35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Femoral Cartilage Segmentation in Knee MRI Scans Using Two Stage Voxel Classification

Adhish Prasoon*, Christian Igel, Marco Loog, François Lauze, Erik B. Dam, Mads Nielsen

***Abstract*— Using more than one classification stage and exploiting class population imbalance allows for incorporating powerful classifiers in tasks requiring large scale training data, even if these classifiers scale badly with the number of training samples. This led us to propose a two-stage classifier for segmenting tibial cartilage in knee MRI scans combining nearest neighbor classification and support vector machines (SVMs). Here we apply it to femoral cartilage segmentation. We describe the similarities and differences between segmenting these two knee cartilages. For further speeding up batch SVM training, we propose loosening the stopping condition in the quadratic program solver before considering moving on to other approximation techniques such as online SVMs. The two-stage approach reached a higher accuracy in comparison to the one-stage state-of-the-art method. It also achieved better inter-scan segmentation reproducibility when compared to a radiologist as well as the current state-of-the-art method.**

***Index Terms*— femoral cartilage, support vector machine, nearest neighbor classifier, online support vector machine, osteoarthritis, magnetic resonance imaging**

I. Introduction

The need to cope with large scale data in medical imaging often limits the use of complex classifiers having excellent generalization ability. An example of such a classifier is a non-linear support vector machine (SVM, [1]), where the training time scales worse than quadratically with the number of training data points. In our previous work [2], we presented a two-stage cascaded classifier approach to overcome this restriction. The proposed classifier was applied to segment tibial cartilage in low-field knee MRI scans and outperformed the state-of-the-art method. As a step towards completing the study we apply the similar approach for segmenting femoral cartilage, also from low-field knee MRI scans. The segmentation of articular cartilage is useful for the quantitative analysis of the deterioration of articular cartilage, which causes osteoarthritis. Osteoarthritis is one of main causes of work disability through out the world specially for the elderly population. Non-invasive assessment of articular cartilage are most commonly done using MRI scans [3].

In a general two-stage classifier for segmentation, we have a classifier trainable on huge data-sets in the first stage. However, the goal of this first stage is not necessarily to achieve best segmentation results, but to maximize sensitivity, that is, to minimize false negatives. The points classified as background by the first stage are labeled accordingly, while all the points classified as foreground go through a second stage of classification. The classifier used at this stage can be a more powerful classifier, which may scale badly with number of training data points. However, if the background population is large compared to the foreground population and a large portion of background population is screened in the first stage, a significantly smaller portion of data points is fed into the second-stage classifier. This makes it possible to use classifiers scaling badly with number of training data points.

In this study, we extend our earlier work and apply the two-stage approach to segmenting femoral cartilage. The approach is compared to the state-of-the-art method that is based on one stage of nearest neighbor classification. We discuss the similarities and differences in segmenting femoral and tibial cartilages as well as the challenges faced due to the even higher amount of training data compared to [2]. Furthermore, we consider images of subjects scanned twice within one week and investigate the inter-scan reproducibility of the proposed classifier in comparison to a radiologist and the current state-of-the-art method.

II. Related Work

Computer-aided segmentation of the articular cartilage from the knee MRI scans is an active research field. Methods either rely on 2D approaches to segment slice by slice or directly use 3D segmentation. Stammberger et. al. [4] used b-splines to segment each slice of MRI scan. Another slice-by-slice cartilage segmentation method based on active shape models was proposed by Solloway et. al. [5]. Folkesson et. al. [6] developed a 3D voxel classification approach which can be considered as state-of-the-art method for fully automatic segmentation. A semi-automated method was developed by Bae et. al. [7]. Their segmentation method is based on graph-cut algorithm. They performed volumetric measurements of the cartilage from high-resolution knee magnetic resonance (MR) images from the Osteoarthritis Initiative (OAI) database and assessed the intra and inter-observer reproducibility of measurements obtained via their method. A semi-automatic method based on radial transformation was proposed by Chang et. al. [8]. Yin et. al. proposed a method for simultaneous segmentation of bone and cartilage surfaces [9]. A fully automatic method was proposed by Seim et. al. who segmented bones and cartilage from MRI scans using statistical shape model and graph based optimization [10].

* Corresponding Author
Adhish Prasoon, Christian Igel, François Lauze and Mads Nielsen are with Department of Computer Science, University of Copenhagen, Denmark (e-mail: adhish@diku.dk; igel@diku.dk; francois@diku.dk; madsn@diku.dk)
Mads Nielsen is also with Biomediq, Copenhagen, Denmark.
Marco Loog is with Pattern Recognition Laboratory, Delft University of Technology, The Netherlands (e-mail: m.loog@tudelft.nl)
Erik Dam is with Biomediq, Denmark (e-mail: erikdam@biomediq.com)

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

Vincent et. al. presented a fully automatic system [11] based on active appearance models. A three stage scheme was proposed by Fripp et. al. [12]. In the first stage automatic segmentation of the bones is performed. In the second stage, the bone cartilage interfaces is extracted. In the final stage segmentation of the cartilages is performed. Dodin et. al. proposed automatic segmentation method for knee MRI scans acquired using 3T scanner and a knee coil [13]. They segmented bone cartilage interface for tibia and femur independently. A level sets based algorithm was proposed in [14] for 3D segmentation, and [15] incorporated multiple spatial inter-relationship on n-dimensional graphs followed by graph optimization that yields a globally optimal solution to segment cartilage.

III. Approach

This section presents the two-stage classifier and describes its application to the segmentation of femoral cartilage. We also consider the challenges associated with extending the study from tibial cartilage segmentation to femoral cartilage segmentation and comment on speeding up SVM training by approximating the SVM solution.

A. Two-stage classifier

Let us assume w.l.o.g. a binary segmentation problem where the population of the positive class is less than the negative class population by at least an order of magnitude. Let X be the input space and $Y = \{-1, 1\}$ the output. A hypothesis $h_1 : X \mapsto Y$ is learned in the first stage using all the training data. Let ℓ be the number of samples and $D_{\text{train}_1} \subset (X \times Y)^\ell$ be the training data. The hypothesis h_1 is tuned to achieve maximum sensitivity and, thus, having minimum false-negatives. This stage should use a learning algorithm which can handle a very large number of training data points. The data points classified as background by first stage classifier are labeled as background and rest of the points $D_{\text{train}_2} = \{(x, y) \,|\, (x, y) \in D_{\text{train}_1} \wedge h_1(x) = 1\}$ are used to train our second stage hypothesis h_2 . The aim of the learning algorithm at this stage is to achieve good segmentation performance. As the number of data points at this stage is just a small fraction of ℓ, we can employ a powerful classifier at this stage, even if it scales badly with training data population. This way, the final two-stage classifier has good generalization ability and can also handle huge training data sets. The two-stage classifier can be summarized as

$$h(x) = \begin{cases} -1 & \text{if } h_1(x) = -1 \\ -1 & \text{if } h_1(x) = 1 \text{ and } h_2(x) = -1 \\ 1 & \text{otherwise} \end{cases} .$$

Figure 1 depicts the general two stage classifier.

B. Automatic segmentation of femoral cartilage

This section presents the application of a two-stage classifier to segmenting femoral cartilage. Whenever needed, we will also refer to tibial cartilage segmentation for comparison.

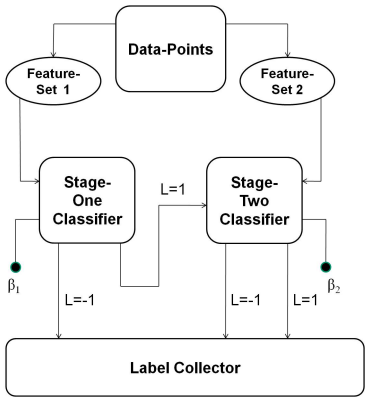


Fig. 1: General concept of our two-stage classifier, where β_1 denotes parameters used to tune the first stage classifier for maximum sensitivity, while the parameters β_2 are used to tune the second stage for best segmentation performance. The labels $L = 1$ and $L = -1$ refer to cartilage and background voxels, respectively.

1) Training data and features: Training data was extracted using 25 scans, which are exactly the same scans as used by state-of-the-art method [6]. Firstly, a region of interest from each MRI scan was extracted. The volume of region of interest is 30% of the volume of the MRI scan. Each MRI scan has around 6.85 million voxels with a region of interest of approximately 2 million voxels. As the background points are too high in number, we sample the background from the ROI and take all the cartilage voxels in our training data. The sampling is performed very densely in the region close to the cartilage, rarely in the region far from the cartilage and the sampling probability varies linearly between two values. For femoral cartilage, the number of all the cartilage points (from 25 scans) is 295,403 while the number of background points is 2,408,864. In our earlier study on tibial cartilage, the number of tibial cartilage and background points were 119,684 and 1,892,696 respectively. As we can see, femoral cartilage is considerably bigger than the tibial, resulting in higher number of cartilage as well as background data points. We use the same set of 178 features which were used by Folkesson et. al. [6] as candidate features. Folkesson et. al. used features selection to find a smaller set of features in order to improve the performance.

2) The two stages for femoral cartilage segmentation: Stage one of our classifier is similar to the state-of-the-art one-stage kNN of Folkesson et. al. [6]. However, they have different aims. The one-stage kNN is trained to select the value of k, a smaller set of features using a feature selection method, and a posterior threshold t that deals with the large class imbalance to achieve best possible segmentation results. Let p_b be the posterior probablitity of a voxel being in background class, then the voxel is classified as background if $p_b > t$ and as cartilage otherwise. The features selected in case of tibial and femoral cartilage are slightly different. The number of selected features for tibial cartilage was 36 while that for femoral cartilage it was 42. The main difference

between kNN used by [6] and our stage-one is the purpose of the classifier. We adjust t in order to achieve maximum sensitivity or minimum false positives. The value of k used by Folkesson et. al. was 100 for both the cartilage and we also use the same value of k.

Stage two of our classifier is an SVM with Gaussian kernel. We employed LIBSVM [16] for training the SVM. The training data comprised the points labeled as cartilage in stage-one. Although stage one used only a selected subset of features, in the second stage the SVM used all the 178 features. In case of the tibial cartilage, we performed nested grid search using cross-validation (splitting the available training data) as performance criterion. We searched for a good combination of kernel width parameter and regularization (commonly denoted by γ and C) on a 13×29 grid. After finding the best of these 377 combinations, we placed a second narrow 3×3 grid around the optimum value of C and γ. At this point we introduced a weight ratio parameter $W \in \{1.0, 1.1, 1.2, \ldots, 1.8\}$, which made it possible to select different regularization parameters. The final regularization parameters of cartilage and background class were $C \cdot W$ and C respectively.

Performing grid search in a big space of 377 combinations was time consuming even with a lot of computing resource. However, when performing the grid search for the second-stage of femoral cartilage segmentation, we placed just a 3×3 grid around the same pair (C, γ) which was found to be optimum during grid-search for the second-stage of tibial cartilage segmentation.The good results that we achieved in case of femoral cartilage segmentation, using the hyper-parameters similar to what we learnt for tibial cartilage segmentation, show the robustness of our two-stage classifier.

C. Speeding-up SVM training: Online learning vs. batch leaning with low accuracy

There were more than 700,000 training data points more in case of femoral cartilage than in case of tibial. On top of that, the first-stage of kNN performed slightly worse in the case of femoral cartilage segmentation when compared to tibial cartilage segmentation, thus the specificity achieved for maximum sensitivity was lower. Thus, the percentage of points screened in the first stage of femoral segmentation was lower than in the tibial case. In fact, the number of training data points for second stage SVM in case of femoral was 688,128, while in case of tibial the second stage SVM had to handle only 262,142 data points. Thus, model selection and final training gets very time consuming.

We consider non-linear SVMs. Training the machines amounts to solving a quadratic program (QP) having time complexity $\Omega(\ell^2)$ [17]. We use iterative sequential minimal optimization, and to speed up SVM training for the femoral cartilage classification, we loosened the stopping criterion from $\varepsilon = 0.001$ to $\varepsilon = 0.5$. We use the common stopping criterion as discussed, e.g., in [18]. Let $\alpha_1, \ldots, \alpha_\ell$ denote the coefficients of the SVM dual objective function f and let g_i denote the partial derivative of f with respect to α_i.

TABLE I: Comparison of classifiers applied for femoral cartilage segmentation. DSC stands for the dice similarity coefficient. The proposed cascaded classifier is referred to as two-stage 2-stage, the state-of-the-art reference algorithm is 1-stage. All values are mean over 114 scans.

Classifier	DSC	Accuracy	Sensitivity	Specificity
2-stage	0.8115	96.3234%	80.8236%	98.0760%
1-stage	0.7984	96.0821%	79.7736%	97.8938%

Then we stop when

$$\max\left(\max_{\alpha_i<C, y_i=1} -g_i, \max_{\alpha_i>0, y_i=-1} g_i\right) - \min\left(\min_{\alpha_i<C, y_i=-1} g_i, \min_{\alpha_i>0, y_i=1} -g_i\right)$$

falls below the threshold ε.

An alternative to this approach is using an online SVM such as LASVM [19]. However, tuning ε is simpler, and we found it to produce more accurate solutions than LASVM within the same time budget in our application.

We also conducted experiments with LASVM to solve the problem in just one stage using all the 178 features and all the training data points. However, we observed no improvement in performance and too long training times.

IV. Evaluation and Results

We evaluate our classifier on a hold-out set of 114 test scans. We used Dice Similarity Coefficient to evaluate the segmentation performance,

$$\text{DSC}(A, B) = \frac{2(|A \cap B|)}{|A| + |B|}$$

where A and B are manual and automatic segmentations. In table I we compare results obtained by our two-stage method with the state-of-the-art one stage kNN. Our method performed statistically significantly better than the one-stage kNN in terms of DSC and accuracy (Wilcoxon rank-sum test, $p < 0.05$), with both better sensitivity and specificity.

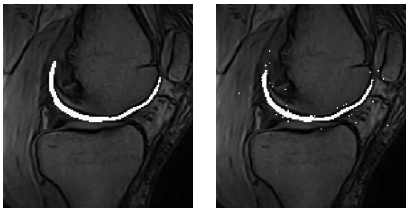
We also evaluated interscan segmentation reproducibility on 31 pairs of scans, each pair obtained within a week. The same radiologist segmented both the scans of each pair. The evaluation was based on the *RMS-CV* score calculated as follows. Let V_s^i and V_r^i be cartilage volumes obtained from scan and re-scan of the ith pair. We calculate the coefficient of variation for ith pair

$$C_v^i = \frac{\sqrt{2}(|V_s^i - V_r^i|)}{V_s^i + V_r^i}$$

which is the same as the ratio of the standard deviation to the mean of two volumes. The *RMS-CV* score is given as $\sqrt{\sum_{i=1}^{31} {C_v^i}^2/31}$. The lower the *RMS-CV* score, the better the reproducibility. Table II summarizes the results obtained, showing that our method performed better than the radiologist as well as the state-of-the-art method.

TABLE II: Comparing interscan femoral cartilage segmentation reproducibility on 31 pairs of scans

Method	RMS-CV
2-stage	0.0785
1-stage	0.0810
Manual	0.1140



(a) manual (b) two-stage

Fig. 2: Slice taken from a 3D MRI scan segmented by (a) a radiologist and (b) our two-stage method. The slice was chosen to demonstrate that, although the proposed method outperforms the state-of-the algorithm [6], there is still room for improvement by (simple) post-processing.

Figure 2 shows a slice segmented by the radiologist and our method. A slice is taken from the 3D segmentation for visualization purpose (actually, radiologists segment the scans in a slice by slice manner). The resulting segmentations suggest that some post-processing can further increase the segmentation results.

V. Discussion

The proposed two-stage classification method is a general tool for scenarios in which the number of training data points is huge and the classes are unbalanced. These scenarios are often found in medical imaging applications and thus such a classifier is particular useful in this field. Its application for segmenting articular cartilage from low field knee MRI scans was very successful.

The two-stage method outperformed the state-of-the-art one-stage kNN and also achieved better interscan segmentation reproducibility when compared to one-stage kNN and the manual segmentations done by a radiologist. However, the testing time of our method was 30-35% more compared to the one-stage kNN for femoral cartilage. For increasing the speed of the SVM training, we found no advantage in using online SVMs over simply reducing the accuracy of batch SVM training (by loosening the stopping condition in the quadratic program solver). Replacing the two-stage classifier by a single online SVM did not lead to better performance given our time budget.

Acknowledgment

Christian Igel gratefully acknowledges support from the European Commission through project AKMI (PCIG10-GA-2011-303655).

References

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[2] A. Prasoon, C. Igel, M. Loog, F. Lauze, E. Dam, and M. Nielsen, "Cascaded classifier for large-scale data applied to automatic segmentation of articular cartilage," in *Proceeding of SPIE Medical Imaging: Image Processing*, 2012, vol. 8314.

[3] H. Graichen, R. Eisenhart-Rothe, T. Vogl, K.-H. Englmeier, and F. Eckstein, "Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging," *Arthritis Rheumatism*, vol. 50, no. 3, pp. 811–816, Mar. 2001.

[4] T. Stammberger, F. Eckstein, M. Michaelis, K. H. Englmeier, and M. Reiser, "Interobserver reproducibility of quantitative cartilage measurements: Comparison of b-spline snakes and manual segmentation," *Magnetic Resonance Imaging*, vol. 17, no. 7, pp. 1033–1042, 1999.

[5] S. Solloway, C. Hutchinson, J. Vaterton, and C. Taylor, "The use of active shape models for making thickness measurements of articular cartilage from MR images," *Magnetic Resonance in Medicine*, vol. 37, pp. 943–952, 1997.

[6] J. Folkesson, E.B. Dam, O.F. Olsen, P.C. Pettersen, and C. Christiansen, "Segmenting articular cartilage automatically using a voxel classification approach," *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 106–115, 2007.

[7] KT Bae, H. Shim, C. Tao, S. Chang, JH Wang, R. Boudreau, and CK Kwoh, "Intra and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method," *Osteoarthritis and Cartilage*, vol. 17, no. 12, pp. 1589 – 1597, 2009.

[8] Ku-Yaw Chang, Shao-Jer Chen, Lih-Shyang Chen, and Cheng-Jung Wu, "Articular cartilage segmentation based on radial transformation," in *9th International Conference on Hybrid Intelligent Systems (HIS 2009), August 12-14, 2009, Shenyang, China*. 2009, pp. 239–242, IEEE Computer Society.

[9] Y. Yin, X. Zhang, D. D. Anderson, T. D. Brown, C. Van Hofwegen, and M. Sonka, "Simultaneous segmentation of the bone and cartilage surfaces of a knee joint in 3D," in *Proceeding of SPIE Medical Imaging 2009: Physics of Medical Imaging*, 2009, vol. 7258, pp. 72591O–72591O–9.

[10] H. Seim, D. Kainmueller, H. Lamecker, M. Bindernagel, J. Malinowski, and S. Zachow, "Model-based auto-segmentation of knee bones and cartilage in MRI data," in *Proceeding of Medical Image Analysis for the Clinic: A Grand Challenge.*, 2010, pp. 215 – 223.

[11] G. Vincent, C. Wolstenholme, I. Scott, and M. Bowes, "Fully automatic segmentation of the knee joint using active appearance models," in *Proceeding of Medical Image Analysis for the Clinic: A Grand Challenge*, 2010, pp. 224 –230.

[12] J. Fripp, S. Crozier, S.K. Warfield, and S. Ourselin, "Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 55–64, January 2010.

[13] P. Dodin, J. Pelletier, J. Martel-Pelletier, and F. Abram, "Automatic human knee cartilage segmentation from 3-D magnetic resonance images," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2699 – 2711, 2010.

[14] E. Hinrichs, B. Appleton, B.C. Lovell, and G.J. Galloway, "Autonomous direct 3D segmentation of articular knee cartilage," in *Australian and New Zealand Intelligent Information Systems*. Queensland University of Technology, 2011, vol. 1, pp. 417 – 420.

[15] Y. Yin, Z. Xiangmin, R. Williams, Wu. Xiaodong, D.D. Anderson, and M. Sonka, "LOGISMOS - layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint," *IEEE Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2023 – 2037, 2010.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[17] L. Bottou and C.-J. Lin, "Support vector machine solvers," in *Large Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, Eds., pp. 1–28. MIT Press, 2007.

[18] C.-J. Lin, "A formal analysis of stopping criteria of decomposition methods for support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1045–1052, 2002.

[19] A. Bordes, S. Ertekin, J. Weston, and L. Bottou, "Fast kernel classifiers with online and active learning," *Journal of Machine Learning Research*, vol. 6, pp. 1579–1619, 2005.