented using graph-cuts approach, by minimizing the energy function consisting of unary and pairwise probabilities, determined during learning stage. Unary probability depends on local histogram of the cartilage, and pairwise probability is the likelihood of finding a boundary at specific location. This produces several local segmentations, which are combined into a single mask with a global graph cut.

The accuracy of the presented algorithm was evaluated using manual segmentations provided by two trained experts. However, due to time constraints, the cartilage was only delineated on every third slice. The results were compared with the output of a semi-automatic algorithm developed by Shim et al [17], and showed slight under-performance. On the other hand, the method does not require any user interaction, and it is able to produce compartmentalized results by assigning a label of the parent bone to the corresponding cartilage. The algorithm performed better when a larger training set was used. This confirms the assumption that training has to be done on a wider range of datasets in order to catch all inter-patient variation. Lastly, the authors proposed classifying the training data into categories such as age, gender, pathology etc, for further separation of the samples.

4. Discussion

The literature reviewed in this study can be classified into the following image processing approaches: edge tracking, intensity-based analysis, supervised learning, statistical models, and graph-cuts. Every method has its pros and cons, depending on the type and quality of the input images and the tolerance for segmentation errors.

Edge tracking is the method of choice when there is a clear contrast between cartilage and the surrounding tissues. It can be safely combined with low-pass filters to remove the noise, because it does not rely directly on the intensity values but rather on the transition between them. It is a very efficient method when it comes to performance (compared with kNN classifiers, for instance). Typically, edge detection is carried out on 2D slices and combined into a mesh using interpolation. Although the method can be virtually extended to 3D, it is not very common due to a high trade-off between the geometric accuracy and the performance [42]. The current state-of-the-art imaging techniques seldom provide optimal contrast between the tissues, thus the applicability of edge detection remains limited to high field strength images. It also fails in regions where two cartilages are adjacent to each other, such as tibio-femoral contact, because of high tissue homogeneity in that area.

Intensity-based algorithms are based on local intensities around the set of given landmarks. The simplest example was presented by Dupuy et al [13] where the region-of-interest is thresholded by picking a value within sampled intensity range.

Watersheds approach is a special case of region growing, which is typically done on gradient images. The image is gradually flooded and the catchment basins between the peaks of the gradient image are separated. The watershed approach has several disadvantages, which include oversegmentation, sensitivity to noise and poor detection of thin structures [27]. Oversegmentation requires the connected components to be merged together later on, which adds an extra step to the segmentation work flow. Sensitivity to noise can be overcome by running the image through a low-pass filter. However, it increases the risk of missing weak edges between the object and the background, resulting in an overflow, which cannot be corrected. The inability to detect thin structures is very important for clinical cases where the cartilage has degenerated to sub-voxel thickness. As the result, the cartilage will appear smaller than it actually is, with eroded borders and holes. Furthermore, watershed transform requires seed placement, which is usually done manually. In future research, it would be interesting to see the results of automatic seed initialization, as it would save a substantial amount of time for the user.

In texture analysis, the voxels are classified as object or background based on intensity distribution of the surrounding voxels. This method requires that the tissues have minimum overlap when it comes to brightness and homogeneity [35]. Since this depends very much on the MRI sequence and the image quality, the texture analysis methods are likely to suffer from lack of robustness.

Several supervised learning methods have been mentioned in the reviewed literature, including k-nearest neighbors classifiers, support vector machines, and ensemble learning (i.e. Adaboost, random forest). One of the main disadvantages of binary classifiers in cartilage segmentation context is that they do not take into account spatial relations within the segmentation. When the algorithm fails, it produces "holes" or floating patches, which have to be corrected by an extra post-processing filter.

The kNN classifiers label each voxel in the image as object or background based on the coordinates of the voxel in multi-dimensional feature space. The voxels in the training set are represented by N-dimensional vectors, where each dimension is a feature that describes the voxel in the context of its

location and neighborhood. There are infinite number of features one can use to train the classifiers, however the most commonly used are signal intensity, position in 3D space, value of the N-order derivatives, eigenvalues of the Hessian, eigenvalues and eigenvectors of the structure tensor, Haar features, etc [6, 36]. Provided the training data is sufficiently large, the object and non-object classes can be represented by two clouds of points with a mean and a variance. Classifying the voxel in the input images is done by simply computing the distance between its multi-dimensional vector and the k-nearest training data vectors. Whether the voxel belongs to the object or the background is decided by weighted majority voting. The kNN classifiers have several disadvantages. First of all, they are computationally heavy, because for every classified voxel the distance from each point in the training set has to be calculated and sorted. This increases with every new image that is being added to the pool, and for every new feature. Folkesson et al states that non-optimized version of their kNN classifier takes approximately 60 minutes, which is almost as slow as manual segmentation [6]. The second disadvantage of this method is that the classification is determined purely by how many points of one class are found in the vicinity, making it sensitive to noise. Lastly, it is not possible to give more trust to the features that are more relevant for classification, because all features have the same weight. In cartilage segmentation context, it would be extremely helpful to inform the algorithm that proximity to the bone is a much more important factor than, for example, absolute voxel intensity. Unfortunately, kNN classifier is not the best solution for this task.

Support vector machines are similar to kNN classifiers in the sense that they are based on computing the distance between points in multi-dimensional space. In SVM, the separation between the object and the background is defined by a hyperplane (Gaussian RBF kernel in Zhang et al [36]), which maximizes the margin between the two classes. Points that lie within the margin borders are called support vectors, and are used to model the separating surface. Although the distributions are likely to overlap with each other, we can avoid unnecessary over-fitting by defining a constant C, which controls how much penalty is given to the training errors. SVM generally produces high accuracy results even in the presence of learning errors due to noise. The major disadvantages are the high computational load during training, and sensitivity to outliers [43].

Ensemble methods, unlike kNN and SVM, demonstrate good results even when the number of samples is less than the number of features [44], which is a common problem in cartilage segmentation. They can also be easily extended to carry out multiobject classification (e.g. bone, cartilage and other tissues). Adaboost combines a set of weak classifiers (e.g. based on Haar features, as seen in Yin et al [2]) into a single strong classifier. It starts by calculating the error of the first weak classifier and assigning more weight to the points that have been misclassified. In the next iteration, the image is analyzed by another weak classifier using the weights calculated in the previous step. The final classifier consists of a weighted sum of all intermediate classifiers [45]. This method produces a good estimate of the true segmentation, although training it can take a long time. In random forests approach, a large number of decision trees are generated, where a random subset of the features is selected for each level in the tree. The final classifier is based on the weighted voting between the classification trees.

The cartilage can also be segmented using statistical models, as seen in Angelini et al [38] and Fripp et al [37]. The idea behind it is that the object can represented by a set of landmarks, sampled from a collection of training data. These landmarks can be specific points on the anatomy, or simply points on the mesh grid. Provided that the number of samples is large enough, each set of landmarks will have a normal distribution, and their eigen components can be extracted using principal component analysis. The eigenvector corresponding to the largest eigenvalue then represents the most significant mode of variation, meaning that this landmark varies the most in the given population sample. In Angelini et al, each landmark is a voxel inside the cartilage. A template of the cartilage was constructed by taking a weighted average of the 3 largest principal components (each one is a separate cartilage mask). This template was then fitted to the test image by measuring how well it correlates with the voxels underneath it.

Active shape models, also called "smart snakes", use stochastic information from the training data to find the best fit for the shape applied to the test image. It starts by defining a rough location,

orientation and scaling of the object model. In cartilage segmentation, we can pre-position the model based on bone segmentation, because it provides a good fit for the bone-cartilage interface. The cartilage borders are then refined by an iterative process where the algorithm searches for the strongest gradient along the normals to each landmark, and checking whether new shape fits the statistical model of the cartilage. In active surface models, the shape is iteratively deformed by minimizing the sum of internal and external energies. The external energy is driven by the gradient across the normals, similar to active shape models. The internal energy dictates the elasticity of the model, defined by its surface area and curvature along the principal directions [46].

For the purpose of cartilage segmentation, statistical models approach is a powerful tool because it uses prior information to give a realistic estimate of the cartilage surface while taking into account the patient-specific anatomy. The internal shape constraints make sure that the model is not affected by random noise in the image. The external energy can incorporate various information, such as the probability of finding the cartilage along the normals based on training data [37]. The disadvantage of these methods is that the energy minimization can be stuck at a local minimum, therefore proper initialization of the model is required. Also, it is important to have enough training data for optimal representation of the variance in the population.

Graph-cuts is another popular method for cartilage segmentation. The principle behind it is simple: let the image be represented by a graph $N = (V, E)$ where each voxel V is a node, and each node is connected to the neighboring nodes by undirected edges E with non-negative capacity. Here, the capacity can be defined as the probability that two nodes belong to the same class. For instance, we know that the transition from cartilage to bone is often characterized by a strong gradient. Therefore, if there is a sharp intensity transition between two voxels, it probably means that they belong to opposite classes and will be asigned a low edge capacity. Suppose s and t are two nodes on the graph representing the object and the background. We can then find such a flow from s to t that utilizes the edge capacities in the most efficient way. A graph cut is defined as a collection of saturated edges that we can reach from s when the flow is maximized. The resulting subsets of the graph then represent the optimal segmentation of object from the background. In image segmentation context, the edge capacity is commonly referred to as the energy cost. Graph-cut minimizes the total energy that is required to separate object from the background.

The graph-cuts algorithm provides a good solution for cartilage segmentation, due to several reasons. First of all, the energy cost can be selected in a flexible manner, for optimal representation of the relationship between the object and the background. Critical features can be assigned a higher weight or can be tuned automatically during training. The computational load is in the low-order polynomial time [34]. In addition to that, this method proves to be robust to noise, because high-cost (low capacity) edges found in the middle of the graph do not block the flow. One of the disadvantages of graph-cuts is that the source and the sink need to be initialized, whether manually by placing seeds [29], or automatically by pre-segmentation [2, 3].

Drawing a parallel between the automatic segmentation methods described in this review, it is evident that most of the algorithms (except for kNN classifier by Folkesson et al) include an additional bone/BCI extraction step prior to cartilage segmentation. Bone segmentation serves multiple purposes:

1. Bone is an important landmark, which defines the region-of-interest when searching for the cartilage.

2. Bones are directly adjacent to the cartilage, therefore extracting the bone-cartilage interface automatically provides a starting point for the outer cartilage edge finding.

3. Sometimes it is necessary to identify which bone is adjacent to the segmented cartilage, for example if we wanted to separate different cartilages from each other.

In addition to that, most of the semi-automatic methods require that the user places a number of seeds close to the cartilage. Essentially, what we are trying to do is to help the algorithm identify the starting area for cartilage detection, since it is a rather subtle structure. This leads to a logical assumption that most of the semi-automatic methods can be extended to fully automatic, provided that there is an accurate and reliable bone segmentation method available. However, bone segmentation is a topic for a separate study.

The applicability of the algorithms to longitudinal studies, where cartilage volume and thickness are monitored throughout a period of time, deserves special attention. This is important for clinicians because it allows early diagnosis of osteoarthritis, so that preventive measures can be taken as early as possible. Volume monitoring is a relatively simple task compared to thickness monitoring, because it is not affected by position and orientation of the knees in the scans. One can simply compute the difference between cartilage mask volumes in the initial session and in the follow-up session. Thickness is typically measured by computing the distance along the normals from the bone-cartilage interface to the articulating surface. The difficult part is to align the cartilages in such way that the corresponding normals in the baseline session and the follow-up session match. Even though the articulating surface might change over time, we expect the bone-cartilage interface to be consistent between the scans. Kauffmann et al [30] suggested registering the knees by first unwrapping the bone-cartilage interfaces into local coordinate systems (cylindrical for femoral and planar for tibial). The cartilages can then be aligned by sliding one surface along the other until the differences in their thickness vector lengths are minimized.

Conclusion

A number of cartilage segmentation approaches have been developed in the past years, with variable degree of automation. In semi-automatic methods, the user helps the algorithm to identify the starting location for cartilage segmentation. In automatic methods, the initialization is typically done by first finding the adjacent bones and/or bone-cartilage interface. Cartilage extraction can be done in several ways. Edge detection methods perform well when there is a clear boundary between the cartilage and the surrounding tissues. Intensity-based methods employ the fact that cartilage and non-cartilage tissues have distinctly separable gray-scale distributions. Supervised learning methods use large pools of ground truth data to specify which local features around the cartilage voxels are important for optimal classification. In statistical models, the shape information is used to fit a stochastic model of the cartilage onto the input image. Graph-cuts minimize the energy required to reach the object from the specified location on the outside. Although manual segmentation still remains the golden standard, the results presented in the recent literature suggest that semi- and fully automatic segmentation algorithms have a potential to become the standard technique for the quantification of cartilage surface and volume, for single cases and for the follow-up studies.

Appendices

Author	**User input**	**Bone/BCI extraction**	**Cartilage segmentation method**	**Tested on**	**Training needed**
Dupuy et al (1996)	ROI definition	x	Thresholding	Femoral, Tibial, Patellar	x
Grau (2004)	Seed placement	x	Watersheds transform	Femoral, Tibial, Patellar	x
Akhtar (2007)	ROI definition, seed placement	x	Edge detection	Femoral	x
Carballido-Gamio (2007)	Seed placement	x	Edge detection	Femoral	x
Duryea (2007)	Seed placement	x	Edge detection, active contours	Femoral, Tibial, Patellar	x
Gougoutas (2004)	Seed placement, delineation	x	Live Wire (graph search)	Patellar	Live Wire training
Bae (2009)	Seed placement	x	Graph-cuts	Femoral, Tibial, Patellar	x
Kauffmann (2003)	Seed placement	x	2D active contours	Femoral, Tibial	x
Dodin (2010)	x	Edge detection	Texture analysis	Femoral, Tibial, Patellar	x
Folkesson (2007)	x	x	kNN classifier	Femoral, Tibial	kNN training
Zhang (2013)	x	Thresholding, connected components labeling	Support vector machine	Femoral, Tibial, Patellar	SVM training

Continued on the next page

Continued from previous page

Author	**User input**	**Bone/BCI extraction**	**Cartilage segmentation method**	**Tested on**	**Training**
Angelini (2003)	x	Edge detection	Adaptive template matching	Femoral, Tibial, Patellar	Atlas generation
Fripp (2010)	x	Active shape models	Statistical shape models, expectation-maximization	Femoral, Tibial, Patellar	SSM training
Yin (2010)	x	Adaboost, random forest	Graph-cuts	Femoral, Tibial, Patellar	Ensemble classifiers training
Lee (2011)	x	Branch-and-mincut, kNN classifier	Graph-cuts, Markov random fields, kNN classifier	Femoral, Tibial, Patellar	kNN training, ground truth generation

Table 1: Overview of the knee cartilage segmentation approaches in literature

References

[1] Julio Carballido-Gamio, Jan S Bauer, Robert Stahl, Keh-Yang Lee, Stefanie Krause, Thomas M Link, and Sharmila Majumdar. **Inter-subject comparison of MRI knee cartilage thickness**. *Medical image analysis*, 12(2):120–135, 2008.

[2] Yin Yin, Xiangmin Zhang, Rachel Williams, Xiaodong Wu, Donald D Anderson, and Milan Sonka. **LOGISMOS—layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint**. *Medical Imaging, IEEE Transactions on*, 29(12):2023–2037, 2010.

[3] Soochahn Lee, Sang Hyun Park, Hackjoon Shim, Il Dong Yun, and Sang Uk Lee. **Optimization of local shape and appearance probabilities for segmentation of knee cartilage in 3-D MR images**. *Computer Vision and Image Understanding*, 115(12):1710–1720, 2011.

[4] Xiaojuan Li, Alex Pai, Gabrielle Blumenkrantz, Julio Carballido-Gamio, Thomas Link, Benjamin Ma, Michael Ries, and Sharmila Majumdar. Spatial distribution and relationship of t1ρ and t2 relaxation times in knee cartilage with osteoarthritis. *Magnetic Resonance in Medicine*, 61(6): 1310–1318, 2009.

[5] TCB Pollard, EG McNally, DC Wilson, DR Wilson, B Mädler, M Watson, HS Gill, and AJ Carr. Localized cartilage assessment with three-dimensional dgemric in asymptomatic hips with normal morphology and cam deformity. *The Journal of Bone & Joint Surgery*, 92(15):2557–2569, 2010.

[6] Jenny Folkesson, Erik B Dam, Ole F Olsen, Morten A Karsdal, Paola C Pettersen, and Claus Christiansen. Automatic quantification of local and global articular cartilage surface curvature: biomarkers for osteoarthritis? *Magnetic Resonance in Medicine*, 59(6):1340–1346, 2008.

[7] Guangju Zhai, Changhai Ding, Flavia Cicuttini, and Graeme Jones. Optimal sampling of mri slices for the assessment of knee cartilage volume for cross-sectional and longitudinal studies. *BMC musculoskeletal disorders*, 6(1):10, 2005.

[8] Robert Stahl, Anthony Luke, Xiaojuan Li, Julio Carballido-Gamio, C Benjamin Ma, Sharmila Majumdar, and Thomas M Link. T1rho, t2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early oa patients—a 3.0-tesla mri study. *European radiology*, 19(1):132–143, 2009.

[9] S Akhtar, CL Poh, and RI Kitney. **An MRI derived articular cartilage visualization framework**. *Osteoarthritis and Cartilage*, 15(9):1070–1085, 2007.

[10] SR Morgan, JC Waterton, RA Maciewicz, JE Leadbetter, SJ Gandy, RJ Moots, P Creamer, and AFP Nash. Magnetic resonance imaging measurement of knee cartilage volume in a multicentre study. *Rheumatology*, 43(1):19–21, 2004.

[11] Lászlo G Nyu and Jayaram K Udupa. On standardizing the mr image intensity scale. *image*, 1081, 1999.

[12] Julio Carballido-Gamio, Gabby B Joseph, John A Lynch, Thomas M Link, and Sharmila Majumdar. Longitudinal analysis of mri t2 knee cartilage laminar organization in a subset of patients from the osteoarthritis initiative: a texture approach. *Magnetic Resonance in Medicine*, 65(4): 1184–1194, 2011.

[13] Damian E Dupuy, Robert M Spillane, Michael S Rosol, Daniel I Rosenthal, William E Palmer, Dennis W Burke, and Andrew E Rosenberg. **Quantification of articular cartilage in the knee with three-dimensional MR imaging**. *Academic radiology*, 3(11):919–924, 1996.

[14] Jurgen Fripp, Stuart Crozier, Simon K Warfield, and Sebastien Ourselin. Automatic segmentation of the bone and extraction of the bone–cartilage interface from magnetic resonance images of the knee. *Physics in Medicine and Biology*, 52(6):1617, 2007.

[15] Gerard E Boyle, Mary Ahern, Jennie Cooke, Niall P Sheehy, and James F Meaney. An interactive taxonomy of mr imaging sequences1. *Radiographics*, 26(6):e24–e24, 2006.

[16] J Duryea, G Neumann, MH Brem, W Koh, F Noorbakhsh, RD Jackson, J Yu, CB Eaton, and P Lang. **Novel fast semi-automated software to segment cartilage for knee MR acquisitions**. *Osteoarthritis and cartilage*, 15(5):487–492, 2007.

[17] Hackjoon Shim, Samuel Chang, Cheng Tao, Jin-Hong Wang, C Kent Kwoh, and Kyongtae T Bae. Knee cartilage: Efficient and reproducible segmentation on high-spatial-resolution mr images with the semiautomated graph-cut algorithm method 1. *Radiology*, 251(2):548–556, 2009.

[18] Arish Asif Qazi, Jenny Folkesson, PC Pettersen, Morten A Karsdal, Claus Christiansen, and Erik B Dam. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. *Osteoarthritis and Cartilage*, 15(10):1199–1206, 2007.

[19] Jenny Folkesson, Erik B Dam, Ole Fogh Olsen, Paola C Pettersen, and Claus Christiansen. **Segmenting articular cartilage automatically using a voxel classification approach**. *Medical Imaging, IEEE Transactions on*, 26(1):106–115, 2007.

[20] Flavia Cicuttini, Andrew Forbes, Kevin Morris, Sandy Darling, Michael Bailey, and Stephen Stuckey. Gender differences in knee cartilage volume as measured by magnetic resonance imaging. *Osteoarthritis and Cartilage*, 7(3):265–271, 1999.

[21] Marc D van Leersum, Mark E Schweitzer, Frank Gannon, Simon Vinitski, Geral Finkel, and Donald G Mitchell. Thickness of patellofemoral articular cartilage as measured on mr imaging: sequence comparison of accuracy, reproducibility, and interobserver variation. *Skeletal radiology*, 24(6):431–435, 1995.

[22] F Eckstein, S Müller, SC Faber, K-H Englmeier, M Reiser, and R Putz. Side differences of knee joint cartilage volume, thickness, and surface area, and correlation with lower limb dominance—an mri-based study. *Osteoarthritis and cartilage*, 10(12):914–921, 2002.

[23] Alexander J Gougoutas, Andrew J Wheaton, Arijitt Borthakur, Erik M Shapiro, J Bruce Kneeland, Jayaram K Udupa, and Ravinder Reddy. **Cartilage volume quantification via Live Wire segmentation¡ sup¿ 1¡/sup¿**. *Academic radiology*, 11(12):1389–1395, 2004.

[24] José G Tamez-Pena, Monica Barbu-McInnis, and Saara Totterman. Knee cartilage extraction and bone-cartilage interface analysis from 3d mri data sets. In *Medical Imaging 2004*, pages 1774–1784. International Society for Optics and Photonics, 2004.

[25] Alberto Sanfeliu and Manuel Lazo Cortés. *Progress in pattern recognition, image analysis and applications*. Springer, 2004.

[26] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, 2004.

[27] Vicente Grau, AUJ Mewes, M Alcaniz, Ron Kikinis, and Simon K Warfield. **Improved watershed transform for medical image segmentation using prior information**. *Medical Imaging, IEEE Transactions on*, 23(4):447–458, 2004.

[28] Tannaz Iranpour-Boroujeni, Atsuya Watanabe, Reza Bashtar, Hiroshi Yoshioka, and Jeffrey Duryea. Quantification of cartilage loss in local regions of knee joints using semi-automated segmentation software: analysis of longitudinal data from the osteoarthritis initiative (oai). *Osteoarthritis and Cartilage*, 19(3):309–314, 2011.

[29] KT Bae, H Shim, C Tao, S Chang, JH Wang, R Boudreau, and CK Kwoh. **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**. *Osteoarthritis and Cartilage*, 17(12):1589–1597, 2009.

[30] Claude Kauffmann, Pierre Gravel, Benoît Godbout, Alain Gravel, Gilles Beaudoin, J-P Raynauld, Johanne Martel-Pelletier, J-P Pelletier, and Jacques A de Guise. **Computer-aided method for quantification of cartilage thickness and volume changes using MRI: validation study using a synthetic model**. *Biomedical Engineering, IEEE Transactions on*, 50(8):978–988, 2003.

[31] Ch Lantuejoul and F Maisonneuve. Geodesic methods in quantitative image analysis. *Pattern recognition*, 17(2):177–187, 1984.

[32] MH Brem, PK Lang, G Neumann, PM Schlechtweg, E Schneider, R Jackson, J Yu, CB Eaton, FF Hennig, H Yoshioka, et al. Magnetic resonance image segmentation using semi-automated software for quantification of knee articular cartilage—initial evaluation of a technique for paired scans. *Skeletal radiology*, 38(5):505–511, 2009.

[33] Alexandre X Falcão, Jayaram K Udupa, Supun Samarasekera, Shoba Sharma, Bruce Elliot Hirsch, and Roberto de A Lotufo. User-steered image segmentation paradigms: Live wire and live lane. *Graphical models and image processing*, 60(4):233–260, 1998.

[34] Yuri Y Boykov and M-P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112. IEEE, 2001.

[35] Pierre Dodin, J-P Pelletier, Johanne Martel-Pelletier, and François Abram. **Automatic human knee cartilage segmentation from 3-D magnetic resonance images**. *Biomedical Engineering, IEEE Transactions on*, 57(11):2699–2711, 2010.

[36] Kunlei Zhang, Wenmiao Lu, and Pina Marziliano. **Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies**. *Magnetic resonance imaging*, 31(10):1731–1743, 2013.

[37] Jurgen Fripp, Stuart Crozier, Simon K Warfield, and Sébastien Ourselin. **Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee**. *Medical Imaging, IEEE Transactions on*, 29(1):55–64, 2010.

[38] Elsa D Angelini and Edward J Ciaccio. **Optimized region finding and edge detection of knee cartilage surfaces from magnetic resonance images**. *Annals of biomedical engineering*, 31(3):336–345, 2003.

[39] Cristian Lorenz and Jens von Berg. Fast automated object detection by recursive casting of search rays. In *International Congress Series*, volume 1281, pages 230–235. Elsevier, 2005.

[40] Ben Glocker, Nikos Komodakis, Nikos Paragios, Christian Glaser, Georgios Tziritas, and Nassir Navab. Primal/dual linear programming and statistical atlases for cartilage segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 536–543. Springer, 2007.

[41] Soochan Lee, S Park, H Shim, I Yun, and S Lee. 3-d segmentation of knee bones on mr images by constrained branch-and-mincut. In *Probabilistic Models for Medical Image Analysis, A MICCAI Workshop*, 2009.

[42] Christian Bähnisch, Peer Stelldinger, and Ullrich Köthe. Fast and accurate 3d edge detection for surface reconstruction. In *Pattern Recognition*, pages 111–120. Springer, 2009.

[43] Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of chemical information and computer sciences*, 43(6): 1947–1958, 2003.

[44] Ramón Díaz-Uriarte and Sara Alvarez De Andres. Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1):3, 2006.

[45] Bing Niu, Yu-Dong Cai, Wen-Cong Lu, Guo-Zheng Li, and Kuo-Chen Chou. Predicting protein structural class with adaboost learner. *Protein and peptide letters*, 13(5):489–492, 2006.

[46] O Coulon, SJ Hickman, GJ Parker, GJ Barker, DH Miller, and SR Arridge. Quantification of spinal cord atrophy from magnetic resonance images via a b-spline active surface model. *Magnetic resonance in medicine*, 47(6):1176–1185, 2002.