Abstract. Osteoarthritis is one of the major socio-economic burdens today. Patients often undergo magnetic resonance image examination which facilitates detection of pathological occurence in cartilage tissue. Images are usually manually examined by radiologists who aim at assessing any pre- and post treatment changes even if it is a time consuming process prone to inter- and intra-reader variability. This thesis presents an algorithm for semi-automatic segmentation of cartilage in magnetic resonance images acquired from knees of patients with clinically verified osteoarthritis. The method requires minimal manual intervention in which an operator quickly outlines cartilage tissue with a rectangular region. Segmentation is conducted within the region using a k-nearest-neighbour classifier that operates on a very small training set deduced from a large training set using an unsupervised clustering algorithm. In order to separate cartilage voxels from background the algorithm exploits the unique geometry of cartilage using a novel similarity map approach that assess intensity differences between voxels in a small neighbourhood to model their spatial similarity distribution. The method is tested on 50 1.5T magnetic resonance images and allows for more accurate analysis of 3D cartilage models in clinical investigations. Results were compared to state-of-the-art techniques and were favoured by clinical professionals.

Keywords: Segmentation, Cartilage, Osteoarthritis

Abstract. Slidgigt er en stor byrde for det moderne sundhedsvæsen. Patienter får ofte taget et MR billede, udfra hvilket bruskforandringer kan undersøges. Billederne bliver normalt undersøgt af radiologer, der vurderer eventuelle ændringer, selvom de kan være svære at identificere og undersøgelser derfor svære at reproducere. Denne afhandling præsenterer en algoritme til semi-automatisk segmentering af brusk i magnetic resonance billeder erhvervet fra knæ af patienter med klinisk verificeret slidgigt. Metoden kræver minimal manuel intervention, hvor en operatør hurtigt skitserer bruskvæv med et rektangulært område. Segmentering gennemføres inden for regionen ved hjælp af en k-nearest-neighbour algoritme på baggrund af et lille træningssæt udledt fra et stort træningssæt af en fuld-automatisk clustering algoritme. For at adskille brusk fra baggrund udnytter algoritmen brusks unikke geomtri ved hjælp af similaritetsmodeller, der bruger lysintensitetsforskelle til at evaluere lighed mellem en voxel og dens naboer. Metoden er afprøvet på 50 1.5T magnetic resonance billeder og giver mulighed for mere præcis analyse af 3D brusk modeller i kliniske undersøgelser. Resultaterne blev sammenlignet med nyere teknikker og foretrukket af klinisk personale.

Nøgleord: Segmentering, Brusk, Slidgigt

Acknowledgements

I would like to thank all the people who have helped and inspired me during the preparation of this thesis. First and foremost I would like to thank my supervisors at the University of Copenhagen, Associate Professer Stig Steenstrup at Niels Bohr Institute and Associate Professor Bente Rona Jensen at Department of Sports and Exercise for making this thesis possible. They have supported me thoughout the process with patience, knowledge and invaluable discussion whilst always allowing room for my own ideas. One simply could not wish for better or friendlier supervisors.

I owe my deepest gratitude towards my family and friends who have supported me throughout my studies even if they learned about machine segmentation more than they ever wanted to know. Special thanks to Peter Krusager for splendid technical and for bearing the brunt of my frustration and rages against the seemingly free will of computers (I'm sure it's Skynet).

It is impossible to list all the individuals who have assisted me in the preparation of this thesis but I offer my regards to all of those who supported me in any respect during the completion of the project. It is a particular pleasure to thank all my colleagues at The Parker Institute, Frederiksberg Hospital. This work would not have been possible without their collective work on the CAROT-study from which this thesis has its data and in my daily work I have enjoyed a very pleasant atmosphere. An especially big thank-you to Dr. Olga Kubassova [Image Analysis Ltd., Leeds, United Kingdom] for always incredibly intelligent criticism and technical inspiration. Her encouragement, guidance and support from the initial to the final level enabled me to develop the necessary understanding of medical image processing. I owe big thanks to MD PhD fellow Henrik Gudbergsen for segmentation of too many MR images, Research Radiogist Rasmus Bouert for setting up the MR sequence and acquiring images, head of research Professor DMSc Henning Bliddal for medical advice, Senior Researcher at the imaging department MD PhD Mikael Boesen for Kellgren-Lawrence scoring of MR images and medical advice on OA, PT PhD Marius Henriksen for medical advice on gait analysis, MD PhD fellow Birgit Falk Riecke for clinical assessment and cand.scient, PhD Robin Christensen for advice on statistics.

A big thank-you to The Parker Institute for funding of study trips to Leeds, UK, during the summers of 2009 and 2010 and for making available a state-of-the-art NVIDIA graphics card and Mac Pro beast throughout my research. I would also like to thank all my friends at the University of Copenhagen for a wonderful time during our first three years at uni.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Osteoarthritis (OA) is a major cause of morbidity and disability, affecting millions of people [1]. More than 10% of the general population is affected by OA [2] with prevalence increasing with age. Approximately 70% of a population over 65 years suffers from OA [3].

OA is a syndrome of joint disorders with cartilage destruction and may be regarded as a common joint failure induced by disease entities of different aetiologies [4]. The development of OA follows complex patterns with gradually evolving disease manifestations including decreased cartilage water content, degradation of cartilage matrix components, collagen fibers and glucose-aminoglycans (GAG) [5, 6]. The process is irreversible due to the poor repair potential of the hyaline cartilage tissue whatever the underlying aetiology [7]. No international agreement has been reached regarding a common definition of OA, but it is generally accepted that disease processes comprise cartilage destruction, remodeling of the subchondral bone and soft tissue involvement which in combination with an inflammatory reaction causes the well known symptoms of pain, swelling, limited range of motion and loss of function. OA is expected to be among the major disabling illnesses in the future [2].

Currently, the treatment of OA focuses on symptom control [8] with joint replacement as a common result of end-stage disease. Disease-modifying drugs (DMD) are still to be found and tested [9]. Therefore, it is desirable to improve methods for early detection of OA. Robust tools for early identification of disease manifestation in cartilage tissue and objective evaluations of medications and physical interventions might support treatment algorithms which allow for preventive treatment, postponing and in a future setting, potentially arresting OA altogether. Studies have shown that volume, thickness, curvature and homogeneity measurements of cartilage from Magnetic Resonance (MR) images are promising biomarkers for OA progression [10, 11, 12, 13, 14]. The first step to obtain such quantitative measures is segmentation of articular cartilage tissue.

The primary objective of this thesis is to develop an algorithm for automated segmentation of knee articular cartilage that is sufficiently robust to replace a manual procedure. While the proposed method may be applicable to other tissues, this thesis is focusing on segmentation of knee articular cartilage. A second objective is to assess the quality and reliability of the proposed algorithm and to demonstrate its applicability in clinical practice.

1.2 MRI for Quantitative Cartilage Assessment

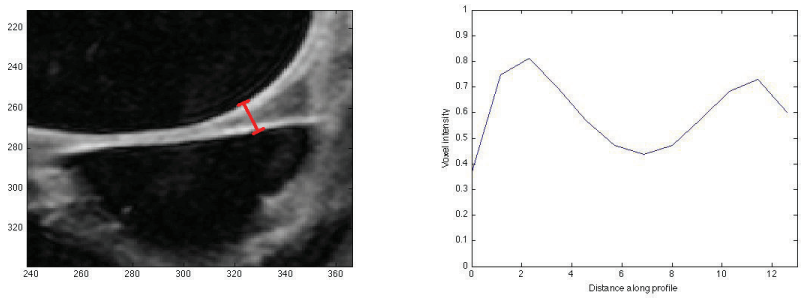
MRI plays in increasingly prominent role in diagnosis of OA because it allows for quantification of disease influence on cartilage morphology. MRI is the most widespread imaging modality for direct, non-invasive assessment of cartilage [15] and structural changes of cartilage tissue are statistically significant in MRI [16]. A significant advantage of MRI is that it allows for acquisition of high-contrast cartilage images.

Nonetheless, clinical use of imaging is challenged by subjectivity in image interpretation and variability of imaging techniques even within modalities. Computer based measurements may improve accuracy of the diagnostic procedure through quantitative tissue-based evaluation of diseases and decrease observer bias inherently introduced in human analysis.

However, machine segmentation is not an easy task. Due to the thin structure of cartilage, detection of early cartilage lesions on sub-millimeter scale requires high precision in order to obtain statistical significant results. Low contrast between cartilage-meniscus and surrounding tissue makes automation of tissue interface detection challenging [17] (Fig. 1a and 1b) and image artifacts like the partial volume averaging effect further reduce the sensitivity of MRI for measurements of thin fissures, cartilage flaps and shallow defects and makes the tissue boundary difficult to delineate (Fig. 2). Ultimately, cartilage degeneration could be overlooked or falsely verified. However, if successfully executed, an efficient computer-based method would eliminate intra- and inter-reader variability, unreasonable time consumption and high cost of a manual method and allow for non-specialized medical personel to obtain patient-specific cartilage anatomical features in clinical practice. Therefore, a significant amount of research has focused on computer vision to aid in the diagnostic procedure. The following section is concerned with the most prominent of such methods.

1.3 Related Work

Due to the lack of sufficiently accurate automated algorithms, manual or semi-automatic methods are typically applied in research where segmentation is needed [18]. Utilized semi-automatic approaches include region growing [19], active contours [13], B-spline snakes [20, 21] and live-wires [22]. However, training and utilization of operators are time consuming [23], prone to inter- and intra-observer variability [24] and often operate on a slice by slice basis. Therefore, manually obtained quantitative cartilage measurements are generally avoided in diagnostics.

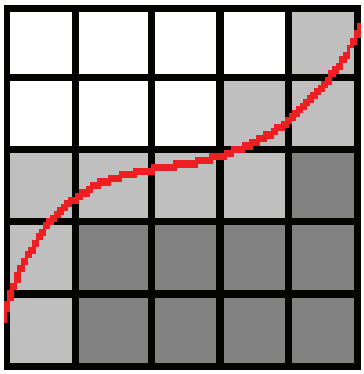


(a) (b)

Figure 1: Tissue interface profile. (b) shows intensities along the line in (a) from bottom to top. The borderline is poorly delineated.

Recently, several promising methods have been proposed for automated segmentation of cartilage from MRI data. Folkesson et al. [24] applied a method based on pixelwise segmentation of features derived from local neighbourhoods. Fully automatic segmentation is conducted using a trained tissue classifier validated on a large sample of pathological low-field MR images. Fripp et al. [18] used a three-dimensional statistical shape model to segment bone. BCI was extracted used a deformable model that utilizes localization and patient-specific tissue estimation followed by utilization of a model of thickness variation to yield cartilage boundary. The method was demonstrated on non-pathological knees but they reported interesting results and the method was fully automatic. Grau et al. [25] used an improved watershed transform incorporating prior information to segment cartilage. The method yielded the most accurate results in the literature but tests were performed on 7 scans from 4 non-pathological subjects and method required around 10 minutes of manual intervention for each cartilage compartment.

Figure 2: Schematic depiction of partial volume averaging, one of the effects that make machine segmentation difficult. The red line delineates the tissue interface between a white and a dark gray tissue. Contrast between tissues suffer from inferior voxel resolution.



1.4 Overview of the Work Presented

In this thesis, a novel method for semi-automatic cartilage segmentation is presented. The proposed method is based on spatial composition of similarity between a voxel and its neighbours and requires minimal manual intervention. We demonstrate the applicability of the method in clinical practice and that it is possible to achieve greater precision compared to previous research even on low-quality pathological images. The method is tested on cartilage sensitive 3D images of participants from an ongoing in-house trial investigating the influence of weight loss in obese, elderly patients with verified Knee OA [26]. In these patients, it is particularly challenging to obtain high quality images due to the increased distance between cartilage tissue and skin surface, resulting in an inferior MR signal to record.

The thesis is organized as follows. In section 2 MR data are presented and we expose the main contribution of this work which is novel voxel features and utilization of the k-means algorithm for training data optimization. In section 3 validation methodology and experimental results are presented and in section 4 these are analyzed and discussed with respect to OA severity. The thesis is concluded by discussing a potentially promising method for early detection of OA which employs the proposed method.

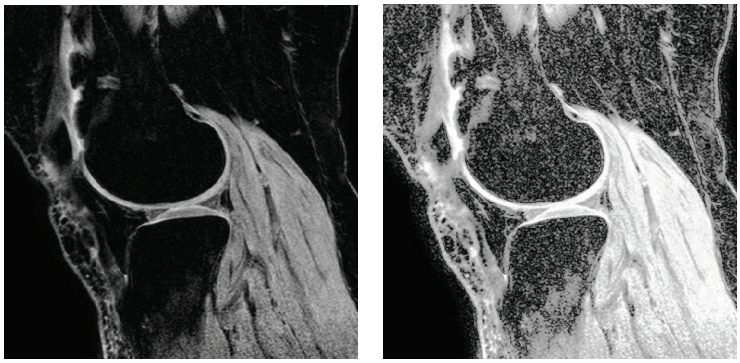
2 Materials and Methods

2.1 Data

50 OA patients were imaged using a 1.5T Philips Gyroscanner (Philips Medical Systems, Best, The Netherlands) running Intera Software 12.1.5.4 (2003). A FLASH Gradient Echo (GR) MR sequence was used with the following settings: Flip angle 20 degrees; repetition time 21 ms; echo time 8.4 ms; field of view 140×140 mm. 80 sagittal 512×512 partitions were obtained at a partition thickness of 1.5 mm with an in-plane resolution of 0.2734×0.2734 mm. To treat all scans analogously, images of left knee are reflected about the center in the sagittal plane.

2.2 Patient Demographics

All patients, 43 of which were females, were diagnosed with OA according to the American College of Rheumatology criteria with clinical signs and symptoms as well as radiologically or arthoscopically verified OA in one or both knees.



(a) Sample of acquired MRI data.

(b) Sample data after preprocessing. Before segmentation, histogram equalization is applied to homogenize grayscales across images.

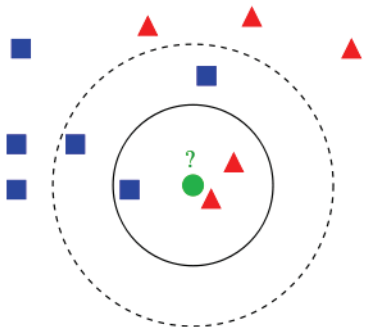
Figure 3: Illustration of MRI data.

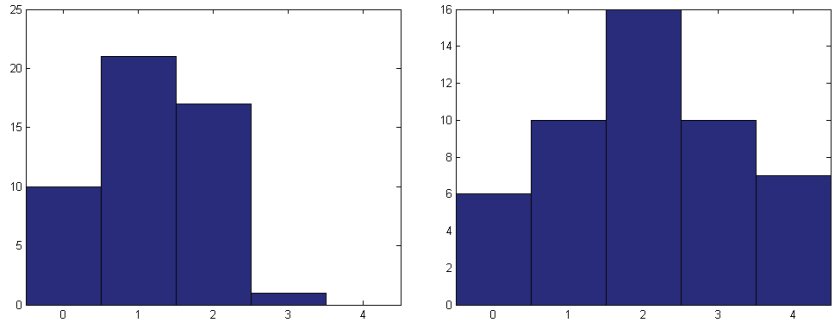
Patients were at least 50 years of age (61.5 ± 6.4 std) and obese with a BMI of at least 30 kg/m^2 (37.3 ± 4.7 std). The MRI scan was performed on the knee with most active OA. The status of the knees on the Kellgren-Lawrence index (KLi) [27] is seen in Fig. 4.

2.3 Method

In this work, a classifier for segmentation of articular cartilage is constructed by adopting a supervised learning approach that learns a function mapping numerical representations of voxels into foreground and background by looking at input-output examples of expert manual segmentation. The goal of the learning process is to construct a classifier that mirrors a human decision engine. We utilize the k-nearest neighbour (kNN) machine learning algorithm with the standard euclidean distance metric. Given training and query data, the kNN algorithm classifies each unlabeled query vector by vote among its k nearest neighbours

Figure 5: 2D schematic depiction of the kNN algorithm. Query point class is determined by a vote among the k nearest neighbours in the training set. The test sample (green circle) should be classified either to the first class of red triangles if $k = 3$ or or to the second class of blue squares if $k = 5$ [29].





(a) Histogram of Kellgren-Lawrence scores of latera compartments.

(b) Histogram of Kellgren-Lawrence scores of meidal compartments. Scoring was not possible on 1 scan due to poor image quality.

Figure 4: Histograms of Kellgren-Lawrence index scores. KL$i = 0$ is healthy, KL$i = 1$ is considered borderline or mild OA and KL$i \geq 2$ is severe OA. OA is more frequently observed in the medial compartment [28] which is in good agreement with our dataset.

in the training set. kNN is one of the most frequently used machine learning algorithms due to its simplicity and ability to model abitrary complex target functions. In addition, it is possible to choose the exact weight of each training vector in the training set, a concept which this work utilizes to great extent as discussed in section 2.6.

Let $\mathbf{M} \in \mathbb{R}^d$ denote a numerical representation of cartilage and let $L \in \{0, 1\}$ denote background and cartilage labels respectively. Training examples are drawn from the joint distribution $E(\mathbf{M}, L)$ provided by expert segmentation and a set of m such examples is referred to as the training data $\mathbf{T}$. The classifier is a function f that maps from numerical representations to labels. Based on training data, we seek an f that correctly predicts the class $L = f(\mathbf{M})$ of new objects $\mathbf{M}$. This is accomplished by constructing an abstract vector space $\mathcal{S}$ of voxel features in which the classifier operates. This vector space is also refered to as the feature space and numerical representations of voxels are refered to as feature vectors. The goal is to find a set of features for which cartilage feature vectors are predisposed to cluster together so that cartilage is efficiently separated from background by the classifier. In the following section a set of features is derived that carry this property to great extend.

2.4 Feature Selection

Knee articular cartilage is located on the surface of bone and consist of several layers, each with it's own composition of chemical components. These are

mainly Type II collagen (15-20% of tissue weight), proteoglycans (PG, 3-10%) and water (65-80%) [30]. Using the FLASH MR seqeunce, the abundance of water yields a higher signal intensity in cartilage than in surrounding tissue. Therefore, cartilage comprise a plane of voxels with high intensity sandwhiched between planes of lower intensity. This structure is assumed to be unique to cartilage in the subchondral region of the image. A numerical representation that enhances the qualitative difference between neighbourhoods of cartilage voxels and neighbourhoods of voxels exterior to the cartilage plane is pursued. The proposed method quantifies the (dis)similarity between a voxel and all voxels in its neighbourhood and exploits the spatial composition of similarity to identify cartilage.

Let $\mathcal{I} : \Omega \longmapsto \mathbb{R}^+$ be a single channel image defined on $\Omega \in \mathbb{R}^3$ and let $\Gamma : \Omega \longmapsto \{0, 1\}$ be a binary grid of object or background labels. Let p, q denote voxels and let $\eta(p) = (q_1, q_2, \ldots, q_n)^T$ denote the neighbourhood of p where n is the number of neighbours. We define the similarity operator S^p that maps the voxel on which its working onto a similarity scale with respect to p. The result is refered to as a *similarity element*,

$$S^p q \equiv G_\sigma * \|\eta(p) - \eta(q)\|_2^2, \tag{1}$$

where G_σ is a normalized gaussian kernel with standard deviation σ and $*$ denotes convolution on the neighbourhood. Applying this operator to the entire neighbourhood, one obtains what we refer to as a *similarity map* that expresses the geometrical configuration of $\eta(p)$,

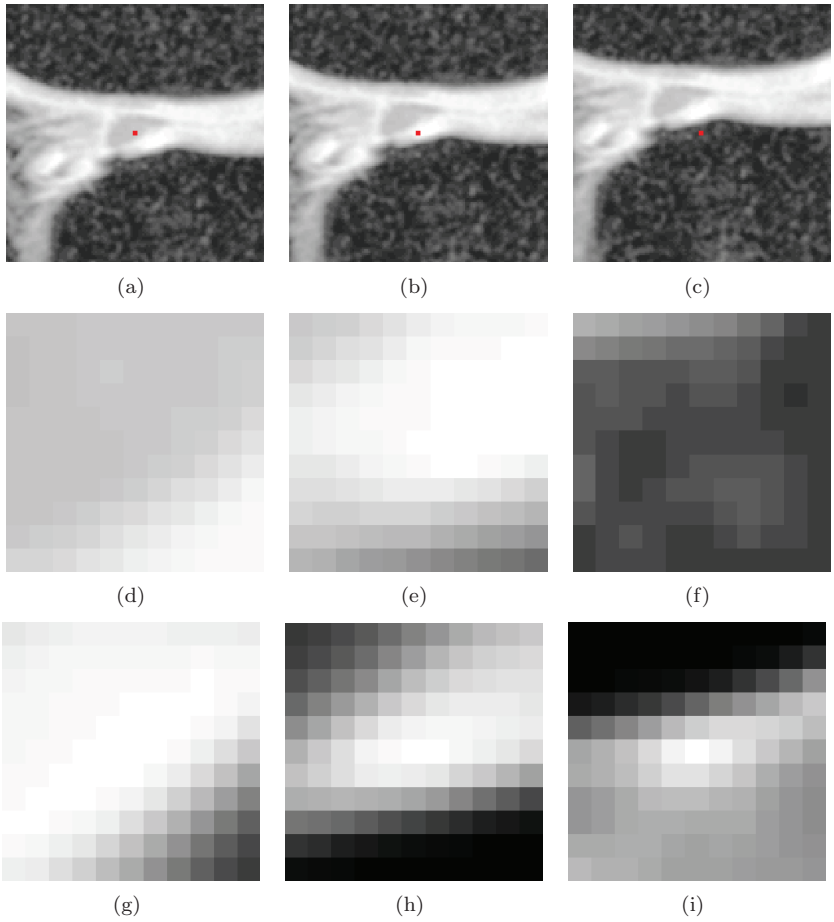
$$\mathbf{M}(p) \equiv S^p \eta(p) = (S^p q_1, S^p q_2, \ldots, S^p q_n)^T. \tag{2}$$

Similarity maps naturally generalize to three dimensions eliminating the problem of limited continuation in 2D methods. Correlation between the geometrical configurations of neighbourhoods is exposed by distances between similarity maps. Therefore they facilitate feature vectors of the proposed method.

2.5 Implementation

The proposed algorithm is implemented in MATLAB R2010a (Mathworks, MA, USA) with computation of similarity maps implemented in C and the NVIDIA CUDA programming interface. CUDA allows for general-purpose arithmetic on NVIDA Graphics Processing Units (GPUs) which provides orders of magnitude increase in floating-point throughput compared to regular CPUs. With an

Figure 6: 11×11 similarity maps. Top row shows the position of p on a preprocessed image, middle row shows the local neighbourhood $\eta(p)$ and bottom row shows similarity maps computed using (2). Left column illustrates how nearby non-cartilage soft tissue appear in the feature space. Center column illustrates a typical cartilage similarity map. Right column illustrates similarity map of nearby bone. The similarity map of cartilage is qualitatively different from those of non-cartilage tissues even close to tissue interfaces.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

NVIDIA GTX 480 graphics card, runtime was decreased by more than an order of magnitude compared to a fully vectorized MATLAB implementation that was tailored for the built-in LAPACK matrix algebra library. Given training sets, segmentation is conducted as follows.

1. Gray values are normalized to the interval $[0; 1]$ and histogram equalization is applied to homogenize grayscales across images (Fig. 3).
2. A Region Of Interest (ROI) is specified by the user as a rough smallest bounding box around cartilage. This is in principle a six-click manual segmentation. While being a manual procedure it only takes seconds to peform.
3. Classification is conducted throughout the ROI by a kNN algorithm that operates on few, specially chosen training vectors and associated weights.

Step 2-3 is repeated for each cartilage compartment. In this work only tibial lateral cartilage is segmented.

2.6 Training Set

In step 3, classification is conducted using few, characteristic training vectors. The idea is that an all-inclusive training set may carry a lot of redundant training vectors that implicitly influence the nearest neighbour scheme wherefore the classifier may miss out on important expert decisions. This section presents a method for selecting the most explanatory feature vectors and utilize the redundancy to assign weight coefficients. The method partitions the training set into clusters using the k-means algorithm and assign weights to the cluster center vectors according the average classification within the cluster. The training set used by the classifier consist of cluster center vectors and their associated weights. This could potentially provide a more meaningful classification due to fact that the training set is analyzed twice and a larger of number of expert decisions are assessed per classsification. This will also siginificantly reduce the number of training vectors and consequently the runtime.

k-means is an unsupervised clustering algorithm that iteratively partitions vectors into k clusters where each vector belongs to the cluster with nearest mean. Formally, given a set of observations $(\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_m)$ the k-means algorithm aims to partition the m observations into k clusters $(k < m)$ $\mathbf{C} = \{C_1, C_2, \ldots, C_k\}$ that minimizes the within-cluster sum of squares,

$$\arg\min \sum_{i=1}^{k} \sum_{\mathbf{M}_j \in \mathbf{C}_i} \|\mathbf{M}_j - \mu_i\|^2, \tag{3}$$

where μ_i is the mean of points in $\mathbf{C}_i$. In this work, MATLAB's `kmeans()` is utilized. Each center vector is assigned a likelihood coefficient expressing the average classification of a vector near $\mathbf{M}_i$ defined as

$$p_i = \frac{1}{k'} \sum_{j=1}^{k'} \delta_{ij} \quad , \quad \delta_{ij} = \begin{cases} 1 & \text{if } \mathbf{M}_{ij} \text{ is cartilage} \\ 0 & \text{if } \mathbf{M}_{ij} \text{ is background} \end{cases} \tag{4}$$

where j runs over the k' nearest neighbours of $\mathbf{M}_i$ in the full training set. The classification rule is obtained by thresholding the mean of k'' nearest p_i. The number of clusters needed to sufficiently describe the training set was observed to be in the order of 10^2. The kNN algorithm operates very fast on such a small training set with a runtime in the order of 10^{-1} seconds even with several thousand voxels to classify.

Due to anatomical bonds, similarity elements are highly correlated and similarity maps are well suited for dimensionality reduction. It has been shown that the principal components obtained with Principal Component Analysis (PCA) are the continuous solutions to the discrete cluster membership indicator for k-means clustering [31]. PCA inherently projects data to the subspace where the global solution of k-means clustering lies and thus facilitates k-means clustering to find near-optimal solutions. This motivates the use of PCA for dimensionality reduction which is discussed in the following section.

2.7 Dimensionality Reduction With PCA

In high dimensional vector spaces, the concept of proximity, distance or nearest neighbour may not be qualitatively meaningful. In addition, the kNN algorithm is $O(mp)$ in runtime complexity where $m = |\mathbf{T}|$ is the number of vectors in the training set and p denotes the cost of evaluating the query predicate for one single object, which is usually dominated by the complexity of the applied distance function. Thus a diluted distance measure and long runtime is often associated with high data dimensionality wherefore it is desirable to project data onto a lower dimensional manifold.

PCA projects feature vectors into a subspace spanned by the most significant eigendirections in $\mathcal{S}$. In essence, vectors are rotated around their mean such

that the greatest variance by any projection of the data comes to lie on the first principal component, the second greatest variance on the second component, and so on. Each principal component is a linear combination of the original variables. Formally, the first principal component is given by

$$\mathbf{w}_1 = \arg\max_{\|\mathbf{w}\|=1} \mathrm{Var}\{\mathbf{w}^T\mathbf{T}\}. \tag{5}$$

where $\mathbf{w}_1$ is the sought component and $\mathbf{T}$ is a matrix of training instances where each row corresponds to a vector. The training set is normalized to zero empirical mean. The kth component is found by subtracting the first $k-1$ components from $\mathbf{T}$,

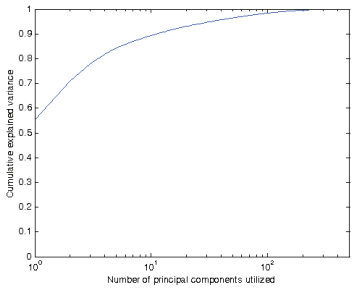
$$\hat{\mathbf{T}}_{k-1} = \mathbf{T} - \sum_{i=1}^{k-1} \mathbf{w}_i \mathbf{w}_i^T \mathbf{T}, \tag{6}$$

and substituting this as the new $\mathbf{T}$,

$$\mathbf{w}_k = \arg\max_{\|\mathbf{w}\|=1} \mathrm{Var}\left\{\mathbf{w}^T\hat{\mathbf{T}}_{k-1}\right\}. \tag{7}$$

In this work, MATLAB's `princomp()` is used to yield the principal components. Only the first 10^2 (order of magnitude) components are needed to explain a significant cut of the variance (Fig. 7) and it was found empirically that $n = 30$ principal components are explaining a sufficient 93% of the variance.

Figure 7: Explained variance by the principal components applied to tibial lateral training set.



3 Experiments

3.1 Validation Methodology

Machine classifier performance is investigated by comparing automatically acquired cartilage volumes to expert binary segmentation with the sensitivity $\frac{T_P}{T_P+F_N}$, specificity $\frac{T_N}{T_N+F_P}$ and Dice Similarity Coefficient (DSC) $\frac{2T_P}{2T_P+F_P+F_N}$

similarity measures. T_P, T_N, F_P and F_N are abbreviations for True Positive (correctly classified cartilage voxels), True Negative (correctly classified background voxels), False Positive (cartilage voxes labeled as background) and False Negative (background voxels labeled as cartilage) respectively. Sensitivity is the ratio of correctly classified cartilage voxels to the total number of cartilage voxels, specificity is the ratio of correctly classified background voxels to the total number of background voxels and DSC is a statistical measure of volume overlap.

Images were manually segmented by an experienced observer using 3DManSeg (Nordic Bioscience, Denmark). Segmentation was conducted on a slice-by-slice basis with 20 images segmented twice. Whether or not manual segmentation is a valid reference is subject to debate in the scientific community. It is generally agreed, however, that while manual segmentation does not provide a "ground truth", the expert "gold standard" segmentation is indeed sufficient and the only feasible method for *in vivo* cartilage studies. Similarity between segmentation and resegmentation is a common indicator of manual segmentation quality. The Coefficient of Variation (CV) of total volume differences between our segmentation and resegmentation is CV $= \sigma(\triangle \mathrm{V})/\overline{\triangle \mathrm{V}} = 6.0\%$ where $\triangle$V is volume differences and $\overline{\triangle \mathrm{V}}$ is the mean volume difference. This is reasonable for a manual procedure.

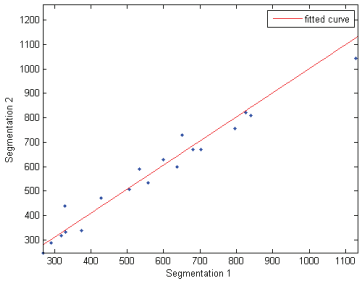


Figure 8: Assessment of manual segmentation; scatter plot of volumes, manual segmentation vs. manual resegmentation. A CV of 6.0% and good fit ($R^2 = 0.967$) indicate satisfactory quality.

3.2 Calibration of Parameters

20 images are randomly selected as training and optimal parameters are automatically estimated from the training set using a leave-one-out approach. The search is conducted in parallel on a Mac Pro with two 2.26GHz Quad-Core Intel Xeon Nehalem processors (8 physical cores, 16 virtual cores) and 16GB DDR3 RAM. While a high classification correct rate yields the best results on paper, it doesn't always provide the most meaningful segmentation. Therefore, a combination of high classification correct rate and manual assessment of cartilage geometry is used to find the optimal set. For tibial lateral cartilage the most intelligent segmentions were achieved with the following parameters: $\eta(p)$ spans 5×5 voxels or 1.3670×1.3670 mm, the standard deviation of the gaussian

neighbourhood convolution kernel is $\sigma = 0.36$ mm and $M(p)$ consist of similarity elements from every other voxel in the sagittal plane in a $21 \times 21 \times 3$ window ($13 \cdot 13 \cdot 3 - 1$ similarity elements) or $5.74 \times 5.74 \times 0.82$ mm. This is in good agreement with the fact that similarity maps would need to span the typical thickness of cartilage in the sagittal plane ($3 - 5$ mm [32]) in order to incorporate sufficient geometrical information.

3.3 Results

Tibial lateral cartilages in 30 images were segmented using the approach described in section 2. Segmentations were conducted on a PC with a 2.4GHz Intel Quadcore processor, 8GB of and a NVIDIA GTX 480 graphics card. The graphics card was used for computation of similarity maps which takes around 1 second for a typical ROI with a couple of thousand voxels. A naive third-party C-implementation of the kNN algorithm was used for classification with a typical runtime in the order of 10^-1 seconds per ROI. User interaction was simulated by computing the smallest bounding box of expert binary segmentation and expanding the ROI by 10 voxels in each direction in the sagittal plane. Results are presented in table 1.

Table 1: Results.

	Tibial Cartilage		
	Sens.	Spec.	DSC
Lateral	0.904 ± 0.053	0.999 ± 0.001	0.850 ± 0.028

Classification with the special training set outperformed an all-inclusive approach. The method appears very general and is probably applicable to other instances of kNN machine learning.

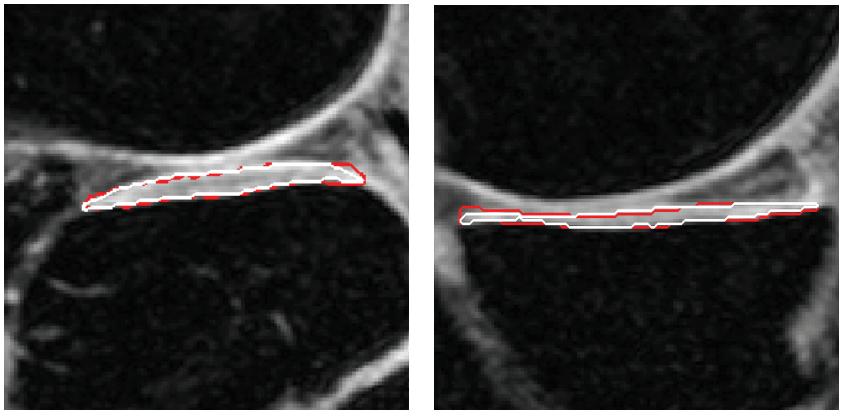
4 Discussion

In this study, a novel method for semi-automatic segmentation of knee articular cartilage has been presented and tested on a dataset of images from both healthy and osteoarthritic subjects. When comparing our method to Grau et al. [25] an equivalent sensivitiy and lower DSC suggest that the proposed method is more prone to over-segmentation. Grau et al. produces the best results but it should be noted that the method was validated on 4 subjects and requires a significant amount of manual intervention.

Table 2: Comparison with earlier studies. Note that methods operate on different data sets and this work only operates on tibial lateral cartilage.

	Tibial Cartilage		
	Sens.	Spec.	DSC
Fripp et al.	0.829 ± 0.207	0.999 ± 0.000	0.826 ± 0.083
Grau et al.	0.900 ± 0.030	0.999 ± 0.000	0.895 ± 0.001
Folkesson et al.	0.868 ± 0.106	0.999 ± 0.010	0.810 ± 0.067
Our Approach (lateral)	0.904 ± 0.053	0.999 ± 0.001	0.850 ± 0.028

Image quality had a noteworthy influence on segmentation accuracy especially in regions of low contrast where the border is more difficult to delineate. Overall, the primary cause of segmentation errors were regions affected by partial voluming and other signal decreases. This occurred primarily around the edges, in thin regions of the cartilage and where high curvature in slice thickness direction was most significant. It is suspected that further image homogenization may counter poor image quality to a certain extent and improve on these special cases; since similarity maps are inherently defined by the intensity scale, any procedure that enhances consistency across scans may potentially improve segmentation accuracy. Future work will investigate this issue.



(a) Segmentation of tibial lateral cartilage. Slight over-segmentation is seen near edges in thin regions of cartilage. Sensitivity, specificity and dice similarity coefficient for this volume is 0.970, 0.999 and 0.922 respetively.

(b) Segmentation of tibial medial cartilage. While this compartment was not rigorously investigated in this work an example is here presented. Slight over-segmentation is seen near the left edge. Sensitivity, specificity and dice similarity coefficient for this scan is 0.912, 0.999 and 0.906 respectively.

Figure 9: Comparison with manual segmentation. Manual segmentation is in white and automatically obtained bounderies are in red.

To study a potential applicability in clinical practice Table 3 shows results as a function of Kellgren-Lawrence score. This allows for analysis of classifier performance on the severe pathological knees.

Table 3: Results ordered by KLi score.

	Tibial Lateral		
	Sens.	Spec.	DSC
KL$i = 0$ (7 images)	0.892 ± 0.052	0.999 ± 0.001	0.849 ± 0.037
KL$i = 1$ (14 images)	0.909 ± 0.051	0.999 ± 0.001	0.850 ± 0.033
KL$i = 2$ (8 images)	0.913 ± 0.063	0.999 ± 0.001	0.854 ± 0.015
KL$i = 3$ (1) images)	0.850 ± 0.000	0.999 ± 0.000	0.842 ± 0.000
KL$i = 4$ (0 images)	-	-	-

Sensitivity is lower on KL$i \geq 3$ images but overall the algorithm performs well even on severe pathological images which are considered challenging to segment due to irregular bone- and cartilage geometry and inter- and intra-tissue pathological variation. This suggests that the proposed method is well suited for clinical application; although more data of KL$i \geq 3$ and additional compartments assessed are needed to draw final conclusions. The expected outcomes include autonomous segmentation for accelerated clinical trials and simpler pre- and post-operative disease monitoring and early detection of pathology-associated changes in thickness and volume for early diagnosis. Small variances convey good robustness and results look promising. To our knowledge, this is the fastest method in the literature including fully automated methods. A conservative estimate of manual intervention time for outlining four ROIs (tibial lateral, tibial medial, femural lateral and femural medial) is two minutes per MR volume which is favourable compared to at least $30 - 40$ minutes of manual segmentation per volume.

The user intervention makes the inevitable variability of knee placement in scanners a non-issue though we would like to automate this step in the future by incorporating a flood-fill algorithm seeded at random voxels within the ROI. If a voxel was found to be cartilage, its 26-neighbourhood would be examined. This process would repeat until $\Gamma(\mathcal{I}) = 1$ is bounded by $\Gamma(\mathcal{I}) = 0$ or ROI boundary. The result would be number of connected regions of cartilage and a significantly reduced number of processed background voxels. In addition, we will investigate how to choose seeds for a fully automated framework.

Further, we would like to investigate two slightly more sophisticated machine learning algorithms. Since the kNN algorithm classifies each unlabeled example by the majority label among its nearest neighbours its performance depends crucially on the distance metric. Due to geometrical bonds, similarity elements are highly correlated. Therefore it seems compelling to apply algorithms which

take this into account. The Multi-Metric Largest-Margin Nearest neighbour (MMLMNN) is a recently proposed method that learns a Mahalanobis distance metric from the feature space to attain an optimal decision rule, and the authors report interesting results [33].

Support Vector Machines (SVM) construct a hyperplane or a set of hyperplanes in the feature space that optimally separate groups in the training set. State-of-the-art SVMs produce some of the most competitive results in the field of machine learning. However, the computationally expensive nature of training SVMs puts constraints on the feasible size of training sets. Carpenter [34] provides a CUDA implementation which offers substantial speedup. This may make SVMs practically feasible in this work. Future work will explore MLMNN's and SVM's applicability to this method.

The proposed method does carry some limitations. The classifier was evaluated only on a single cartilage compartment due to difficulty of parameter estimation. Future work will include studying classifier performance on remaining cartilages. Further, it has not been possible to assess the reproducibility which is greatly desired in the validation process as this is often touted as the main advantage automated methods carry over the manual procedures. Finally, the method may require seperate training sets for different MR sequences.

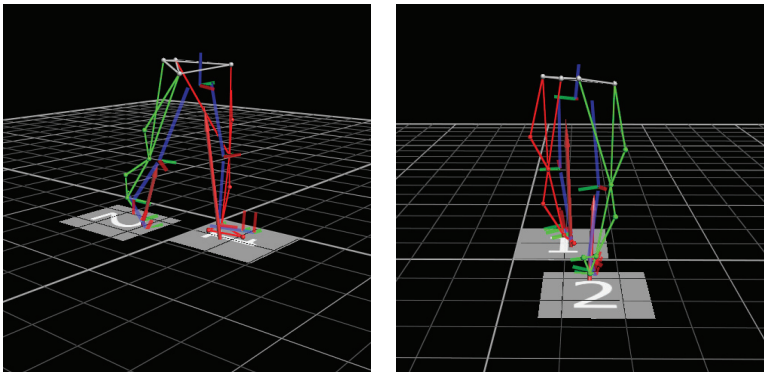
5 The Future of OA Treatment: Early Intervention Based On Patient-specific Cartilage Load Distribution

Accurate, patient-specific cartilage geometry may potentially improve quantitative cartilage health assessment tools already applied in knee OA research [35] since it allows for precise diagnosis as well as monitoring of new treatment modalities. To highlight clinical applicability of accurate, automatic segmentation of cartilage, the following section reviews a method for fusion of cartilage segmentation and motion capture gait analysis.

It is generally accepted that knee OA is biomechanically driven and caused by abnormal mechanical loading on the cartilage tissue in a context of systemic susceptibility [36]. Knee joint loads during walking are of particular interest since walking is the most natural way of human locomotion and causes repetitive joint loadings. The knee joint loads during walking have been proven to discriminate between healthy and knee OA patients [37], knee OA disease severity levels [37] and predict disease progression [38].

However, the current knowledge about the pathological mechanics of knee

OA is based on generic models of joint loads [39]. Modelling patient-specific surface pressure distribution on the joint surface during walking could prove to be effective in identifying patient-specific OA risk factors. This approach have previously been used during static conditions (standing) [40], but the articular joint contact stress during movements would prove even more important, because of the established discriminative and predictive of joint loads joint loadings during walking. Early identification of abnormal pressure distribution may allow for preventive treatment which may decrease human and socioeconomic costs of treatments and joint replacements.



(a) Patient seen from 2 o'clock. (b) Patient seen from 6 o'clock

Figure 10: Motion capture gait analysis using a Vicon MX system at The Parker Institute, Frederiksberg Hospital, Copenhagen, Denmark. Left leg is in red and right leg is in green. The red arrows indicate direction and magnitude of force applied to pressure pads.

A motion capture system with force platforms will record movement and ground reaction forces, from which joint moments of force are calculated using an inverse dynamics approach and a statically determinate model of joint forces resolved in an arbitrarily located joint center. In a patient-specific model, an array of linear springs distributed across the joint surface would model the cartilage with geometry derived from MRI segmentation using the proposed method. Discrete Element Analysis (DEA) would compute the pressure distribution throughout the gait cycle and contact stresses throughout the joint surfaces can be determined from spring deformation during loading.

Such a patient-specific model of joint contact stresses can in cross sectional studies be related to the health of the joint (cartilage, bone etc. from MRI), symptoms (patient reported), and function (muscle strength, walking performance etc). In prospective studies, baseline joint surface stresses may predict disease progression more effectively than the current available methods, with the ultimate goal being clinical use.

6 Conclusion

In this thesis, a method for semi-automatic segmentation of knee articular cartilage has been presented. The results are obtained from a large cohort of pathological knees and ranges from comparable to better than previous automated methods. With future work seeking to improve on the presented results, the proposed method holds great potential for clinical applicability.

References

[1] Jackson, D., Simon, T., Aberman, H.: Symptomatic articular cartilage degeneration: The impact of the new millenium. Clinical Orthopaedics And Related Research **391** (2001) 14–25

[2] Woolf, A.D., Pfleger, B.: Burden of major musculoskeletal conditions. Bulletin of the World Health Organization **81** (2003) 646 – 656

[3] Bagge, E., Bjelle, A., Edén, S., Svanborg, A.: Osteoarthritis in the elderly: clinical and radiological findings in 79 and 85 year olds. Ann Rheum Dis **50** (1991) 535–9

[4] Buckwalter, J.A., Mankin, H.J., Grodzinsky, A.J.: Articular cartilage and osteoarthritis. Instr Course Lect **54** (2005) 465–80

[5] Lorenzo, P., Bayliss, M.T., Heinegård, D.: Altered patterns and synthesis of extracellular matrix macromolecules in early osteoarthritis. Matrix Biol **23** (2004) 381–91

[6] Felson, D.T.: An update on the pathogenesis and epidemiology of osteoarthritis. Radiol Clin North Am **42** (2004) 1–9, v

[7] Buckwalter, J.A.: Articular cartilage: injuries and potential for healing. J Orthop Sports Phys Ther **28** (1998) 192–202

[8] Bliddal, H., Christensen, R.: The treatment and prevention of knee osteoarthritis: a tool for clinical decision-making. Expert Opin Pharmacother **10** (2009) 1793–804

[9] Dougados, M.: How can one develop disease-modifying drugs in osteoarthritis? Curr Rheumatol Rep **7** (2005) 22–8

[10] Dam, E., Loog, M., Christiansen, C., Byrjalsen, I., Folkesson, J., Nielsen, M., Qazi, A., Pettersen, P., Garnero, P., Karsdal, M.: Identification of progressors in osteoarthritis by combining biochemical and mri-based markers. Arthritis Research And Therapy **11** (2009) R115 See related editorial by Williams, http://arthritis-research.com/content/11/5/130.

[11] Tummala, S., Dam, E.: Surface smoothness: cartilage biomarkers for knee oa beyond the radiologist. Proceeding SPIE **7623** (2010)

[12] Qazi, A.A., Folkesson, J., Pettersen, P., Christiansen, C., Dam, E.B., Karsdal, M.A.: Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. Osteoarthritis and Cartilage (2007)

[13] Cohen, Z.A., McCarthy, D.M., Kwak, S.D., Legrand, P., Fogarasi, F., Ciaccio, E.J., Ateshian, G.A.: Knee cartilage topography, thickness, and contact areas from mri: in-vitro calibration and in-vivo measurements. Osteoarthritis and Cartilage **7** (1999) 95 – 109

[14] Hohe, J., Ateshian, G., Rejser, M., Englmeier, K.H., Eckstein, F.: Surface size, curvature analysis, and assessment of knee joint incongruity with mri in vivo. Magnetic Resonance in Medicine **47** (2002) 554–561

[15] Graichen, H., von Eisenhart-Rothe, R., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging: technical validation for use in analysis of cartilage volume and further morphologic parameters. Arthritis Rheum **50** (2004) 811–6

[16] Pessis, E., Drape, J.L., Ravaud, P., Chevrot, A., Ayral, M.D.X.: Assessment of progression in knee osteoarthritis: Results of a 1 year stud comparing arthroscopy and mri. Osteoarthritis and Cartilage **11** (2003) 161–369

[17] Lang, P., Noorbakhsh, F., Yoshioka, H.: Mr imaging of articular cartilage: Current state and recent developments. Radiologic Clinics of North America **43** (2005) 629 – 639

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

[19] Eckstein, F., Schnier, M., Haubner, M., Pribsh, J., Glaser, C., Englmeier, K.H., Reiser, M.: Accuracy of cartilage volume and thickness measurements with magnetic resonance imaging. Clinical Orthopaedics And Related Research **352** (1998) 137–148

[20] Stammberger, T., Eckstein, F., Englmeier, K.H., Reiser, M.: Determination of 3d cartilage thickness data from mr imaging: Computational method and reproducibility in the living determination of 3d cartilage thickness data from mr imaging: Computational method and reproducibility in the living. Magnetic Resonance in Medicine **41** (1999) 529–536

[21] Kornaat, P., Reeder, S., Koo, S., Brittain, J., Yu, H., Andriacchi, T., Gold, G.: Mr imaging of articular cartilage at 1.5 t and 3.0 t: Comparison of spgr and ssfp sequences,. Osteoarthritis and Cartilage **13** (2005) 338 – 344

[22] Gougoutas, A.J., Wheaton, A.J., Borthakur, A., Shapiro, E.M., Kneeland, J.B., Udupa, J.K., Reddy, R.: Cartilage volume quantification via live wire segmentation. Academic radiology **11** (2004) 1389–1395

[23] Jaremko, J.L., Cheng, R.W.T., Lambert, R.G.W., Habib, A.F., Ronsky, J.L.: Reliability of an efficient mri-based method for estimation of knee cartilage volume using surface registration. Osteoarthritis Cartilage **14** (2006) 914–22

[24] Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P.C., Christiansen, C.: Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans Med Imaging **26** (2007) 106–15

[25] Grau, V., Mewes, A.J.U., Raya, M.A., Kikinis, R., Warfield, S.K.: Improved watershed transform for medical image segmentation using prior information. IEEE Transactions on Medical Imaging **23** (2004) 447–458

[26] Riecke, B., Christensen, R., Christensen, P., Leeds, A., Boesen, M., Lohmander, L., Astrup, A., Bliddal, H.: Comparing two low-energy diets for the treatment of knee osteoarthritis symptoms in obese patients: a pragmatic randomized clinical trial. Osteoarthritis and Cartilage **18** (2010) 746 – 754

[27] KELLGREN, J.H., LAWRENCE, J.S.: Radiological assessment of osteoarthrosis. Ann Rheum Dis **16** (1957) 494–502

[28] Dunn, T.C., Lu, Y., Jin, H., Ries, M.D., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. Radiology **232** (2004) 592–8

[29] Image courtesy of Wikimedia Commons: http://en.wikipedia.org/wiki/file:knnclassification.svg (2010)

[30] Freeman: Adult Articular Cartilage. Pitman Medical Publishing, London (1979)

[31] Ding, C., He, X.: K-means clustering via principal component analysis. Proceedings of the International Conference of Machine Learning (2004) 225–232

[32] Ding, M., Odgaard, A., Hvid, I.: Changes in the three-dimensional microstructure of human tibial cancellous bone in early osteoarthritis. J Bone Joint Surg Br **85** (2003) 906–12

[33] Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research **10** (2009) 207–244

[34] Carpenter, A.: cusvm: A cuda implementation of support vector classification and regression. http://patternsonascreen.net/cuSVM.html (2010)

[35] Peterfy, C.G., Schneider, E., Nevitt, M.: The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthritis Cartilage **16** (2008) 1433–41

[36] Hunter, D.J., Felson, D.T.: Osteoarthritis. BMJ **332** (2006) 639–42

[37] Henriksen, M., Graven-Nielsen, T., Aaboe, J., Andriacchi, T.P., Bliddal, H.: Gait changes in patients with knee osteoarthritis are replicated by experimental knee pain. Arthritis Care Res (Hoboken) **62** (2010) 501–9

[38] Miyazaki, T., Wada, M., Kawahara, H., Sato, M., Baba, H., Shimada, S.: Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. Ann Rheum Dis **61** (2002) 617–22

[39] Schipplein, O.D., Andriacchi, T.P.: Interaction between active and passive knee stabilizers during level walking. J Orthop Res **9** (1991) 113–9

[40] Segal, N.A., Anderson, D.D., Iyer, K.S., Baker, J., Torner, J.C., Lynch, J.A., Felson, D.T., Lewis, C.E., Brown, T.D.: Baseline articular contact stress levels predict incident symptomatic knee osteoarthritis development in the most cohort. J Orthop Res **27** (2009) 1562–8