Automatic Atlas-based Three-label Cartilage Segmentation from MR Knee Images

Liang Shan[a], Christopher Zach[b], Cecil Charles[c], Marc Niethammer[a,d]

[a]*Department of Computer Science, University of North Carolina at Chapel Hill, USA*
[b]*Microsoft Research Cambridge, UK*
[c]*Department of Radiology, Duke University, USA*
[d]*Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, USA*

Abstract

Osteoarthritis (OA) is the most common form of joint disease and often characterized by cartilage changes. Accurate quantitative methods are needed to rapidly screen large image databases to assess changes in cartilage morphology. We therefore propose a new automatic atlas-based cartilage segmentation method for future automatic OA studies.

Atlas-based segmentation methods have been demonstrated to be robust and accurate in brain imaging and therefore also hold high promise to allow for reliable and high-quality segmentations of cartilage. Nevertheless, atlas-based methods have not been well explored for cartilage segmentation. A particular challenge is the thinness of cartilage, its relatively small volume in comparison to surrounding tissue and the difficulty to locate cartilage interfaces – for example the interface between femoral and tibial cartilage.

This paper focuses on the segmentation of femoral and tibial cartilage, proposing a multi-atlas segmentation strategy with non-local patch-based label fusion which can robustly identify candidate regions of cartilage. This method is combined with a novel three-label segmentation method which guarantees the spatial separation of femoral and tibial cartilage, and ensures spatial regularity while preserving the thin cartilage shape through anisotropic regularization. Our segmentation energy is convex and therefore guarantees globally optimal solutions.

We perform an extensive validation of the proposed method on 706 images of the Pfizer Longitudinal Study. Our validation includes comparisons of different atlas segmentation strategies, different local classifiers, and different types of regularizers. To compare to other cartilage segmentation approaches we validate based on the 50 images of the SKI10 dataset.

Keywords: Cartilage, atlas, segmentation, three-label, automatic

1. Introduction

Osteoarthritis (OA) is the most common form of joint disease and a major cause of long-term disability in the United States of America [29]. Cartilage loss is believed to be the dominating factor in OA. As magnetic resonance imaging (MRI) is able to evaluate cartilage volume and thickness and allows reproducible quantification of cartilage morphology [10, 9] it is increasingly accepted as a primary method to evaluate progression of OA. An accurate cartilage segmentation from magnetic resonance (MR) knee images is crucial to study OA and would be of particular use for future clinical trials to test so far non-existing disease-modifying OA drugs. Already today, large image databases exist for OA studies which are well suited to design and test automatic cartilage segmentation algorithms capable of processing thousands of images. For example, the Pfizer Longitudinal Study (PLS) dataset contains 158 subjects, each with five time points. The Osteoarthritis Initiative (OAI) dataset includes 4,796 subjects with multiple time points. Due to the large size of image databases, a fully automatic segmentation and analysis method is essential. In this paper, we therefore propose a new cartilage segmentation method from knee MR images, which requires no user interaction (besides quality control). The method is a step towards automatic analysis of large OA image databases.

Recently, several automatic methods have been proposed for cartilage segmentation. Folkesson *et al.* [11] proposed a voxel-based hierarchical classification scheme for cartilage segmentation. Fripp *et al.* [12] used active shape models for bone segmentation in order to extract the bone-cartilage interface followed by tissue classification. A graph-based simultaneous segmentation of bone and cartilage was developed by Yin *et al* [32]. Vincent *et al.* [27] applied multi-start and hierarchical active appearance modeling to segment cartilage. Texture analysis [7] has also been employed in cartilage segmentation. Seim *et al.* [20] utilized prior knowledge on the variation of cartilage thickness. Voxel-based classification approaches have been investigated for segmenting multi-contrast MR data in [17, 34].

To allow for localized analysis and the suppression of un-

likely voxels in a segmentation, introducing a spatial prior is desirable. This can be achieved through an atlas-based analysis method. In particular, multi-atlas segmentation strategies [18] have shown to be robust and reliable image segmentation methods. While such methods have been successfully used in brain imaging, they have so far rarely been used for cartilage segmentation. The work by Glocker et al. [13], which used a statistical shape atlas from a set of pre-aligned knee images, and the work by Tamez-Peña et al. [26] using a multi-atlas-based method, are two exceptions. Our work in this paper is most closely related to [26] as both methods make use of multi-atlas segmentation strategies. However, we significantly extend the prior work by [26]. In particular:

1) We propose a convex three-label segmentation method which allows for anisotropic spatial regularization. This is a generally applicable segmentation method. Applied to the segmentation of femoral and tibial cartilage, it guarantees their spatial separation while ensuring spatially smooth solutions accounting for the cartilage thinness through anisotropic regularization. We incorporate spatial priors via atlas information (see 2) and local segmentation label likelihoods through appearance classification comparing both k nearest neighbors (kNN) classification and classification by a support vector machine (SVM).

2) We compare different atlas-based segmentation methods: using a single average-shape atlas as well as multiple atlases with various label fusion strategies as segmentation priors.

3) We perform an extensive validation on over 700 images with varying levels of OA disease progression using data from both the Pfizer Longitudinal Study (PLS) and from SKI10 [14] to compare to existing methods.

These contributions are significant as:

1) Due to its convexity our segmentation method allows the efficient computation of globally optimal solutions for *three* segmentation labels. Furthermore, we demonstrate that anisotropic regularization within this segmentation model is less sensitive to parameter settings than isotropic regularization and yields more accurate segmentations.

2) We show that using non-local patch-based label fusion from multiple atlases to obtain segmentation priors improves segmentation results significantly over using a single atlas or a local label fusion strategy.

3) Our validation dataset (with more than 700 images) is at least one order of magnitude larger than most prior cartilage segmentation validation studies, hence demonstrating the ability of our proposed segmentation method to automatically achieve accurate cartilage segmentations for large imaging studies. The
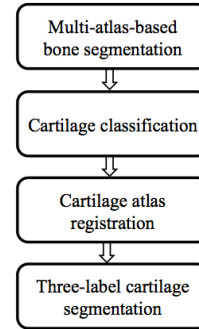

Figure 1: Cartilage segmentation pipeline.

required robustness of the segmentation method is achieved by using a multi-atlas segmentation strategy. The obtained accuracy can be attributed to the combination of local classification, multi-atlas label fusion, three-label segmentation and anisotropic regularization.

Figure 1 illustrates the proposed cartilage segmentation method. The method starts with multi-atlas-based bone segmentation to guide the cartilage atlas registration. The cartilage spatial prior is then obtained from either multi-atlas or average-shape-atlas registration. A probabilistic classification is performed to compute local likelihoods. The three-label segmentation makes the final decision from the spatial priors and the local likelihoods jointly, allowing for anisotropic spatial regularization. The method described in this paper is an extension of the preliminary ideas we presented in recent conference papers [24, 22, 23]. This paper offers more details, additional experiments, and a much larger validation study.

The remainder of this paper is organized as follows: Section 2 clarifies the atlas terminology and briefly reviews the atlas-based segmentation approaches. Section 3 presents the three-label segmentation framework with isotropic and anisotropic regularization. Section 4 describes the multi-atlas-based bone segmentation method. The probabilistic cartilage classification is explained in section 5. Sections 6 and 7 discuss the average-shape-atlas-based and multi-atlas-based cartilage segmentation, respectively. Experimental results on the PLS dataset are shown in section 9. We compare the proposed method to other methods in section 10 by making use of the SKI10 dataset. The paper closes with conclusions and future work.

2. Atlas Terminology

An atlas [1], in the context of atlas-based segmentation, is defined as the pairing of an original structural image and the corresponding segmentation. Atlas-based segmentation methods can be categorized into three groups [15], namely single-atlas-based, average-shape atlas-based and multi-atlas-based methods. The work by Glocker et al. [13]

falls into the second group. The work by Tamez-Peña *et al.* [26] belongs to the multi-atlas category.

In the single-atlas-based method, a single labeled image is chosen as the atlas and registered to the query image. The atlas label is propagated following the same transform to generate the segmentation for the query image. The drawbacks of the single-atlas-based segmentation include the possibility that the atlas used is anatomically unrepresentative of the query image and occasional registration failures because the method critically depends on the success of only one registration. To alleviate the problem of being non-representative, average-shape-atlas-based methods have been proposed, where a reference image is selected to build the atlas from a set of labeled images. However, here success still depends on the success of a single registration. Furthermore, the choice of reference image is important for segmentation accuracy and frequently addressed by building an average atlas-image through registration – which in itself is not a trivial task. Alternatively, in multi-atlas-based segmentation, multiple labeled images are registered to the query image independently, hereby avoiding reliance on one registration while allowing to represent anatomical variations. The downside of multi-atlas-based segmentation is high computation cost as multiple registrations are required. In spite of the expensive computation, multi-atlas-based segmentation has been popular and successful in brain imaging. In particular, Rohlfing *et al.* [18] demonstrated that multi-atlas-based segmentation is more accurate than the two other atlas-based segmentation methods. We will therefore follow a multi-atlas strategy in what follows.

3. Three-label segmentation method

A binary segmentation consists of only two labels, i.e., foreground and background, and tends to merge touching objects if spatial regularity is enforced. A multi-label segmentation can keep objects separated and is therefore particularly suited to segment touching objects. Since the femoral and tibial cartilage (as well as bones for severe OA patients) may touch each other in the MR images, we make use of a three-label segmentation method to avoid possible mergings. Figure 2 demonstrates the limitations of a binary versus a three-label segmentation method for a synthetic bone case.

The three-label case is a specialization of our previous multi-label segmentation method [33] which allows for a symmetric formulation with respect to the background segmentation class. A multi-label segmentation is a mapping from an image domain Ω to a label space represented by a set of non-negative integers, i.e. $\mathcal{L} = \{0, ..., L-1\}$. The labeling function $\Lambda : \Omega \to \mathcal{L}, \mathbf{x} \mapsto \Lambda(\mathbf{x})$ maps a pixel $\mathbf{x}$ in the image domain Ω to label $\Lambda(\mathbf{x})$ in label space. The goal is to find a labeling function that minimizes an energy functional of the form:

$$E(\Lambda) = \int c(\mathbf{x}, \Lambda(\mathbf{x})) + V(\nabla\Lambda, \nabla^2\Lambda, ...) \ d\mathbf{x} \qquad (1)$$
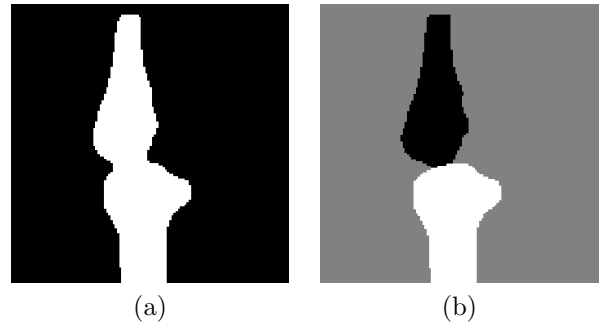

(a)   (b)

Figure 2: Synthetic example. (a) Binary segmentation result. Femur and tibia are segmented as one object and the boundary in the joint region is not captured well due to regularization effects. (b) Proposed three-label segmentation. The boundaries between bones and background are preserved.

where $c(\mathbf{x}, \Lambda(\mathbf{x}))$ is the cost of assigning label $\Lambda(\mathbf{x})$ to pixel $\mathbf{x}$ and $V(\cdot)$ is a regularizing term. The different labelings can be encoded through a level function u

$$u(\mathbf{x}, l) = \begin{cases} 1 & \text{if } \Lambda(\mathbf{x}) < l, \\ 0 & \text{otherwise}, \end{cases} \qquad (2)$$

which maps the Cartesian product of the image domain Ω and the labeling space $\mathcal{L}$ to $\{0, 1\}$. By definition, we have $u(\mathbf{x}, 0) = 0$ and $u(\mathbf{x}, L) = 1$. Of note, u does not directly encode labels, but instead defines them through its discontinuity set. Figure 3 illustrates the relation between u and Λ for the three-label case. This setup is in general asymmetric with respect to the labels, since the design of the level function implies a specific label ordering. However, for the three-label case, the background label can be symmetrically positioned between the two object labels (i.e., femoral cartilage and tibial cartilage) hence resulting in a method which treats the two cartilage classes symmetrically.

Minimizing the segmentation energy functional

$$E(u) = \int_{\mathcal{D}} g \|\nabla_{\mathbf{x}} u\| + c |\nabla_l u| \ d\mathbf{x} dl, \ \mathcal{D} = \Omega \times \mathcal{L}$$
$$u \in [0, 1], \ u(\mathbf{x}, 0) = 0, \ u(\mathbf{x}, L) = 1, \qquad (3)$$

with respect to u, results in an essentially binary and monotonically increasing level function u indicating the multi-label image segmentation. Here, $\nabla_{\mathbf{x}} u$ is the spatial gradient of u, $\nabla_{\mathbf{x}} u = (\partial u/\partial x, \partial u/\partial y, \partial u/\partial z)^T$ and $\nabla_l u$ is the gradient in label direction, $\nabla_l u = \partial u/\partial l$; g controls the isotropic regularization and c defines the labeling cost. In our implementation, we set g to a non-negative constant. The formulation is convex and therefore a global optimum can be computed. We apply an iterative gradient descent/accent scheme for the optimization. See Appendix A for more details on the numerical solution to (3). The solution u is essentially binary[1] and monoton-

---

[1] Since u is in [0,1], the returned solution can be fractional. An equivalent binary optimal solution can be obtained by thresholding u.
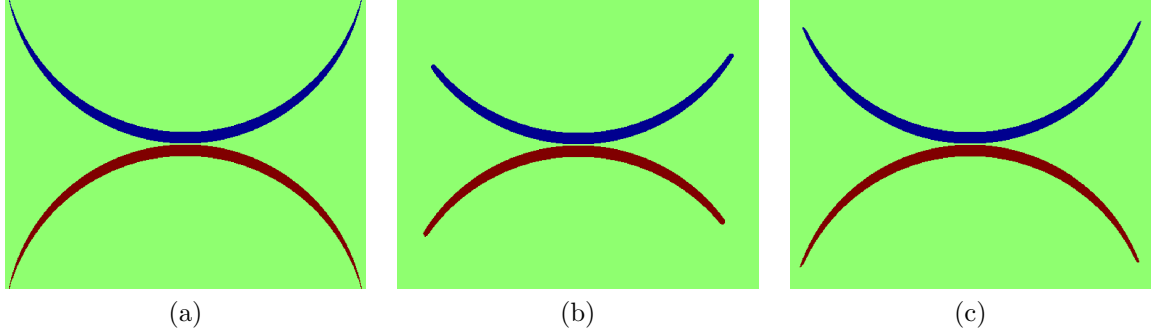
Figure 4: Synthetic example: (a) original image to be segmented; (b) and (c) three-label segmentation results with isotropic and anisotropic regularization respectively. Anisotropic regularization avoids over-regularization at the tips of the synthetic shape.
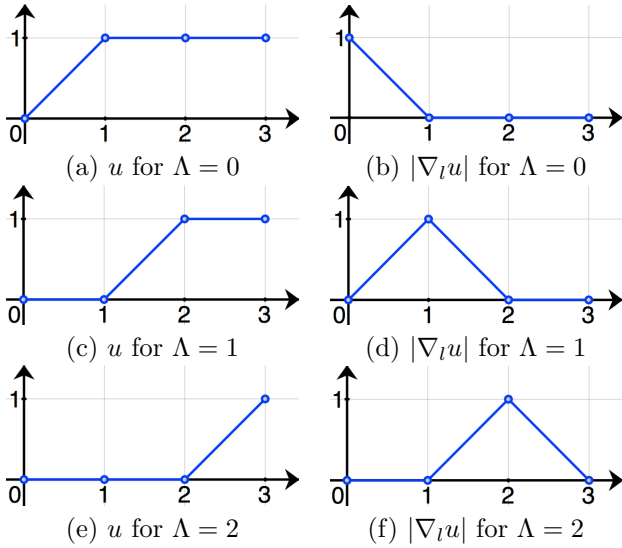

(a) u for $\Lambda = 0$ (b) $|\nabla_l u|$ for $\Lambda = 0$

(c) u for $\Lambda = 1$ (d) $|\nabla_l u|$ for $\Lambda = 1$

(e) u for $\Lambda = 2$ (f) $|\nabla_l u|$ for $\Lambda = 2$

Figure 3: Values of u and $|\nabla_l u|$ for different label assignments in a three-label segmentation (abscissa l). Assuming a discretization with forward differences. $|\nabla_l u|$ determines the label assignment.
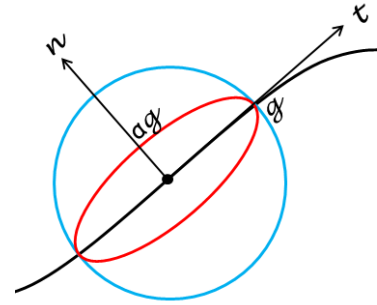

Figure 5: Difference between isotropic and anisotropic regularization. The black curve is an edge in an image. The regularization is illustrated at a pixel (the dot). The blue circle indicates the isotropic case where regularization is enforced equally in every direction. The red ellipse shows the anisotropic situation where less regularization is applied in the normal direction and more in the tangent direction.

ically increasing. The three-label segmentation can then be computed from the discontinuity set of u.

The isotropic regularization in model (3) treats all directions equally, which is not an ideal choice for long and thin objects like the cartilage. To customize the segmentation model (3) for cartilage segmentation, we replace the isotropic regularization term, g, by an anisotropic one

$$E(u) = \int_{\mathcal{D}} \|\mathbf{G}\nabla_{\mathbf{x}}u\| + c|\nabla_l u| \; d\mathbf{x}dl, \; \mathcal{D} = \Omega \times \mathcal{L}, \quad (4)$$
$$u \in [0, 1], \; u(\mathbf{x}, 0) = 0, \; u(\mathbf{x}, L) = 1,$$

where $\mathbf{G}$ is a positive-definite matrix determining the amount of regularization. This avoids over-regularization at the boundaries of the cartilage layers and therefore allows for a more faithful segmentation. Figure 4 illustrates the problem with isotropic regularization which tends to shrink the segmentation boundary by cutting thin objects short and the benefit from anisotropic regularization. We

choose $\mathbf{G}$ as

$$\mathbf{G} = g\left[\mathbf{I} + (\alpha - 1)\mathbf{n}\mathbf{n}^T\right], \alpha \in [0, 1], \quad (5)$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{n}$ is a unit vector indicating the direction of less regularization (the normal direction to the cartilage surface). See Fig. 5 for an illustration of isotropic versus anisotropic regularization. Since the normal direction to the cartilage surface is not known a-priori, we approximate it by the normal direction to the bone-cartilage interface which can be determined from the segmentations of femur and tibia (see section 8). The energy functional (4) is also convex and therefore a global optimum can also be computed. Again, we apply an iterative gradient descent/accent scheme for the optimization. See Appendix A for more details on the numerical solution to (4). The solution u is also essentially binary and monotonically increasing.

Section 4 describes how to use the model (3) for bone segmentation using a multi-atlas-based approach to obtain spatial priors. Details on using model (4) for cartilage segmentation are discussed in sections 5, 6 and 7.

4. Multi-atlas-based Bone Segmentation

The labeling cost c in (3) for each label l in $\{\text{FB}, \text{BG}, \text{TB}\}$ ("FB", "BG" and "TB" denote the femoral bone, the background and the tibial bone respectively) are defined by log-likelihoods for each label given image I at a voxel location $\mathbf{x}$:

$$c(\mathbf{x}, l) = -log(P(l|I(\mathbf{x}))) = -log\left(\frac{p(I(\mathbf{x})|l) \cdot P(l)}{p(I(\mathbf{x}))}\right).$$
(6)

Note that the background label "BG" is placed in the label order between the femur label "FB" and the tibia label "TB" in order to achieve a symmetric formulation.

The likelihood terms $p(I(\mathbf{x})|\text{FB})$ and $p(I(\mathbf{x})|\text{TB})$ are computed from image intensities. Since bones appear dark in T1 weighted MR images, we assume a simple model (7) to estimate bone likelihoods,

$$p(I(\mathbf{x})|\text{FB}) = p(I(\mathbf{x})|\text{TB}) = exp(-\beta I(\mathbf{x})),$$
(7)

where β is set to 0.02 in our implementation assuming $I(\mathbf{x}) \in [0, 100]$.

To compute the prior terms $p(\text{FB})$ and $p(\text{TB})$ in (6), we employ a multi-atlas registration approach followed by label fusion. Suppose we have N atlases A_i and their bone segmentations S_i^{FB} and S_i^{TB} ($i = 1, 2, ..., N$). Registration from an atlas A_i to a query image I is an affine registration T_i^{affine} followed by a B-Spline registration T_i^{bspline}. Averaging all N propagated atlas labels yields a spatial prior of femur and tibia for the query image:

$$p(\text{FB}) = \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{bspline}} \circ T_i^{\text{affine}} \circ S_i^{\text{FB}}\right),$$

$$p(\text{TB}) = \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{bspline}} \circ T_i^{\text{affine}} \circ S_i^{\text{TB}}\right).$$
(8)

Now that we have computed the spatial priors and the local likelihoods, we integrate them into (6) and solve (3) to obtain the three-label bone segmentation. The bone segmentation will help locate the cartilage in atlas-based cartilage segmentation.

5. Probabilistic classification

We use the three-label segmentation with anisotropic regularization for cartilage segmentation to account for thin cartilage layers. The labeling cost c for each label l in $\{\text{FC}, \text{BG}, \text{TC}\}$ ("FC", "BG" and "TC" denote the femoral cartilage, the background and the tibial cartilage respectively) are defined by log-likelihoods for each label:

$$c(\mathbf{x}, l) = -log(P(l|\mathbf{f}(\mathbf{x}))) = -log\left(\frac{p(\mathbf{f}(\mathbf{x})|l) \cdot p(l)}{p(\mathbf{f}(\mathbf{x}))}\right),$$
(9)

where $\mathbf{f}(\mathbf{x})$ denotes a *feature* vector at a voxel location $\mathbf{x}$. Again the background label "BG" is placed between the femoral cartilage label "FC" and the tibial cartilage label "TC" in order to achieve a symmetric formulation.

We compute the spatial prior $p(l)$ in two different ways: using an average-shape-atlas registration and a multi-atlas registration (see sections 6 and 7). We compare the performance of both approaches in section 9. The local likelihood term $p(\mathbf{f}(\mathbf{x})|l)$ is obtained from a probabilistic classification based on local image appearance. We investigate classification based on a probabilistic k nearest neighbors (kNN) [8] as well as by a support vector machine (SVM) [4]. For classification we use a reduced set of features compared to [11]: intensities on three scales, first-order derivatives in three directions on three scales and second-order derivatives in the axial direction on three scales. The three different scales are obtained by convolving with Gaussian kernels of $\sigma = 0.3\ mm, 0.6\ mm$ and $1.0\ mm$. All features are normalized to be centered at 0 and have unit standard deviation.

An important difference from [11] and [17] is the probabilistic nature of our classification which allows for an easy incorporation of the classification result into our Bayesian framework. Further, the final segmentations are generated by a segmentation method with anisotropic regularization, whereas no regularization was used in [11] nor [17]. We demonstrate in section 9 that spatial regularization helps improve the segmentation accuracy and anisotropic regularization yields better accuracy than isotropic regularization.

5.1. Classification using kNN

We estimate the data likelihoods for femoral and tibial cartilage, $p(\mathbf{f}(\mathbf{x})|l)$, of (9) by probabilistic kNN classification [8]. We use a one-versus-other classification strategy and the expert segmentations of femoral and tibial cartilage to build the kNN classifier. Specifically, let "FC" denote the femoral cartilage class, "TC" the tibial cartilage and "BG" the background class. The training samples of class FC are the voxels labeled as femoral cartilage. Similarly, the training samples of class TC are the voxels labeled as tibial cartilage. The training samples of class BG are the voxels surrounding the femoral and tibial cartilage within a specified distance. The outputs of the probabilistic kNN classifier given a query voxel $\mathbf{x}$ with its feature vector $\mathbf{f}(\mathbf{x})$ are:

$$p(\mathbf{f}(\mathbf{x})|\text{FC}) = \frac{n^{\text{FC}}(\mathbf{f}(\mathbf{x}))}{k},$$

$$p(\mathbf{f}(\mathbf{x})|\text{BG}) = \frac{n^{\text{BG}}(\mathbf{f}(\mathbf{x}))}{k},$$

$$p(\mathbf{f}(\mathbf{x})|\text{TC}) = \frac{n^{\text{TC}}(\mathbf{f}(\mathbf{x}))}{k}.$$
(10)

Here n^{FC}, n^{TC}, n^{BG} denote the number of votes for the femoral cartilage, tibial cartilage, and background class respectively; k is the number of nearest neighbors of concern. Since kNN is sensitive to the number of training samples, we scale the outputs according to the training class sizes to balance the three classes.

5.2. Classification using an SVM

An alternative approach to compute the local likelihoods is to use a support vector machine (SVM) [4], which constructs a hyperplane maximally separating classes given a labeled training set. Koo *et al.* [17] proposed to use two-class SVM to segment cartilage automatically from multi-contrast MR images. We apply LIBSVM [3] to perform probabilistic three-class SVM classification with the features described above. The results are local likelihoods for the background, the femoral and the tibial cartilage, i.e., $p(\mathbf{f}(\mathbf{x})|\text{BG})$, $p(\mathbf{f}(\mathbf{x})|\text{FC})$ and $p(\mathbf{f}(\mathbf{x})|\text{TC})$. We compare the SVM and the kNN probabilistic classification methods in section 9.

6. Average-shape-atlas-based Cartilage Segmentation

This section discusses how to build a probabilistic bone and cartilage atlas by averaging registered expert segmentations and computing the cartilage spatial priors by registration of the atlas. The atlas within this section captures the spatial relationships between the bone and the cartilage.

Suppose we have N images with expert segmentations. We pick the segmentation of one image as the reference to bring all the segmentations to the same position. Specially, we register the femur segmentation S_i^{FB} ($i = 1, 2, ..., N$) and the tibial segmentation S_i^{TB} ($i = 1, 2, ..., N$) to the reference femur and tibia segmentations with affine transforms T_i^{FB} and T_i^{TB} respectively. The femoral and tibial cartilage segmentations S_i^{FC} and S_i^{FC} are propagated accordingly. The average bone and cartilage atlas A_{avg} (including A_{avg}^{FB}, A_{avg}^{TB}, A_{avg}^{FC} and A_{avg}^{TC}) is computed by

$$
\begin{aligned}
A_{avg}^{\text{FB}} &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{FB}} \circ S_i^{\text{FB}} \right), \\
A_{avg}^{\text{TB}} &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{TB}} \circ S_i^{\text{TB}} \right), \\
A_{avg}^{\text{FC}} &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{FB}} \circ S_i^{\text{FC}} \right), \\
A_{avg}^{\text{TC}} &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{TB}} \circ S_i^{\text{TC}} \right).
\end{aligned}
\tag{11}
$$

Given a query image I, we have computed the bone segmentation S^{FB} and S^{TB} from section 4. The atlas femur A^{FB} and tibia A^{FB} are registered to the segmentation of femur S^{FB} and tibia S^{TB} with affine transforms T^{FB} and T^{TB}. The spatial prior for each cartilage is then computed by propagating each cartilage atlas with the corresponding transform,

$$
\begin{aligned}
p(\text{FC}) &= T^{\text{FB}} \circ A_{avg}^{\text{FC}}, \\
p(\text{TC}) &= T^{\text{TB}} \circ A_{avg}^{\text{TC}}, \\
p(\text{BG}) &= 1 - p(\text{FC}) - p(\text{TC}).
\end{aligned}
\tag{12}
$$

These spatial priors and the local likelihoods from section 5 are integrated into (9) and the cartilage segmentation is obtained by optimizing the three-label segmentation energy with anisotropic regularization (4).

7. Multi-atlas-based Cartilage Segmentation

This section presents an alternative approach to computing the spatial prior for cartilage. We make use of multi-atlas registration, rather than average-shape-atlas registration as described in section 6. Each atlas is an individual expert bone and cartilage segmentation in this section. Three popular label fusion methods are discussed in this section, i.e., majority voting, locally-weighted and non-local patch-based fusion.

We have N atlases A_i including their femur segmentations S_i^{FB}, tibia segmentations S_i^{TB}, femoral cartilage segmentations S_i^{FC} and tibial cartilage segmentations S_i^{TC} ($i = 1, 2, ..., N$). For a query image I, we have the bone segmentation S^{FB} and S^{TB} from section 4.

The atlas bone segmentations S_i^{FB} and S_i^{TB} are registered to the bone segmentations S^{FB} and S^{TB} of the query image separately by affine transforms T_i^{FB} and T_i^{TB}.

We can simply take the average of the registered cartilage atlas segmentations to compute the spatial priors, which is majority voting [18] label fusion:

$$
\begin{aligned}
p(\text{FC}) &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{FB}} \circ S_i^{\text{FC}} \right), \\
p(\text{TC}) &= \frac{1}{N} \sum_{i=1}^{N} \left(T_i^{\text{TB}} \circ S_i^{\text{TC}} \right).
\end{aligned}
\tag{13}
$$

We can also apply a locally-weighted label fusion strategy [15], which was shown to yield a better segmentation accuracy than a majority voting strategy. In this case, we choose to favor the atlases which locally agree better with the cartilage likelihoods $p(\mathbf{f}(\mathbf{x})|\text{FC})$ and $p(\mathbf{f}(\mathbf{x})|\text{TC})$ from the probabilistic classification in section 5. The spatially varying weighting functions λ_i^{FC} for the femoral cartilage and λ_i^{TC} for the tibial cartilage are calculated as

$$
\begin{aligned}
\lambda_i^{\text{FC}}(\mathbf{x}) &= \frac{1}{\alpha \left| T_i^{\text{FB}} \circ S_i^{\text{FC}} - p(\mathbf{f}(\mathbf{x})|\text{FC}) \right| + \epsilon}, \\
\lambda_i^{\text{TC}}(\mathbf{x}) &= \frac{1}{\alpha \left| T_i^{\text{TB}} \circ S_i^{\text{TC}} - p(\mathbf{f}(\mathbf{x})|\text{TC}) \right| + \epsilon},
\end{aligned}
\tag{14}
$$

followed by a small amount of diffusion smoothing. We choose $\alpha = 0.2$ and $\epsilon = 0.001$ in our implementation. The spatial prior for each cartilage is then the weighted average of the propagated atlas cartilage segmentations

$$
\begin{aligned}
p(\text{FC}) &= \sum_{i=1}^{N} \frac{\lambda_i^{\text{FC}}}{\sum_{i=1}^{N} \lambda_i^{\text{FC}}} \left(T_i^{\text{FB}} \circ S_i^{\text{FC}} \right), \\
p(\text{TC}) &= \sum_{i=1}^{N} \frac{\lambda_i^{\text{TC}}}{\sum_{i=1}^{N} \lambda_i^{\text{TC}}} \left(T_i^{\text{TB}} \circ S_i^{\text{TC}} \right).
\end{aligned}
\tag{15}
$$

Recently, non-local patch-based label fusion techniques have been proposed [5, 19]. Instead of deciding the label from the same voxel location in each propagated atlas, these methods obtain a label using the surrounding patches in a predefined neighborhood across the training atlases. Weights are assigned to these patches according to the distances between the target patch and the selected patches. This allows local robustness to registrations error.

Let $p^{\text{FC}}(\mathbf{x})$ and $p^{\text{TC}}(\mathbf{x})$, respectively, denote the spatial prior of femoral cartilage (i.e. $p(\text{FC})$) and tibial cartilage, (i.e. $p(\text{TC})$) at voxel $\mathbf{x}$. We calculate the probabilities by weighted averages of the propagated labels in a pre-specified search neighborhood $\mathcal{N}(\mathbf{x})$ across N warped atlases. The weights are determined by local patch similarities. For simplicity, let $\tilde{S}_i^{\text{FC}} = T_i^{\text{FB}} \circ S_i^{\text{FC}}$ and $\tilde{I}_i^{\text{FC}} = T_i^{\text{FB}} \circ I_i$. Here, i is the atlas index, running from 1 to N, S_i^{FC} refers to the femoral cartilage segmentation of the i-th atlas, and I_i is the i-th atlas appearance. For the femoral cartilage, we have

$$p^{\text{FC}}(\mathbf{x}) = \frac{\sum_{i=1}^{N} \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w^{\text{FC}}(\mathbf{x}, \mathbf{y}) \tilde{S}_i^{\text{FC}}(\mathbf{y})}{\sum_{i=1}^{N} \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} w^{\text{FC}}(\mathbf{x}, \mathbf{y})}, \qquad (16)$$

$$w^{\text{FC}}(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{\sum_{\mathbf{x}' \in \mathcal{P}(\mathbf{x})} \sum_{\mathbf{y}' \in \mathcal{P}(\mathbf{y})} \left(I(\mathbf{x}') - \tilde{I}_i^{\text{FC}}(\mathbf{y}') \right)^2}{h^{\text{FC}}(\mathbf{x})} \right), \tag{17}$$

where $\mathbf{x}'$ is a voxel in the patch $\mathcal{P}(\mathbf{x})$ centered at $\mathbf{x}$ (similarly $\mathbf{y}'$ a voxel in the patch $\mathcal{P}(\mathbf{y})$ centered at $\mathbf{y}$) and $h^{\text{FC}}(\mathbf{x})$ is defined by

$$h^{\text{FC}}(\mathbf{x}) = \min_{\substack{1 \le i \le N \\ \mathbf{y} \in \mathcal{N}(\mathbf{x})}} \sum_{\substack{\mathbf{x}' \in \mathcal{P}(\mathbf{x}) \\ \mathbf{y}' \in \mathcal{P}(\mathbf{y})}} \left(I(\mathbf{x}') - \tilde{I}_i^{\text{FC}}(\mathbf{y}') \right)^2 + \epsilon. \tag{18}$$

Substitute "FB" with "TB" and "FC" with "TC" in superscripts of the equations above for the calculation of $p^{\text{TC}}(\mathbf{x})$.

The three label fusion strategies, namely majority voting, locally-weighted and non-local patch-based fusion, are compared in section 9. The non-local patch-based method is shown to result in the best average segmentation accuracy.

These spatial priors and the local likelihoods from section 5 are integrated into (9) and the cartilage segmentation is obtained by optimizing the three-label segmentation energy with anisotropic regularization (4).

8. Overall segmentation pipeline

The automatic cartilage segmentation requires expert segmentations of femur, tibia, femoral and tibial cartilage on a set of training images. Given a query knee image, we first correct the MRI bias field [25], scale image intensities to a common range, and then perform edge-preserving smoothing using curvature flow [21].

In the multi-atlas-based bone segmentation, the atlases are registered to the query images with an affine transform followed by a B-spline transform based on mutual information. We compute the average of the propagated atlas bone segmentations as the bone spatial priors. The bone likelihoods are then calculated from the image intensities using (7). The priors and the likelihoods are combined in (6) and then integrated in the three-label segmentation (3), the global optimal solution of which produces the bone segmentation.

Once we have the bone segmentation, we perform the probabilistic classification (kNN or SVM) of knee cartilage in the joint region. The spatial priors for the cartilage can be obtained through registration of an average bone and cartilage atlas, which requires only one registration, or through a multi-atlas registration of cartilage, which needs a number of registrations. If a multi-atlas-based method is chosen, propagated atlas labels are fused (using majority voting, locally-weighted or non-local patch-based label fusion) to obtain the spatial priors. The normal direction $\mathbf{n}$ in (5) is computed by taking the gradient of the diffusion smoothed three-label bone segmentation result in-between the joint area. Finally, the local likelihoods and the spatial priors are integrated into the three-label segmentation to generate the cartilage segmentation.

9. Experimental results

9.1. Data description

Our main dataset is a subset of the PLS dataset, containing 706 T1-weighted (3D SPGR) images for 155 subjects, imaged at baseline, 3, 6, 12, and 24 months at a resolution of $1.00 \times 0.31 \times 0.31 \; mm^3$. Some subjects have missing scans. The Kellgren-Lawrence grades (KLG) [16] were determined for all subjects from the baseline scans, classifying 82 as normal control subjects (KLG0), 40 as KLG2 and 33 as KLG3.

Expert cartilage segmentations (drawn by a domain expert, Dr. Felix Eckstein) are available for all images. The femoral cartilage segmentation is drawn only on the weight-bearing part while the tibial cartilage segmentation covers the entire region. Therefore, we expect partial femoral cartilage and full tibial cartilage segmentation results.

9.2. Bone validation

Bone segmentation is validated on 18 images because expert segmentations for bones are only available for the baseline images of 18 subjects. We validate our multi-atlas-based bone segmentation method in a leave-one-out manner. Each test image is segmented using the other 17 images as atlases. The segmentation accuracy is evaluated

Table 1: Statistics (mean and standard deviation (STD)) of DSC of bone segmentation on 18 test images with and without spatial regularization.

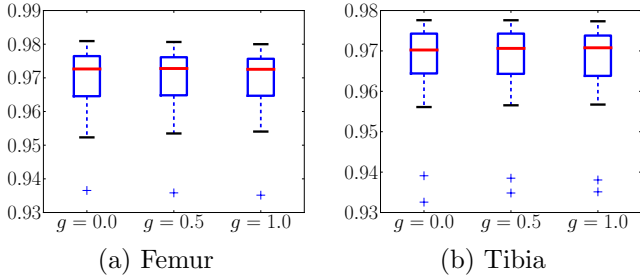|		$g = 0$	$g = 0.5$	$g = 1.0$
Femur	Mean	0.969	0.970	0.969
	STD	0.011	0.011	0.011
Tibia	Mean	0.966	0.967	0.966
	STD	0.013	0.012	0.012

(a) Femur   (b) Tibia

Figure 6: Box plots of DSC for femur and tibia with different amount of regularization on 18 test images. The center red line is the median and the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

with respect to the expert segmentations using the Dice similarity coefficient (DSC) [6] defined as

$$\text{DSC} = \frac{2\,|S \cap R|}{|S| + |R|}, \tag{19}$$

where S and R represent two segmentations. Table 1 and Fig. 6 show the validation results of the bone segmentation with and without regularization (corresponding to $g > 0$ and $g = 0$ in model (3) respectively). No significant improvement is observed by introducing spatial regularization to the bone segmentation, because the multi-atlas-based spatial prior nicely locates the bones. We can see from Fig. 7 that the multi-atlas-based prior captures the bone very well and our segmentation result is very close to the expert segmentation especially in the joint region. We use the bone segmentation with spatial regularization $g = 0.5$ to compute the cartilage segmentations for the remaining experiments in this section.

9.3. Cartilage validation

Figure 8 illustrates the beneficial behavior of our three-label segmentation method compared to a binary segmentation which treats femoral and tibial cartilage as one object. While the three-label method is able to keep femoral and tibial cartilage separated due to the joint estimation of the segmentation, the binary segmentation approach cannot guarantee this separation.

We build an average shape atlas of bone and cartilage from the expert bone and cartilage segmentations of the 18 images. Figure 9 shows an example slice of the average probabilistic bone and cartilage atlas and the 3-dimensional rendering. The cartilage is well located on top of the bone.
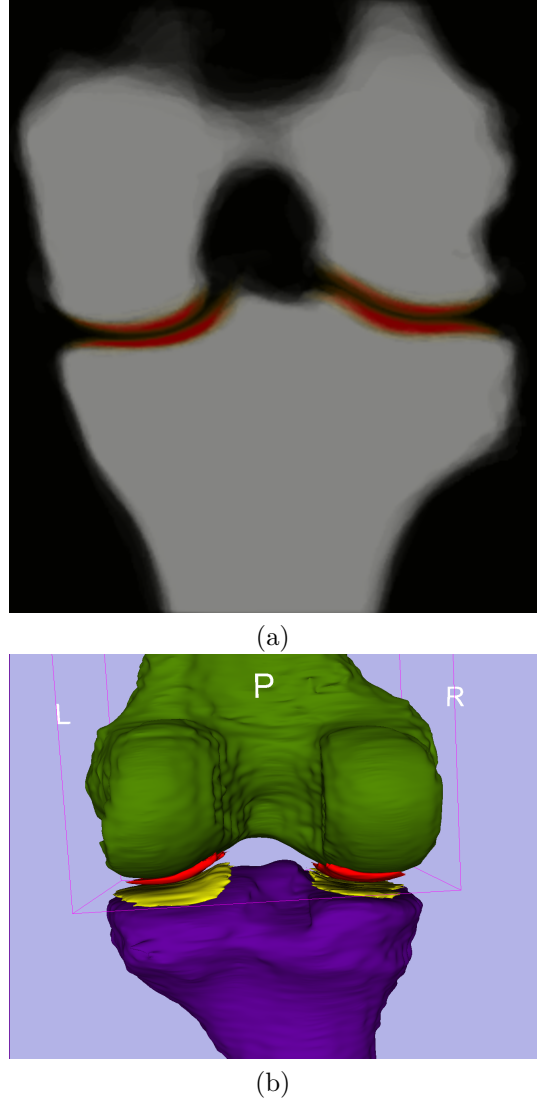

(a)

(b)

Figure 9: Atlas built from 18 images. (a) is a slice of the probabilistic atlas of femoral and tibial bone and cartilage (red) overlaid on the bone in coronal view. Saturated red denotes high probability. (b) is a 3-dimensional rendering of the thresholded atlas of femur (green), tibia (purple), femoral (red) and tibial cartilage (yellow).

In the average-shape-atlas-based cartilage segmentation, we use the atlas built from 18 images (each from a different subject) to segment cartilage of the remaining 137 subjects. Within the 18 subjects, we test in a leave-one-out manner where each subject is segmented using the atlas built from the other 17 subjects. The same strategy is applied in the multi-atlas-based cartilage segmentation. We use all 18 images as atlases to segment cartilage of the other 137 subjects. The 18 subjects are tested in a leave-one-out fashion. Each subject is segmented using the other 17 images as atlases. The training images for kNN and SVM are chosen in the same way.

In the non-local patch-based label fusion, we upsample the images to approximately isotropic resolution and search for similar $5 \times 5 \times 5$ patches within a $9 \times 9 \times 9$ neighborhood.
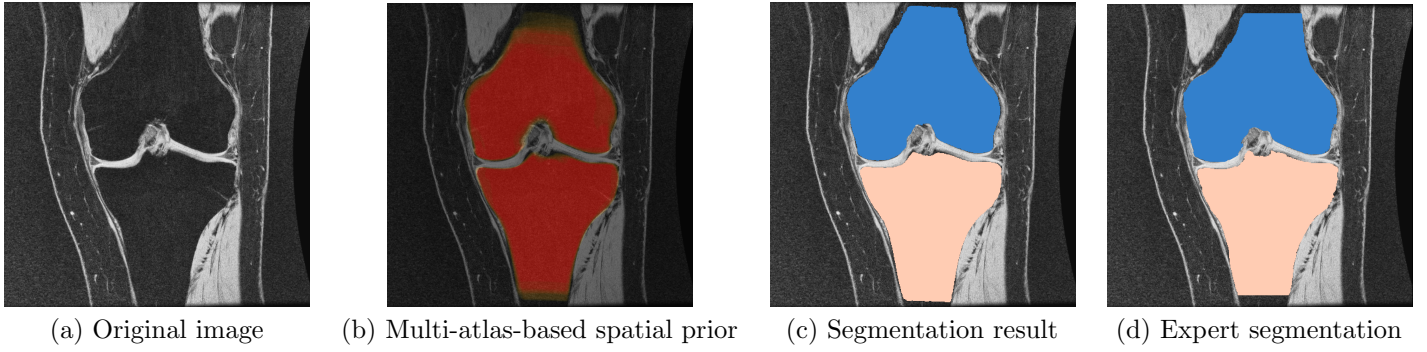
8

(a) Original image   (b) Multi-atlas-based spatial prior   (c) Segmentation result   (d) Expert segmentation

Figure 7: Bone segmentation of one example slice in coronal view.

(a) Binary segmentation   (b) Three-label segmentation   (c) Expert segmentation
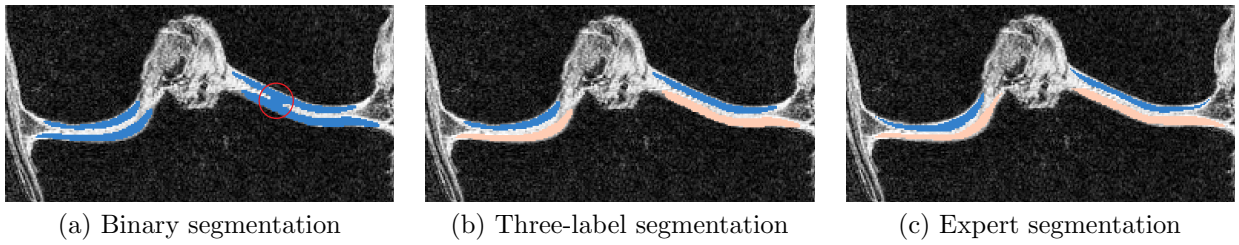
Figure 8: Example comparing binary and three-label segmentation methods. (a) is the binary segmentation result. (b) is the three-label segmentation result in which femoral and tibial cartilage have distinct labels. (c) is the expert segmentation. In (a), as the red circle indicates, the lateral (right) femoral cartilage and tibial cartilage are segmented as one object and the joint boundary is not well captured. The three-label segmentation (b) keeps the femoral and tibial cartilage separate and is therefore superior to binary segmentation.
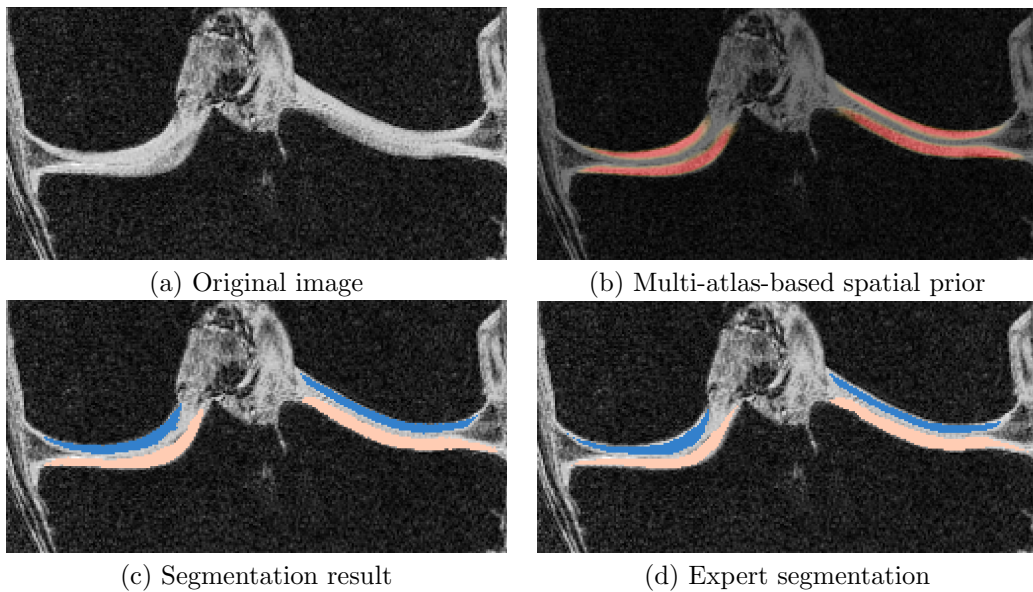

(a) Original image   (b) Multi-atlas-based spatial prior

(c) Segmentation result   (d) Expert segmentation

Figure 10: Cartilage segmentation of one example slice in coronal view. Only joint region is shown.

(a) Average-atlas

(b) Multi-atlas with majority voting

(c) Multi-atlas with locally-weighted fusion

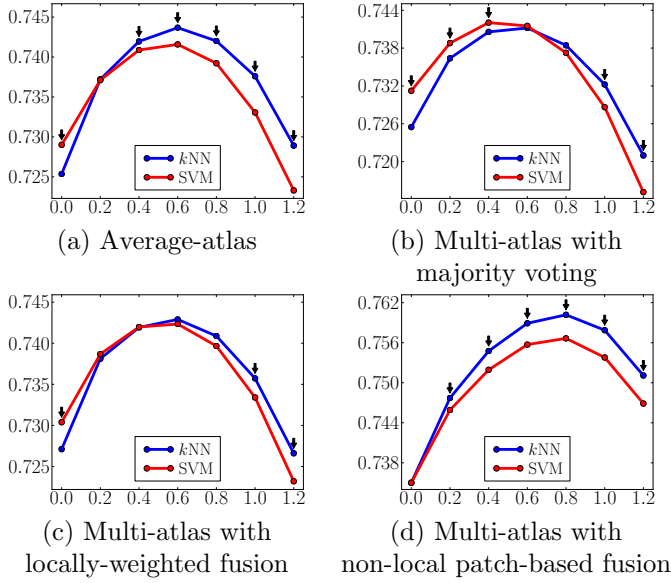(d) Multi-atlas with non-local patch-based fusion

Figure 11: Comparison of kNN and SVM based on the mean DSC (ordinate) with varying amount of isotropic regularization (abscissa g) under different atlas choices for the **femoral** cartilage. The black downarrows ($\Downarrow$) indicate statistically significant differences between the two methods at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

(a) Average-atlas

(b) Multi-atlas with majority voting

(c) Multi-atlas with locally-weighted fusion

(d) Multi-atlas with non-local patch-based fusion

Figure 12: Comparison of kNN and SVM based on the mean DSC (ordinate) with varying amount of isotropic regularization (abscissa g) under different atlas choices for the **tibial** cartilage. The black downarrows ($\Downarrow$) indicate statistically significant superiority of SVM to kNN at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

Figures 11 and 12 compare the two local classification methods, i.e., kNN versus SVM, for the femoral and the tibial cartilage, under different atlas choices with varying amount of isotropic spatial regularizations. Note that the femoral cartilage is only segmented in the weight-bearing region and hence the DSC for the femoral cartilage is more sensitive to mis-segmentations than the tibial cartilage. For the femoral cartilage, kNN and SVM generate similar mean DSC. The SVM improves the mean DSC by a considerable amount over the kNN for the tibial cartilage. A possible reason for the similar performance for the femoral cartilage might be that the main disagreement between the automatic and the expert segmentation is along the anterior-posterior direction delineating the weight-bearing region, which may overwhelm any improvement obtained by SVM over kNN. SVM performs better than kNN for tibial cartilage which is segmented in its entirety.

Figure 13 compares the different atlas choices, including the average-shape atlas, multiple atlases with majority voting, locally-weighted and non-local patch-based label fusion, under the different parameter settings of isotropic regularization. The former three yield very similar mean DSC. Non-local patch-based label fusion outperforms the other three considerably. Figure 14 compare the four atlas choices under the different parameter settings of anisotropic regularization. Again, non-local patch-based label fusion outperforms the other three considerably.

Figure 15 shows the advantage of anisotropic regularization. The isotropic regularization has a tendency to cut long and thin objects short as shown in Fig. 15 (a) at the medial femoral cartilage. Anisotropic regularization, on
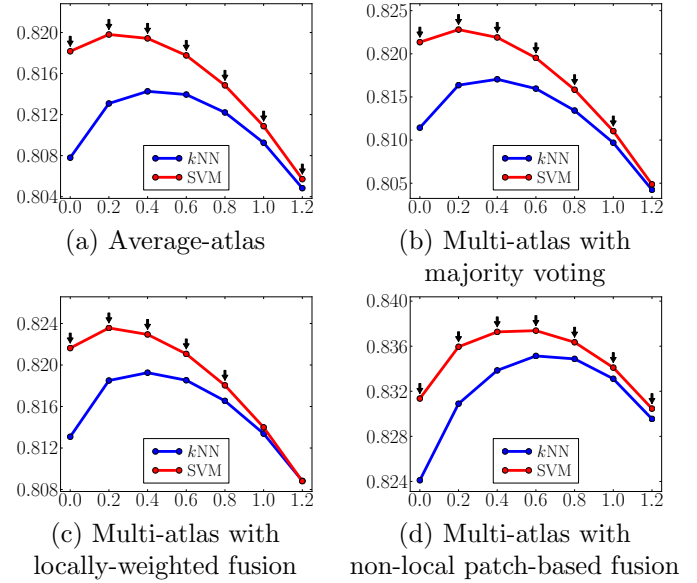

(a) Femoral cartilage kNN

(b) Femoral cartilage SVM

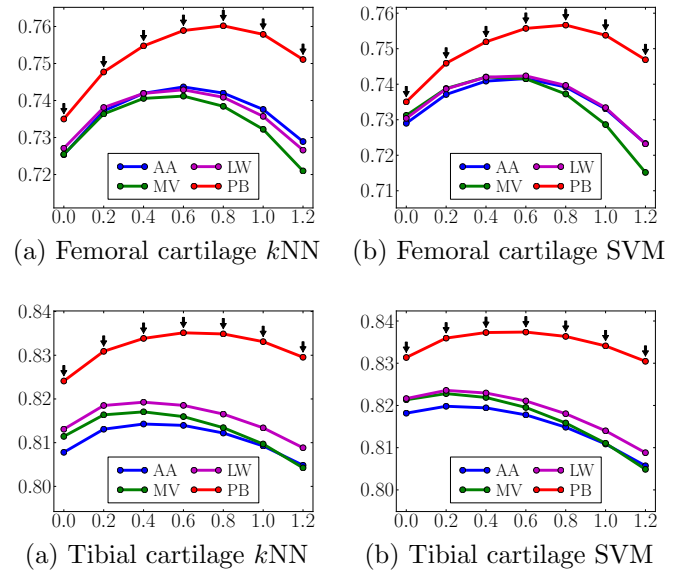(a) Tibial cartilage kNN

(b) Tibial cartilage SVM

Figure 13: Comparisons of mean DSC (ordinate) from different atlas choices for different amount of **isotropic** regularization (abscissa g). AA average-shape-atlas. MV multi-atlas with majority voting. LW multi-atlas with locally-weighted fusion. PB multi-atlas with non-local patch-based fusion. The black downarrows ($\Downarrow$) indicate statistically significant superiority of PB to the other three methods at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

the other hand, avoids this problem (see Fig. 15 (b)) resulting in a better segmentation of the medial femoral cartilage. Besides avoiding unrealistic segmentation results, anisotropic regularization is also less sensitive to parameter settings than isotropic regularization. This is illus-
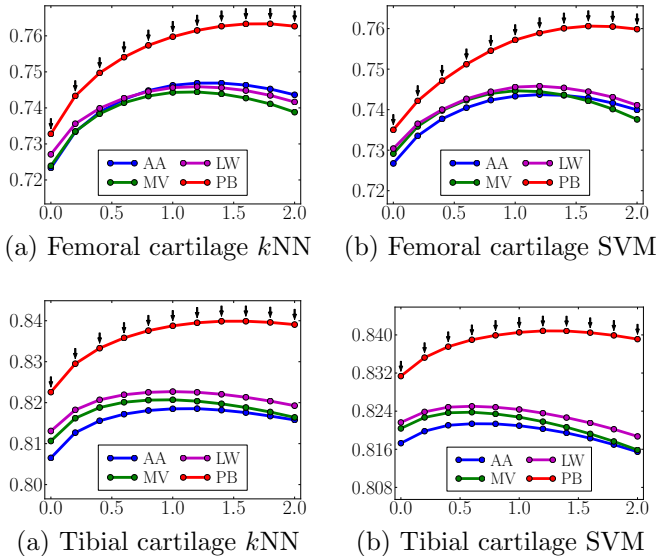
(a) Femoral cartilage kNN   (b) Femoral cartilage SVM

(a) Tibial cartilage kNN   (b) Tibial cartilage SVM

Figure 14: Comparisons of mean DSC (ordinate) from different atlas choices for different amount of **anisotropic** regularization (abscissa g). The parameter α controlling the anisotropy is set to be 0.2. AA average-shape-atlas. MV multi-atlas with majority voting. LW multi-atlas with locally-weighted fusion. PB multi-atlas with non-local patch-based fusion. The black downarrows ($\Downarrow$) indicate statistically significant superiority of PB to the other three methods at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

(a) Femoral cartilage   (b) Tibial cartilage

Figure 16: Change of mean DSC for femoral and tibial cartilage with isotropic and anisotropic regularization over the amount of regularization g (abscissa). The parameter α is set to be 0.2 for all anisotropic tests. All tests use SVM and non-local patch-based label fusion. The black downarrows ($\Downarrow$) indicate statistically significant differences between the two methods at corresponding spatial regularization settings via paired t-tests at a significance level of 0.05.

(a) Femoral cartilage   (b) Tibial cartilage
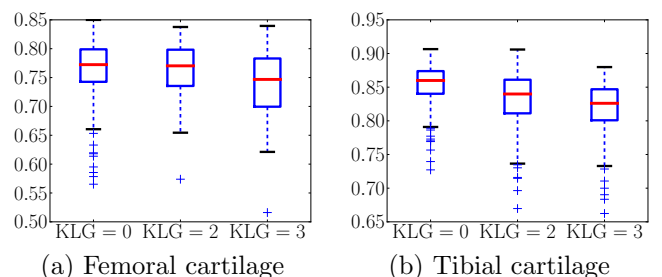
Figure 17: Boxplots of DSC's for different KLG's. We choose the best strategy combination, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.

Table 2: Statistics summary (mean, median and standard deviation) of DSC under the best strategy combination: SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$ from the PLS dataset.

	Mean	Median	STD
Femoral cartilage	0.760	0.768	0.048
Tibial cartilage	0.841	0.847	0.037

trated in Fig. 16 (a) and (b). Note that the anisotropic regularizer is parametrized in such a way that its regularization is reduced in the normal direction, but equal to the isotropic regularization in the plane orthogonal to the normal and the results are therefore comparable (see Fig. 5). The faster drop-off in the isotropic case indicates a stronger dependency on the parameter settings for isotropic regularization.

For anisotropic regularization, to select the "optimal" parameter g for the PLS dataset, we tried different values of $g \in [0, 2.0]$ and found that $g = 1.4$ yields the best average DSC (0.764 and 0.840 for femoral and tibial cartilage respectively) for the 18 training subjects. We apply this "optimal" parameter setting to the test data (the remaining 137 subjects). This setting yields an average DSC of 0.760 for femoral cartilage and 0.841 for tibial cartilage. We use the same parameter setting for a completely independent data set, SKI10 [14], and obtain an average DSC of 0.856 and 0.859 for femoral and tibial cartilage respectively. Even though the "optimal" g could be different for different datasets, the choice of $g = 1.4$ appeared to be a good candidate for cartilage segmentation.

As anisotropic regularization is less sensitive to parameter settings, the choice of g does not have a strong impact on segmentation results. Parameter α is set to be 0.2 in our experiments but other values ($\alpha \in [0, 1]$) can also be used. The influence of α will be studied in the future.

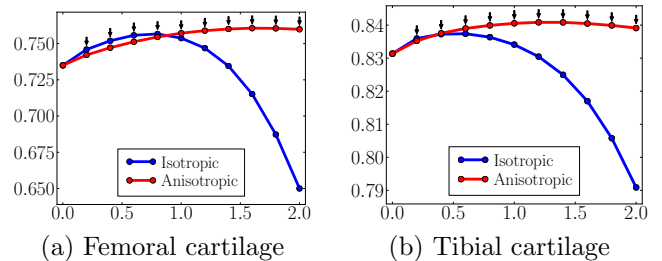To further illustrate segmentation behavior, we show the box plots of the DSC for different progression levels (i.e.,

KLGs) for femoral and tibial cartilage in Fig.17. As expected, we observe a slight deterioration in segmentation accuracy for larger KL grades as it is more challenging to segment pathological knee cartilage.

Figure 18 shows scatter plots of segmentation volumes of the proposed method versus the expert segmentation. The correlation between the volume measured from the expert segmentation and the automatic algorithm achieves a Pearson's correlation coefficient of 0.77 for all subjects (KLG0: 0.85, KLG2: 0.68, KLG3: 0.74) for the femoral cartilage. For the tibial cartilage, the Pearson's correlation coefficient is 0.87 for all subjects (KLG0: 0.89, KLG2: 0.80, KLG3: 0.89).

The local cartilage thickness is computed from the cartilage segmentation using a Laplace-equation approach [30]. We compute the correlation coefficient of local thickness maps from the expert and the proposed segmentations for each image. Figure 19 shows box plots of Pearson's correlation coefficients for different KLG's. Thicknesses of the automatic and the expert segmentations are strongly cor-

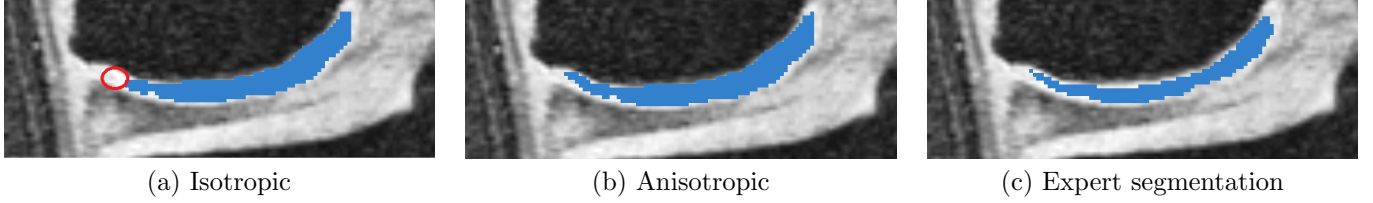|		
(a) Isotropic	(b) Anisotropic	(c) Expert segmentation

Figure 15: Improvement by anisotropic regularization. (a) uses isotropic regularization and misses circled region. (b) uses anisotropic regularization and captures the missing region in (a). (c) is the expert segmentation.

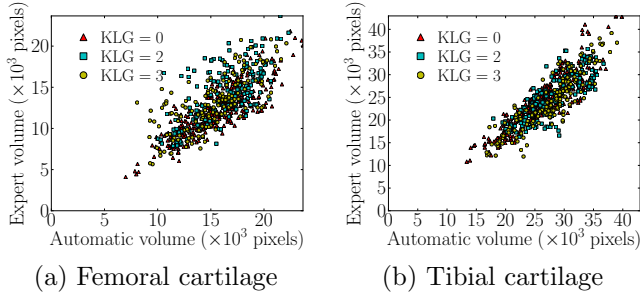|	
(a) Femoral cartilage	(b) Tibial cartilage

Figure 18: Scatter plots of segmentation volumes (number of pixels). We choose the best strategy combination, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.

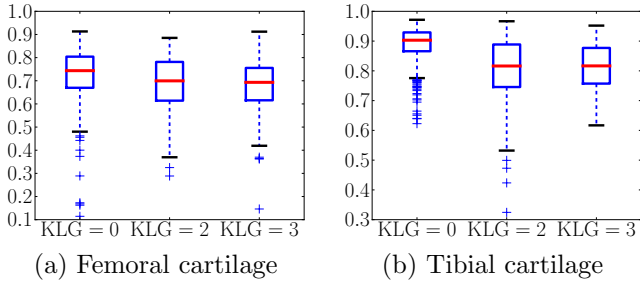|	
(a) Femoral cartilage	(b) Tibial cartilage

Figure 19: Boxplots of Pearson's correlation coefficients of local cartilage thickness for different KLG's. We choose the best strategy combination, SVM and non-local patch-based label fusion with an anisotropic regularization with $g = 1.4$ and $\alpha = 0.2$.

9.4. Running time

The overall running time for the segmentation of an MR image is hours. Each atlas registration takes 10 - 30 minutes. If registrations are done sequentially, this step takes up to 9 hours (as there are 18 atlases). The patch-based label fusion step is completed in 10 minutes. The local tissue classification takes about 20 minutes. The computation time for the three-label segmentation varies from minutes to hours. Using anisotropic regularization is slower than isotropic. The running time also depends on the amount of regularization. Large regularization requires more iterations to converge. The segmentation step can be sped up

drastically by a GPU implementation, which will be part of the future work.

10. Comparison to other methods

We quantitatively compare methods based on the SKI10 dataset and qualitatively discuss methods which have so far not been tested on SKI10.

10.1. Comparisons based on the SKI10 dataset

To compare to other algorithms we use the data from the cartilage segmentation challenge SKI10 [14]. We randomly pick 15 images from the provided 60 training images as atlases to limit computational cost (in principle all 60 images could be used as atlases). SKI10 uses a combined score based on volume difference and volume overlap error for cartilage *and* bone to score different methods. At time of writing, SKI10 included results for 16 different methods. We restrict ourselves to comparisons between the top 8 methods. The proposed method ranks 5/16 overall. However, as we will discuss below our proposed method performs as well as the top method on volume overlap error for cartilage segmentation (or equivalently Dice coefficient) which as we argue is the most important of the performance measures. For simplicity we denote the methods as Rank 1 to Rank 8 to simplify readability. Tables 3 and 4 contains references and names of the methods as available.

Note that the SKI10 dataset is very challenging as its data was collected from pre-surgery cases, which exhibit severe cartilage damage. It should therefore be regarded as complementing the OAI and the PLS data for validation which cover a much broader range of cartilage degeneration and damage. In particular, the performance of an algorithm on the PLS or OAI data may be more informative for future clinical drug trials aimed at showing small changes in cartilage in relation to therapy.

Figure 20 shows different measures for femoral and tibial cartilage from the top 8 methods . The volumetric difference and volumetric overlap error (VOE) are defined as follows, given a segmentation S and a reference segmentation R.

$$\text{VOE} = 100 \left(1 - \frac{|S \cap R|}{|S \cup R|} \right), \tag{20}$$

$$\text{VD} = 100 \frac{|S| - |R|}{|R|}. \tag{21}$$

related. Note that correlations for femoral cartilage with respect to volume and thickness are generally lower than for the tibial cartilage due to the fact that only the weight-bearing region of the femoral cartilage is being segmented.

12

The challenge defined a scoring system based on inter-observer variations of VD and VOE. On a range from 0 to 100 (meaning a perfect segmentation), a second rater's outcome corresponds to 75, a result with error twice as high gets 50 and so on. The Dice coefficient can be computed from the VOE as follows

$$\text{DSC} = \frac{200 - 2 \times \text{VOE}}{200 - \text{VOE}}. \qquad (22)$$

Our method achieves excellent performance on VOE and DSC. The VD is best at zero: our method performs well on the femoral cartilage but not as well on the tibial cartilage compared to other methods. Note that a low VD, which only compares the segmentation volumes, may not indicate a good segmentation since a good score may be achieved for a similar volume at incorrect locations. As VOE and DSC measure local differences we regard them as more informative than VD for the assessment of cartilage segmentation differences.

Table 3 compares our method to other methods based on the different SKI10 validation measures. Specifically, we test if scores of competing methods are significantly better than for our method. Our method achieves statistically significantly better accuracy than most of the other methods regarding VOE and DSC before and after multiple comparison correction. Table 4 shows that the proposed method has the second best DSC values for femoral and tibial cartilage, which are only marginally lower than for the first ranked method. In particular, we do not observe statistically significant performance difference in VOE and DSC for femoral and tibial cartilage with respect to the top two ranked methods after correction for multiple comparisons. *Before* multiple comparison correction also no statistically significant differences were found expect for an improved performance of our method for femoral cartilage segmentation with respect to the second ranked method by Seim *et al.* [20]. This suggests that our method can be regarded as one of the top-performing methods for femoral and cartilage segmentation on the SKI10 dataset.

Interestingly, the top-performing method is based on an active appearance model [27]. While this puts the method at an advantage for producing segmentations which are within the trained shape and appearance spaces. Variation outside these spaces cannot be properly captured if no subsequent relaxation step is used. Our method can be regarded as softly constraining the space of plausible segmentations through the use of multiple atlases and non-local patch-based label fusion. However, given that atlas information is only included as a prior into our overall segmentation method, our method remains flexible enough to also capture cartilage variations not strictly contained in the atlas set.

Note that the SKI10 [14] images were acquired for knee surgery planning and therefore most images exhibit serious cartilage loss. As the cartilage segmentations for SKI10 were performed semi-automatically, they mostly capture cartilage well, but occasionally tend towards oversegmen-

(a) Femoral cartilage score (b) Tibial cartilage score

(c) Femoral cartilage VD (d) Tibial cartilage VD

(e) Femoral cartilage VOE (f) Tibial cartilage VOE

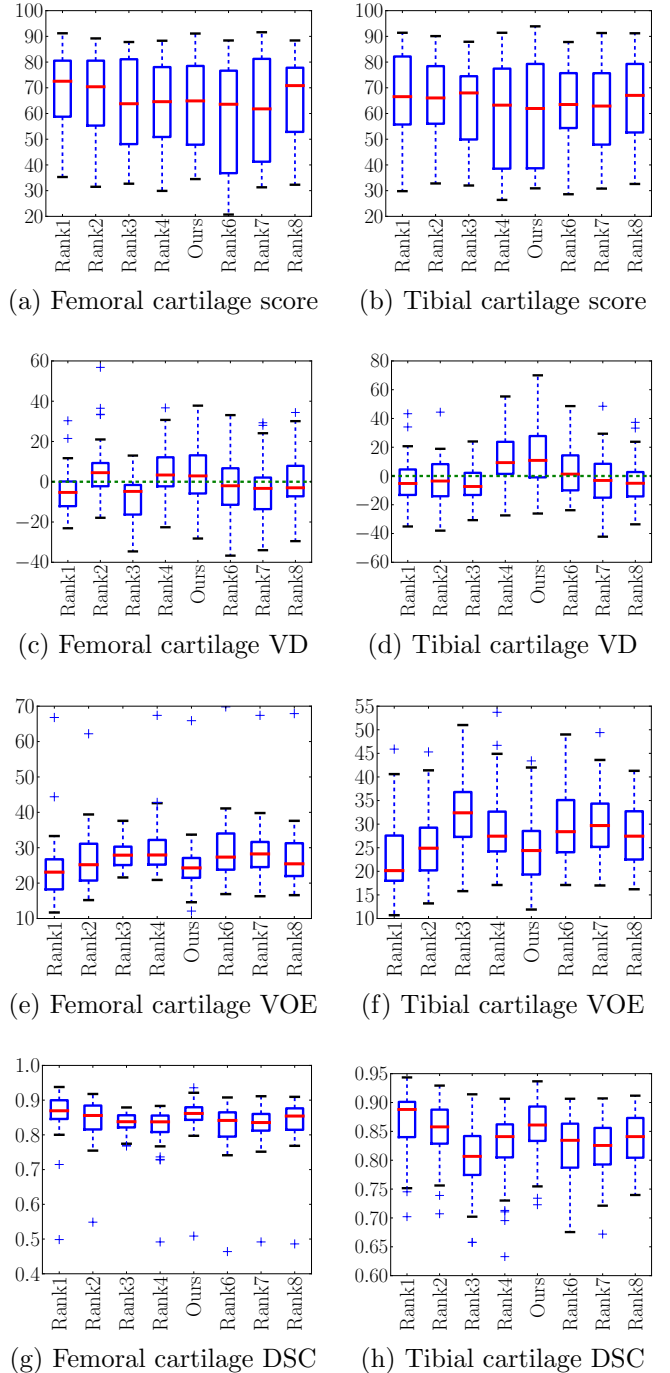(g) Femoral cartilage DSC (h) Tibial cartilage DSC

Figure 20: Box plots of segmentation measures for femoral and tibial cartilage from top 8 ranking methods on SKI10 website. The center red line is the median and the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

tation at pathological regions; e.g., segmenting across regions of total cartilage loss or segmenting osteophytes. Figure 21 shows an example illustrating total cartilage loss and the challenge to define a reliable gold standard segmentation.

Table 3: Results of statistical tests (paired t tests for score, VOE and DSC, Wilcoxon signed-rank tests for VD) between different methods. Our method is compared to the other top ranking methods in terms of different measures. Symbol "+" denotes statistically significant superiority of our method; "−" denotes inferiority; "NS" denotes a statistically insignificant difference ($p > 0.05$). Each table entry consists of two symbols, before and after the correction for multiple comparisons. Rank 7 was also submitted by the authors but using a slightly different combination, i.e., probabilistic kNN and locally-weighted label fusion.

Rank	Team	Femoral cartilage				Tibial cartilage			
		Score	VD	VOE	DSC	Score	VD	VOE	DSC
1	Imorphics [27]	NS/NS	+/+	NS/NS	NS/NS	NS/NS	−/−	NS/NS	NS/NS
2	ZIB [20]	NS/NS	+/NS	+/NS	+/NS	NS/NS	−/−	NS/NS	NS/NS
3	UPMC_IBML	NS/NS	+/+	+/NS	+/NS	NS/NS	−/−	+/+	+/+
4	SNU_SPL	NS/NS	NS/NS	+/+	+/+	NS/NS	NS/NS	+/+	+/+
6	UIiibiKnee [31]	NS/NS	−/−	+/+	+/+	NS/NS	−/−	+/+	+/+
7	shan_unc	NS/NS	+/+	+/+	+/+	NS/NS	−/−	+/+	+/+
8	BioMedIA [28]	NS/NS	+/+	+/+	+/+	NS/NS	−/−	+/+	+/+

Table 4: Statistics summary (mean, median and standard deviation) of DSC from the top ranking methods. Rank 7 was also submitted by the authors but using a slightly different combination, i.e., probabilistic kNN and locally-weighted label fusion.

Rank	Team	Femoral cartilage			Tibial cartilage		
		Mean	Median	STD	Mean	Median	STD
1	Imorphics [27]	0.861	0.869	0.065	0.865	0.888	0.054
2	ZIB [20]	0.845	0.856	0.058	0.850	0.858	0.049
3	UPMC_IBML	0.836	0.838	0.028	0.805	0.807	0.057
4	SNU_SPL	0.821	0.838	0.059	0.824	0.841	0.058
5	**shan_unc (proposed)**	**0.856**	**0.862**	**0.057**	**0.859**	**0.861**	**0.047**
6	UIiibiKnee [31]	0.824	0.842	0.067	0.825	0.834	0.056
7	shan_unc	0.828	0.836	0.060	0.820	0.826	0.051
8	BioMedIA [28]	0.840	0.854	0.062	0.836	0.841	0.048

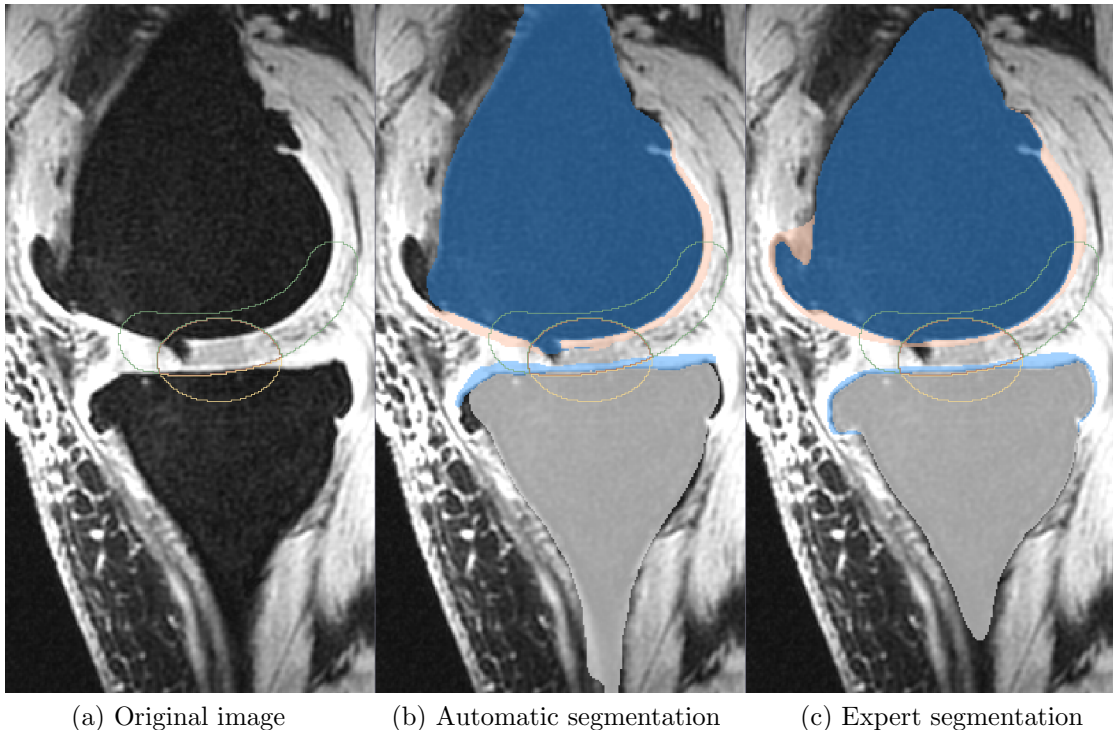(a) Original image   (b) Automatic segmentation   (c) Expert segmentation

Figure 21: An example slice from SKI10 [14] training dataset. (a) is the original image. (b) and (c) are automatic and expert segmentations, respectively. Femur: dark blue, tibia: light grey, femoral cartilage: pink, tibial cartilage: light blue. Yellow contour: validation region for the femoral cartilage. Green contour: validation for the tibial cartilage. Red contour: validation region for both cartilage. A cartilage lesion is present in the femoral cartilage shown in the weight-bearing region (touching region) in the original image. Our segmentation successfully delineates it, but the expert segmentation fails to do so.

The methods that have not been tested on SKI10 [14] dataset are not directly comparable to our method because of different datasets. Note that our method compares favorably to other methods, however, none of the competing methods was validated on datasets as large as ours (with more than 700 images for the PLS data alone). For example, Folkesson *et al.* [11] tested on 139 images, Fripp *et al.* [12] 20 images, Tamez-Peña *et al.* used [26] 12 images and Yin *et al.* [32] 60 images. Hence, our validation dataset is an order of magnitude larger than for most other existing studies.

11. Conclusion and Future Work

In this paper, we proposed a fully-automatic cartilage segmentation approach. We used a multi-atlas-based bone segmentation to guide the registration of a cartilage atlas. We investigated cartilage segmentation using an average shape atlas or multiple atlases with various label fusion techniques to obtain spatial cartilage priors within a novel three-label segmentation framework which incorporates anisotropic regularization to improve segmentation performance for the thin femoral and tibial cartilage layers.

We demonstrated that the multi-atlas-based segmentation strategy is appropriate for cartilage segmentation and performs as well as the top-ranking methods on the SKI10 dataset. The proposed method is robust, because multi-atlas-based methods can overcome occasional registration failures. This is a critical aspect when moving toward the analysis of larger datasets, such as the OAI dataset. We also demonstrated that the multi-atlas-based segmentation with non-local patch-based label fusion performs better than other label fusion strategies for cartilage segmentation.

The proposed three-label segmentation framework is novel, general and guarantees separation of the cartilage layers. The anisotropic regularization is a customization for cartilage segmentation but has general applicability for the segmentation of thin objects. An important advantage of the segmentation framework is its convexity which guarantees that a globally optimal solution can be computed for given segmentation parameters.

The major drawback of our method is a typical disadvantage of multi-atlas-based methods, namely their high computational cost. To alleviate this problem, atlas selection heuristics have been proposed. These heuristics select only a subset of promising training subjects for atlas registration and label fusion [1]. Such a selection strategy can be integrated into our segmentation method and is expected to further increase segmentation performance. We will explore atlas selection for cartilage segmentation in our future work.

Most crucially, our future work will focus on studying cartilage thickness changes longitudinally.

Appendix A. Numerical solution of (4)

This section discusses an iterative scheme to optimize (4). Solving (3) is a special case with $\mathbf{G} = g\mathbf{I}$ ($\mathbf{I}$ is the identity matrix). We introduce two dual variables $\mathbf{p}$ (vector field) and q (scalar field) and rewrite (4) as

$$E(u, \mathbf{p}, q) = \int_{\mathcal{D}} \langle \mathbf{p}, \mathbf{G}\nabla_{\mathbf{x}}u \rangle + q\nabla_l u \; d\mathbf{x}dl,$$
$$\text{subject to } \|\mathbf{p}\| \leq 1, \; |q| \leq c, \qquad \text{(A.1)}$$

in which $\langle \cdot, \cdot \rangle$ represents inner products. Minimizing (4) with respect to u is equivalent to minimizing (A.1) with respect to u and maximizing it with respect to $\mathbf{p}$ and q. The gradient descent/ascent update scheme of (A.1) is

$$\mathbf{p}_t = -\mathbf{G}\nabla_x u, \; \|\mathbf{p}\| \leq 1 \qquad \text{(A.2)}$$

$$q_t = -\nabla_l u, \; |q| \leq c \qquad \text{(A.3)}$$

$$u_t = -\operatorname{div}_{\mathbf{x}}(\mathbf{G}\mathbf{p}) - \nabla_l q \qquad \text{(A.4)}$$

The iterative scheme will lead to a global optimum upon convergence [2] because of the convexity of (4). Let S and T denote the source and sink sets: $S = \Omega \times \{0\}$, $T = \Omega \times \{3\}$. The region without sources or sinks is denoted as $\mathring{\mathcal{D}} = \mathcal{D} \setminus (S \cup T) = \Omega \times \{1, 2\}$. The dual energy is

$$E^*(u) = \int_S \operatorname{div}(\mathbf{G}\mathbf{p}) + \nabla_l q \; d\mathbf{x}dl$$
$$+ \int_{\mathring{\mathcal{D}}} \min(0, \operatorname{div}(\mathbf{G}\mathbf{p}) + \nabla_l q) \; d\mathbf{x}dl \qquad \text{(A.5)}$$

We terminate the iterations when the duality gap between the primal energy (4) and the dual energy (A.5) is sufficiently small. After convergence, the solution u is essentially binary and monotonically increasing. The three-label segmentation can be easily recovered from the discontinuity set of u.

Appendix B. Acknowledgments

References

[1] Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46, 726–738.

[2] Appleton, B., Talbot, H., 2006. Globally mimimal surfaces by continuous maximal flows. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 106–118.

[3] Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–197.

[5] Coupé, P., Manjḳn, J., Fonov, V., Prusessner, J., Robles, M., Collins, D., 2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 59, 3736–3747.

[6] Dice, L., 1945. Measures of the amount of ecologic association between species. Ecology , 297–302.

[7] Dodin, P., Pelletier, J.P., Martel-Pelletier, J., Abram, F., 2010. Automatic human knee cartilage segmentation from 3-d magnetic resonance images. IEEE Transactions on Biomedical Engineering 57, 2699–2711.

[8] Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification (second edition). Wiley-Interscience.

[9] Eckstein, F., Cicuttini, F., Raynauld, J.P., Waterton, J.C., Peterfy, C., 2006. Magnetic resonance imaging (mri) of articular cartilage in knee osteoarthritis (oa): morphological assessment. Osteoarthritis and Cartilage 14, A46–A75.

[10] Eckstein, F., Schnier, M., Haubner, M., et al., 1998. Accuracy of carilage volume and thicknesss measurements with magnetic resonance imaging. Clinical Orthopaedics and Related Research 352, 137–148.

[11] Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P.C., Christiansen, C., 2007. Segmenting articular cartilage automatically using a voxel classification approach. IEEE Transactions on Medical Imaging 26, 106–115.

[12] Fripp, J., Crozier, S., Warfield, S.K., Ourselin, S., 2010. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. IEEE Transaction on Medical Imaging 29, 21–27.

[13] Glocker, B., Komodakis, N., Paragios, N., Glaser, C., Tziritas, G., Navab, N., 2007. Primal/dual linear programming and statistical atlases for cartilage segmentation. Medical Image Computing and Computer-Assisted Intervention LNCS 4792, 536–543.

[14] Heimann, T., Morrison, B., Styner, M., Niethammer, M., Warfield, S., 2010. Segmentation of knee images: a grand challenge. Proc. MICCAI Workshop on Medical Image Analysis for the Clinic , 207–214.

[15] Išgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B., 2009. Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans. IEEE Transactions on Medical Imaging 28, 1000–1010.

[16] Kellgren, J., Lawrence, J., 1957. Radiological assessment of osteoarthritis. Annals of Rheumatic Diseases 16, 494–502.

[17] Koo, S., Hargreaves, B.A., Gold, G.E., . Automatic segmentation of articular cartilage from MRI. Patent, US 2009/0306496 .

[18] Rohlfing, T., Brandt, R., Menzel, R., C. R. Maurer, J., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21, 1428–1442.

[19] Rousseau, F., Habas, P., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. IEEE Transactions on Medical Imaging 30, 1852–1862.

[20] Seim, H., Kainmueller, D., Lamecker, H., Bindernagel, M., 2010. Model-based auto-segmentation of knee bones and cartilage in mri data. Medical Image Analysis for the Clinic: A Grand Challenge , 215–223.

[21] Sethian, J.A., 1999. Level set methods and fast marching methods. Cambridge Press.

[22] Shan, L., Charles, C., Niethammer, M., 2012a. Automatic atlas-based three-label cartilage segmentation from MR knee images. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis , 241–246.

[23] Shan, L., Charles, C., Niethammer, M., 2012b. Automatic multi-atlas-based cartilage segmentation from knee MR images. IEEE International Sysmposium on Biomedical Imaging , 1028–1031.

[24] Shan, L., Zach, C., Niethammer, M., 2010. Automatic three-label bone segmentation from knee MR images. IEEE International Symposium on Biomedical Imaging , 1325–1328.

[25] Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity non-uniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97.

[26] Tamez-Peña, J., Farber, J., González, P.C., Schrever, E., Schneider, E., Totterman, S., 2012. Unsupervised segmentation and quantification of anatomical knee features: data from the osteoarthritis initiative. IEEE Transactions on Biomedical Engineering 59, 1177–86.

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

[28] Wang, Z., Donoghue, C., Rueckert, D., 2013. Patch-based segmentation without registration: application to knee MRI. Machine Learning in Medical Imaging (MLMI) MICCAI Workshop .

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

[30] Yezzi, A.J., Prince, J.L., 2003. An Eulerian PDE approach for computing tissue thickness. IEEE Transactions on Medical Imaging 22, 1332–1339.

[31] Yin, Y., Williams, R., Anderson, D.D., Sonka, M., 2010a. Hierarchical decision framework with a priori shape models for knee joint cartilage segmentation - MICCAI grand challenge. Medical Image Analysis for the Clinic: A Grand Challenge , 241–250.

[32] Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D.D., Sonka, M., 2010b. Logismos-layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. IEEE Transaction on Medical Imaging 29, 2023–2037.

[33] Zach, C., Niethammer, M., Frahm, J.M., 2009. Continuous maximal flows and Wulff shapes: application to MRFs. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 1911–1918.

[34] Zhang, K., Lu, W., 2011. Automatic human knee cartilage segmentation from multi-contrast mr images using extreme learning machines and discriminative random fields. MLMI'11 Proceedings of the Second International Conference on Machine Learning in Medical Imaging , 335–343.