Magnetic Resonance Imaging 31 (2013) 1731–1743

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.mrijournal.com

Automatic knee cartilage segmentation from multi-contrast MR images using support vector machine classification with spatial dependencies

Kunlei Zhang [a,*], Wenmiao Lu [b], Pina Marziliano [a]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore
[b] Beckman Institute, University of Illinois, Urbana-Champaign, IL 61822, USA

ARTICLE INFO

Article history:
Received 17 December 2012
Revised 28 May 2013
Accepted 10 June 2013

Keywords:
Magnetic resonance imaging (MRI)
Multi-contrast Segmentation
Knee cartilage
Discriminative random field (DRF)
Support vector machine (SVM)

ABSTRACT

Accurate segmentation of knee cartilage is required to obtain quantitative cartilage measurements, which is crucial for the assessment of knee pathology caused by musculoskeletal diseases or sudden injuries. This paper presents an automatic knee cartilage segmentation technique which exploits a rich set of image features from multi-contrast magnetic resonance (MR) images and the spatial dependencies between neighbouring voxels. The image features and the spatial dependencies are modelled into a support vector machine (SVM)-based *association potential* and a discriminative random field (DRF)-based *interaction potential*. Subsequently, both potentials are incorporated into an inference graphical model such that the knee cartilage segmentation is cast into an optimal labelling problem which can be efficiently solved by loopy belief propagation. The effectiveness of the proposed technique is validated on a database of multi-contrast MR images. The experimental results show that using diverse forms of image and anatomical structure information as the features are helpful in improving the segmentation, and the joint SVM-DRF model is superior to the classification models based solely on DRF or SVM in terms of accuracy when the same features are used. The developed segmentation technique achieves good performance compared with gold standard segmentations and obtained higher average *DSC* values than the state-of-the-art automatic cartilage segmentation studies.

1. Introduction

Musculoskeletal diseases and articular disorders are one of the major health problems in causing work disability. In particular, the human knee joint is commonly affected by acute injury or osteoarthritis (OA) which is mainly characterized by the articular cartilage pathology. Quantitative cartilage measurements such as the thickness, volume and surface area, are required to properly assess the damage to cartilage. Accurate segmentation of the knee cartilage is the key to obtain these measurements. Because a segmentation system has been the essential task for the cartilage assessment, this study is focused on developing cartilage segmentation methods. A recent review of algorithms for medical image segmentation can be found in [1].

Magnetic resonance (MR) imaging is the most effective imaging modality to detect anatomical changes in knee joints for its excellent tissue contrasts. Fig. 1 illustrates a two-dimensional (2D) sagittal fat suppressed (FS) spoiled gradient recall (SPGR) MR image of the knee joint which is composed of three bones (patella, femur and tibia), three corresponding cartilage compartments (patellar cartilage, femoral cartilage and tibial cartilage), and other tissues including muscle, ligament, tendon, menisci and adipose tissue, etc. MR imaging has commonly been used for knee cartilage segmentation and assessment in the literature [2].

In general, fully manual and semi-automatic cartilage segmentation approaches [3–15] for routine clinical use are laborious and time consuming. Moreover, the intra/inter-observer variability complicates the interpretation of the results. On the other hand, automatic cartilage segmentation from MR images is a challenging task due to the thin variable morphology of cartilages, low contrast between cartilages and other soft tissues, MR artifacts, and intensity inhomogeneity (see Fig. 1). In this work, we address the problem of automatic knee cartilage segmentation using a classification approach, in which we utilize a rich set of image features provided by multiple spectral MR images taken with different sequences (referred to as *multi-contrast MR images*[1]) and incorporate the image features with the spatial dependences between neighbouring voxels in an inference graphical model.

1.1. Prior works on automatic segmentation of knee cartilage

In recent years, a number of automatic cartilage segmentation works have been reported. Most of them used atlas/template registration, optimal graph, statistical or deformable model-based

* Corresponding author.
E-mail address: kunleizhang@gmail.com (K. Zhang).

[1] They are also called *multi-spectral* or multi-modal MR images in some works.

0730-725X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mri.2013.06.005

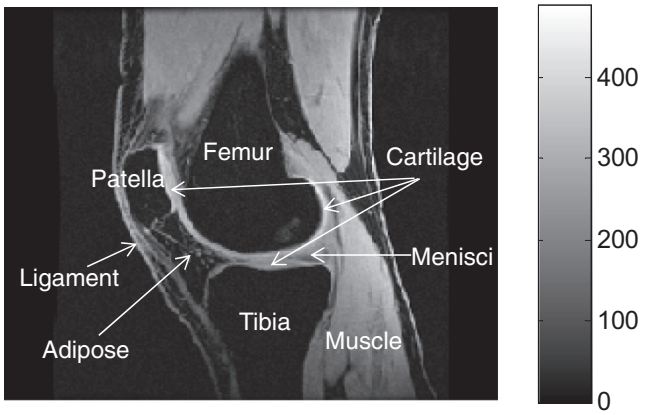

Fig. 1. An illustration of the knee MR imaging: one 2D slice from sagittal FS SPGR MR sequence with anatomical structures annotated.

approaches which typically require a large training dataset to get the prior knowledge, e.g., shape prior, atlas. Glocker et al. [16] segmented only the patellar cartilage by fitting a statistical atlas to MR data in a nonrigid registration scheme. Fripp et al. [17] developed a hierarchical segmentation scheme which consists of the automatic bone segmentation using a three-dimensional (3D) active shape model, the bone-cartilage interface (BCI) extraction, and cartilage segmentation from the BCI using a hybrid deformable model. Dodin et al. [18] presented a hierarchical automatic segmentation algorithm for OA knee cartilage volume quantification. Yin et al. [19] used a LOGISMOS (layered optimal graph image segmentation of multiple objects and surfaces) framework which simultaneously segments all the bone and cartilage surfaces from volumetric MR images. Lee et al. [20] presented a hierarchical scheme for automatic knee segmentation from MR images which was composed of bone segmentation using a modified branch-and-mincut algorithm, BCI detection using a voxel-based binary classification of local appearance and position, and the segmentation of cartilages based on the optimization of estimated local shape and appearance probabilities in localized Markov random fields. At the 2010 MICCAI conference, a competition for knee segmentation was held at a workshop “Medical Image Analysis for the Clinic – A Grand Challenge" [21]. Model-based methods such as the works in [22,23] were ranked the highest in terms of accuracy. Vincent et al. [22] presented an automatic system for segmenting bone and cartilage based on active appearance models with a minimum description length groupwise image registration method for generating high quality correspondences. In Seim et al. [23], the bone surfaces are first reconstructed based on statistical shape models and graph-based optimization, and the cartilages are then segmented with a multi object technique using prior knowledge on the variation of cartilage thickness.

Without formulating the prior knowledge, image segmentation can also be treated as a statistical classification problem in which each voxel belongs to a specific class. Folkesson et al.[24] employed a two step *k*-nearest neighbour (*k*-NN) voxel classifier to automatically separate cartilages from non-cartilages. Koo et al. [25] segmented cartilages automatically with multi-contrast MR images using support vector machines (SVM).However, these works assumed that data instances were independent and identically distributed (i.i.d.) in the classification, which may not be appropriate for the task of cartilage segmentation from MR images. As pointed out by Kumar and Hebert tect[26], class labels are not independent in most real-world spatial classification problems where correlations in labels of neighbouring instances exist in data with multi-dimensional structure, such as images and volumetric data.

Actually, most voxel labels strongly depend on their neighbours in knee MR images. This motivates us to incorporate contextual information in the form of spatial dependencies of both the MR images and the class labels in the classification for cartilage segmentation. Probabilistic graphical models such as conditional random field (CRF) [27] and discriminative random fields (DRF) [26] have been used to model spatial contextual correlations in many applications. As DRF is a powerful method for modelling interactions in both the image data and the labels, in this study we employ the DRF framework to model the spatial contextual dependencies, which has shown to yield more reliable results.

1.2. Overview of our work

In this study, we provide a novel solution to the problem of automatic knee cartilage segmentation with multi-contrast MR images using supervised classification with spatial dependencies. An automatic segmentation system specialized for the knee cartilage segmentation task is presented, which works on multi-contrast MR images and involves a preprocessing pipeline, the extraction of a rich set of features, the design of a classification model based on the joint SVM-DRF, and the parameter learning and inference on this model. Not only local image appearance but also (global) geometrical information of anatomical structures are exploited as features. The geometrical features are extracted based on the automatically segmented bones with the help of multi-contrast MR images. To enjoy the strong generalization properties of SVM as well as incorporate spatial dependencies between neighbouring voxels, we employ a joint SVM-DRF as classification model by constructing an *association potential* modelling on the decision function of SVM classifier and a DRF-based *interaction potential*. Subsequently, both potentials are incorporated into an inference graphical model such that the knee cartilage segmentation is cast into an optimal labelling problem. Although similar combination of SVM and DRF has been used for prostate cancer localization [28] and brain tumor extraction [29,30], there are significant differences between this study and these works: (1) we specialize it in our system for cartilage segmentation; (2) the work in [29] used DRF just in an independent stage to regularize the result produced by SVM, but not in an integrated way of DRF and SVM; (3) the formulation of *association potential* in our work is different from those in [28,30] where the estimate of *association potential* parameters are required first and then the *interaction potential* parameters are estimated, while the *association potential* parameters and the *interaction potential* parameters can be learned simultaneously; (4) we employ loopy belief propagation (LBP) for efficient inference.

Fig. 2 shows the overview of the presented segmentation system, whereby each part will be introduced in detail in the following sections. A brief version of the proposed segmentation scheme has been introduced in [31] with some preliminary results. This paper further strengthens our segmentation system by illustrating the multi-contrast MR imaging of the knee, describing deeply both the classification models and feature extraction, conducting the evaluation on comprehensive multi-contrast MR datasets, and analyzing statistical significance of segmentation results, etc.

The rest of this paper is organized as follows. Section 2 presents multi-contrast MR imaging of the knee, and provides the details of the multi-contrast MR datasets used in this work. Section 3 describes segmentation methodologies where the extraction of features is first introduced and then the classification models along with parameter learning and inference over the models are elaborated. Section 4 validates the proposed knee cartilage segmentation scheme and compares its performance to other methods. A discussion and summary of this work is given in Section 5.

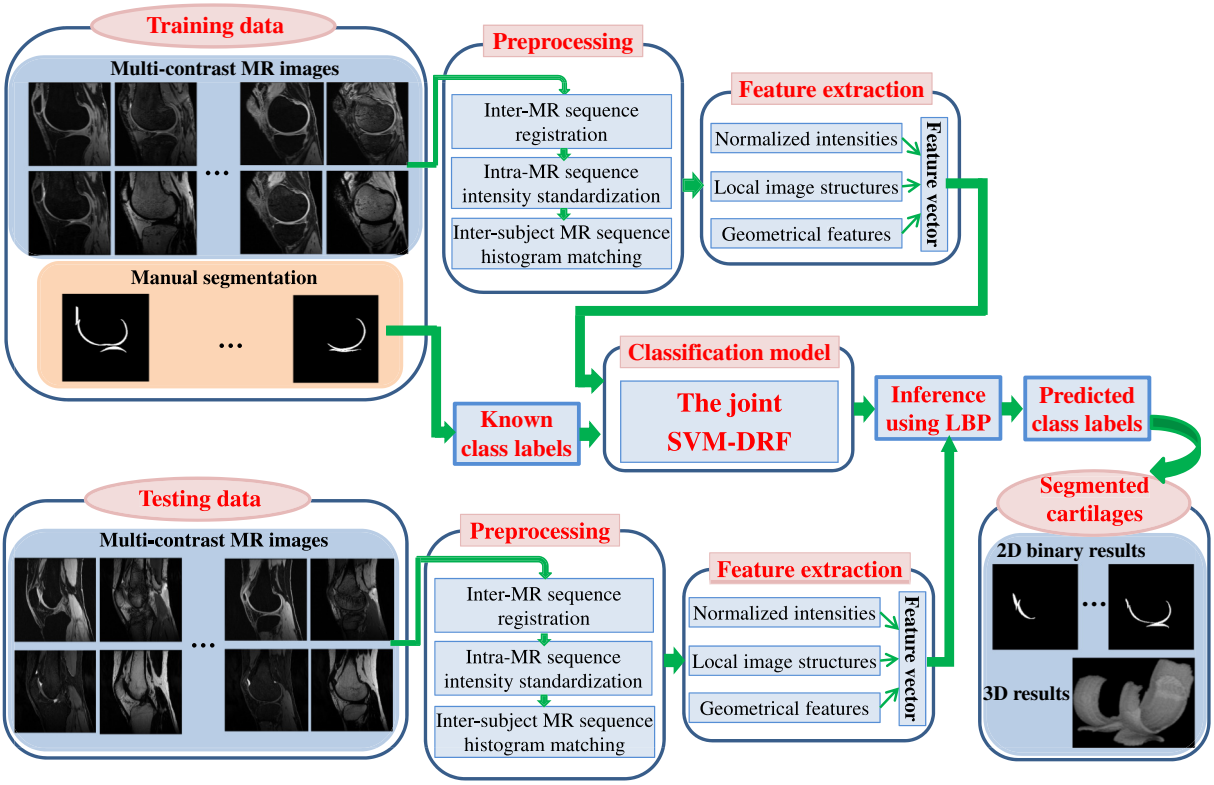

Fig. 2. The overview of the presented automatic cartilage segmentation system. First, multi-contrast MR knee images and corresponding manual cartilage segmentations are collected as training data. Second, the multi-contrast MR data undergo a preprocessing pipeline. Third, a set of features is extracted from multi-contrast MR data to form a feature vector for each voxel. Fourth, the unified SVM and DRF classification model is trained. Fifth, after the testing data undergo the same preprocessing and feature extraction as in the training stage, LBP algorithm is used for efficient inference of the optimal labels based on the trained classification model. Finally, cartilage segmentations are directly obtained from the predicted class labels.

2. Multi-contrast MR images of the knee

2.1. Multi-contrast MR imaging of the knee

Contrast resolution is the relative appearance of different anatomical structures in a medical image. In addition to hydrogen (water) density, many other mechanisms, e.g., relaxation time, result in contrast between various anatomical structures in MR imaging. Bones, cartilages and other structures in the knee joint possess different tissue properties. For example, they have various T1 (spin–lattice relaxation), T2 (spin-spin relaxation) time constants at a given magnetic strength. MR imaging of the knee takes advantage of these tissue properties to increase the contrast between different structures. The commonly used contrast mechanisms in MR imaging are T1-weighted, T2-weighted, T2/T1-weighted, and proton density imaging with or without fat suppression. An illustration of contrast mechanisms of knee MR imaging is given in Fig. 3. One can observe that it is difficult to separate different anatomical structures based on

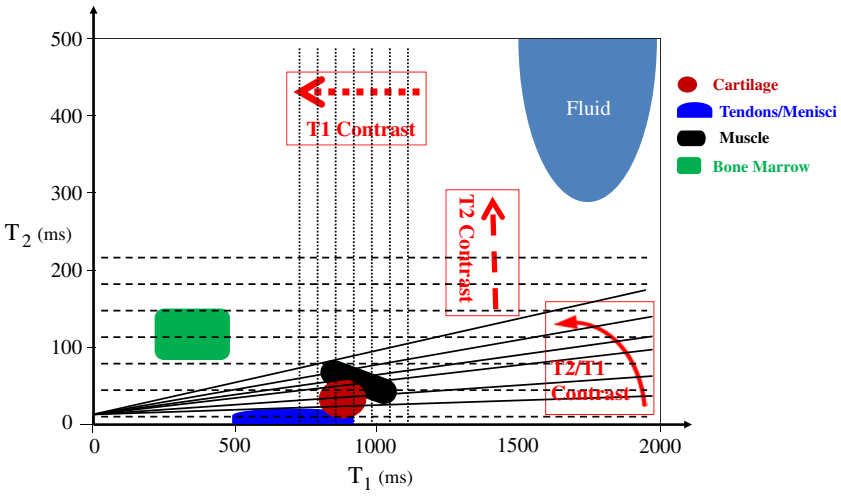

Fig. 3. An illustration of contrast mechanisms of knee MR imaging: cartilages and surrounding tissues with T1 contrast, T2 contrast and T2/T1 contrast, respectively.

single contrast mechanism. For example, considering the T1 contrast alone in Fig. 3, there is big overlapping between the cartilage, the muscle and the tendons/meniscus, which can also be observed in both the T2 contrast and T2/T1 contrast.

Multi-contrast MR images have the advantage of providing rich spectral correlation which produces information that spatial correlation cannot provide. These benefits cannot obtained from a single contrast MR sequence. Nowadays, there have been many MR sequences utilizing different contrast mechanisms which conduces toward different contrasts between knee joint structures. Excellent reviews on current MR imaging sequences of articular cartilage and the whole knee joint can be found in [32]. Fig. 4 illustrates examples of various contrast human knee MR sequences including T1-weighted FS SPGR [33], T2/T1-weighted fast imaging employing steady-state acquisition (FIESTA) [34], and T2/T1-weighted iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL) gradient echo (GRE) water and fat images [35].

The FS SPGR sequence is currently considered as the standard for quantitative morphology imaging of cartilage [32] since it provides good spatial resolution, contrast-to-noise ratio (CNR) and signal-to-noise ratio (SNR). In FSSPGR MR images, although there is high contrast between bone and cartilage, the contrast between cartilage and adjacent joint fluid is quite low, which is often not adequate to highlight the cartilage defects. On the other hand, the FIESTA sequences provide comparable or higher cartilage SNR efficiency as well as significantly higher cartilage-fluid contrast compared with SPGR sequences [34]. However, the fluid and fat both appear bright on FIESTA images, which may cause abnormalities to appear similar to normal fat and thus obscure underlying pathology. In addition, gradient-echo techniques can be combined with IDEAL, a water-fat separation method [35], to obtain IDEAL GRE sequence which has high SNR and yields excellent quality of water and fat images with high resolution. There are also many other MR imaging sequences for knee joint such as dual-echo steady-state (DESS), vastly interpolated projection reconstruction (VIPR),T2 relaxation mapping, etc., which are not included in this work. The reader is referred to [32] for a review of MR imaging of articular cartilage. Actually, as claimed in [32], there has yet to be established an ideal MR imaging sequence for knee cartilage which can provide an optimal combination of spatial resolution, CNR, minimal artifacts and SNR efficiency.

Compared with a single MR sequence, multi-contrast MR images provide more information on the contrast of different tissues in the knee and thus are helpful in achieving clearer separation of various anatomical structures. To illustrate this, Figs. 5 and 6 give the normalized intensity histograms of cartilages and the surrounding structures in the cases of both single contrast and multi-contrast MR images, respectively, based on the same MRI slice shown in Fig. 4. Note the high degree of overlap between these tissues' intensity distributions in Fig. 5A–D, which makes the desired joint structures, like bones and cartilages, difficult to separate from other joint structures based solely on the intensity from single contrast MR images. The usefulness of multi-contrast MR images in the separation of these structures can be observed from Fig. 6A–D, where the overlap between their intensity distributions is much less compared to the single contrast case in Fig. 5. This motivates the use of multi-contrast MR images for the task of knee segmentation in this study.

Nevertheless, there still is a certain amount of overlapping between these structures. For example, some parts of cartilages still overlap with muscle and other tissues. In other words, the intensity alone is insufficient to accurately extract cartilages from the knee joint. We thus need to exploit other image structure and anatomical information from multi-contrast MR images as features to get more accurate cartilage segmentations, and the details will be presented in Section 1.

2.2. Multi-contrast MRI data acquisition

In this study, we use a database of multi-contrast MR images acquired from 11 volunteers with unknown knee health conditions. Four MR sequences including T1-weighted FS SPGR, T2/T1-weighted FIESTA, T2/T1-weighted IDEAL GRE water and fat imaging, as

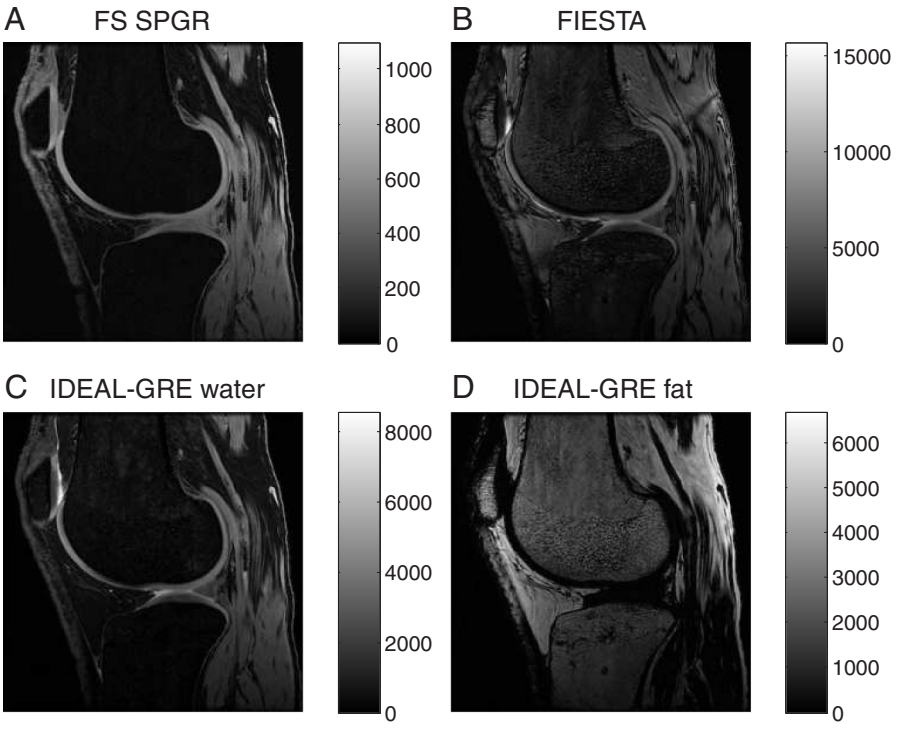

Fig. 4. One example 2D slice from multi-contrast human knee MR sequences where each sequence provides a unique mechanism to produce different contrasts between tissues.

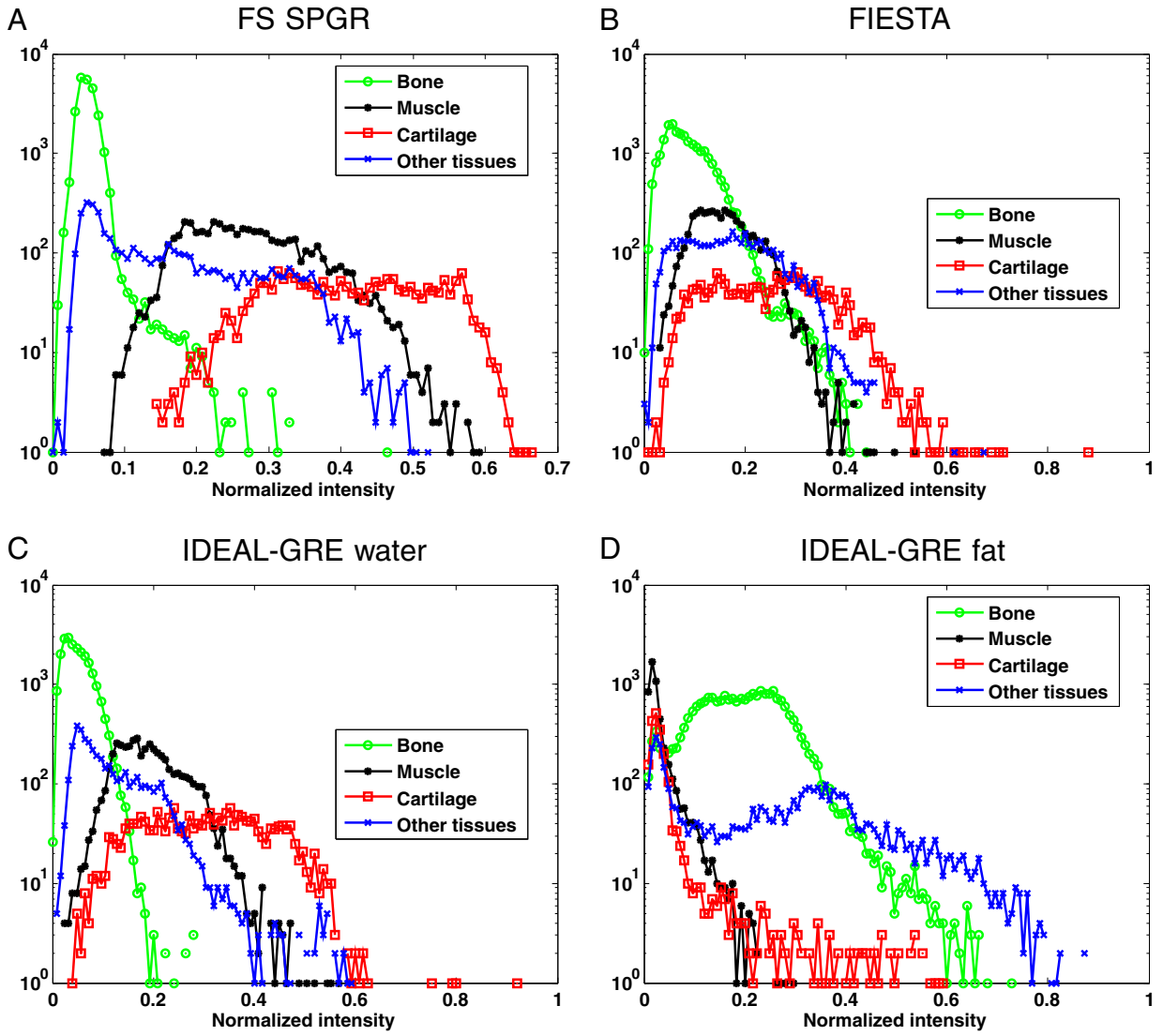

Fig. 5. Normalized intensity histograms of cartilages and the surrounding tissues in the case of single contrast MR sequence based on the same MRI slice shown in Fig. 4.

described in Section 1, were taken for each knee joint using a 3.0 T magnet scanner (GE Healthcare, Waukesha, WI, USA) within 30 minutes. Different MR sequences were acquired using different sets of parameters: FS SPGR sequences were obtained with echo time (TE) 4 ms, repetition time (TR) 28 ms, and flip angle (FA) 25°; FIESTA with TE 3.5 ms, TR 7.2 ms and FA 25°; and IDEAL-GRE (water and fat) with TE 3.4 ms, TR 9.4 ms and FA 30°. In addition, several parameters were common: all images were acquired in the sagittal plane with slice thickness 1.5 mm, in-plane spacing 0.625 mm × 0.625 mm and matrix size 256 × 56. It should be noted that the multi-contrast MR sequences are required to be homologous, that is, multiple MR sequences should be imaged in the same plane (say, the sagittal plane) and have the same slice number with each slice from each MR sequence at the same location in the knee.

Each knee was manually segmented by a medical expert on the FS SPGR sequence resulting in the form of binary mask images of the cartilage. The manual segmentations were later used as gold standard in our experiments.

3. Knee cartilage segmentation

In this section, we will elaborate the cartilage segmentation based on voxel classification approaches. Before performing segmentation, multi-contrast MR images need to be preprocessed as explained next.

3.1. Preprocessing

We perform a preprocessing pipeline on the knee MR datasets described in Section 2. First, the four MR sequences were spatially aligned using an automatic rigid multi-modal registration algorithm with mutual information metric in the Insight Segmentation and Registration Toolkit (ITK)[2]. Subsequently, intensity values of each MR sequence for all subjects were normalized to [0, 1], which brings them within the same dynamic range, and thus improves the stability of the classifier. In addition, we performed the image histogram matching across all subjects for each MR sequence separately to allow for inter-subject classification. All the preprocessing steps were implemented automatically without user interaction.

[2] National Library of Medicine Insight Segmentation and Registration Toolkit, http://www.itk.org.

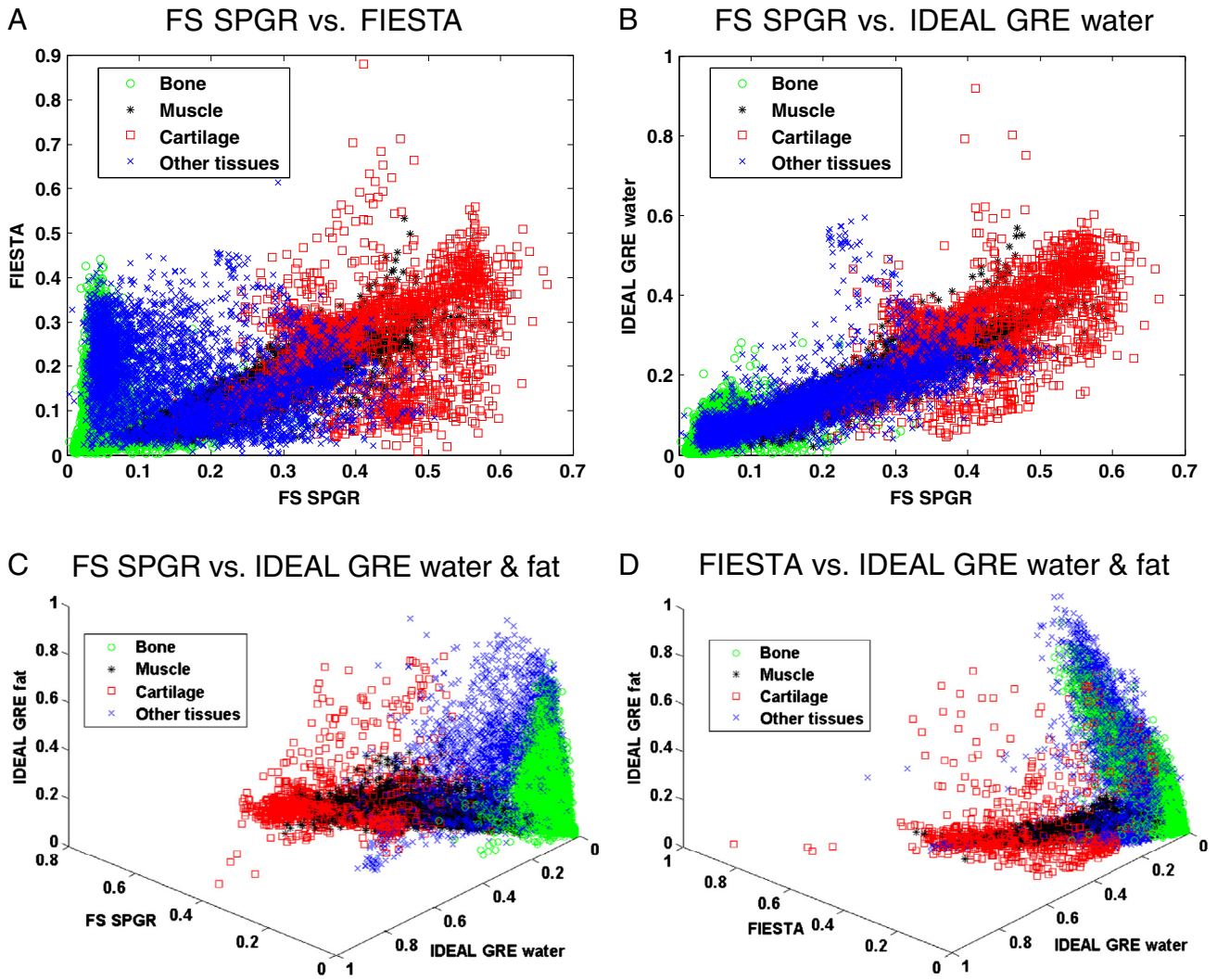

Fig. 6. Normalized intensity histograms of cartilages and the surrounding tissues in the case of multi-contrast MR images based on the same MRI slice shown in Fig. 4.

3.2. Segmentation approaches

Throughout this paper, let $\mathrm{x} = \{x_i\}_{i \in S}$ represent the observed data from a volumetric MR image, where the ith voxel is typically represented by the d-dimensional feature vector $\gamma_i(\mathbf{x})$, and S is the set of elements for the MR image. Let the corresponding labels of the image voxels be given by $\mathbf{y} = \{y_i\}_{i \in S}$, where $y_i \in \{-1,1\}$ is the class label (1 represents cartilage and -1 represents the others) of the ith voxel. The aim of the classification based segmentation is to infer the most likely joint class labels $\mathbf{y}^* = \{y_i^*\}_{i \in S}$ based on the observed MR image $\mathbf{x}$ via a trained classification model.

Next, we first introduce the features used in this work and then describe the classification models including DRF, SVM, the joint SVM-DRF along with parameter learning and inference over this model.

3.2.1. Feature extraction

As has been illustrated in Figs. 5 and 6, features based solely on intensity values are far from able to accurately extract cartilages. Hence, we also exploit features based on local image structures which contain local object appearance and shape information. Appendix A gives more details on the description of local image structure-based measurements. We choose to use gradients of Gaussian smoothed images which are obtained by convolving the MR images with Gaussian derivative at a proper scale (1 mm), and three principal components (eigenvalues) of the Hessian image. In spite of benefiting from multi-contrast MR images, local image features alone have limited effect on distinguishing cartilages another joint structures due to overlapping intensity distributions and ambiguous boundaries, and therefore they are still insufficient for accurate segmentation results.

Hence, we further extract the geometrical information of anatomical structures in the knee joint with the help of multi-contrast MR images. Geometrical features can be considered as a type of global shape feature which is crucial for object-oriented medical image segmentation. Inspired by the work of [25], bones in the knee images are automatically segmented in 3D from the multi-contrast MR sequences. Since bones are well-separated from other joint structures in the combination of FS SPGR and IDEAL-GRE water and fat sequences (see Fig. 6c), they can thus be robustly extracted by using a threshold-based method followed by connected-components labelling and distance transform. Fig. 7 shows an example segmentation surface of three main knee bones, i.e., the femur, tibia and patella.

Based on the segmented femur, posterior center (the grey box) and anterior center (the black box) of the lateral femoral condyle, posterior center (the grey circle) and anterior center (the black circle) of the medial femoral condyle are first detected by exploiting the coordinate information of voxels belonging to the femur. Then, we easily locate both the lateral femur center (the white box) which is the mean of posterior center and anterior center of the lateral femoral condyle,

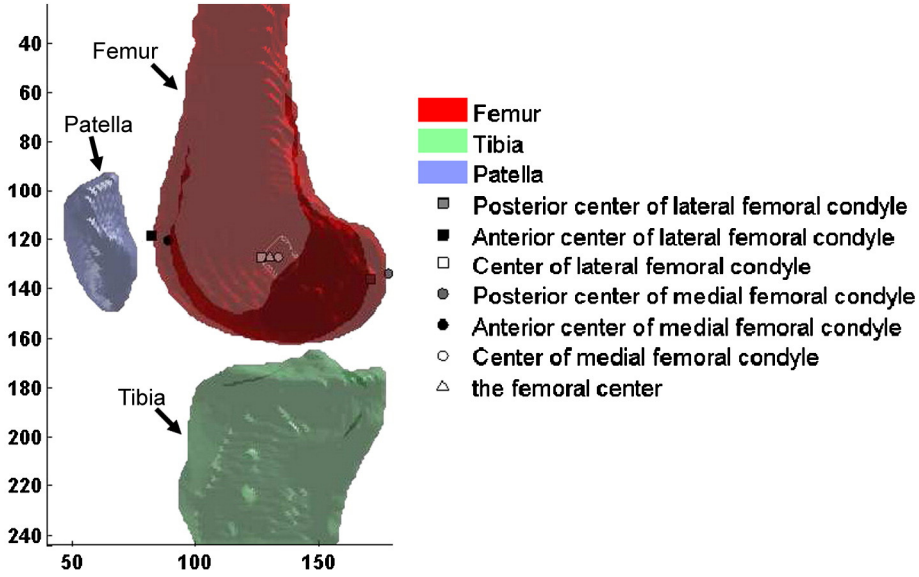

Fig. 7. Segmented bone surfaces and detected femoral bone features: an example on subject 9.

and medial femur center(the white circle) which is the mean of posterior center and anterior center of the medial femoral condyle. Finally, the femoral center (the white triangle) which is the mean of the lateral and medial femur centers is determined. Note all the centers in Fig. 7 are referred to the 3D center.

Subsequently, we calculate the geometrical features for each voxel. Relying on the segmented bone surfaces and the detected femoral features, for each voxel, the Euclidean distance (in 3D space) from the closest bone surface, the angle between the main magnetic direction of the MRI and the line that connects the voxel to the femoral center, and the relative location along the medial and lateral center of the distal femur are calculated as the geometrical features. These geometrical features help to extract cartilages from the other knee joint structures which have very similar appearance to the cartilage. For example, the distance from the closest bone is helpful in separating cartilage from muscle since cartilages are always adjacent to the end of the bones and muscle further away [25]. The feature "angle" is also useful in distinguishing different tissues since they have different locations and directions in the knee joint which result in different angle values.

Thus, the full set of features adopted in this paper consists of normalized intensity values and local image structure-based features from all four MR sequences, as well as geometrical features. Accordingly, each image voxel i is mapped to a 31-dimensional vector $\gamma_i(\mathbf{x})$ whose elements are composed of the aforementioned features. The experimental results show that all the features are helpful in improving segmentation accuracy of the cartilage. In particular, the use of geometrical features highly increased the accuracy of cartilage segmentation in this work as will be presented in Section 4. It should be noted that the three MR sequences, i.e., FS SPGR, IDEAL-GRE water and fat, are necessary in the proposed scheme to automatically segment the bones and subsequently extract the geometrical features to get accurate cartilage segmentation results.

3.2.2. Classification models

We now briefly describe the basics of DRF and SVM on the basis of which the joint SVM-DRF classification model is introduced.

3.2.2.1. Support vector machines. As a supervised machine learning algorithm, SVM [36] can be used as a discriminative classifier. In the SVM, the goal is to find an optimal separating hyperplane as the decision boundary in the feature space using the support vectors from the training set. Given a training set $\{(\gamma_i(\mathbf{x}), y_i)|i = 1, 2,\cdots, n\}$ where n is the number of training samples, the optimal α_i's can be learned through solving the following quadratic programming optimization problem represented in dual Lagrangian form

$$\max_{\alpha}\left[\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \kappa\left(\gamma_i(\mathbf{x}),\gamma_j(\mathbf{x})\right)\right] \quad \text{subject to } \sum_{i=1}^{n}\alpha_i y_i = 0 \tag{1}$$

where $0 \le \alpha_i \le C$ is the Lagrange multiplier, C is a positive constant that bounds the misclassification error, and κ: $R^d \times R^d \to R$ is the kernel function that maps the observations into a Hilbert space. The most commonly used kernel for classification is the Gaussian radial basis function (RBF)

$$\kappa\left(\mathbf{z}_i,\mathbf{z}_j\right) = exp\left(-\sigma||\mathbf{z}_i-\mathbf{z}_j||^2\right). \tag{2}$$

Then the decision function for a test data $\mathbf{x}'$ can be described as

$$f_{SVM}\left(\gamma_i\left(\mathbf{x}'\right)\right) = \sum_{t=1}^{T}\alpha_t y_t \kappa\left(\gamma_t(\mathbf{x}),\gamma_i\left(\mathbf{x}'\right)\right) + b \tag{3}$$

where the bias b is scalar which can also be learned, $\gamma_t(\mathbf{x})$ denotes a support vector which is the example closest to the separating hyperplane, and T is the number of the support vectors.

Although the powerfully discriminative ability and strong generalization property make SVM to be a frequently used algorithm in image classification tasks, it has obvious drawbacks. SVM makes the critical i.i.d. assumption on data, and especially it only considers dependencies between the features of a data instance and its class label but cannot incorporate spatial contextual interactions among neighbouring instances. This is not appropriate for applications like medical image segmentation since neighbouring data instances are dependent on each other.

3.2.2.2. Conditional and discriminative random fields. CRF is a discriminative approach which directly models the posterior distribution $P(\mathbf{y}|\mathbf{x})$ as a Gibbs field. CRF relaxes the strong conditional independence assumption of the observations and allows the modelling of complex dependencies (i) between the label of an instance and its features, (ii) between the labels of adjacent instances, and (iii) between the labels of adjacent instances and their features.

CRF was designed for labelling and segmenting sequential data, which assumes one-dimensional chain-structure. Accordingly, DRF was proposed by [26] as a multi-dimensional extension of CRF for lattice-structured data (Although the terminology "CRF or conditional random field" is often used in the literature even in the case of high-dimensional data, we particularly use "DRF or discriminative random field" in image-based applications.). In the DRF model, the joint distribution over the labels $\mathbf{y} = \{y_i\}_{i\in S}$ given the observations $\mathbf{x}$ can be written as

$$P\left(\mathbf{y}|\mathbf{x}\right) = \frac{1}{Z}\prod_{i\in S} A(y_i,\mathbf{x})\prod_{i\in S}\prod_{j\in N_i} I\left(y_i,y_j,\mathbf{x}\right) \tag{4}$$

where

$$Z = \sum_{\mathbf{y}}\left(\prod_{i\in S} A(y_i,\mathbf{x})\prod_{i\in S}\prod_{j\in N_i} I\left(y_i,y_j,\mathbf{x}\right)\right)$$

is a normalization constant called the partition function, $A(y_i, \mathbf{x})$ is the *association potential* modelling dependencies between the ith label y_i and the set of all (or a part of) observations $\mathbf{x}$, $I(y_i, y_j, \mathbf{x})$, is the *interaction potential* modelling dependencies between the labels y_i, y_j and the observations $\mathbf{x}$, and $\mathcal{N}_i$ stands for a neighbourhood of voxel i.

DRF is a powerful method for modelling dependencies in spatial data, but the *association potential* based on a logistic regression classifier has restricted classification capability and often cannot estimate appropriate parameters for the cases with unbalanced class labels, high-dimensional feature spaces, or highly correlated features [30]. Due to this, in some tasks DRF is not able to produce results as accurate as powerful classification models such as SVM.

3.2.2.3. The joint SVM-DRF model. Although SVM has strong generalization classification ability, it may produce undesirable results for the cartilage segmentation task because in SVM individual image voxels are assumed be to independent and interactions among adjacent voxels are not incorporated. Conversely, DRF considers these interactions but does not have the appealing generalization properties as SVM. To improve upon the cartilage segmentation based on the independent labellings by SVM classification, we use the DRF framework to incorporate spatial dependencies between neighbouring voxels in classification. Specifically, we adopt the following unified formulation of SVM and DRF as the classification model, and the $p(\mathbf{y}|\mathbf{x})$ is formulated as

$$P\left(\mathbf{y}|\mathbf{x}\right) = \frac{1}{Z}\prod_{i\in S} A\left(y_i,\hat{f}_{SVM}(\boldsymbol{\gamma}_i(\mathbf{x}))\right)\prod_{i\in S}\prod_{j\in N_i} I\left(y_i,y_j,\mathbf{x}\right) \tag{5}$$

In the formulation, the new *association potential* $A\left(y_i,\hat{f}_{SVM}(\boldsymbol{\gamma}_i(\mathbf{x}))\right)$ is obtained by converting the output of the SVM decision function $f_{SVM}(\cdot)$ in Eq. (3) to a posterior probability using the *sigmoid* function[3] as follows:

$$\begin{aligned} A\left(y_i,\hat{f}_{SVM}(\boldsymbol{\gamma}_i(\mathbf{x}))\right) &= \frac{1}{1+exp\left(-y_i\left(w_0+w_1\hat{f}_{SVM}(\boldsymbol{\gamma}_i(\mathbf{x}))\right)\right)} \\ &= \eta\left(y_i\mathbf{w}^T\mathbf{f}(\boldsymbol{\gamma}_i(\mathbf{x}))\right) \end{aligned} \tag{6}$$

where $\mathbf{w} = [w_0, w_1]^T$ is the vector of *association potential* parameters to be learned, and

$$\mathbf{f}(\boldsymbol{\gamma}_i(\mathbf{x})) = \left[1\hat{f}_{SVM}(\boldsymbol{\gamma}_i(\mathbf{x}))\right]^T.$$

[3] *Sigmoid* function is defined by the formula $(t) = \frac{1}{1+exp(-t)}$.

The *interaction potential*

$$I\left(y_i,y_j,\mathbf{x}\right) = exp\left(y_iy_j\mathbf{v}^T\boldsymbol{\gamma}_{ij}(\mathbf{x})\right) \tag{7}$$

has the same expression as in [26], which models of the interactions in neighbouring labels y_i, y_j and make them data-adaptive. In this study, we choose a ten-neighbourhood system, i.e., $\mathcal{N}_i = 10$ for each voxel i in a volumetric MR image, which defines a local dependency structure. The $\boldsymbol{\gamma}_{ij}(\mathbf{x})$ in Eq. (7) denotes the feature vector for a pair of neighbouring voxels i and j, which is computed from the observations $\mathbf{x}$. We set $\boldsymbol{\gamma}_{ij}(\mathbf{x})$ by taking the absolute difference of $\boldsymbol{\gamma}_i(\mathbf{x})$ and $\boldsymbol{\gamma}_j(\mathbf{x})$, that is, $\boldsymbol{\gamma}_{ij}(\mathbf{x}) = |gamma_i(\mathbf{x}) - \boldsymbol{\gamma}_j(\mathbf{x})|$ which penalizes for high absolute differences between the features of neighbouring voxels. This is reasonable for the knee segmentation task since neighbouring voxels with highly different features tend to have different class labels, e.g., the adjacent cartilage and bone voxels. The vector $\mathbf{v}$ is *interaction potential* parameter which can be learned in the training stage.

The use of a kernel SVM classifier to learn the *association potential* in Eq. (6) can dramatically strengthen the classification ability of the original DRF since DRF-based works like [30,37] showed that the discriminative power in the *association potential* is crucial to the overall performance of the DRF framework. Therefore, the joint SVM-DRF model not only enjoys the strongly generalized classification ability of SVM but also enables the incorporation of spatial dependencies between neighbouring voxels via the DRF model.

Although the joint SVM-DRF can apply to both 2D and 3D image, we build a 3D lattice graph structure for the DRF model in the implementation. Thus, not only the contextual dependencies in (2D) intra-slices but also the continuation information between slices (3D) can be incorporated into the joint SVM-DRF, which is expected to improve the results over the classification model based solely on SVM.

3.2.3. Learning

Given a training set $\{(\mathbf{x}^m, \mathbf{y}^m)|m = 1,\cdots,M|\}$ where M is the number of training images, the classification model in Eq. (5) can be learned by using a sequential strategy. We first solve the SVM quadratic programming problem in Eq. (1) based on the *LibSVM* implementation [38]. With the resulting SVM decision function $\hat{f}_{SVM}(\cdot)$, the parameters $\Theta = \{\mathbf{w}, \mathbf{v}\}$ are then simultaneously learned by maximizing the pseudolikelihood, i.e.,

$$\hat{\Theta} = \underset{\Theta}{argmax}\prod_{m=1}^{M}\prod_{i\in S} P\left(y_i^m\middle|y_{N_i}^m,\mathbf{x}^m,\mathbf{w},\mathbf{v}\right) \tag{8}$$

where y_i^m is the observed label for the ith voxel in the mth training image $\mathbf{x}^m$, and

$$P\left(y_i^m\middle|y_{N_i}^m,\mathbf{x}^m,\Theta\right) = \frac{1}{Z_i^m}\eta\left(y_i^m\mathbf{w}^T\mathbf{f}(\boldsymbol{\gamma}_i(\mathbf{x}^m))\right)\prod_{j\in N_i} exp\left(y_i^my_j^m\mathbf{v}^T\boldsymbol{\gamma}_{ij}(\mathbf{x}^m)\right)$$

with

$$Z_i^m = \sum_{y_i^m\in\{-1,1\}}\left[\eta\left(y_i^m\mathbf{w}^T\mathbf{f}(\boldsymbol{\gamma}_i(\mathbf{x}^m))\right)\prod_{j\in N_i} exp\left(y_i^my_j^m\mathbf{v}^T{}_{ij}(\mathbf{x}^m)\right)\right].$$

We then have the negative logarithm pseudolikelihood

$$L = \sum_{m=1}^{M}\sum_{i\in S}\left[log(Z_i^m) - log\left(\eta\left(y_i^m\mathbf{w}^T\mathbf{f}(\boldsymbol{\gamma}_i(\mathbf{x}^m))\right)\right) - \sum_{j\in N_i} y_i^my_j^m\mathbf{v}^T\boldsymbol{\gamma}_{ij}(\mathbf{x}^m)\right]. \tag{9}$$

Inspired by the works in [26,28], to ensure that the negative logarithm pseudolikelihood in Eq. (9) is convex and to prevent over-

fitting, we further utilize the L_2 -regularization, which involves adding a penalty term in the form of a sum of squares of all the parameters in order to discourage the parameters from reaching large values. Hence, the parameters can be estimated by minimizing the following *penalized* negative logarithm pseudolikelihood

$$\hat{\theta} = \underset{\theta}{argmin}\left[L + \frac{\lambda_1}{2}\mathbf{w}^T\mathbf{w} + \frac{\lambda_2}{2}\mathbf{v}^T\mathbf{v}\right] \tag{10}$$

where λ_1 and λ_2 are nonnegative regularizing constants which can be determined by a cross-validation[4] method.

Similar to [26], the penalized logarithm pseudo-likelihood in Eq. (10) can be easily minimized using gradient descent method since it is jointly convex with respect to the parameters θ. We herein adopt a quasi-Newton strategy where limited-memory BFGS updates are used in computing the step direction, and a bracketing line-search is used to find a step length satisfying the strong Wolfe conditions [40].

3.2.4. Inference

Given a test knee with multi-contrast MR images $\mathbf{x}'$, we first perform the same preprocessing and feature extraction as in the training stage. Then, the inference problem is to find an optimal label configuration $\mathbf{y}^*$ based on the learned classification model and its parameters^via solving a maximum a posteriori (MAP) problem:

$$\mathbf{y}^* = \underset{\mathbf{y}}{argmax}\ P\left(\mathbf{y}\middle|\mathbf{x}, \hat{\theta}\right). \tag{11}$$

In this study, we employ the sum-product loopy belief propagation (LBP) inference algorithm [41] to efficiently solve this problem. LBP is a message passing algorithm which performs approximate inference on general graphical models. It calculates the marginal distribution for each unobserved node (voxel), conditional on the observed nodes. Starting from some initial set of belief propagation messages, we iterate through all the voxels and repeatedly apply the belief propagation updates to the messages. The label configuration is obtained by computing the maximum of the marginals. In the implementation, we utilize the software of UGM (Matlab code for undirected graphical models) provided by Dr. Mark Schmidt[5].

3.2.5. Multi-class classification for femoral, tibial and patellar cartilage segmentation

Our goal is to segment the three cartilage compartments: femoral cartilage, tibial cartilage and patellar cartilage. We employ a one-versus-all [42] strategy to directly combine three two-class classifiers (one binary classifier trained to separate femoral cartilage from the rest, one binary classifier trained to separate tibial cartilage from the rest, and one binary classifier trained to separate patellar cartilage from the rest) into a multi-class classifier for the segmentation of individual cartilage compartments. Each binary classifier is trained in a way described in Section 3. Both of the qualitative and quantitative segmentation results for femoral, tibial and patellar cartilage will be presented in Section 2.

4. Experiments and results

In this section, we evaluate our cartilage segmentation scheme on multi-contrast MR datasets described in Section 2.

[4] See the details on cross-validation method from [39] and http://en.wikipedia.org/wiki/Cross-validation_(statistics).

[5] UGM (2011), http://www.di.ens.fr/mschmidt/Software/UGM.html.

4.1. Experiment setup

4.1.1. Training and testing strategy

We conduct the experiments in a leave-one-out cross-validation strategy which involves using a single subject's data as the test data, and the remaining subjects' data as the training data. This is repeated such that each subject is used once as the test data.

4.1.2. Parameter tuning

In our experiments, the popular Gaussian RBF defined in Eq. (2) is used as the kernel in SVM. It is known that the performance of SVM is sensitive to the two parameters: the regularization parameter C and the Gaussian RBF's variance σ. In order to achieve good generalization performance and seek for an optimal classification accuracy, both C and σ need to be chosen appropriately. As in [38], we employ a grid-search on tC and σ in the range of $\{2^{-20},2^{-18},\cdots,2^{18},2^{20}\}$ using a 5-fold cross-validation method. In more details, we use a two-step grid-search strategy: (1) a coarse grid-search to identify a small candidate region on the original large grid, and (2) a fine grid-search on the small region. This can save substantial time compared with conducting a complete grid-search directly on the original grid. The pair of (C,σ) with the highest cross-validation accuracy is picked. Similarly, λ_1 and λ_2 in Eq. (10) are chosen from the range $\{10^{-10},10^{-8},\cdots,10^{8},10^{10}\}$ by performing a grid-search using a 5-fold cross-validation method.

4.1.3. Evaluation metrics

Both the gold standard segmentations and the automatic segmentations are provided visually in the forms of 2D slices, and their differences can thus be directly perceived. In addition to visual evaluation, the cartilage segmentations automatically obtained are quantitatively compared to the gold standard using *Sensitivity*, *Specificity* and *Dice similarity coefficient* (*DSC*) measures, which are usually used in cartilage segmentation works [16-19,24]. Let *TP* denote true positive, *TN* denote true negative, *FP* denote false positive, and *FN* denote false negative counts for the voxels, then the $Sensitivity = TP/(TP + FN)$ is the true positive fraction, $Specificity = TN/(FP + TN)$ is the true negative fraction and $DSC = 2\text{xtitTP}/(2TP + FP + FN)$ is a spatial overlap index. *DSC* ranges between 0 and 1, where 1 signifies perfect overlap, and 0 means no overlap. Moreover, a paired *t*-test with a *p*-value is used to measure the statistical significance of these quantitative results. In statistical significance testing, the *p*-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed. It is generally considered to be statistical significance when the *p*-value is less than a significance level, which is often set at 0.05.

4.1.4. Compared methods

The effectiveness of different features is first compared based on the joint SVM-DRF classification model. Then, we compare the results obtained by the joint SVM-DRF with those based solely on SVM or DRF.

4.2. Results

4.2.1. Comparison of features

To validate the performance improvements through the use of multi-contrast MR images and different types of features, five different feature vectors for each voxel are constructed as follows:

- Feature vector 1 (FV1) consists of 4-dimensional normalized intensity values of multi-contrast MR images (from all four MR sequences);
- Feature vector 2 (FV2) consists of 1-dimensional normalized intensity values of single contrast MR images from the FS SPGR sequence and 6-dimensional local image structure based features;

Table 1
Five types of feature vectors.

Feature vectors	Dimensions	Features			
		NISC[a]	*NIMC*[b]	*LISF*[c]	*GF*[d]
FV1	4		√[e]		
FV2	7	√		√	
FV3	28		√	√	
FV4	7		√		√
FV5	31		√	√	√

[a] NISC denotes the normalized intensity values of single contrast MR images.
[b] NIMC denotes the normalized intensity values of multi-contrast MR images.
[c] LISF denotes local image structure based features.
[d] GF denotes geometrical features.
[e] √ denotes this type of feature is contained in the corresponding feature vector.

- Feature vector 3 (FV3) consists of 4-dimensional normalized intensity values of multi-contrast MR images and 24-dimensional local image structure based features;
- Feature vector 4 (FV4) consists of 4-dimensional normalized intensity values of multi-contrast MR images and 3-dimensional geometrical information of multi-contrast MR images;
- Feature vector 5 (FV5) consists of 4-dimensional normalized intensity values of multi-contrast MR images, 24-dimensional local image structure based features, and 3-dimensional geometrical features of multi-contrast MR images.

The details of all five feature vectors are provided in Table 1.

Fig. 8 depicts segmentation results for all three cartilage compartments using the joint SVM-DRF model with FV1, FV2, FV3, FV4 and FV5, respectively, on examples slices from several subjects. These example segmentation results allow us to visually observe the benefits of using multi-contrast MR images as well as the use of features based on local image appearance and geometrical information. Although segmentation results obtained using FV1 give high *TP* (red zones) value, extitFP (cyan zones) value is also very high. This means over-segmentation is caused and indicates that intensity alone is not sufficient to separate cartilages from the surroundings due to overlapping intensity distributions. Through comparing the segmentation results obtained using FV2 and FV3, the effect of employing multi-contrast MR images can be observed since the results obtained using FV3 give lower *FN* (yellow zones) value than those using FV2. One can easily see that the results obtained using FV4 are much better than those using FV3, which thus demonstrates the key role of geometrical information in locating cartilages in the MR images. The importance of local image features on improving the segmentation performance can also be observed by comparing the segmentation results obtained using FV4 and FV5, because the results obtained using FV5 give slightly lower *FN* value than those using FV4 while both of them have similar *FP* value.

Table 2
Mean ± standard deviation of *Sensitivity*, *Specificity* and *DSC* on all subjects using the joint SVM-DRF model with FV1, FV2, FV3, FV4 and FV5, respectively.

SVM-DRF with	Cartilages	Quantitative measurements		
		Sensitivity	Specificity	DSC
FV1	Femoral	0.117 ± 0.187	0.997 ± 0.002	0.200 ± 0.184
	Tibial	0.009 ± 0.170	1.000 ± 0.000	0.019 ± 0.167
	Patellar	0.525 ± 0.162	0.995 ± 0.005	0.651 ± 0.155
FV2	Femoral	0.330 ± 0.144	0.993 ± 0.006	0.451 ± 0.142
	Tibial	0.378 ± 0.129	0.994 ± 0.006	0.477 ± 0.134
	Patellar	0.867 ± 0.130	0.253 ± 0.010	0.308 ± 0.126
FV3	Femoral	0.581 ± 0.133	0.993 ± 0.005	0.678 ± 0.128
	Tibial	0.521 ± 0.095	0.992 ± 0.006	0.514 ± 0.114
	Patellar	0.355 ± 0.124	0.992 ± 0.006	0.336 ± 0.116
FV4	Femoral	0.780 ± 0.117	0.994 ± 0.004	0.825 ± 0.105
	Tibial	0.777 ± 0.124	0.996 ± 0.003	0.784 ± 0.119
	Patellar	0.800 ± 0.115	0.991 ± 0.007	0.812 ± 0.096
FV5	Femoral	0.826 ± 0.108	0.996 ± 0.002	**0.864 ± 0.087**
	Tibial	0.860 ± 0.122	0.995 ± 0.004	**0.880 ± 0.102**
	Patellar	0.819 ± 0.110	0.997 ± 0.002	**0.841 ± 0.074**

Table 2 gives the average quantitative measurements on all subjects for femoral, tibial and patellar cartilage using the joint SVM-DRF model with FV1, FV2, FV3, FV4 and FV5, respectively. It can be noted that results obtained using FV5 give the highest average *Sensitivity* and *DSC* values. Furthermore, *p*-values from Table 3 show that the differences in the average *DSC* between FV5 and other four feature vectors are statistically significant since all *p*-values are much less than 0.05. It should be noted that *p*-values are computed based on the overall *DSC* values for the combined three cartilage

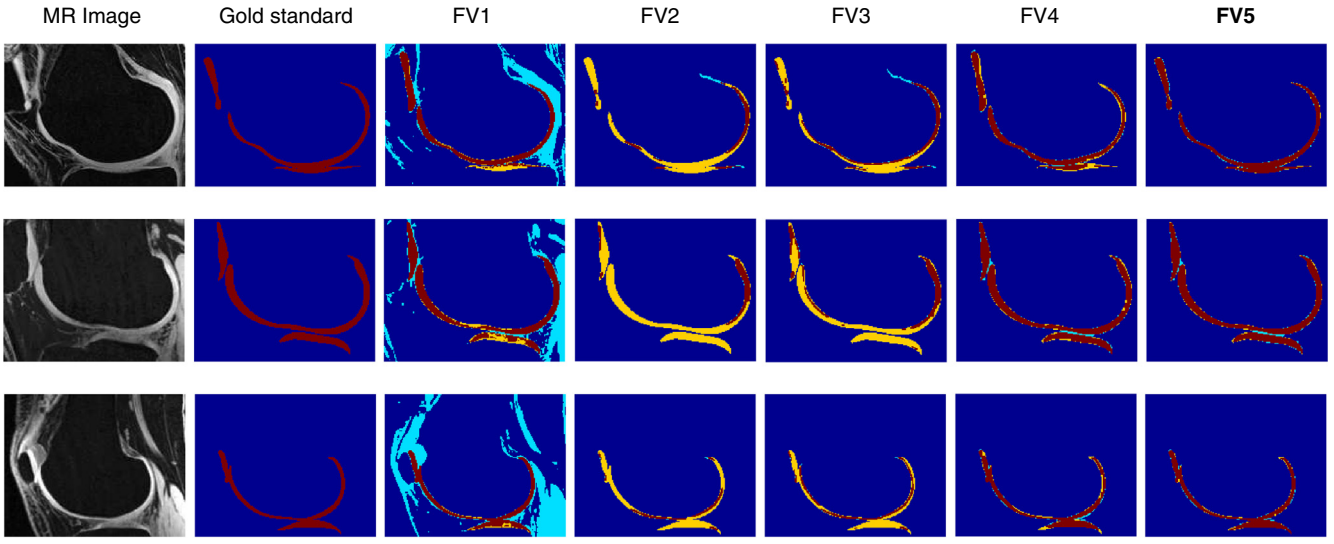

Fig. 8. Segmentation results for example slices using the joint SVM-DRF model with five types of feature vectors: (column 1) The MR images of example slices from different knees, (column 2) the gold standard, (column 3) segmentation using FV1, (column 4) segmentation using FV2, (column 5) segmentation using FV3, (column 6) segmentation using FV4, (column 7) segmentation using FV5. Red zones are *TP*, cyan zones are *FP*, yellow zones are *FN*, and blue zones are *TN* counts for the voxels.

Table 3
p-Values on *DSC* for the overall three cartilage compartments using the joint SVM-DRF with FV1, FV2, FV3, FV4, FV5.

SVM-DRF with	FV5 vs. FV4	FV5 vs. FV3	FV5 vs. FV2	FV5 vs. FV1
p-Value	$p < 10^{-2}$	$p < 10^{-10}$	$p < 10^{-10}$	$p < 10^{-30}$

compartments. Therefore, we will conduct the remaining experiments using FV5 only.

4.2.2. Comparison of classification models

We now compare the segmentation results obtained by the three classification models: the joint SVM-DRF, SVM, DRF, when using FV5. Fig. 9 illustrates the segmentation results for the same slices shown in Fig. 8. The segmentations using SVM show scattered *FN* and *FP* values, since SVM assumes that the image voxels are i.i.d and ignores the spatial contextual interactions. The segmentations using DRF give dense as well as high *FP* and *FN* values, since DRF incorporates spatial dependencies but lacks the generalization ability. The joint SVM-DRF model provides relatively low *FP* and *FN* values, since it not only enjoys the generalized classification ability of SVM but also incorporates the spatial correlations among neighbouring image voxels via the DRF framework.

To compare the segmentation results obtained by using the three models quantitatively, Table 4 provides average quantitative measurements on all subjects for femoral, tibial and patellar cartilage, respectively. Also, the *p*-values of the *DSC* in Table 5 show that the advantage of combining SVM with DRF as opposed to using merely DRF or SVM as the classification model is statistically significant.

Although the parameter tuning in the training phase is time-consuming, it can be done offline. Once the presented system is trained, it can be used to automatically segment cartilages from new datasets (which should undergo the same preprocessing and feature extraction as in the training stage). We ran the experiments on a 48-core high performance computer with Linux system. In the testing phase, the computational time for SVM classification is approximately 10 minutes, and the computational time for DRF inference is around 1 minute, for each cartilage compartment. The computation load mainly depends on the size of the MR volume.

Table 4
Mean ± standard deviation of *Sensitivity*, *Specificity*, and *DSC* of all knees using DRF, SVM, and the joint SVM-DRF, respectively, with FV5.

Methods	Cartilages	Quantitative measurements		
		Sensitivity	Specificity	DSC
DRF	Femoral	0.739 ± 0.100	0.957 ± 0.008	0.551 ± 0.087
	Tibial	0.644 ± 0.101	0.935 ± 0.009	0.456 ± 0.095
	Patellar	0.770 ± 0.081	0.893 ± 0.012	0.674 ± 0.068
SVM	Femoral	0.777 ± 0.114	0.993 ± 0.007	0.812 ± 0.116
	Tibial	0.815 ± 0.131	0.994 ± 0.004	0.824 ± 0.125
	Patellar	0.781 ± 0.117	0.994 ± 0.005	0.810 ± 0.106
SVM-DRF	Femoral	0.826 ± 0.108	0.996 ± 0.002	0.864 ± 0.087
	Tibial	0.860 ± 0.122	0.995 ± 0.004	0.880 ± 0.102
	Patellar	0.819 ± 0.110	0.997 ± 0.002	0.841 ± 0.074

In summary, the visual and quantitative results presented above show that our approach is able to provide promising cartilage segmentations. It can also be observed that the joint SVM-DRF model for classification is superior to the individual DRF or SVM, and that using multi-contrast MR images as well as diverse forms of image and anatomical structure information as the features are helpful in improving the accuracy of cartilage segmentation.

5. Discussion and conclusion

As reviewed in Section 1, a number of automatic cartilage segmentation works [16-19,22,23] have been reported in recent years by using atlas registration, optimal graph, statistical or deformable model-based approaches which typically require a large training dataset to get the particular prior knowledge, e.g., atlas/template, shape prior. Most of these approaches highly rely on the datasets used and thus it is not trivial to reproduce them. Unlike atlas/template, shape prior, statistical or deformable model based segmentation methods which typically formulate prior information and require training datasets from both healthy subjects and patients with OA, our approach does not rely on such prior information related to the subject demographics and pathological state of the knee. Instead, the spatial dependences between neighbouring voxels and a rich set of features are exploited from the multi-contrast MR images to achieve a more robust cartilage segmentation. Once the joint SVM-DRF model is

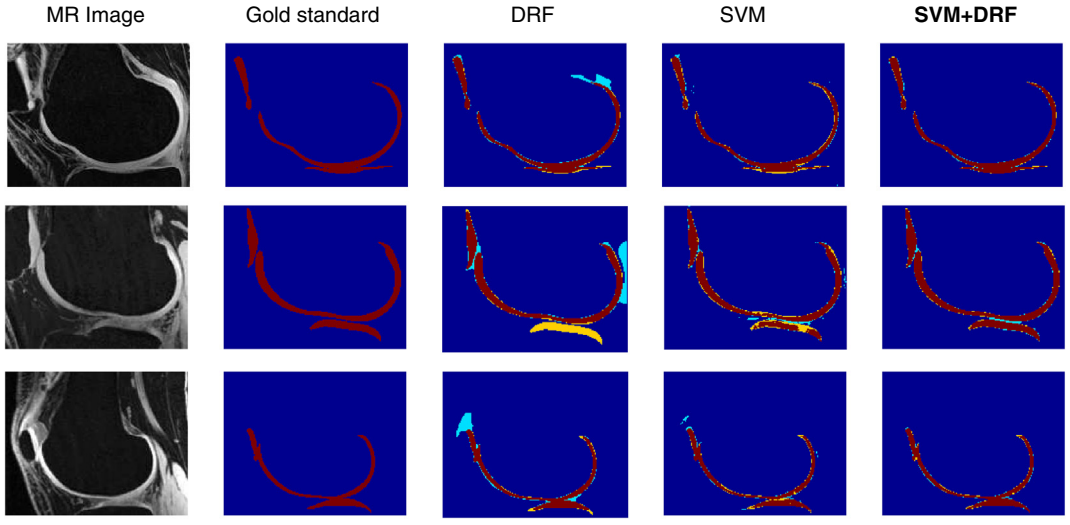

Fig. 9. Segmentation results for example slices using DRF, SVM and the joint SVM-DRF, respectively, with FV5: (column 1) The MR images of example slices from different subjects, (column 2) the gold standard, (column 3) segmentation using DRF, (column 4) segmentation using SVM, (column 5) segmentation using the joint SVM-DRF. Red zones are *TP*, cyan zones are *FP*, yellow zones are *FN*, and blue zones are *TN* counts for the voxels.

Table 5
p-Values of the *DSC* using DRF, SVM, and the joint SVM-DRF, with FV5.

FV5 with	SVM + DRF vs. SVM	SVM + DRF vs. DRF	SVM vs. DRF
p-Value	$p < 10^{-2}$	$p < 10^{-10}$	$p < 10^{-10}$

trained, it can be used to automatically segment cartilages on any new datasets from healthy or pathological subjects.

Table 6 shows the comparison of different automatic methods for cartilage segmentation. It can be observed that the presented technique in this study outperforms the state-of-the-art works [16,17,19,20,24] for automatic segmentation of patellar, femoral and tibial cartilages in terms of the average *DSC* value. Nevertheless, a more direct and reasonable comparison of these works is actually difficult since the characteristics of approaches and study population differ from one study to another.

Notably, multi-contrast MR data in the proposed segmentation system currently must include the three MR sequences, i.e., FS SPGR, IDEAL-GRE water and fat, to automatically segment the bones and subsequently extract the geometrical features to get accurate cartilage segmentation results. Additional choices of MR sequences with different contrast mechanisms can be FIESTA, DESS, T2 relaxation time mapping, T1ρ mapping, etc.

As stated in [17] that "there is still significant clinical and research interest in the segmentation of healthy cartilage tissue", that is the motivation that the proposed scheme has only been validated on nonpathological knees. The capability of the proposed scheme to be further extended to pathological knees is promising. Also, the use of a larger database as training is believed to yield a more accurate segmentation of cartilages.

In conclusion, we have presented an automatic cartilage segmentation system which works on the multi-contrast MR images. The proposed technique employs the SVM classification with spatial dependencies between neighbouring voxels via a DRF framework. Each voxel is represented as a high-dimensional feature vector consisting of normalized intensity values, features based on local image structure and geometrical features exploited from multi-contrast MR images; subsequently, each voxel is classified by solving an optimal labelling problem in an inference graphical model which has an SVM-based *association potential* and a DRF-based *interaction potential*. The advantages of our approach are achieved via effective incorporation of both the SVM-based classification with a rich feature set and the spatial contextual interactions in both the image data and the class labels via the DRF mode. Using a comprehensive multi-contrast knee database, we demonstrated that the combination of SVM and DRF for cartilage segmentation outperforms the commonly used classification models based solely on DRF or SVM in terms of accuracy, and that using diverse forms of image and anatomical structure information as the features are helpful in improving the segmentation.

Table 6
Comparison of different automatic methods for knee cartilage segmentation.

Works	Methods	Average DSC for		
		Patellar	Femoral	Tibial
Glocker et al. [16]	Atlas registration	0.84	–	–
Folkesson et al. [24]	*k*-NN classification	–	0.77	0.81
Fripp et al. [17]	Hybrid deformable model	0.83	0.85	0.82
Yin et al. [19]	Layered optimal graph	0.80	0.84	0.80
Lee et al. [20]	Optimization of local shape and appearance	0.82	0.82	0.81
The proposed	SVM ± DRF classification	0.84	0.86	0.88

Acknowledgments

The authors are thankful Dr. Seungbum Koo and the volunteers from Stanford University for help in acquiring multi-contrast MR scans using a 3 T MR system in the Richard M. Lucas Center. K. Zhang is also grateful to the financial support of China Scholarship Council (CSC).

Appendix A. Features based on local image structure

Local structures of an image are able to capture local object appearance and shape information. To exploit features based on local image structures, we consider a Taylor expansion of a volumetric image $\Gamma(\mathbf{x})$ around the voxel $\mathbf{x}_0$:

$$\Gamma(\mathbf{x}) \approx \Gamma(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)^T \nabla\Gamma(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \mathbf{H}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) \tag{12}$$

where $\nabla\Gamma = [\Gamma_x, \Gamma_y, \Gamma_z]^T$ denotes the gradient vector, the Hessian is defined by

$$\mathbf{H} = \begin{bmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{bmatrix}, \tag{13}$$

and $\{\Gamma_s(\mathbf{x}) | s \in \{x,y,z,xx,xy,xz,yy,yz,zz\}\}$ denotes the first- and second-order derivatives in x, y and z. Consequently, the local image structures up to second order can be described by the intensity, the gradient, and the Hessian.

We first calculate the $\Gamma_s(\mathbf{x})$ by convolving the image $\Gamma(\mathbf{x})$ with Gaussian derivatives at a proper scale τ,

$$\Gamma_s(\mathbf{x}) = \Gamma(\mathbf{x}) \otimes D_s\{g(\tau)\} \tag{14}$$

where $\otimes$ and $D\{\cdot\}$ denote the convolution and differential operator, respectively, and $g(\tau)$ denotes a Gaussian with the scale τ. The Hessian $\mathbf{H}$ can then be obtained according to Eq. (A.2). The three eigenvalues κ_1, κ_2 and κ_3 of a Hessian image can characterize the resemblance of the local structure to a tube ($|\kappa_1| \ll |\kappa_2| \leq |\kappa_3|$), a sheet ($|\kappa_1| \leq |\kappa_2| \ll |\kappa_3|$) and a blob ($|\kappa_1| \leq |\kappa_2| \leq |\kappa_3|$) [43]. Knee cartilage can locally be described as a thin disc, which corresponds to finding positions with one large and two small eigenvalues of the Hessian. Therefore, besides the image derivatives, the three eigenvalues of the Hessian are also incorporated as features in this work.

References

[1] Ma Z, Tavares JMR, Jorge RN, Mascarenhas T. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. Comput Methods Biomech Biomed Engin 2010;13(2):235–46.
[2] Eckstein F, Cicuttini F, Raynauld J, Waterton J, Peterfy C. Magnetic resonance imaging (MRI) of articular cartilage in knee osteoarthritis (OA): morphological assessment. Osteoarthritis Cartilage 2006;14:46–75.
[3] Gougoutas A, Wheaton A, Borthakur A, Shapiro E, Kneeland J, Udupa J, et al. Cartilage volume quantification via live wire segmentation. Acad Radiol 2004;11(12):1389–95.
[4] Grau V, Mewes A, Alcaniz M, Kikinis R, Warfield S. Improved watershed transform for medical image segmentation using prior information. IEEE Trans Med Imaging 2004;23(4):447–58.
[5] Jaremko J, Cheng R, Lambert R, Habib A, Ronsky J. Reliability of an efficient MRI-based method for estimation of knee cartilage volume using surface registration. Osteoarthritis Cartilage 2006;14(9):914–22.
[6] Kapur T, Beardsley P, Gibson S, Grimson W, Wells W. Model-based segmentation of clinical knee MRI. IEEE International Workshop on Model-Based 3D Image, Analysis; 1998. p. 97–106.
[7] König L, Groher M, Keil A, Glaser C, Reiser M, Navab N. Semi-automatic segmentation of the patellar cartilage in MRI. Computer science, Bildverarbeitung für die Medizin (BVM); 2007. p. 404–8.
[8] Lynch J, Zaim S, Zhao J, Stork A, Peterfy C, Genant H. Cartilage segmentation of 3D MRI scans of the osteoarthritic knee combining user knowledge and active contours. Proc SPIE Med Imaging: Image Process 2000;3979:925–35.

[9] Pakina S, Tamez-Pena J, Tottermanc S, Parkerd K. Segmentation, surface extraction and thickness computation of articular cartilage, in: Proc. SPIE Med Imaging: Image Process 2002;4684:155–66.
[10] Pirnog C. Articular cartilage segmentation and tracking in sequential MR images of the knee, Ph.D. thesis, ETH Zurich (2005).
[11] Shim H, Chang S, Tao C, Wang J, Kwoh C, Bae K. Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the semi-automated graph-cut algorithm method. Radiology 2009;251(2):548–56.
[12] Solloway S, Hutchinson C, Waterton J, Taylor C. The use of active shape models for making thickness measurements of articular cartilage from MR images. Magn Reson Med 1997;37(6):943–52.
[13] Stammberger T, Eckstein F, Michaelis M, Englmeier K, Reiser M. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. Magn Reson Imaging 1999;17(7):1033–42.
[14] Tamez-Pena J, Barbu-McInnis M, Totterman S. Knee cartilage extraction and bone-cartilage interface analysis from 3D MRI data sets, in: Proc. SPIE Med Imaging: Image Process 2004;5370:1774–84.
[15] Warfield S, Kaus M, Jolesz F, Kikinis R. Adaptive, template moderated, spatially varying statistical classification. Med Image Anal 2000;4(1):43–55.
[16] Glocker B, Komodakis N, Paragios N, Glaser C, Tziritas G, Navab N. Primal/dual linear programming and statistical atlases for cartilage segmentation. Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2007. p. 536–43.
[17] Fripp J, Crozier S, Warfield S, Ourselin S. Automatic segmentation and quantitative analysis of the articular cartilages from magnetic resonance images of the knee. IEEE Trans Med Imaging 2010;29(1):55–64.
[18] Dodin P, Pelletier J, Martel-Pelletier J, Abram F. Automatic human knee cartilage segmentation from 3-D magnetic resonance images. IEEE Trans Biomed Eng 2010;57(11):2699–711.
[19] Yin Y, Zhang X, Williams R, Wu X, Anderson D, Sonka M. LOGISMOS-layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. IEEE Trans Med Imaging 2010;29(12):2023–37.
[20] Lee S, Park S, Shim H, Yun I, Lee S. Optimization of local shape and appearance probabilities for segmentation of knee cartilage in 3-D MR images. Computer Vision and Image Understanding 2011;115(12):1710–20.
[21] Heimann T, Morrison B, Styner M, Niethammer M, Warfield S. Segmentation of knee images: a grand challenge. Proc Medical Image Analysis for the Clinic: A Grand Challenge, in conjunction with MICCAI 2010; 2010. p. 207–14.
[22] Vincent G, Wolstenholme C, Scott I, Bowes M. Fully automatic segmentation of the knee joint using active appearance models, in: Proc. Medical Image Analysis for the Clinic: A Grand Challenge, in conjunction with MICCAI 2010; 2010. p. 224–30.
[23] Seim H, Kainmueller D, Lamecker H, Bindernagel M, Malinowski J, Zachow S. Model-based auto-segmentation of knee bones and cartilage in MRI data. Proc. Medical Image Analysis for the Clinic: A Grand Challenge, in conjunction with MICCAI 2010; 2010. p. 215–23.
[24] Folkesson J, Dam E, Olsen O, Pettersen P, Christiansen C. Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans Med Imaging 2007;26(1):106–15.
[25] Koo S, Hargreaves B, Gold G. Automatic segmentation of articular cartilage from MRI, U.S. Patent 20,090/306,496 (Dec. 2009).
[26] Kumar S, Hebert M. Discriminative random fields. Int J Comput Vis 2006;68(2):179–201.
[27] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proc. International Conference on Machine Learning; 2001. p. 282–9.
[28] Artan Y, Haider M, Langer D, van der Kwast T, Evans A, Yang Y, et al. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. IEEE Trans Image Process 2010;19(9):2444–55.
[29] Bauer S, Nolte L, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization, Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2011. p. 354–61.
[30] Lee C, Schmidt M, Murtha A, Bistritz A, Sander J, Greiner R. Segmenting brain tumors with conditional random fields and support vector machines. Proc. Computer Vision for Biomedical Image Applications; 2005. p. 469–78.
[31] Zhang K, Deng J, Lu W. Segmenting human knee cartilage automatically from multi-contrast MR images using support vector machines and discriminative random fields. Proc. IEEE International Conference on Image Processing; 2011. p. 733–6.
[32] Gold G, Chen C, Koo S, Hargreaves B, Bangerter N. Recent advances in MRI of articular cartilage. Am J Roentgenol 2009;193(3):628–38.
[33] Disler D, McCauley T, Kelman C, Fuchs M, Ratner L, Wirth C, et al. Fat-suppressed three-dimensional spoiled gradient-echo MR imaging of hyaline cartilage defects in the knee: comparison with standard MR imaging and arthroscopy. Am J Roentgenol 1996;167(1):127–32.
[34] Hargreaves B, Gold G, Beaulieu C, Vasanawala S, Nishimura D, Pauly J. Comparison of new sequences for high-resolution cartilage imaging. Magn Reson Med 2003;49(4):700–9.
[35] Reeder S, Wen Z, Yu H, Pineda A, Gold G, Markl M, et al. Multicoil Dixon chemical species separation with an iterative least-squares estimation method. Magn Reson Med 2004;51(1):35–45.
[36] Vapnik V, Vapnik V. Statistical learning theory. New York (NY): Wiley; 1998.
[37] Hoiem D, Rother C, Winn J. 3d layoutcrf for multi-view object class recognition and segmentation, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition; 2007. p. 1–8.
[38] Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Tech 2011;27(2):1–27 software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.
[39] Geisser S. Predictive inference: An introduction, Vol. 55. New York (NY): Chapman & Hall/CRC; 1993.
[40] Heath MT. Scientific Computing: an introductory survey. McGraw-Hill; 2002.
[41] Murphy K, Weiss Y, Jordan M. Loopy belief propagation for approximate inference: an empirical study. Proc. of Uncertainty in AI; 1999. p. 467–75.
[42] Rifkin R, Klautau A, et al. In defense of one-vs-all classification. J Mach Learn Res 2004;5:101–41.
[43] Sato Y, Westin C, Bhalerao A, Nakajima S, Shiraga S, Tamura M, et al. Tissue classification based on 3D local intensity structures for volume rendering. IEEE Trans Vis Comput Graph 2000;6(2):160–80.