

9TH MICCAI CONFERENCE



JOINT DISEASE WORKSHOP

Quantitative Automated Musculoskeletal Analysis

OCTOBER 5TH · 2006 · COPENHAGEN · DENMARK

EDITORS:

Erik Dam, Sharmila Majumdar & J Christopher Buckland-Wright

MICCAI 2006 WORKSHOP
P R O C E E D I N G S

www.itu.dk/image/JOINT



MICCAI

MICCAI 2006, Copenhagen, Denmark
Workshop Proceedings

Joint Disease Workshop – *Quantitative Automated Musculoskeletal Analysis*

Erik B. Dam, Sharmila Majumdar and J. Christopher Buckland-Wright

© All rights reserved by the authors

ISBN 10: 87-7611-151-2
ISBN 13: 978-87-7611-151-9

Printed by Samfundslitteratur Grafik

Preface

The idea for the MICCAI Joint Disease Workshop came from the desire to bridge the technically oriented community of MICCAI with the clinically oriented researchers working within the field of joint disease. Even though diseases such as osteoarthritis and osteoporosis have a huge impact on our society — both in terms of the socio-economy and reduced quality of life — they have previously been only sparsely addressed at MICCAI. The potential for fruitful collaboration is even bigger given the more and more advanced image analysis methodology used in recent years in clinically oriented research focusing on automation and advanced morphological/structural analysis of the relevant anatomical structures.

The submissions for the workshop has an impressive range of topics including development of MRI pulse sequences, gait analysis, shape models and morphological analysis, and optimized segmentation algorithms — with applications in rheumatoid arthritis, osteoarthritis, and osteoporosis. These are all topics that naturally belong within the MICCAI community. The future will tell whether the focused topic of joint disease will earn more attention at future MICCAI conferences — either in a workshop setting or at the main conference.

This workshop would not have been possible without the enthusiasm of the co-chairs Sharmila Majumdar (University of California, San Francisco, Department of Radiology) and Christopher Buckland-Wright (King's College London, Division of Applied Biomedical Research). Also, thanks are due to Arish Qazi (IT University of Copenhagen) for organizing submissions and reviews, and to Felix Eckstein (Paracelsus Private Medical University, Salzburg, Institute of Anatomy and Musculoskeletal Research) for accepting to come to Copenhagen and present some of the world-leading research that he is involved in.

Last but not least, sincere thanks are due to the program committee members for making the review process pleasant and fruitful:

Scott Acton, University of Virginia
Monica Barbu-McInnis, VirtualScopics
Marleen de Bruijne, IT University of Copenhagen
Graeme Jones, Menzies Research Institute, Tasmania,
Jenny Folkesson, IT University of Copenhagen
Bram van Ginneken, Image Sciences Institute, Utrecht
Garry Gold, Stanford University
Michael Kaus, Philips Research Laboratories
Nancy Lane, University of California, Davis
Philipp Lang, Harvard Medical School
Thomas Link, University of California, San Francisco
Stefan Lohmander, Lund University Hospital
Anne Karien Marijnissen, Utrecht University Hospital
Steven Millington, Medical University of Vienna
Sébastien Ourselin, BioMedIA Lab, CSIRO, Sidney
Paola Pettersen, Center for Clinical and Basic Research, Copenhagen
Julia Schnabel, University College London
Koen Vincken, Image Sciences Institute, Utrecht
Simon Warfield, Brigham & Women's Hospital
John Waterton, AstraZeneca
Tomos Williams, University of Manchester

Copenhagen, August 2006

Erik Dam
General Chair

Contents

Oral Presentations

Improving the Segmentation Accuracy of Fractured Vertebrae with Dynamically Sequenced Active Appearance Models.....	1
---	---

Martin Roberts, Tim Cootes, and Judith Adams

Position Normalization in Automatic Cartilage Segmentation.....	9
---	---

Jenny Folkesson, Erik Dam, Ole Olsen, Paola Pettersen, and Claus Christiansen

Automatic Curvature Analysis of the Articular Cartilage Surface.....	17
--	----

Jenny Folkesson, Erik Dam, Ole Olsen, Paola Pettersen, and Claus Christiansen

Anatomically Equivalent Focal Regions Defined on the Bone Increases Precision when Measuring Cartilage Thickness from Knee MRI	25
--	----

Tomos Williams, Andrew Holmes, John Waterton, Rose Maciewicz, Anthony Nash, and Chris Taylor

Automatic Detection of Erosions in Rheumatoid Arthritis.....	33
--	----

Georg Langs, Philipp Peloschek, Horst Bischof, and Franz Kainberger

Quantitative vertebral morphometry using neighbor-conditional shape models	41
--	----

Marleen de Bruijne, Michael T. Lund, Paola Pettersen, Laszlo B. Tanko, and Mads Nielsen

Automatic Cartilage Thickness Quantification using a Statistical Shape Model.....	42
---	----

Erik Dam, Jenny Folkesson, Paola Pettersen, and Claus Christiansen

Poster Presentations

MR image segmentation using phase information and a novel multiscale scheme	50
<i>Pierrick Bourgeat, Jurgen Fripp, Peter Stanwell, Saadallah Ramadan, and Sebastien Ourselin</i>	
Automatic Quantification of Cartilage Homogeneity	51
<i>Arish Asif Qazi, Ole Fogh Olsen, Erik Dam, Jenny Folkesson, Paola Pettersen, and Claus Christiansen</i>	
Knee Images Digital Analysis: a quantitative method for individual radiographic features of knee osteoarthritis.	59
<i>Anne Marijnissen, Koen Vincken, Petra Vos, Johannes Bijlsma, Wilbert Bartels, and Floris Lafeber</i>	
Improved Parameter Extraction From Dynamic Contrast-Enhanced MRI Data in RA Studies	64
<i>Olga Kubassova, Roger Boyle, and Alexandra Radjenovic</i>	
Novel Method for Quantitative Evaluation of Segmentation Outputs for Dynamic Contrast-Enhanced MRI Data in RA Studies	72
<i>Olga Kubassova, Roger Boyle, and Alexandra Radjenovic</i>	
3D Shape Description of the Bicipital Groove: Correlation to Pathology	80
<i>Aaron Ward, Ghassan Hamarneh, and Mark Schweitzer</i>	
Efficient Automatic Cartilage Segmentation	88
<i>Erik Dam, Jenny Folkesson, Marco Loog, Paola Pettersen, and Claus Christiansen</i>	
3D Shape Analysis of the Supraspinatus Muscle	96
<i>Aaron Ward, Ghassan Hamarneh, Reem Ashry, and Mark Schweitzer</i>	
Three dimensional dynamic model with different tibial plateau shapes: Analyzing tibio-femoral movement	104
<i>Ekin Akalan, Mehmet Ozkan, and Yener Temelli</i>	
Author Index	111

Improving the Segmentation Accuracy of Fractured Vertebrae with Dynamically Sequenced Active Appearance Models

M.G. Roberts, T.F. Cootes, J.E. Adams

Department of Imaging Science and Biomedical Engineering
University of Manchester, Manchester, M13 9PL, UK
`martin.roberts@manchester.ac.uk`

Abstract. The accurate identification of vertebral fractures is important in diagnosing the early stages of osteoporosis. Good mean segmentation accuracy has been previously obtained for vertebrae on lateral dual x-ray absorptiometry (DXA) scans, by modelling the spine as a sequence of overlapping triplets of vertebrae. However the Active Appearance Models (AAM) used for each triplet sub-model were prone to occasionally converge to incorrect local minima, mainly in the case of severe (grade 3) fractures. The robustness of the AAM search sequence has been improved by two means: 1) ensuring a better starting solution by using information on the approximate centre of each vertebra; 2) including an alternative starting solution at each vertebral level to represent a possible severe fracture. The best of a set of normal and fractured alternative sub-model solutions was selected at each iteration. This provides a robust and accurate search method, even with multiple severe fractures present. A mean segmentation accuracy of 0.7mm was achieved on normal vertebrae, rising to 1.2mm on severely fractured vertebrae.

1 Introduction

Osteoporosis is a progressive skeletal disease characterized by a reduction in bone mass, resulting in an increased risk of fractures. Vertebral fractures are the most common, and occur in younger patients. The presence of vertebral fractures significantly increases the risk of further vertebral and non-vertebral fractures [5]. The accurate identification of prevalent vertebral fractures is therefore clinically important. However there is no precise definition of exactly what constitutes a vertebral fracture, though a variety of methods of describing such fractures have been developed [1]. These include semi-quantitative methods, which involve some subjective judgement by an expert radiologist and 3 grades of fracture based on the percentage of height reduction; and fully quantitative morphometric methods. The latter require the manual annotation of six (or more) points on each vertebra. The manual marking is time consuming, and subtle shape information is lost in the reduction of shape to 6 points.

Our ultimate aim is to define more reliable quantitative fracture classification methods based on a complete definition of the vertebra's shape, and surrounding

texture. The first step must therefore be to achieve a reliable automatic segmentation. Some success in automatically locating vertebrae has been reported by several authors [6, 7, 4] using statistical models of shape and appearance; and by [2] using shape particle filtering.

We have previously used an Active Appearance Model (AAM) [10] approach to segment vertebrae starting from an approximate manual initialisation on the top, bottom and middle of the spine [8]. Although this approach achieved good accuracy on normal vertebrae, and most mild to moderate fractures, there were problems in reliably locating severe (grade 3) fractures. This was not simply because of inadequacies in the shape model training, as selecting random train/test set splits and refitting the shape model to the annotated data in the test set indicated that the shape model could fit to within 1mm of the manual segmentation in up to 90% of points in severe fractures. Many of the problem cases appeared to be caused by local minima. A potential failure mode with severe fractures was for the AAM search to fit the top of the fractured vertebra to the bottom of the neighbouring vertebra above (or vice versa). This tends to occur if the edge of the severely fractured vertebra is collapsed too far below the starting solution. See for example Figure 2.

In this paper we address this problem by using a combination of a better initialisation method, and by using an alternative starting solution corresponding to a severe fracture.

2 Materials and Methods

2.1 Data - DXA Images

Assessment for vertebral fracture is traditionally carried out using spinal radiographs. This study used dual x-ray absorptiometry (DXA) images however. Despite the images being noisier and of lower resolution, DXA has several advantages, such as a substantially lower radiation dose, and a lack of projective effects. Figure 1a shows a typical DXA scan with some endplate fractures, and Figure 1b shows the solution superimposed. The models cover the lumbar starting at L4 and continuing up the thorax up to vertebra T7. The dataset used comprised 240 DXA images containing 215 fractures. The dataset used was as for [8], together with an additional 38 images recently acquired from patients attending for bone mineral density assessment. These additional images were selected because they contained fractured vertebrae useful in training the shape model. The newer 38 images were manually annotated using an in-house tool, by an experienced radiographer (SC), with supervision by the first author. Each vertebral contour uses between 32 to 40 points around the vertebral body (32 points for T10 and above), with 8 further points around the pedicles for L4 to T10.

2.2 Summary of previous work in AAM segmentation of vertebrae

Many problems in medical image interpretation require an automated segmentation, but the images may provide noisy data, and typically complex structure.

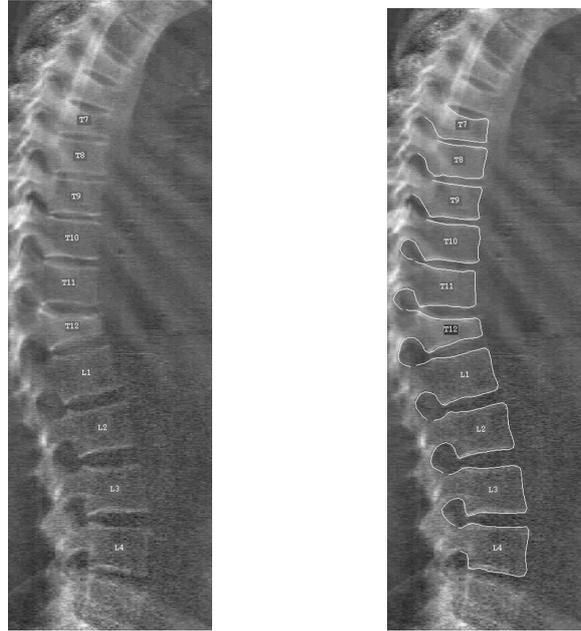


Fig. 1. Lateral DXA image of a spine displaying some early symptoms of osteoporosis (e.g. T12 fracture). a) shows the raw image; b) shows the model solution superimposed

Model based methods offer solutions to these difficulties, by enforcing strong priors learned from a set of annotated training images. A widespread such current approach is the Active Appearance Model (AAM) [10]. However under-training of the model may mean that it is insufficiently adaptable on a local level, especially when pathologies are present. In previous work [7] we showed that this problem could at least be mitigated by using multiple sub-models. The sub-structures were linked by partially overlapping them, and fitting them in a sequence. The constrained form of the AAM [9] was used, with constraints associated with the overlapping points. An extension to this approach was devised [8], which uses a “best-fit-first” heuristic to obtain a good ordering of the sub-model fitting sequence. This approach defers poorly fitting or noisy regions until they have been constrained by their neighbours, and improved the accuracy and robustness of the fit [8]. The sub-model combination algorithm also uses a global shape model of the entire spine, but this is used to guide the initialisation of later sub-models in the fitting sequence (given the earlier sub-model solutions), and does not absolutely constrain the solution.

2.3 AAM Initialisation

As AAMs are local search methods it is necessary to initialise the shape models approximately around the vertebral locations. Our original initialisation method [8] used a 3 point manual initialisation, wherein the user clicked on the bottom of L4, top of T12, and top of T7. However analysis of the results indicated that

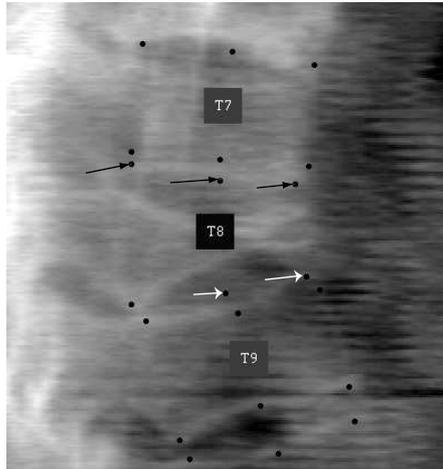


Fig. 2. Edge misfit for grade 3 fracture. The fitted standard 6-point morphometric corner and mid-points are shown. The arrows indicate where the AAM search fitted the top of T8 to the bottom of T7, and the anterior bottom corner of T8 to the top of T9. T7 has been displaced as a consequence

there were a significant number of convergence failures for severe fractures, due to edge confusions with neighbouring vertebrae (in around 25% of cases). We hypothesised that some of these failures might be avoided by allowing more information in the initialisation. Furthermore it was observed that the user would have to count up the vertebrae one at once, to locate the required three vertebrae to click on, and so it would be as easy for the user to simply click on the (approximate) centre of each vertebrae in turn. This would give a 10-point initialisation, which might provide a better start point for intermediate vertebrae, especially where pathologies have displaced a vertebra from its expected position.

For both 10-point and 3-point initialisation methods the shape is initialised by calculating the global shape model's best fit to the input set of positions (in a least squares sense). However conditions such as scoliosis or high kyphosis can cause the global model fit to be poor at particular vertebrae. So with the 10-point method there is a correction phase in which individual vertebrae are first rigidly translated to be centred on their input positions. But this in turn can occasionally introduce inconsistency, by causing vertebrae to overlap, so finally any overlapping vertebrae are locally reduced in height until they are separated.

2.4 Multiple Sub-Model Selection

The dynamic model sequencing method of [8] operates as follows. A set of N candidate sub-models (i.e. different triplets) is maintained. Each candidate is provisionally fitted, then the submodel with the best quality of fit (see below) is imposed into the global solution, and then removed from the candidate list. If possible a new candidate is added to replace the one just imposed. The nearest remaining neighbour (if any) is used as the new candidate sub-model. These

iterations continue until all sub-models have been fitted and had their solutions imposed, so the candidate list is empty. We used an initial set of 5 candidate sub-models, starting with the triplets centred on T8,T10,T12,L2 and L3. The candidate set size of 5 is a compromise with speed constraints, as each candidate is refitted at every iteration.

We extended this method by adding an alternative starting solution for each sub-model in the candidate list. At each iteration an alternative fractured variant is generated for the central vertebra in each candidate triplet sub-model. Each of these $2N$ AAM searches is tentatively run in turn, using a constrained AAM so that any previously determined points (with therefore high constraint weights) are forced to be close to their previous values. The best fitting of all the searches is selected as the solution to impose at this iteration, and its submodel is removed from the list of current candidates. The points of the central vertebra of the triplet are saved in an overall solution vector, as are points in either neighbour which have not been previously determined. The central points now have high constraint weights attached, to ensure consistency for subsequent searches on neighbouring triplets. If the best solution obtained was from a fractured starting solution, and the shape is indicative of a fracture according to standard morphometric height reduction criteria [3], then the AAM searches for the two neighbouring vertebrae are also re-run if they had already been imposed. This is because the neighbour's triplet search might have been previously thrown out by the fractured vertebra being erroneously fitted to the neighbour (e.g. see Fig 2). Note that the point constraints on the neighbour are relaxed before re-running its AAM search. Finally a new candidate sub-model is now added as before, and this completes an iteration of the fitting sequence.

The alternative fractured initial solution is generated thus. Firstly the posterior height H_p of the triplet's central vertebra is reduced by 15%, to allow for a potential misfit onto both neighbours in the worst case. Next its mid-height is reduced to $0.6H_p$, and its anterior height to $0.75H_p$, as most grade 3 endplate fractures also display signs of wedging. This fixes the corners and mid-points in the central vertebra (i.e. the standard 6 morphometric points). The corners and mid-points in neighbouring vertebrae are temporarily fixed at their current values, and the alternative solution (for the central vertebra of the triplet) is then initialised to the best triplet shape model fit to these (18) fixed points.

In comparing solutions for different triplets, or alternative solutions of the same triplet, it is necessary to have a comparison criterion. An obvious candidate to use is essentially the residual sum of squares of the fit, S say. In [8] we mapped S onto its associated cdf, as when comparing different sub-models with different numbers of points it is not meaningful to directly compare values of the sum of squares. So S is converted to a value on a standard Gaussian approximating the associated χ^2 distribution, but with further empirical scaling parameters for its overall mean and variance. See [8] for details.

2.5 Experiments

Miss-8-out tests were performed over the 240 images. On each experiment the user initialisation was simulated by using the known equivalent marked points

and adding random offsets to them. For 3-point initialisation these were zero-mean Gaussian errors with SD of 1mm in the y-direction (along the spine) and 3mm in the x-direction. For 10-point initialisation the SD of the error in the y-direction was increased to 2mm, as it is clearly less precise to judge the centre than an edge. Ten replications (i.e. random initialisations) of each image were performed.

3 Results

The accuracy of the search was characterised by calculating the absolute point-to-line distance error for each point on the vertebral body. The error is the distance from each point on the located vertebra contour to the nearest point (in the same vertebra) on the “true” contour. The true contour is the smooth bezier spline passing through the manually annotated points.

Table 1 compares accuracy results for 3-point initialisation, 10-point initialisation, and 10-point initialisation with the alternative solution. The results are decomposed into points within normal or the three grades of fractured vertebrae. Data are given for the mean, the percentage of point errors in excess of 2mm, and the percentage of vertebral edges erroneously fitted to the converse edge of a neighbour (i.e. as in Fig 2). The threshold of 2mm would be around 2.5SDs of manual precision, and can be viewed as a point failure indicator. The results are for points in vertebrae L3 up to T8, and the extreme vertebrae (T7 and L4) have been excluded from the results. This is for several reasons. Firstly these vertebrae have explicit initialisation information in the 3-point initialisation method, and so the change in initialisation should have less effect on their results. Secondly the extreme vertebrae are special cases in the triplet fitting algorithm, in that they do not have their own central triplet, and are fitted as part of the T8 or L3 triplet respectively. The minor complication in the algorithm to generate alternatives for the extrema has not been made at this provisional stage. Thirdly, although the shape models are generally adequate even up to 90% of the grade 3 fractures, L4 specifically does suffer from an undertraining problem in our dataset as fractures of L4 have a very low prevalence.

The change in initialisation method improves the accuracy on fractured vertebrae generally. For example the grade 2 mean error reduces from 1.37mm to 0.93mm, and the point failure rate is reduced from 19% to 10%. The effect of including the alternative fractured solution is largely confined to grade 3 fractures, as might be expected. Use of a bootstrap resampling method on the differences due to the use of secondary fractured alternative gives a 98% confidence interval of [0.12,0.55] for the improvement in mean error on grade 3 fracture points. As this interval does not span zero, a corollary is that the improvement is statistically significant at the 2% level. Overall the combination of 10-point initialisation and an alternative starting solution largely eliminates the tendency for edge confusions with neighbouring vertebrae. The point failure rate for grade 3 fractures is reduced from 25% to 14%, and the mean error is reduced from 1.97mm to 1.17mm.

	3-point Initialisation			10-point Initialisation			10-point Initialisation with alternative		
Fracture Grade	Mean Acc	%ge pts failing	%ge edges misfit	Mean Acc	%ge pts failing	%ge edges misfit	Mean Acc	%ge pts failing	%ge edges misfit
Normal	0.74	4.6%	4.2%	0.70	3.8%	2.1%	0.69	3.7%	1.7%
Grade 1	1.03	12.0%	9.5%	0.88	8.6%	6.1%	0.85	7.3%	5.9%
Grade 2	1.37	18.8%	10.0%	0.93	10.1%	6.0%	0.93	9.5%	4.4%
Grade 3	1.97	28.7%	24.7%	1.48	20.5%	14.9%	1.17	13.7%	8.7%
All Fractures	1.38	18.5%	13.7%	1.08	12.5%	8.5%	0.96	9.9%	6.2%

Table 1. Search Accuracy Results by Fracture Status for the 3 initialisation strategies. NB Fracture grades 1-3 correspond to height reduction thresholds of 20%, 25% and 40% respectively. The percentage of points failing columns give the number of points for which the error exceeded 2mm

4 Discussion

Fractured vertebrae present a more challenging problem for an approach using statistical models, as the variation is much greater for pathological cases than for normals. Not only is there more variation in shape but more local minima exist. Patients with multiple fractures typically also have low BMD and this tends to lead to a particularly poor SNR on DXA. Nevertheless with 10-point initialisation and the use of an alternative fractured starting solution, over 90% of points in fractured vertebrae from L3 to T8 are located to within 2mm with the dynamically sequenced multi-AAM method. Even severe fractures are located with a mean accuracy close to manual precision, and with good reliability (c. 86% of points).

Obviously there are more general methods of searching using multiple candidates, typically using some form of stochastic search. For example de Bruijne et al [2] used shape particle filtering to segment vertebrae on lumbar radiographs. However our results are significantly more accurate, despite the higher noise with DXA, and the fact that thoracic vertebrae have a higher fracture prevalence, and suffer from more soft tissue clutter. On the other hand we rely on an approximate manual initialisation. Given this initialisation our results indicate that it is not necessary to have large numbers of randomised candidates for vertebrae. The modest number of alternatives that appear to be required can be handled within an AAM framework. In fact we also experimented with a third alternative representing a moderate fracture, but this produced no further improvement. Clearly various hybrid approaches can be envisaged, such as running some form of stochastic search (e.g. shape particle filtering) to obtain plausible starting solutions, followed by AAM for finer accuracy on the best few candidate solutions. Fully automatic searches such as [2] might be combined with subsequent AAM search in pharmaceutical trials or large epidemiological studies where there are thousands of images and manual initialisation is therefore less desirable.

4.1 Conclusions

Using a solution initialised on the centre of each vertebra, rather than a 3-point initialisation, significantly improves accuracy on fractured vertebrae, by reducing local minima caused by edge confusion. Including an alternative severe fracture as a secondary starting solution also improves accuracy and robustness with severe fractures. The overall accuracy of our results is now comparable to manual precision, even on fractured vertebrae, in all but a handful of extreme cases of severe pathologies or very poor images. Although we anticipate further small improvements for pathological cases as the training set is extended, our training set and algorithms have now reached the point of being fit for supervised clinical use. Automatic determination of the detailed vertebral outline could enable the development of more powerful quantitative classifiers, that can capture some of the more subtle aspects of visual expert reading of vertebral fractures, such as the characteristic collapse of the endplate.

References

1. Guermazi A, Mohr A, Grigorian M, Taouli B, and Genant HK. Identification of vertebral fractures in osteoporosis. *Seminars in Musculoskeletal Radiology*, 6(3):241–252, 2002.
2. de Bruijne M and Nielsen M. Image segmentation by shape particle filtering. In *International Conference on Pattern Recognition*, pages 722–725. IEEE Computer Society Press, 2004.
3. R Eastell, SL Cedel, HW Wahner, BL Riggs, and LJ Melton. Classification of vertebral fractures. *J Bone Miner Res*, 6(3):207–215, 1991.
4. B. Howe, A. Gururajan, H. Sari-Sarraf, and R. Long. Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized hough transform and extensions to active appearance models. In *Proc IEEE 6th SSIAI*, pages 182–186, 2004.
5. Melton LJ III, Atkinson EJ, Cooper C, O’Fallon WM, and Riggs BL. Vertebral fractures predict subsequent fractures. *Osteoporosis Int*, 10:214–221, 1999.
6. Smyth PP, Taylor CJ, and Adams JE. Vertebral shape: automatic measurement with active shape models. *Radiology*, 211:571–578, 1999.
7. M.G. Roberts, T.F. Cootes, and J.E. Adams. Linking sequences of active appearance sub-models via constraints: an application in automated vertebral morphometry. In *14th British Machine Vision Conference*, pages 349–358, 2003.
8. M.G. Roberts, T.F. Cootes, and J.E. Adams. Vertebral shape: Automatic measurement with dynamically sequenced active appearance models. In *8th MICCAI Conference*, volume 2, pages 733–740, 2005.
9. Cootes TF and Taylor CJ. Constrained active appearance models. In *8th International Conference on Computer Vision*, volume 1, pages 748–754. IEEE Computer Society Press, July 2001.
10. Cootes TF, Edwards GJ, and Taylor CJ. Active appearance models. In Burkhardt H and Neumann B, editors, *5th European Conference on Computer Vision*, volume 2, pages 484–498. Springer (Berlin), 1998.

Position Normalization in Automatic Cartilage Segmentation

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

¹ Image Analysis Group, IT University of Copenhagen, Denmark, jenny@itu.dk,
² Center for Clinical and Basic Research, Ballerup, Denmark.

Abstract. In clinical studies of osteoarthritis using magnetic resonance imaging, the placement of the test subject in the scanner tends to vary and this can affect the outcome of automatic image analysis methods for articular cartilage assessment, particularly in multi-center studies. We have developed an automatic iterative method that corrects for position variations by combining the two steps: shifting the cartilage towards the expected position and performing a voxel classification with the normalized position as a feature. By applying this placement adjustment scheme to an automatic knee cartilage segmentation method we show that the inter-scan reproducibility is much improved and is now as good as that of a highly trained radiologist.

1 Introduction

Osteoarthritis (OA) is one of the major health issues among the elderly population [1]. One of the main effects of OA is the degradation of the articular cartilage, causing pain and loss of mobility of the joints.

Magnetic resonance imaging (MRI) is the only imaging modality for direct, non-invasive segmentation of the articular cartilage [2] where cartilage deterioration can be detected [3]. Among MRI sequences, the most established are fat-suppressed gradient-echo T1 sequence using a 1.5T or a 3T magnet which yields high image quality. Low-field dedicated extremity MRI produces images with lower quality but to a very low cost. They can provide similar information on bone erosions and synovitis as expensive high-field MRI units [4], and if a low-field scanner can be used for articular cartilage assessment as well, costs for making clinical studies would be reduced significantly.

In quantitative assessment of articular cartilage using MRI, the most crucial step is the segmentation. The cartilage can be manually segmented slice-by-slice by experts, but for routine clinical use manual methods are too time consuming and they are prone to inter- and intra-observer variability. In order to overcome these problems much effort has been put into development of semi- or fully automatic segmentation methods, both in 2D [5],[6] and directly in 3D [7],[8],[9]. When assessing the cartilage directly in 3D the problem of limited continuation between slices that is present in 2D techniques is eliminated.

An uncommitted segmentation scheme is often desired to achieve invariant performance under irrelevant transformations of the images such as translation, rotation, resolution of and orientation and position in the measuring device. A mathematical well-founded and operational approach to this is the scale space methodologies as is also applied in the automatic segmentation framework of Folkesson et al. based on approximate k NN classification [9]. In segmentation by classification it is also well recognized that the geometrical covariance between feature points on a space of features derived mainly from Gaussian scale space derivatives can support the task significantly [10]. Similar we would like to introduce position variance relative to the training data in a normalized way. Images and in particular medical scans are obviously not acquired with random relation between objects of interest and the field of view. This manifests itself in the application [9] by having non-normalized position as a very strong feature for classification between cartilage and background when compared to invariant geometric scale space features. On the other hand patients will be placed slightly different depending on clinical staff. This can be corrected using an iterative scheme for position normalization and at the same keep the relative position as strong feature. In this paper, we introduce an iterative position normalization scheme for automatic segmentation tasks and our evaluation shows that it increases the robustness of an automatic segmentation method significantly.

2 Methods

2.1 Acquisition and Population

The test subjects are between 22-79 years old with an average age of 56 years, 59% females, and there are both healthy and osteoarthritic knees according to the Kellgren-Lawrence index (KLi) [11], a radiographic score from 0-4 where $KLi = 0$ is healthy, $KLi = 1$ is considered borderline or mild OA, and $KLi \geq 2$ is severe OA. We examine 25-114 knees, 25 for training of the segmentation method and 114 for evaluation. Of the 114, 31 knees have been re-scanned and the reproducibility is evaluated by comparing the first and second scanning. In the test set there are 51, 28, 13 and 22 scans that have $KLi = 0, 1, 2$ and 3 respectively.

MRI is performed with an Esaote C-Span low-field 0.18T scanner dedicated to imaging of extremities yielding a Turbo 3D T1 sequence (40° flip angle, T_R 50 ms, T_E 16 ms). Approximate acquisition time is 10 minutes and the scan size, after automatically removing background that contain no information, is $104 \times 170 \times 170$ voxels. The spatial resolution of the scans is approximately $0.8 \times 0.7 \times 0.7mm^3$.

31 knees were re-scanned after approximately one week in order to examine segmentation precision. All the scans have been manually delineated by a radiologist in order to establish the accuracy of the automatic method and the same 31 scans were delineated twice with the purpose of examining the intra-rater variability of the manual delineations. An example of how a MRI slice and the manual delineation looks like can be seen in the top rows of Figures 2 and 3.

2.2 The Cartilage Segmentation Method

The cartilage segmentation method we examine is fully automatic and consists of two binary approximate k NN classifiers [9] implemented in an Approximate Nearest Neighbor (ANN) framework developed by Mount et al. [12].

Three classes are separated, tibial medial cartilage, femoral medial cartilage and background. The method focuses on the medial compartments since OA is more often observed there [13]. One binary classifier is trained to separate tibial cartilage from the rest and one is trained to separate femoral cartilage from the rest, and these classifiers are combined with a rejection threshold [14]. A voxel is classified as belonging to one cartilage class if the posterior probability for this is higher than for the other cartilage class and higher than the rejection threshold,

$$j \in \begin{cases} \omega_{tm}, P_{tm,j} > P_{fm,j} \text{ and } P_{tm,j} > T; \\ \omega_{fm}, P_{fm,j} > P_{tm,j} \text{ and } P_{fm,j} > T; \\ \omega_b \quad \text{otherwise,} \end{cases} \quad (1)$$

where a voxel is denoted j and belongs to class ω_i (tm and fm is tibial and femoral medial cartilage and b is background). The rejection threshold, T , is optimized to maximize the Dice Similarity Coefficient (DSC) [15] between manual and automatic segmentations.

The classifiers are trained on 25 scans using feature selection, which is sequential forward selection followed by sequential backward selection with the area under the ROC curve [16] as criterion function. The selected features are: the image intensities smoothed (Gaussian) on three different scales, the position in the image, eigenvalues and the eigenvector corresponding to the largest eigenvalue of the structure tensor and the Hessian, and Gaussian derivatives up to third order. These features are ordered in decreasing significance determined by the feature selection.

2.3 Position Normalization Applied to the Segmentation

The placement of the knee varies slightly in clinical studies but is still a strong cue to the location of cartilage, which is evident in the described segmentation method where the position in the scan is selected as one of the most significant features. Even though the global location is a strong cue the minor variation in placement is a source of errors. Segmentation methods that rely on manual interaction is usually less sensitive to knee placement, we however seek to eliminate manual labor in segmentation tasks thus placement variations is an issue that needs attention. Figure 1 shows how knee position in the scan affects the automatic segmentation method.

One way of correcting for knee placement is to manually determine where in the scan the cartilage is, but this can take time with 3D images since a human expert typically search through the scans on a slice-by-slice basis. And when the segmentation method itself is automatic, an automatic adjustment is advantageous.

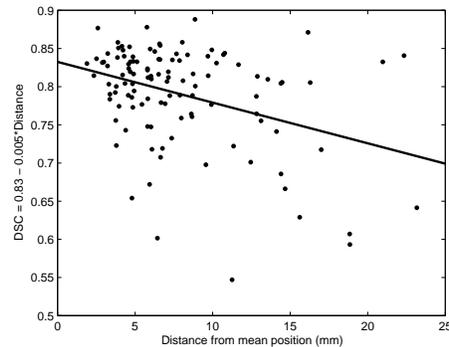


Fig. 1. The DSC between manual and automatic segmentation as a function of the distance to the mean position for the 114 scans with a line least-squares fitted to the points, illustrating how position variance affects the segmentation performance.

In order to adjust the segmentation method to become more robust to variations in knee placement we have developed an iterative scheme, which consists of two steps. First, the coordinates of the scan are shifted so that the cartilage center of mass found from the segmentation is positioned at the location for the center of mass for the cartilage points in the training set. Then the dilated volume of the segmentation is classified with the other features unchanged. The dilation extends the boundary outwards by three voxels and by only classifying the voxels inside this volume, which is typically only a few percent of the total scan volume, the computation time is not significantly increased. In order to determine if the selected region is a reasonable choice we repeated the classification with all the voxels in the image, yielding the same results with much longer computation time. The outcome is combined according to (1) and the largest connected component is selected as the cartilage segmentation.

3 Results

The automatic segmentation yields an average sensitivity, specificity and DSC are 81.1% ($\pm 11.0\%$ s.d.), 99.9% ($\pm 0.04\%$ s.d.) and 0.79 (± 0.07 s.d.) respectively in comparison with manual segmentations. As to inter-scan reproducibility of the volumes from the automatic segmentations, the linear correlation coefficient between the first and second scanning is 0.86 for the 31 knees, with an average volume difference of 9.3%.

After applying position normalization, the average sensitivity, specificity and DSC are 83.9% ($\pm 8.37\%$ s.d.), 99.9% ($\pm 0.04\%$ s.d.) and 0.80 ($\pm 0.06\%$ s.d.) respectively and it converges in only one iteration. Compared to the initial segmentation there is a significant increase in sensitivity ($p < 1.0 * 10^{-7}$) and in DSC ($p < 2.5 * 10^{-3}$) according to a paired t-test. In order to illustrate how the

segmentations are affected, the best and the worst results from the position correction scheme are shown in Figures 2 and 3. In the best case the DSC increases with 0.17 and for the worst scan it decreases with 0.017.

The reproducibility of the segmentation is improved, with an increase of the linear correlation coefficient from 0.86 to 0.93 and the average volume difference decreases from 9.3% to 6.2%. These reproducibility values can be compared to the volumes from the manual segmentations by a radiologist for the same data set. The linear correlation coefficient is 0.95, and the radiologist has an average volume difference of 6.8%.

The radiologist re-delineated the tibial medial cartilage in 31 scans in order to determine intra-rater variability for the manual segmentations. The average DSC between the two manual segmentations is 0.86, which explains the fairly low values of the DSC in our evaluation because the method is trained on manual segmentations by the expert and therefore attempts to mimic the expert. Assuming most misclassifications occur at boundaries, thin structures will typically have lower DSC. The corresponding DSC of the automatic segmentation versus expert for the tibial cartilage of the 31 scans is 0.82.

4 Summary and Discussion

We have developed an iterative method for normalization of the position used as a feature in segmentation of the articular cartilage in the knee joint using MRI. We maintain the strong cue given by position and simultaneously achieve robustness (but not mathematical invariance) to the variance in the operators placement of the knee in the scanner.

Our position normalization scheme converges in only one iteration, after which the inter-scan reproducibility is improved, the linear correlation coefficient for the volumes between the first and second scan occasion increases from 0.86 to 0.93, and the volume difference decreases from 9.3% to 6.2%. The corresponding values for the radiologist are 0.95 for the correlation coefficient and 6.8% volume difference. There is a small but significant increase in both sensitivity and in DSC for the 114 scans evaluated. From Figures 2 and 3 it can be seen that the best case is an efficient improvement the segmentation and the worst case is a practically unaltered segmentation, thus the segmentation is often positively and never negatively affected by our method. The scan with the worst result is from a severely osteoarthritic knee which can be difficult even for a highly trained expert to segment.

Thus by only slightly increasing computation time our position normalization scheme increases the sensitivity and DSC of the segmentation method, but more importantly, the inter-scan reproducibility of the method is much improved and is now as good as that of a radiologist, which is highly relevant in clinical studies where the knee placement in the scanner will vary. Because the method is completely automatic and improves the segmentation reproducibility without increasing the computational complexity, it can become a cost efficient tool in clinical studies where automated segmentation methods are being used.

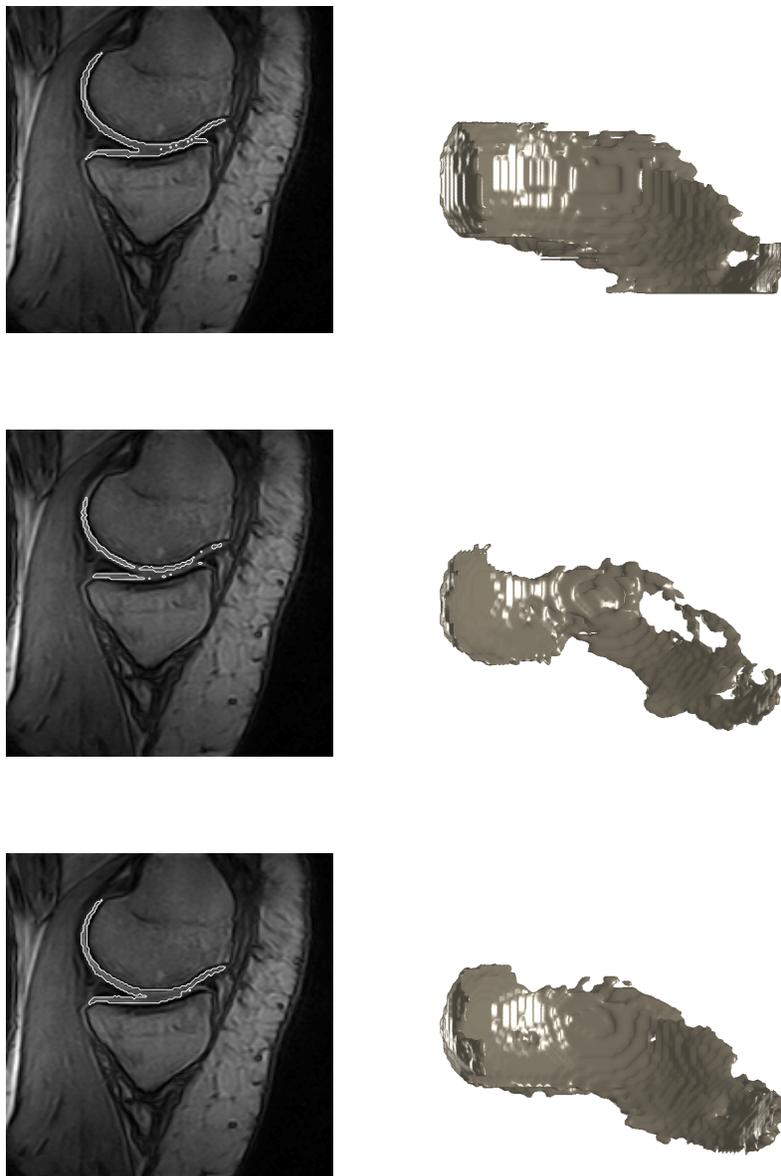


Fig. 2. The scan most improved by the position correction scheme, where the DSC increases from 0.61 to 0.77. Top row shows the manual segmentation, the second row shows the original segmentation and the third row shows the segmentation after position correction. The 3D views are seen from above, and the 2D images are a sagittal slice of the segmentation.

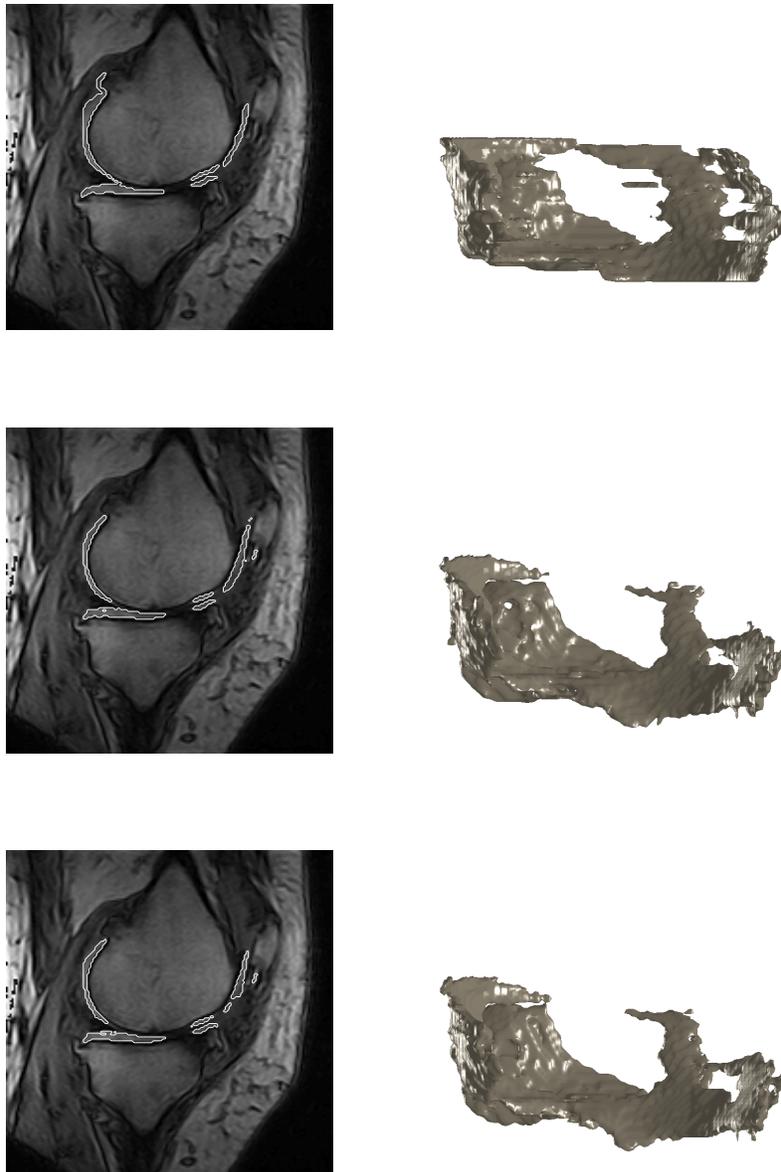


Fig. 3. The worst case scenario of applying position correction. The knee is severely osteoarthric ($KLi = 3$). For this scan there is no improvement in DSC. The manual segmentation is in the top row, the second row shows initial segmentation and the third row shows the segmentation after position correction.

References

1. Felson, D., Zhang, Y., Hannah, M., Naimark, A., Weissman, B., Aliabadi, P., Levy, D.: The incidence and natural history of knee osteoarthritis in the elderly. the framingham osteoarthritis study. *Arthritis and Rheumatism***38** (1995) 1500–1505
2. Graichen, H., Eisenhart-Rothe, R.V., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. *Arthritis and Rheumatism***50**(3) (2004) 811–816
3. Pessis, E., Drape, J.L., Ravaud, P., Chevrot, A., Ayrat, M.D.X.: Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and mri. *Osteoarthritis and Cartilage* **11** (2003) 361–369
4. Ejbjerg, B., Narvestad, E., adn H.S. Thomsen, S.J., Ostergaard, M.: Optimised, low cost, low field dedicated extremity mri is highly specific and sensitive for synovitis and bone erosions in rheumatoid arthritis wrist and finger joints: a comparison with conventional high-field mri and radiography. *Annals of the Rheumatic Diseases***13** (2005)
5. Lynch, J.A., Zaim, S., Zhao, J., Stork, A., Peterfy, C.G., Genant, H.K.: Cartilage segmentation of 3d mri scans of the osteoarthritic knee combining user knowledge and active contours. Volume 3979., *SPIE* (2000) 925–935
6. Stammberger, T., Eckstein, F., Michaelis, M., Englmeier, K.H., Reiser, M.: Interobserver reproducibility of quantitative cartilage measurements: Comparison of b-spline snakes and manual segmentation. *Magnetic Resonance Imaging* **17**(7) (1999) 1033–1042
7. Grau, V., Mewes, A., Alcañiz, M., Kikinis, R., Warfield, S.: Improved watershed transform for medical image segmentation using prior information. *IEEE Transactions on Medical Imaging* **23**(4) (2004)
8. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis* (4) (2000) 43–55
9. Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P., Christiansen, C.: Automatic segmentation of the articular cartilage in knee mri using a hierarchical multi-class classification scheme. In: *MICCAI*. (2005) 327–334
10. Kendall, D.: Shape manifolds, procrustean metrics and complex projective spaces. *Bull. London Math Soc.* **16** (1984) 81–121
11. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases* **16**(4) (1957)
12. Arya, S., Mount, D., Netanyahu, N., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. Number 5, *ACM-SIAM. Discrete Algorithms* (1994) 573–582
13. Dunn, T., Lu, Y., Jin, H., Ries, M., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. *Radiology***232**(2) (2004) 592–598
14. Folkesson, J., Olsen, O.F., Dam, E.B., Pettersen, P., Christiansen, C.: Combining binary classifiers for automatic cartilage segmentation in knee mri. In: *First International Workshop CVBIA*. (2005) 230–239
15. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* **26** (1945) 297–302
16. Jain, A., Duin, R., Mao, J.: Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence***22**(1) (2000)

Automatic Curvature Analysis of the Articular Cartilage Surface

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

¹ Image Analysis Group, IT University of Copenhagen, Denmark, jenny@itu.dk,
² Center for Clinical and Basic Research, Ballerup, Denmark.

Abstract. In osteoarthritis (OA) the articular cartilage degenerates, thereby losing its structure and integrity. Curvature analysis of the cartilage surfaces has been suggested as a potential disease marker for OA but until now there has been few results to support that suggestion. We present two methods for surface curvature analysis, one that estimates curvature on lower scales using mean curvature flow, and one method based on normal directions of and distances between surface points given by a cartilage shape model for high scale curvature estimates. We show that both methods can distinguish between healthy and osteoarthritic groups, and the shape model based method can even distinguish healthy from mild OA with high reproducibility, indicating the potential of surface curvature in becoming powerful new disease marker for cartilage degeneration.

1 Introduction

Magnetic resonance imaging (MRI) is becoming one of the leading imaging modalities in osteoarthritis (OA) research as it allows for non-invasive quantification of the articular cartilage and detection of cartilage degeneration, which is the characteristic symptom of the disease [1]. OA is second to heart disease in causing work disability and is associated with a large socioeconomic impact on health care systems [2].

Typical quantitative disease markers for OA is the articular cartilage volume and thickness, where the volume on its own can be a relatively poor disease marker [3]. Cartilage thickness measures combined with methods for finding correspondences in anatomy can give a more localized measure of cartilage degeneration, and be more sensitive to changes especially in load bearing parts. Williams et al. [4] have found sub-millimeter changes in local thickness measurement in a longitudinal study of OA risque subjects, but their method so far lacks statistical evaluation and relies on manual labor.

Standard clinical treatment of OA today does not reverse the cartilage degeneration which makes it is important to detect the disease at an early stage, and in clinical studies of new treatments it is equally important to detect changes across populations as soon as possible due to the high costs associated with drug

development. This leads to a search for new disease markers that can alone or combined with well established ones detect OA earlier and more reliably.

Joint congruity is determined by the surface shape, and describes the extent of which contacting joint surfaces match each other. It has been suggested that joint incongruity is related to high peak stress and such mechanical factors can influence the initiation and progression of OA [5]. Curvature analysis of the articular surface can be used as an estimation of the joint incongruity since high curvature on the overall surface will most likely lead to mismatching of the surfaces as the joint bends.

Hohe et al. [6] have analyzed the curvature of knee cartilage surfaces from MRI as an incongruity measure, by first segmenting the cartilage slice-by-slice using b-splines then estimating the principal curvatures locally from a b-spline interpolation on a 5×5 neighborhood of surface points $6mm$ apart. They found an average mean curvature of $29.6m^{-1}$ ($\pm 9.9m^{-1}$ s.d) for the tibial medial cartilage surface and $-0.9m^{-1}$ ($\pm 3.8m^{-1}$ s.d.) for the central part of the same surface, with inter-scan reproducibility values up to $4.7m^{-1}$ root mean squared SD, on 14 healthy subjects.

Terukina et al. [7] performed an in-vitro study of the curvature in 2D by slicing the knee joint through the sagittal plane and fitting a circle to three equidistant points within $1cm$ on the cartilage surface then taking the inverse of the radius for the curvature. They found an average curvature of $4.4m^{-1}$ for the femoral condyle in their study intended for cartilage replacement, emphasizing the importance of cartilage congruity in such interventions.

Even though curvature analysis shows potential as a OA disease marker, there has so far not been any study showing how or if the curvature differs between healthy and osteoarthritic knees. Joint congruity is related to the overall surface shape, thus mainly concern high scale curvature analysis. But low-scale curvature analysis is also interesting since it could be related to local shape changes and thus to focal cartilage lesions. In this paper, we present two automatic methods for high and low scale curvature analysis and demonstrate their abilities to distinguish between healthy and OA populations.

2 Curvature Estimation

We would like to analyze the curvature of the articular cartilage surface and see if the curvature differs between healthy and OA cartilage. We would also like to examine a wide range of scales in order to obtain the most relevant and robust estimates.

Looking at existing methods, Terukina et al. [7] keep a constant distance between points ($d = 0.5cm$) for their circle estimates. The curvature is the inverse of the radius of the circle which is uniquely defined by taking three neighboring points on the curve and let the arclength between them go to zero. By not shrinking the distance at high curvature, two points can be placed at the same approximate location giving an upper bound for the curvature value of $(d/2)^{-1} = 400m^{-1}$ for $d = 0.5cm$ (see Figure 1), giving a crude estimate at

low scales (high curvature). Another drawback is its 2D nature, where curvature can only be estimated in the selected plane of view.



Fig. 1. Fitting a circle to three equidistant points when two of the points are approaching each other (from left to right).

For non-invasive 3D curvature estimates the method of Hohe et al. [6] is more interesting and forcing b-splines to control points of the segmented surface at specific distances is a way of selecting a scale for the curvature analysis but the locations of the control points are affected by the image resolution, and by discontinuities between slices since their segmentation method is in 2D.

We apply curvature flow to super sampled volumes from a fully automatic cartilage segmentation [8] for a relatively smooth deformation of the surface from low towards higher scales, and evaluate the ability to discriminate between a healthy and an OA group during the flow for this low scale analysis. We also fit a deformable shape model to the cartilage and obtain a curvature measure from the normals and locations of boundary points for high scale curvature analysis, something the shape model is suitable for due to inherent regularization. Using the shape model we also examine the importance of evaluating the curvature in an anatomically well defined region of interest.

2.1 Curvature Estimation by Mean Curvature Flow

The mean curvature flow for a surface \mathbf{S} is $\mathbf{S}_t = \kappa_M \mathbf{N}$, where κ_M is the mean curvature (the mean of the two principal curvatures), t is time and its subscript denotes differentiation, and \mathbf{N} is the normal of \mathbf{S} . The standard method for implementing surface evolution is the level set method [9], with a function $\phi(x, y, z; t)$ which is an implicit representation of the surface at time t so that $\mathbf{S}(t) = \{(x, y, z) \mid \phi(x, y, z; t) = 0\}$. In the level set formulation the mean curvature flow is described by

$$\phi_t = \kappa_M |\nabla \phi| = \left[\nabla \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right] |\nabla \phi|,$$

where $\nabla \phi$ is the gradient, thus the mean curvature can be written

$$\begin{aligned} \kappa_M = & (\phi_x^2(\phi_{yy} + \phi_{zz}) + \phi_y^2(\phi_{xx} + \phi_{zz}) + \phi_z^2(\phi_{xx} + \phi_{yy}) \\ & - 2(\phi_x \phi_y \phi_{xy} + \phi_x \phi_z \phi_{xz} + \phi_y \phi_z \phi_{yz})) / (\phi_x^2 + \phi_y^2 + \phi_z^2)^{3/2} \end{aligned} \quad (1)$$

in terms of derivatives of ϕ , and the derivatives are calculated using finite differences.

For curvature flow of curves in \mathbb{R}^2 the Gage-Hamilton and Grayson theorems assures that convex non-intersecting curves will shrink smoothly to a point. This property does not extend to surfaces in \mathbb{R}^3 where topology changes can occur. However existence, stability and uniqueness of viscosity solutions of the mean curvature motion for hypersurfaces in level sets have been proved [10] [11].

The time behavior of the curvature for curves is described by a reaction diffusion equation which is non-trivial to solve [9]. For our surfaces, we examine the mean curvature behavior throughout the flow, starting from tibial medial cartilage volumes from automatic segmentations [8], by taking the average of the absolute value of the mean curvature in (1) at the tibial medial articular surface. Reinitializations are made every 7 iterations and the time step is 0.15. We super sample the scans, dividing each voxel into 125, with a new resolution of approximately $0.16mm$ side length, meaning that the curvature flow will initially mainly reduce partial volume effects.

2.2 Curvature Estimation by Shape Model Boundary Points

Using curvature flow the surface curvature can be analyzed at low scales, but when moving to higher scales as the flow continues the volume loses its cartilage like appearance and may change topology. Therefore we have developed another scheme for high scale curvature analysis where a deformable m-rep shape model is fitted to the cartilage. The m-rep represents an object by a mesh of medial atoms, each associated with a position, radius and directions to the boundary [12]. Besides curvature analysis, the shape model is also used for finding anatomical correspondences and local thickness measures in a related study [13].

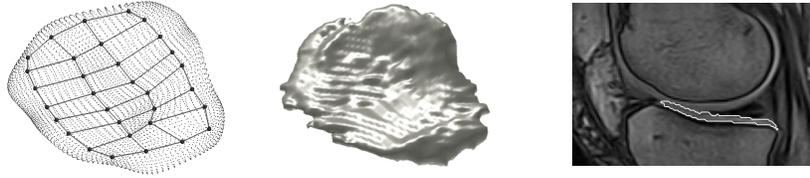


Fig. 2. An m-rep surface of a tibial medial cartilage sheet to the left, large points indicate medial atoms and small indicate boundary points. The segmented cartilage it was fitted to is in the middle, and part of a sagittal slice with the tibial medial cartilage delineated to the right.

The tibial medial cartilage model is fitted to the same set of automatic segmentations and consists of a mesh of 4×8 medial atoms, from which boundary points are interpolated on the articular surface with an average distance of

approximately $0.7mm$, see Figure 2. The deformable shape model framework ensures a regular, smooth boundary which makes it well suited for high scale curvature analysis. We define the load bearing (central) part of the articular surface as a region of interest and estimate the curvature locally in a 5×5 neighborhood by taking the angle between the normals and divide it by distance between them. Assuming short distances and small angles due to the regularization, this is an approximation to the curvature, $\kappa = \frac{d\theta}{ds}$, where θ is the angle between two normals and s is the arc length. The average of the absolute values of local estimations κ of the boundary points in the region of interest is chosen as the quantitative curvature estimate. The principal curvatures are the maximal and minimal curvature associated with corresponding directions. For the model there are only 8 discrete directions to choose from, and the mean of the curvature estimates in these directions can be seen as an approximation to the mean curvature in the region.

3 Results

Our data set consists of 139 (25 are used for training, 114 for testing) knee Turbo 3D T1 scans acquired from an Esaote C-Span low-field 0.18T scanner dedicated to imaging of extremities (40° flip angle, T_R 50 ms, T_E 16 ms), with approximate spatial resolution $0.8 \times 0.7 \times 0.7mm^3$. The test subjects are between 22-79 years old with an average age of 56 years, 59% females, and there are both healthy and osteoarthritic knees according to the Kellgren-Lawrence index (KLi) [14], a radiographic score from 0-4 where $KLi = 0$ is healthy, $KLi = 1$ is considered borderline or mild OA, and $KLi \geq 2$ is severe OA. In the test set there are 51, 28, 13 and 22 scans that have $KLi = 0, 1, 2$ and 3 respectively. For reproducibility evaluation, 31 knees were re-scanned after approximately one week, making the total number of scans 170. The spatial resolution is approximately $0.8 \times 0.7 \times 0.7mm^3$. We analyze the medial tibial cartilage in this study since OA is most often observed in medial compartments [15].

In the mean curvature flow, there is a rapid decrease in the average mean curvature on the articular surface initially as can be seen in Figure 3. During this time the inter-scan reproducibility (defined as percent pairwise measurement differences) is low, and an unpaired t-test cannot separate a healthy group ($KLi = 0$) from an OA group ($KLi \geq 1$) at a statistical significance level at 5%. This could be a result of the cancelation of partial volume effects present in the initial volume. As the flow propagates these values stabilize with curvature values of approximately $500m^{-1}$ and reproducibility of 5% pairwise difference (linear correlation coefficient around 0.4). In the separation between healthy and OA groups there is a minimum in p-values after approximately 35 iterations, where the p-value is $= 0.0011$ with inter-scan reproducibility of 6.2% pairwise difference with a linear correlation coefficient of 0.45 and average mean curvature of $660m^{-1}$ ($\pm 77m^{-1}$ s.d). At no time during the flow is it possible to separate healthy from borderline OA ($KLi = 1$) groups. The increasing p-values after around 160 iterations could be due to the inherent property of the flow that sur-

faces shrink to spheres, and that the flow have moved from a smoothing phase to a deformation phase. Figure 4 illustrates how a cartilage sheet deforms during mean curvature flow.

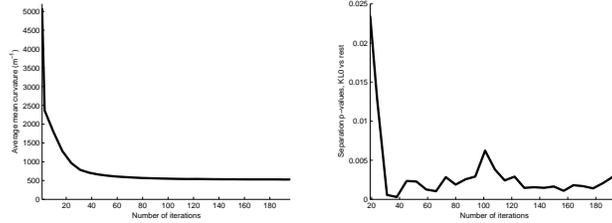


Fig. 3. Left: the average mean curvature of the articular surface. Right: p-values from t-test between healthy and OA groups.



Fig. 4. The appearance of a cartilage sheet during mean curvature flow. From left to right: after 0, 35, 100, and 200 iterations.

The curvature estimation on the boundary points of the m-rep gives an average of $45m^{-1}$ ($\pm 9.7m^{-1}$ s.d.), which is a factor 10 lower than the values of the curvature flow method. The m-rep based method can separate the healthy from the OA group ($p = 2.6 * 10^{-4}$) and can even separate healthy from borderline OA ($p = 0.0085$). The linear correlation coefficient for the inter-scan reproducibility is 0.68, with a pairwise difference in measurements of 11%. When estimating the curvature on the entire articular surface instead of only the load bearing part, the mean curvature is approximately the same, $46m^{-1}$ ($\pm 8.1m^{-1}$ s.d), but the method is no longer able to separate the healthy from borderline OA group ($p = 0.086$). The linear correlation is decreased from 0.68 to 0.59 in the inter-scan reproducibility.

These results can be compared to a well established OA disease marker, the cartilage volume normalized for bone size. Performing an unpaired t-test with

the normalized volumes from manual segmentations by a radiologist yields a p-value of 0.095 for the healthy vs. OA groups and 0.22 for the healthy vs. mild OA groups.

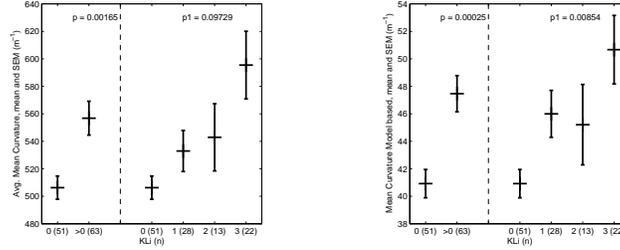


Fig. 5. Separation between healthy and OA groups using curvature estimation. Left: from curvature flow after 35 iterations, right: model based curvature.

4 Discussion

We have demonstrated two methods for curvature analysis on the articular cartilage surface. Using shortening flow it is possible to separate a healthy from an OA group ($p = 0.0011$) using the average mean curvature of the articular surface points where the best separation is after 35 iterations. The inter-scan reproducibility at that point has a correlation coefficient of 0.45. The method cannot at any time during the flow separate healthy from borderline OA groups. The model based method can separate the healthy from the OA group ($p = 0.00026$) and from the borderline OA group ($p = 0.0085$), with a correlation of 0.68 in reproducibility. In comparison, the normalized cartilage volume gives p-values of 0.095 and 0.22 for the healthy vs. OA and mild OA groups respectively. The ability to discriminate mild OA and the higher correlation for the model based method could be explained by the shape models robust method of finding relevant points on the articular surface.

The regularization of the model makes it suitable for high scale curvature analysis, finding curvature values around $50m^{-1}$ thus structures in the cm range, which corresponds to evaluating the overall shape of the surface and thus congruity. The mean curvature flow evaluates lower scales (curvatures in the range $500m^{-1}$ thus objects in the mm range) and could be interesting when searching for local changes or lesion on the cartilage surface. Future work will involve a longitudinal study and ground truth evaluation.

Because significant differences between healthy and mild to severe OA populations can be detected using curvature analysis with higher statistical significance than a well established measure, the cartilage volume normalized for bone size, these fully automatic methods can potentially become useful in clinical studies as markers for cartilage degeneration.

References

1. Graichen, H., Eisenhart-Rothe, R.V., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. *Arthritis and Rheumatism* **50** (2004) 811–816
2. Jackson, D., Simon, T., Aberman, H.: Symptomatic articular cartilage degeneration: the impact in the new millenium. *Clinical orthopaedics and related research* **391** (2001) 14–25
3. Gandy, S., Dieppe, P., Keen, M., Maciewicz, R., Watt, I., Waterton, J.: No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. *Osteoarthritis and Cartilage* (2002) 929–937
4. Williams, T.G., Taylor, C.J., Gao, Z., Waterton, J.C.: Corresponding articular cartilage thickness measurements in the knee joint by modelling the underlying bone. *MICCAI* (2003)
5. Bullough, P.: The geometry of diarthrodial joints, its physiologic maintenance, and the possible significance of age related changes in geometry to load distribution and the development of osteoarthritis. *Clinical Orthopaedics* **156** (1981) 61–66
6. Hohe, J., Ateshian, G., Reiser, M., Englmeier, K.H., Eckstein, F.: Surface size, curvature analysis, and assessment of knee joint incongruity with mri in vivo. *Magnetic Resonance in Medicine* (2002) 554–561
7. Terukina, M., Fujioka, H., Yoshiya, S., Kurosaka, M., Makino, T., Matsui, N., Tanaka, J.: Analysis of the thickness and curvature of articular cartilage of the femoral condyle. *Arthroscopy* **19** (2003) 969–973
8. Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P., Christiansen, C.: Automatic segmentation of the articular cartilage in knee mri using a hierarchical multi-class classification scheme. In: *MICCAI*. (2005) 327–334
9. Sethian, J.: *Level Set Methods and Fast Marching Methods*. 2 edn. Cambridge University Press (1999)
10. Evans, L., Spruck, J.: Motion of level sets by mean curvature i. *Journal of Differential Geometry* **33** (1991) 635–681
11. Chen, Y., Giga, Y., Goto, S.: Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Journal of Differential Geometry* **33** (1991) 749–786
12. Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Gash, G., Glotzer, J.M., Jiroutek, M.R., Lu, C., Muller, K.E., Tracton, G., yushkevich, P., Chaney, E.L.: Deformable m-reps for 3d medical image segmentation. *IJCV* **55** (2/3) (2003)
13. Dam, E.B., Folkesson, J., Pettersen, P., Christiansen, C.: Automatic cartilage thickness quantification using a statistical shape model. In: *MICCAI Joint Disease Workshop*. (2006)
14. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthritis. *Annals of the Rheumatic Diseases* **16** (1957)
15. Dunn, T., Lu, Y., Jin, H., Ries, M., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. *Radiology* **232** (2004) 592–598

Anatomically Equivalent Focal Regions Defined on the Bone Increases Precision when Measuring Cartilage Thickness from Knee MRI

Tomos G. Williams¹, Andrew P. Holmes², John C. Waterton², Rose A. Maciewicz², Anthony P. Nash², and Chris J. Taylor¹

¹ Imaging Science and Biomedical Engineering,
University of Manchester, Manchester, M13 9PT, U.K.

`tomos.williams@manchester.ac.uk`

² AstraZeneca, Alderley Park, Macclesfield, Cheshire, SK10 4TG, U.K.

Abstract. MRI cartilage measurement techniques need to be sufficiently sensitive to detect small, focal changes if they are to be used as biomarkers for OA drug development. Detailed cartilage thickness maps were constructed from MRI's of 19 healthy female volunteers. Anatomical correspondence between the volunteers was achieved by constructing optimal statistical shape models of the bones. Cartilage coverage across the cohort was used to define the region of pre-morbid sub-chondral bone and trimming boundaries which excluded the edges of the cartilage sheets. Functional sub-regions of the joint were drawn on the mean bone shapes. The regions of interest were propagated to all individuals, in an anatomically consistent manner, using the model-based correspondences. Mean cartilage thickness was measured within each region of interest. Excluding the edges of the cartilage sheet from the mean thickness measures increased the reproducibility, and hence sensitivity, of the measures.

1 Introduction

Magnetic Resonance Imaging (MRI) of the knee has been shown to be an effective method for quantifying articular cartilage [1]. To be useful as biomarkers for assessment of disease progression and drug efficacy in Osteoarthritis (OA), MR based cartilage quantitative methods must be sufficiently sensitive to detect small, focal changes in cartilage thickness. Cartilage degradation in OA is localised and focal analysis is required to ensure that changes are detected [2]. Others have identified the need to analyse the functional compartments of the joint independently [3] and to exclude the edges of the joint where cartilage is difficult to measure and may not be as involved in the disease process [4]. More sensitive measures allow effects of treatments on disease to be assessed with fewer subjects over shorter time periods, reducing the drug development lead-times and cost. Cartilage quantitative methods must also be scalable to enable analysis of large clinical trials.

In this paper we present a method which enables cartilage thickness to be measured in anatomically equivalent focal regions of the knee in a population

of volunteers. We demonstrate that reproducibility of mean cartilage thickness measures is improved by dividing the compartments into their functional subdivisions and excluding the edges of the cartilage sheets in an anatomically consistent manner. Improved reproducibility increases the sensitivity of the measures, making them more likely to detect the effects of osteoarthritis on articular cartilage in population studies.

2 Method

2.1 Data Acquisition

Twenty healthy female volunteers (age range 22 to 58, mean 38) were imaged using 1.0T or 1.5T at three different centres in the UK. At each visit, two sets of MR images were taken; a fat-suppressed T1-weighted 3D gradient echo sequence for visualisation of the hyaline cartilage, and a T2-weighted sequence to permit definition of the endosteal bone surface. Both images were acquired with $0.625mm \times 0.625mm \times 1.6mm$ resolution. One volunteer was excluded from the analysis due to an incorrect image acquisition. Further details of the study are presented in [5].

The femur, tibia and patella hyaline cartilage was segmented manually from the T1-weighted images by two trained, non-expert segmenters using a region growing algorithm implemented in TOSCA (IBM, Winchester). Each image was segmented twice by both segmenters. The endosteal surfaces of the femur, tibia and patella were segmented from the T2-weighted images by a third trained, non-expert segmenter using EndPoint software (Imorphics, Manchester, UK). Closed, triangular surface representations of the cartilage and bone were produced using a semi-automatic approach described previously [6].

Differences between the bone and cartilage tissue properties result in finite mis-registration artifact between the two MR sequences known as the *Chemical Shift*. In addition, there is the possibility of movement of the joint during scanning, causing further mis-alignment. To ensure alignment of the bone and cartilage from the two MR sequences, these artefacts were corrected by rigid registration. A multi-resolution search pattern was used to find the rigid transform of the bone surface which minimised the standard deviation of the distances from a dense set of points on the bone surface to the inner cartilage surface.

2.2 Measuring Cartilage Thickness

We wish to measure cartilage thickness at any point on the surface of the knee joint. The bone, rather than the cartilage itself, is used as the reference surface from which cartilage thickness is measured. This is because it is more stable over time than the cartilage and has a consistent topology for each individual. At each measurement point on the bone, a 3D normal to the surface was defined. If the 3D normal did not intersect the cartilage surface within 4mm of the bone, then the measurement point was labelled as uncovered by cartilage. If points of

intersection of the bone normals with the inner and outer cartilage surface were found, thickness was measured as the distance between these intersecting points. By measuring cartilage at a dense set of points on the bone surface, a detailed map of cartilage coverage and thickness can be constructed.

2.3 Establishing Correspondence Between Thickness Maps

In order to combine thickness measurements across a population, the individual cartilage thickness maps need to be brought into anatomical correspondence. Anatomical correspondence is achieved by defining a set of dense, corresponding points on each bone shape. As discussed by Davies *et al.* [7], this is the dual of constructing statistical shape models. Statistical shape models capture the variation in a training set of examples [8]. They consist of a mean shape, represented as a dense set of points, and a set of modes of shape variation. Each example in the training set can be represented as the mean shape and a linear combination of the modes of variation.

In order to define anatomical correspondences, statistical models of the patella, femur and tibia bones were built using the Minimum Description Length (MDL) method of Davies *et al.* [7]. The approach automatically finds a set of corresponding points on each example shape. It is based on the premise that anatomical correspondences produce simpler, linear statistical models. Model complexity is measured using its minimum description length criterion.

An overview of the approach is presented here. A more detailed description of the MDL method applied to constructing statistical shape models of the bones can be found in [9]. Each bone surface was mapped onto a unit sphere using the diffusion method of Brechbühler [10]. This provides a common reference for all shapes and a basis for manipulation of the correspondences. A set of equally spaced points were defined on each unit sphere and mapped back onto each bone surface by finding their precise locations on their residing triangles in barycentric coordinates, and computing the equivalent location in the corresponding triangle on the original surface. This provided initial positions for a corresponding number of points on each surface.

A multi-resolution, iterative optimisation approach was used to move the correspondences on the bone surfaces in order to bring them into anatomical alignment. During each iteration, each shape example was chosen in random order and the position of its corresponding points and alignment pose were optimised with respect to the model description length. When the model reached convergence, the number of corresponding points was increased by linear interpolation and optimisation continued at the higher resolution. When the model converged at the highest resolution, optimisation was terminated.

2.4 Aggregate Cartilage Coverage and Thickness Maps

We wish to produce summary maps of cartilage coverage and thickness for a population of volunteers. Cartilage coverage and thickness was measured, for each cartilage segmentation, at the dense corresponding points on the bone surface

found using the MDL method. Since measurements were made at corresponding points, they can be combined across the population to produce aggregate maps and displayed on the mean bone shapes. Aggregate coverage was computed as the point-wise percentage of individual coverage. Aggregate thickness maps were similarly constructed by calculating point-wise summary statistics of the individual thickness maps.

2.5 Defining Regions of Interest

In characterising cartilage quantity, changes to both the thickness and extent of the cartilage sheet are of interest. Computing the mean cartilage thickness over the currently covered bone area does not account for denuded areas where previously existing cartilage has disappeared. It is therefore necessary to estimate the total subchondral bone region i.e. the area covered by cartilage in the pre-morbid condition [11]. Others have used manual identification of the subchondral area in each image slice [3].

The edges of the cartilage sheets are least involved in OA disease and are difficult to segment, particularly, given the partial volume effects where the cartilage sheet curves into the image plane. Others have manually delineated the central, load bearing regions of the joint in each image [4]. Morphological changes to cartilage due to Osteoarthritis tend to be localised. Others have identified the need for regional analysis of cartilage but current methods require manual delineation of regions during image segmentation [3].

Manual identification of regions of interest for each image is time consuming and requires a significant level of expertise. Our framework allows us to define regions of interest on the mean bone surface and, using the dense correspondences, project them onto any individual bone shape maintaining their anatomical fidelity.

The total subchondral area, denoted tAB (total Area of Bone), was defined for each cartilage compartment as the connected set of measurements points exhibited coverage in any of the segmentations. This is an overestimation of the subchondral area in any individual. To limit measurement to the central load-bearing sub-regions of each cartilage sheet, trimming boundaries were defined on the mean bone shapes. The trimmed region for each cartilage compartment, denoted $Trm AB$, was drawn manually as a closed polygon on the mean bone surface, such that they tightly encompassed the points which exhibited cartilage coverage in at least 90% of the segmentations from the healthy female volunteers. The subchondral and trimmed regions were defined along the connecting edges of the correspondences and could therefore be propagated to all individuals' bone surfaces.

Sub-regions of the femoral and tibial articular surfaces were defined on the mean bone shapes according to the bearing surfaces of the different joint compartments as described in [11]. As with the coverage boundaries, the dividing boundaries of the functional regions were drawn along the connecting edges of the corresponding points. This enabled them to be propagated from the mean bone shape to the individual bones maintaining their anatomical fidelity.

2.6 Regional Cartilage Morphology Measures

The most basic quantitative measure of cartilage is volume, calculated directly from the cartilage segmentations. Volume of cartilage, denoted VC , was computed for the major compartments from the closed cartilage surfaces using Gauss' theorem [12]. To normalise for variation in joint size, cartilage volume can be divided by the subchondral bone area to provide a measure, denoted VC/tAB (Volume of Cartilage divided by total Bone Area), with units of millimetres [13]. Since cartilage compartments were segmented as a whole, these measures cannot be restricted to sub-regions of the joint. Using our framework however, it is possible, to measure regional mean thickness using the dense, corresponded cartilage thickness measurements. We calculated mean cartilage thickness within a region of interest, denoted $ThCAB$ (Thickness of Cartilage within specified Bone Area), as the mean of cartilage thickness measures within the region weighted by their surrounding area. For each functional sub-compartment, the mean cartilage thickness was computed over both subchondral ($ThCtAB$) and trimmed ($ThC^{Trm}AB$) coverage regions.

3 Results

The MDL statistical shape models provided a set of 4098 anatomically corresponding points each for the patella, femur and tibia. In order to provide detailed thickness maps and enable accurate delineation of the sub-regions of the joint, linear interpolation of the correspondence points was performed on the larger femoral and tibial bone surfaces. This resulted in mean connected neighbour separations of 1.09mm, 1.32mm and 1.07mm for the patella, femur and tibia bones respectively.

Individual cartilage coverage and thickness maps were produced for each cartilage segmentation. This yielded a total of 76 (19 volunteers \times two segmenters \times two repeat segmentations) coverage and thickness maps. Figure 1(a) shows the aggregate coverage map, shown on the mean bone shapes, and the subchondral and trimmed regions. The subchondral area allows compartmental volumes to be normalised for inter-subject variation in joint sizes, and provides anatomically consistent regions over which to compute mean thickness for each individual. The trimmed regions, which are completely encompassed by the subchondral region, allow mean thickness to be measured over the central, load-bearing regions of the joint, excluding the edges of the cartilage sheets.

Figure 1(b) shows the range of cartilage thickness. As expected for a healthy cohort, the load-bearing regions of the joint exhibit thicker cartilage in particular the patella, femoral patellar groove and the load bearing regions of the tibial compartments. Standard deviation map demonstrates consistent and low variation, approximately 1mm, for the majority of the surface area.

Figure 2 shows the functional regions of interest shown on the mean bone shapes and propagated to two individual bones using the dense correspondences. The trimming boundary is shown for reference. The anatomical integrity of the regions is largely maintained when propagated onto the individual bone surfaces.

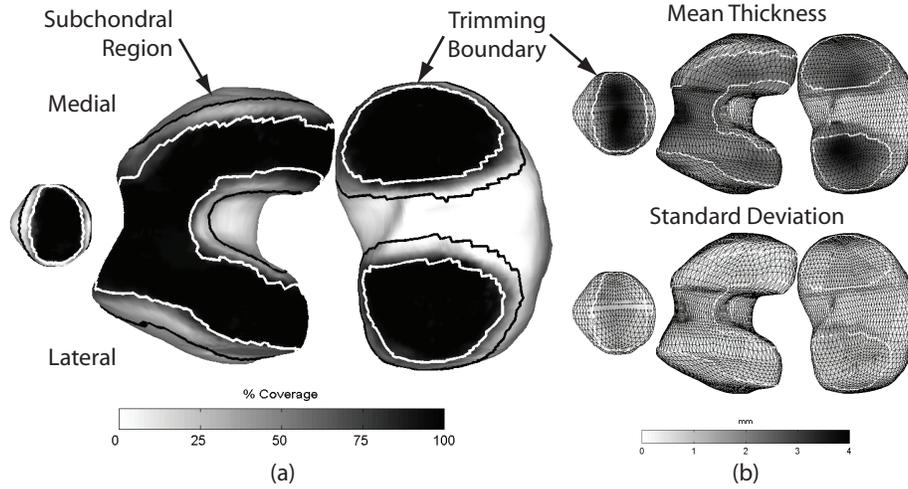


Fig. 1. (a) Aggregate cartilage coverage map with the subchondral (black) and trimmed (white) boundaries. (b) Normal range of cartilage thickness.

The intra-segmenter and inter-subject variability of mean thickness measures is shown in Table 1. They demonstrate that repeatability is improved when measurement is restricted to the trimmed regions, implying that measurements taken at the edges of the cartilage are less reliable [4].

Morphological Measure CoV(%)	Intra-Segmenter				Inter-Subject			
	F	P	LT	MT	F	P	LT	MT
VC/tAB	2.91	3.18	3.23	4.80	13.8	15.4	23.6	15.9
$ThCtAB$	2.86	2.91	3.26	4.89	13.4	16.0	24.0	15.9
$ThC^{Trim}AB$	2.79	2.58	2.91	4.04	12.6	17.1	17.9	14.6

Table 1. Intra-Segmenter and Inter-Subject coefficient of variation.

Table 2 presents the normal range of mean cartilage thickness for all functional regions derived from the 19 healthy female volunteers. The results indicate that cartilage is thicker in the central, load bearing regions. Cartilage is thicker in the patella, femoral patella groove and the lateral tibia which is consistent with others' published results [1].

4 Discussion

We have introduced a method that automatically identifies dense, anatomical correspondences invariant to individual joint shape or size. The method offers operational advantages by allowing regions of interest to be defined on the mean

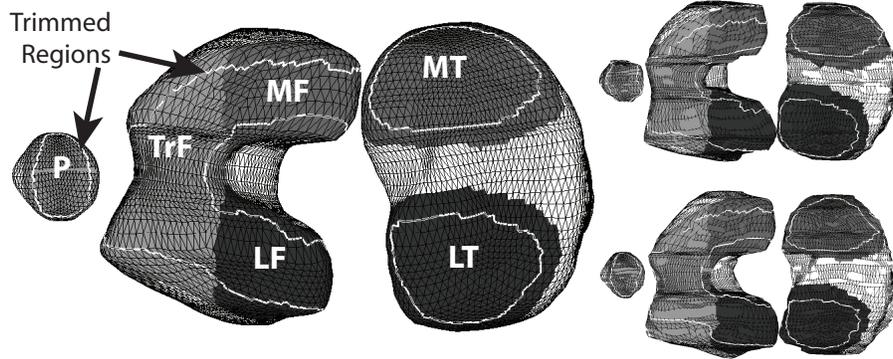


Fig. 2. Regions of interest defined on the mean bone shapes (top) and propagated to two of the individuals' bone shapes. P: Patella, TrF: Trochlearfemur, MF: Medial Femur LF: Lateral Femur, MT: Medial Tibia LT: Lateral Tibia.

(mm)	F	TrF	LF	MF	P	LT	MT
VC/tAB	1.33 (0.16)	-	-	-	1.88 (0.29)	1.51 (0.32)	1.23 (0.18)
$ThCtAB$	1.19 (0.14)	1.35 (0.22)	1.06 (0.18)	1.14 (0.22)	1.89 (0.30)	1.60 (0.35)	1.36 (0.21)
$ThC^{Trim}AB$	1.90 (0.22)	2.00 (0.28)	1.80 (0.23)	1.87 (0.28)	2.82 (0.48)	2.54 (0.44)	1.91 (0.27)

Table 2. Normal range, mean (standard deviation), of mean thickness measures.

bone shapes, and propagated to individuals in an anatomically consistent manner. Regional analysis allows changes in cartilage thickness to be detected independently in functional sub-regions, and also allows analysis to be restricted to anatomically consistent central, load-bearing regions of the joint, where change is more likely to occur. The intra-segmenter coefficients of variation for the mean thickness measurements for the femoral, patellar, lateral tibial and medial tibial compartments were reduced when measurement was restricted to the trimmed regions. This suggests that segmentations were more repeatable in the central, load-bearing regions of the joint. Similarly, inter-subject variation for the femur, lateral tibia and medial tibia compartments decreased demonstrating less inter-subject variability in the central regions of the joint making the measures more sensitive, and increasing the likelihood of detecting change in cohort studies.

Acknowledgments

We acknowledge the study investigators S. J. Gandy, R. J. Moots, P. Creamer and A. J. Silman.

References

1. Eckstein, F., Glaser, C.: Measuring cartilage morphology with quantitative magnetic resonance imaging. *Seminars in Musculoskeletal Radiology* **8**(4) (2004) 329–353
2. Cicuttini, F., Ding, C.H., Wluka, A., Davis, S., Ebeling, P.R., Jones, G.: Association of cartilage defects with loss of knee cartilage in healthy, middle-age adults - a prospective study. *Arthritis and Rheumatism* **52**(7) (2005) 2033–2039
3. Eckstein, F., Hudelmaier, M., Wirth, W., Kiefer, B., Jackson, R., Yu, J., Eaton, C., Schneider, E.: Double echo steady state (dese) magnetic resonance imaging of knee articular cartilage at 3 tesla - a pilot study for the osteoarthritis initiative. *Annals of the Rheumatic Diseases* **65**(4) (2005) 433–441
4. Koo, S., Gold, G.E., Andriacchi, T.P.: Considerations in measuring cartilage thickness using MRI: factors influencing reproducibility and accuracy. *Osteoarthritis and Cartilage* **13**(9) (2005) 782–789
5. Morgan, S.R., Nash, A.F.P., Waterton, J.C., Maciewicz, R.A., Leadbetter, J.E., Gandy, S.J., Moots, R.J., Creamer, P., Silman, A.J.: Magnetic resonance imaging measurement of knee cartilage volume in a multi-centre study. *Rheumatology* **42** (2003) 1–4
6. Williams, T.G., Taylor, C.J., Waterton, J.C., Holmes, A.: Population analysis of knee cartilage thickness maps using model based correspondences. In: *ISBI 2004 2004 IEEE Int. Symp. on Biomedical Imaging, Arlington, VA, USA., IEEE* (2004) 193–196
7. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.T.: 3D statistical shape models using direct optimisation of description length. In: *ECCV 2002: 7th European Conference on Computer Vision, LNCS 2352/2002. Volume III., Copenhagen, Denmark, Springer* (2002) 3–21
8. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models — their training and application. *Computer Vision and Image Understanding* **61**(1) (1995) 38–59
9. Williams, T.G., Taylor, C.J., Gao, Z., Waterton, J.C.: Corresponding articular cartilage thickness measurements in the knee joint by modelling the underlying bone. In Taylor, C., Noble, J.A., eds.: *IPMI 2003 Proc. 18th Int. Conf. on Information Processing in Medical Imaging, Lecture Notes in Computer Science 2732, Ambleside, UK, Springer* (2003) 126–135
10. Brechbühler, C., Gerig, G., Kubler, O.: Parametrization of closed surfaces for 3-D shape-description. *Computer Vision and Image Understanding* **61**(2) (1995) 154–170
11. Eckstein, F., Ateshian, G., Burgkart, R., Burstein, D., Cicuttini, F., Dardzinski, B., Gray, M., Link, T.M., Majumdar, S., Mosher, T., Peterfy, C., Totterman, S., Waterton, J., Winalski, C.S., Felson, D.: Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis. *Osteoarthritis and Cartilage* **(to appear)** (2006)
12. Hughes, S.W., D'Arcy, T.J., Maxwell, D.J., Saunders, J.E., Ruff, C.F., Chiu, W.S.C., Sheppard, R.J.: Application of a new discreet (*sic*) form of gauss' theorem for measuring volume. *Phys. Med. Biol.* **41** (1996) 1809–1821
13. Burgkart, R., Glaser, C., Hinterwimmer, S., Hudelmaier, M., Englmeier, K.H., Reiser, M., Eckstein, F.: Feasibility of T and Z scores from magnetic resonance imaging data for quantification of cartilage loss in osteoarthritis. *Arthritis and Rheumatism* **48**(10) (2003) 2829–35

Automatic Detection of Erosions in Rheumatoid Arthritis Assessment

Georg Langs^{1,3}, Philipp Peloschek², Horst Bischof¹, and Franz Kainberger² *

¹ Institute for Computer Graphics and Vision, Graz University of Technology,
Inffeldg. 16 2.OG, A-8010 Graz, Austria,

langs@prip.tuwien.ac.at, bischof@icg.tu-graz.ac.at

² Department of Radiology, Medical University of Vienna,
Währinger Gürtel 18-20 A-1090 Vienna, Austria,

philipp.peloschek@meduniwien.ac.at, franz.kainberger@meduniwien.ac.at

³ Pattern Recognition and Image Processing Group, Vienna University of Technology,
Favoritenstr. 9 / 1832, A-1040 Wien

Abstract. In this paper a method for automatic detection of erosive destructions caused by rheumatoid arthritis is proposed. Based on hand radiographs the algorithm detects erosions by means of an appearance model learned from a training set of bones. The model is utilized to classify the texture of the bone in the vicinity of the contour with respect to erosive destructions. Quantitative results of the algorithm are reported for a set of 17 radiographs of moderately and mildly diseased hands.

1 Introduction

Rheumatoid arthritis (RA) is a chronic systemic inflammatory disease involving primarily the synovial membranes and articular structures of multiple, mainly peripheral joints. The disease is progressive and results in pain, stiffness, and swelling of joints, which show deformity and ankylosis in late stages. Recurring inflammation of affected joints (i.e. arthritis) leads to degradation of cartilage and bone erosions. This affects physical function and mobility, causes substantial short-term and long-term morbidity and a significantly shorter life expectancy compared with the general population.

The precise quantification of cartilage and bone destruction (i.e. joint space narrowing and erosions) caused by rheumatoid arthritis is a decisive factor for the treatment with aggressive treatment strategies (e.g. methotrexate) or biologic agents like anti-Tumor Necrosis Factor alpha (anti-TNF alpha) drugs. Furthermore the anti-inflammatory effects of different agents, which retard radiographic progression, have to be compared across clinical trials. Radiography is used as the standard method to monitor the long-term progression of RA.

State of the art in quantification Several manual quantification methods (i.e. scoring systems) have been published over the last 30 years [1, 2]. They are time consuming, require specialized training and suffer from significant inter- and intrareader variation [3] limiting long term assessment of disease progression.

* This research has been supported by the Austrian Science Fund (FWF) under the grant P17083-N04 (AAMIR).

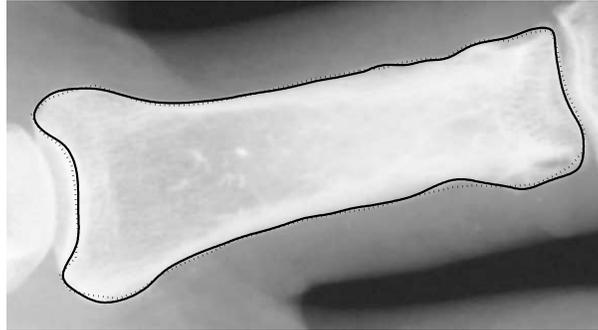


Fig. 1. Proximal phalanx 3: ASM result (dashed), and refined contour used for analysis (solid).

Recently the availability of digital image acquisition systems has prompted the development of new, increasingly automatic, quantitative measurements of radiographic features like bone axes, and density, or joint space width. Several groups have developed computer-assisted image analysis systems to measure osseous changes on radiographs over time [4–8]. Only a small number of these methods integrate a certain degree of automation with user input. They are restricted to angle- and distance measurements, and include manual annotation on the image as part of the measurement procedure. Although in a number of recent clinical studies [9, 10] erosions gave more discriminative information, the assessment of erosive changes on the bone contour is performed manually and thus suffers from the aforementioned limitations.

Automatic assessment of erosive changes In this paper we present a fully automatic method to assess erosive changes caused by RA. The method identifies the bone contours in hand radiographs and subsequently analyses them with respect to erosive bone destructions. Erosions are detected and their extent is determined. As a result the algorithm provides the medical expert with a repeatable measurement of the extent of erosive destruction caused to each joint.

Paper structure The paper is structured as follows: In Sec. 2.1 the automatic delineation of bone contours is described. The model based approach to erosion detection in Sec. 2.2 and Sec. 2.3 forms the main part of the work. In Sec. 3 experimental results are reported, and in Sec. 4 a final discussion is given.

2 Methodology

The automatic assessment of erosive changes is performed in two main steps. First the bone contours are delineated automatically, then the contour is analyzed w.r.t. erosive destructions. The accurate identification of the bone contour is based on ASM driven snakes [11], and will be explained only briefly. The main innovation of the paper is the automatic erosion detection based on local bone appearance.

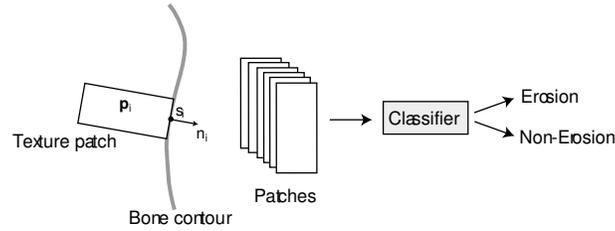


Fig. 2. Patches are extracted from the bone contour and are classified with respect to the presence of erosive destructions.

2.1 Locating bone contours

Given a hand radiograph the locations of the bones are determined by local linear mapping nets and an accurate delineation of the bone contour is accomplished with a model driven procedure [11]. After joint positions have been estimated coarsely, *active shape models (ASMs)* identify the bone contours and subsequently the contour estimate is refined by ASM driven snakes. Thereby a very accurate contour delineation is obtained. Due to the landmark based nature of ASMs positions on the bone can be identified consistently over different instances of the same bone. This allows for a repeatable definition of joint regions and follow up analysis.

Active shape models (ASMs) are based on shapes represented by a finite set of n landmarks. Given a training set of n_T annotated shapes, e.g. contours of proximal phalanges, each shape can be represented by a $2n$ dimensional vector \mathbf{s}_i generated by concatenation of the x and y coordinates for 2 dimensional data. In order to achieve a compact representation principal component analysis (PCA) is used on the set $\mathcal{S} = \{\mathbf{s}_i, i = 1, \dots, n_T\}$ and thereby a new coordinate system that represents each of the vectors

$$\mathbf{s}_i = \bar{\mathbf{s}} + \sum_{j=1}^{n_p} a_j \mathbf{e}_j, \quad (1)$$

is created. The modes \mathbf{e}_j are the eigenvectors of the covariance matrix sorted according to decreasing eigenvalue λ_j . $\bar{\mathbf{s}}$ is the mean shape and n_p can be chosen to fulfill a given accuracy constraint. n_p was chosen so that 95% of the training set variability are represented by the model. The eigenvalues λ_j correspond to the variance of the data in the direction \mathbf{e}_j . It can be viewed as a mean shape of an object and a set of valid deformations that, if applied, generate new instances of shapes. Due to the model constraint these shapes remain in the learned class of objects. The local texture information at the landmark positions is extracted in the form of gray level profiles orthogonal to the contour. ASMs are described in detail in [12, 13].

ASM search starts with a coarse estimate of the object location, and iteratively updates the landmark positions. The search is guided by the gray-level

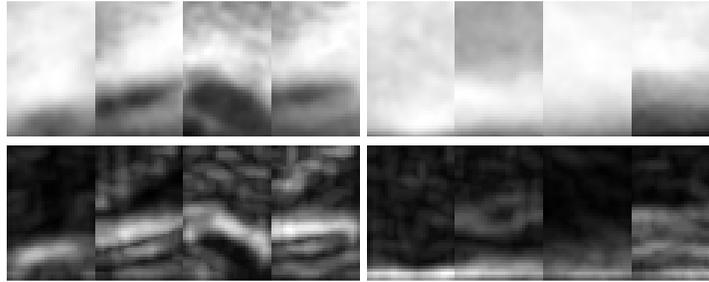


Fig. 3. Examples of bone texture patches, with the bone contour on the lower patch border; left: affected by erosive destructions, right: not affected; upper row: radiography texture, lower row: local gray-value range images.

information in the image and the shape constraint imposed by the eigenspace model. The result is the landmark vector \mathbf{s} defining the positions of the model landmarks in the search image.

Contour refinement Although the shape constraint imposed by ASM search fosters a highly reliable detection of the landmark positions and allows for the consistent identification of landmarks in different instances of the bone, it can prohibit high accuracy, due to a possibly limited representative power of the training set. This is especially relevant in the presence of pathology that cannot be adequately captured by the linear model underlying ASMs. Therefore the contour delineation \mathbf{s} is refined by a snake [11, 14] that adapts to the bone contour while increasing the flexibility of the bone model. During refinement the elasticity of the snake is increased, allowing for a gradual fitting of deviations from the model. It results in a dense contour estimate $\hat{\mathbf{s}}$. Fig. 1 shows the bone contour of a proximal phalanx detected by ASM search (dashed line) and subsequently refined by a snake (solid line). The mean/median error to a manual ground truth annotation is in the range of 0.14/0.10 mm. Note that for better stability in the base region a ridge close to the subchondral cortical plate was used as reference.

2.2 Modeling local bone appearance

Extracting appearance After the bone contour has been delineated resulting in a dense sampling of the contour points $\hat{\mathbf{s}}$ for each point $s_i \in \hat{\mathbf{s}}$ on the bone contour, the bone texture is extracted in the form of a rectangular patch \mathbf{p}_i with borders parallel and orthogonal to the bone contour normal vector n_i at the position s_i . This results in a set of patches $(\mathbf{p}_i)_{i=1,\dots,n}$, where $\mathbf{p}_i \in \mathbf{R}^{v \times k}$ i.e. a gray value patch of size $v \times k$ extracted from the image. See Fig. 2 for a schematic overview of the texture patch extraction. For the detection of erosive changes local range images give relevant information. They are generated from the patches by calculating the range of gray values in a local neighborhood for each patch pixel. In Fig. 3 examples of patches with and without erosive destructions are depicted.

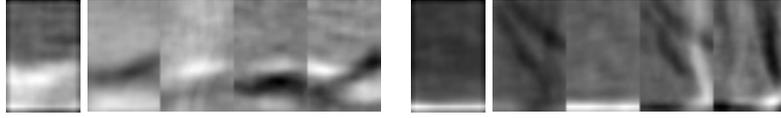


Fig. 4. Appearance models: centers and first 4 modes of variation; left: erosion- model; right: non-erosion model.

Modeling appearance During training each point on the bone contour is assigned one of two classes: *erosion* and *non-erosion*. For these two classes the appearance of the training examples is modeled by *Gaussian mixture models (GMMs)*[15]. That is on a training set of radiographs the bone contours are annotated manually w.r.t. the two classes by a medical expert. The resulting corresponding two sets of local range patches $(\mathbf{p}_i)_{i=1,\dots,n_1}^{ERO}$ and $(\mathbf{p}_i)_{i=1,\dots,n_2}^{NONERO}$, resp. are used to build two different models. A GMM comprising N Gaussians defines a probability density function

$$p(\mathbf{x}|\theta) = \sum_{j=1}^N \alpha_j \mathcal{N}(\mathbf{x}|\mu_j, \Sigma_j), \quad \text{and hence} \quad \theta = \langle \mu_1, \Sigma_1, \dots, \mu_N, \Sigma_N \rangle, \quad (2)$$

where \mathbf{x} is an observed patch extracted from the bone contour as described above. For each Gaussian distribution μ_j is the mean vector or *centre* and Σ_j is the full covariance matrix. Both define the probability density function. The ordered set of μ_j and Σ_j defines the GMM and will be denoted by θ . These parameters are derived from the training data by the EM - algorithm [15].

The GMMs capture the appearance of the intact bone and that of erosions by a clustering in the space \mathbf{R}^{vk} spanned by the vectorized patches. The model is generative. For each Gaussian the mean μ_j corresponds to a mean patch or cluster center and the distribution is spanned by the eigenvectors of Σ_j . It enables the model to represent a variety of erosions and bone textures in a compact manner. However, the direct appearance based approach has limits, on which we will comment in Sec. 3. In Fig. 4 centers and the first 4 eigenvectors of texture variation of the two distributions θ_{ERO} and θ_{NONERO} are depicted.

2.3 Detecting erosions

In order to detect erosions in a radiograph, patches are extracted along the bone contours analogous to model building. For each patch \mathbf{p}_i the classification is performed based on the two probability density functions $p(\mathbf{p}_i|\theta_{ERO})$ and $p(\mathbf{p}_i|\theta_{NONERO})$ learned during training:

$$l(\mathbf{p}_i) = \begin{cases} \text{Erosion,} & p(\theta_{ERO})p(\mathbf{p}_i|\theta_{ERO}) > p(\theta_{NONERO})p(\mathbf{p}_i|\theta_{NONERO}) \\ \text{Non Erosion,} & p(\theta_{ERO})p(\mathbf{p}_i|\theta_{ERO}) < p(\theta_{NONERO})p(\mathbf{p}_i|\theta_{NONERO}) \end{cases} \quad (3)$$

where $l(\mathbf{p}_i)$ is a label assigned to each point on the bone contour stating whether it belongs to an erosion or not. $p(\theta_{ERO})$ (where $p(\theta_{NONERO}) +$

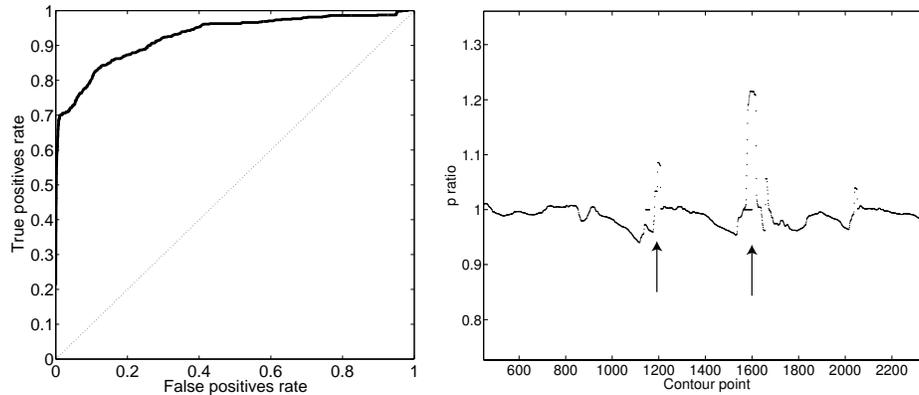


Fig. 5. Left: detection rate for 4-fold cross validation on hand radiographs; right: example of the ratios between $p(\mathbf{p}_i|\theta_{ERO})$ and $p(\mathbf{p}_i|\theta_{NONERO})$ for a set of contour points. Two erosions are indicated by arrows.

$p(\theta_{ERO} = 1)$ is the prior probability for an erosion to occur. It is estimated from the training set. The ROC - curve in Fig. 5 is reported w.r.t. this prior.

The resulting $l(\mathbf{p}_i)$ is an indicator for the presence of erosions on each bone contour point. It can be used to quantify the extent of the erosive destruction directly. Moreover the label is visualized in the radiograph in order to provide the radiologist with the possibility to verify the detected erosions.

3 Experiments

Setup Evaluation results of the method are reported for a set of 17 hand radiographs. Image resolution is 0.0846mm/pixel. Manual annotations of the erosions by a medical expert serve as relative groundtruth. Bone contours for moderately and mildly diseased proximal phalanges (PP2, PP3, PP4, and PP5) were delineated by ASMs comprising 64 landmarks, that had been trained on a different set of 40 radiographs. A refinement of these contours by ASM driven snakes served as reference contour for the patch extraction.

In order to investigate the performance quantitatively the bone contour was classified with respect to erosive changes in approximately 1 pixel intervals. The evaluation was restricted to unequivocal erosions showing all radiographic signs (cortical break, focal osteopenia and irregular surface). On a set of 7896 labeled patches of size 30×50 pixels (676 with erosions and 7250 without erosions) extracted from the 17 images 17-fold cross validation was performed. In order to judge the capability of the method to quantify the extent of erosions manual annotations and automatic detection were compared.

Results In Fig. 5 left an ROC curve for the detection results from 17-fold cross validation is depicted. Contour points were classified as described above. With a false positives rate of 16% the classification reaches a true positives rate of 85%.

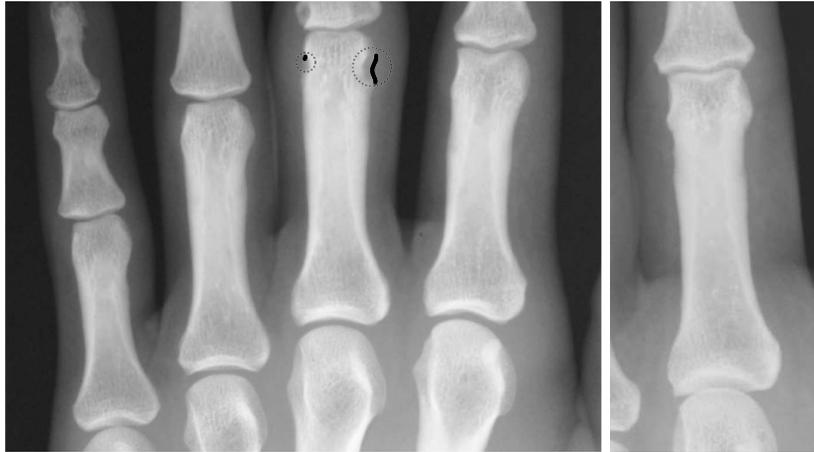


Fig. 6. Left: Unequivocal erosive destructions indicated by circles and automatic detection result (black lines) after analysis of PP2-PP5; For reference a healthy PP3 is depicted on the right.

Fig. 5 right shows an example of the ratio between the two posteriori probabilities $p(\mathbf{p}_i|\theta_{ERRO})$ and $p(\mathbf{p}_i|\theta_{NONERO})$ for a hand. The x-axis corresponds to the indices of the extracted texture patches. Note the two peaks at the locations of erosions indicated by arrows. Fig. 6 shows a radiograph with erosions automatically marked by the algorithm. The algorithm performs poor in cases when the texture is not sufficiently discriminative for a classification by the appearance model e.g. erosions that only diminish the contour contrast, or do not generate sharp structures within the bone. Radiologists classify these cases temporarily as *pre erosions*, and definite diagnosis is made during follow up.

4 Conclusion

In this paper a method for the automatic detection of erosive destructions caused by rheumatoid arthritis is proposed. The algorithm delineates bone contours on hand radiographs automatically and subsequently analyses the bone texture. By utilizing an appearance model that is learned from training examples the contour points are classified with respect to *erosion* and *non-erosion*. The procedure provides the medical expert with repeatable measurements. Thus inter- and intra-reader variation limiting therapy monitoring and multi center studies can be minimized. The algorithm has been evaluated on the proximal phalanges of 17 hand radiographs. An extension to other anatomical regions can be accomplished in a straightforward way, since both shape and appearance models are learned during a training phase.

Future work will focus on the enhancement of erosion detection in cases when the basic appearance model is not sufficient for reliable discrimination between anatomy and pathology. In the context of a current trial the method will be evaluated in a clinical setting, to assess performance w.r.t. human experts [16].

References

1. Sharp, J., Lidsky, M., Collins, L., Moreland, J.: Methods of scoring the progression of radiologic changes in rheumatoid arthritis. *Arthritis and Rheumatism* **14** (1971) 706–720
2. Larsen, A., Dale, K., Eek, M.: Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. *Acta Radiol Diagn* **18** (1977) 481–491
3. van der Heijde, D., Boonen, A., Boers, M., Kostense, P., van der Linden, S.: Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology* **38** (1999) 1213–1220
4. Angwin, J., Lloyd, A., Heald, G., Nepom, G., Binks, M., MF, M.J.: Radiographic hand joint space width assessed by computer is a sensitive measure of change in early rheumatoid arthritis. *Journal of Rheumatology* **31** (2004) 1050–1061
5. Genant, H., Jiang, Y., Peterfy, C., Lu, Y., Redei, J., Countryman, P.: Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs. *Arthritis Rheum* **41** (1998) 1583–1590
6. Sharp, J., Gardner, J., Bennett, E.: Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis. *Arthritis Rheum* **43** (2000) 1378–1386
7. Duryea, J., Jiang, Y., Zakharevich, M., Genant, H.: Neural network based algorithm to quantify joint space width in joints of the hand for arthritis assessment. *Med. Phys.* **27**(5) (2000) 1185–1194
8. Sailer, J., Scharitzer, M., Giurea, P.P.A., Imhof, H., Grampp, S.: Quantification of axial alignment of the lower extremity on conventional and digital total leg radiographs. *Eur Radiol* **15** (2005) 170–173
9. Keystone, E.C., Kavanaugh, A.F., Sharp, J.T., Tannenbaum, H., Hua, Y., Teoh, L.S., Fischkoff, S.A., Chartash, E.K.: Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: A randomized, placebo-controlled, 52-week trial. *Arthritis and Rheumatism* **50**(5) (2004) 1400–1411
10. B. Bresnihan, B., Newmark, R., Robbins, S., Genant, H.: Effects of anakinra monotherapy on joint damage in patients with ra. extension of a 24-week randomized, placebo-controlled trial. *Journal of Rheumatology* (2004)
11. Langs, G., Peloschek, P., Bischof, H.: ASM driven snakes in rheumatoid arthritis assessment. In: Proceedings of 13th Scandinavian Conference on Image Analysis, SCIA 2003, Goeteborg, Sweden. LNCS 2749, Springer (2003) 454–461
12. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Training models of shape from sets of examples. In: Proceedings of BMVC'92. (1992) 266–275
13. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: The use of active shape models for locating structures in medical images. *Image and Vision Computing* **12**(6) (1994) 355–366
14. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal on Computer Vision* **1** (1988) 321–331
15. Bishop, C.: *Neural Networks for Pattern Recognition*. Oxford Univ. Press (1995)
16. Sharp, J., Wolfe, F., Lassere, M., MaartenBoers, von der Heijde, D., Larsen, A., Paulus, H., Rau, R., Strand, V.: Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. *Journal of Rheumatology* **31**(6) (2004) 1062–1072

Quantitative Vertebral Morphometry Using Neighbor-Conditional Shape Models

Marleen de Bruijne¹, Michael T. Lund¹
László B. Tankó², Paola Pettersen², and Mads Nielsen¹

¹ IT University of Copenhagen, Denmark

² Center for Clinical and Basic Research, Ballerup, Denmark
marleen@itu.dk

Abstract. A novel method for vertebral fracture quantification from X-ray images is presented. Using pairwise conditional shape models trained on a set of healthy spines, the most likely normal vertebra shapes are estimated conditional on all other vertebrae in the image. The differences between the true shape and the reconstructed normal shape is subsequently used as a measure of abnormality. In contrast with the current (semi-)quantitative grading strategies this method takes the full shape into account, it uses a patient-specific reference by combining population-based information on biological variation in vertebra shape and vertebra interrelations, and it provides a continuous measure of deformity. The method is demonstrated on 212 lateral spine radiographs with in total 78 fractures. The distance between prediction and true shape is 1.0 mm for unfractured vertebrae and 3.7 mm for fractures, which makes it possible to diagnose and assess the severity of a fracture.

The full paper is available in the MICCAI proceedings:
R. Larsen, M. Nielsen, and J. Sparring, eds.
9th International Conference on Medical Image Computing & Computer-Assisted Intervention – MICCAI 2006, Copenhagen, Denmark, October 1-6
Lecture Notes in Computer Science, Springer, 2006

Automatic Cartilage Thickness Quantification using a Statistical Shape Model

Erik B. Dam^{1,2}, Jenny Folkesson¹,
Paola C. Pettersen², and Claus Christiansen²

¹ IT University of Copenhagen

² Center for Clinical and Basic Research

Abstract. Cartilage thickness is a central indicator for monitoring osteoarthritis (OA) progression. We present a novel, automatic method for quantification of cartilage thickness from magnetic resonance imaging. First, an automatic voxel classification produces a binary segmentation of the cartilage sheet. Second, a statistical shape model is deformed into the binary segmentation. Finally, the thickness map is extracted from the shape model.

We evaluate the cartilage thickness quantification on a collection of knee scans with both healthy and OA subjects. The method shows high reproducibility and proves to be able to separate healthy from OA subjects.

1 Introduction

Osteoarthritis (OA) is a major health problem for the elderly population [1]. OA causes pain, reduced range of motion, and eventually disability — all factors reducing quality of life. Currently, no treatment shows consistent, document effect on OA [2]. A central problem in developing new treatments is the lack of convincing quantification methods with high accuracy, precision and sensitivity.

A central process in OA progression is cartilage breakdown. During the early stages of OA, local cartilage lesions cause local swelling and then thinning, eventually leading to holes in the cartilage. In severe OA, large areas of cartilage are entirely missing. Cartilage volume and thickness are obvious measures for quantifying this progression. Volume can capture the overall progression, whereas thickness also has a potential for capturing the local lesions in the earlier stages. In addition, analysis of thickness maps also gives insight into the localizations.

Much research has been devoted to quantifying this progression from radiographs. This has obvious limitation due to the loss of information in the 2D projection combined with the fact that cartilage is not visible in X-ray. We focus on quantification from Magnetic Resonance Imaging (MRI) which allows non-invasive visualization of the cartilage [3] and direct cartilage assessment [4].

MRI sequences with near-isotropic voxels from a 3D sequence are best suited for cartilage quantification [5] and for 3D modeling in general, giving voxel sizes of 0.5mm or larger. Thickness quantification is therefore challenging since even healthy cartilage is only 2-3mm thick. This inspires use of statistical, smooth

shape models in order to allow robust segmentation and to avoid voxelation effects. A major challenge is then the shape variability of the cartilage sheet for severe OA with topology changes (holes) and large areas missing. The method we propose is entirely automatic and consists of:

- Voxel classification of the knee cartilage resulting in a binary segmentation based on the method in [6]. This step is reviewed briefly in section 3.
- Deformation of a statistical cartilage shape model in order to give a regularized representation of the cartilage segmentation. The shape deformation is described in section 4 with a brief review of the method for building the statistical shape model [7, 8].
- Extraction of the thickness map from the shape model. Since we are using a medial shape model, this step is quite simple. However, in section 5 special attention is put on the potential problems with holes in the cartilage sheets.

We evaluate the thickness quantification method on a collection of 114+31+25 knee MRI (see section 2 for details). The evaluation in section 6 focuses on reproducibility and ability to distinguish healthy from OA test subjects.

1.1 Related Work

Somewhat surprisingly, cartilage volume is relatively poor for quantification of OA progression and in some studies there is even no detectable volume change over time for OA patients [9] and [10]. Methods for quantification of thickness show more promise. In [11], a statistical shape model of the underlying tibial bone gives a cartilage coordinate map. From manual segmentations they then produce a thickness map. Even if the method still requires manual segmentation of both bone and cartilage, the basic approach is very sound. In [10] they are able to detect significant thickness loss on OA patients over a six month period.

A number of semi-automatic thickness quantification methods have been published. Some methods are essentially 2D such as the slice-wise active contour approach in [12] and the B-spline approach in [13]. In the latter they evaluate the accuracy of the thickness quantification for the shoulder by comparing with measurements on extracted cartilage specimens and get differences between 15% and 20%. In [14], they evaluate the B-spline approach from [15] which requires 2.5 hours of interaction per knee joint and get an inter-observer variability of around 7% for the mean thickness.

In [16], from an automated 3D approach that requires around 10 minutes of manual corrections of the segmentations per knee, they also produce cartilage thickness maps from segmentations of both bone and cartilage similar to [11]. Their limited validation indicates that the OA subjects have thinner cartilage.

Finally, an almost fully automatic method based on a graph searching segmentation algorithm [17] followed by mean thickness quantification is evaluated on ankle joints in [18]. The evaluation on 8 cadaveric ankles show accurate thickness measurements. Presumably the method could also be used for knees.

To the best of our knowledge, no evaluation of another fully automatic method for knee cartilage thickness quantification showing difference between OA patient and healthy has been published yet.

2 Data Collection

We evaluate our thickness quantification method on a collection of knee MRI prospectively acquired on an Esaote C-Span low-field 0.18 T scanner dedicated to imaging of extremities using a sagittal Turbo 3D T1 sequence (flip angle 40° , TR 50ms, TE 16ms) with a voxel size of $0.70 \times 0.70 \times 0.78mm^3$. Approximate scan time is 10 minutes.

The collection includes both left and right knees — right knee scans are reflected in order to apply the same methodology to all scans. The test subjects were males and females of ages between 21 and 72 years with no to quite severe OA symptoms (scores 0 to 3 on the Kellgren and Lawrence scale (KL) [19]). The collection contains 25 knee we use for training of the automated methods and 114 knees for evaluation. Among the 114, 31 knees were rescanned a week later. For all scans, the cartilage was segmented by a radiologist by slice-wise outlining (see figure 2). We focus on the medial compartment of the tibial cartilage sheet since the correlation between degradation and clinical symptoms is predominant in the medial compartments [20].

3 Automatic Cartilage Classification

The fully automatic classification step is based on previous work [6] and is a k Nearest Neighbor (kNN) voxel classifier trained on the manual segmentations from the 25 training set scans. A feature selection scheme selects a feature vector of approximately 50 features from a collection of potential features consisting of voxel position, intensity, Gaussian derivatives up to order three, eigenvalues and eigenvectors for the Hessian, and eigenvalues and eigenvectors for the Structure Tensor — all except position evaluated at multiple scales.

The segmentation results in classification of the medial compartments of tibial and femoral cartilage (an example tibial segmentation is in figure 2).

4 Automatic Shape Model Deformation

In order to regularize the voxel classification results such that thickness measurements at sub-voxel accuracy is possible, we use a deformable shape model. Since the aim is thickness measurements, a medial model seems appropriate.

We use the m-rep as our smooth shape representation [21]. The m-rep is a sampled medial sheet composed on medial atoms defined as a 4-tuple $\mathbf{m} = \{\mathbf{x}, r, \mathbf{F}, \theta\}$, consisting of: $\mathbf{x} \in \mathbb{R}^3$ and $r \in \mathbb{R}^+$, the center and radius of the sphere, $\mathbf{F} \in \mathbf{SO}(3)$ an orthonormal local frame, and $\theta \in [0, \pi)$ the object angle determining the angulation of the two implied opposing boundary points to the local frame. A subdivision surface method interpolates the implied boundary points and produce a smooth surface. See figure 2 for an example m-rep.

The m-rep model \mathbf{M} is deformed into a binary image I (a cartilage segmentation) by optimizing the objective function:

$$F(\mathbf{M}, I) = L(\mathbf{M}, I) + \alpha G(\mathbf{M}) + R(\mathbf{M}) \quad (1)$$

The *image match*, L , is the distance from the model boundary to the cartilage segmentation boundary (calculated by a Danielsson distance transform [22]) integrated over the implied model boundary. The *geometric typicality*, G , is a shape prior given by the Mahalanobis distance from a statistical shape model. The *model regularization*, R , is an addition to the standard m-rep framework ensuring regularization of the model coordinate system and model boundary [8].

The model is initialized by translation and scaling to fit mean and volume of the classification. Then F is optimized by conjugated gradient descent where global similarity transformations and deformations given by the cartilage model principal modes of variation are allowed. Finally, the parameters $\{\mathbf{x}, r, \mathbf{F}, \theta\}$ are optimized for each atom to fine-tune boundaries and thickness locally.

The statistical shape model used in the geometric typicality, G , is constructed in a bootstrap shape model building framework [7]. Here a generic, flat model is first deformed into each training shape using equation 1 with $\alpha = 0$. Then a mean model with modes of deformation is extracted using principal geodesic analysis [23]. This provisional statistical shape model is then used as the starting point for a new fitting of the training shapes, and after a few iterations of fitting and extraction of mean model and main modes of variation, the process converges to the final statistical shape model. We build our tibial cartilage sheet shape model from the 25 knee scans also used for training of the automatic segmentation.

5 Cartilage Thickness Map

It is trivial to construct a thickness map from a medial model since the model has pairs of opposing boundary points with coordinates given by the shape model. However, special attention is due to holes and gaps in the cartilage sheets.

The image match function, L , is the distance from the shape model boundary to the cartilage boundary in the binary image. This will not allow the shape model to “crawl” into the ends of thin shapes since there is a shape boundary closer than the end. These model boundary parts (red in figure 1) can be detected as points where the shape model boundary normal is approximately perpendicular to the cartilage boundary distance map gradient and requires addition of a *Crawl* term to the basic shape distance transform $Dist$ in the formulation of L that integrates over the model boundary $\mathcal{B}(\mathbf{M})$ using a parameterization s :

$$L(\mathbf{M}, I) = \int_{\mathcal{B}(\mathbf{M})} (1 - isRed(s)) |Dist(\mathbf{mb}(s))| + isRed(s) Crawl(s) ds$$

$$isRed(s) = \min(1, (\max(0, 2 - \frac{1}{\theta_{cr}} \left| \text{angle}(\mathbf{mbn}(s), \mathbf{g}(s)) - \frac{\pi}{2} \right|)))$$

$$Crawl(s) = -sign(Dist(\mathbf{mb}(s))) \frac{\partial \mathbf{mb}(s)}{\partial t} \cdot \mathbf{mbn}(s)$$

Here $isRed \in [0, 1]$ is 1 if the model boundary normal \mathbf{mbn} is less than θ_{cr} from being perpendicular to the gradient \mathbf{g} to the cartilage boundary distance map $Dist$ and goes linearly down to 0 at $2\theta_{cr}$. We use $\theta_{cr} = 0.4$.

In *Crawl* the parameter t in the partial derivative on \mathbf{mb} is an optimization evolution parameter — making the model boundary move in the normal direction. The sign of the shape distance function ensures that the term applies both for model boundary points inside the shape crawling into protrusions and for points outside crawling into indentations. Thereby, this additional term allows modeling of thin shapes as well as shapes with holes.

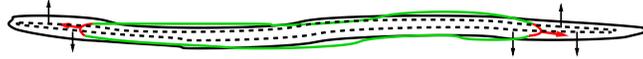


Fig. 1. For end model boundary points (red), the shape boundary (black) distance function (dotted iso-distance curve) offers no further help for the deformation of the model. The model boundary points at the ends with a normal approximately perpendicular to the shape boundary distance gradient are assigned an outwards force in the optimization that allows the model to crawl to the shape end. The principle is identical in 3D.

This crawling/digging is crucial for modeling thin cartilage sheets with holes and gaps. The statistical shape model ensures that the shape model keeps the topology of healthy cartilage — while the digging ensures that holes are actually modeled as “very thin” parts of the shape. Thereby the correspondence is preserved while allowing a cartilage thickness map with zero thickness. Another behavior encountered in knees with severe osteoarthritis is that the tibial cartilage is all gone in the lateral side (away from the center of the knee). By using a statistical shape model trained mainly on healthy cartilage we ensure that the shape model is biased towards representing the full cartilage sheet — with zero thickness at the center of the gap and gradually increasing thickness from there due to the regularization. Figure 2 illustrates both effects.

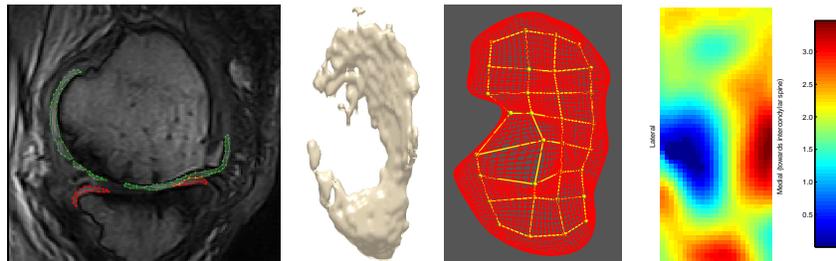


Fig. 2. The thickness map for a knee with severe OA ($KL=3$) showing osteophytes and holes in the cartilage. A sagittal slice from the knee MR with manual outlines of tibial cartilage in red and femoral cartilage in green (left). Next, the corresponding automatic segmentation seen from above with a central part of the cartilage missing. The m-rep shape model, with the grid of medial atoms in yellow and implied boundary mesh in red, captures a probable, full cartilage sheet (third). The thickness map shows severe thinning, with thickness 0mm in the gap center.

6 Evaluation

The automatic segmentation step followed by the automatic shape model fitting step provides a cartilage thickness map. Here we evaluate the mean thickness as an osteoarthritis progression quantification measure. Our mean thickness is measured across the cartilage sheet excluding the rim (also done in [10, 18]). In section 6.1 we look further into focal measures.

We evaluate the thickness quantification for reproducibility using the 31 scan-rescan pairs. Furthermore, we evaluate the ability to capture progression in OA by testing whether the healthy subjects have significantly thicker cartilage than the OA subjects according to a t-test. The evaluation results are in figure 3.

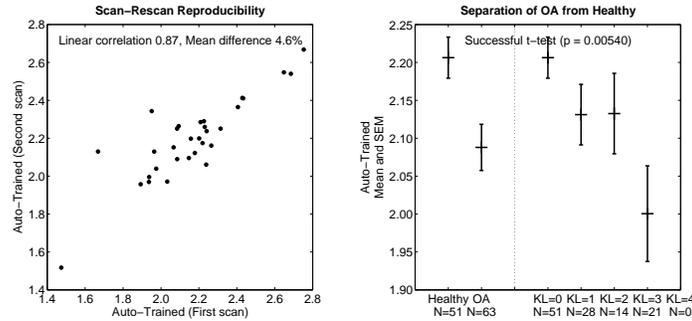


Fig. 3. The reproducibility of the thickness quantification method is evaluated on the 31 scan-rescan pairs (left). The OA subjects have on average thinner cartilage, 2.09mm compared to 2.21mm for healthy (statistically significant, $p < 0.01$) (right, left of dotted line). The mean thickness goes down to 2.00mm for the KL 3 group (right, right of dotted line). The cartilage thickness is normalized by the width of the tibial plateau.

The evaluation shows that our thickness quantification is very reproducible and able to separate healthy from OA (significant, $p < 0.01$). The mean thickness for healthy is 2.21mm and for the OA group 2.09mm. There is also a progression of cartilage thinning with OA progression from Kellgren & Lawrence index 0 to 3. The table below compares these results to volume quantification.

Quantification	Reproducibility		Separation
	Correlation	Mean diff	T-test p value
Volume manual segmentation	0.82	11.1%	0.002
Volume automatic segmentation	0.75	10.1%	0.001
Mean thickness using shape model	0.87	4.6%	0.005

While both volume and thickness quantification allow separation of the groups of healthy and OA subject ($p < 0.01$), the thickness quantification is far more reproducible. The statistical shape model proves to add robustness to the thickness quantification. However, equally interesting, the thickness map allows investigation of the focal distribution of the progression of OA.

6.1 Thickness Map Focal Statistics

The thickness map defined by the shape model allows focal statistics across the test population. Figure 4 shows some examples. The mean and standard deviation maps look more or less as expected. The cartilage is thickest towards the intercondylar spine, and the largest variation is in the load bearing region. Also, fairly large variation is seen at the outer edge of the thickness maps. This could be due to the cartilage gaps that occur there in severe OA.

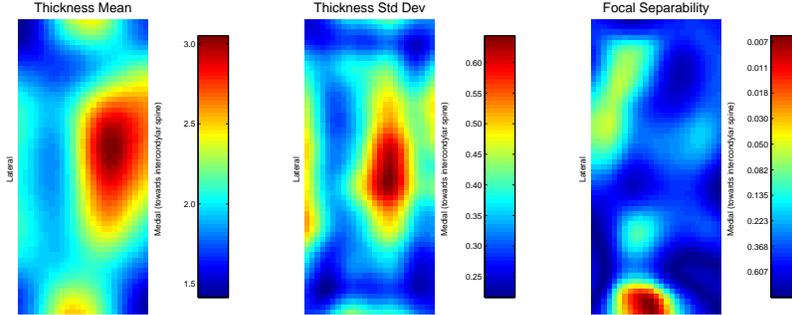


Fig. 4. Statistics across the 114 evaluation scans for each position in the shape model. The mean thickness map (left) shows up to 3mm cartilage on average close to the intercondylar spine. The thickness standard deviation map (center) shows largest variation in the load bearing region as expected. The separation map (ability to separate OA from healthy given as t -test p values) is harder to interpret.

The separation map is less clear with no obvious anatomical interpretation. One could expect that the load-bearing region would be the region that separates OA from healthy subject. This is not the case. The reason for this could be that the correspondence given by the shape model is flawed — but the intuitively nice maps for mean and standard deviation indicate that this is not the case. Alternatively, the reason could be that progression of KL score is more related to local cartilage lesions scattered across the sheet than to gradual thinning in a specific region. Future research will look into this.

7 Conclusion

We present a novel, fully automatic method for measuring thickness of tibial cartilage based on deformation of a statistical shape model into the result of a voxel classification. The use of the statistical shape model ensures the topology of healthy cartilage while the shape deformation allow zero thickness within the sheet giving proper modeling of gaps in the cartilage sheet. We evaluate the method for quantification of progression of OA. The method is able to separate healthy from OA subjects ($p < 0.01$) and has a very high reproducibility (the precision of 4.6% corresponds to 0.09mm compared to the voxel size of 0.7mm).

Thereby, the method looks promising for use in clinical studies. Future work will involve evaluation of the method in longitudinal studies.

References

1. Felson, D., Zhang, Y., Hannah, M., Naimark, A., Weissman, B., Aliabadi, P., Levy, D.: The incidence and natural history of knee osteoarthritis in the elderly. the framingham osteoarthritis study. *Arthritis and Rheumatism* **38** (1995)
2. Altman, R.: Measurement of structure (disease) modification in osteoarthritis. *Osteoarthritis and Cartilage* (2004)
3. Graichen, H., Eisenhart-Rothe, R.V., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. *Arthritis and Rheumatism* **50** (2004)
4. Pessis, E., Drape, J.L., Ravaud, P., Chevrot, A., Ayrat, M.D.X.: Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and mri. *Osteoarthritis and Cartilage* **11** (2003)
5. Xia, Y.: The total volume and the complete thickness of articular cartilage determined by mri. *Osteoarthritis and Cartilage* **11** (2003)
6. Folkesson, J., Olsen, O.F., Dam, E.B., Pettersen, P.C., Christiansen, C.: Combining binary classifiers for automatic cartilage segmentation in knee mri. In: *ICCV, Computer Vision for Biomedical Image Applications*. (2005)
7. Dam, E.B., Fletcher, P.T., Pizer, D.S.M., Tracton, G., Rosenman, D.J.: Prostate shape modeling based on principal geodesic analysis bootstrapping. In: *Proceedings of MICCAI 2004*. Volume 3216–3217 of LNCS., Springer (2004)
8. Dam, E.B., Fletcher, P.T., Pizer, S.M.: Automatic shape modeling based on principal geodesic analysis bootstrapping. *Medical Image Analysis* (2006) In review.
9. Gandy, S., Dieppe, P., Keen, M., Maciewicz, R., Watt, I., Waterton, J.: No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. *Osteoarthritis and Cartilage* (2002) 929–937
10. Williams, T., Holmes, A., Maciewicz, R., Waterton, J., Taylor, C., Creamer, P., Nash, A.: Cartilage loss in osteoarthritis detected by statistical shape analysis of magnetic resonance images. *Osteoarthritis and Cartilage* **13, supplement A** (2005)
11. Williams, T., Taylor, C., Gao, Z., Waterton, J.: Corresponding articular cartilage thickness measurements in the knee joint by modelling the underlying bone. In: *MICCAI*. Number 2879 in LNCS (2003)
12. Raynauld, J.P., Kauffmann, C., Beaudoin, G., Berthiaume, M.J., de Guisei, J.A., Bloch, D.A., Camacho, F., Godbout, B., Altman, R.D., Hochberg, M., Meyer, J.M., Cline, G., Pelletier, J.P., Martel-Pelletier, J.: Reliability of a quantification imaging system using magnetic resonance images to measure cartilage thickness and volume in human normal and osteoarthritic knees. *Osteoarthritis and Cartilage* **11** (2003)
13. Graichen, H., Jakob, J., von Eisenhart-Rothe, R., Englmeier, K.H., Reiser, M., Eckstein, F.: Validation of cartilage volume and thickness measurements in the human shoulder with quantitative magnetic resonance imaging. *Osteoarthritis and Cartilage* **11** (2003)
14. Koo, S., Gold, G., Andriacchi, T.: Considerations in measuring cartilage thickness using mri: factors influencing reproducibility and accuracy. *Osteoarthritis and Cartilage* **13** (2005)
15. Stammberger, T., Eckstein, F., Englmeier, K., Reiser, M.: Determination of 3d cartilage thickness data from mr imaging: computational method and reproducibility in the living. *Magn Reson Med* **41** (1999)
16. Barbu-McInnis, M., Tamez-Pena, J.G., Totterman, S.: Focal cartilage defect progression detection: Measurement of precision and variation in natural characteristics of cartilage thickness maps derived from 3d mri data. In: *IEEE Int Conf on Image Processing*. (2004)
17. Li, K., Millington, S., Wu, X., Chen, D.Z., Sonka, M.: Simultaneous segmentation of multiple closed surfaces using optimal graph searching. In: *Information Processing in Medical Imaging: 19th International Conference*. Volume 3565 of LNCS., Springer (2005)
18. Millington, S., Li, K., Wu, X., Hurwitz, S., Sonka, M.: Automated simultaneous 3d segmentation of multiple cartilage surfaces using optimal graph searching on mri images. *Osteoarthritis and Cartilage* **13, supplement A** (2005)
19. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* **16** (1957)
20. Dunn, T., Lu, Y., Jin, H., Ries, M., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. *Radiology* **232** (2004)
21. Joshi, S., Pizer, S., Fletcher, P.T., Yushkevich, P., Thall, A., Marron, J.S.: Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *Transactions on Medical Imaging* **21** (2002)
22. Danielsson, P.E.: Euclidean distance mapping. *Computer Graphics and Image Processing* **14** (1980)
23. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: *Proc. of Computer Vision and Pattern Recognition*. (2003)

MR image segmentation using phase information and a novel multiscale scheme

Pierrick Bourgeat¹, Jurgen Fripp^{1,3}, Peter Stanwell², Saadallah Ramadan²,
and Sébastien Ourselin¹

¹ BioMedIA Lab, Autonomous Systems Laboratory, CSIRO ICT Centre, Sydney,
Australia.

{[pierrick.bourgeat](mailto:pierrick.bourgeat@csiro.au), [jurgen.fripp](mailto:jurgen.fripp@csiro.au), [sebastien.ourselin](mailto:sebastien.ourselin@csiro.au)}@csiro.au

² Institute for Magnetic Resonance Research, Sydney, Australia.

{[stan](mailto:stan@imrr.usyd.edu.au), [saad](mailto:saad@imrr.usyd.edu.au)}@imrr.usyd.edu.au

³ School of ITEE, University of Queensland, Australia.

Abstract. This paper considers the problem of automatic classification of textured tissues in 3D MRI. More specifically, it aims at validating the use of features extracted from the phase of the MR signal to improve texture discrimination in bone segmentation. This extra information provides better segmentation, compared to only using magnitude features. We also present a novel multiscale scheme to improve the speed of pixel-wise based classification algorithm, such as support vector machines. This algorithm dramatically increases the speed of the segmentation process by an order of magnitude through a reduction of the number of pixels that needs to be classified in the image.

The full paper is available in the MICCAI proceedings:

R. Larsen, M. Nielsen, and J. Sporring, eds. 9th International Conference on Medical Image Computing & Computer-Assisted Intervention - MICCAI 2006, Copenhagen, Denmark, October 1-6 Lecture Notes in Computer Science, Springer, 2006

Automatic Quantification of Cartilage Homogeneity

Arish Asif Qazi¹, Ole Fogh Olsen¹, Erik Dam^{1,2}, Jenny Folkesson¹, Paola Pettersen², and Claus Christiansen²

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

² Center for Clinical and Basic Research, Ballerup, Denmark.

Abstract. Osteoarthritis (OA) is a degenerative joint disease that involves the wearing down of the articular cartilage. A typical problem has been quantifying progression and early detection of the disease. In this study we develop a fully automatic method for investigating knee cartilage homogeneity on 114 manually and automatically segmented T1 knee MR images from subjects with no, mild or severe OA symptoms. To measure homogeneity we characterize the tibial and femoral compartments in each cartilage by several statistical measures and then evaluate their ability to quantify OA progression. The discriminatory power of each measure for separating the group of healthy subjects from group of subjects having OA is tested statistically. Our method outperforms a standard measure like volume in separating healthy subjects from subjects having OA. We show that our method is reproducible through a scan-rescan evaluation from additional 31 MR images.

1 Introduction

The majority of the elderly population will at some point encounter osteoarthritis (OA) resulting in pain and reduced range of motion in mainly the knee and hip joints. For the worst cases, joint replacement surgery is even required. Currently, available treatments are directed towards relief of symptoms and at present no drugs have shown to consistently modify joint structure or reverse joint pathology [1]. New, accurate and precise methods are needed in order to quantify the disease progression in clinical studies that determines the effect of potential treatments.

A hallmark process in OA is cartilage breakdown. Therefore, typical disease progression quantifications have traditionally been:

- Joint gap (between the femur and tibia on the knee joint) is an indirect measure of cartilage thinning from radiographs (X-ray) where the cartilage itself is not visible [2].
- Cartilage volume and thickness measured from magnetic resonance imaging (MRI) where the cartilage is visible [3, 4].

However before thinning even begins, the cartilage loses its firm structure [5]. This structure is composed by the three layers of collagen fibers that allows

absorption/expulsion of water and thereby ensures shock absorption. When this structure is breaking down, the first stage is swelling of the cartilage which is followed by thinning in the later stages. Therefore, measures focusing on volume or thickness may not be adequate to capture the early stages of OA.

The early loss of integrity could ideally be measured directly by quantifying the decrease in alignment of the collagen fibers in the three layers of the cartilage. In the future, this will possibly be done through analysis of very high resolution diffusion tensor MRI [6]. Due to limitations in resolution and acquisition times, we settle for analysis of the intensities observed in regular MRI as a first step. Instead of measuring integrity of cartilage layer alignment directly, we therefore quantify cartilage "homogeneity".

In this paper, we investigate a number of measures for measuring cartilage homogeneity and evaluate their ability to quantify OA progression. Since there is no ground truth for cartilage homogeneity available, we cannot evaluate accuracy directly. Possibly, a validation could be performed based on histological slices of cartilage. However, due to the highly invasive nature of this, it is not suitable as a standard in vivo evaluation technique.

We therefore evaluate accuracy indirectly by investigating the ability of the measures to separate healthy knees from knees with some degree of OA from a collection of 114 knees. We evaluate the robustness of the method by comparing the measures obtained from both manual and automatic segmentations of the knees. Finally we evaluate the reproducibility of the method through a scan-rescan evaluation where measures obtained from 31 pairs of scans acquired a week inbetween are compared. The investigation shows that a subset of the proposed measures separates healthy from OA more accurately than volume. Furthermore they are able to detect early stages of OA.

1.1 Related Work

Several studies have shown correlation between OA progression and the mean intensities from T2 MR scans [7]. Analysis of structure has previously been applied to anatomical structures other than the cartilage. For example the structure of trabecular bone can be quantified by fractal signature analysis [8]. Texture analysis is also used for osteoporosis progression analysis in the form of Minkowski functionals.

The thin cartilage structure with a typical thickness of only a few voxels complicates the analysis. However, a few studies have done early feasibility studies on the use of diffusion tensor MRI for cartilage analysis. In a small in vitro study, it has been shown that the eigenvector orientations of the local diffusion tensor can separate cartilage layers that to some degree reflects the alignment of the collagenous fiber network in the articular cartilage [6].

2 Method

Image acquisition was done on a Esaote C-Span lowfield 0.18 T scanner dedicated to imaging of extremities acquires Turbo 3D T1 scans (40° flip angle,

T_R 50 ms, T_E 16 ms). The scans are made through the sagittal plane with the resolution $0.7 \times 0.7 \times 0.8 \text{ mm}^3$. The dimensions of all the scans are $256 \times 256 \times 104$ voxels. The MR scans have been manually segmented by radiologists. Figure 2 illustrates a slice in which the tibial and femoral medial cartilage has been manually segmented.

The 114 scans in the data set are of both left and right knees. A week later 31 knees were rescanned for the precision and reproducibility study. The test subjects are both males and females aged between 21 and 72 years. They have no, mild or severe OA symptoms, diagnosed by radiologists as being 0,1,2 and 3 on the Kellgren and Lawrence Index [9]. Among the 114 test subjects 51 are healthy (KL:0) and 63 have OA (KL:1-3).

2.1 Cartilage Homogeneity

We define cartilage homogeneity to be a measure of the variation of the intensities inside the cartilage compartment. To quantify homogeneity we calculate a number of measures based on two different types of statistical methods.

First Order Statistical Methods

The first order statistical methods rely on approximating the probability of observing a particular intensity at a randomly chosen location in the image. The measures are calculated from the gray scale histogram of the image defined by: $H(i) = \frac{n_i}{N}$; $i = 0, 1, \dots, L - 1$ where N is the number of pixels in the image, n_i is the number of occurrences for intensity i and L is the number of gray levels in the image. Using the equation for histogram we define the following measures:

Mean (average intensity value): $\mu = \sum_{i=0}^{L-1} iH(i)$

Standard Deviation (contrast of image): $\sigma = \sqrt{\sum_{i=0}^{L-1} (i - \mu)^2 H(i)}$

Uniformity (energy of image): $\sum_{i=0}^{L-1} H(i)^2$

Entropy Normalized (randomness): $-\sum_{i=0}^{L-1} \frac{H(i) \log H(i)}{\log(q)}$

When measuring entropy and uniformity we smooth the histogram by decreasing the number of bins. $H(i)$ then corresponds to the number of occurrences of intensities within a range defined by the bin width. The bin width is determined by the Freedman-Diaconis [10] rule and is chosen to be 100. Furthermore to ensure invariance to change in the image quantization levels we normalize entropy by \log_2 (base 2) of the quantization levels ($q = 4096$).

Second Order Statistical Methods

The second order statistical methods measure the joint probability distribution of pairs of voxels in the image. This involves the estimation of the discrete

³ We define $0 \log(0) = 0$

second-order probability function $C_{d\theta}(i, j)$ which represents the probability of occurrence of a voxel pair with gray levels i and j given the spacing between the pair of voxels is d along a given direction θ . $C_{d\theta}(i, j)$ is called the gray level co-occurrence matrix (GLCM) [11]. Due to the small size of the cartilage region we only consider the immediate neighboring voxels. Furthermore to reduce the dimensionality we assume that the joint probabilities are direction independent. Therefore we have an averaged $L \times L$ GLCM where L is the number of gray levels in the image.

In order to quantify this spatial dependence of gray level values, Haralick [11] suggests 14 measures which are extracted from the GLCM. A subset of the four most relevant measures is chosen:

$$\text{Contrast: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 C[i, j] \quad \text{Homogeneity: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{C[i, j]}{1+|i-j|}$$

$$\text{Correlation: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{(i-\mu)(j-\mu)C[i, j]}{\sigma^2} \quad \text{Energy: } \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} C[i, j]^2$$

3 Results

3.1 Manual Segmentations

We have a total of 114 manually segmented knee MR Images. For each knee we have the segmentations both for the tibial and the femoral medial cartilage. The values of the measures for both tibial and femoral medial cartilage in each image are calculated and the measures are grouped according to the KL values of the subjects: being healthy or having OA. To determine the amount of discrimination a measure provides we test the null hypothesis that the two distributions (healthy and OA) have the same mean using the t-test. The resulting p -value is the probability with which the hypothesis can be rejected. If this p -value is less than a chosen level of significance α then we will reject the null hypothesis. Using $\alpha = 0.05$ the following measures succeeded to discriminate health from OA: standard deviation, uniformity and entropy from first order statistics and contrast from second order.

Table 1 lists the p -values for the hypothesis testing of all the first and second order measures calculated from manual segmentations of the tibial medial cartilage. The p -values are low suggesting that the measures can significantly discriminate healthy from OA. Volume of the knee cartilage is considered a good measure for monitoring OA progression [12]. To assess and compare the discriminatory power of our measures we also include p -values for volume (as in [4] volume of each cartilage is normalized by width of the subject's tibial plateau). Figure 1 shows the comparison between volume and entropy in capturing OA progression (due to space constraints we only show the comparison between volume and entropy). We can see that entropy ($p = 0.0005$) outperforms volume ($p = 0.002$) in separating healthy from OA. Furthermore it can also detect early stages of OA. That is by separating KL:0 from KL:1 ($p = 0.001$).

The first row of Table 2 shows the results for manual segmentations of the femoral medial cartilage. The p -values are high which means that the method

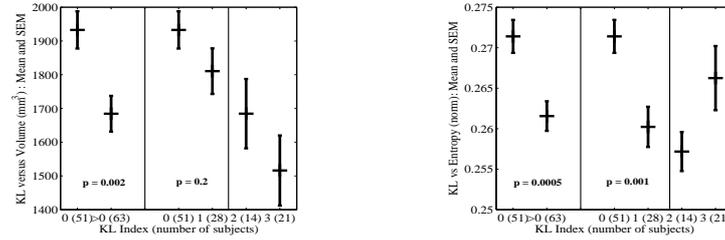


Fig. 1. Comparison of volume (left) and entropy (right) as a separator for healthy versus OA for manual segmentations of the tibial medial cartilage. The first part of each graph shows the error bars (which represent Standard Error of the Mean (SEM)) for healthy (KL:0) versus OA (KL:0-3). The second part shows the error bars for each group (KL:0-3) separately. Entropy ($p = 0.0005$) outperforms volume ($p = 0.002$) in separating healthy from OA and also in detecting early stages of OA (KL:0-KL:1).

Table 1. P -Values for the hypothesis testing of all the first (left column) and second order (right column) measures calculated from manual segmentations of tibial medial cartilage.

Measure	P -Value	Measure	P -Value
Volume	0.002	GLCM:Contrast	0.007
Mean	0.1	GLCM:Corrleation	0.8
Standard deviation(σ)	0.0006	GLCM:Energy	0.3
Uniformity	0.002	GLCM:Homogeneity	0.1
Entropy	0.0005		

fails to discriminate. [13] shows that accuracy of the thickness measurements from MR images is better in the weight bearing regions of femoral. It is because the weight bearing region of the femur sustains contact with the tibial cartilage during the gait cycle. The load bearing region is approximated by the intersection between femoral cartilage and the result of applying a 3D morphological dilation operator to the tibial region 10 times.

Figure 2 shows a femoral medial cartilage and its segmented load bearing region. The second row of Table 2 shows the results for the load bearing region of the femoral medial region. All our results for the femoral cartilage in the subsequent sections are calculated using only the load bearing region.

3.2 Automatic Segmentations

For evaluating the robustness of our automatic method the same data set of 114 images is segmented automatically using the algorithm in [14]. The segmentation is based on an three class approximate k Nearest Neighbor (k NN) classification scheme. The p -values are low suggesting that our method is robust. For e.g. entropy ($p = 0.00004$) outperforms volume ($p = 0.001$) in separating healthy from OA. The first and second row of Table 3 shows the results for tibial and femoral (load bearing) medial regions respectively.

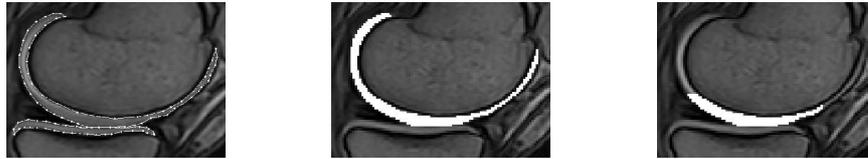


Fig. 2. A slice from a knee MR scan (left), in which the tibial medial and femoral medial cartilage is manually segmented by radiologists. Sagittal view of segmented femur (middle) and its load bearing region (right).

Table 2. *P*-Values for manual segmentations of femoral medial cartilage. The first row shows the values for the whole femur while the second row shows the values for the load bearing region.

Type	Volume	Std dev(σ)	Uniformity	Entropy	GLCM:Contrast
Whole femur	0.3	0.4	0.06	0.2	0.1
Load bearing	0.3	0.05	0.01	0.02	0.1

3.3 Reproducibility

To evaluate the reproducibility of the approach we have an additional set of 31 re-scanned manually segmented images. For each measure we calculate the mean percentage difference (%) and the correlation coefficient (r) with previous scans. Table 4 shows the results for manual and automatic segmentations of the tibial and the femoral medial cartilage respectively.

The correlation coefficient can be close to 1 even if there is a considerable variation in the data [15]. Another way to get a correct visual assessment of the relationship between two measures is to plot their differences against their mean. This resultant plot is known as a Bland-Altman plot [15]. According to Bland and Altman; if 95% of the differences lie within two standard deviations then we can conclude that the method is repeatable. Figure 3 shows a Bland-Altman plot for entropy of the manual segmentations of tibial and femoral medial cartilages. We can see that the percentage differences are very low. For entropy it is less than 5%. Therefore we can conclude that our method is highly reproducible.

Table 3. *P*-Values for automatic segmentations. The first row shows the results for tibial medial cartilage. The second row shows the results for femoral medial cartilage (load bearing region).

Type	Volume	Std dev(σ)	Uniformity	Entropy	GLCM:Contrast
Tibial	0.001	0.00004	0.0002	0.00004	0.0003
Femoral	0.004	0.02	0.003	0.002	0.1

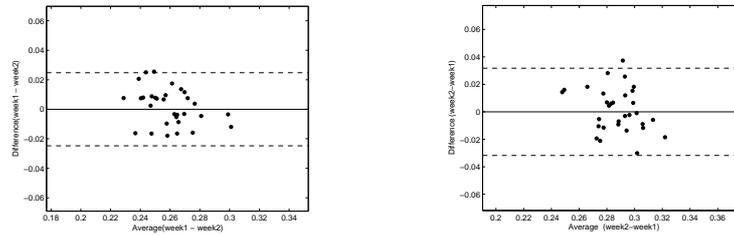


Fig. 3. Bland-Altman plot for entropy for manual segmentations of the tibial medial (left) and the femoral medial (right:load bearing region) cartilage. The dotted line represents two standard deviations of the difference.

Table 4. Reproducibility results (week1-week2) for manual (prefixed *m*) and automatic segmentations (prefixed *a*). The first two rows show the percentage differences (%) and correlation coefficients (*r*) for tibial medial cartilage. The next two rows show the results for femoral medial cartilage (load bearing region).

Type	Volume	Std dev(σ)	Uniformity	Entropy	GLCM:Contrast
m-Tibial (%)	11.1	9.3	9.0	4.3	21.9
m-Tibial (<i>r</i>)	0.82	0.82	0.78	0.78	0.81
m-Femoral (%)	9.6	12.7	12.3	4.8	32.1
m-Femoral (<i>r</i>)	0.91	0.6	0.63	0.64	0.55
a-Tibial (%)	10.1	8.1	8.3	3.6	16.1
a-Tibial (<i>r</i>)	0.75	0.81	0.8	0.78	0.86
a-Femoral (%)	45.1	11.7	12.3	4.8	27.5
a-Femoral (<i>r</i>)	0.86	0.65	0.62	0.64	0.63

4 Conclusion

The results of our evaluation show that the texture of the cartilage is not homogenous and can be a marker for disease progression. We use a fully automatic method to quantify homogeneity. We do this by characterizing the manual and automatic segmentations of 114 knee cartilages by several statistical measures. To determine if the measures can quantify OA progression we do a t-test to evaluate their ability to separate healthy from OA; from a subset of measures that succeed we choose entropy to be the measure that (tibial cartilage: man seg ($p = 0.0005$), auto seg ($p = 0.00004$)) can discriminate healthy from OA more confidently than a standard measure like volume [12] (tibial cartilage: man seg ($p = 0.002$), auto seg ($p = 0.001$)). Furthermore the measures are also able to detect early progression of OA. We also show that our method is reproducible (tibial: % ≤ 4.3 for entropy). Future work may involve:

- probing the surroundings of the cartilage.
- longitudinal studies.
- comparison with histological analysis of cartilage structure.

References

1. Altman, R.: Measurement of structure (disease) modification in osteoarthritis. *Osteoarthritis and Cartilage* **12** (2004) 69–76
2. Brandt, K., Fife, R., Braunstein, E.: Radiographic grading of the severity of knee osteoarthritis: relation of the Kellgren and Lawrence grade to a grade based on joint space narrowing, and correlation with arthroscopic evidence of articular cartilage degeneration. *Arthritis Rheum* **34** (1991) 1381–1386
3. Graichen, H., Eisenhart-Rothe, R., Vogl, T., Englmeier, K., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. *Arthritis and Rheumatism* **50** (2004) 811–816
4. Folkesson, J., Olsen, O., Dam, E., Pettersen, P., Christiansen, C., Alexandersen, P.: Robust volume estimation of articular cartilage from knee MRI scans. *Osteoarthritis and Cartilage* **13** (2005) 121–122
5. Hollander, A., Pidous, I., Reiner, A., Rorabeck, C., Bourne, R., Poole, A.: Damage to type II collagen in ageing and osteoarthritis starts at the articular surface, originates around chondrocytes, and extends into cartilage with progressive degeneration. *Clinical Investigation* **96** (1995) 2859–2869
6. Filidoro, L., Dietrich, O., Weber, J., Rauch, E., Oerther, T., Wick, M., Reiser, M., Glaser, C.: High-resolution diffusion tensor imaging of human patellar cartilage: Feasibility and preliminary findings. *Magnetic Resonance in Medicine* **53** (2005) 993–998
7. Naish, J.H., Vincent, G., Bowes, M., Kothari, M., White, D., Waterton, J.C., Taylor, C.: A method to monitor local changes in MRI signal intensity in articular cartilage: A potential marker for cartilage degeneration in osteoarthritis. *MICCAI* **3217** (2004) 959–966
8. Buckland-Wright, J., Lynch, J., Macfarlane, D.: Fractal signature analysis measures cancellous bone organisation in macroradiographs of patients with knee osteoarthritis. *Annals of Rheumatological Diseases* **55** (1996b) 749–755
9. Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthrosis. *Annals of the Rheumatic Diseases* **16** (1957)
10. Freedman, D., Diaconis, P.: On the histogram as a density estimator: l_2 theory. *Wahrsch. Verw. Geb* **51** (1981) 453–476
11. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3** (1973) 610–621
12. Jean, P., Johanne, M., Marie, J., Gilles, B., Denis, C., Boulos, H., Hyman, T., Joan, M., John, F., Gary, A., Jean, P.: Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. *Arthritis Research and Therapy* **8** (2005)
13. Koo, S., Gold, G., Andriacchi, T.: Considerations in measuring cartilage thickness using MRI: factors influencing reproducibility and accuracy. *Osteoarthritis and Cartilage* **13** (2005) 7825–789
14. Folkesson, J., Olsen, O., Dam, E., Pettersen, P., Christiansen, C.: Combining binary classifiers for automatic cartilage segmentation in knee MRI. *First International Workshop CVBIA* (2005) 230–239
15. Altman, D., Bland, J.: Measurement in medicine: the analysis of method comparison studies. *The Statistician* **32** (1983) 307–317

Knee Images Digital Analysis: a quantitative method for individual radiographic features of knee osteoarthritis.

Marijnissen A.C.A.¹, Vincken K.L.², Vos P.A.J.M.¹, Bijlsma J.W.J.¹, Bartels L.W.², and Lafeber F.P.J.G.¹.

¹Rheumatology & Clin. Immunology, F02.127, UMC Utrecht, PO Box 85500, 3508 GA Utrecht, the Netherlands

²Image Sciences Institute, Q0S.459, UMC Utrecht, PO Box 85500, 3508 GA Utrecht, the Netherlands

A.C.A.Marijnissen@umcutrecht.nl

Abstract. Objective evaluation of structural changes in osteoarthritis is essential for diagnosis and evaluation of disease progression. Since conventional radiography is still the gold standard, in the present study a newly developed digital method to analyze standard radiographs of knees was evaluated. Joint space width (JSW), osteophyte area, subchondral sclerosis, deviation of the angle of the joint, and eminentia height were measured using the interactive application Knee Images Digital Analysis (KIDA) on a standard PC. Enlargements on screen can be performed, when required. The application provides multiple measures for all parameters as continue variables. Two observers evaluated the radiographs on two different occasions with one-week interval. The observers were blinded to the source of the radiographs and their previous measurements. Intra- and interobserver variation was evaluated. The results demonstrate KIDA to be a reliable method to quantify and document (for follow-up) the radiographic parameters of knee osteoarthritis.

Introduction

Osteoarthritis (OA) is a degenerative joint disease characterized by damage of the articular cartilage, changes in the subchondral bone and secondary inflammation. Objective quantification of these structural changes is essential for diagnosis, evaluation of disease progression, and assessment of efficacy of treatment. Conventional radiography is still the gold standard for imaging the OA joint. However, reliable objective quantification of radiographs is still difficult. Grading systems have been developed for several joints on the basis of radiographically observed changes related to OA. The most frequently used grading system for measuring the severity of OA in several joints is the one of Kellgren and Lawrence¹, which emphasises the development of osteophytes. Up to now, only joint space width can be given as continuous variable. Therefore, in the present study a newly

developed digital method was evaluated to analyze different OA characteristics on standard radiographs of knees in a quantitative way.

Methods

Standardized radiographs of normal (n=10) and OA (n=55) knees were taken according to the semi-flexed method of Buckland-Wright² (55kV, 5mAs, FFD=1.0 m with the knee against the detector). Radiographs were taken with an aluminium step wedge alongside in order to quantify bone density and correct for possible magnification of the radiograph. Joint space width (JSW), osteophyte area, and subchondral sclerosis, all at several predefined locations; and deviation of the angle of the joint, and eminentia height were measured using the newly developed Knee Images Digital Analysis (KIDA) on a standard PC (Figure 1).



Fig. 1. Presentation of Knee Images Digital Analysis

The interactive computer measurement consists of 6 steps:

- 1) detection of the reference wedge, followed by calculation of the maximum value in the (linear) region of correct exposure in the characteristic curve of the X-ray film or detector.
- 2) indication of the edges of the joint necessary for further evaluation
- 3) defining the bone-cartilage interface at 4 points in the lateral and medial femur and tibia,
- 4) defining the top of the eminentia,
- 5) indicating the osteophytes at each of the compartments, and
- 6) indicating the minimal JSW.

Enlargements on screen can be performed, when required. Subsequently, the computer application provides multiple measures for joint space width (e.g. minimum, lateral, medial, average), for subchondral bone density (lateral, medial, average), height of the eminentia, joint angle deviation, and osteophyte areas (at four locations) as continue variables. The quantitative results can easily be statistically analyzed.

Two observers evaluated all radiographs on two different occasions with one-week interval. The observers were blinded to the source of the radiographs and their previous measurements. Intra- and inter-observer variation was evaluated using the statistical method of Bland and Altman³. Individual KIDA data were compared to the overall Kellgren & Lawrence grade. The Wilcoxon rank test was used to analyse differences between healthy and osteoarthritic knees. Spearman correlation was used for comparisons of individual KIDA parameters with each other and with the Kellgren & Lawrence grade.

Results

KIDA can be learned within an hour. A full evaluation including data storage requires less than 10 minutes per radiograph. Evaluation of KIDA shows small intra- and inter-observer variations for all relevant parameters. E.g. observer A found a minimum JSW of 2.8 ± 1.7 mm with a mean difference between the two observations of -0.02 mm (range $-0.8 - 0.59$) (see figure 2).

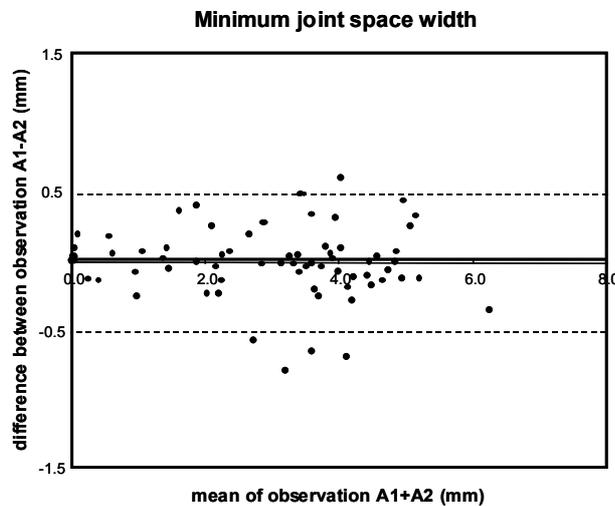


Fig. 2. Intra- and interobserver variations of individual parameters of Knee Images Digital Analysis according to the Bland and Altman method.

Observer A and B found a minimum JSW of 2.9 ± 1.7 mm with a mean difference between the two observations of -0.11 mm (range $-2.45 - 1.12$). A mean eminentia height of 11.7 ± 1.8 mm was measured with a mean difference between the two observers of -0.22 mm (range $-10.23 - 3.58$). Mean subchondral bone density was 29.7 ± 5.1 Alu Eq. with a mean difference between two observers of -0.11 Alu Eq. (range $-1.05 - 1.67$). Mean osteophyte area was 9.3 ± 6.4 mm² with a mean difference between two observers of 1.23 mm² (range $-5.37 - 13.39$), mean angle of the joint was $-2.5 \pm 2.7^\circ$ with a mean difference between two observers of -0.04 (range $-5.43 - 6.31$).

Significant correlations were found between subchondral sclerosis and minimum JSW ($R = -0.56$; $p < 0.001$) and between osteophyte formation and minimum JSW ($R = -0.55$; $p < 0.001$). In addition, significant correlations were found between individual KIDA parameters and the overall Kellgren & Lawrence osteoarthritis grade ($R = -0.57$ with $p < 0.0001$ for minimum JSW, $R = 0.27$ with $p < 0.02$ for subchondral sclerosis, and $R = 0.57$ with $p < 0.0001$ for osteophyte area).

Differences between OA and healthy individuals in individual parameters can be objectively measured using KIDA (Figure 3).

Healthy versus osteoarthritis

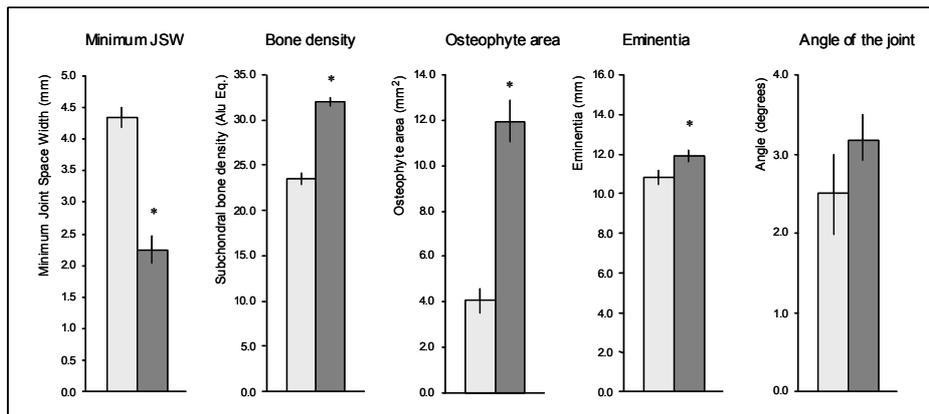


Fig. 3. Differences between healthy and osteoarthritis knees. Mean \pm SEM are depicted for healthy knees (light grey) and osteoarthritis knees (dark grey).

* Statistically significant differences in individual parameters between osteoarthritis and healthy knees are depicted.

Discussion

These results demonstrate KIDA to be a reliable method to quantify and document (for follow-up) the radiographic parameters of knee OA. Evaluation of sensitivity to changes is a next step. Data on joint space narrowing are identified to be ~0.2mm per year, which is measurable by the newly developed KIDA method. However, so far nothing is known about the rate of changes in subchondral bone density, osteophytes, eminentia and joint angle deviation in the OA process, since these parameters could not be quantified before. KIDA may be worthwhile to the diagnosis and evaluation of progression of disease, as well as to the evaluation of treatment efficacy in clinical trials.

References

1. Kellgren and Lawrence. *Ann Rheum Dis.* 1957; 494-501.
2. Buckland-Wright. *J. Rheumatol.* 1999; 2664-74.
3. Bland and Altman. *Lancet* 1986; 307-10.

This study was supported by the Dutch Arthritis Association

Improved Parameter Extraction From Dynamic Contrast-Enhanced MRI Data in RA Studies

Olga Kubassova¹, Roger D. Boyle¹, and Aleksandra Radjenovic²

¹ School of Computing, University of Leeds, UK, {olga,roger}@comp.leeds.ac.uk

² Academic Unit of Medical Physics, University of Leeds, Leeds General Infirmary, Leeds, UK, sasha@medphysics.leeds.ac.uk *

Abstract. A fully automated method for analysis of Dynamic Contrast Enhanced MRI scans of the metacarpophalangeal joints of the hand and assessing the extent and magnitude of inflammatory activity in rheumatoid arthritis is put forward. The method incorporates automated segmentation, spatial registration, and a modelling of the underlying physical procedure. It affords robust and consistent quantitation of the spatial and temporal properties of 4D datasets, allows a more robust and consistent extraction of various parameters, and provides information on procedure completion, hitherto unavailable, that is of value in guiding the procedure and informing the reliability of the parameter estimates. The technique is demonstrated on 10 DEMRI studies and has potential for wider application.

1 Introduction

Rheumatoid arthritis [RA] is an inflammatory disease associated with pathological alteration of microcirculation within the synovial lining of the diarthroidal joints such as metacarpophalangeal [MCP] joints of the hand. In Dynamic Contrast Enhanced MRI [DEMRI], temporal variation of MRI signal intensity occurs following intra-venous administration of the contrast agent diethylene triamine pentacetic acid [Gd-DTPA]. The time course of signal changes is proportional to the underlying changes in local bulk tissue concentration of Gd-DTPA, which in turn depends on the degree of inflammatory activity. Therefore, the extent of inflammation and monitoring treatment-induced changes in RA can be assessed by DEMRI [1].

In DEMRI a sequences of 3D images of the MCP joints is produced over a period of time resulting in a four dimensional dataset (Figure 1). DEMRI acquisition for this application has been performed on a 1.5T MRI scanner (Gyrosan ACS NT, Phillips Medical Systems, Best, The Netherlands), using a 3D T1 weighted spoiled gradient echo sequence: repetition time/echo time/flip angle = 14/3.8/40°; field of view = 100mm, 6 slices, 3mm slice thickness, 20 dynamic scans at 7.1 seconds intervals with 128 × 256 image matrix. The total scanning

* We are grateful to Elizabeth Berry and Steven Tanner, University of Leeds, for access to data and provision of expert opinion.

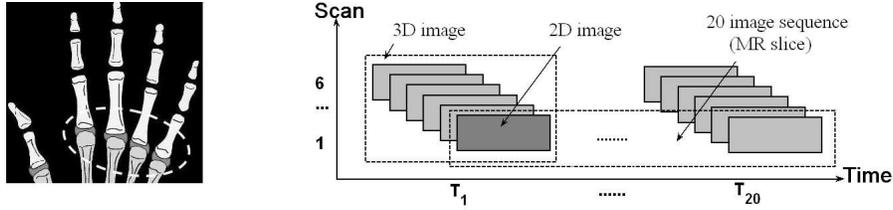


Fig. 1. Anatomy of the area under inspection, and the dataset structure.

time was 142 seconds. Patients were positioned prone, with their arm extended in front of the head and a linear circular 11cm diameter surface coil placed on the dorsum of the hand. A typical image is shown in the top left of Figure 3, where the imaging volume encompasses 4 ($2^{nd} - 5^{th}$) metacarpophalangeals, surrounded by cartilage, blood vessels, and muscle. Full details of the procedure may be found at [1, 2]; the structure of the study is shown in Figure 1 (right).

The contrast agent Gd-DTPA induces selective enhancement of signal intensity from about the $5^{th} - 6^{th}$ time instants. Signal enhancement occurs only in well perfused tissues and where capillary walls allow contrast extravasation. Inflamed synovium will enhance significantly, as will blood vessels. Surrounding muscle will exhibit a low degree of enhancement while cortical bone and cartilage will experience no enhancement.

DEMRI data can be analysed by either pharmacokinetic or ‘black box’ [BB] methods. The former are computationally expensive, but provide a framework that can be used to link MRI physics and the underlying patho-physiology that governs Gd-DTPA kinetics, while the latter measure the effects in terms of heuristics, describing the evolution of enhancement. BB techniques are popular and useful in every day practise; they enable automation of computations, but do not fully use information about the data. They will take the signal intensity [SI] series at each pixel (20 long in our case) and consider various properties that may be extracted from it, without making any assumptions about underlying physical or physiological processes [1, 3]. A technique, which incorporates information about the data into BB methods is introduced in this paper.

2 Current analysis

An existing semi-interactive approach for DEMRI data analysis [1] uses the commercially available software ANALYZE [4], which employs a 3×3 spatial average filter, manual segmentation of the joint exteriors and bone interiors, and no image registration.

Each SI curve $I = (s_1, s_2, \dots, s_{20})$ is normalised over a baseline, computed as the mean of the first 3 values: $b = (s_1 + s_2 + s_3)/3$, $\hat{I} = (s_1/b, s_2/b, \dots, s_{20}/b)$. Computation of three BB variables is performed at each pixel of each 2D image. The *Initial Rate of Enhancement* [IRE] – informally, the greatest rate of

intensity increase – is estimated by passing an averaging window of length n over the signal and determining the gradient of the linear best fit in each window. The maximal such gradient is taken as IRE , and the instant at which this occurs recorded as T_{onset} ($n = 5$ has been used with success). *Maximum Enhancement* [ME], is found as a maximum of mean intensity values calculated in each window.

It is customary then to assemble *Statistical Parameter Maps* [SPMs], which are 2D images depicting these parameters: thus an SPM is a representation of 20 images, summarising some property of interest derived from one 2D section of the data set. Examples can be seen in Figure 3 (which is best viewed in colour).

A further statistic N_{total} is the number of ‘interesting’ pixels – those with $T_{onset} < 8.5$ and $ME > 1.2$. Failure to satisfy both these conditions would indicate a SI curve without appreciable take-up, or one at which the behaviour was not in the expected time interval.

Whereas the method enables partially automated computation of the variables, there are some drawbacks and limitations. Firstly, the estimated T_{onset} corresponds to the time at which intensity is increasing fastest, and is clearly larger than the actual time of onset of enhancement; this estimate is, of course, trivial to improve given this time and the gradient IRE . Secondly, there is often a proportion of curves in which the maximal intensity has not been reached. The method does not allow identification and possible exclusion of these. Estimates of ME and IRE , and assessment of tissue condition, at these locations will not be accurate. Lastly, the definition of ‘pixels of interest’ is very crude in practise. These issues are in no sense fatal flaws in a useful procedure, but do represent some scope for improvement. What is absent from this approach is a fully coherent application of the physical model driving the data acquisition.

3 Methods

3.1 Segmentation and registration

Two aspects of the current technique lend themselves to immediate improvement: spatial registration and automated segmentation of regions not of direct interest.

Registration: The data acquired from the patients suffering from severe RA often exhibit spatial movement. A known intensity-based non-rigid registration algorithm, which models the mapping between images as a locally affine but globally smooth warp [5] has been applied successfully to register all images within a slice to the first one, thereby factoring out the effects of movement. The beneficial effects of this are obvious in examining the SPMs derived from pre- and post-registered data, and are illustrated in Figure 3.

Segmentation: Various interactive and semi-interactive approaches are currently used in isolating the boundary of the hand, the blood vessels, and the bone interiors to ensure that processing concentrates only on areas in which disease may be detected and quantified. In earlier work [6] we have presented an approach to automating this, which we have successfully deployed here also.

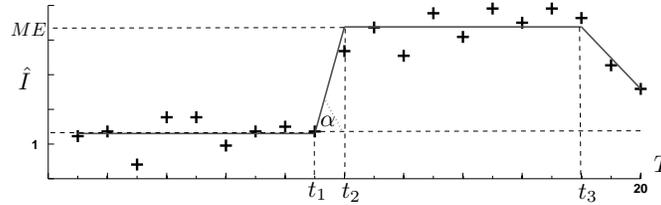


Fig. 2. A sample SI curve approximated with the model M_3 . t_1 is onset of enhancement; t_2 , plateau; t_3 , beginning of relaxation; ME , maximum enhancement; α , gradient of intensity climb.

3.2 Modelling the Procedure

The behaviour of the SI curves may be explained by an underlying physical model. Starting from a baseline, we expect particular tissues to absorb the contrast agent, and their intensity to climb from around the 5th time instant, usually to around the 7th – 15th instant. Following excitation, the tissue regains its equilibrium, which may be seen as an intensity plateau. Finally, the contrast agent leaks into the blood, which will cause the intensity values to decrease. The commencement of this phase is tissue dependent, and is not always observed. Therefore, the tissue under examination might exhibit one of 4 broad behaviours, which we will label and define as:

- M_0 – steady state: pixels at which no enhancement of note can be detected (these will include bone interiors).
- M_1 – base/climb: pixels which clearly enhance but do not reach their maximum intensity within the 20 recorded instants.
- M_2 – base/climb/plateau: pixels at which the maximum is reached and an intensity plateau develops.
- M_3 – base/climb/{plateau}/decline: pixels at which the Gd-DTPA has dissipated and the intensity has detectably started to drop.

We will use this understanding of the underlying procedure to model the SI curves, as an aid to noise reduction. This should permit improved accuracy in the estimates of BB variables which will be extracted from the parameters of the fitted model rather than from the raw curves. Some earlier work [3] has focused on fitting the ‘best’ model to the observed SI behaviour, while others [7] suggest requiring the data to fit an expected model. We seek a compromise between these approaches.

The simplest idea is to impose a piecewise linear interpretation of each of the models $M_0 - M_3$. Figure 2 shows an example \hat{I} signal; the noise levels here are characteristic of the data, suggesting that a linear fit is unlikely to be inferior to any more sophisticated model. The approximation illustrates the phases of the most general case (M_3), not all of which are present in $M_0 - M_2$.

In the same way as the established procedure, we normalise all SI curves to a baseline of 1. Acknowledging that models will be fitted through noise, we require

an estimate of this. Deviations from the baseline b in the first three time instants provide explicit noise measurements. Further, it is expected that there is no Gd-DTPA take-up in the tissues identified as within the bone interiors; therefore, signals corresponding to these pixels may be approximated by a constant (the signal mean), with variations being alternative explicit noise measurements. It is not obvious that these two different noise measurements will correspond, but we have constructed distributions from a wide range of samples and determined using a Kolmogorov-Smirnov [KS] test [8] that it is reasonable to take them as the same. Accordingly, we aggregate these explicit noise measurements to derive a more reliable empirical noise distribution.

We proceed by attempting to fit each of the $M_0 - M_3$ models in a least-squares sense, simultaneously minimising over the model parameters. Each such ‘fit’ implies 20 noise measurements, which we compare for plausibility with the empirically derived noise distribution. This is done using a KS test, and may result in any number of the models being deemed ‘plausible’ – in practise, we find that one at least is always acceptable, and at most locations, only one. Note that we are interested in matching noise distribution and not minimising noise observation; the latter would preclude the simpler models such as M_1, M_2 in favour of M_3 . In the event of more than one model being acceptable, a secondary KS statistic permits discrimination of which is ‘best’.

It is reasonable to suppose at any point that neighbouring tissue will behave in a similar way. Arguing that the models M_1, M_2, M_3 represent ‘ordered’ behaviour, we have passed a selective median filter over the labels represented. The selectivity is that only pixels labelled 1, 2 or 3 are computed in the median.

This results in a relabelling of some pixels which are no longer in our sense ‘best fits’. In the overwhelming majority of cases, these are (in the sense of the KS statistic) second best, suggesting that the noise distribution will not be perturbed radically. We have considered the aggregate distribution of such and compared it to the adopted error model; the KS statistic allows us to deduce it is the same distribution as that of the model to which we are working. Figure 5 shows one illustration of the output of this procedure, which is further discussed in Section 4.

We now restrict our interest to pixels at which the chosen model is one of M_1, M_2, M_3 . ME is immediately available as the maximum intensity value of the model, and we further disregard (as in the earlier technique) locations at which $ME < 1.2$. IRE is chosen as the gradient of the intensity climb, and T_{onset} as the time instant at which the intensity climb commences (t_1).

4 Results

We do not have a formal noise model for the procedure we are observing, but have already determined in Section 3 that the noise exhibited in the first 3 time instants is statistically comparable to that observed at unaffected image locations (bone interiors). It is not immediately clear that noise characteristics will remain unchanged during the procedure; we have considered the noise distribu-

Method	Study 1	Study 2	Study 3	Study 4	Study 5	Study 6
New	0.58 (22%)	0.59 (6%)	0.42 (10%)	0.39 (13%)	0.39 (2%)	0.54 (7%)
Old	0.68	0.63	0.51	0.57	0.57	0.68

Table 1. N_{total} computed for various studies using old and new approaches. In brackets we include the percentage of these pixels at which we consider the procedure to be incomplete.

tion across all T_i during the observed intensity increases: the KS test suggests that the distribution of these errors is not distinguishable from the model we have constructed. Of course, this is in part by definition of the model selection process we have imposed, but we take this as evidence that the procedure is reasonable.

Figure 3 illustrates 3 instances of an SPM of ME : at top right, the original approach, at bottom left, the same data having benefited from co-registration to the first time instant, and at bottom right ME has been derived using our method. There is an incremental increase in the quality of these pictures. The study chosen exhibits significant patient movement, and the effect of registration to compensate for this should be clear. We note that with the new technique the skin enhancement in the SPM has been reduced, and blood vessels partly ignored in the map obtained with the original method are in clear evidence. This will allow their identification and elimination. SPMs obtained with the new method are smoother and less ‘messy’, providing clearer shape of the bone contours, blood vessels, and disease-affected areas.

Considering T_{onset} , it is trivial to adjust the estimate of the original method better to estimate the actual onset time. Our method provides an alternative approach (t_1 of Figure 2) and we have compared these. In the great majority of cases these differ in magnitude by at most 1 and so there is arguably no difference, but there are still many cases in which the difference is appreciable (around 20% of locations of interest). On inspection, most of these represent instances in which our modelling generates a better interpretation of \hat{I} , and our estimate of T_{onset} is more reliable. A full analysis of the details is at [2].

This method allows a more accurate estimate of N_{total} , which helps in tracking disease progression. Originally, this judgement was made by requiring $ME > 1.2$ and $T_{onset} < 8.5$. Our approach permits an improved measure: we label as ‘uninteresting’ any pixel which assumes the label M_0 , or at which ME (as measured by the fitted model) is less than 1.2. Further, we are able to identify pixels at which the procedure is yet to complete, providing an indicator of its overall success. Table 1 shows the number of such curves (N_{total}) normalised to the total number processed in the image (128×256) for a selection of studies. There is a clear qualitative difference in these numbers.

We are further able to map the Gd-DTPA take-up by plotting the adoption of models $M_1 - M_3$. Blood vessels usually assume M_3 , indicating (as expected) that intensities have peaked and are decaying. Most SI curves corresponding to

disease-affected tissues normally assign models M_2 or M_3 , indicating a plateau of intensity and full absorption of the Gd. However, some require the model M_1 , suggesting continuous intensity increase and that the Gd has not been fully absorbed by the tissues during the time permitted for the procedure. Figure 5 (best viewed in colour) illustrates this for different studies; a preponderance of red indicates the procedure is incomplete. The example on the right is clearly inferior. Information of this nature may clearly be of use in tuning the procedure, and in informing more reliable estimates of the BB variables.

5 Discussion and conclusions

We have considered a particular application and presented improvements to an existing semi-automated analysis technique. One of these is an application of an established registration algorithm [5], a second is a purpose-designed segmentation approach [6], and the third is to suggest a modelling of the underlying process that permits more precise extraction of parameters of interest. Each of these three improves incrementally on the results of the original technique. Further evaluation within a clinical setting is needed to examine the impact of these incremental improvements on diagnostic utility of the proposed method.

We consider that all the improvements we suggest have potential for wider application. Currently, we are working on a more sophisticated approach to filtering the output of the model fitting, and developing robust and fully automatic segmentation of blood vessels.

References

1. Radjenovic, A.: Measurement of physiological variables by dynamic Gd-DTPA enhanced MRI. PhD thesis, School of Medicine, University of Leeds (2003)
2. Kubassova, O., Boyle, R.D., Radjenovic, A.: Statistical analysis of DEMRI data of MCP joint. Technical Report 2006.03, University of Leeds (2006) http://www.scs.leeds.ac.uk/research/pubs/reports/2006/2006_03.pdf.
3. Jackson, A.: Analysis of DEMRI. *Radiology* **77**(2) (2004) 154–166
4. Robb, A., Hanson, D.P., Karwoski, R.A., Larson, A.G., Workman, E.L., Stacy, M.C.: Analyze: a comprehensive, operator-interactive software package for multidimensional medical image display and analysis. *Computerized Medical Imaging and Graphics* **6**(13) (1989) 433–454
5. Periaswamy, S., Farid, H.: Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging* **22**(7) (2003) 865–874
6. Kubassova, O., Boyle, R.D., Pyatnizkiy, M.: Bone segmentation in metacarpophalangeal MR data. *Proc. ICAPR* **2** (2005) 726–735
7. Tofts, P.S., Kermode, A.G.: Measurement of the blood-brain barrier permeability and leakage space using DEMRI. *Magnetic Resonance in Medicine* **17**(2) (1991) 357–367
8. Massey, F.J.: The Kolmogorov-Smirnov test for goodness of fit. *American Statistical Association* **46**(253) (1956) 68–77

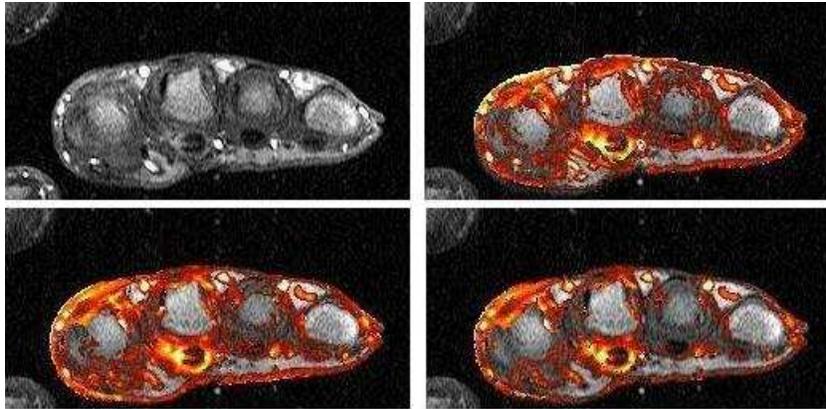


Fig. 3. *This image is best viewed in colour.* A typical MR image (top L). Thereafter, SPMs of *ME*: original approach (top R), with spatial registration (bottom L), obtained using the new method (bottom R).

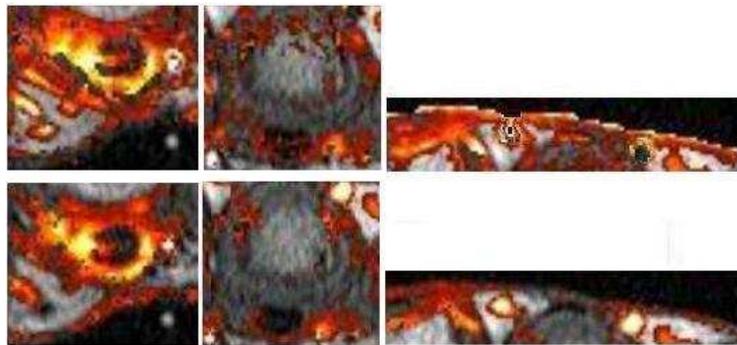


Fig. 4. *This image is best viewed in colour.* SPMs of *ME* with original (1st row) and new (2nd row) approaches. Blow ups show behaviour changes within blood vessels and bone interiors, and reduced skin enhancement.

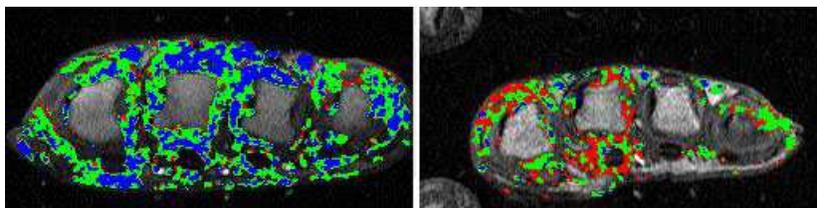


Fig. 5. *This image is best viewed in colour.* Gd take-up maps obtained for slices from 2 different studies. Pixels adopting M_1 are in red, M_2 in green, and M_3 in blue. Red suggests the procedure is incomplete.

Novel Method for Quantitative Evaluation of Segmentation Outputs for Dynamic Contrast-Enhanced MRI Data in RA Studies

Olga Kubassova¹, Roger D. Boyle¹, and Aleksandra Radjenovic²

¹ School of Computing, University of Leeds, UK, {olga,roger}@comp.leeds.ac.uk

² Academic Unit of Medical Physics, University of Leeds, Leeds General Infirmary, Leeds, UK, sasha@medphysics.leeds.ac.uk *

Abstract. We introduce two new metrics for evaluation of segmentation outputs obtained from Dynamic Contrast-Enhanced MRI data. Considering a live application area, we demonstrate the shortcomings of currently accepted algorithms. We present a new supervised approach as an enhanced derivation of a state-of-the-art method, and a novel unsupervised approach, which enables automation of segmentation output assessment. The proposed discrepancy measure considers local blur, partial volume effects, intensity variations, subtle contrast changes and the inconsistency of human experts. We consider the approaches presented to be an improvement on those prevailing, and worthy of wider experiment and application.

1 Introduction

Evaluation of segmentation outputs is clearly important, but despite many years attention has no consensus on approach. Supervised approaches (when segmentation outputs are compared with manually outlined ground truth [GT]) [1–3] are labour intensive and of uncertain reliability, while unsupervised approaches (judging quality without GT knowledge) [4–7] are usually demonstrated on synthetic datasets, rarely agree with each other, and put serious constraints on image properties. These constraints are not generally valid for medical imagery, which is complicated by low contrast and intensity, local blur, patient movement artefacts, and the presence of ambiguous boundaries. The main purpose of this research is to consider currently accepted methods, to show their limitations in application to medical imagery, and to propose new consistent and efficient evaluation metrics for a particular application area.

In this study, we consider Dynamic Contrast-Enhanced MRI [DEMRI] data of the metacarpophalangeal [MCP] joint, involved in hand rheumatoid arthritis [RA] studies. Imaging was performed using a 3D T1 weighted spoiled gradient echo sequence: TR/TE/flip angle = 14/3.8/40⁰, FOV = 100mm, 6 slices, 3mm

* We are grateful to Elizabeth Berry and Steven Tanner, University of Leeds, for access to data and provision of expert opinion.

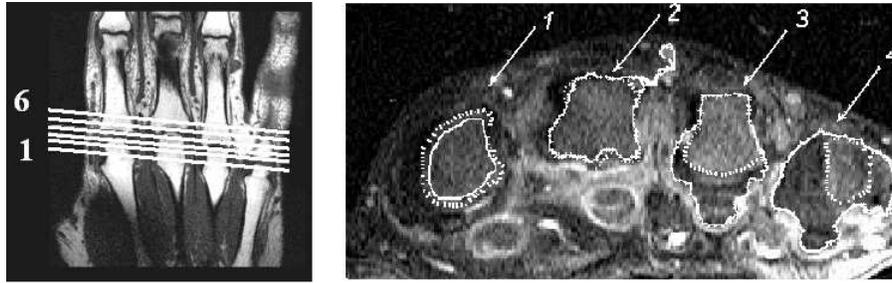


Fig. 1. Left: Positioning of the imaging volume in the hand RA study. (Taken with permission from citesasha03). Right: A typical 2D image of the MCP joint; segmentation outputs of algorithm [9] are shown in solid and ground truth in dotted.

slice thickness, 20 dynamic scans at 7.1 seconds intervals with 128×256 image matrix; the imaging volume encompasses four ($2^{nd} - 5^{th}$) metacarpophalangeals [8]. A contrast enhancer (Gd-DTPA – Gadolinium diethylene triamine pentacetic acid) is injected and provokes selective enhancement of disease affected parts of the joint and blood vessels. Full details of the procedure may be found in [8]. A typical 2D image is shown in Figure 1 (right).

Our application is fraught with problems caused by patient movement, intensity variations within the dataset, partial volume effects, and ambiguity of region boundaries, making segmentation and its evaluation a challenging task. Reliable segmentation is required in order to locate the boundaries of bone interiors, which are good anchors for temporal registration, and which assist in studying the effects of RA. Evaluation metrics are needed both to assess algorithms' performance and the quality of the segmentation outputs either in the presence of GT information or automatically.

Sections 2.1 and 2.2 of this paper present state-of-the-art and new supervised evaluation methods; in Section 2.3 we introduce a novel unsupervised metric. Objective comparison of the currently accepted and novel methods is discussed in Section 3. Section 4 concludes.

2 Methods

2.1 State-of-the-art evaluation techniques, and limitations

To measure quantitatively the results of a segmentation method, it is common to compute a discrepancy metric between the algorithm's output and the GT overlay [4]. Popular metrics are based on a percentage of mutual overlap between these regions [M_{MO}] [1, 3, 10] or a percentage of misclassified pixels in region interiors computed using Hausdorff distance [M_{HD}] [2, 11].

A debate persists on which evaluation metric is preferred [12–14]. In [13] M_{MO} is preferred because M_{HD} does not provide adequate evaluation of seg-

mentation in the case of under-segmented regions, favours over-segmented regions, and its results are highly dependent on confidence in GT. Conversely, it has been shown in [2, 14] that evaluation with M_{MO} does not correspond to human observers' opinion. We will present limitations of these metrics in applications such as ours, and then propose a new one, which performs consistently on this type of data.

- The metrics do not afford an application-adaptable threshold for a degree of tolerance in segmentation error, and therefore cannot deal with local blur, partial volume effects, ambiguity of region boundaries, or inconsistency of human expert judgements. Moreover, evaluation with M_{MO} does not always correspond to expert opinion. For instance, the 4 regions in Figure 1 have M_{MO} of 81%, 74%, 53%, and 11%. With a threshold value of 80%, only the 1st region is considered 'correct', whereas in fact the boundaries of the 2nd and the 3rd partially intersect with the GT and therefore reflect some of the bones' properties, which from a clinician's viewpoint may make this sort of segmentation useful.
- M_{MO} is of most value only when the distances from GT to the segmentation output are unimodal with low variance, and so do not include very local significant deviations such as the boundary of the 2nd region in Figure 1. If the boundary pixels form a 'tail' the evaluation can be confusing – because the 'tail' area is small, M_{MO} is still high.
- These metrics cannot be applied to non-closed boundaries.
- Interpretation of evaluation results is unclear. It is customary to set a threshold for M_{MO} to 50% to indicate an instance of correct segmentation [10], but as shown this is not always adequate. Further, significantly different segmentations can generate very similar M_{MO} measures, making comparability of results an issue.

Despite these drawbacks, these metrics are widely used for evaluation of segmentation algorithms executed on medical imagery [10]. We propose an enhanced metric as a derivation of M_{HD} , which can overcome these limitations and produce comprehensive, reliable and easy to interpret results.

2.2 An enhanced supervised metric

The *Hausdorff distance* [HD] between segmentation output A and GT B is computed by finding the minimum distance from each element of one to some element of the other, and then finding the maximum such;

$$h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} d(a, b) \right\} \quad (1)$$

where $d(a, b)$ is a distance metric, commonly the Euclidean distance between a and b . The HD is asymmetric; usually $h(A, B) \neq h(B, A)$. A general definition of the HD between two sets and corresponding evaluation metric is then

$$M_{HD}(A, B) = \max\{h(A, B), h(B, A)\} \quad (2)$$

We introduce a tolerance threshold t to describe how separated boundaries may acceptably become. For any two boundaries, we can compute the number of pixels of one that are within a threshold t of the other. This number, normalised by the total number of pixels in both boundaries, provides a metric $H_t(A, B)$, which gives the percentage of pixel-wise ‘closeness’ between two boundaries. Let N_A be the number of pixels in boundary A ; then let $A_t(B)$ be the pixels of A within a distance t of a pixel of B . If N_{A_t} is the cardinality of $A_t(B)$, and N_B and N_{B_t} are defined similarly, we will write

$$H_t(A, B) = \frac{N_{A_t} + N_{B_t}}{N_A + N_B} \quad (3)$$

For a given boundary, this metric will increase monotonically with t , and converge to 1. As it measures the distance between boundaries of the regions instead of the regions themselves, it permits evaluation of open boundaries.

The parameter t is an interval of tolerance, within which pixels from one boundary are considered as being in the vicinity of the other: this reflects the acceptable error of segmentation. It can be adjusted for the desired segmentation quality; for example, the width of ambiguous boundary sections, or the opinion of experts. The tolerance can be extracted from the domain, and therefore reflects an application-dependent acceptable segmentation error.

This metric produces easy to interpret and comprehensible results. It permits comparison of different algorithms on various datasets or tuning a method’s parameters. Using it, we can detect the number of pixels which coincide precisely with the GT overlay, or assess the width of ambiguity.

2.3 Unsupervised evaluation

Difficulties in acquiring GT such as imprecise definition, paucity of information, and time consumption have motivated research into unsupervised evaluation [4, 5], but most of this work has been demonstrated on binary or synthetic data, with limited experiments on real world imagery. Generally unsupervised evaluation metrics are based on the location, shape, size, contrast, or intensity of segmented regions, and require these regions to have homogeneous intensities, bright contrast, or compact shapes. DEMRI data with its unfavourable characteristics might not be evaluated adequately by existing metrics; we are unaware of unsupervised evaluation metrics being successfully employed in domains with the characteristics of ours.

We assume that as the result of segmentation some of the pixels from the machine segmented boundary will be allocated in the vicinity of the actual boundary of the region and some not. Thus, some pixels can be thought of as of ‘high detection confidence’, others might belong to the ambiguous sections of the boundary, or might not represent boundary at all.

In the vicinity of the ‘reliable’ pixels, we expect a contrast change, which will be adapted to average image intensity and local image/region contrast. The metric will then evaluate segmentation algorithms based on their ability to detect a large number of reliable pixels in region’s boundaries.

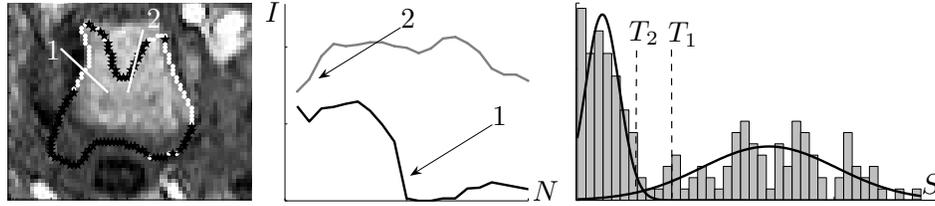


Fig. 2. Left: Segmentation output split into reliable (shown in white) and unreliable (shown in black) parts. Two normals of length N are drawn across the boundary. Middle: The intensity profiles along the normals. Right: Distribution of the boundary pixels' strengths (S).

As a guide to scale, we take the bounding box of a detected boundary and compute the length of its diagonal D ; in our application this is reasonable as regions (while not circular) are not elongated. We will construct normals of length $2N$, where $2N = D/2, D/4, D/8$, along which we will consider intensity profiles. Such profiles may or may not correspond to that expected at a 'good edge' – see Figure 2 (left).

To estimate the 'reliability' of a boundary pixel, along each profile we build a weighted sum to measure its strength. If I_1 denotes the intensity profile within the region and I_2 the profile without, we will select $\alpha_1 > \alpha_2 > \dots > \alpha_N > 0$, $\sum_i \alpha_i = 1$, and set $S(N)$ by Equation 4.

$$S(N) = \left| \sum_{i=1}^N \alpha_i \cdot I_{1i} - \sum_{i=1}^N \alpha_i \cdot I_{2i} \right| \quad (4)$$

The weights α_i may be selected in many ways: for simplicity we have chosen them to be linearly decreasing to 0.

When for a given profile length we examine a distribution of these strengths, it is frequently seen to be bimodal: the reliable pixels display high strength and the others low – an example is in Figure 2 (right). This is a useful property of this form of application. We model this bimodality as a mixture of two Gaussians by applying the EM algorithm [15]. A Kolmogorov-Smirnov [KS] test [16] shows that in the majority of cases a mixture of two Gaussians is a reasonable approximation for this bimodal distribution.

At this stage we need a discriminator between 'reliable' and 'unreliable' boundary pixels. The model delivers μ_1, σ_1 for the 'high' strengths, and μ_2, σ_2 for the 'low' strengths. Two simple choices of threshold, seeking 95% confidence, are:

$$\begin{aligned} T_1 &= \mu_1 - 1.96\sigma_1 \\ T_2 &= \mu_2 + 1.96\sigma_2 \end{aligned} \quad (5)$$

The choice of 1.96 may obviously be changed. If the distributions are well separated we will expect $T_1 \geq T_2$; if this is not the case we do not satisfy the

Metrics	MO	H_2	H_3
M_1	0.16	0.15	0.18
M_2	0.18	0.19	0.21
M_3	0.28	0.26	0.31
M_4	0.32	0.32	0.33
New Metric	0.84	0.93	0.97

Table 1. Correlation between results of supervised and unsupervised metrics

assumption that requires segmentation output to contain a number of ‘reliable’ and a number of clearly ‘mis-detected’ boundary pixels. A number of circumstances can cause $T_1 < T_2$ which we do not consider in this paper. We select T_1 as a threshold; this produces a strict condition. This strictness is enhanced by requiring a ‘reliable’ pixel to meet this condition at all three scales chosen.

After each boundary pixel is assessed as reliable or not, we pass a median filter of size 3×3 across the border, which has a simple but beneficial effect on the connectivity of reliable pixels. The quality of segmentation result is then computed as the number of reliable boundary pixels normalised by the number of pixels in the region boundary, giving a measure in the range $[0, 1]$.

3 Experiments and Discussion

The dataset used for assessing performance had 400 regions of different noise, contrast, and intensity level, with open, closed, ambiguous boundaries segmented by 3 algorithms of different behaviour and underlying design criteria. They were: region growing [9], adaptive segmentation based [17], and snake [18] algorithms. Corresponding GT overlays were manually outlined for each region by several human experts. Per-pixel difference between these judgements averaged 2-3 pixels, which suggests an appropriate t in evaluating H_t .

The performance of an unsupervised metric can be assessed by comparison of its results with those of a supervised metric applied to the same dataset. In these experiments we consider the metrics we have developed alongside 4 currently used unsupervised approaches. They are: Levine and Nazif inter-region uniformity criterion [19] (will be referred to as M_1); Weszka and Rosenfeld intra-region uniformity metric [20] (M_2); Borsotti criterion [21] applied for medical imaging [5] (M_3); and Zeboudj contrast criterion [22] (M_4).

The quality of the regions from the dataset has been evaluated by each metric, then the correlation coefficients between vectors of evaluation results (400 long) have been computed for each supervised and unsupervised approach [5]. Table 1 summarises the results.

This experiment demonstrates that the newly proposed unsupervised metric produces reliable results, which closely correspond to the evaluation provided by supervised approaches. M_1 , M_2 and M_3 perform adequately on DEMRI data only if contrast is sharp, regions are homogeneous and of compact shapes. M_4

operates on contrast changes in the neighbourhoods of boundary pixels, but cannot cope with noisy images.

We have chosen regions which can be adequately evaluated by both M_{MO} and H_t . The supervised metrics generate very similar results, confirming that the novel way to assess segmentation does not contradict the state-of-the-art approach, but when necessary can cope better with regions of unfavourable quality and shape.

For the most challenging regions we have evaluated the quality of the reliable boundaries detected by the new metric. On average, such boundaries are 0.86 pixels away from the average GT, and more than 90% of pixels identified as reliable coincide with GT. This suggests that the proposed metric is highly efficient when precise segmentation is required.

The new metric can also be used in assessment of algorithm performance; we evaluated 3 algorithms [9, 17, 18]. In our application, images from early in the sequence provide much higher contrast; later, the enhancing agent will reduce contrast in diseased areas. Correspondingly, segmentation outputs' quality decreases. The algorithms [9, 18] are highly parameter-dependent and tend to over-segment images from the end of the sequence and to under-segment those from the beginning; [17] is largely free of this parameter selection problem and provided accurate segmentation in most cases. Detailed discussion can be found in [17]. For algorithms [9, 17, 18] on average H_3 is (0.6, 0.9, 0.7) and H_0 is (0.3, 0.8, 0.6). These measurements allow thorough assessment of segmentation quality at different t , tuning the algorithms' parameters, and finding an appropriate algorithm for an application.

4 Conclusion

In this paper we demonstrated that currently used metrics are of restricted applicability and often cannot adequately assess segmentation outputs obtained on some types of medical data. We have suggested a means of enhancing an existing supervised metric to enable evaluation of closed, open, and ambiguous boundaries in images complicated by local blur and patient movement artefacts. The metric has been demonstrated in the assessment of segmentation of DEMRI data of the MCP joint (but has been seen to perform consistently on other medical or real world images too).

We have proposed a novel unsupervised metric, which enables automated assessment of segmentation outputs, and has outperformed existing unsupervised approaches. The metric can also be used in automatic detection of 'reliable' boundaries, which might assist in disease diagnoses in a variety of studies. The metric has shown reliable performance in our domain and on images of similar characteristics; currently we are working on experiments for wider application.

References

1. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26**(3) (1945) 297–302

2. Yasnoff, W.A., Mui, J.K., Bacus, J.W.: Error measures for scene segmentation. *Pattern Recognition Letters* **9** (1977) 217–231
3. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M.C., Kaus, M.R., Haker, S.J., III, W.M.W., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index. *Radiology* **11**(2) (2004) 178–189
4. Zhang, Y.J.: A survey on evaluation methods for image segmentation. *Pattern Recognition* **29**(8) (1996) 1335–1346
5. Chabrier, S., Emile, B., Rosenberger, C., Laurent, H.: Unsupervised performance evaluation of image segmentation. *EURASIP Journal on Applied Signal Processing* **2006** (2006) 1–12
6. Forbes, L.A., Draper, B.A.: Inconsistencies in edge detector evaluation. *Computer Vision and Pattern Recognition* **2**(5) (2000) 398–404
7. Sezgin, M., Marmara, T.: Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* **13** (2004) 146–165
8. Radjenovic, A.: Measurement of physiological variables by dynamic Gd-DTPA enhanced MRI. PhD thesis, School of Medicine, University of Leeds (2003)
9. Gonzalez, R.C., Woods, R.E., Eddins, S.: *Digital image processing*. Prentice Hall, New York (2004) pp. 404–407.
10. Bowyer, K.W.: Validation of medical image analysis techniques. In Sonka, M., Fitzpatrick, J.M., eds.: *Handbook of Medical Imaging*. Volume 2. 1 edn. Press Monograph, Bellingham, WA (2000) 567–607
11. Chalana, V., Kim, Y.: A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Transactions on Medical Imaging* **16**(5) (1997) 642–652
12. Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldof, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A., Fisher, R.B.: An experimental comparison of range segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(7) (1996) 673–689
13. Chabrier, S., Laurent, H., Emile, B., Rosenberger, C., March, P.: A comparative study of supervised evaluation criteria for image segmentation. *12th European Signal Processing Conference, Vienna, Austria* **12** (2004) 1143–1146
14. Sendov, B.: Hausdorff distance and image processing. *Russian Mathematical Survey* **59**(2) (2004) 319–328
15. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(3) (2002) 381–396
16. Massey, F.J.: The Kolmogorov-Smirnov test for goodness of fit. *American Statistical Association* **46**(253) (1956) 68–77
17. Kubassova, O., Boyle, R.D., Pyatnizkiy, M.: Bone segmentation in metacarpophalangeal MR data. *3rd International Conference on Advances in Pattern Recognition, Bath, UK* **2** (2005) 726–735
18. Kass, M., Witkin, A., Terzopoulos, D.: Snakes, active contour models. *International Journal of Computer Vision* **1**(4) (1987) 321–331
19. Levine, M.D., Nazif, A.M.: Dynamic measurement of computer generated image segmentations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7**(2) (1985) 155–164
20. Weszka, J.S., Rosenfeld, A.: Threshold evaluation techniques. *IEEE Transactions on Systems, Man and Cybernetics* **8**(8) (1978) 622–629
21. Borsotti, M., Campadelli, P., Schettini, R.: Quantitative evaluation of colour image segmentation results. *Pattern Recognition Letters* **19**(8) (1998) 741–747
22. Jourlin, M., Pinoli, J.C., Zeboudj, R.: Contrast definition and contour detection for logarithmic images. *Journal of Microscopy* **156** (1989) 33–40

3D Shape Description of the Bicipital Groove: Correlation to Pathology

Aaron D. Ward¹, Ghassan Hamarneh¹, and Mark E. Schweitzer²

¹ Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, V5A 1S6,
Canada

{award, hamarneh}@cs.sfu.ca, <http://mial.fas.sfu.ca/>

² Department of Radiology and Orthopedic Surgery, NYU School of Medicine,
Hospital for Joint Diseases, New York, NY, 10016, U.S.A.
{mark.schweitzer}@nyumc.org

Abstract. The bicipital groove (BG) of the proximal humerus retains the tendon of the long head of the biceps. It is understood that the shape of the BG is related to the probability of injury to the long biceps tendon (LBT). Measurements taken of the BG in previous studies from dry bones and radiographs (henceforth *classical measurements*) are of single cross sections of the humerus, and may therefore overlook important BG shape characteristics. In this study, we test the hypothesis that a novel, medial axis-based 3D shape descriptor captures all relevant features measured in previous work, plus more. To this end, we review previous studies wherein classical measurements have been taken on large numbers of BGs, yielding a distribution that reveals the nature of a normal BG. We develop an automated approach to replicating those measurements on MRI to determine, for each of our data sets, the deviation from the mean of all the classical measurements. We train a classifier by pairing our 3D representations with these deviations to evaluate the potential for computer aided diagnosis of BG pathology based on our 3D shape descriptor.

1 Introduction

The bicipital groove (BG) of the proximal humerus is located on the humeral head, and is formed by the greater and lesser tuberosities (figure 1). The long biceps tendon (LBT) is retained by the BG as the arm moves. Abnormal shape of the BG can induce injury of the LBT. Relevant shape measurements taken in previous work (henceforth *classical measurements*) include BG depth, width, and medial wall angle (figure 2) [1–5]. A deep, narrow BG can irritate the LBT, causing tenosynovitis. A shallow, wide BG can favour dislocation of the LBT. Also, the presence of the supratubercular ridge of Meyer (henceforth *the ridge*) is understood to greatly favour dislocation [1, 2]. In previous work, measurements are taken at a single 2D cross section of the humeral head. Due to large intra-subject variation in BG shape (figure 1 (b),(c)), such measurements are prone to overlooking important shape features, motivating the need for a 3D shape descriptor that captures information along the entire length of the BG.

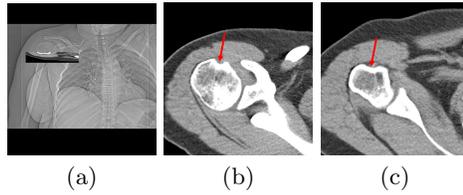


Fig. 1. (a) A radiograph intended to show the location of the BG within the body (indicated by the small intersecting axial cross section). (b) A proximal axial cross section showing the shape of the BG, indicated by an arrow (CT scan shown for clarity of illustration). (c) A distal axial cross section from the same patient; note the large difference in BG shape within a single patient; this is typical.

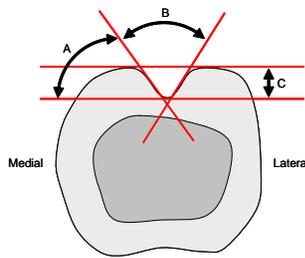


Fig. 2. BG measurements taken from a single cross section in previous literature. A: Medial opening angle. B: Total opening angle (capturing width). C: Depth. (Adapted from [5].)

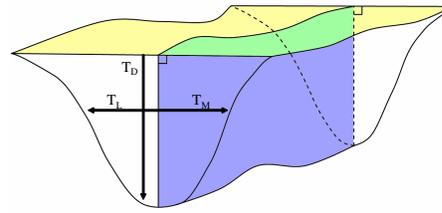


Fig. 3. Depiction of the shape descriptor used in this study [6]. An intertubercular sheet (yellow) is computed to join the tuberosities. A medial sheet (blue) is orthogonal to the intertubercular sheet and intersects as near to the deepest BG point as possible while remaining smooth. Magnitudes of vectors (T_M , T_L , T_D) emanating from sampled points on these sheets and terminating at the BG surface form several 2D thickness fields (medial wall, lateral wall, depth, and width) capturing BG shape.

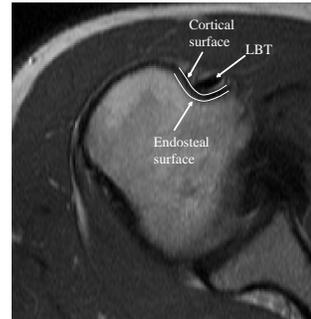


Fig. 4. Appearance of bone on T2-weighted MRI. The inner surface where bone meets bone marrow is the endosteal surface. The outer surface where bone meets surrounding tissue is the cortical surface. We must distinguish between these during surface extraction, because previous population studies of BG shape consist of measurements taken on dry bones and radiographs. Thus, our classical measurements should measure the cortical surface.

In previous work [6], we demonstrated that a 3D, medial axis-based shape descriptor captures medically relevant shape information as 2D thickness fields computed relative to a medial sheet positioned to approximately bisect the BG (figure 3). In the current paper, we investigate the performance of machine learning classifiers in determining normal vs. pathological BG shapes based on this descriptor. That is, the input is a set of thickness fields representing a BG shape, and the expected output is a proper classification of BG shape (e.g. “normal”, “abnormal medial wall angle”, “presence of the ridge”). The hypothesis is that the 3D shape descriptor is sufficient for automated determination of the deviation of BG shape from normal. This is a first step in answering the subsequent question, to be addressed in future work, of whether the 3D shape descriptor is sufficient for automated determination of the probability of injury to the LBT. To test the hypothesis, we examine previous studies reporting classical measurements taken on large numbers of BGs (dry bones and radiographs). This yields a distribution that allows, for a set of classical measurements taken from a BG, the determination of the deviation from the mean of the distribution. Our automated approach to classical measurement on MRI allows the determination of where each of our BGs fits within the distribution given by previous work, yielding the inputs to our tested classifiers. Our main motivation for pursuing a machine learning/classification approach to this problem is that identification of osseous spurs, the ridge, and the angle of the medial BG wall is difficult due to lack of precise definitions of these structures. This motivates the need for a 3D shape descriptor inherently capturing these features inherently which can be used in classifier training for identification of pathologies in these structures.

In the course of this work, we face the question: why not directly perform a study correlating BG shape to incidence of LBT injury? The reason is that there is an indirect relationship between the anatomical structure whose shape we are computing (BG) and that whose probability of injury we are trying to estimate (LBT). It is understood that an abnormal BG shape *may predispose* an individual to LBT injury. For example, an osseous spur (bony abnormality) inside of the BG can cause the LBT to fray and tear [2]. However, it is entirely possible that at the time of the MRI scan, a patient with an osseous spur in his BG has a perfectly healthy LBT; his LBT injury, if it is going to happen, has not happened *yet*. This presents a problem when training a classifier to learn the probability of LBT injury based on BG shape: with a practically feasible sample size, such cases (i.e. healthy LBT in abnormally-shaped BG and vice versa) can confuse the classifier. Given a sufficiently large sample, one expects the effect of this confusion to be minimized, but collecting such a sample is cost prohibitive.

The indirect relationship between the BG shape and incidence of LBT injury is one of this study’s primary motivations. By computing the classical measurements of BG shape and placing each data set in context of the distribution of measurements from previous population studies, we can determine the deviation of each BG shape from normal. This allows us to identify and handle cases where the patient has an injured LBT in a normally-shaped BG and vice-versa. The other primary motivation of this study is to verify that the shape descriptor suf-

ficiently captures, at minimum, the classical measurements of BG shape. If this shape descriptor can be used to effectively train a classifier to perform BG shape diagnosis, then we can proceed with confidence in a future study establishing a relationship between this shape descriptor and the probability of LBT injury.

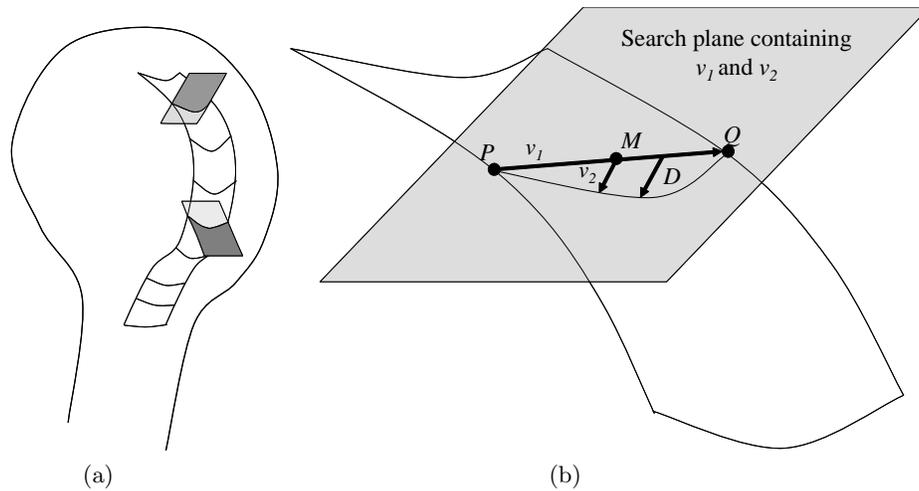


Fig. 5. Because the BG curves with the contour of the humeral head, it is important in measuring depth to ensure that depth is measured along vectors orthogonal to the intertubercular sheet, lying in search planes exemplified in (a). (b) is an enlarged view of a single search plane from (a) showing the details of the definition of the search plane at a given slice of the BG. Segment PQ is defined to join the endpoints of the intertubercular sheet, and vector v_1 is defined from P to Q . Vector v_2 is defined to originate from the midpoint M of PQ and be *normal to the intertubercular sheet* at M . The search plane is defined by v_1 and v_2 . Rays are cast from points sampled uniformly along and orthogonal to PQ , constrained to lie in the search plane, and terminating at the BG surface. The length of the longest such ray (indicated by D) is determined to be the depth of the BG on this slice.

The remainder of this paper is organized as follows. In section 2 our data sets and experimental approach are described. In section 3 we give our results and a discussion, and in section 4 we give some concluding remarks.

2 Material and Methods

This study is based on 32 T2-weighted MRIs of the shoulder taken at 1.5T. 10 of these data sets correspond to patients diagnosed with a normal LBT; 22 are diagnosed as abnormal (subluxation, dislocation, or tear). Our approach is as follows (see figure 6).

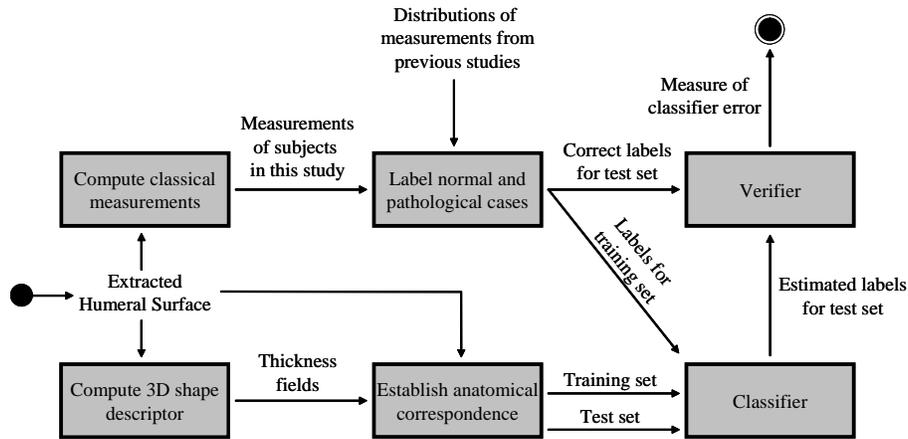


Fig. 6. Procedure followed in this investigation. Extracted BG surface points are used to compute a 3D shape descriptor (figure 3), and also used to compute “classical” measurements of BG shape used in previous work (figure 2). Based on distributions of classical measurements determined in previous studies, a diagnosis of normal versus pathological shape is made for each data set based on its classical measurements. These diagnoses form labels for a set of training shapes given to a classifier, which attempts to correctly label BG shapes in the test set. We evaluate the accuracy of the classifier based on a comparison of these labels to known labels for the test set.

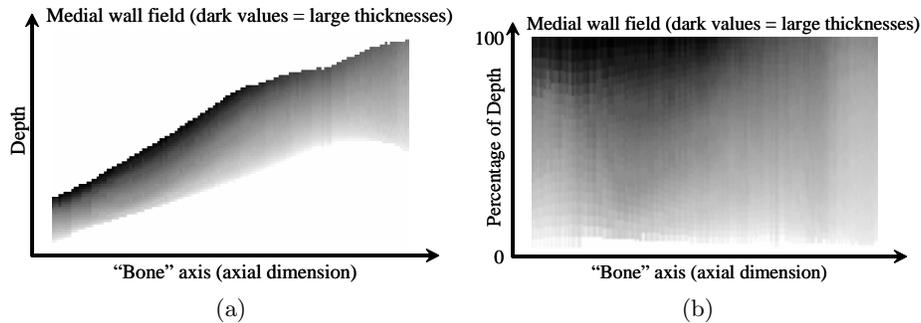


Fig. 7. (a) A medial wall field rendered as a 2D image, with one axis along the bone (axial) dimension, and the other along the depth of the BG. (b) The same field after rescaling along the depth dimension to transform it into a percentage of depth dimension for alignment.

1. **Compute classical measurements.**¹ Given a set of points sampled from a BG surface, we automatically compute the width, depth, and medial wall angle (figure 2) of each data set. Care must be taken to ensure that the

¹ The authors wish to acknowledge and thank Eli Gibson for his efforts in discussion and implementation of this approach to automated classical measurement.

cortical surface of the bone is extracted, to ensure consistency with previous population studies of dry bones and radiographs (figure 4). Segmentation is not the focus of our study and is done manually. To take the measurements, the BG surfaces are first rotated so that slicing axially yields cross sections orthogonal to the humerus. An intertubercular sheet is fitted to the tuberosities (figure 3). At each axial slice S , the intertubercular sheet appears as a line segment PQ with endpoints P and Q touching the tuberosities. The depth D of the BG on S is determined as shown in figure 5. The width of the BG at S is the length of PQ . Starting from the deepest point on the BG surface on S , points sampled along the medial wall define endpoints of line segments approximating the wall. Angles of these line segments with respect to the segment PQ are recorded for S . So, for each BG, we have a set of depth and width values and a set of medial wall angles. Since measurements taken in previous work are of a single slice, and previous authors are not specific in describing how the slice is chosen [7, 2, 4], we aggregate all of our measurements by taking the mean for a single data set. This yields a single depth, width, and medial wall angle for each data set for comparison purposes. This single-slice limitation further reinforces the need for a 3D shape descriptor.

2. **Label normal and pathological cases.** Previous studies took measurements of depth, width, and medial wall angle on dry bones and radiographs [7, 2, 4]; 130 patients in total. The results are as follows. Medial wall angle mean: 60.02° , standard deviation (SD): 15.32° . Depth mean: 4.19mm, SD: 0.96mm. Width mean: 7.9mm, SD: 1.42mm. To provide a binary classification of each BG to the classifier for training, we must specify a standard deviation cutoff defining normal vs. abnormal. We choose a threshold (1.5 SD for our data, for all measurement types) that results in half of the data being normal, presenting the greatest challenge to the classifier. Thus, attempts to classify at random result in the poorest possible performance (e.g. if 90% of the data sets were normal, a classifier could achieve 90% accuracy by simply classifying all test sets as normal). Each data set is also labeled according to expert observation of the presence of the ridge.
3. **Computation of 3D shape descriptor.** An intertubercular sheet is fitted to close the BG, and a medial sheet is computed orthogonal to the intertubercular sheet on each slice. Medial and lateral wall, depth and width fields are computed relative to the sheets (figure 3). Our previous publication [6] gives further details.
4. **Anatomical correspondence.** To prepare the thickness fields for machine learning, we establish anatomically meaningful correspondence between elements. Point (i, j) in any thickness field should correspond anatomically with points (i, j) in the thickness fields of all other data sets. Establishing correspondence is challenging for the BG as it lacks meaningful anatomical landmarks. Due to the large slice thickness, the proximal end of the BG is not reliably determined, and a method for determining the distal end is debated [4]. Our approach is indirect: since the BG is formed by the tuberosities of the humerus we align the *humeri*, consequently aligning the BGs. Our

shape descriptor is invariant to rigid transformations except axial translation, and it is not invariant to changes in scale. To establish correspondence, we find the parameters of the best-fit sphere to the humeral head using the Hough transform. We align all thickness fields such that the axial coordinate of these sphere centers are the same, thus aligning the bones. We scale thickness fields to normalize for humeral head size, according to the spheres' radii, resulting in a set of thickness fields (e.g. figure 7(a)), normalized for scale and aligned along the bone (axial) dimension. Finally, we rescale the field along the depth dimension to make the fields rectangular by resampling thickness values from 0 to 100% along the depth of the BG (figure 7(b)).

<i>Classical Measurement</i>	<i>Classifier</i>	<i>Error</i>	<i>No. Principal components</i>
Width	Quadratic Bayes	0.1875	6
Depth	Min. LS Linear	0.1875	6
Medial wall angle	Quadratic Bayes	0.3750	15
Supratubercular ridge	Min. LS Linear	0.1250	6

Table 1. Results of testing of classification, showing the methods that gave the best performance against our data. The error indicates the proportion of data sets that were mis-classified.

- Classification.** We performed dimensionality reduction using PCA on the 1000D vectors formed by the thickness fields. We then trained several classifiers against the dimensions of the thickness field data capturing 95% of the variation. We also trained classifiers to recognize the presence of the supratubercular ridge of Meyer [1] from the depth fields. Testing was performed in a leave-one-out fashion, with classification errors averaged over all rounds.

3 Results

Table 1 shows the results of classification. Accuracy was over 80% for most classifications; different types of classifiers performed best for different tasks. Using classifiers in the *PRTools v.4* package ², we obtained best results with the minimum least square linear classifier and quadratic Bayes normal classifier. No classifier performed adequately in diagnosis of medial wall angle abnormalities. Vagueness regarding the slice locations of BG measurements taken in previous studies may also adversely affect classification. Considering these obstacles, the results are encouraging; they suggest that the majority of important BG shape features are captured by our representation. Especially encouraging is the classification performance for the ridge, which can be difficult for the human expert to identify.

² PRTools v.4, Delft U. of Technology

4 Conclusion

In this work, we investigated the ability of a 3D shape descriptor for the BG to capture aspects of shape known to be related to LBT injury. We showed this by demonstrating that classification algorithms can be trained, using our shape descriptor, to perform accurate diagnosis of BG shape abnormality. The outcome of this investigation is that classification performance using this shape descriptor is acceptable, given the practical obstacles of small sample size and lack of precise literature specifying how some classical measurements were taken in previous studies. The auxiliary aim of this work is to illustrate practical considerations that need to be addressed in a computational study of musculoskeletal disorders on real data, such as development of strategies for handling small sample sizes and anatomical alignment of structures that may lack clear anatomical landmarks. Future work includes establishing the relationship between the 3D shape of the BG and the incidence of injury to the LBT.

References

1. Meyer, A.W.: Spontaneous dislocation and destruction of tendon of long head of biceps brachii. Fifty-nine instances. *Archives of Surgery* **17** (1928) 493–506
2. Hitchcock, H.H., Bechtol, C.O.: Painful shoulder. Observations on the role of the tendon of the biceps brachii in its causation. *Journal of Bone and Joint Surgery, American Volume* **30A** (1948) 263–273
3. O'Donoghue, D.: Subluxing biceps tendon in the athlete. *American Journal of Sports Medicine* **1**(3) (1973) 20–29
4. Ueberham, K., Le Floch-Prigent, P.: Intertubercular sulcus of the humerus: Biometry and morphology of 100 dry bones. *Surg. and Radiol. Anat.* **20**(5) (1998) 351–354
5. Pfahler, M., Branner, S., Refior, H.J.: The role of the bicipital groove in tendopathy of the long biceps tendon. *J. Shoulder and Elbow Surgery* **8**(5) (1999) 419–424
6. Ward, A.D., Schweitzer, M.E., Hamarneh, G.: 3D shape description of the bicipital groove of the proximal humerus. *Proceedings of SPIE Medical Imaging: Physiology, Function, and Structure from Medical Images* **6143** (2006) 1–9
7. Ahovuo, J.: Radiographic anatomy of the intertubercular groove of the humerus. *European Journal of Radiology* **5**(2) (1985) 83–86

Efficient Automatic Cartilage Segmentation

Erik B. Dam^{1,2}, Jenny Folkesson¹, Marco Loog¹,
Paola C. Pettersen², and Claus Christiansen²

¹ IT University of Copenhagen

² Center for Clinical and Basic Research

Abstract. Cartilage volume is the most obvious quantification of cartilage breakdown when monitoring osteoarthritis progression. We present a novel algorithm for speeding up voxel classification by an order of magnitude. This new classification scheme is used for fully automatic segmentation of tibial and femoral articular knee cartilage.

We evaluate the method on a collection of 114+31+25 knee MR scans of both healthy and OA subjects, and show that the segmentations are identical to the segmentation from a straight-forward voxel classification method. Furthermore, the method shows high reproducibility and proves to be able to separate healthy from OA subjects.

1 Introduction

Osteoarthritis (OA) is a very common age-related joint disease; in fact the majority of the elderly population will experience OA [1]. The total economic burden of arthritis is estimated at 1–2.5% of the gross national product of Western countries; OA accounts for the major share of this burden [2]. The disease causes pain, joint swelling, reduced range of motion, and disability. Thereby, it is a severe cause for reduced quality of life.

Currently, no treatment shows consistent, document effect on OA [3]. A central problem in developing new treatments is the lack of convincing quantification methods with high accuracy, precision and sensitivity. For clinical trials, automatic and precise progression quantification is needed in order to eliminate observer variability and thereby allowing smaller study populations and accurate quantification is needed to allow shorter study duration.

A central process in OA progression is cartilage breakdown. During the early stages of OA, local cartilage lesions occur causing local swelling and then thinning, which in later stages leads to holes in the cartilage. In severe OA, large areas of cartilage are entirely missing. Cartilage volume is the most obvious measure for quantifying this overall breakdown.

Much research has been devoted to quantifying this progression from radiographs. This has limitation due to the loss of information in the 2D projection combined with the fact that cartilage is not visible in X-ray. We therefore focus on quantification from Magnetic Resonance Imaging (MRI) which allows non-invasive visualization of the cartilage [4] and direct cartilage assessment [5].

This work is standing on the shoulders of previous work on cartilage segmentation. In [6, 7], we presented a fully automatic cartilage segmentation method based on supervised learning. This method is reviewed in section 3.

The contribution of this paper is an efficient algorithm for performing voxel classification. The algorithm is generally applicable, and is here evaluated on the task of performing cartilage segmentation. The algorithm, described in detail in section 4, works by random sampling of a sparse set of seed voxels, followed by voxel classification performed in regions where cartilage is located only. Thereby, the computation time needed for performing classification is reduced by an order of magnitude.

We evaluate the automatic volume quantification method on a collection of 114+31+25 knee MRI (see section 2 for details). The evaluation in section 5 focuses on the efficiency of the new method as well as the the equivalence of the segmentation results between the new and the original methods. We also evaluate reproducibility and ability to distinguish healthy from OA test subjects.

1.1 Related Work

Somewhat surprisingly, cartilage volume is relatively poor for quantification of OA progression and in some studies there is even no detectable volume change over time for OA patients [8] and [9]. However, even if cartilage volume alone is not the best quantification of OA progression, cartilage segmentation is still a necessary step in quantification of more sensitive measures such as cartilage thickness.

A number of semi-automatic cartilage segmentation methods have been published. Some methods are essentially 2D such as the slice-wise active contour approach in [10] and the B-spline approach in [11]. In the latter they evaluate the accuracy of the thickness quantification for the shoulder by comparing with measurements on extracted cartilage specimens and get differences between 15% and 20%. In [12], they evaluate the B-spline approach from [13] which requires 2.5 hours of interaction per knee joint and get an inter-observer variability of around 6% for volume quantification. Another study also using the same B-spline approach get a scan-rescan reproducibility of 5.5% for volume measurement of the medial tibial cartilage for high-resolution MRI [14].

In [15], from an automated 3D approach that requires around 10 minutes of manual corrections of the segmentations per knee, they produce segmentations of both bone and cartilage. Their limited validation indicates that the OA subjects have thinner cartilage.

Finally, an almost fully automatic method based on a graph searching segmentation algorithm [16] followed by mean thickness quantification is evaluated on ankle joints in [17]. The method requires a simple, manual initialization and then requires 4 minutes and 30 seconds of computation on average. The evaluation on 8 cadaveric ankles show accurate thickness measurements. Presumably the method could also be adapted to knees.

2 Data Collection

We evaluate our thickness quantification method on a collection of knee MRI prospectively acquired on an Esaote C-Span low-field 0.18 T scanner dedicated to imaging of extremities using a sagittal Turbo 3D T1 sequence (flip angle 40° , TR 50ms, TE 16ms) with a voxel size of $0.70 \times 0.70 \times 0.78\text{mm}^3$. Approximate scan time is 10 minutes.

The collection includes both left and right knees — right knee scans are reflected in order to apply the same methodology to all scans. The test subjects were males and females of ages between 21 and 72 years with no to quite severe OA symptoms (scores 0 to 3 on the Kellgren and Lawrence scale [18]). The collection contains 25 knee we use for training of the automated methods and 114 knees for evaluation. Among the 114 knees, 31 were rescanned a week later. For all scans, the cartilage was segmented by a radiologist by slice-wise outlining (see figure 1). We focus on the medial compartments of the tibial and femoral cartilage sheets since the correlation between degradation and clinical symptoms is predominant in the medial compartments [19].

3 Automatic Cartilage Classification

The fully automatic classification step is based on previous work [6] and is a k Nearest Neighbor (kNN) voxel classifier trained on the manual segmentations from the 25 training set scans. A feature selection scheme selects a feature vector of around 40 features from a collection of potential features consisting of voxel position, intensity, Gaussian derivatives up to order three, eigenvalues and eigenvectors for the Hessian, and eigenvalues and eigenvectors for the Structure Tensor — all except position evaluated at multiple scales. In the implementation we use an approximate nearest neighbor classifier [20].

A separate one-versus-all classifier is used for tibial and femoral cartilage. The resulting posterior probability maps are then combined into a single label map where each voxel is classified as the tibial or femoral cartilage depending on which posterior probability is highest provided that probability is above a threshold (98%) — otherwise the voxel is classified as background. This is followed by extraction of the largest connected component for each compartment in order to clean the segmentation and dispose of stray, false positives. The segmentation thereby results in a segmentation of the medial compartments of tibial and femoral cartilage (see example in figure 1).

4 Efficient Voxel Classification

The segmentation method based on voxel classification described above is very appealing in many respects. It is very accurate (compared to manual expert segmentations), very reproducible (evaluated on scan-rescan pairs), and completely automatic. However, it is fairly slow — the entire computation takes more than an hour.

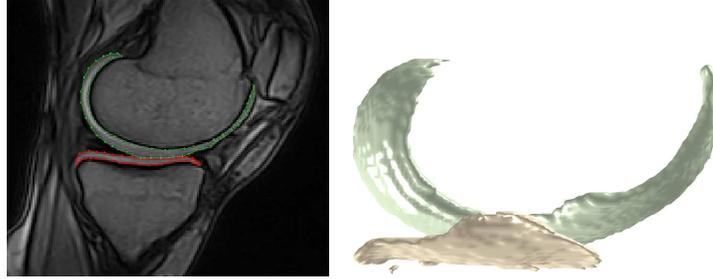


Fig. 1. Scan and manual cartilage segmentations for an example knee. A sagittal slice from the knee MR with manual outlines of tibial cartilage in red and femoral cartilage in green (left). The same cartilage sheets are visualized seen from below the center of the knee (right).

Even though computation power is relatively in-expensive, such long computation times are inconvenient in clinical studies using thousands of scans. The main computational steps and the approximate computation times are:

Initialization of KNN classifier	1 min
Feature extraction	9 min
Voxel classification	60 min

The initialization of the KNN classifier is not really important. It is fairly quick, and in a settings where multiple scans are to be segmented, the classifier only has to be initialized once. The feature extraction step is somewhat expensive computationally. However, in this paper we will not go into strategies for optimizing this step.

The voxel classification is the big culprit in terms of computation time. Every single voxel in the scan is classified as being tibial cartilage, femoral cartilage, or background. Since a scan has more than two million voxels, this takes time.

The basic idea behind the efficient voxel classification algorithm, that we present here, is simply not to classify all voxels but to focus mainly on the cartilage voxels. The algorithm is conceptually very simple. We sample a set of starting voxels randomly. Each starting voxel is then classified as either cartilage or background. If the voxel is cartilage, we continue with classification of the neighboring voxels. This expansion process continues until we find no more cartilage voxels.

This results in a number of connected regions of cartilage. Provided that our initial sampling of starting voxels hits each cartilage sheet in at least a single voxel, the resulting segmentation will be exactly like the one resulting from a full classification of all voxels. This can be ensured by simply making the initial random sampling not too sparse. For knee scans, we know that some parts of the tibial/femoral cartilage sheets will be fairly near the center of scan. We therefore sample fairly densely at the center of the scan and gradually more sparsely away from the center. Using a sampling probability of 5% for each voxel at the center,

the probability of missing all 4000 tibial cartilage voxels is in the order of one to the number of atoms in the universe (and even lower for the femoral cartilage).

The method in algorithmic detail is in algorithm 1. The process is illustrated for an example scan in figure 2. On this scan, the method samples 5912 seed points for the tibial cartilage of which 89 actually hit the tibial cartilage sheet (for the femoral cartilage sheet: 5913 and 137). A total of 9957 voxels out of 2 million are classified in order to find the 4316 tibial cartilage voxels (20826 are visited for the 11230 femoral cartilage voxels). In total, the Sample-Expand method thereby only performs 1% of the single voxel classifications performed by the original voxel classification segmentation method.

Algorithm 1 *Sample-Expand Voxel Classification*

Input: The original scan volume and a function that allows classification of a single voxel into background or an object label.

Output: A volume identical to the scan in size, with each voxel labelled as *Not Visited*, *Background*, or an object label L_i (where $L_1 \dots L_n$ correspond to the n objects in the scan).

1. Initialization:
 - (a) Initialize empty voxel queue VQ .
 - (b) Allocate label volume LV (size equal to scan volume) and initialize all voxels to *Not Visited*.
2. Sample seed voxels:
 - (a) Give the center voxel probability 5% and the corner voxels probability 0%. The remaining voxels get a probability between depending on the distance from the center voxel.
 - (b) Sample seed voxels such that each voxel is chosen with the probability given above. Enqueue all seed voxels into VQ .
3. While voxel queue VQ not empty:
 - (a) Dequeue voxel V from VQ .
 - (b) Classify the voxel into *Background* or an object label L using the single-voxel classifier.
 - (c) If the voxel is *Background*: mark $LV(V)$ as *Background*
 - (d) If the voxel is object: mark $LV(V)$ as L and enqueue all neighbors of V into VQ .

Scan Dimension: The algorithm is general for any scan dimension. The neighborhood list must simply be adapted. We use the term voxel here since the cartilage MR scans are 3D.

Neighborhood: The neighboring voxels in step 3d above can be chosen as either only direct neighbors or also diagonal neighbors (in 3D, 6 or 26 neighbors). This affects the connectivity of the resulting segmentation for special cases.

Sampling Strategy: For cartilage, we sample more densely at the center of the scan. For other structures, uniform sampling could be more appropriate.

Queue: The queue can either be by a traditional queue (first in, first out) or a stack (last in, first out). This is of no importance and simply corresponds to whether the voxels are searched used a breadth-first or depth-first traversal. We use a stack.

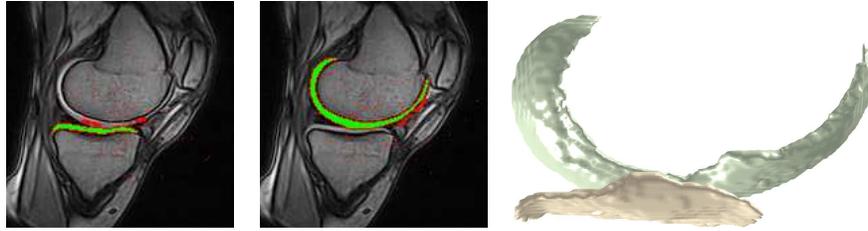


Fig. 2. *The Sample-Expand voxel classification algorithm is illustrated on the scan from figure 1. The voxels visited and classified during classification of the tibial cartilage are shown as green for cartilage and red for background (left) and next the same for the classification of the femoral cartilage (middle). The stray red voxels are seed voxels. The lumps of red voxels are areas that looked tibial/femoral but was later overruled by the other sheet. Most voxels are never visited by the classification method. Finally, a 3D illustration of the resulting tibial and femoral cartilage segmentation (right).*

5 Evaluation

We evaluate the segmentation method on the 114 scans and the 31 additional rescans from the evaluation set described in section 2.

First, we evaluate the correctness of the efficient Sample-Expand voxel classification algorithm. For all 114+31 evaluation scans, the segmentations resulting from the original and the new efficient method are identical.

Secondly, we evaluate the computational efficiency. For each scan, there are the three computational steps listed in section 4. As mentioned above, we could ignore the initialization of the KNN classifier since it only has to be done once — however, we have chosen to include it in the timing. We timed the computation time for each step for each of the 114 scans for both the original method and our new improved method based on the Sample-Expand classification algorithm on a standard desktop 2.8 Ghz PC. The mean times in seconds are listed below with the mean count of voxels being visited during the voxel classification stage by the methods:

	Original Method	Sample-Expand
Initialization of KNN classifier	44 s	44 s
Feature extraction	506 s	506 s
Voxel classification	3605 s	68 s
Voxels classified	100 %	0.8 %

Finally, we evaluate the volume quantification from the segmentations for reproducibility using the 31 scan-rescan pairs. Furthermore, we evaluate the ability to capture progression in OA by testing whether the healthy subjects have a larger cartilage volume than the OA subjects according to a t-test. The evaluation results for the medial tibial compartment are in figure 3. We focus on

the medial tibial compartment since tibial and femoral measurements are highly correlated and the tibial are more reproducible [21].

The evaluation shows that our volume quantification is quite reproducible (10% mean difference between the scan-rescan pairs) and that it is able to separate healthy from OA (statistically significant, $p = 0.001$). There is also a clear progression of cartilage loss with OA progression from Kellgren & Lawrence index 0 to 3.

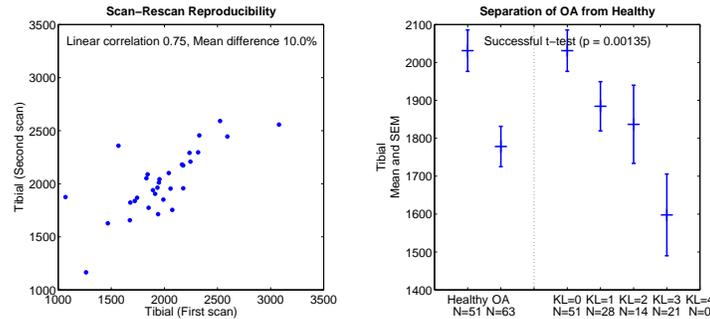


Fig. 3. The reproducibility of the automatic tibial cartilage volume quantification method is evaluated on the 31 scan-rescan pairs (left). The OA subjects have less cartilage on average, 1778mm^3 compared to 2031mm^3 for healthy (statistically significant, $p = 0.001$) (right, left of dotted line). The mean volume goes down to 1598mm^3 for the KL 3 group (right, right of dotted line). The cartilage volume is normalized by the width of the tibial plateau.

6 Conclusion

We present a novel algorithm for voxel classification based on sparse sampling followed by classification of voxels in regions with non-background only. The algorithm allows a segmentation speed-up from 70 to 10 minutes when used for segmentation of articular tibial and femoral cartilage from knee MRI — producing segmentations identical to the result of the original method.

For the classification step alone, the speed up is a factor of more than 50. This is due to the fact that only 0.8% of the voxels are actually classified. Thereby the algorithm overhead needed for bookkeeping of the voxel queue and the map of visited voxels is not problematic.

We evaluate the method for quantification of progression of osteoarthritis. The segmentation method, in which the sparse voxel classification method is incorporated, is able to separate healthy from OA subjects ($p = 0.001$) and has a fairly high reproducibility (the precision is 10% compared to 11% for the expert, manual segmentations). Thereby, the method looks very promising for use in clinical studies.

Future methodological work will include further optimization by only computing the features at the voxels that are actually classified. Future clinical work will involve evaluation of the method in longitudinal studies.

References

- [1] Buckwalter, J., Saltzman, C., Brown, T.: The impact of osteoarthritis - implications for research. *Clin Orthop Rel Research* **427** (2004)
- [2] Reginster, J.: The prevalence and burden of arthritis. *Rheumatology* **41** (2002)
- [3] Altman, R.: Measurement of structure (disease) modification in osteoarthritis. *Osteoarthritis and Cartilage* (2004)
- [4] Graichen, H., Eisenhart-Rothe, R.V., Vogl, T., Englmeier, K.H., Eckstein, F.: Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging. *Arthritis and Rheumatism* **50** (2004)
- [5] Pessis, E., Drape, J.L., Ravaut, P., Chevrot, A., Ayral, M.D.X.: Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and mri. *Osteoarthritis and Cartilage* **11** (2003)
- [6] Folkesson, J., Olsen, O.F., Dam, E.B., Pettersen, P.C., Christiansen, C.: Combining binary classifiers for automatic cartilage segmentation in knee mri. In: *ICCV, Computer Vision for Biomedical Image Applications*. (2005)
- [7] Folkesson, J., Dam, E., Olsen, O., Pettersen, P., Alexandersen, P., Christiansen, C.: Robust volume estimation of articular cartilage from knee mr scans. *Osteoarthritis and Cartilage* **13, supplement A** (2005)
- [8] Gandy, S., Dieppe, P., Keen, M., Maciewicz, R., Watt, I., Waterton, J.: No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. *Osteoarthritis and Cartilage* (2002) 929–937
- [9] Williams, T., Holmes, A., Maciewicz, R., Waterton, J., Taylor, C., Creamer, P., Nash, A.: Cartilage loss in osteoarthritis detected by statistical shape analysis of magnetic resonance images. *Osteoarthritis and Cartilage* **13, supplement A** (2005)
- [10] Raynauld, J.P., Kauffmann, C., Beaudoin, G., Berthiaume, M.J., de Guisei, J.A., Bloch, D.A., Camacho, F., Godbout, B., Altman, R.D., Hochberg, M., Meyer, J.M., Cline, G., Pelletier, J.P., Martel-Pelletier, J.: Reliability of a quantification imaging system using magnetic resonance images to measure cartilage thickness and volume in human normal and osteoarthritic knees. *Osteoarthritis and Cartilage* **11** (2003)
- [11] Graichen, H., Jakob, J., von Eisenhart-Rothe, R., Englmeier, K.H., Reiser, M., Eckstein, F.: Validation of cartilage volume and thickness measurements in the human shoulder with quantitative magnetic resonance imaging. *Osteoarthritis and Cartilage* **11** (2003)
- [12] Koo, S., Gold, G., Andriacchi, T.: Considerations in measuring cartilage thickness using mri: factors influencing reproducibility and accuracy. *Osteoarthritis and Cartilage* **13** (2005)
- [13] Stammberger, T., Eckstein, F., Englmeier, K., Reiser, M.: Determination of 3d cartilage thickness data from mr imaging: computational method and reproducibility in the living. *Magn Reson Med* **41** (1999)
- [14] Burgkart, Glaser, Hyhlik-Dürr, Englmeier, Reiser, Eckstein: Magnetic resonance imaging-based assessment of cartilage loss in severe osteoarthritis. *Arthritis & Rheumatism* **44** (2001)
- [15] Barbu-McInnis, M., Tamez-Pena, J.G., Totterman, S.: Focal cartilage defect progression detection: Measurement of precision and variation in natural characteristics of cartilage thickness maps derived from 3d mri data. In: *IEEE Int Conf on Image Processing*. (2004)
- [16] Li, K., Millington, S., Wu, X., Chen, D.Z., Sonka, M.: Simultaneous segmentation of multiple closed surfaces using optimal graph searching. In: *Information Processing in Medical Imaging: 19th International Conference. Volume 3565 of LNCS.*, Springer (2005)
- [17] Millington, S., Li, K., Wu, X., Hurwitz, S., Sonka, M.: Automated simultaneous 3d segmentation of multiple cartilage surfaces using optimal graph searching on mri images. *Osteoarthritis and Cartilage* **13, supplement A** (2005)
- [18] Kellgren, J., Lawrence, J.: Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* **16** (1957)
- [19] Dunn, T., Lu, Y., Jin, H., Ries, M., Majumdar, S.: T2 relaxation time of cartilage at mr imaging: comparison with severity of knee osteoarthritis. *Radiology* **232** (2004)
- [20] Arya, S., Mount, D., Netanyahu, N., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. In: *ACM-SIAM. Discrete Algorithms. Number 5* (1994)
- [21] Cicuttini, Wluka, Wang, Stuckey: Longitudinal study of changes in tibial and femoral cartilage in knee osteoarthritis. *Arthritis & Rheumatism* **50** (2004)

3D Shape Analysis of the Supraspinatus Muscle

Aaron D. Ward¹, Ghassan Hamarneh¹, Reem Ashry², and Mark E. Schweitzer²

¹ Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, V5A 1S6,
Canada

{award, hamarneh}@cs.sfu.ca, <http://mial.fas.sfu.ca/>

² Department of Radiology and Orthopedic Surgery, NYU School of Medicine,
Hospital for Joint Diseases, New York, NY, 10016, U.S.A.
reem.ashry@gmail.com, mark.schweitzer@nyumc.org

Abstract. Pathology of the supraspinatus muscle can involve tearing, which often leads to atrophy and/or retraction of the muscle. Retraction can be corrected through a pull forward operation in surgery, whereas atrophy is generally not correctable. It is therefore important to distinguish between retraction and atrophy. However, since both of these conditions are characterized by a reduction in size, we put forth a pilot study examining changes in 3D shape as they relate to pathological conditions. After segmenting the supraspinatus muscle surface from MRIs representing 57 patients, we compute several different 3D shape measures of the surfaces, and conclude that there are statistically significant differences in shape and size between pathology groups.

1 Introduction

The supraspinatus muscle is one of several muscles making up the rotator cuff in the shoulder (figure 1). Disorders of the supraspinatus muscle may involve tearing, which can lead to muscle retraction, atrophy, or both [1]. It is important to be able to distinguish between retraction and atrophy because retraction is a condition that is repairable by pulling the muscle forward in surgery, whereas atrophy is not a condition correctable by surgery. Since both of these conditions result in a reduction of the apparent size of the muscle, 3D shape analysis of the muscle is useful in order to discover shape characterizations that may assist the physician in distinguishing between these groups. Although shoulder arthroscopy is considered the gold standard for the evaluation of the rotator cuff, MR has an exceptionally high accuracy which has been accepted as a standard of reference for several prior papers (e.g. [2–5]).

In this study, we extract the 3D surfaces of the supraspinatus muscle from MRIs of a set of patients. Each patient’s data set is labeled according to pathology, forming several groups of patients. For each group, we compute a set of measurements of the 3D surfaces and report the differences observed between the groups.

The remainder of this paper is organized as follows. In sections 2 and 3 we discuss the specifics of the data sets used, and describe our methods for extracting the 3D surface of the muscle and the computation of the shape characteristics

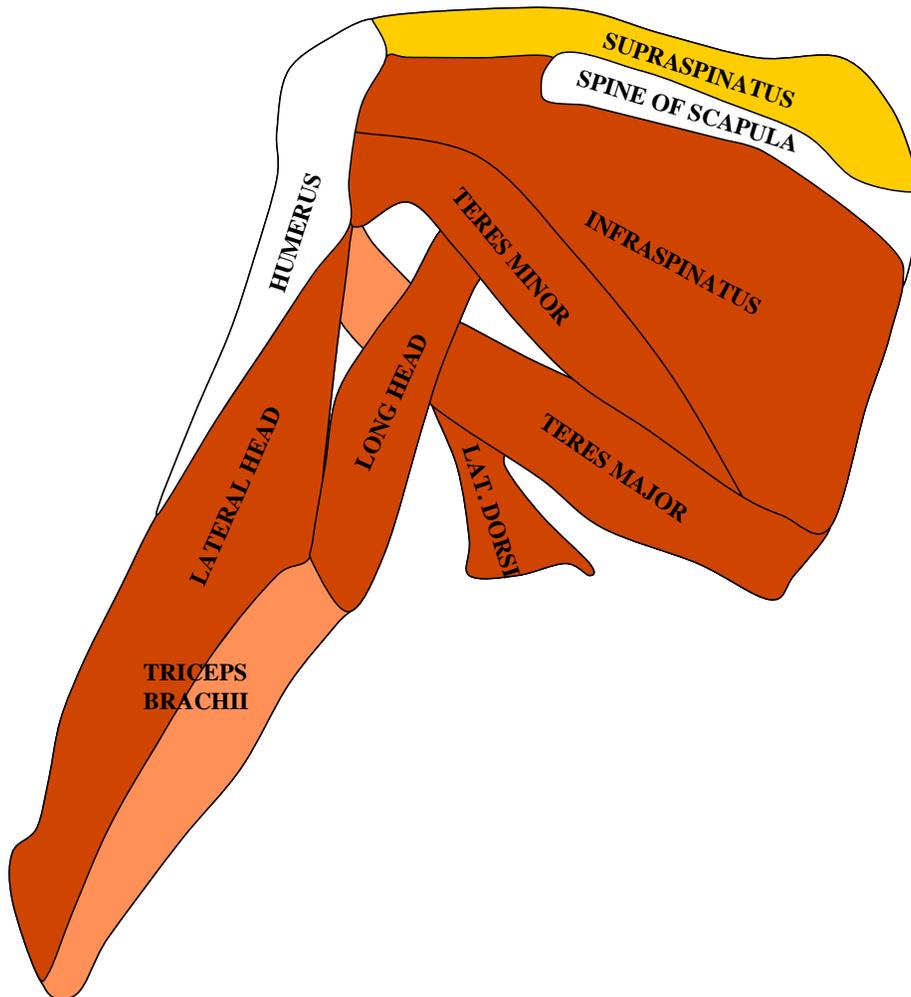


Fig. 1. Diagram of shoulder anatomy indicating location of the supraspinatus. Adapted from Grey's Anatomy [6].

of the surfaces. In section 4 we give our results, and in section 5 we make some concluding remarks and give some possible future avenues of research based on this work.

2 Material

We acquired MR images of the shoulder from 57 patients at 1.5T. Patients were consistently imaged in supine position, relaxed, and in minimal external rotation in order to normalize for effects of pose and gravity on the shape of the muscle.

The in-plane (sagittal) resolution of the data was 0.3-0.6 mm and the out-of-plane resolution was 3-5 mm. The patients were selected according to diagnoses made by examining the MRIs of the shoulder. The group of patients with torn supraspinati is composed of patients suffering different severities of disease, under the assumption that some of these patients would have visible muscle shape changes and some would not. The retraction group comprises patients with observed relevant mechanical changes to the muscle, and the atrophy group had relevant physiological changes. The control group was composed of patients with unstable shoulders, because they represent a more relevant cross section of the population than would normal subjects.

3 Methods

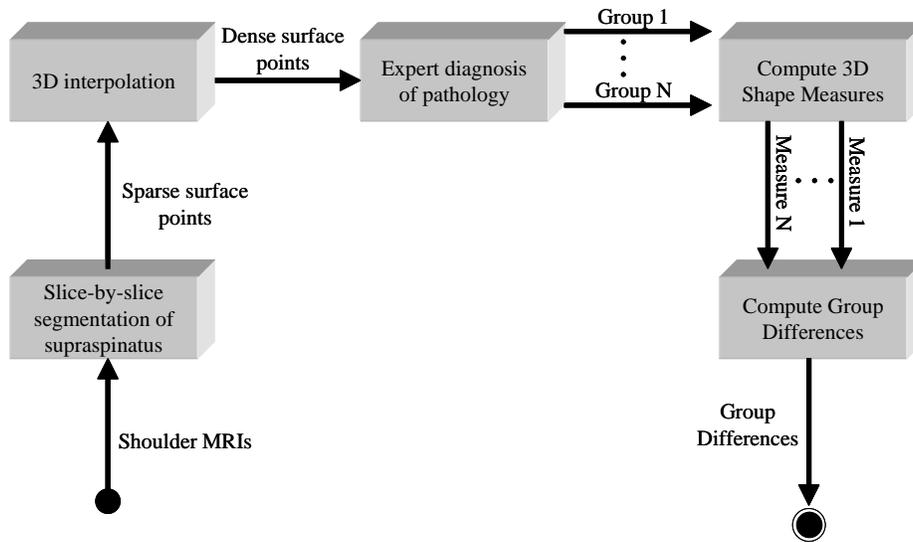


Fig. 2. Overall process used in this study. Beginning with a set of 1.5T MRIs of the shoulder, we segment the supraspinatus from each 3D image in a slice-by-slice manner. This yields a set of points that are dense in the imaging plane, but sparse out of the imaging plane due to large intra-slice spacing (figure 3(a)). We then interpolate these points to form a dense set of points forming a 3D surface of the supraspinatus using the Interpolation Module of the segmentation editor in the Amira software (Mercury Computer Systems, Inc) (figure 3(b)). Next, we divide the cases into groups according to expert diagnosis of pathology. We then compute an aggregate (mean) of several 3D shape measures for each group. Finally, we compute and report on differences between the groups.

The overall processing performed on these images is given in figure 2. Expert manual segmentation of the supraspinatus muscle was performed on the sagittal MR images in a slice-by-slice manner. The segmentation tool employed allowed the expert to select control points lying on the surface of the supraspinatus on each slice, and fit a parametric cubic spline curve to these points to guarantee smoothness. The tool performed the spline fitting interactively so that the expert could manipulate the control points until the curve accurately followed the contour of the muscle. It has been shown that intra- and inter-observer variation in supraspinatus contouring is less than 5% [7].

Due to the 3-5 mm inter-slice spacing in the data, the result of this slice-by-slice segmentation is a set of points which are dense within the imaging planes but sparse in the out-of-plane direction (figure 3(a)). 3D interpolation was therefore performed in order to obtain a set of points lying on the object surface that is dense along all axes using the Interpolation Module of the segmentation editor in the Amira software (Mercury Computer Systems, Inc), based on implicit, level-set based shape representation, similar to work by Turk and O'Brien [8] (figure 3(b)).

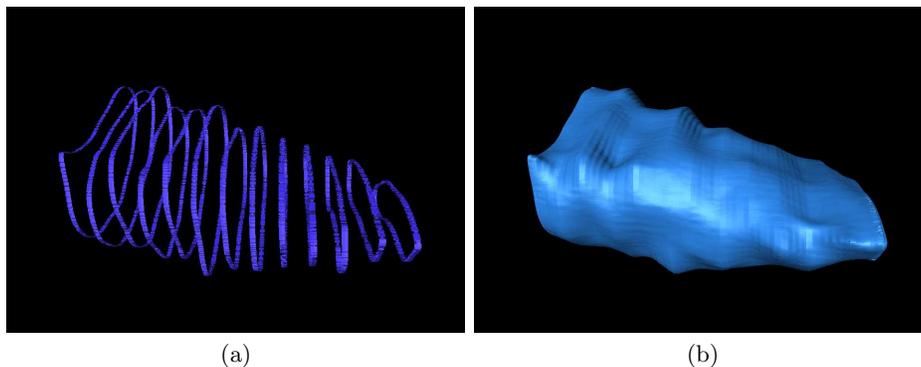


Fig. 3. Supraspinatus muscle surface extraction. (a) Contours resulting from expert segmentation of the supraspinatus, rendered using physical space coordinates. Large spaces between contours are due to the low out-of-plane resolution of the data. (Contours appear non-parallel because of perspective projection.) (b) Result of 3D interpolation of the contours in (a), yielding a dense set of points lying on the surface of the supraspinatus, rendered as a surface.

Next, the condition of the supraspinatus of each patient was assessed by an expert, and assigned to one of the following four groups: normal (N), full thickness tear (T), tear and atrophy (TA), tear and atrophy and retraction (TAR). These four groups represent those for which we wish to compute the shape differences.

We computed 11 different 3D shape measures for each data set:

- **Ratios of eigenvalues (3 measures):** We performed principal components analysis (PCA) [9] against the points lying on the surface of each shape, yielding three eigenvectors and corresponding eigenvalues. The eigenvalues $\lambda_1, \lambda_2, \lambda_3$ represent the variance of the supraspinatus surface points in the direction of the eigenvectors (which describe the main directions of variation) for each supraspinatus. They give an approximation to each supraspinatus by an ellipsoid, where $\lambda_1, \lambda_2, \lambda_3$ represent the lengths of the major axes of such an ellipsoid. Computing the three values of the ratios $\frac{\lambda_1}{\lambda_2}, \frac{\lambda_1}{\lambda_3}$, and $\frac{\lambda_2}{\lambda_3}$ yields measures of elongation of the object. For a spherical object we expect that $\lambda_1 \approx \lambda_2 \approx \lambda_3$. For a cylindrical object we expect that $\lambda_1 \gg \lambda_2 \approx \lambda_3$, and for a disk-like object we expect that $\lambda_1 \approx \lambda_2 \gg \lambda_3$.
- **Mean and standard deviation of distances to centroid (2 measures):** Here, we compute the centroid of all of the supraspinatus surface points, and compute the mean Euclidean distance from each surface point to this centroid, as a measure of size. We also compute the standard deviation of these distances as a measure of surface roughness/non-sphericity.
- **3D moment invariants (3 measures):** We compute three 3D moments that have been shown to be invariant to translation and rotation [10]. They are computed as follows:

$$J_1 = \mu_{200} + \mu_{020} + \mu_{002}$$

$$J_2 = \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2$$

$$J_3 = \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^2 - \mu_{020}\mu_{101}^2 - \mu_{200}\mu_{011}^2$$

where m_{pqr} is the 3D moment and μ_{pqr} is the 3D central moment as follows:

$$m_{pqr} = \sum_x \sum_y \sum_z x^p y^q z^r p(x, y, z)$$

$$\mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r p(x, y, z)$$

$$\bar{x} = m_{100}/m_{000}$$

$$\bar{y} = m_{010}/m_{000}$$

$$\bar{z} = m_{001}/m_{000}$$

$$p(x, y, z) = \begin{cases} 1 & \text{if } (x, y, z) \text{ is a surface point.} \\ 0 & \text{otherwise.} \end{cases}$$

- **Surface area, volume, and their ratio (3 measures):** We compute the surface area and volume of each supraspinatus in physical units, and take the ratio of surface area to volume.

4 Results

Table 1 gives the mean values of the measurements taken for each group. To gain some insight into how these values differ, we computed ratios of measurements of abnormals to normals (table 2). For example, in the first row, we take the mean of the measurements of all abnormal groups to get a single aggregate measure for the abnormals, and take the ratio to the normals. We can see, for example, in the first row of table 2 that the ratio of 0.4 for measurement $J3$ indicates that the mean of the normals was 250% larger than the normals. In subsequent rows of table 2 we give comparisons between specific pathology groups and the normals. For each measurement type, we performed a one-way ANOVA to test the null hypothesis that the means of the normal and pathological groups were the same. The p -values resulting from these tests are given in table 4. Measurements with p -values rejecting the null hypothesis ($p < 0.05$) are $\frac{\lambda_1}{\lambda_3}$, mean of distances to centroid, the 3D moment invariants, surface area, volume, and the ratio of surface area to volume. From this it appears that characteristics distinguishing normal cases from pathological cases are elongation, size, volume.

5 Conclusion

In this work, we presented a pilot study involving 57 cases of varying supraspinatus pathology: normal, tearing, retraction, atrophy, and allowable combinations of these diagnoses. We computed 11 shape characteristics based on 3D points from extracted surface muscles from MRI, and performed a statistical analysis to determine whether or not the measurements of the groups were significantly statistically different. The results indicate that there are significant differences between the groups, and the measures giving the best performance suggest that elongation, surface area and volume are good characterizations of shape for this anatomy.

Future work in this area includes the use of these shape characterizations to train and measure the performance of a classifier attempting to aid in diagnosis of pathology based on the 3D shape of the supraspinatus. Such a classifier would be of great use to a physician attempting to determine whether or not surgery is required to pull the muscle forward (supraspinatus retraction) or if surgery would be ineffective (supraspinatus atrophy). Another interesting area for future study would be to assess the impact on shape analysis of positional difference during imaging (e.g. the influence of internal and external rotation).

References

1. Robertson, P.L., Schweitzer, M.E., Mitchell, D.G., Schlesinger, F., Epstein, R.E., Frieman, B.G., Fenlin, J.M.: Rotator cuff disorders: interobserver and intraobserver variation in diagnosis with MR imaging. *Radiology* **194**(3) (1995) 831–835
2. Zlatkin, M., Iannotti, J., Roberts, M., Esterhai, J., Dalinka, M., Kressel, H., Schwartz, J., Lenkinski, R.: Rotator cuff tears: Diagnostic performance of MR imaging. *Radiology* **172** (1989) 223–229

	1	2	3	4	5	6	7	8	9	10	11
<i>N</i>	2.4	5.7	2.5	2.0	0.7	1.2×10^5	5.3×10^9	7.6×10^{13}	77.0	24.8	0.3
<i>T</i>	2.6	6.7	2.6	1.8	0.6	8.7×10^4	2.7×10^9	2.6×10^{13}	62.1	16.3	0.2
<i>TA</i>	4.0	10.4	2.7	1.8	0.7	6.5×10^4	1.5×10^9	1.3×10^{13}	58.2	13.7	0.2
<i>TAR</i>	3.0	7.4	3.0	1.6	0.6	3.7×10^4	4.6×10^8	1.7×10^{12}	42.7	7.9	0.2

Table 1. Mean values of each of the measurements for each of the groups. Refer to table 3 for meanings of the numbered column headings.

	1	2	3	4	5	6	7	8	9	10	11
<i>mean(Abnormal)/N</i>	1.2	1.3	1.1	0.9	0.9	0.7	0.5	0.4	0.8	0.6	0.8
<i>T/N</i>	1.1	1.2	1.1	0.9	0.9	0.7	0.5	0.3	0.8	0.7	0.8
<i>TA/N</i>	1.6	1.8	1.1	0.9	1.0	0.6	0.3	0.2	0.8	0.6	0.7
<i>TAR/N</i>	1.2	1.3	1.2	0.8	0.8	0.3	0.09	0.02	0.6	0.3	0.6

Table 2. Comparison of the mean values of the measurements for each group. The values in the first row are computed by first taking the mean of the measurements across all abnormal groups. We then take the ratio of the mean measurement for the abnormal to the mean of the normals. The remaining rows show the ratios for all of the individual abnormal groups. Refer to table 3 for meanings of the numbered column headings.

<i>Measurement number</i>	<i>Description</i>
1	λ_1 / λ_2
2	λ_1 / λ_3
3	λ_2 / λ_3
4	Mean of distances to centroid (cm)
5	Standard deviation of distances to centroid (cm)
6	J_1
7	J_2
8	J_3
9	Surface area (cm ²)
10	Volume (cm ³)
11	Surface area / Volume (1 / cm)

Table 3. Meanings of the column headings given in tables 1 and 2. This legend is given in a separate table here due to space considerations.

	1	2	3	4	5	6	7	8	9	10	11
<i>p-value</i>	0.0733	0.0071	0.5849	0.0198	0.3304	0.0093	0.0118	0.0129	0.0010	0.0007	0.00001

Table 4. p-values resulting from one-way ANOVA test for each measurement, testing the null hypothesis that the means of all of the groups are the same. Refer to table 3 for meanings of the numbered column headings.

3. Bachmann, G.F., Melzer, C., Heinrichs, C.M., Moehring, B., Rominger, M.B.: Diagnosis of rotator cuff lesions: Comparison of US and MRI on 38 joint specimens. *European Radiology* **7**(2) (1997) 192–197
4. Quinn, S.F., Sheley, R.C., Demlow, T.A., Szumowski, J.: Rotator cuff tendon tears: Evaluation with fat-suppressed MR imaging with arthroscopic correlation in 100 patients. *Radiology* **195** (1995) 497–500
5. Swen, W.A.A., Jacobs, J.W.G., Algra, P.R., Manoliu, R.A., Rijkmans, J., Willems, W.J., Bijlsma, J.W.J.: Sonography and magnetic resonance imaging equivalent for the assessment of full-thickness rotator cuff tears. *Arthritis and Rheumatism* **42**(10) (1999) 2231–2238
6. Gray, H.: *Anatomy of the Human Body*, 20th ed. Lea and Febiger, Philadelphia, U.S.A. (1918)
7. Lehtinen, J.T., Tingart, M.J., Apreleva, M., Zurakowski, D., Palmer, W., Warner, J.J.: Practical assessment of rotator cuff muscle volumes using shoulder MRI. *Acta Orthop Scand* **74**(6) (2003) 722–729
8. Turk, G., O'Brien, J.F.: Shape transformation using variational implicit functions. *Proceedings of ACM SIGGRAPH* (1999) 335–342
9. Jolliffe, I.T.: *Principal Components Analysis*. Springer-Verlag, New York, U.S.A. (1986)
10. Sadjadi, F.A., Hall, E.L.: Three-dimensional moment invariants. *IEEE PAMI* **2**(2) (1980) 127–136

Three dimensional dynamic model with different tibial plateau shapes: Analyzing tibio-femoral movement

Akalan E.A.¹, Özkan M.², Temelli Y.^{1*}

¹ Gait analysis Laboratory, Dept. of Orthopedics and Traumatology, Physical Therapy Center,

Dept. of Developmental Child Neurology of Istanbul Medical University

^{1*} Dept. of Orthopedics and Traumatology of Istanbul Medical University

² Institute of Biomedical Engineering of Boğaziçi University

Abstract

The objective of the present study is to create a dynamic 3D knee model which represents tibio-femoral joint surfaces, bones and ligaments by consideration of their geometric and material properties to simulate 0°-90° passive knee flexion. Tibial plateaus of the tibia and condyles of femur are modeled as ellipsoids as described in the literature. The contact forces between tibia and femur are defined as frictionless mathematical model. Anterior, posterior cruciate ligaments, medial, lateral collateral ligaments are represented as non linear elastic springs. Knee flexion with and without internal-external torque are simulated, and the results are compared with the literature for sloped and flattened medial tibial plateau models. As a result, normal internal rotation of tibia and adduction ranges are achieved for unloaded condition in flattened model, but the knee flexion with forced internal/external rotation are out of normal range for both models.

Keywords: Knee, kinematics, anatomical dynamic modeling

1. Introduction

It is well known that the mathematical models play an important role for the understanding of complicated biological structures. The human knee has a complex anatomical structure and complicated three dimensional movements. Not only a faithful description of normal function, but also identification of and treatment of dysfunction presents many problems [23].

The mechanical functions of the structures at the tibio-femoral joint include guiding the relative motion of the tibia and femur and transmitting load between these bones. Surgeons seek to preserve or restore both of these mechanical functions during joint reconstruction or replacement. During surgery, both the geometry of joint structures and their mechanical properties are often changed. How geometric changes to the joint structures affect the three dimensional movement of the knee has not been explained completely. It is not clear that which anatomical structures guide the knee in passive flexion and how their geometric arrangement produces the unique path of passive knee motion [22]. It has proved challenging to measure and then to depict knee joint motion [9]. The four bar theory based kinematical models developed by Zuppinger [1904], Menschik [1974] and Huson [1974]. In this type of model force action in the structures of the joint is not considered.

Crowninshield [1976] studied on force action between structures but kinematic behavior of the knee is considerably simplified. Morrison represented the knee as a simple hinge joint [1]. In the model of Crowninshield [1976], the motions in the joint were based on experimental data in the literature. However the contribution of the curved joint surfaces to the mechanical behavior was ignored in these models [9,14, 25].

Andriacchi [1977] developed a model to analyze the movement and the force changes of the knee by employing finite element method. The ligaments, joint capsules modeled as nonlinear springs while the joint surfaces were modeled by a number of flat surfaces. The studies of the kinematic knee modeling continued until wide spreading of MRI scanning [9, 20].

MRI screening provided a huge improvement on analyzing 3D knee kinematics. Freeman [2000] defined the natural knee movements by MRI scanning. As a result of this study, defining the natural movement of the knee play a very important role for understanding the effectiveness of prosthesis, rehabilitation and surgery on joint pathologies. Unfortunately simulation of passive knee movement by representing natural anatomic structures and their 3D geometries has not been published yet. Even though tibial plateau was represented by flat [1, 11, 22] and uneven [9, 11, 12, 17] surfaces by different

studies, there is no study worked on superiority of uneven and flat surfaces. There has already been extensive work on the kinematics of tibio-femoral joint, but geometric representation of Freeman and Pinskerova (2005) was not studied which is the most detailed geometric shape representation of the femoral and tibial surface to analyze tibio-femoral flexion.

The objective of our study is to create a dynamic 3D knee model which represents tibio-femoral joint surfaces, bones and ligaments by consideration of their geometric and material properties to simulate 0°-90° passive knee flexion.

2. Materials and Methods

The model created as the characteristics; 1.80 cm tall, 80 kilograms, 18 years old human's volume rendered shell files, the locations of their center of mass, inertial moments of the right femur and right tibia are provided from BRG*. All the files are imported into MSC. ADAMS software [21, 24]. The foot segment was also imported to the ADAMS software except the shell file to decrease the load to the computer during simulation.

2.1 Geometric and contact conditions

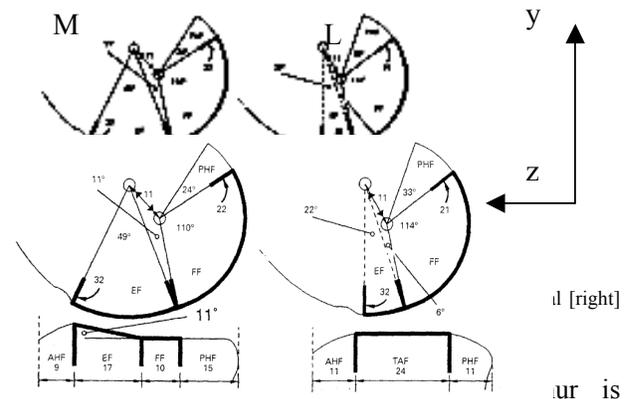
Femoral condyles represented as spheres as defined by Freeman and Pinskerova [2005]. In this study, medial femoral condyle is represented as two ellipsoids; 22mm for flexor facet radii, 32mm for extensor facet radii. Lateral femoral condyle ellipsoid is represented as two spheres 21mm for flexor facet radii, 32mm for extensor facet radii. According to the work medial femoral condyle divided in to 3 sections; Extensor Facets 49°, Flexor facets 110° and Posterior horn facets 24°. The lateral femoral condyle has 3 sections; Extensor Facets [EF] 32°, Flexor facets [FF] 114° and Posterior Horn Facets [PHF] 33°. Medial tibial plateau has 4 sections; Anterior Horn Facets [AHF] 9mm, Extensor facet [EF] 17mm, Flexor Facets [FF] 10mm, Posterior Horn Facet [PHF] 15mm. Extensor facet slopes upwards and forwards by 11° relative to posterior, roughly horizontal surface. Lateral tibial plateau has 3 sections; Anterior Horn Facet 11mm, Tibial Articular Facet [TAF] 24mm, Posterior Horn Facet 11mm [9, 11] (Figure 1).

In the present work femur is assumed to be fixed and tibia moves relative to the femur and gravity is assumed to be opposite direction in order to contact

continued between femur and tibia during flexion. The tibia is assumed to begin its motion from rest while the knee was fully extended.

Friction forces are neglected because of the extremely low coefficient of friction of articular surfaces [1].

In the absence of joint axial compressive loads the effect of meniscectomy on joint motion is minimal compared with that of cutting ligaments [18]. Since loading conditions are limited to those where the knee joint is not subjected to external axial compressive loads, the menisci thus not included in the present model [1].



formalized as:

$$F_n = k [g^e] + \text{step}[g, 0, 0, d_{\max}, c_{\max}] dg/dt \quad [1]$$

where k [46.58 N/mm] is the stiffness coefficient, g is penetration [mm], e [4] is a force exponent, d_{\max} [1cm] is the penetration limit c_{\max} is the maximum damping coefficient [97.19 N/mm/sec], dg/dt is the penetration velocity.

Patton [1993] took the penetration length as 1 cm for foot, but Blankevoort et al [6] revealed that the articulate thickness for the knee is 2 mm, so 50% of the cartilage thickness as penetration length is taken.

The single component force is applied to flex the knee from the center of mass of the tibia as presented in Abdel-Rahman and Hefzy [1998] and Blankevoort et.al. [1, 5-7].

$$F_q = A e^{-4.75[t/t_0]^2} \sin[\pi t/t_0] \quad [2]$$

where A and t_0 are the amplitude and the pulse duration, respectively. Forcing pulses of this can be simulated experimentally. Forcing pulse duration is assumed as 140N and 0.1 respectively.

A co-ordinate system for the normal knee based on posterior femoral circles has been proposed by McPherson et. al. [2004]. The origin is located at the center of the posterior spherical portion of the

* LifeMOD Biomechanics Modeler database.

medial femoral condyle so that the origin of the system approximately coincides with the center of rotation of the knee as defined in Freeman and Pinskerova [2005].

2.2 Ligamentous forces

The model includes 13 nonlinear spring elements which represent different ligamentous structures and capsular tissue posterior of the knee joint. Four of them stand for the respective anterior and posterior fiber bundles of anterior cruciate ligament [ACL] and the posterior cruciate ligament [PCL]; another four represent the deep, oblique, anterior and posterior fiber bundles of the medial collateral ligament [MCL], one element represents the lateral collateral ligament [LCL], and four elements represent the medial, lateral, oblique fiber bundles of posterior part of the capsule [CAP]. The local coordinates of the femoral and tibial insertion sites of the ligamentous structures are specified according to the data available in the literature [3, 5, 17].

The Ligament assumed to be a line element extending from the femoral origin to tibial insertion, wrapping around the bone surfaces is not taken into account.

In the present study the ligaments are determined according to the force length relationship as Wissman et al [1980]

$$F_j = \begin{cases} 0; & \epsilon_j \leq 0 \\ K1_j(L_j - L_{0j})^2; & 0 = \epsilon_j \leq 2\epsilon_1 \\ K2_j[L_j - (1 + \epsilon_1)L_{0j}]; & \epsilon \geq 2\epsilon_1 \end{cases}$$

where ϵ_j is the strain in the j th element, $K1_j$ and $K2_j$ are the stiffness coefficients of the j th spring element for the parabolic and linear regions, respectively, and L_j and L_{0j} are its current and slack lengths, respectively. The linear range threshold is specified as $\epsilon_1: 0.03$ [1, 4, 15].

Values of the stiffness coefficients of the spring elements used to model the different ligamentous structures are taken from the data available in the literature [1]. The slack length of each spring element is obtained by assuming an extension ratio ϵ_j at full extension and using it to evaluate the spring element's slack length, L_{0j} , from its length at full extension which can be calculated from the coordinates of the attaching points. The values of the extension ratios are specified according to the data available in the literature [3,5]. It is verified that the selected extension ratios did not produce nonanatomical strains [i.e., strain levels that indicate ligamentous failure] over the whole range of motion.

3. Results

The comparison of internal and external rotation during passive knee flexion data are shown in Figure 2. The behavior of the graphics of flat surface simulation and natural rotation of tibia are different. Tibial rotation is lower than normal during knee flexion in Freeman and Pinskerova [2005] based simulation. Tibial internal rotation is around 2 - 4° between 10 - 40° of knee flexion and then tibia rotated externally between 40 - 65°. Even though tibial rotation is trying to catch the normal rotation after 65° knee flexion, tibia has reached only 12° internal rotation which is below the normal according to Wilson et al. 2000.

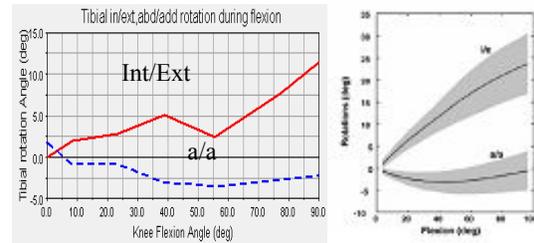


Figure 2. Comparison of int-ext rotation during knee flexion. Freeman and Pinskrova based simulation (left), and Knee rotaion from Wilson et. al.(right)

In simulation based on Freeman and Pinskerova translations of tibial attachment of pACL is in range except the antero-posterior translation in first 20° flexion. Tibia translated forward during the first 20° where it is expected to translate to the backward (Figure 3).

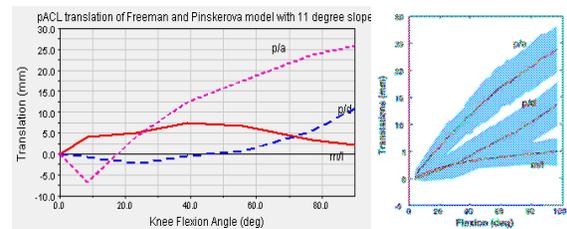


Figure 3. Comparison of simulation with 11° slope [left] and Wilson et al's 2000. Freeman and Pinskrova based simulation (left), and Knee rotaion from Wilson et. al.(right)

11° slope of tibial external facet is removed and the surface of the external facet and the posterior part of the tibia provided to be on the same level. So that magnitude of tibial internal rotation increased (Figure 4).

The internal rotation and adduction during the knee flexion is in range of normal knee which revealed in Wilson et al 24.

The translation of tibial insertion point of pACL relative to femur during knee flexion is in the range of normal translation described in Wilson et al's work

(Figure 5). The backward translation of pACL attachment reduced.

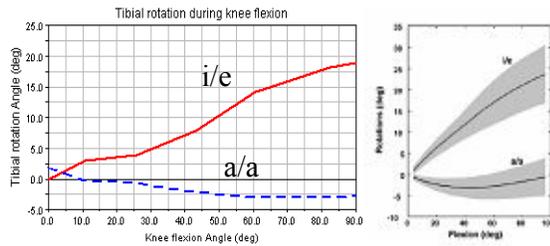


Figure 4. Natural internal rotation during knee flexion after 11° slope is removed

The translation of tibial insertion point of pACL relative to femur during knee flexion is in the range of normal translation described in Wilson et al's work (Figure 5). The backward translation of pACL attachment reduced.

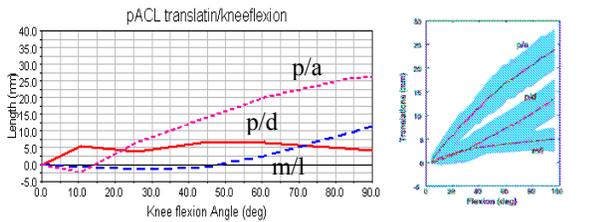
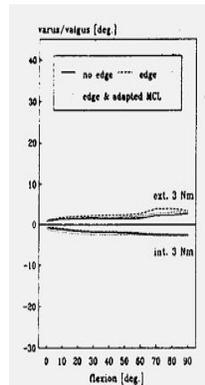


Figure 5. The translation of pACL during knee flexion. Freeman and Pinskrova based simulation (left), and Knee rotation from Wilson et al.(right)

Position of medial and lateral condyle contact point is simulated as in Pinskerova 2000. Medial femoral EF to FF rock occurred at 36° which is late according to the literature. Femoral FF is in contact with tibial FF from 36° to 90° knee flexion. Lateral Femoral EF to FF rock occurred around 5° and femoral FF is in contact with the tibia from 5 to 90° knee flexion in sloped model.

EF to FF rock is occurred at 10° flexion in flattened tibial surface. After 10° FF contact lasts till 90° as in the literature [11]. EF to FF rock occurred at 30° flexion and FF contact with femoral FF till 120° flexion as revealed in Pinskerova et al work 2000 for the flattened medial tibial surface.

To analyze the behavior of the knee under loading conditions, 3Nm internal and external force applied from the location of tibial center of mass as described in Blankevoort et al [4-7]. The results for the model with and without 11° slope is compared with the literature results (Figure 6) [4,5].



Tibial internal rotation is increased linearly and reached 60° tibial rotation at 90° knee flexion which is above the normal range, external rotation with 3Nm external torque is in normal range for the Iwaki and Pinskerova based simulation [11].

Even though tibial internal rotation with 3Nm torque is in normal range, tibial external rotation is lower than normal after 35° of flexion for external rotation torque in simulation without 11° tibial slope.

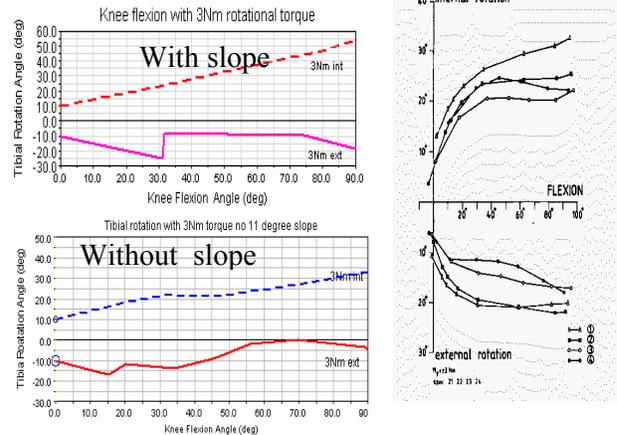
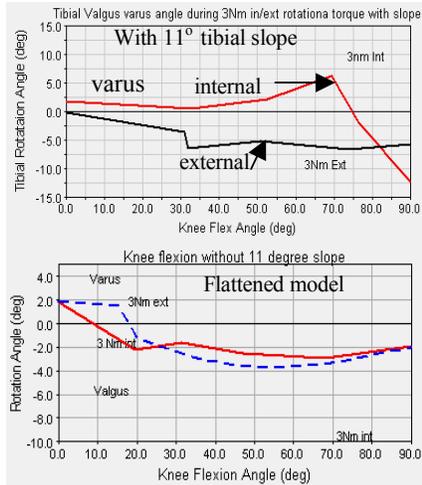


Figure 6; Comparison of 3Nm loading torque with blankevoort et al Freeman and Pinskrova based simulation (left), and Knee rotation from Blankevoort et al.[1991](right)

Knee varus / valgus rotation during the application of 3 Nm internal/external rotational torque is also studied. The results are compared with Blankevoort et al [6] (Figure 6).

According to the results, valgus occurs with internal load and varus occurs with external load till 75° knee flexion for 11° tibial slope. Valgus rotation is observed After 75° flexion for sloped model.

Valgus rotation is seen for both internal and external load after 30° flexion for flattened medial tibial surface model. Before the 30° flexion valgus rotation has occurred by applying internal rotation torque(Figure 7).



(a)

(b)

Figure 7. Comparison of internal and external 3Nm loading torque on a) valgus / varus rotation with and without 11° tibial slope, b) Blankevoort et al. (right).

4. Discussion

A review of the literature reveals that there is no published anatomical dynamic knee model based on Freeman et al works [4] which describes the anatomy of the articular surfaces and their movement in the normal tibio-femoral joint by some combination of MRI, CT, RSA or fluoroscopy. During building anatomical dynamic knee model, natural tibial rotation is achieved by removing 11° tibial slope. The aim of the study is to observe the effect of 11° tibial slope to the tibio-femoral movement and to build the most carefully sectioned tibio-femoral dynamic knee model by guidance of Freeman et al 2005.

Tibial internal rotation during knee flexion is lower than the normal with 11° slope. The slope on the medial tibial surface decreases the rotation by blocking tibial internal rotation and allowing external rotation in 0 - 40° knee flexion. 11° slope tackles the medial flexor facet and pushes the tibia into its external rotation between 40 - 65°. Contact point moves to the posterior tibial surface and faster internal rotation is occurred after 65° knee flexion.

For the flat surface tibia demonstrated natural internal rotation. Abduction and adduction rotation with and without 11° medial tibial slope are in the allowed limits (Figure 2, 4).

Translations of most posterior location of anterior cruciate ligament is followed. Antero-posterior translation is in normal range except first 20° part of knee flexion in original model. Tibia moves forward instead of backward in the sloped model in 20° knee flexion. The backward movement might be

resource from hitting the femoral EF to the tibial slope and produces a contra force to lead forward motion. It may be a cause of damping coefficient of contact formulation which make tibia jump on femoral EF. Smoother forward movement occurs due to the absence of the slope, in flat tibial surface.

Position of each part of the lateral and medial plateau relative to femur during knee flexion is specially revealed in Pinskerova et al 2000. According to the literature the medial components; between 0 - 10°: Femoral EF contacts with tibial EF, from 10 - 30°: EF should be rock to FF, from 30 - 120°: Femoral FF should in contact with tibial FF.

EF to FF femoral rock occurs at 36° in simulation with sloped model which was late relative to the normal but EF to FF rock occurs at 30° knee flexion in flattened tibial model. After that, the rock contact match with literature [11, 17].

For the lateral compartment; between 0 - 10°: EF, or FF in absence of EF, should be in contact which means EF to FF rock needs to be between the related range [5]. Femoral EF to FF rock is occurred at 5°, 10° knee flexion in both sloped and in flattened model respectively which are within the allowed limits. between 10 to 90°: FF is in contact with the tibia, over 90°: tibial contact shared with PHF [17]. FF is in contact with tibial FF after the 10° as literature for both models.

To analyze loading conditions on both model, 3Nm rotational torque was applied [4, 5]. Higher and a linear manner tibial internal rotation is seen with internal load and lower external rotation observed for external load in sloped model. The reason of increased internal rotation might be tight PCL. Changing force during knee flexion for aPCL was demonstrated in Figure 8.

Tibial internal rotation is in range during internal load in flattened tibial model. However the external rotation was decreased in flattened model.

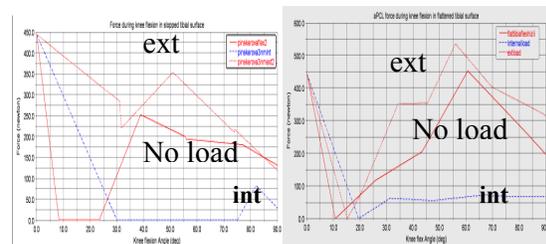


Figure 8. Comparison of knee flexion and aPCL force relationship for sloped (left) and flattened (right) models.

aPCL force increases while during external tibial rotation which conflicts with Moglo et. al. for both models [15]. Generally, the strain of the ligaments quite high than in the literature [1] for the present simulations. The difference of internal-external

rotation between the model and the literature needs to be minimized by variations of the reference strains of the ligaments as revealed in Blankevoort et al. 1991.

The coefficients of the ligaments are obtained directly from the Abdel-Rahman and Hefzy 1998, but the co-ordinates of the ligament attachments taken from Crowninshield et al 1976. In Crowninshield et al 1976 medial collateral ligament is represented as four fibers which are anterior, posterior, deep and oblique although Abdel-Rahman and Hefzy 1998 MCL represented as three fibers, posterior fiber is missing.

In our study anterior fiber coefficients is used for posterior and anterior fibers to compensate the absence of posterior fiber. Arcuate popliteal ligament of posterior capsule which represented in Abdel-Rahman and Hefzy 1998 is not included because of absence of location co-ordinates in Crowninshield et al 1976. The reason of obtaining coordinates of ligament attachments from Crowninshield et al 1976 is the origin co-ordinates are not defined clearly in Abdel-Rahman and Hefzy 1998.

Valgus rotation occurs by applying external rotational torque, varus by internal rotational torque for normal knee [4, 5]. Valgus rotation and first 75° varus rotation occurred by $\pm 3\text{Nm}$ rotational torque are in normal range for sloped model. The reason of valgus rotation after 75° knee flexion is the sliding of medial tibial plateau backward and loosens contact with flexor facet.

Valgus rotation is seen with internal rotation torque as normal although varus rotation is only seen in 30° flexion in flattened tibial model. Valgus rotation is observed by external rotation load after 30°. The reason of valgus rotation might be resource from excessive ligament strain or anatomical inefficiencies.

The coronal and transverse plane representation is not clear as sagittal plane in Freeman, Pinskerova and Iwaki [9,11,17]. The radii of the spheres for representing tibio-femoral articular surface are also used for coronal plane radii. The radius differences might perform the unnatural valgus rotation by application of external load.

The one of the limitations of the present study is neglecting friction force because of the extremely low coefficient of friction of the articular surface [1]. The MCL is also modeled as a straight line segment connecting the femoral and tibial attachments, while the natural MCL wraps around the tibial plateau [22]. Meniscus and joint capsules are not modeled because of their complex structures [2]. Some loads (around 4kg which is weight of shank and foot) are still applied to the knee as it was flexed due to provide continued contact between tibia and femur by assuming the gravity as upward direction.

The results show that tibial internal rotation during flexion is within the normal limits for the model without 11° anterior tibial slope. Posterior translation of tibial attachment of pACL and sliding and rolling motion of the tibia over femur is near normal range in flattened tibial plateau model. Both models have showed different behaviors in loading conditions. Flattened tibial plateau model is simulated the knee motion within the normal range for unloaded condition.

The primary feature of the three-dimensional dynamic anatomical modeling of the knee is variation of ligament strain to achieve reasonable loading behavior for the knee as revealed in Blankevoort et al. (1991). Modeling of the meniscus, friction, defined in detail contact should be well studied.

5. Acknowledgement

This project is supported in part by TUBITAK (The Scientific and Technological Research Council of Turkey) grant number 104S508. The project has been coordinated at the Institute of Biomedical Engineering in Boğaziçi University. Gait analysis experiments have been conducted at the Motion Analysis Laboratory of Istanbul University, School of Medicine.

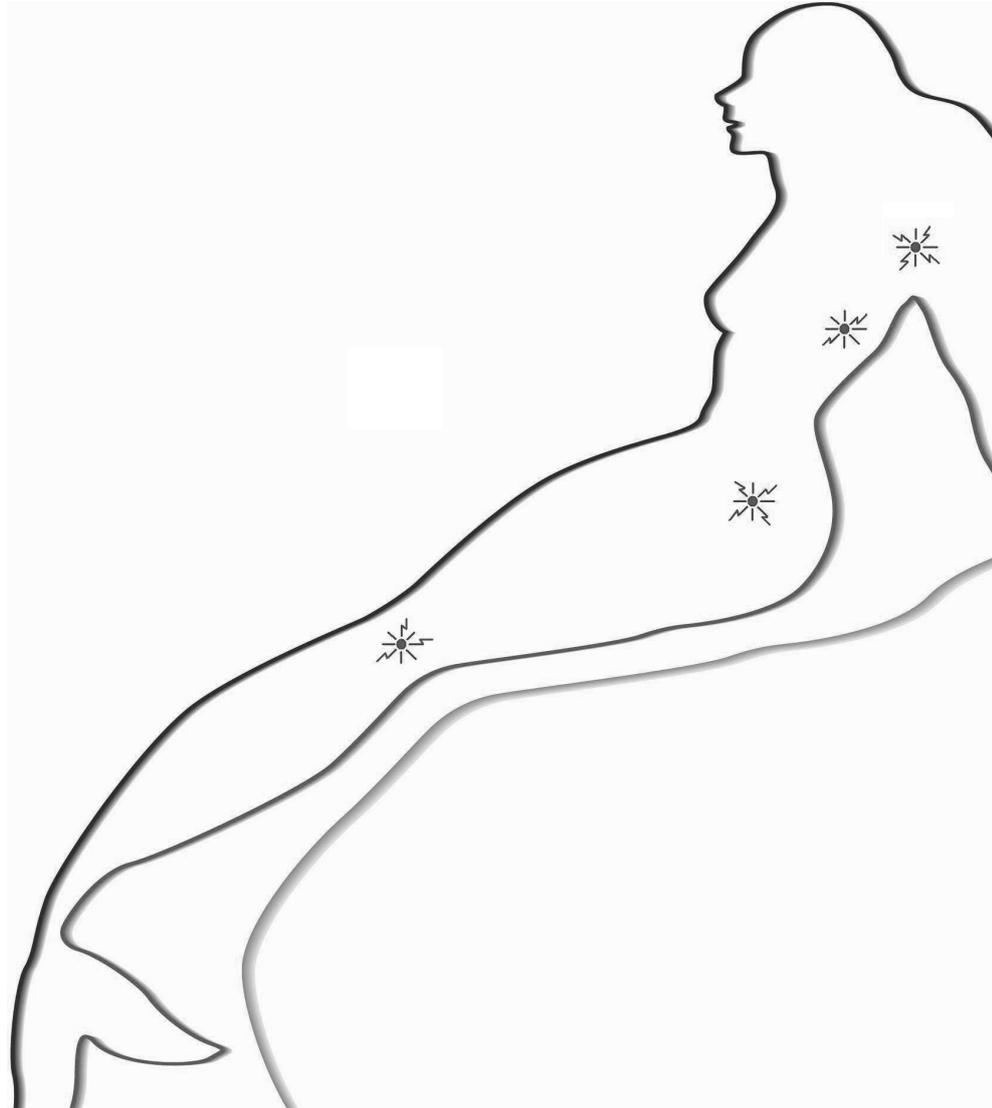
6. References

1. Abdel-Rahman, E. M., Hefzy, M. S. 1998. Three-dimensional dynamic behavior of the human knee joint under impact loading. *J. Biomech* 20, 276-90.
2. Andriacchi, T. P., Mikosz, R. P., Hampton, S. J., Galante, O. A., 1977. Statically indeterminate model of the human knee joint, *Biomechanics symposium AMD.23*, 227-239.
3. Beynon, B., Yu, J., Huston, D., Fleming, B., Johnson, R., Haugh, L., Pope, M.H.1996. A sagittal plane model of the knee and cruciate ligaments with application of a sensitivity analysis. *J. Biomech. Eng.* 118, 227-239.
4. Blankevoort, L., Huiskes, R., De Lange, A. 1991. Recruitment of the knee joint ligaments *ASME J. Biomech Eng.* 113(1), 94-103.
5. Blankevoort, L., Huiskes, R., De Lange, A., 1988. The envelope of passive knee joint motion. *Journal of Biomechanics* 21, 705-720.
6. Blankevoort, L., Kuiper, J. H., Huiskes, R., Grootenboer, H. J.1991. Articular contact in a three-dimensional model of knee. *J. Biomech* 24, 1019-31.

7. Blankvoort, L., Huijskes, R., 1991. Ligament-bone interaction in a three-dimensional model of the knee. *J. Biomedical Eng.* 113, 263-269
8. Crowninshield, R., Pope, H., Johnson, R. J. 1976. An analytical model of the knee. *J. Biomech* 9, 397-405.
9. Freeman, M. A. R., Pinskerova, V. 2005.[0] The movement of the normal tibio-femoral joint. *J. Biomech.* 38 (2), 197-208.
10. Huson, A., Biomechanische probleme des kniegelenks. 1974. *Orthopade* 3, 119-126.
11. Iwaki, H., Pinskerova, V., Freeman, M.A.R. 2000. Tibia-femoral movement 1: the shapes and relative movements of the femur and tibia in the unloaded cadaver knee. *J Bone Joint Surg. (Br)* 82-B, 1189-195.
12. Kapandji, I. A. 1970. The knee. In: Kapandji, I. A. *The physiology of Joints*, 2nd edition. Churchill Livingstone, 72-134
13. Mc Pherson, A., Karrholm, J., Pinskerova, V., Sosna, A., Martelli, S., 2005. Imaging knee motion using MRI, RSA/CT and 3Ddigitization. *Journal of Biomech* 37, 38(2), 263-8
14. Menschik, A., Teil, Z. 1974. Mechanic dess Knigelenks. *Orthop.* 112, 481-495
15. Moglo, K. E., Shirazi-Adl, A., 2003. Cruciate coupling and screw home mechanism in passive knee joint during extension-flexion *J. Biomech.* 38 (5), 1075-83
16. Patton, J. L. 1993. Forward dynamic modeling of human locomotion. Master thesis. pp:32
17. Pinskerova, V., Iwaki, H., Freeman, M. 2000. The shapes and relative movements of the femur and tibia in loaded cadaveric knee: A study using MRI as an anatomical tool. In: Insall, J.N, Scott, W.N. *Surgery of the knee* eds. 3rd edition. Philadelphia: W.B Saunders Co.
18. Seedhom, B.B., Suda, Y. 2000. Axis of tibial rotation and its change with flexion angle. *Clin Ort.* 341,178-182.
19. Shelburne, K. B., Pandy, M.C. 1997. A musculoskeletal model of the knee for evaluating ligament forces during isometric contractions. *J Biomech.* 30,163-76
20. Smith, P.N., Refshauge, K.M., Scarvell J.M. 2003. Development of concepts of knee kinemtaics. *Arch. Phys. Med. Rehabil.* 84(12),1895-1902
21. The genesis of the LifeMOD® Biomechanics Modeler Virtual Biomechanics Brochure
22. Wilson, D. R., Feikes, J. D., Zavatsky, A. B., O'Connor, J. J. 2000. The components of passive movement are coupled to flexion angle. *J. Biomech.* 33, 465-73
23. Wismans, J., Veldpaus and Jansen, J. 1980.[0] A three dimensional mathematical model of the knee joint. *J Biomech.* 13, 677-85
24. www.adams.com
25. Zuppinger, H. 1904. Die aktive Flexion in unbelasteten Kniegelenk Zürich Habil. Wiesbaden, Bergmann, pp.703-763

Author Index

Adams, Judith	1	Nash, Anthony	25
Akalan, Ekin	104	Nielsen, Mads	41
Ashry, Reem	96	Olsen, Ole Fogh	9, 17, 51
Bartels, Wilbert	59	Ourselin, Sebastien	50
Bijlsma, Johannes	59	Ozkan, Mehmet	104
Bischof, Horst	33	Peloschek, Philipp	33
Bourgeat, Pierrick	50	Pettersen, Paola	9, 17, 41, 42, 51, 88
Boyle, Roger	64, 72	Qazi, Arish Asif	51
Christiansen, Claus	9, 17, 42, 51, 88	Radjenovic, Alexandra	64, 72
Cootes, Tim	1	Ramadan, Saadallah	50
Dam, Erik	9, 17, 42, 51, 88	Roberts, Martin	1
Folkesson, Jenny	9, 17, 42, 51, 88	Schweitzer, Mark	80, 96
Fripp, Jurgen	50	Stanwell, Peter	50
Hamarneh, Ghassan	80, 96	Tanko, Laszlo B.	41
Holmes, Andrew	25	Taylor, Chris	25
Kainberger, Franz	33	Temelli, Yener	104
Kubassova, Olga	64, 72	Vincken, Koen	59
Lafeber, Floris	59	Vos, Petra	59
Langs, Georg	33	Ward, Aaron	80, 96
Loog, Marco	88	Waterton, John	25
Lund, Michael T.	41	Williams, Tomos	25
Maciewicz, Rose	25	de Bruijne, Marleen	41
Marijnissen, Anne	59		



ISBN 10: 87-7611-151-2
ISBN 13: 978-87-7611-151-9